Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 991 (2003) 47–59
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Net analyte signal as a deconvolution-oriented resolution criterion in the optimisation of chromatographic techniques

G. Vivó-Truyols, J.R. Torres-Lapasió*, M.C. García-Alvarez-Coque
Departament de Química Analítica, Facultat de Química, Universitat de València, 46100 Burjassot, València, Spain

Received 28 May 2002; received in revised form 30 December 2002; accepted 9 January 2003

Abstract

The performance of two multivariate calibration measurements, multivariate selectivity (SEL_s) and scalar net analyte signal (scalar NAS), as chromatographic objective functions (COFs), was investigated. Since both assessments are straightforwardly related to the quantification of analytes in the presence of interferents, they were expected to confer new features in the optimisation of compound resolution, not present in conventional assessments. These capabilities are especially interesting in situations of low resolution, where peak deconvolution becomes an attractive alternative. For comparison purposes, chromatographic resolution (R_s) and peak purity (p_s) were used as reference COFs. In order to correlate COFs with the probability of deconvolution error, an artificial peak crossing was used to generate 73 different peak arrangements, which were deconvolved using three different methods. SEL_s exhibited the best correlation, which allowed predicting properly the risk of obtaining inaccurate deconvolutions. The optimisation of a poorly resolved mixture of 16 aromatic compounds by reversed-phase liquid chromatography with methanol–water and acetonitrile–water mobile phases was examined to investigate the differences in performance among the resolution criteria. In situations like these, SEL_s tends to consider acceptable mobile phase compositions with partial coelution, which permits however the deconvolution with low errors. In contrast, p_s selects compositions where the resolution of some compounds is sacrificed to enhance the separation of others. Scalar NAS was not so favourable as expected, since it depends on sampling frequency and peak widening. SEL_s was not affected by these factors.

Keywords: Optimisation; Objective function; Resolution; Peak deconvolution; Net analyte signal

1. Introduction

Computer-assisted interpretive optimisations are becoming very popular in chromatography [1–3]. These approaches are based on the possibility of predicting the chromatogram of a given mixture at varying experimental factors (e.g., mobile phase composition, pH, or temperature). The goal of the optimisation process is to maximise an objective function, ideally perfectly correlated with the overall peak separation. Besides, the chromatographic objective function (COF) may collect additional aims, more or less subjective, such as short analysis times or desirable peak shapes (i.e., high efficiencies and low asymmetries). A COF that takes into account simultaneously all these features is however difficult to obtain. Some proposals are found in the literature, such as those suggested by Berridge [4], Schoenmak-

*Corresponding author. Tel.: +34-96-354-3003; fax: +34-96-354-4436.
E-mail address: jose.r.torres@uv.es (J.R. Torres-Lapasió).

0021-9673/03/$ – see front matter
doi:10.1016/S0021-9673(03)00172-9

ers [5], Watson and Carr [6], or Glajch et al. [7], which are based on the measurement of chromatographic resolution, R_s. The use of R_s presents several limitations. For instance, it is indefinitely sensitive to peak separation (even when no overlap exists), which obligates to introduce some artefacts, such as weights or truncations. Moreover, it is not a single-peak criterion, which constitutes a strong restriction when the goal is not the separation of all the components of the sample, but the resolution of only some particular compounds [8], or an optimisation based on complementary mobile phases [9]. Examples of COFs based on other peak properties different of R_s are those based on the information theory [10], peak purity [8,9,11], or diverse valley-to-peak ratio measurements [11]. All these COFs try to forecast the chromatographic conditions offering the best separation.

The limited selectivity of chromatographic techniques often leads to situations where partial peak overlap remains in the whole studied experimental domain. In cases like these, chemometric methods can aid the chromatographer to retrieve the contributions of the individual compounds present in each peak cluster. Solute contributions to the gross chromatographic signal are thus obtained by mathematical treatments, which complement the partial separation achieved by the chromatographic process. The most common family of chemometric techniques to perform such mathematical treatments in chromatography is that based on the accommodation of the raw signal to a forced elution profile, using peak models. Deconvolution of chromatographic peaks is an essential strategy for separation modes where obtaining acceptable selectivity is troublesome, as is the case of chiral chromatography [12]. The recent concern in the chromatographic literature about research in new peak models [13–18], and deconvolution strategies [19–21], constitutes a symptom of the growing interest of such approaches.

The difficulty of deconvolution problems can be assessed with multivariate figures of merit, which are useful tools to evaluate the performance of analytical methods [22]. Net analyte signal (NAS) is likely the most important concept related to these figures of merit. NAS can be defined as the part of the signal of a given analyte free of interference from other compounds [23,24].

Up to now, COFs have been mainly oriented to assess the separation degree under a more or less subjective chromatographic perspective, and indeed they work properly in situations of complete separation. However, in problems where even the best chromatograms are deficiently resolved—which is not uncommon in practice—conventional COFs will not be necessarily well correlated with the probability of obtaining an unbiased deconvolution.

NAS—and its derived figures of merit—have been proposed in chromatography as alternative measurements of chromatographic resolution [25,26], but they have not been used as COFs in optimisation yet. In this work, NAS and derived figures of merit are proposed as objective functions, and their performance is critically evaluated using several cases of study in high-performance liquid chromatography. As deconvolution-oriented resolution criterion, NAS is compared with other chromatographic measurements, and the probability of obtaining biased deconvolved elution profiles is examined. The study is developed in two steps: first, the correlations of several resolution functions with the deconvolution error in an artificial peak crossing are established. Secondly, these functions are applied to the optimisation of mobile phase composition in some cases of study. The final aims are to determine, on the one hand, whether the deconvolution errors are well correlated with NAS-derived COFs, and on the other, if this new kind of COFs are able to find optimal compositions as other classical COFs do.

2. Theory

2.1. Elementary resolution criteria

For computing the overall resolution, the first step is to establish an individual criterion that quantifies the separation for each peak or peak pair. The peak separation measurements examined in this work as resolution quantifiers are (see Fig. 1).

(i) Chromatographic resolution, R_s, which for peak pair s–b is:

$$R_s = \frac{t_{R,b} - t_{R,s}}{B_s + A_b} \quad (1)$$

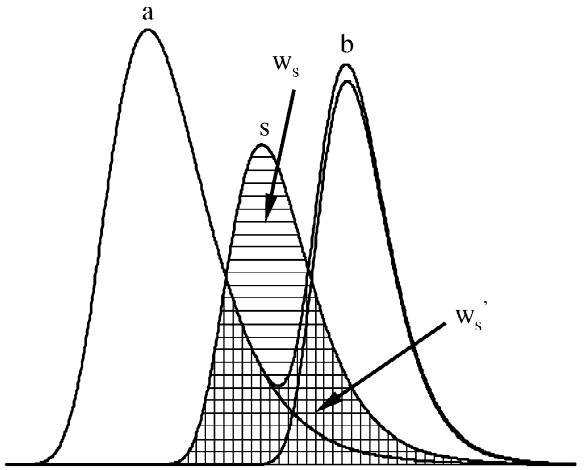

Fig. 1. Meaning of the peak purity criterion for *s*, which is the compound of interest.

where B_s is the distance from the maximum of peak *s* to its tail, and A_b the distance from the front of peak *b* to its maximum, both measured at 10% of peak height. Finally, $t_{R,b}$ and $t_{R,s}$ are the retention times for peaks *b* and *s*, respectively.

(ii) Peak purity, expressed as free area fraction [9]:

$$p_s = 1 - \frac{w_s'}{w_s} \tag{2}$$

This assessment, which has been demonstrated to have an especially good performance [11], is defined as the fraction of *s* not overlapped by its interferents (*a* and *b* in Fig. 1). In Eq. (2), w_s is the area of peak *s*, and w_s', the area under that peak overlapped by the chromatogram of the other compounds in the sample. Peak purity varies between 0 and 1: values close to 1 mean complete resolution, and close to 0, full overlap.

(iii) Multivariate measurements (first-order selectivity and scalar net analyte signal): Since these assessments are less intuitive, a more detailed description is required. The three peaks example will be considered again (Fig. 1). Any of these peaks can be considered as a vector with *t* elements (see Fig. 2a, vectors **a**, **b** and **s**). In this way, if the vectors of the interferents (*a* and *b*) are adjoined to build a matrix **W**, the fraction of **s** (the signal of analyte *s*) that is interfered by the signals of *a* and *b* (vectors **a** and **b**) will be given by [26]:

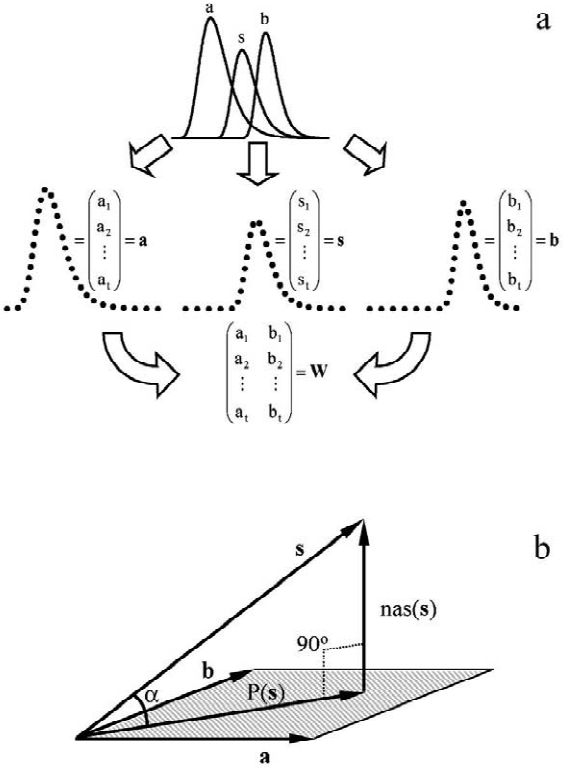

Fig. 2. Representation of a chromatogram under a vectorial perspective (a) to set out the multivariate figures of merit studied in this work. The corresponding geometrical representation is given in (b): the decomposition of the signal of *s*, onto the vectorial space of its interferents, *a* and *b*, gives rise to an explainable contribution—P(**s**)—and an unexplainable contribution: the net analyte signal of solute *s*—nas(**s**).

$$\mathrm{P}(\mathbf{s}) = \left[\mathbf{W}\ \mathbf{W}^{+}\right] \mathbf{s} \tag{3}$$

where $\mathbf{W}^{+}$ denotes the Moore–Penrose generalised inverse of **W**. The fraction of *s* free of interference, namely the NAS of *s*, can be calculated as (see Fig. 2b):

$$\mathrm{nas}(\mathbf{s}) = \mathbf{s} - \mathrm{P}(\mathbf{s}) \tag{4}$$

According to Lorber [23], only nas(**s**) contains useful information to calibrate the analyte. From this framework, several multivariate measurements can be defined. One of them is the multivariate first-order selectivity, which for solute *s* is given by [25,26]:

$$\mathrm{SEL}_s = \frac{||\mathrm{nas}(\mathbf{s})||}{||\mathbf{s}||} \quad (5)$$

where $||\mathbf{s}||$ denotes the norm of s. Several definitions of $||\mathbf{s}||$ are possible depending on the value given to a characteristic parameter k [24,27]:

$$||\mathbf{s}||_k = \left(\sum_{j=1}^{t} (|s_j|)^k \right)^{\frac{1}{k}} \quad (6)$$

Since a similar definition holds for $||\mathrm{nas}(\mathbf{s})||$, several SEL_s measurements can be defined. However, only with $k=2$, SEL_s equals $\sin(\alpha)$ (Fig. 2b), and varies between 0 and 1. Similarly to p_s, values close to 0 entail a strong overlap between the analyte of interest and its interferents, whereas values close to 1 mean that compound s is chromatographically well resolved.

Note that both $||\mathrm{nas}(\mathbf{s})||$ and $||\mathbf{s}||$ are directly proportional to the peak size of solute s, and equal zero in its absence. Consequently, SEL_s will be independent of both concentration and detector sensitivity. Usually, this independence is a beneficial feature, but in some circumstances it can be undesirable. For instance, when the detector shows a different sensitivity to each compound, or when the concentration of the eluted compounds are extremely different (e.g., separation of an impurity from a major component). In such cases, more realistic optimisations can be carried out correcting SEL_s by including an additional term in Eq. (5), in order to introduce a dependence on the concentration. Such a descriptor will be called weighted multivariate selectivity, WSEL_s:

$$\mathrm{WSEL}_s = \frac{||\mathrm{nas}(\mathbf{s})||}{||\mathbf{s}||} w_s \quad (7)$$

The weight w_s can be set to the peak area of each solute to consider both concentrations and sensitivities. Therefore, solutes presenting low concentrations and/or sensitivities give rise to low w_s values, and thus, the significance of obtaining acceptable selectivities for these compounds is increased.

Another straightforward candidate measurement to be used as resolution criterion, which certainly depends on the signal size, is the norm of $\mathrm{nas}(\mathbf{s})$, $||\mathrm{nas}(\mathbf{s})||$. This assessment—that has been called scalar net analyte signal (scalar NAS) [28]—depends linearly on the analyte peak size, but presents some disadvantages, which will be discussed in Section 4.1.

2.2. Global resolution

The elementary resolution measurements for all peaks should be reduced to a single numerical value to express the overall separation of all compounds. For this purpose, the resolution criteria should be normalised; otherwise, artefacts will be introduced in the global resolution [29]. SEL_s and p_s do not present such problems, but the values of w_s in WSEL_s (Eq. (7)) require normalisation with respect to the maximal peak area. For elementary assessments varying between 0 and 1, the most adequate reduction function is the product of the elementary values:

$$\mathrm{RG}_p = \prod_{s=1}^{n} p_s \quad (8)$$

$$\mathrm{RG}_{\mathrm{SEL}} = \prod_{s=1}^{n} \mathrm{SEL}_s \quad (9)$$

where n is the number of solutes in the sample. In contrast, R_s values are not normalised. The following reduction function is useful to make the values in different situations comparable [29]:

$$\mathrm{RG}_{\mathrm{Rs}} = \frac{\prod_{i=2}^{n} R_{s,i}}{\left(\frac{\sum_{i=2}^{n} R_{s,i}}{n-1} \right)^{n-1}} \quad (10)$$

These reduction functions (Eqs. (8)–(10)) are used as COFs throughout this work.

2.3. Prediction of chromatograms

The implementation of COFs in interpretive optimisation strategies requires the development of a simulation system, able to predict the chromatogram at any arbitrary set of experimental factors.

Retention in reversed-phase liquid chromatography (RPLC) with aqueous–organic mixtures can

be predicted as a function of mobile phase composition, as follows [30]:

$$\log k = c_0 + c_1\varphi + c_2\varphi^2 \quad (11)$$

where k is the retention factor, φ the volumetric fraction of organic solvent, and c_0, c_1 and c_2 regression parameters. Other peak properties as efficiency and asymmetry factor, can be locally modelled.

Once solved the problem of predicting the retention—and optionally, the efficiency and asymmetry—a mathematical peak function is required to simulate chromatograms. A model suitable for asymmetrical peaks, in which the standard deviation of a Gaussian function is modified using a polynomial, is [13]:

$$h(t) = h_0 \exp\left[-\frac{1}{2}\cdot\left(\frac{t - t_R}{s_0 + s_1(t - t_R)}\right)^2 \right] \quad (12)$$

h_0 being the peak height, t_R the retention time, and s_0 and s_1 coefficients related with peak efficiency and asymmetry factor. Eq. (12) allows an adequate description of chromatographic peaks. It however presents the disadvantage of producing a growth of the baseline outside the peak neighbourhood for strongly asymmetrical peaks. This drawback was tackled by just setting the signal to the minimal values outside the peak window.

3. Experimental

Several cases of study were generated using the retention data of a set of 16 aromatic compounds eluted in methanol–water and acetonitrile–water, which were taken from a report of Rosés and Bosch [31]. For both aqueous–organic systems, mobile phase compositions were 30, 40, 50, 60, 70, 80, 90 and 100% (v/v) organic solvent. The solutes were chromatographed with a thermostatted LiChrospher 100 RP-18 column (250×4.0 mm I.D., 5 μm particle size), and detected at 282 nm.

Data treatment was done using laboratory built-in routines, written in MATLAB 4.2c (The Mathworks, Natick, MA, USA).

4. Results and discussion

4.1. Drawbacks of scalar net analyte signal as resolution assessment

As commented in Section 2.1, some drawbacks make the use of the scalar NAS as resolution criterion difficult. First, $\|\mathrm{nas}(\mathbf{s})\|$ depends on the number of points in the chromatogram: the higher the sampling frequency, the higher the scalar NAS. Secondly, an undesirable dependence between $\|\mathrm{nas}(\mathbf{s})\|$ and the peak shape exists: the scores of narrower peaks, as those obtained at shorter retention times, are artificially high. As a consequence, mobile phase composition will affect the scalar NAS.

The first drawback can be fully eliminated by multiplying $\|\mathrm{nas}(\mathbf{s})\|$ by $\Delta t^{1/k}$, where Δt is the sampling interval, and k the parameter that defines the type of norm (Eq. (6)). The second problem (i.e., the peak widening effect) can be controlled by varying the parameter k: the closer to one, the smaller the dependence on peak shape. Although at $k=1$ the problem of the dependence of $\|\mathrm{nas}(\mathbf{s})\|$ with peak shape is fully resolved, a third drawback is collaterally generated: abnormal growths of scalar NAS are detected at moderate overlaps. This effect is observed for $k<2$, and is mainly due to the unfulfilment of the Euclidean metrics. In such cases, $\|\mathrm{nas}(\mathbf{s})\|$ reaches values slightly greater than those measured in situations of baseline resolution. All these problems recommend withdrawing the use of the scalar NAS as resolution criterion in chromatography. For these reasons, this criterion will not be considered further.

4.2. Influence of the type of norm in SEL_s

The former two drawbacks are not present in SEL_s, since the ratio of norms in Eq. (5) makes this measurement independent of the number of points in the chromatogram, and also of its shape. However, slight growths with norms calculated using $k<2$ are similarly observed. In order to study the most adequate type of norm for computing SEL_s, chromatograms with different overlap were generated by crossing an artificial peak with another identical. Fig. 3 depicts the values of SEL_s as a function of peak separation. As can be seen, in situations of moderate

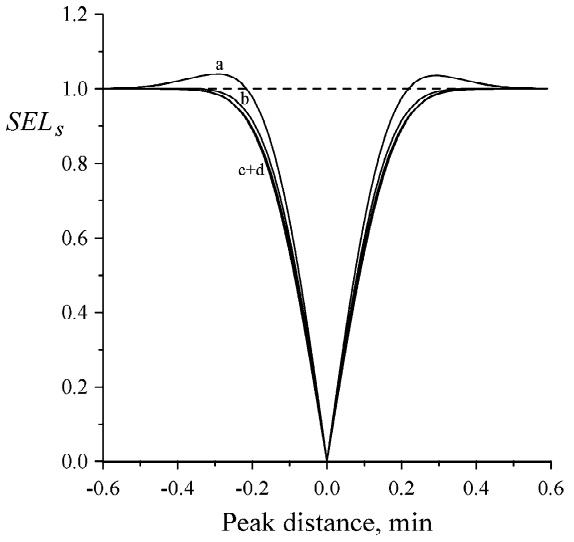

Fig. 3. Dependence of the multivariate selectivity with peak separation, when a Gaussian peak eluted at 3.5 min with $N=1000$ is crossed by another peak of identical properties. Along the migration, peak width was kept constant. The lines correspond to different values of k in Eq. (6): (a) 1, (b) 1.5, (c) 2, and (d) 3. A reference line with $SEL_s=1$ is overlaid.

overlap, values of SEL_s greater than those obtained with completely isolated peaks ($SEL_s=1$) are observed at $k<2$ (curves *a* and *b*). The growth at $k=1.5$ is considerably smaller, although evident at appropriate magnifications. The origin of this artefact is the same as that commented in the scalar NAS criterion, formerly discussed. Such results advise about using Euclidean norms for the computation of SEL_s, and accordingly, $k=2$ was adopted for further studies in next sections. Besides, Euclidean norms allow the development of simple expressions, analytically more useful [25].

4.3. Performance of COFs in the prediction of deconvolution errors

Before considering the behaviour of R_s, p_s and SEL_s in multi-solute mixtures, a simple case of two-peak crossing will be comprehensively studied. The greater simplicity will allow getting more neat conclusions about the relationship between these COFs and deconvolution errors.

A binary mixture was used to generate a set of 73 simulated chromatograms with different peak separation. These simulations correspond to a peak crossing, where one of the peaks was kept fixed while the other was gradually shifted up to cross the first. This example allowed examining a large number of situations with intermediate resolution. In addition, the use of artificial signals prevented the influence of non-idealities, such as the lack of fit or the presence of heteroscedastic noise, that would bias the interpretability of the results. In order to avoid biases attributable to lack of fit, the same peak model was used for building and deconvolving the artificial chromatograms.

The theoretical signals corresponded to a hypothetical experiment where both solutes eluted at 10 min with a chromatographic plate count of $N=3000$. The peaks were simulated with different asymmetry ($B/A=1.2$ and 2.2), but the same area; peak heights were 1 and 0.86, respectively. The crossings were generated keeping the shape of both peaks constant, and changing the position of the second between $t_R=9$ and 11 min, stepped by 0.05 min. Situations of strong overlap (from 9.8 to 10.2 min) were sampled in more detail, with time increments of 0.01 min. Each case of study was built by combining the individual contributions of both peaks (Eq. (12)), and adding normally distributed noise with amplitude of 0.01 standard deviation units to the resulting chromatogram.

Each of the 73 chromatograms was deconvolved five times, using a different seed for noise generation to obtain the peak parameters of each peak (Eq. (12)). Attending to the amount of known information, three different deconvolution methods were applied to the 73×5 cases.

(i) Fully free deconvolution. In this method, the retention times (t_R), heights (h_0) and peak shape parameters (s_0 and s_1), were fitted without any restriction. This strategy is suitable when the peak shape of the solutes cannot be determined a priori from additional experiments (e.g., when no standards are available, or when strong changes in peak shape are observed).

(ii) Constrained peak profile deconvolution. In this case, the true s_0 and s_1 values were kept constant, and only the retention times and heights were fitted. This strategy is the most practical in real cases, since flow-rate or sample injection irregularities introduce fluctuations in retention times. Thus, it should be

used in those cases where peak shapes are well known and do not vary appreciably among injections [18,21].

(iii) Fully constrained deconvolution. This method, common in spectroscopic analysis, consists of fitting only peak heights. In chromatography, owing to the frequent irregularities in retention times, the suitability of this method is rather questionable, but it is included in this study to prospect the possibilities of the deconvolution in ideal conditions, and get an estimation of the best possible results.

Methods (i) and (ii) are non-linear with respect to the parameters, requiring thus non-linear fitting procedures. Such numerical methods are iterative and present the risk of being trapped in local solutions. A suitable method, which hybridises an unconstrained local search with a classical genetic algorithm, has been recently proposed to avoid such kind of drawback [32,33]. This numerical method, called LOGA, was used in approaches (i) and (ii), but not in (iii). The reason is that the latter approach is rigorously linear, and a direct minimisation of the sum of the squared residuals (SSR) can be performed in a single step, obtaining the exact solution without any risk of ambiguities.

Note that the function to be minimised in any of these deconvolution problems is SSR. However, a low value does not necessarily mean a full agreement between the retrieved and theoretical peak profiles, since a given chromatogram can be usually modelled with more than one peak arrangement. A more appropriate error definition consists of relating directly the magnitude of the difference between deconvolved and theoretical individual peak profiles. A possible measurement of disagreement can be computed as the squared Euclidean norm of the vector difference between the deconvolved and theoretical profiles, since any chromatogram can be outlined as a vector (see Fig. 2a). This value will be named individual squared sum of residuals, SSR_i. The mean between the two error values obtained (one by solute) was computed for each of the 73×5 cases. In a further step, the five mean error values in each of the 73 overlapping cases, were averaged. Finally, the square root was calculated ($SSR_i^{1/2}$) and plotted in Fig. 4 for the three deconvolution methods.

Since the areas of both peaks are equal, peak purities of both compounds are identical. Owing to

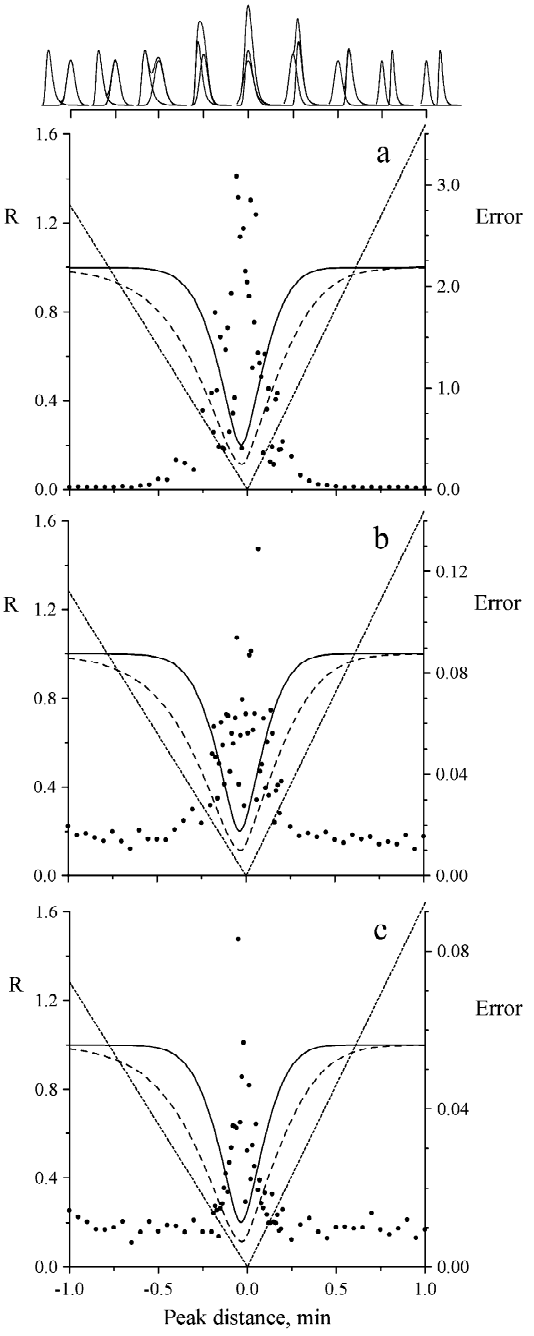

Fig. 4. Resolution and deconvolution errors as a function of peak separation, in a two-compound peak crossing. Three resolution measurements are plotted (left *y*-axis): R_s (short dashed line), p_s (long dashed line), and SEL_s (solid line). Dots represent the deconvolution errors (right *y*-axis), calculated as $SSR_i^{1/2}$, and cases a, b and c correspond to deconvolution methods (i), (ii) and (iii), respectively (see Section 4.3). The top of the figure shows the peak arrangements corresponding to the *x*-axis ticks.

this, only the peak purity values (p_s) of one compound were plotted in the figure. The same holds for SEL_s, even for peaks showing different areas. Finally, for R_s, there is only a single value available because this measurement is intrinsically associated to peak pairs.

As expected, the higher the amount of known information in the deconvolution process, the lower the probability of recovering biased profiles (note the differences in scale in the right *y*-axis of Fig. 4a–c). Method (i) gives poorer results than methods (ii) and (iii), which gave errors of similar magnitude. Note that in method (iii), a residual error of about 0.01 remains, even at baseline resolved chromatograms. Since this method is purely linear, the error corresponds to the magnitude of the added noise in standard deviation units. This result is not obtained by chance: the equivalence between $SSR_i^{1/2}$ and the noise can be rigorously demonstrated. Values of $SSR_i^{1/2}$ are especially useful to evaluate whether the deconvolution errors come from the presence of white noise or, on the contrary, are a true bias originated by non-linearities, or correlated signals. The figure illustrates this effect: in methods (i) and (ii), the remaining error in chromatograms resolved up to the baseline is slightly larger than 0.01, although the scale of the plots in method (i) does not allow to appreciate this appropriately.

In situations of moderate overlap, the three approaches yielded good results. Methods (i) and (ii) succeeded for peak distances above 0.5 min. This threshold of reliable deconvolution decreased to 0.25 min for method (iii). The existence of a threshold, which means in practice that deconvolutions are safe up to a certain peak overlap, evidences that R_s is not a good predictor of the deconvolution errors. Since R_s increases indefinitely with peak separation, no breakpoint can be correlated with a sudden increase in the deconvolution error.

The results show that both p_s and SEL_s are correlated with the deconvolution errors. Since the correlation is somewhat worse for peak purity, it can be concluded that multivariate selectivity is the best criterion in the prediction of deconvolution errors. For the former, the disagreement between the quantification of resolution and deconvolution errors is more perceptible in situations of moderate overlap, where a high probability of good results is expected. Therefore, if the p_s criterion is implemented as COF in an optimisation strategy, chromatograms with partial resolution will be over-penalised, although the individual peaks could be quantified with low errors by deconvolution approaches. These considerations should be taken into account if no full resolution is expected in the considered experimental domain.

The disagreement between the quantification of SEL_s values and deconvolution errors is negligible: the values of multivariate selectivity keep close to one when the probability of success in the deconvolution method is acceptable. Since the error pattern is identical in shape to SEL_s (although evidently opposite), multivariate selectivity gives the best description of the probability of success using any of the three deconvolution methods. Method (iii) yielded a sharper error pattern (Fig. 4c). Consequently, for this method, the SEL_s criterion is not so perfectly correlated as in methods (i) and (ii), but nevertheless it is still the best descriptor of deconvolution errors.

Note that situations in which $SEL_s = 1$ do not mean necessarily that baseline resolution has been reached, although good deconvolution results are obtained. Thus, SEL_s is a complementary criterion to be considered after checking with p_s that full resolution is not possible without the assistance of deconvolution techniques.

4.4. Implementation of SEL_s as COF: cases of study

In the previous section, the behaviour of SEL_s and other criteria was studied and correlated with peak deconvolution errors. The following step in the implementation of SEL_s as COF is studying its performance on real optimisation problems. For this purpose, a mixture of 16 aromatic compounds eluted in isocratic RPLC with mobile phases of methanol–water and acetonitrile–water was considered [31]. Plate counts were set to the reported mean values: 4430 and 5550 for methanol–water and acetonitrile–water mobile phases, respectively. Peak asymmetries were also assumed to be constant in the experimental domain and set to $B/A = 1.2$.

The chromatographic retention was modelled using Eq. (11). A set of chromatograms were simulated for a regular distribution of mobile phases in the experimental domain (from 30 to 100%, v/v,

organic modifier in both methanol and acetonitrile systems). For each simulated chromatogram, RG_p, RG_{SEL} and RG_{Rs} were calculated (Eqs. (8)–(10)). The results with methanol and acetonitrile are plotted in Fig. 5.

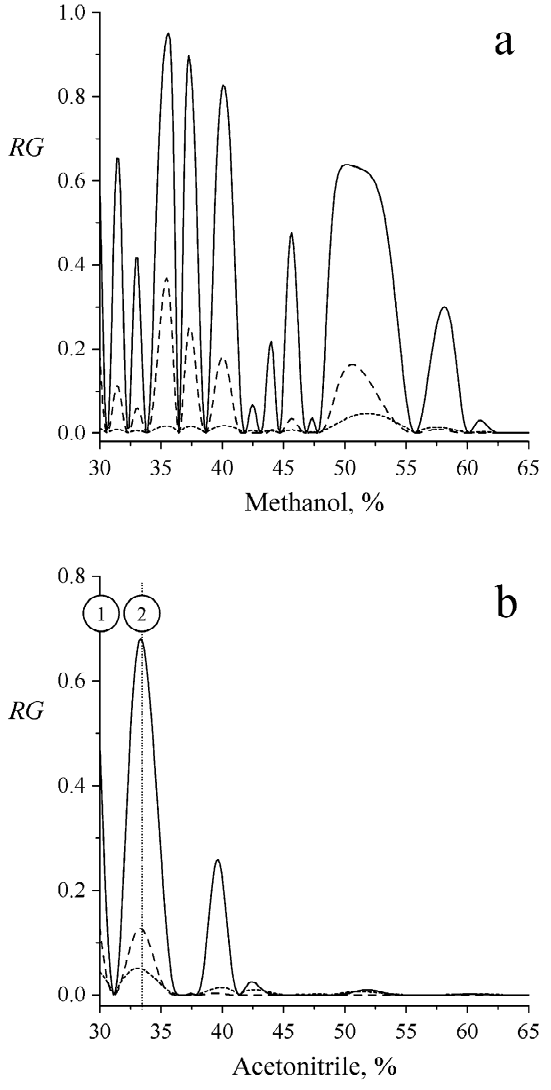

Fig. 5. Resolution maps corresponding to a mixture of 16 aromatic compounds eluted in isocratic RPLC with: (a) methanol–water, and (b) acetonitrile–water mobile phases. The plotted COFs are: RG_{SEL} (solid line, Eq. (9)), RG_p (long dashed line, Eq. (8)) and RG_{Rs} (short dashed line, Eq. (10)). Only the chromatographically significant domains of organic solvent are shown. Optimal compositions for the acetonitrile–water system are labelled as 1 (30% acetonitrile) and 2 (33.4%), encircled.

The maps obtained with methanol (Fig. 5a) show that the three criteria are sensitive to peak crossing, which originates several resolution drops to zero. There is a general trend of deterioration of resolution at increasing modifier concentrations, due to the shorter analysis times with mobile phases of higher elution strength. Note that each criterion gives a different significance to each composition. Also, peak purity gives smaller values than multivariate selectivity (see also Fig. 4), although both criteria advise the same optimal mobile phase composition. RG_{SEL} is a descriptor that combines the capability of RG_p of finding composition ranges yielding acceptable or full separation, with the ability of locating the optimal mobile phase in situations where full resolution is not possible and a mathematical deconvolution becomes an attractive alternative. RG_{Rs} yields biased resolution patterns at increasing methanol concentration. This artefact is due to the influence of the mean resolution term in the denominator of Eq. (10), since it decreases at increasing elution strength. This explains the abnormally high values of RG_{Rs} at higher concentrations of methanol.

The second example (Fig. 5b) illustrates how RG_p and RG_{SEL} weight the optima in a remarkably different way. Whereas RG_p locates two best compositions with practically identical performance (labelled with indexes 1 and 2), RG_{SEL} assigns a noteworthy different value to these compositions, so that one of them (33.4% acetonitrile) becomes clearly preferred. Fig. 6 shows the corresponding chromatograms. At 30% acetonitrile (Fig. 6a) an acceptable separation is achieved for peaks 4–6 and 12–13, but peaks 7–8 overlap fully. The chromatogram at 33.4% exhibits a moderate overlap for peaks 4–6 and a relatively high overlap for peaks 7–8 and 12–13 (Fig. 6b). A chromatographer would consider both situations unsatisfactory, since some solutes remain unresolved. In fact, the peak purity criterion tends to correlate well with the chromatographer appraisal [11]. According to this, it considers any kind of overlap undesirable. However, such situations do not constitute a serious drawback under the point of view of peak deconvolution. This is the key to understand why RG_{SEL} evaluates so differently the chromatograms. Thus, according to the RG_p criterion (see Table 1), no mobile phase is able to resolve satisfactorily the mixture ($p_s \approx 0.4$ for peaks 7 and 8

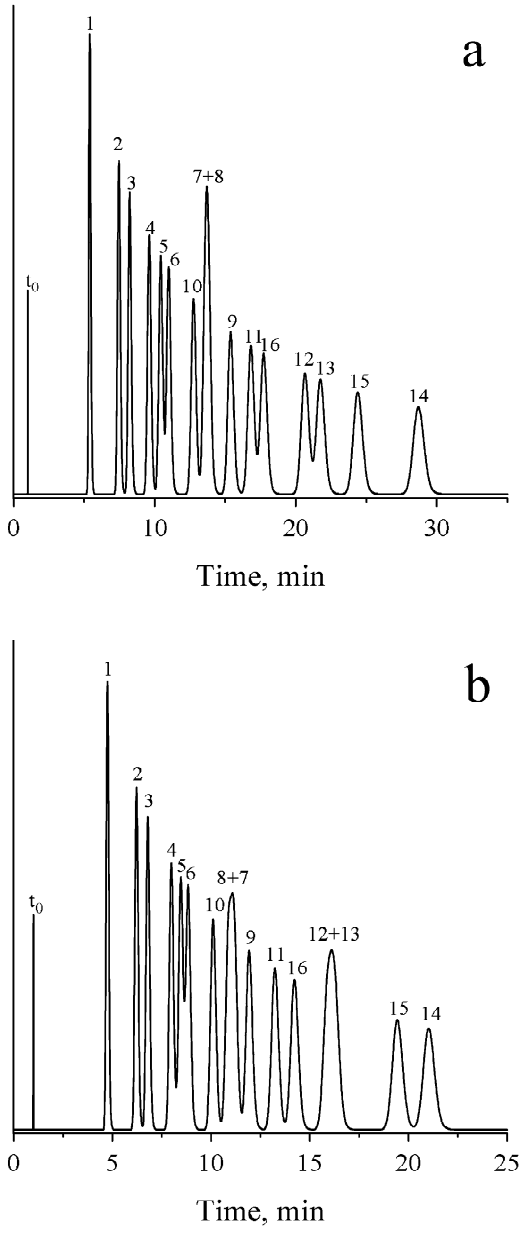

Fig. 6. Optimal chromatograms for a mixture of 16 aromatic compounds. Mobile phases: (a) 30%, and (b) 33.4% acetonitrile. Solutes: (1) phenol, (2) 4-nitrophenol, (3) 3-nitrophenol, (4) 2-methylphenol, (5) 2-chlorophenol, (6) 2,4-dinitrophenol, (7) 2-nitrophenol, (8) 3-chlorophenol, (9) 4-bromophenol, (10) 4-chlorophenol, (11) 2,4-dimethylphenol, (12) 2,6-dichlorophenol, (13) 4-chloro-3-methylphenol, (14) 2,4-dichlorophenol, (15) benzene, and (16) nitrobenzene.

at 30% organic solvent, and <0.7 for peaks 7, 8, 12 and 13 at 33.4%), but SEL_s concludes that retrieving all the solutes in the mixture by performing a deconvolution is possible, with relatively low errors ($SEL_s = 0.7$ for peaks 7 and 8 at 30% acetonitrile, and ≥ 0.9 for all solutes in the mixture at 33.4%). The feasibility of performing deconvolutions of the overlapped peaks with acceptable errors is demonstrated below.

Deconvolution of the critical peaks in the chromatograms shown in Fig. 6a and b was carried out with the three methods outlined in Section 4.3, at two noise levels (0.05 and 0.025 standard deviation units). Only peaks 4–6, 7–8 and 12–13 were deconvolved; peak pair 11–16 was not included, since it was well resolved at 33.4% acetonitrile and, in addition, its resolution at 30% yielded results similar to those obtained for peak pair 12–13, which is a more interesting case of study. The errors were calculated as $SSR_i^{1/2}$ and as the relative difference between the retrieved and true peak areas (ε_{area}). Ten chromatograms were generated changing the noise seed and deconvolved independently. The results are given in Table 1. To make the differences among criteria more evident, the discussion is focused mainly on the results of method (i).

The analysis of Table 1 confirms that the chromatographic resolution of peaks 4–6 is sufficient at both optima to obtain a reliable quantification. Only the errors achieved with method (i) in peaks 5 and 6 are slightly larger at 33.4% acetonitrile ($\varepsilon_{area} = 1.23$–$1.26\%$ and 0.72–0.79% at 0.05 and 0.025 noise levels, respectively). The other deconvolution methods yielded similar results with both optimal chromatograms at the two noise levels. The similar deconvolution errors obtained at both optima are not well represented by the different values of p_s, and especially, R_s.

Peak pairs 7–8 and 12–13 constitute a more interesting example to illustrate the differences in performance amidst the three studied COFs. Chromatogram at 33.4% acetonitrile (Fig. 6b) shows both peak pairs highly merged. The chromatographer would score this composition worse than 30% acetonitrile. However, deconvolution errors originate a quite different perspective. At 30% acetonitrile, method (i) failed dramatically with pair 7–8 ($\varepsilon_{area} =$

Table 1
Elementary resolutions and deconvolution errors for several peaks eluted with acetonitrile–water (Fig. 6), using methods (i)–(iii) (see Section 4.3)[a]

Solute	Resolution						Method	Noise[b] = 0.05				Noise[b] = 0.025			
	R_s		p_s		SEL_s			$SSR_i^{1/2}$		ε_{area} (%)		$SSR_i^{1/2}$		ε_{area} (%)	
	30%	33.4%	30%	33.4%	30%	33.4%		30%	33.4%	30%	33.4%	30%	33.4%	30%	33.4%
4	1.46	1.05	0.997	0.973	1.000	1.000	i	0.11	0.10	0.40	0.35	0.04	0.06	0.15	0.19
							ii	0.08	0.08	0.33	0.28	0.02	0.04	0.09	0.15
							iii	0.06	0.05	0.33	0.28	0.015	0.03	0.09	0.15
5	0.94	0.76	0.951	0.869	1.000	0.997	i	0.12	0.31	0.43	1.23	0.06	0.19	0.18	0.72
							ii	0.08	0.07	0.24	0.22	0.03	0.04	0.14	0.13
							iii	0.05	0.05	0.26	0.24	0.02	0.02	0.13	0.13
6	0.94	0.76	0.954	0.896	1.000	0.997	i	0.12	0.29	0.49	1.26	0.07	0.18	0.21	0.79
							ii	0.09	0.08	0.33	0.27	0.04	0.03	0.13	0.10
							iii	0.06	0.05	0.33	0.27	0.02	0.015	0.13	0.08
7	0.27	0.47	0.414	0.666	0.692	0.929	i	11.9	3.98	94.5	23.0	9.55	2.82	73.6	18.0
							ii	0.56	0.11	2.89	0.41	0.20	0.04	1.33	0.19
							iii	0.08	0.05	0.51	0.26	0.03	0.02	0.19	0.11
8	0.27	0.47	0.431	0.675	0.692	0.929	i	11.9	3.97	94.6	23.0	9.56	2.82	73.6	18.0
							ii	0.56	0.13	3.15	0.48	0.21	0.05	1.25	0.20
							iii	0.08	0.08	0.54	0.47	0.03	0.02	0.14	0.12
12	0.92	0.41	0.950	0.623	1.000	0.890	i	0.11	4.21	0.64	33.5	0.06	3.29	0.37	22.3
							ii	0.06	0.18	0.33	0.82	0.03	0.05	0.16	0.26
							iii	0.05	0.07	0.33	0.49	0.02	0.02	0.16	0.17
13	0.92	0.41	0.950	0.623	1.000	0.890	i	0.10	4.21	0.54	33.1	0.07	3.29	0.39	22.3
							ii	0.07	0.18	0.35	1.09	0.04	0.04	0.21	0.19
							iii	0.04	0.07	0.34	0.49	0.03	0.02	0.21	0.14

[a] Acetonitrile concentrations are given in % (v/v).
[b] Standard deviation units.

74–95%), whereas solutes 12–13 remained separated enough to be retrieved with very low errors (0.4–0.6%). Although an important bias remained at 33.4% acetonitrile in the deconvolution of both peak pairs (18–33%), this composition is globally the best. The same conclusion can be derived at both noise levels for methods (ii) and (iii), although the errors were much smaller.

Although ε_{area} values are easier to interpret, $SSR_i^{1/2}$ presents the advantage of having an intrinsically associated threshold, which has interesting practical consequences. When $SSR_i^{1/2}$ differs significantly from the standard deviation of the white noise (i.e., the $SSR_i^{1/2}$ threshold), the presence of a residual bias in the recovered signals can be suspected. Thus, solutes 4–6 with method (iii) are unbiasedly resolved, since $SSR_i^{1/2}$ practically equals the value of the white noise. Meanwhile, methods (i) and (ii), which are non-linear, yield $SSR_i^{1/2}$ values greater than the threshold. For the other peak clusters (i.e., peaks 7–8 and 12–13), method (iii) gives values similar or slightly greater than the standard deviation of the noise. Since a linear deconvolution method is used, the larger $SSR_i^{1/2}$ indicates that the bias has an origin not attributable to the deconvolution method: the presence of a higher correlation between signals.

A second interesting advantage related to the use of $SSR_i^{1/2}$ is its capability of comprehensively comparing signals. A low ε_{area} does not necessarily mean that the recovered profile fits the theoretical one, because the bias in retention times [methods (i) and (ii)] or in peak shapes (method (i)) are out of the scope of such a measurement. In contrast, $SSR_i^{1/2}$ quantifies the magnitude of the differences between the recovered and theoretical signals, and conse-

quently, is affected by errors in each experimental point.

5. Conclusions

Multivariate measurements are suitable choices as COFs in interpretive optimisations. One of the possible criteria in this category is the scalar NAS. It however presents the drawback of being dependent on the number of points in the chromatogram, as well as on peak widening at increasing retention times. The multivariate selectivity (SEL_s) presents the interesting advantage of detecting the presence of bias in the deconvolution of overlapped signals. This new COF is well suited to find, with minimal errors, the best mobile phase composition for the estimation of individual peak profiles using either linear or non-linear deconvolution methods.

Other criteria studied in this work for comparative purposes were the chromatographic resolution (R_s) and the peak purity (p_s). The former does not give useful correlations with the deconvolution error, whereas the latter, being better, does not yield so good correlations as SEL_s. Note that p_s is not a deconvolution-oriented resolution measurement. The disagreement between p_s and the deconvolution error is more appreciable at intermediate overlaps; p_s tends to penalise these situations in the same way as a chromatographer would do. However, these cases are the most relevant under the point of view of deconvolution. If a given mixture cannot be separated at any condition, a global criterion based on SEL_s tends to advise peak arrangements where some solutes overlap partially. In contrast, other conventional COFs yield more unbalanced arrangements, tending to recommend situations of stronger overlap in a smaller number of solutes.

SEL_s is a measurement associated to each peak, a feature highly desirable in the implementation of any measurement as COF. Several definitions of SEL_s are possible, depending on the kind of norm. The best studied choice was the Euclidean norm. Its derived definition of SEL_s yields normalised values, which is another advisable advantage in the implementation of this function as COF.

When the true peak profiles are known, the most appropriate error measurement for deconvolution purposes is the square root of the sum of the squared residuals between the individual peak profiles (retrieved and theoretical), $SSR_i^{1/2}$, since it allows to establish rigorously the presence of bias as a result of the deconvolution. In addition, this quantity is straightforwardly related with the white noise, which allows setting a quality threshold. $SSR_i^{1/2}$ cannot be measured in real chromatograms, whereas interpretive optimisations give a chance to predict the theoretical individual peak profiles.

Acknowledgements

This work was supported by Project BQU2001-3047 (Ministerio de Ciencia y Tecnología of Spain and FEDER funds) and Project CTIDIB/2002/226 (Generalitat Valenciana). J.R.T.L. and G.V.T. thank the MCYT for a Ramón y Cajal position, and the Generalitat Valenciana for an FPI grant, respectively.

References

[1] L.R. Snyder, J.W. Dolan, D.C. Lommen, J. Chromatogr. 485 (1989) 65.
[2] J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, J.J. Baeza-Baeza, Anal. Chim. Acta 348 (1997) 187.
[3] S. Heinisch, E. Lesellier, C. Podevin, J.L. Rocca, A. Tchapla, Chromatographia 44 (1997) 529.
[4] J.C. Berridge, J. Chromatogr. 244 (1892) 1.
[5] P.J. Schoenmakers, Optimization of Chromatographic Selectivity. A Guide to Method Development, Elsevier, Amsterdam, 1986.
[6] M.W. Watson, P.W. Carr, Anal. Chem. 51 (1979) 1835.
[7] J. Glajch, J.J. Kirkland, K.M. Squire, J.M. Minor, J. Chromatogr. 199 (1980) 57.
[8] R.D. Caballero, S.J. López-Grío, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, J. Liq. Chromatogr. Rel. Technol. 24 (2001) 1895.
[9] G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, J. Chromatogr. A 876 (2000) 17.
[10] R. Cela, C.G. Barroso, J.A. Pérez-Bustamante, J. Chromatogr. 485 (1989) 477.
[11] S.J. López-Grío, G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Anal. Chim. Acta 433 (2001) 187.
[12] J. Krupcík, P. Oswald, I. Spánik, P. Májek, M. Bajdichová, P. Sandra, D.W. Armstrong, Analusis 28 (2001) 859.
[13] J.R. Torres-Lapasió, J.J. Baeza-Baeza, M.C. García-Alvarez-Coque, Anal. Chem. 69 (1997) 3822.
[14] P. Nikitas, A. Pappa-Louisi, A. Papageorgiou, J. Chromatogr. A 912 (2001) 13.

[15] T.L. Pap, Zs. Papai, J. Chromatogr. A 930 (2001) 53.
[16] V. Di Marco, G.G. Bombi, J. Chromatogr. A 931 (2001) 1.
[17] J. Li, J. Chromatogr. A 952 (2002) 63.
[18] R.D. Caballero, M.C. García-Alvarez-Coque, J.J. Baeza-Baeza, J. Chromatogr. A 954 (2002) 59.
[19] W. Cai, F. Yu, X. Shao, Z. Pan, Anal. Lett. 33 (2000) 373.
[20] X.G. Shao, Z.H. Chen, X.Q. Lin, Chemometr. Intell. Lab. Syst. 50 (2000) 91.
[21] G. Vivó-Truyols, R.D. Caballero, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, J. Chromatogr. A 958 (2002) 35.
[22] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A.
[23] A. Lorber, Anal. Chem. 58 (1986) 1167.
[24] J.H. Kalivas, P.M. Lang, Chemometr. Intell. Lab. Syst. 32 (1996) 135.
[25] K. Faber, A. Lorber, B.R. Kowalski, J. Chemometr. 11 (1997) 419.
[26] N.J. Messick, J.H. Kalivas, P.M. Lang, Anal. Chem. 68 (1996) 1572.
[27] J.H. Kalivas, P.M. Lang, Mathematical Analysis of Spectral Orthogonality, Marcel Dekker, New York, 1993.
[28] N.M. Faber, Chemometr. Intell. Lab. Syst. 50 (2000) 107.
[29] S. Carda-Broch, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Anal. Chim. Acta 396 (1999) 61.
[30] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, J. Chromatogr. 185 (1979) 179.
[31] M. Rosés, E. Bosch, Anal. Chim. Acta 274 (1993) 147.
[32] G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Chemometr. Intell. Lab. Syst. 59 (2001) 89.
[33] G. Vivó-Truyols, J.R. Torres-Lapasió, A. Garrido-Frenich, M.C. García-Alvarez-Coque, Chemometr. Intell. Lab. Syst. 59 (2001) 107.