Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1096 (2005) 123–132
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Robust interpretive optimisation in high-performance liquid chromatography considering uncertainties in peak position☆

G. Vivó-Truyols, V. Concha-Herrera, J.R. Torres-Lapasió *, M.C. García-Alvarez-Coque

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

Available online 14 April 2005

Abstract

In the context of interpretive chromatographic optimisation, robustness is usually calculated by introducing deliberated shifts in the nominal optimal conditions and evaluating their effects on the monitored objective function, mimicking thus the experimental procedures used in method validation. However, such strategy ignores a major source of error: the uncertainties associated to the modelling step, that may give rise to deceiving results when conditions that were expected to yield baseline separation are reproduced in the chromatograph. Two approaches, based on the peak purity concept, are here proposed to evaluate the robustness of the objective function under the perspective of measurement errors and modelling. The first approach implements these uncertainties as an extra band broadening for each chromatographic peak. The second one implements them as peak fluctuations in simulated replicated assays, which gives rise to a distribution of peak purities, easily computed through Monte-Carlo simulations. Both approaches predict satisfactorily a decreased separation capability, with respect to the conventional approach, for those situations where the uncertainties in peak position make the objective function critical. The first approach is less optimistic and formally less rigorous than the second one, but its computation is simpler. It can be used to map the critical resolution regions, to be comprehensively appraised further by the slower, although more rigorous, Monte-Carlo approach.

Keywords: Reversed-phase liquid chromatography; Robustness; Interpretive optimisation; Modelling of retention

1. Introduction

Interpretive chromatographic optimisations are those supported by models (or more complex algorithms), which are applied to predict the quality of the separation as a function of the experimental factors being optimised. The optimal condition corresponds to the combination of experimental variables whose predicted chromatogram shows the best separation among peaks. The optimisation procedure consists of maximising a numerical expression of separation quality, namely the "objective function". However, the condition found will be useless if it is so critical that the predicted optimal chromatogram cannot be obtained in practice, owing to unavoidable random shifts from the nominal values in the experimental factors. The aim of this work is proposing an objective function able to discriminate the practical translation of the optimised conditions to the chromatograph. This mimics what is formally done in method validation, but the evaluation is carried out during the proper method development step, saving time and effort. Accordingly, it can be considered as a "pre-validation", which will prevent to choose an optimal condition giving rise to a chromatogram unfeasible to be reproduced in practice.

Several validation strategies focused on the evaluation of robustness in analytical methods can be found in the literature [1–5]. The simplest procedures monitor experimentally the effects of deliberated shifts in the optimal conditions. Interpretive chromatographic optimisation approaches allow an alternative measurement of robustness as the variation that suffers the objective function, by simulating the mentioned shifts. A logical approach to quantify robustness is, thus, the use of partial derivatives of the objective function with respect to the experimental factors [6]. In this case, computer simulation just mimic errors in setting the optimal conditions, as could be equally done experimentally.

☆ This work was presented in the 25th International Symposium on Chromatography, held in Paris in October 2004.

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

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.03.057

However, the interpretive approach indicated above for robustness evaluation is too optimistic: it leaves aside the uncertainty introduced by the retention model, which constitutes in some instances an important source of error, even in the absence of lack of fit. It is obvious that the predictions of the objective function have a limited precision, since the models that support them always involve some uncertainty associated to the fitting process, that comes from the model itself, the experimental data and the regression procedure. The uncertainties associated to the models can be estimated in most cases, at least when they are fitted in a least-squares fashion, provided that some degrees of freedom are kept and the fitted models are unbiased.

As can be observed, the uncertainties in the predictions of retention have been traditionally studied as an independent problem of the sensitivity of the objective function to changes in the experimental factors. We have found no previous reports where both concepts were interconnected; particularly, the robustness has not been evaluated from the susceptibility to changes in the objective function originated by the propagation of the uncertainties associated to the retention model. This research is aimed to develop a statistical theory oriented to yield more robust optimisation methods, based on these considerations. For this purpose, the retention model uncertainties were propagated into the final resolution measurements, using two approaches. Both methods were applied to the optimisation of the separation of several amino acid derivatives of *o*-phthalaldehyde (OPA) and *N*-acetylcysteine (NAC) in reversed-phase liquid chromatography.

2. Theory

2.1. Conventional optimisation strategy

Interpretive optimisations are well known in the chromatographic field [7], and only a brief description will be given here. These approaches are based on the computation of a parameter depicting the resolution of the analysed mixture, which is monitored as a function of the experimental factor(s). In the example of concern, the factor being optimised was the concentration of organic modifier in the aqueous–organic mobile phase, which is usually selected owing to its major influence on retention in a reversed-phase system.

In a first step, the retention of each compound is modelled from a small number of experiments. For this purpose, the OPA–NAC derivatives of six amino acids (asparagine, serine, glutamine, histidine, arginine and threonine) were eluted in several mobile phases containing acetonitrile in the range 5.0–17.5% (v/v). The retention data from this experimental design were fitted to two well-known models [8,9]:

$$\log k = a + b\varphi \tag{1}$$

$$\log k = a + b\varphi + c\varphi^2 \tag{2}$$

where $\log k$ is the decimal logarithm of the retention factor, φ the volumetric fraction of organic modifier in the mobile phase, and a, b and c are fitting parameters. Several discussions can be found in the literature on which of these two equations is correct to model the retention [10,11]. In this work, a partial ANOVA test [12] was used to check the significance of the curvature of $\log k$ versus φ. Only when this curvature was found statistically significant, Eq. (2) was used.

When Eqs. (1) or (2) are linearly fitted (which is normally the case), the logarithmic transformation converts a fraction of the random error in systematic [13]. This effect can be compensated with the introduction of weights in the linear fitting [14], which has allowed the improvement of predictions in chromatographic optimisation [15]. In the case of a logarithmic transformation, the weights are computed as follows [14]:

$$w = \frac{1}{(\partial \log k/\partial t_R)^2} = (t_0 2.303k)^2 \tag{3}$$

This weighting strategy has been applied to obtain a homoskedastic error distribution for the predicted response. However, good predictions of chromatograms need to model not only the retention of compounds but also the efficiency and peak asymmetry. In this work, linear models relating the efficiency and asymmetry factors to the mobile phase composition were locally fitted, following the strategy outlined elsewhere [16].

Once the retention, and incidentally other properties related to peak shape, have been modelled, the second step consists of the simulation of chromatograms, which are computed following a regular distribution of mobile phase compositions. These synthetic chromatograms are built by adding, compound by compound, the signals predicted according to a mixed linear-exponential modified Gaussian model. The basis of the model [17] is:

$$h(t) = h_0 \exp\left[-\frac{1}{2}\left(\frac{t - t_R}{s_0 + s_1(t - t_R)}\right)^2\right] \tag{4}$$

where h_0 is the maximal peak height, $h(t)$ the height at time t, t_R the solute retention time, and s_0 and s_1 are the standard deviation and a distorting parameter, respectively. The exponential modification consists of substituting Eq. (4) by exponential decays calculated in such a way that the continuity of the mixed-function is guaranteed at 10% peak height. Peak parameters in Eq. (4) and the exponential auxiliary functions are calculated from the peak area, and the predicted retention time, peak efficiency and asymmetry factor at 10% peak height. More details are given in Ref. [18].

The outlined simulation procedure is applied to compute the elementary resolutions in a predefined set of conditions included within the experimental design. As a measurement of resolution, we selected the peak purity [19]:

$$p_{j,i} = 1 - \frac{o'_{j,i}}{o_{j,i}} \tag{5}$$

where subindexes j and i refer to the mobile phase composition and compound, respectively, $o'_{j,i}$ is the peak area of compound i overlapped by the chromatogram of its interferents, and $o_{j,i}$, the total area of the considered peak. Some work has been published illustrating the advantages of using peak purity instead of R_S [19]. The most relevant feature of p is that it addresses a single value to each compound, instead of to each pair of neighbouring peaks, which is an essential feature for some developments in chromatographic optimisation.

Next, for each experimental condition, the individual peak purities should be reduced to a single value expressing the overall separation of all compounds:

$$P_j = \prod_{i=1}^{\mathrm{ns}} p_{j,i} \tag{6}$$

where P_j is the overall purity at condition j and ns is the number of solutes in the mixture. The optimal experimental condition (in the example shown, the optimal acetonitrile concentration) is that one yielding the maximal value of overall purity.

2.2. Measurement of peak position uncertainties

In this work, the fitting of the retention data to Eqs. (1) and (2) is used with a double purpose: the prediction of retention times and the estimation of the uncertainties in peak position. The procedure for obtaining these uncertainties applies the general rules of error propagation in least-squares fitting, and for this reason, only a brief explanation will be given here. For further details, see Ref. [20].

Accordingly, the standard deviation in peak position for solute i at a given experimental condition j is given by:

$$s_{j,i} = \sqrt{\mathbf{x}_j \mathbf{V} \mathbf{x}_j^{\mathrm{T}}} \tag{7}$$

where $\mathbf{x}_j$ is a row vector including the derivatives of the retention time with respect to each parameter (evaluated at condition j), $\mathbf{V}$ the variance–covariance matrix of the model parameters, and $\mathbf{x}_j^{\mathrm{T}}$, the transpose of $\mathbf{x}_j$. For Eqs. (1) and (2):

$$\mathbf{x}_j = t_0 2.303\, k_j \begin{bmatrix} 1 & \varphi_j \end{bmatrix} \tag{8}$$

$$\mathbf{x}_j = t_0 2.303\, k_j \begin{bmatrix} 1 & \varphi_j & \varphi_j^2 \end{bmatrix} \tag{9}$$

respectively. In these equations, k_j is the predicted retention factor at condition j.

The matrix $\mathbf{V}$ in Eq. (7) can be estimated from the regression as a side result (provided that at least one degree of freedom remains):

$$\mathbf{V} = s_y^2 (\mathbf{X}^{\mathrm{T}} \mathbf{W} \mathbf{X})^{-1} \tag{10}$$

where $\mathbf{X}$ (i.e. the design matrix) contains the derivatives of $\log k$ for compound i with respect to each parameter (columns), evaluated at each point of the experimental design (rows), and $\mathbf{W}$ is a diagonal matrix containing the weights calculated according to Eq. (3). The term s_y^2 is the pure experimental error in the response y (in our case, the retention time, t_R), measured as a variance. In the absence of lack of fit, for solute i, s_y^2 can be approximated to:

$$s_{yi}^2 = \frac{\sum_{j=1}^{\mathrm{nc}} (t_{\mathrm{R}j,i} - \hat{t}_{\mathrm{R}j,i})^2}{\mathrm{nc} - \mathrm{np}} \tag{11}$$

where nc is the number of experimental points (i.e. mobile phases) available for solute i, $t_{\mathrm{R}j,i}$ and $\hat{t}_{\mathrm{R}j,i}$ the experimental and predicted retention times, respectively, and np the number of parameters in the solute retention model. Note that all factors affecting the retention times (e.g. pump irregularities and mobile phase mispreparation) will broaden the uncertainty region around the regression line.

Let's suppose a peak of solute i that is expected to appear at a certain time, $\hat{t}_{\mathrm{R}j,i}$, according to the retention model for the jth mobile phase. The treatment given above (Eqs. (7)–(11)) allows to estimate the uncertainty $s_{j,i}$ in $\hat{t}_{\mathrm{R}j,i}$. The probability of finding the peak shifted to a time t instead of $\hat{t}_{\mathrm{R}j,i}$ can be estimated from the t-Student distribution (referred here as tstud to avoid confusions with time symbols). Thus, if the tested retention is standardised taking $\hat{t}_{\mathrm{R}j,i}$ and $s_{j,i}$ as the mean time and standard deviation, the following estimator can be calculated:

$$\mathrm{tstud}_{\mathrm{exp}} = \frac{t - \hat{t}_{\mathrm{R}j,i}}{s_{j,i}} \tag{12}$$

Then, the probability of finding the peak shifted in an amount $(t - \hat{t}_{\mathrm{R}j,i})$ will be given by:

$$\pi_{j,i}(t) = \mathrm{tpdf}(\mathrm{tstud}_{\mathrm{exp}}, \mathrm{df}_i) \tag{13}$$

where $\pi_{j,i}(t)$ is the probability of finding solute i between t and $t + \mathrm{d}t$, at the experimental condition j, and tpdf represents the probability density function for the t-Student distribution with df_i degrees of freedom. As a consequence, the 95% confidence interval for finding solute i at the experimental condition j will be given by:

$$t_{j,i} = \hat{t}_{\mathrm{R}j,i} \pm \mathrm{tstud}(\alpha = 0.05, \mathrm{df}_i) s_{j,i} = \hat{t}_{\mathrm{R}j,i} \pm c_{j,i} \tag{14}$$

where tstud is the t-Student statistic at the specified confidence level of $\alpha = 0.05$ in a two-sided test with df_i degrees of freedom, and $s_{j,i}$ is given by Eq. (7).

2.3. Strategies for chromatographic optimisation

Two optimisation criteria considering uncertainties in peak position, and based on the peak purity concept, were developed and contrasted with the equivalent unrobust assessment.

2.3.1. Approach (i): measurement of conventional peak purities

This approach has been the subject of previous reports [15,16,19], and was outlined in Section 2.1. It has been included here, as commented, to establish a reference to

appraise the results of the so called "robust optimisation approaches" (Approaches (ii) and (iii)).

2.3.2. Approach (ii): measurement of peak purity from oversized chromatograms

This approach implements the uncertainties in peak position as an extra band broadening for each chromatographic peak. Accordingly, two sources of variance are assumed: the band broadening of the peak originated by the chromatographic process and the uncertainties in retention time associated to the predictions achieved with the retention model. Since both variances are uncorrelated, the total variance in the location of a certain compound i at the experimental condition j is calculated as follows:

$$s_{0j,i}^2 = s_{cj,i}^2 + (f s_{j,i})^2 \tag{15}$$

The peak simulated using $s_{0j,i}$ as standard deviation (the term to be introduced in the mixed linear-exponential Gaussian, see Section 2.1) will be called in this work "oversized chromatographic peak". As can be seen, $s_{0j,i}$ gathers two contributions: $s_{cj,i}$ (the band broadening due to the diffusion of the solute inside the column, which can be calculated from the column efficiency), and $s_{j,i}$ (the uncertainty in peak position, calculated according to Eq. (7)). The factor f accounts for the difference between the normal and the t-Student distributions, since the uncertainty in peak position follows the latter one, due to the low number of degrees of freedom. Usually, the experimental designs include four to six experiments, and the model includes two or three parameters (Eqs. (1) and (2)). Accordingly, df = 2 or 3 are available in most cases. The correction factor f will be then 1.2–1.3. For more degrees of freedom, f approximates to 1.0 (e.g. f is 1.11 and 1.05 for df = 5 and 10, respectively).

Approach (ii) is identical to Approach (i), except in the use of oversized chromatograms. Since the peaks are now wider, the values of elementary purity (Eq. (5)) are systematically decreased with regard to Approach (i). When these elementary values are multiplied each other (Eq. (6)), a robust measurement of peak purity is obtained.

2.3.3. Approach (iii): computation of the mean peak purity through Monte-Carlo simulations

From a probabilistic standpoint, the simulated chromatogram at a given experimental condition, as computed in Section 2.1, represents only the most likely peak configuration in that condition. However, replicated experimental chromatograms can be expected to show peak fluctuations, since the position of each peak may vary within its confidence interval. Accordingly, if the chromatographer would repeat the injection a number of times and were able to measure the peak purity for each compound at that experimental condition, a distribution of peak purities will be observed for each compound. This experimental procedure can be mimicked by computer simulation.

Approach (iii) lies on the idea of obtaining a range of peak purities for a single compound at a given experimental condition, by simulating random shifts within the confidence interval. The smaller the deviation in retention time, the greater its probability. In this approach, the averaged purity (considering all possible peak configurations) is computed at each experimental condition j. This averaged purity is formally defined as:

$$\bar{p}_{j,i} = \int_{t_{R,j,1}-c_{j,1}}^{t_{R,j,1}+c_{j,1}} \int_{t_{R,j,2}-c_{j,2}}^{t_{R,j,2}+c_{j,2}} \cdots \int_{t_{R,j,ns}-c_{j,ns}}^{t_{R,j,ns}+c_{j,ns}} p_{j,i}(t_1, t_2, \ldots, t_{ns}) \times \left[\prod_{k=1}^{ns} \pi_{j,k}(t_k)\right] dt_1\, dt_2 \cdots dt_{ns} \tag{16}$$

where $p_{j,i}(t_1, t_2, \ldots, t_{ns})$ is the peak purity for solute i, when the ns solutes are located at times $t_1, t_2, \ldots, t_{ns}$, respectively, and $\pi_{j,k}(t_k)$ is the probability (see Eq. (13)) of finding solute k at the experimental condition j in the interval t_k to $t_k + dt_k$. Note that the limits of each integral are defined in such a way that each t_k value always falls within its confidence interval (defined in Eq. (14)).

For each mobile phase composition (monitored by index j), there will be a predicted chromatogram, where all peaks will remain in their theoretical positions. The probabilistic treatment will disturb these positions giving rise to different peak configurations. Accordingly, there will be a unique "experimental condition" (the considered mobile phase), and multiple "peak configurations" that will result from random fluctuations around the theoretical positions of the peaks within their respective confidence intervals. Naturally, each configuration will have a different probability of being observed. All these configurations are scanned by Eq. (16), in which each integral examines all possible locations of solute k within its confidence interval.

Unfortunately, there is no way of computing the analytical solution of Eq. (16), since $\bar{p}_{i,j}$ is only numerically available. Also, the computation time of the trapezoidal integration of the ns-folded integral increases exponentially with the number of solutes involved, giving rise to unpractical computation times for chromatograms including more than two peaks, which is usually the case. However, finding the solution by means of a Monte-Carlo simulation is relatively simple. Let's suppose that we wish calculating $\bar{p}_{j,i}$. The Monte-Carlo assay is built in such a way that nl peak configurations are scanned. At each l configuration, the peak position of each compound is randomly shifted within its confidence interval, yielding a vector of ns retention times: $t_{l,1}, t_{l,2}, \ldots, t_{l,ns}$. The elementary peak purity for solute i (and its associated probability) is calculated at each of the nl configurations, according to Eqs. (5) and (13), respectively. Then, an approximation to $\bar{p}_{j,i}$ can be obtained as:

$$\bar{p}_{j,i} = \mu_{1j,i} \approx \frac{\sum_{l=1}^{nl} \left(p_{j,i}(t_{l,1}, t_{l,2}, \ldots, t_{l,ns}) \prod_{k=1}^{ns} \pi_{j,i}(t_{l,k})\right)}{\sum_{l=1}^{nl} \prod_{k=1}^{ns} \pi_{j,i}(t_{l,k})} \tag{17}$$

The values of $\mu_{1j,i}$, which are normalised, are multiplied to obtain a single value representative of the overall separation of all compounds at a given experimental condition. Since this approach is more rigorous, the derived purity measurement will be more reliable than in Approaches (i) and (ii). Note, however, that the precision of this approximation depends critically on the number of averaged configurations, nl.

2.4. Probability density function of the robust peak purity according to Approach (iii)

In Eq. (17), $\mu_{1j,i}$ (i.e. the averaged peak purity) is the first moment of what can be defined for solute i at the experimental condition j, as the "probability density function of peak purities" ($\rho_{pj,i}$), which would express the probability of obtaining a purity value between p and $p + \Delta p$ for that solute in the simulated experimental condition. The density function cannot be formally outlined, but is numerically accessible through Monte-Carlo simulations. Other moments different from the first one (i.e. the mean) can also be computed to characterise this distribution. For instance, the second moment is given by:

$$\mu_{2j,i} \approx \frac{\sum_{l=1}^{\mathrm{nl}} \left([p_{j,i}(t_{l,1}, t_{l,2}, \ldots, t_{l,\mathrm{ns}})]^2 \prod_{k=1}^{\mathrm{ns}} \pi_{j,i}(t_{l,k}) \right)}{\sum_{l=1}^{\mathrm{nl}} \prod_{k=1}^{\mathrm{ns}} \pi_{j,i}(t_{l,k})} \tag{18}$$

which is easily related to the standard deviation ($\mathrm{sp}_{j,i}$) of the probability distribution function [12]:

$$\mathrm{sp}_{j,i} = \sqrt{\mu_{2j,i} - \mu_{1j,i}^2} \tag{19}$$

Higher moments related to the skewness and kurtosis can be equally computed, but will not be considered here. The meaning of Eq. (19) will be commented in Section 4.

3. Experimental

3.1. Reagents

The probe compounds were six L-amino acids, obtained from several sources: arginine (Arg), asparagine (Asn), glutamine (Gln), histidine (His), serine (Ser), and threonine (Thr). A few drops of 1 M HCl (Panreac, Barcelona, Spain) were added to the amino acids to facilitate dissolution. In all cases, the concentration of the stock solutions was ca. 1.5×10^{-3} M, which was reduced to $(2.0–6.0) \times 10^{-6}$ M before the injection into the chromatograph. The derivatisation reagent contained *o*-phthalaldehyde, *N*-acetylcysteine (Fluka, Buchs, Switzerland), and boric acid/borate buffer, which was obtained by adding sodium hydroxide (AnalaR, Poole, UK) to a boric acid solution (Probus, Badalona, Spain) up to reach pH 9.5.

Aqueous–organic mobile phases were prepared with acetonitrile (HPLC grade, Scharlab, Barcelona), 1.0×10^{-2} M trisodium citrate dihydrate (Merck) and 0.1 M HCl, and filtered through 47 mm diameter membranes (Micron Separations, Westboro, MA, USA). The solutions of the OPA–NAC derivatives to be injected were also filtered through 17 mm diameter membranes of the same manufacturer and porous size (0.45 μm). All reagents were of analytical grade. Nanopure water was used throughout (Barnstead, Sybron, Boston, MA, USA).

3.2. Derivatisation and chromatographic procedures

The derivatisation reagent was prepared by dissolving OPA in a small amount of ethanol (Merck, Darmstadt, Germany), diluting it with the boric acid/borate buffer and NAC solution. The final concentrations were 2.5×10^{-4} M *o*-phthalaldehyde, 4.0×10^{-4} M *N*-acetylcysteine, and 0.1 M boric acid/borate buffer. The OPA–NAC reagent was renewed weekly and stored at 4 °C protected from light by covering it with an aluminium foil.

The amino acid derivatives were obtained by mixing an aliquot of each amino acid solution with 3 ml of the OPA–NAC reagent. The resulting mixture was diluted with water up to a final volume of 10 ml. After 10 min, 20 μl of this solution was injected into the chromatograph and eluted isocratically with acetonitrile–water. Acidity of the mobile phases was fixed at pH 6.5 with 5.0×10^{-3} M citric acid/citrate buffer. The concentration of acetonitrile in the experimental design, measured as volumetric fraction, was in the range 7.5–15%.

3.3. Apparatus and software

The chromatographic system, from Agilent (Model HP 1100, Waldbronn, Germany), was equipped with a quaternary

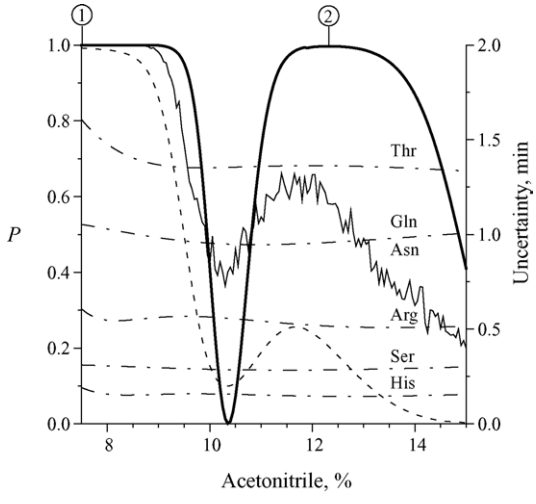

Fig. 1. Overall peak purity for the separation of the six amino acid derivatives according to Approach (i) (conventional definition, thick solid line), and the two robust approaches: Approach (ii) (dashed line), and Approach (iii) (nl = 1000, thin solid line). The uncertainty at 95% confidence level is overlaid (right axis, dotted-dashed lines) for each compound. The two resolution maxima are labelled as "1" and "2".

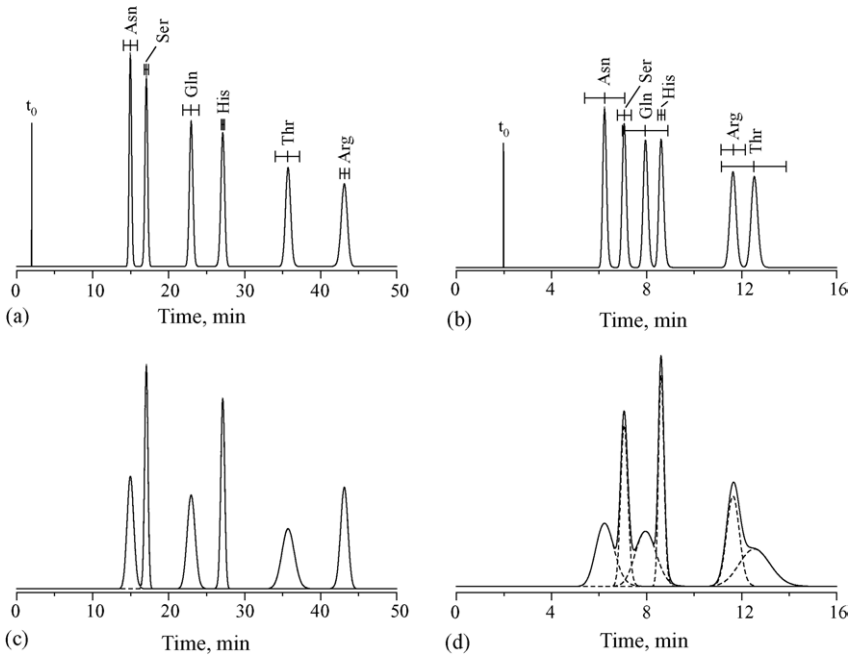

Fig. 2. Conventional (a and b) and oversized (c and d) chromatograms for the mixture of amino acid derivatives at the two most favourable compositions pointed out in Fig. 1: (a and c) 7.5% and (b and d) 12.3% acetonitrile. The 95% confidence intervals are overlaid on the top.

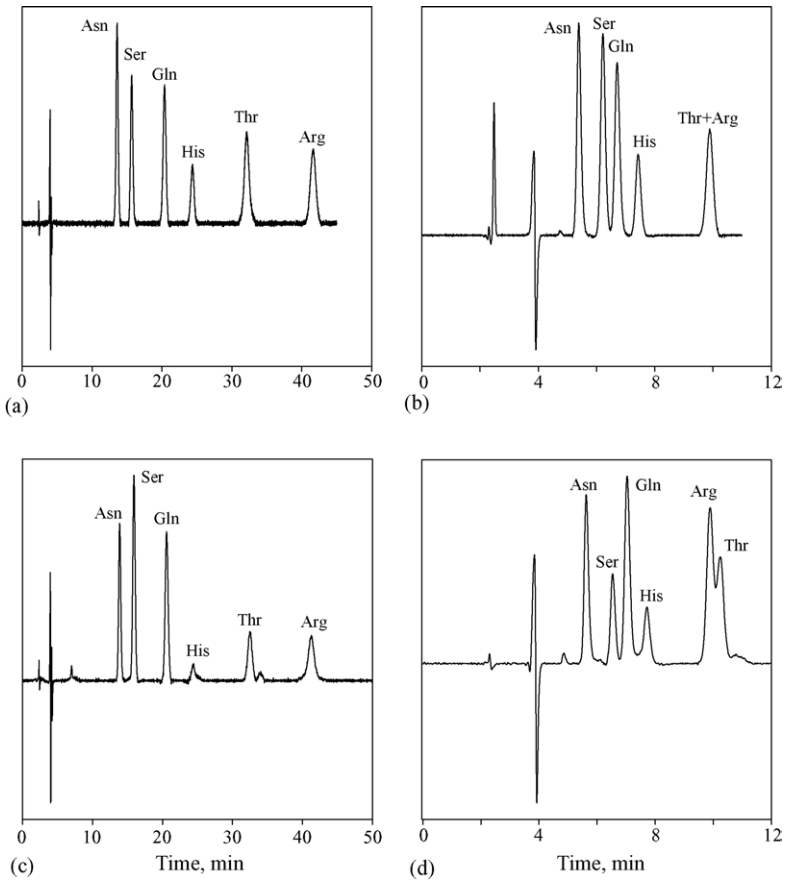

Fig. 3. Experimental chromatograms of the mixture of amino acids eluted with: (a and c) 7.5% and (b and d) 12.5% acetonitrile. Chromatograms a and b were taken during the modelling step, whereas chromatograms c and d were taken later, renewing all solutions although using the same instrument and column.

pump, a diode array detector, an automatic sampler, and a temperature controller. The separation was carried out with a 250 mm × 4.6 mm I.D. Inertsil ODS3 column (Análisis Vínicos, Tomelloso, Spain), connected to a 30 mm × 4.0 mm I.D. Kromasil C18 guard column (Scharlab). In both cases, the particle size was 5 μm. The experiments were run at a nominal constant temperature of 25.0 °C. The flow-rate was set to 1.0 ml/min and the dead time was measured from the first deviation of the baseline. Home built-in routines, written in MATLAB 6.5 (The Mathworks, Natick, MA, USA), were developed for data treatment.

4. Results and discussion

Fig. 1 depicts the values of global purity for the conventional, oversized and probabilistic approaches, in the optimisation of the resolution of the six amino acid derivatives. The uncertainties in peak position for each compound are overlaid.

As can be seen, the conventional (i.e. unrobust) optimisation approach yields two regions of maximal overall purity at 7.5% and 12.3% acetonitrile (labelled as 1 and 2 in Fig. 1), with $P \approx 1$, which indicates that baseline resolution can be almost achieved for all peaks. The P value obtained according to the robust approaches (ii and iii) appraises both situations in a completely different way. The overall purity for situation 2 is dramatically decreased, whereas for situation 1, it remains unaltered, close to 1.0 (i.e. denoting full resolution). The analysis of Fig. 2a and b, which show the optimal chromatograms for both situations, can explain this. The 95% confidence intervals for the predicted retention times are depicted as segments overlaid above each peak. At 12.3% acetonitrile (Fig. 2b), these intervals are rather large compared with the peak width. Thus, although for the optimal chromatogram there is an apparent baseline separation, a significant risk of overlapping is present, since the prediction model is not precise enough with regard to the peak separation. At 7.5% acetonitrile (Fig. 2a), peaks are eluted at longer retention times. This longer elution compensates the increment in peak position uncertainties (note that the scale in Fig. 2a and b is not the same, which masks the increased uncertainties). The final effect is that the relative importance of the uncertainties decreases at the region of lower acetonitrile content in the experimental design.

Approaches (ii) and (iii) appraise the consequences of uncertainties in peak position on the chromatographic resolution, by decreasing the expected separation capability for those situations where the uncertainties make the objective function too sensitive to variations in the experimental factors. Fig. 2c and d illustrate in more detail how Approach (ii) works. In this figure, the "oversized chromatogram" is plotted for 7.5% and 12.3% acetonitrile. The former condition is clearly more robust according to the oversized chromatogram criterion: position uncertainties are not translated in peak merging, since peak separation is sufficient. In contrast, for the latter condition, the smaller distance between neighbouring peaks, together with the greater band broadening led to a decrease in peak purities in the oversized chromatogram, with regard to the values obtained with the conventional optimisation strategy.

In Fig. 1, a minimum in overall purity is observed at 10.45% acetonitrile, which denotes the peak crossing of arginine and threonine. Comparing the conventional approach with the robust ones, it can be observed that the former gives rise to more extreme values, with purity measurements ranging between zero and one, corresponding to full overlapping and full resolution, respectively. This contrasts with the qualifications provided by the robust approaches, which are significantly more moderate: situations of full resolution are

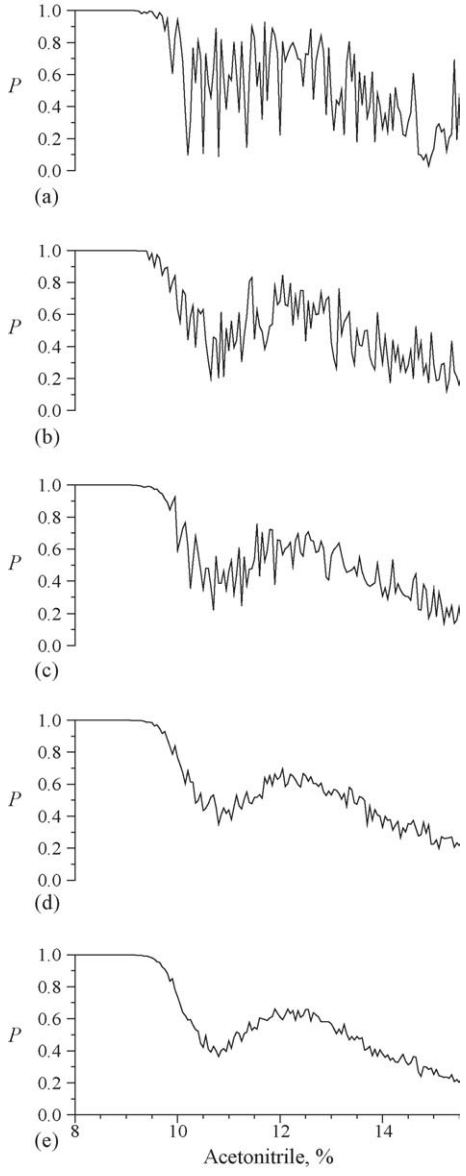

Fig. 4. Effect of the number of simulated replicates (nl) on the estimation of a robust purity map (Eq. (17)) for the mixture of amino acid derivatives, according to Approach (iii): (a) 10, (b) 50, (c) 100, (d) 500, and (e) 1000.

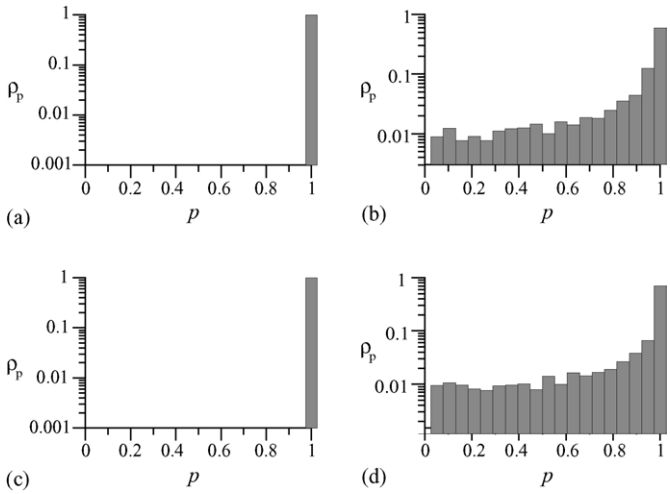

Fig. 5. Probability density function (ρ_p) of peak purity for glycine (a and b) and threonine (c and d) at the two maxima labelled in Fig. 1: 7.5% (a and c) and 12.3% (b and d) acetonitrile. A value of nl = 20 000 was used to increase the precision.

spoiled by peak shifts (i.e. the associated purities are smaller than the conventional ones), but also full overlapping can be benefited by the effects of peak shifts (i.e. the associated purities become greater). In this way, peak shifts coming from uncertainties associated to the predictions are detrimental in situations of full resolution (i.e. isolated peak clusters may merge), which makes robust resolution measurements smaller than the conventional ones. On the contrary, situations of full overlapping are not so negatively qualified by robust measurements, since shifts from the predicted positions will separate the cluster, enhancing the resolution. Hence, robust measurements appraise chromatographic separations in less extreme terms.

Fig. 3 shows experimental chromatograms corresponding to 7.5% and 12.5% (v/v) acetonitrile, using the same instrument and column, but renewing the solutions of amino acids, reagents and mobile phases. Chromatograms a and b were taken during the modelling step. Note that the latter composition is slightly different from that giving the maximal resolution (Fig. 2), but belongs to a region where the resolution scarcely varies. From Figs. 2 and 3, it can be seen that the reliability of the 7.5% acetonitrile mobile phase is larger than that of 12.5%, which agrees with the conclusion of the proposed robust optimisation methods.

Strictly speaking, only Approach (iii) is able to evaluate correctly how the uncertainty in peak position is propagated into the value of the chromatographic objective function. Approach (ii) does not fit properly in the probabilistic theory: the "oversized chromatogram" is an abstraction: it actually does not exist. Approach (ii) should be considered, thus, just as an approximation to Eq. (17), and its results should be interpreted accordingly.

The main problem of Approach (iii) is the increased computation time when high precision is required, which is controlled through the value of nl in Eq. (17). Fig. 4 shows the

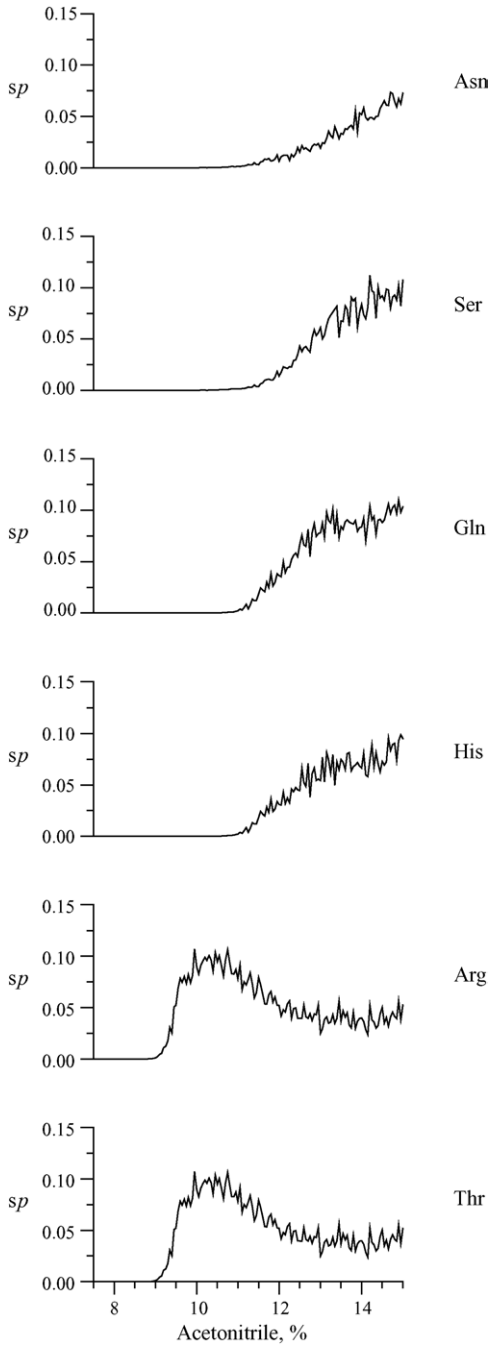

Fig. 6. Standard deviation in the elementary peak purities computed according to Eq. (19), as a function of mobile phase composition.

overall purity profiles obtained with Eq. (17), at progressively greater values of nl. The computation time with a Pentium IV 2.40 GHz computer was (nl, min): (10, 0.5); (50, 2.1); (100, 4.0); (500, 19.5); (1000, 39.3). A proper selection of nl should attend the number of involved peaks and the peak position uncertainty/peak width ratio: the higher this ratio, the higher the value of nl.

As commented previously, the useful information that can be extracted from the Monte-Carlo assay can go beyond a mere evaluation of the robust peak purity. Fig. 5 shows the

probability density function of the peak purity for glycine and threonine at the two critical points of the resolution map, depicted in Fig. 1. The number of configurations at 7.5% and 12.3% acetonitrile was the same: nl = 20 000. As can be seen, both situations yield the maximal value of the estimation of $\overline{p}_{i,j}$ (i.e. $\mu_{1,i,j}$) = 1. Meanwhile, the distributions are broader at 12.3% acetonitrile, which explains why the mean value of the distribution (Eq. (17)) decreases.

An estimation of the consequences of the uncertainties committed in retention time predictions on the elementary purities is presented in Fig. 6, where Eq. (19) is plotted vs. mobile phase composition. A sudden increase in the standard deviation is observed at certain concentrations that are inversely correlated with the solute polarity. Above these concentrations, the precision in the modelling step is not enough to assure accurate values in peak purity for the considered compound.

When the peaks are close enough, the uncertainty in retention time starts to affect the peak purity, and the values of $sp_{j,i}$ increase suddenly. The particular behaviour of arginine and threonine is due to the coelution at 10.45% acetonitrile, which exalts the effect of uncertainties, giving rise to a maximum at that composition. At higher elution strengths, coelution decreases, which brings as consequence a diminution of the influence of prediction errors on peak purity. Around 12.3% acetonitrile, all compounds present some uncertainty. Therefore, the reliability of the maximum labelled as "2" in Fig. 1 is doubtful under the point of view of robustness.

5. Conclusions

The scope of robustness in the field of the optimisation of chromatographic resolution goes beyond its conventional application. This concept has been frequently restricted to a simple quantification of the effects that deliberated errors committed in the transference of the optimal conditions to the chromatograph produce on the quality of the separation. These effects are frequently established experimentally, but can be also obtained from resolution surfaces, by simulating mispreparations. The latter strategy is of interest in the initial steps of method development, whereas the former is mainly put in practice in the final validation step. However, errors in resolution surfaces may come from other sources. This is the case of the errors originated by the propagation of uncertainties associated to the fitting of the retention models, whose consequences can be significantly larger. Ignoring the effects of these uncertainties may be translated in deceiving predictions: an apparently perfect baseline separation can be found impossible to reproduce in the chromatograph.

Two approaches are here proposed to implement the effects of these uncertainties on resolution. Both of them yielded more reliable results than conventional resolution assessments. Approach (ii) was found to be less optimistic and formally less rigorous than Approach (iii). In the absence of more considerations, Approach (ii) is recommended only for a fast evaluation of the robustness of resolution (it took less than 10 s in the examined example). However, since more detrimental effects worsening predictions are present in practice (e.g. mispreparation of the mobile phase, temperature effects, and pump fluctuations), a pessimistic approach scoring positively a given separation condition will be more reliable.

In the construction of retention models, we have two alternatives: fitting the data without weights to get heteroskedastic error distributions, which will yield worse predictions for eluents with low modifier contents, or fitting the data using weights, which will yield homoskedastic error distributions. In comparison to the former method, the latter gets a general decrease in the prediction error, enhancing greatly the predictions at low modifier contents although worsening the predictions for eluents with high modifier contents. Now, it is necessary to make a decision about committing larger errors with low or high modifier contents. Since the probability of getting resolution is larger at intermediate and low modifier contents (owing to the smaller probability of getting peak cluttering), the logical conclusion is to use weights as a general rule. The introduction of weights will affect the fitting, and therefore, the robust approaches. In the case of leaving the data unweighted, the influence will be larger for faster eluents, but these are the least interesting under the point of view of resolution, and what is worst: they will lead to unrobust predictions for those mobile phases yielding larger retention, and thus, resolution.

The main advantage of Approach (ii) is the simplicity of computation. Approach (iii) requires a significantly higher computational effort, correlated to the value of nl used in Eq. (17). The precision of the calculations with Approach (iii) can be compromised if a too low nl value is selected. Therefore, the user should find a balance between the precision of the Monte-Carlo assay and the cost in terms of computation time.

As a general recommendation, we suggest applying Approach (ii) first in order to get an overview of the robust resolution map and the influence of incidental composition errors, and then use Approach (iii), restricted to the maximal robust resolution regions. Both approaches can be implemented with other resolution assessments, different to peak purity.

Acknowledgements

This work was supported by Project CTQ2004–02760/BQU (Ministerio de Educación y Ciencia of Spain) and Groups Grant 04/16 (Generalitat Valenciana). VCH, who is on leave from the University of Zacatecas (Mexico), thanks a Ph.D. fellowship from PROMEP (Mexican Government). GVT thanks the Generalitat Valenciana for an FPI grant. JRTL thanks the MCYT and the Generalitat Valenciana for a Ramón y Cajal position.

References

[1] P.F. Vanbel, B.L. Tilquin, P.J. Schoenmakers, J. Chromatogr. A 697 (1995) 3.
[2] H. Fabre, J. Pharm. Biomed. Anal. 14 (1996) 1125.
[3] Y. Vander Heyden, D.L. Massart, Y. Zhu, J. Hoogmartens, J. DeBeer, J. Pharm. Biomed. Anal. 14 (1996) 1313.
[4] Y. Vander Heyden, K. De Braekeleer, Y. Zhu, E. Roets, J. Hoogmartens, J. De Beer, D.L. Massart, J. Pharm. Biomed. Anal. 20 (1999) 875.
[5] S. Goga-Remont, S. Heinisch, J.L. Rocca, J. Chromatogr. A 868 (2000) 13.
[6] J.R. Torres-Lapasió, D.L. Massart, J.J. Baeza-Baeza, M.C. García-Alvarez-Coque, Chromatographia 51 (2000) 101.
[7] A.M. Siouffi, R. Phan-Tan-Luu, J. Chromatogr. A 892 (2000) 75.
[8] P.J. Schoenmakers, H.A.H. Billiet, R. Tijssen, L. de Galan, J. Chromatogr. 149 (1978) 519.
[9] L.R. Snyder, J.W. Dolan, D.C. Lommen, J. Chromatogr. 485 (1989) 65.
[10] T. Baczek, M. Markuszewski, R. Kaliszan, M. van-Straten, H.A. Claessens, J. High Resolut. Chromatogr. 23 (2000) 667.
[11] A. Pappa-Louisi, P. Nikitas, Chromatographia 57 (2003) 169.
[12] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, 1998.
[13] S.S. Rao, Optimisation: Theory and Applications, Wiley, New York, 1984, pp. 274–283.
[14] J.J. Baeza-Baeza, G. Ramis-Ramos, Anal. Chim. Acta 316 (1995) 173.
[15] J.R. Torres-Lapasió, M. Rosés, E. Bosch, M.C. García-Alvarez-Coque, J. Chromatogr. A 886 (2000) 31.
[16] G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, J. Chromatogr. A 876 (2000) 17.
[17] J.R. Torres-Lapasió, J.J. Baeza-Baeza, M.C. García-Alvarez-Coque, Anal. Chem. 69 (1997) 3822.
[18] G. Vivó-Truyols, J.R. Torres-Lapasió, A.M. Van Nederkassel, Y. Vander Heyden, D.L. Massart, J. Chromatogr. A 1096 (2005) 133.
[19] S. Carda-Broch, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Anal. Chim. Acta 396 (1999) 61.
[20] N.R. Draper, H. Smith, Applied Regression Analysis, third ed., Wiley, New York, 1998.