Journal of Chromatography, 385 (1987) 181–200
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 962

EVALUATION OF PEAK-RECOGNITION TECHNIQUES IN LIQUID CHROMATOGRAPHY WITH PHOTODIODE ARRAY DETECTION

J. K. STRASTERS*, H. A. H. BILLIET and L. DE GALAN
Laboratorium voor Analytische Scheikunde, Technische Hogeschool, Delft, De Vries van Heystplantsoen 2, 2628 RZ, Delft (The Netherlands)
and
B. G. M. VANDEGINSTE and G. KATEMAN
Laboratorium voor Analytische Scheikunde, Katholieke Universiteit Nijmegen, Toernooiveld, 6525 ED, Nijmegen (The Netherlands)

SUMMARY

Four mathematical techniques concerned with peak recognition in liquid chromatography by means of photodiode array detection are evaluated for application in the iterative regression design optimization of the mobile phase. The techniques are multicomponent analysis, multicomponent analysis with a non-negativity constraint, target factor analysis and iterative target testing.

The results obtained for two-component clusters depend to a different degree on chromatographic resolution and changes in the UV spectra caused by a variation of the eluent composition. Multicomponent analysis is independent of resolution but highly sensitive to minute changes in the spectra. Iterative target testing needs no spectral information, but requires a reasonable resolution. Provided adequate spectra are available, target factor analysis allows excellent recognition down to very low resolution.

INTRODUCTION

The optimization of a liquid chromatographic separation can be performed by a systematic variation of the composition of the eluent. When the optimization is performed by an iterative regression design[1,2] with the advantage of a limited number of chromatograms, it is essential that the retention times of all solutes in each chromatogram are known. This can be achieved by separate injection of all solutes, provided they are known and available. However, a more efficient method can be developed when corresponding solutes can be recognized directly in consecutive chromatograms of the sample. Such recognition is prerequisite when we are dealing with an unknown sample.

A one-dimensional detection system, such as single-wavelength UV or RI detection, does not provide enough information to be used in a recognition procedure. The limitations of dual-wavelength detection, *i.e.* the ratio method, have been re-

0021-9673/87/$03.50

ported[3]. An extension towards detection systems with a higher dimensionality, such as the multi-wavelength linear photodiode array detector, is indicated[4,5].

In a previous publication[6] an approach has been described based on visual evaluation and comparison of spectra collected during elution of a mixture using mobile phases with different compositions. As a first step towards further automation of the optimization scheme, we will now attempt to evaluate the observed chromatograms by means of mathematical techniques.

The ensuing problems can be divided into two categories. First, in the case of a complete separation, we have a direct comparison of spectra of pure components, either mutually or with reference spectra. Contrary to library searches with reduced (coded) IR or mass spectra, a more extensive comparison between the spectra is needed owing to a lack of specificity of UV spectra. Since we are dealing with libraries of limited size, the time required for the comparisons is not an important factor. Previously, direct correlation has been used to compare UV spectra. Wegener *et al.*[7] have used several statistical techniques and direct correlation for the identification of cosmetic dyes. Fell[8] proposed the use of higher order derivatives to emphasize small differences between the spectra.

Second, when optimizing a chromatographic system, complete separation of all components is unlikely in the first chromatograms recorded. Hence, poorly resolved components must be recognized either by matching mixed spectra with those from a library, or by extracting pure component spectra from the measured mixture spectra. When the peaks of the chromatogram under consideration are represented by sets of spectra one can perform an analysis of variance, thus avoiding strict assumptions on peak-shape. Again two situations arise: either all contributing solutes (in a peak or in the sample) and their spectra are known, or for one or more components no spectrum is available in the set of reference spectra. In the latter case there is a possibility that the unknown spectra can be derived from another chromatogram of the same sample, or one can use one of the deconvolution techniques currently available[9–11].

Here we are concerned with the performance of certain well known techniques, such as multicomponent analysis[12,13] and target factor analysis[14,15], both utilizing spectra from a library, and the iterative target transformation analysis[10], which can be used when no preliminary information is available. Special emphasis is given to the two major problems encountered, *i.e.* the influence of the chromatographic resolution and the change in the spectral characteristics of the solutes due to a change in the mobile phase composition.

A separate problem is the determination of the actual number of components involved in each cluster of peaks. The answer to this question is usually derived by performing an analysis of variance, *i.e.* principal component analysis (PCA) or factor analysis (FA) coupled with a discrimination criterion[14]. The best results in this respect are achieved by using the technique of cross-validation[16]. A residual error function is generally adequate when the experimental error is well known and normally distributed.

THEORY

In this section we will briefly discuss some well known mathematical and sta-

tistical techniques with special emphasis on the problems mentioned in the introduction.

Direct comparison of UV spectra

When we want to recognize more or less completely separated component peaks in different chromatograms, the problem is reduced to a direct comparison of UV spectra. Since the recorded spectra are in digital form and show little fine structure, a comparison of all corresponding absorbances expressed in a correlation coefficient will yield the best results[17,18]. For systems with a relatively small number of components the required calculation time is of less importance. The correlation coefficient r can be expressed as follows:

$$r = \frac{(\Sigma x_i y_i - \Sigma x_i \Sigma y_i/n)}{\sqrt{\{(\Sigma(x_i)^2 - (\Sigma x_i)^2/n) \cdot (\Sigma(y_i)^2 - (\Sigma y_i)^2/n)\}}} \quad (1)$$

where x_i and y_i represent the absorbances of the spectra x and y measured at wavelength i. Introduction of the denominator into eqn. 1 normalizes the spectra in such a way that the sum of squared absorbances for each spectrum equals 1. The correlation coefficient thus compares spectra for their shape, but not for their magnitude. Other statistical tests employed[7] are closely related to the correlation coefficient. For instance the sum of squares, SS, of the differences between two normalized and averaged spectra can be expressed as:

$$SS = 2 \cdot (1 - r) \quad (2)$$

The following sections are concerned with chromatographic peaks observed for two or more poorly resolved components, thus requiring a more extensive pretreatment of the data, known as spectral deconvolution.

Multicomponent analysis (MCA)

When we have a spectral description of all possible components contributing to the observed spectra, the application of a multicomponent analysis is straightforward. After a general description with regard to spectral analysis by Blackburn[12], it has been extensively used for different applications in the past two decades[13]. Using a matrix notation the general problem can be described by

$$S \cdot C = A \quad (3)$$

where S is a $(w \times n)$ matrix with n columns of reference spectra defined for w wavelengths and A is a $(w \times t)$ matrix with t columns representing mixture spectra recorded during the elution of a chromatographic peak. These two matrices are coupled through the $(n \times t)$ matrix C containing the contributions of the individual reference spectra needed for a reproduction of the measured spectra after summation. In an ideal situation the n rows of matrix C will either contain zeroes, indicating the absence of a component, or follow a more or less gaussian elution profile, since the measured spectra are ordered with respect to time.

MCA is straightforward when the number and nature of the reference spectra agree with the actual components present in the elution profile. In the more realistic situation that the number of candidate reference spectra exceeds the number of components, we ask from MCA that it uses (and hence selects) out of that larger number only those solutes actually present in the profile.

The least-square solution for eqn. 3 can be expressed as the pseudo-inverse:

$$C = (S^T \cdot S)^{-1} \cdot S^T \cdot A \tag{4}$$

Various alternatives to determine matrix C, based on different transformations of the data, have been proposed[19].

Experimental conditions can introduce deviations in the measured spectra, which will produce errors in the estimated contributions of all library spectra used. In general, random errors will be small and will influence the calculated concentrations of all components more or less equally. Systematic errors such as a shift will tend to influence only the estimated concentrations of certain components, thus suggesting the presence or partial absence of certain components.

Burns[20] improved the accuracy of MCA by imposing a certain peak-shape, something we prefer to avoid. He also applied an algorithm of Lawson and Hanson[19] to prevent the occurrence of negative concentrations. Since these negative concentrations are often introduced in connection with erroneous positive amounts of other components, application of this non-negative least-squares solution will improve the overall results.

Target factor analysis (TFA)

Because MCA considers in turn every mixture spectrum recorded during elution, all pure-component spectra must be available collectively. In contrast, we may test for the presence of individual components sequentially by an evaluation of the variation in subsequent mixture spectra. This technique, an application of factor analysis, is described by Malinowski (among others), both in general[14] and for high-performance liquid chromatography with UV detection (HPLC–UV)[15].

Again, we use the general description (eqn. 3) as derived by the application of the linearity principles of the Lambert–Beer law. By eliminating those components not present in the set of spectra under consideration we reduce the matrices to S′, a ($w \times m$) matrix consisting of the spectra of the m components actually present, and C′, a ($m \times t$) matrix containing the elution profiles of those m components:

$$A = S' \cdot C' \tag{5}$$

In this way we can describe all observations (the absorptions of matrix A) with only a limited number of elements (the spectra and elution profiles in the matrices S′ and C′).

When we regard the spectra as t points in a w-dimensional space, with the absorptions observed at different wavelengths represented on the respective axes, only a limited part of the total space can be occupied by spectra resulting from mixtures of a limited number of components, owing to the relations between the observed absorbances at different wavelengths for the same component. When only

one component is present, all points are situated on a straight line through the origin, since the concentration varies but the relative absorptions do not. Similarly two components will define a two-dimensional surface and m components an m-dimensional hyperspace.

Through mathematical techniques (principal component analysis, factor analysis) we can derive a description of the m-dimensional hyperspace connected with a set of observations. This abstract description has the same structure as eqn. 5 (and is therefore a form of data-reduction) and can be transformed to the true-component spectra or elution profiles:

$$A = R \cdot E^T = R \cdot T \cdot T^{-1} \cdot E^T = S' \cdot C' \quad (6)$$

R is an ($w \times m$) matrix, giving a general description of all spectra present in the mixture: every spectrum of the original set can be derived from a linear combination of the columns of R. This holds also true for the pure-component spectra, hence an appropriate transformation of R through the ($m \times m$) matrix T will result in the calculation of S'. The same applies for the ($t \times m$) matrix E containing the abstract description of the elution profiles (as columns).

The actual target test determines whether we can expect a given pure-component spectrum to be situated in the hyperspace described by R. This can be achieved by projecting a known spectrum onto the hyperspace and comparing the projection and the original spectrum. If they resemble each other closely enough the component is thought to be present. From a chromatographer's point of view, the problem is that all components must be identified before the corresponding elution profiles can be found from the inverse transformation T^{-1}, since the calculation of every row of T^{-1} depends on all columns of T. The individual calculation of these columns is performed by:

$$t = (R^T \cdot R)^{-1} \cdot R^T \cdot s \quad (7)$$

where s represents a spectrum from the collection of reference spectra. Mark the similarity between eqns. 4 and 7, indicating that MCA is nothing but a projection of the unknown spectrum in the hyperspace defined by the reference spectra.

Iterative target transformation (ITT)

The methods described thus far have as a major limitation that all pure-component spectra connected with a chromatographic peak should be available before the individual elution profiles can be determined. More often than not the chromatographer is only partly aware of what is present in a mixture. By using the general description of eqn. 6 and imposing a number of boundary conditions (non-negative concentrations and individual absorptions), one can approximate the original spectra by means of an extrapolation of the observed variation, either in the abstract description of the spectra or the elution profiles. Applications of this so-called "self-modelling curve resolution" have been described for two- and three-component profiles[11].

Recently, a more general approach for the HPLC–UV combination was described by Vandeginste *et al.*[10], the so-called "iterative target transformation". The

method is applicable to clusters of more than three components and does not require any spectral knowledge with regard to the pure-component spectra. The only assumptions made are connected with the shape of the elution profile: it should be non-negative at all times and exhibit an unimodal distribution (only one maximum).

The method may be briefly described as follows. After a rotation of the abstract description of the elution profile (Varimax) a first approximation of the contributing solute-profiles is derived. This approximation determines the first target to use in a target test, equivalent to eqn. 7 but using the matrix E instead of R:

$$t = (E^T \cdot E)^{-1} \cdot E^T \cdot c_1 \tag{8a}$$

$$c_1' = E \cdot t \tag{8b}$$

where c_1 represents the first target and c_1' its projection. The resulting projection is refined according to the demands formulated above and again subjected to a target test. This process is repeated until no further refinement is possible or no further iteration is observed. In this way the elution profiles of all components present in the peak are determined. Having thus found E, we can use the inverse transformation (eqn. 6) to determine the corresponding pure-component spectra.

EXPERIMENTAL

Instrumentation

The chromatographic experiments were performed using a Novapak C_{18} column (15 cm × 3.9 mm I.D., 5 μm particles) and a M6000A pump, both from Millipore Waters (Milford, MA, U.S.A.). The detector was the HP-1040A fast-scanning LDA detector (Hewlett-Packard, Waldbronn, F.R.G.) connected to an HP-85 desktop computer, equipped with input/output, plotter/printer, mass storage and advance programming ROMs, 16 kbyte additional memory, HP-IB IEEE-488 interface and RS-232C serial interface. The data were temporarily stored on 5¼ in. flexible disks using a HP82910M disk-drive.

The collection of recorded spectra was transferred from the HP-85 to a PDP11/03 system (Datacare, Zeist, The Netherlands) by means of the serial interfaces on both computers. All calculations were performed on the PDP11/03, which was equipped with two 8 in. disk-drives, a RD51 hard disk unit, 4006-1 Computer Display Terminal (Tektronix, OR, U.S.A.) and HP7470A graphics plotter with serial interface.

Chromatographic data

For the calculations to be described in the next section we used spectra and chromatograms obtained for eight chlorinated phenols. The identities, applied concentrations and some retention times are listed in Table I. Throughout the following discussion the components will be referred to by their number in this table. With the basic optimization method in mind, we collected spectra in seven mobile phases of approximately isoelutropic composition, by separately injecting all components and storing the recorded spectra on the upslope, apex and downslope of the detected peaks. Although the spectra were recorded between 190 and 400 nm, only the interval

TABLE I

IDENTITIES OF THE CHLORINATED PHENOLS

Listed are the concentration C used in the experiments and the retention times observed in acetonitrile–water (35:65) (t_{R1}) and tetrahydrofuran–water (35:65) (t_{R2}). The water was acidified with phosphoric acid (0.001 M).

No.	*Component*	*C (mg/ml)*	t_{R1} *(min)*	t_{R2} *(min)*
1	*p*-Chloro-*o*-cresol	0.49	5.87	7.70
2	2,5-Dichlorophenol	0.25	5.84	9.46
3	*p*-Chloro-*m*-cresol	0.25	4.92	6.48
4	2,3-Dichlorophenol	0.25	5.21	6.48
5	3,5-Dichlorophenol	0.50	7.96	13.33
6	2,4-Dichloro-5-methylphenol	0.50	9.35	11.40
7	*p*-Chlorophenol	0.26	3.50	5.32
8	*o*-Chlorophenol	0.25	3.10	4.33

between 230 and 400 nm was used, owing to excessive noise at lower wavelengths caused by the absorption of the mobile phase. The spectra in acetonitrile–water (35:65) are presented in Fig. 1.

In addition, mixture spectra were recorded during the elution of chromatograms of various resolutions of components 5 and 6. The corresponding mobile phase compositions are listed in Table II. The elution profiles recorded at 230 nm are

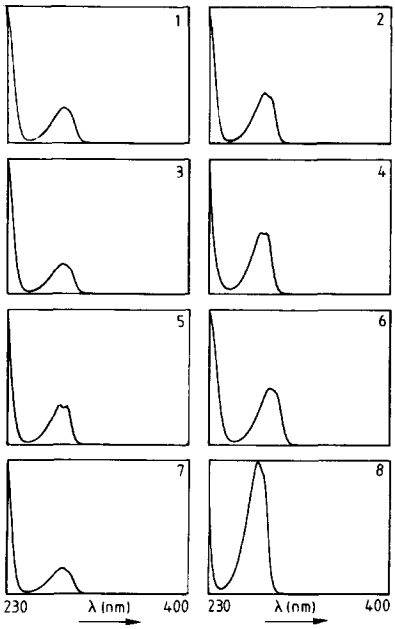

Fig. 1. The UV spectra of eight chlorinated phenols listed in Table I. The components were dissolved in acetonitrile–water (35:65).

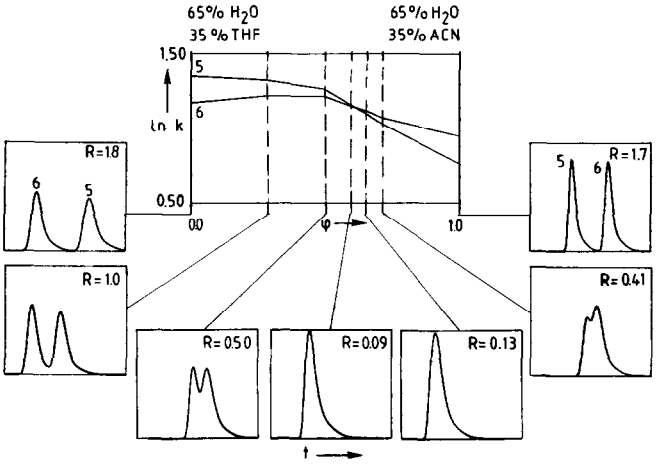

Fig. 2. The retention behaviour of components 5 and 6 as a function of the mixing ratio φ listed in Table II, coupled with chromatograms recorded during elution with the indicated mobile phase compositions. The chromatographic resolution is indicated with R. THF = tetrahydrofuran, ACN = acetonitrile.

displayed in Fig. 2. For a mathematical treatment of the data we selected 45 evenly spaced mixture spectra across every profile.

Software

The software used in the analysis of the spectrum-clusters and for the comparison of spectra was written in Fortran IV. The applied algorithm for the multicomponent analysis was directly derived from the pseudo-inverse (eqn. 3). The algorithm for the non-negative version of the MCA is described by Lawson and Hanson[19]. The software for the application of the target test was developed using the description by Malinowsky and Howery[14] and was extended to the iterative target test according to Vandeginste *et al.*[10].

As far as internal memory allowed, the matrix calculations were used from the Scientific Subroutine Package from DEC (Marlboro, MA, U.S.A.), as was the Varimax subroutine. In all other cases the calculations were performed by simple algorithms, if necessary with intermediate storage on disk. The eigenvalues and eigenvectors of the covariance-matrix of the data-matrix were determined by the HQRII algorithm[21].

All calculated elution profiles were expressed as quantities of pure-component spectra, normalized such that the sum of all squared absorptions of a pure-component spectrum equals 1. This means that the vectors connected with these pure-component spectra have unit length. In the case of the ITT we transformed the calculated spectra and elution profiles in such a way that they complied to this condition as well.

RESULTS AND DISCUSSION

In order to evaluate the above techniques with respect to spectral recognition and determination of retention times, there are three major influences to be con-

TABLE II

MIXING RATIOS φ AND CORRESPONDING MOBILE PHASE COMPOSITIONS, USED TO ELUTE THE MIXTURE OF COMPONENTS 5 AND 6 AND RECORD THE SPECTRA OF ALL CHLORINATED PHENOLS LISTED IN TABLE I

The water was acidified with phosphoric acid (0.001 *M*).

φ	*Acetonitrile (%)*	*Tetrahydrofuran (%)*	*Water (%)*
1.000	35	0	65
0.714	25	10	65
0.657	23	12	65
0.600	21	14	65
0.500	17	18	65
0.286	10	25	65
0.000	0	35	65

sidered: first, the dependence of the spectral characteristics on the mobile phase composition; second, the amount of chromatographic resolution; and third, the spectral similarity of the components involved (spectral resolution). The first two factors can be studied and varied systematically. The spectral similarity is inherent in the group of components under examination (Fig. 1). Except for solute number 8, the correlation coefficient between pure-component spectra varies from 0.85 to 0.998. The influence of chromatographic resolution was investigated for the two-component system containing the chlorinated phenols 5 and 6 (Table II, Fig. 2), which cross-over when going from a 35% acetonitrile binary to a 35% tetrahydrofuran binary. Unavoidably, the variation in chromatographic resolution thus achieved is accompanied by a variation in spectral characteristics as a result of the changing mobile phase.

To avoid differences in injection volumes, sampling times and chromatographic reproducibility, reference elution profiles for components 5 and 6 were not derived from separate injections of the solutes. Instead, we used the theoretically most reliable profiles to be found from spectral deconvolution: the profiles resulting after MCA using only the spectra of components 5 and 6 recorded in the same mobile phase

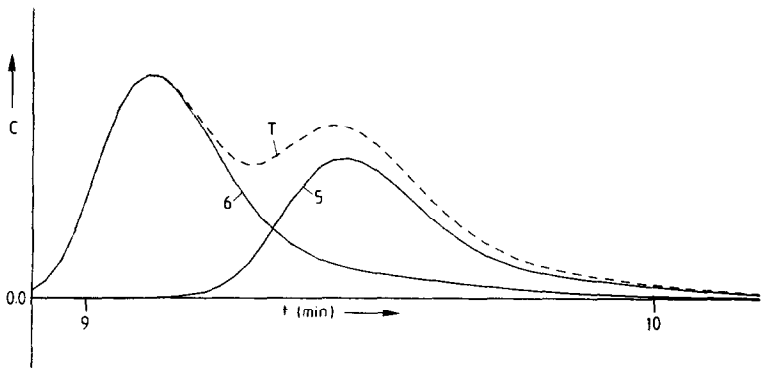

Fig. 3. The recorded chromatogram, T, of a mixture of components 5 and 6, and the individual component profiles calculated by means of MCA, using only the correct pure component spectra of 5 and 6. Mobile phase: acetonitrile–tetrahydrofuran–water (17.5:17.5:65) (φ = 0.5).

used to elute the mixture. An example is presented in Fig. 3. Consecutive elution profiles resulting from the application of the different spectral deconvolutions were compared with the reference profiles in the following way: the sum of squared differences was calculated after normalizing both profiles to the norm of the reference profile. In this way not only the shape but also the observed deviations in the amount of absorbance were involved in the evaluation.

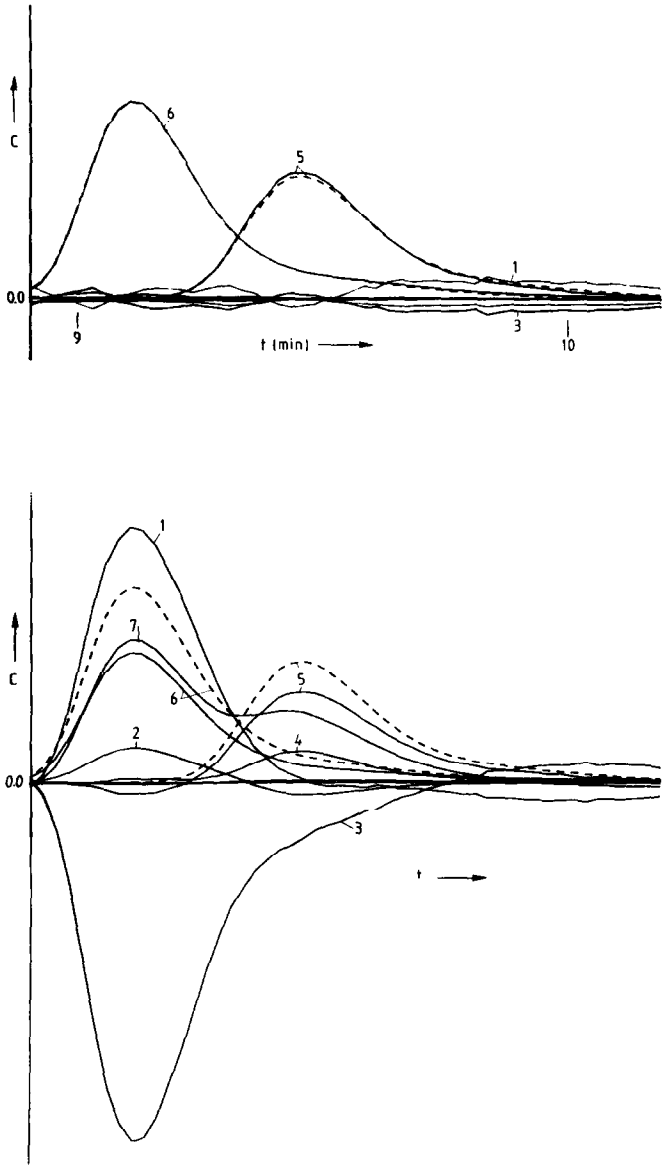

Fig. 4. The true profiles (---) of components 5 and 6 as determined for $\varphi = 0.5$ and the individual component profiles (——) calculated by means of MCA, using the spectra of all eight chlorophenols. (a) Profiles calculated using the correct spectra, recorded in the same solvent used to elute the mixture. (b) Profiles calculated using approximate spectra, recorded in a different solvent ($\varphi = 0$).

The influence of the mobile phase composition

First, we investigated the performance of MCA using an extended set of pure-component spectra, a situation that would occur when we know which components are present in the mixture, but when we do not know where they are situated in the chromatogram. This method, however, is very susceptible to experimental and systematic errors, especially in the case of large sets of reference spectra.

Fairly good results are obtained when we perform the MCA with the reference spectra of all chlorinated phenols, recorded in the same mobile phase used to elute the mixture. The results for the mixture containing 17% acetonitrile and 18% tetrahydrofuran (φ = 0.5) are displayed in Fig. 4a. As we can see, there is only a minor disturbance from components not present in the mixture, caused by experimental errors in the measured spectra and mathematical round-off. Attention is drawn to the reverse profiles of components 1 and 3: the spectra are so much alike that a difference of a positive and negative contribution is used to describe the observed experimental error.

A systematic error can be introduced by a difference in composition between the mobile phase used to elute the mixture, and the solvent used for the determination of the pure component spectra. Fig. 5 shows the difference between the spectra of components 5 and 6 in acetonitrile–water (35:65) and tetrahydrofuran–water (35:65), respectively. A small shift of 1 to 2 nm can be observed.

Because of this influence, quite a different picture emerges when we perform the same calculations on the profile recorded for φ = 0.5 with the spectra recorded in 35% tetrahydrofuran (φ = 0.0) (Fig. 4b). The small shift in the spectra causes large errors in the estimated concentrations, both positive and negative, because the procedure tries to eliminate the differences between measured spectra and linear combinations of reference spectra with additional contributions of the other components. As a consequence we can no longer determine which or even how many components are present in this cluster.

Some improvement can be obtained with the non-negative MCA, as is shown in Fig. 6a. Again we used the reference spectra in 35% tetrahydrofuran but, by

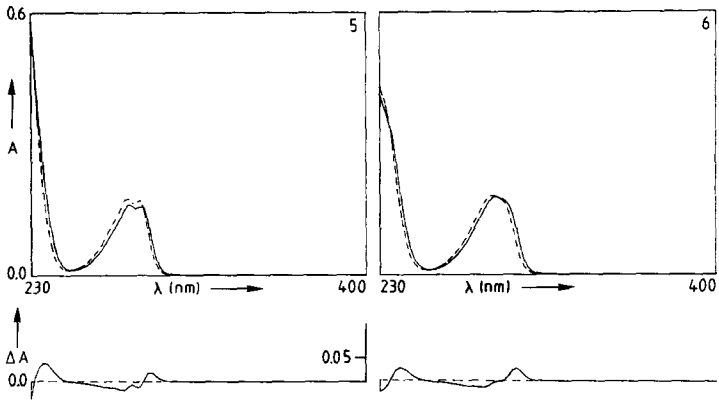

Fig. 5. The pure component spectra of components 5 and 6, recorded in acetonitrile–water (35:65) (---) and in tetrahydrofuran–water (35:65) (——), respectively, as well as the difference spectra.

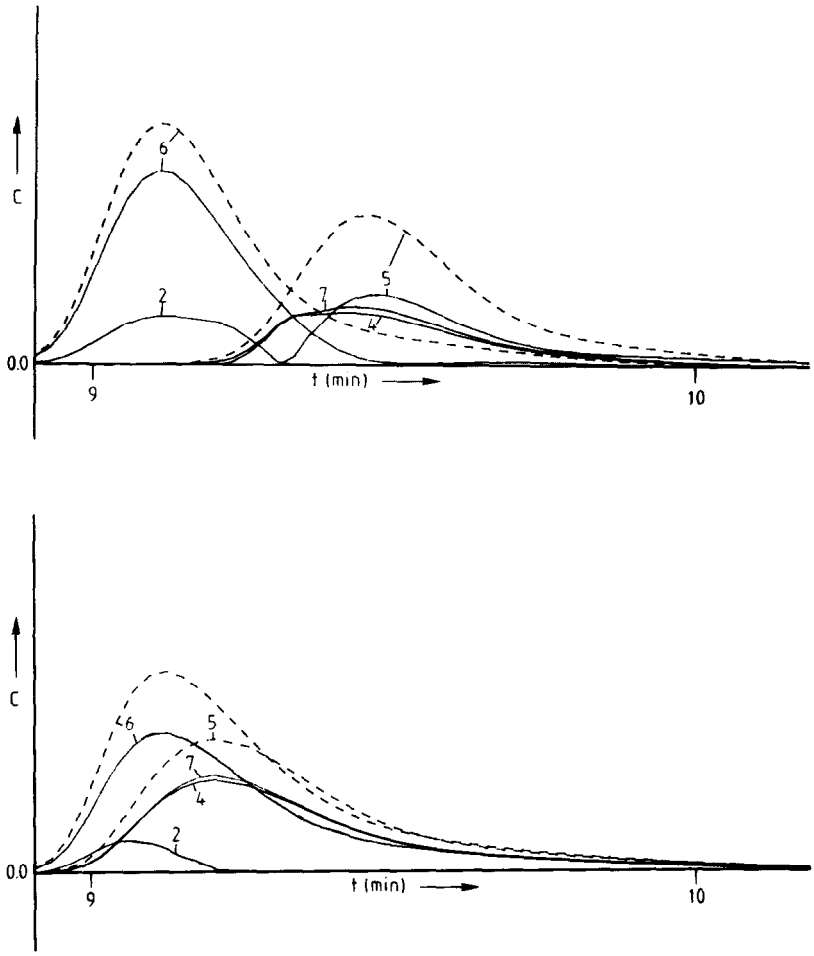

Fig. 6. The true profiles (---) of components 5 and 6 and the individual component profiles (——) calculated by means of MCA and applying a non-negativity criterion, using approximate spectra of all eight chlorophenols recorded at $\varphi = 0$. The elution profiles were recorded at (a) $\varphi = 0.5$ and (b) $\varphi = 0.6$.

eliminating the negative contributions of component 3, the compensating contributions of component 1 are automatically reduced. Still the results are far from satisfactory. Component 6 is correctly identified, but component 5 is at best uncertain and there seem to be more than two components present. Fig. 6b demonstrates an even more dramatic example for the same solutes at lower resolution. Here, component 5 has completely disappeared and has been replaced by a combination of the components 4 and 7. Apparently, MCA only performs well when the exact spectra in the mobile phase used are available.

Far better results are observed when we perform a preselection of appropriate spectra (components) by means of a target test. As described in the theoretical section, we project our targets on the description derived from the mixture spectra. As the first step we conclude, from a principal component analysis (PCA), that there are

two components present in this cluster, hence we reduce our abstract description to two vectors: every observed spectrum can be reconstructed as a linear combination of these two vectors. Similarly, the true solute spectra can also be reconstructed as a linear combination of the two vectors. After performing a target test with all eight reference spectra from our reference file the two most likely candidates are determined. The corresponding transformation matrix is inverted and used to determine the corresponding elution profiles. As an example Fig. 7 displays the elution profiles derived from the profile recorded for $\varphi = 0.5$, after a target test with the spectra of all eight components in 35% tetrahydrofuran and selecting 5 and 6 as the true components, since these components display the highest correlation when used as targets.

Obviously TFA provides a much better estimate of the true elution profiles than MCA or non-negative MCA. The main reason is that the disturbances caused by the other components in the reference file are removed beforehand. In principle we can perform MCA with the selected targets, but, as is shown by eqn. 6, when we have determined the necessary transformations for the spectra we might as well determine the corresponding elution profiles by means of the inverse transformation. In fact we are investigating all measured spectra in one calculation, thus compensating partly for the time required for the principal component analysis.

As before, the remaining difference in Fig. 7 between the estimated and the true elution profiles is due to mobile phase effects on the solute spectra. When these deviations are expressed as the sum of squared absorbance differences (SS), their value increases with increassing difference in mobile phase composition used for recording the reference spectra and the sample spectra, respectively. This is illustrated in Fig. 8, where the curves refer to the observed deviations in the profiles of components 5 and 6 for $\varphi = 0.5$, presented in Fig. 7, after TFA was performed with reference spectra recorded in different solvents (φ ranging from 0.0 to 1.0). It is important to note, however, that for this set of components TFA invariably selects the correct solutes 5 and 6 from the total set of eight chlorinated phenols tested.

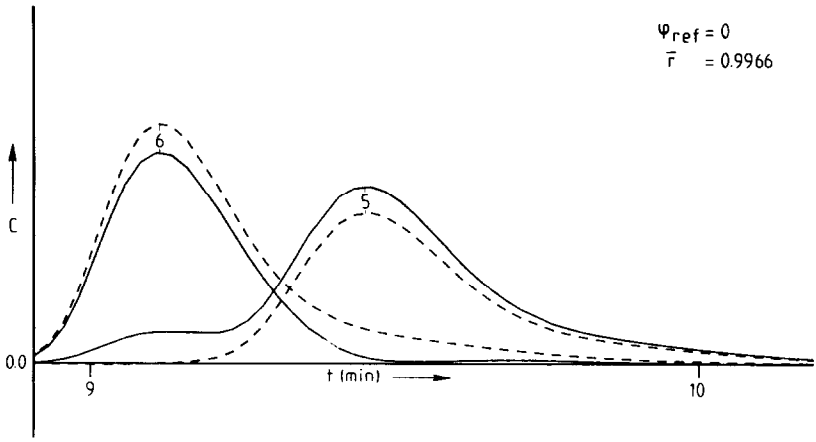

Fig. 7. The true profiles (---) of components 5 and 6 and the individual component profiles (——) calculated by means of TFA, after selecting approximate spectra of components 5 and 6 recorded at $\varphi = 0$.

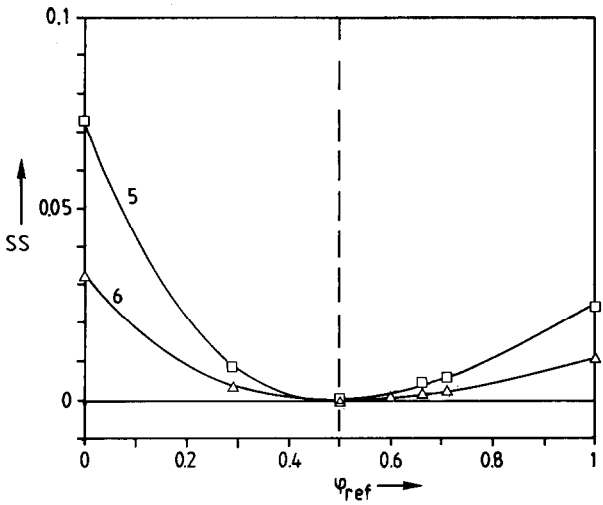

Fig. 8. The observed difference between true profiles and profiles calculated by means of TFA of components 5 and 6 at $\varphi = 0.5$, when using reference spectra recorded in various solvents, indicated by the mixing ratio φ_{ref}. The difference is expressed as the sum of squared differences, SS, between true and calculated profile.

Consequently, although elution profiles and, hence, quantitative analysis is hampered by a poor knowledge of the exact spectra, the correct identification is not. Obviously, for the present purpose of optimization, the latter observation is extremely important.

When investigating ITT we are dealing with the reverse approach. Now we start with the determination of the elution profiles and use these to derive the pure component spectra. Again two points are important with regard to peak recognition for chromatographic optimization: first, the identity of the component as indicated by its spectral characteristics; and second, the corresponding retention time, or more generally the quality of the elution profile. Because of the nature of the technique, the simultaneous contemplation of a collection of mixture spectra, these two characteristics are closely related: when the calculated elution profile deviates from the true profile the calculated component spectrum will be in error as well.

When the components are reasonably well separated and consequently the elution profiles are fairly well defined, the major differences between the calculated spectra and the corresponding spectra in the reference set are mostly caused by the difference in the respective mobile phase compositions. As was to be expected from Fig. 5, this deviation is not dramatic. This is further illustrated in Fig. 9, which shows correlation coefficients between the calculated spectrum of component 5 for $\varphi = 0.714$ ($R = 0.41$) after application of ITT, and reference spectra of all eight solutes recorded in various mobile phases defined by the mixing ratio φ. The performance in the total recognition procedure will be mainly determined by the spectral similarities between the components in the reference set. In this example component 5 is always identified correctly, independent of the mobile phase used to record the reference spectra; the same result was found for the other component in the cluster, *i.e.* component 6.

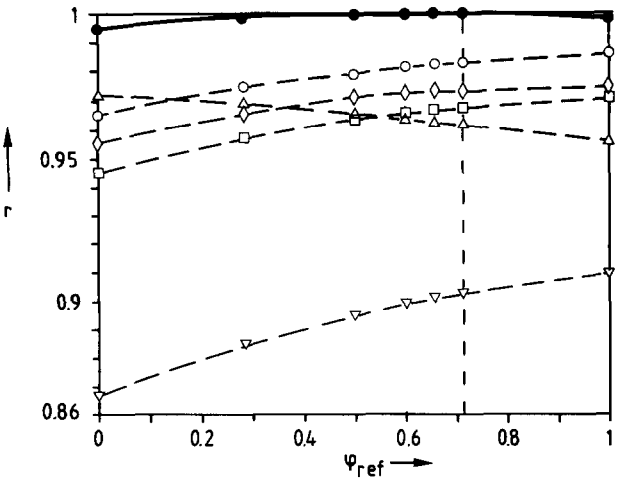

Fig. 9. The correlation coefficient r between a spectrum, obtained by ITT from the elution profile recorded at $\varphi = 0.7$ ($R = 0.4$), and reference spectra of six chlorinated phenols recorded in various solvents indicated by φ_{ref}. Displayed are the correlation coefficients resulting from comparisons with spectra of components 1 (□), 2 (○), 3 (◇), 4 (△), 5 (●) and 6 (▽). Component 5 is correctly selected as the most probable solute in all cases.

The influence of the chromatographic resolution

The second major influence on the performance of the spectral deconvolution techniques is the extent of chromatographic separation between the components. Because of the asymmetric peak-shape of the chlorinated phenols we express the chromatographic resolution by means of the first and second moments of the reference profiles instead of retention times and peak-widths:

$$R = |M_{16} - M_{15}|/2 \cdot (\sqrt{M_{25}} + \sqrt{M_{26}}) \quad (9)$$

where M_{15} and M_{16} are the first (central) moments of components 5 and 6, and M_{25} and M_{26} are the second moments of these components, respectively.

It follows from theoretical considerations that the performance of MCA and MCA0 is independent of the chromatographic resolution. Since every measured spectrum is evaluated apart from the others, there is an independent determination of the concentrations in every mixture spectrum. It is for this reason that only MCA can be applied to unresolved solutes (as in UV spectrometry), although it only performs well when the reference spectra correspond exactly in number and in nature with those of the components in the mixture.

In contrast, such an exact match is not needed for TFA and ITT, but conversely some chromatographic resolution is essential and results become better with increasing resolution. Indeed, when two solutes approach each other more closely in the chromatogram, the mixture spectra recorded across the elution profile display fewer variations, thus increasing the difficulties in the determination of the correct number of components. In the preliminary PCA the first component is emphasized and the second one diminishes and becomes more subject to experimental error. Because the second principal component discriminates between the spectra of dif-

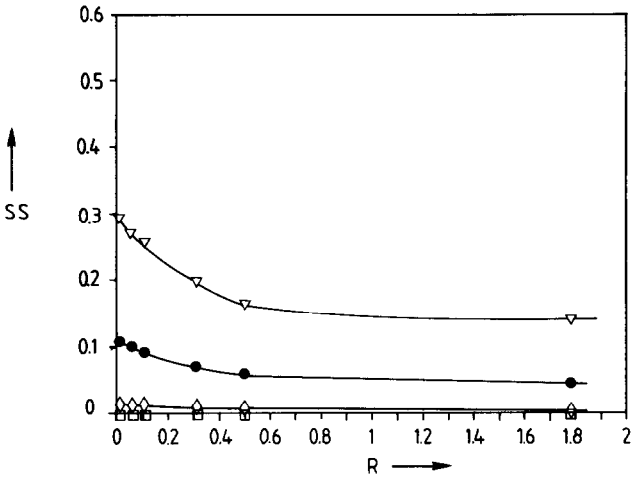

Fig. 10. The observed difference between true profiles and profiles calculated by means of TFA of component 5, resulting after simulation of an elution profile of a mixture of components 5 and 6 at $\varphi = 1$. The profiles display a variation in chromatographic resolution R. The difference is expressed as the sum of squared differences, SS, between true and calculated profile. The reference spectra of components 5 and 6 were recorded in solvents corresponding to $\varphi = 1$ (□), $\varphi = 0.6$ (◇), $\varphi = 0.3$ (●) and $\varphi = 0$ (▽).

ferent components, these errors can cause distortions in the calculated elution profiles even when the solutes are correctly identified in TFA. In ITT the distorted profiles produce equally distorted spectra making solute recognition much more difficult.

In order to separate the influence of the mobile phase composition from the influence of the resolution on the results of TFA, different degrees of chromatographic resolution were simulated by summing the individual profiles of the components

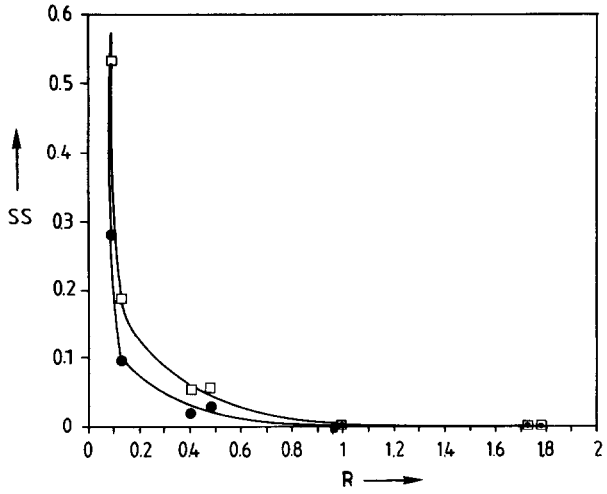

Fig. 11. The observed difference between true profiles and profiles calculated by means of ITT of components 5 (□) and 6 (●). The investigated profiles display a variation in chromatographic resolution R. The difference is expressed as the sum of squared differences, SS, between true and calculated profile.

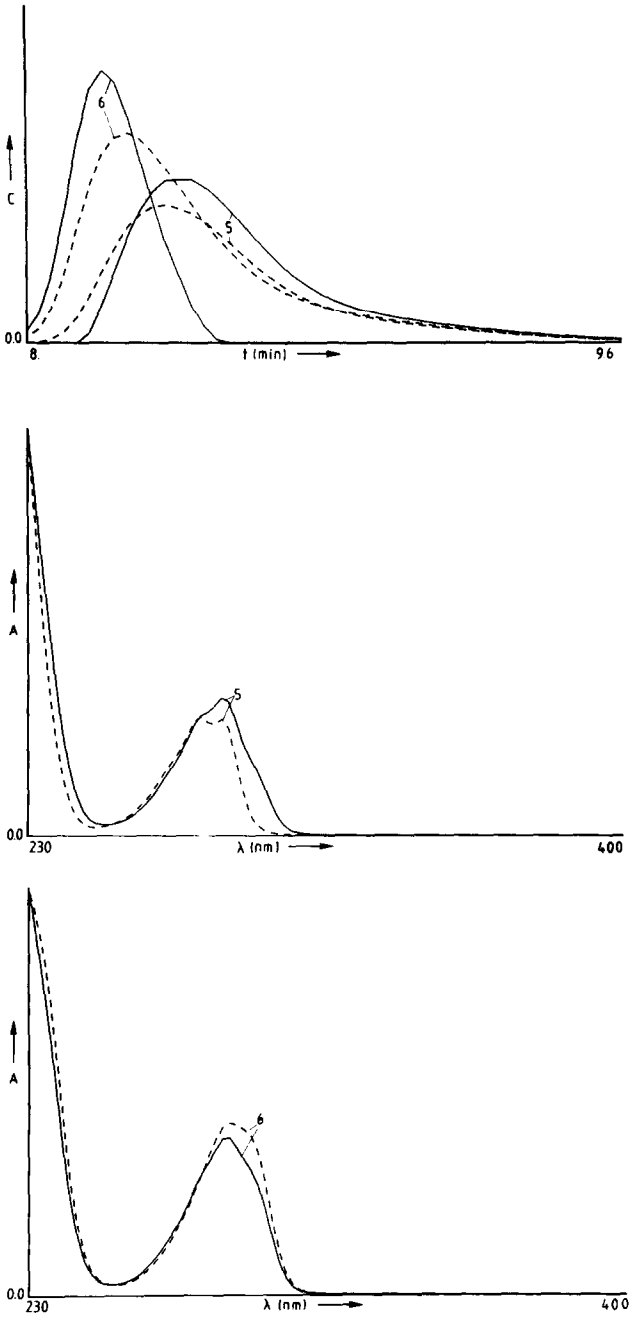

Fig. 12. The results of ITT performed on a profile of components 5 and 6, showing severe overlap (R = 0.1). The mixture was eluted at φ = 0.6. (a) The true (---) and calculated (——) elution profiles. (b) The true (---) and calculated (——) spectrum corresponding to component 5. (c) As b but for component 6.

5 and 6 with various relative positions. A target test was performed with reference spectra recorded in different mobile phases. The resulting deviations in the calculated elution profiles are shown as the sum of squared differences in Fig. 10. When the reference spectra agree exactly with those in the solvent used in eluting the mixture, which is the case at $\varphi = 1$, the elution profiles are reconstructed with remarkable precision down to a resolution as low as 0.006 (the smallest value tested). When the reference spectra do not match exactly, the reconstructed profiles deviate as shown by the larger values of SS at high resolution. Also, the deviations increase somewhat more at very low resolution ($R < 0.1$).

It might be argued that these distortions are not important for qualitative investigations (*e.g.* the optimization strategy) as long as the components are identified correctly and the retention times do not deviate too much. When two (or more) components overlap severely they have almost identical retention times, hence the application in the optimization strategy should pose no problem. The major problem in cases of extreme overlap remains the correct determination of the number of components by PCA.

When ITT is applied, the correct calculation of the elution profiles is much more important because of the earlier mentioned connection between calculated profiles and the derived spectra. Unfortunately ITT is more sensitive to the degree of overlap than TFA. In comparison to Fig. 10, the SS values in Fig. 11 are higher and rapidly increase further when the resolution decreases below 0.2. As an example Fig. 12a shows the reconstructed profiles in the case of a severe degree of overlap ($R = 0.09$). Not only is the shape of the profiles distorted (evidenced by the absence of one peak-tail), there is also a shift in the location of the maxima. As expected, the distortions in the profiles lead to deviating spectra (Fig. 12b and c). As a result, Fig. 13 shows that a correct identification of component 5 is no longer possible, since

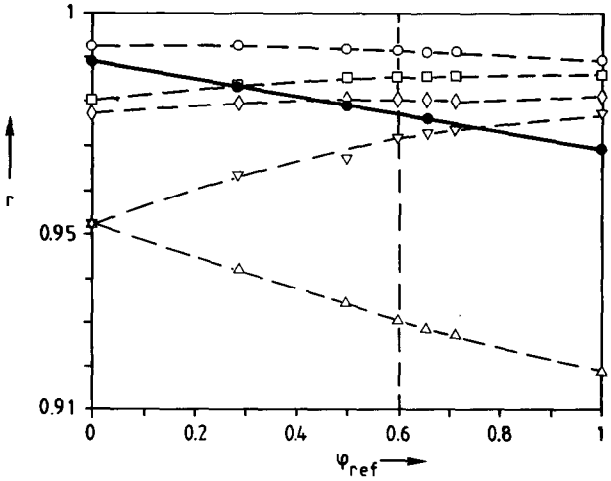

Fig. 13. The correlation coefficient r, resulting from a comparison of a spectrum, calculated by means of ITT from the profile recorded at $\varphi = 0.6$ ($R = 0.1$), and reference spectra of six chlorinated phenols recorded in various solvents indicated by φ_{ref}. Displayed are the correlation coefficients resulting from comparisons with spectra of components 1 (□), 2 (○), 3 (◇), 4 (△), 5 (●) and 6 (▽). Instead of component 5, which was actually present, component 2 is selected as the most probable solute in all cases.

component 2 shows a greater similarity with the calculated spectrum, independent of the solvent used to record the reference spectra.

CONCLUSIONS

From the analysis of various mathematical techniques for the deconvolution of a two-component elution profile the following conclusions can be drawn.

Although the application of MCA is independent of the chromatographic resolution, it has some major disadvantages when used with larger reference sets. It is highly sensitive to a difference between the actual spectra and the reference spectra caused by a change in the mobile phase composition. This leads to large deviations of the estimated profiles and more seriously, impairs the correct identification of the two solutes present in the profile. Although we observed some improvements when applying a non-negativity criterium, the results are still inadequate for an unambiguous recognition. Finally, MCA can only be applied when the set of reference spectra includes at least those of the components actually present in the peakcluster.

In the case of TFA the influence of the change in the spectral characteristics is much less pronounced, mainly because the preliminary PCA limits the reconstruction to the number of components actually present, even though their identity remains to be ascertained. Furthermore we can test for the presence of a component without knowing the spectra of the other components involved. Because we do need all spectra for a reconstruction of the elution profiles, however, this is only a minor advantage with respect to the chromatographic optimization strategy, which needs retention times as well as identities. As the examples have indicated the method becomes somewhat less accurate in cases of extreme overlap, although within the investigated group of eight chlorophenols positive recognition is still possible down to $R = 0.006$. Consequently retention times can be determined with high accuracy, which is important for optimization purposes. The ultimate value of TFA will rely on the ability of PCA to determine the number of components correctly.

Obviously, the main advantage of ITT is the potential to determine elution profiles and spectra without any previous knowledge. When the solutes are reasonably well separated the influence of the mobile phase composition is again minor. The major limitation of the method is connected with the sensitivity to the required resolution. When resolution drops below a minimum value, dependent on the spectral characteristics of the components involved, the method still yields approximate elution profiles, but the derived spectra are too inaccurate for reliable solute recognition.

Thus if there is adequate resolution we can apply ITT, which requires the least knowledge of the sample. When this fails, and unfortunately this is not always apparent, we have to use previously collected pure-component spectra, *e.g.* from another chromatogram with more resolution. The method of choice is then TFA, where distortions caused by minor differences in the spectra can usually be ignored. When there is hardly any resolution (*i.e.* when we cannot determine the correct number of components) the only possibility is the application of MCA preferably with boundary conditions (non-negativity criterion). The results, however, cannot be trusted if there is a large difference between the reference and the experimental spectra as a result of the mobile phase effects.

ACKNOWLEDGEMENT

The authors thank J. M. Guyt for providing and analysing the chromatographic data.

REFERENCES

1 P. J. Schoenmakers, A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Chromatographia*, 15 (1982) 688.
2 A. C. J. H. Drouen, H. A. H. Billiet, P. J. Schoenmakers and L. de Galan, *Chromatographia*, 16 (1982) 48.
3 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 56 (1984) 971.
4 G. Horlick and E. G. Codding, *Anal. Chem.*, 45 (1973) 1490.
5 D. G. Jones, *Anal. Chem.*, 57 (1985) 1057A.
6 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 57 (1985) 962.
7 J. W. M. Wegener, H. J. M. Grunbauer, R. J. Fordham and W. Karcher, *J. Liq. Chromatogr.*, 7 (1984) 809.
8 A. F. Fell, *Anal. Proc.*, 17 (1980) 512.
9 W. H. Lawton and E. A. Sylvestre, *Technometrics*, 13 (1971) 617.
10 B. G. M. Vandeginste, W. Derks and G. Kateman, *Anal. Chim. Acta*, 173 (1985) 253.
11 B. G. M. Vandeginste, R. Essers, T. Bosman, J. Reijnen and G. Kateman, *Anal. Chem.*, 57 (1985) 971.
12 J. A. Blackburn, *Anal. Chem.*, 37 (1965) 1000.
13 M. Otto and W. Wegscheider, *Anal. Chem.*, 57 (1985) 63.
14 E. R. Malinowski and D. G. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.
15 M. McCue and E. R. Malinowsky, *Appl. Spectroscopy*, 37 (1983) 463.
16 H. T. Eastman and W. J. Krzanowski, *Technometrics*, 24 (1982) 73.
17 J. C. Reid and E. C. Wong, *Appl. Spectroscopy*, 20 (1966) 220.
18 K. Tanabe and S. Saëki, *Anal. Chem.*, 47 (1975) 118.
19 L. Lawson and R. Hanson, *Solving Least Squares Problems*, Prentice Hall, Englewood Cliffs, NJ, 1974.
20 D. H. Burns, *PhD Thesis*, University of Washington, 1984.
21 Y. Beppu and I. Ninomiya, *Computers Chem.*, 6 (1982) 87.