ELSEVIER

Journal of Chromatography A, 855 (1999) 487–499

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Application of several modified peak purity assays to real complex multicomponent mixtures by high-performance liquid chromatography with diode-array detection

A. Garrido Frenich[a,*], J.R. Torres-Lapasió[b], K. De Braekeleer[b], D.L. Massart[b], J.L. Martínez Vidal[a], M. Martínez Galera[a]

[a]*Department of Hydrogeology and Analytical Chemistry, University of Almería, 04071 Almería, Spain*
[b]*ChemoAC, Vrije Universiteit Brussel, Laarbeeklan 103, B-1090 Brussels, Belgium*

Received 30 December 1998; received in revised form 3 June 1999; accepted 3 June 1999

Abstract

Simple to use interactive self-modelling mixture analysis (SIMPLISMA), orthogonal projection (OPA) and Needle Search (NS) approaches have been applied to the determination of a number of compounds present in a complex multicomponent system. None of these three approaches succeeded completely when they were tested using the whole data matrix. When OPA and NS were applied to three simpler submatrices, obtained by dividing the total data matrix, and where a smaller number of compounds were present, better performance was achieved.

Keywords: Needle Search; Orthogonal projection; SIMPLISMA; Curve resolution; Peak purity; Chemometrics

1. Introduction

The application of high-performance liquid chromatography (HPLC) coupled with diode-array detection (DAD) is becoming more and more important for the analysis of multicomponent systems. This hyphenated technique generates data matrices where the columns correspond to chromatograms and the rows to spectra. When more than one component elutes, the matrices contain linear combinations of the pure spectra of the eluting compounds in its rows, and combinations of the pure elution profiles in its columns. In spite of the high separation efficiency of the technique, the occurrence of overlapped peaks is often unavoidable. This is particularly critical in environmental analysis, since samples contain large numbers of analytes and coelution often occurs. In those cases the problem of ascertaining the number of compounds present under a chromatographic profile is not an easy task.

An important step forward in the assessment of peak purity in hyphenated techniques has been the development of multivariate approaches for data analysis. In general, approaches for peak purity analysis can be classified in two categories [1]: (a) those based on the comparison of rows or columns of the data matrix with a reference profile (spectrum or chromatogram) and (b) those based on principal component analysis (PCA). Methods such as the orthogonal projection approach (OPA) [2–5] or simple to use interactive self-modelling mixture analysis (SIMPLISMA) [4,6–10] belong to the fist

*Corresponding author.

0021-9673/99/$ – see front matter
PII: S0021-9673(99)00709-8

category. The main difference between them consists in the selection of the reference profile. The criterion used for this purpose is called dissimilarity and purity for the OPA and SIMPLISMA approaches, respectively. Both techniques were initially developed for the determination of the number of compounds present in a multicomponent system, although they have subsequently been extended by adding a least squares (in SIMPLISMA) or an iterative least squares (in OPA) stage to resolve the data matrix into the individual concentration profiles and spectra.

Approaches belonging to the second category are, for instance, heuristic evolving latent projection (HELP) [11–13], fixed size window evolving factor analysis (FSW EFA) [14,15], eigenstructure tracking analysis (ETA) [9,16,17], singular value evolving profile (SVEP) [18,19], and iterative target transformation factor analysis (ITTFA) [20–23].

The Needle Search (NS) algorithm is an alternative multivariate approach for the assessment of peak purity. Recently, promising results have been obtained with NS in the characterization of HPLC–DAD data [23]. It was initially proposed for the selection of initial chromatographic profiles, known as targets, to be refined by using iterative target transformation approaches [24]. It is also based on PCA, so does not require a priori information about the number and identity of the compounds present in the data matrix. The interpretation of the results is easy: the number of targets found by NS equals the number of compounds eluting and their position coincides with the retention times of the compounds.

In general, the above multivariate peak purity approaches have been applied successfully to the resolution of systems with two or three overlapped compounds and where one of them was the main component and the others acted as minor components up to impurity levels. In this work the power of SIMPLISMA, OPA and NS for detecting multiple closely eluting compounds is discussed. In particular, the performance of these three approaches, introducing slight modifications in the OPA and NS, is compared for a very complex real system that contains 12 overlapped compounds. Here we investigate the advantage of reducing peak clusters to submatrices where a smaller number of species are present.

2. Theory

HPLC–DAD provides a data matrix $X(m \times n)$ where the m rows are spectra measured at different time intervals and the n columns are chromatograms measured at different wavelengths. This allows extraction of useful information applying the multivariate approaches in both data matrix directions. Usually, the pure spectra of the components are rather similar, and selective spectral regions cannot be found. However, often there are selective chromatographic regions, at least for the main compound eluting. Here, we apply peak purity methods in the chromatographic direction to search for the purest rows of **X**, i.e. the analysis times where ideally only one compound is eluting. In this case the selected times yield the 'pure spectra', i.e. approximations of the spectral profiles of the pure compounds.

A detailed mathematical explanation of the methods used has been given elsewhere [1,23]. Here only a brief description of each method is given.

For clarity, the following notation is used in this paper. Matrices will be denoted by bold upper case letters (e.g., **X**), vectors by bold italic lower case letters (e.g., ***y***) and scalars by italic lower case letters (e.g., k).

2.1. SIMPLISMA

SIMPLISMA is a stepwise approach that identifies the 'purest spectra'. The purity of spectrum i in plot j, p_{ij}, is defined as

$$p_{ij} = w_{ij}\frac{\sigma_i}{\mu_i + \text{offset}} \quad \text{for } i = 1, \ldots, m \tag{1}$$

where σ_i and μ_i represent, respectively, the standard deviation and the mean of each spectrum. The offset is a positive number, usually 1–3% of the maximum absorbance, and is added to the denominator to avoid spectra with low mean absorbance (i.e. noise spectra) obtaining a high purity value. The weight of each spectrum i in plot j, w_{ij}, is defined as the determinant of the dispersion matrix of $\mathbf{Y}_{ij}$:

$$w_{ij} = \det(\mathrm{Y}_{ij}\,\mathrm{Y}_{ij}^{\mathrm{T}}) \tag{2}$$

In the first plot, before the selection of the first 'pure spectrum', Y_{i1} contains only one row, $\boldsymbol{z}_i$, which

is the normalized spectrum. Therefore, the weight for the spectrum at time i, w_{i1}, is equal to the square of the length of the normalized spectrum, $\|\boldsymbol{z}_i\|^2$. The normalization of the spectra takes into account the offset. This means that spectra with low mean absorbance have a smaller weight than spectra with a high mean absorbance in the first plot. After the selection of the spectrum with the highest purity value the weight factor used in the second plot, w_{i2}, will compare each normalized spectrum, $\boldsymbol{z}_i$, with the first selected and normalized spectrum $\boldsymbol{z}_{s1}$. Spectra which are similar to the first selected spectrum will obtain a low weight and due to this the effect of the first selected compound will be eliminated in the second plot. The weights in the third plot are calculated using the information of the two previous selected spectra. This process continues, including previous selected and normalized spectra in the matrix $\mathbf{Y}_i$, as long as there are peaks observed in the corrected purity plots.

In each step, the purity of each spectrum is plotted as a function of the analysis time and the resulting plot is known as a purity chromatogram. A random profile of the purity plot indicates that the correct number of spectra has been selected and the process finishes.

SIMPLISMA is an interactive approach because the user selects the offset and decides whether a spectrum ought to be selected or not. At the end, the number of selected 'pure spectra' indicates the number of compounds present in the system.

2.2. OPA

OPA is also an interactive approach, which determines the number of compounds by visual inspection of successive dissimilarity plots. These plots are obtained by comparing the spectra of the data matrix $\mathbf{X}$ with one or more reference spectra. The number and the origin of the reference spectra used to calculate the dissimilarities depends on the number of the plot. The dissimilarity for each spectrum i in the first plot, dis_{i1}, is calculated as the determinant of the dispersion matrix of Y_{i1}, which initially contains the mean spectrum (considered as the reference spectrum) of the data matrix normalized to unit length, and the non-normalized measured spectrum i from the complete data set $\boldsymbol{x}_i$

$$\mathrm{dis}_{i1} = \det(\mathrm{Y}_{i1}\,\mathrm{Y}_{i1}^{\mathrm{T}}) \quad \text{for } i = 1, \ldots m \tag{3}$$

The dissimilarity of each spectrum with respect to the mean spectrum is determined and plotted as a function of the analysis time. The spectrum $\boldsymbol{x}_{s1}$ that yields the highest dissimilarity with respect to the mean spectrum is selected, normalized to unit length, $\boldsymbol{z}_{s1}$, and taken as reference, instead of the mean spectrum, for the calculation of the dissimilarity values in the second plot.

In the second step, the dis_{i2} of each individual spectrum of $\mathbf{X}$ with respect to $\boldsymbol{z}_{s1}$ is calculated. As before, the dis_{i2} is calculated as the determinant of the dispersion matrix of Y_{i2}, which now contains the previously selected spectrum $\boldsymbol{z}_{s1}$ as reference spectrum, and a measured spectrum $\boldsymbol{x}_i$. The most dissimilar spectrum, $\boldsymbol{x}_{s2}$, is selected, normalized, $\boldsymbol{z}_{s2}$, and included as a second reference in each matrix Y_{i3} for the determination of the third dissimilarity plot.

The procedure continues by comparing each spectrum from $\mathbf{X}$ with the normalized spectra that have already been selected, including the one most dissimilar to $\mathbf{Y}_{ij}$ (j being the number of the plot) and calculating the determinant of the dispersion matrix of $\mathbf{Y}_{ij}$ until the dissimilarity plot shows a random profile indicating that only noise is left. The number of selected spectra should equal the number of absorbing compounds in the system.

2.2.1. Modified OPA

This method can be modified by introducing in each step the normalized spectrum of a standard of the component that elutes at the same retention time as the most dissimilar spectrum selected. Initially, the dis_{i1} of each spectrum with respect to the mean spectrum is determined, as in the previous OPA method. Then, instead of selecting the spectrum with the highest dissimilarity and taking it as reference for the calculation of the next dis_{i2} values, the normalized spectrum of the standard of the component eluting at the time of the most dissimilar spectrum detected is introduced. The process continues until all reference spectra of the components in the data set have been introduced. The goal is to detect the existence of impurities or smaller intensity peaks under the main compounds. Reference spectra for each component were available, permitting the application of this modified method.

2.3. Needle Search

Needle Search is a powerful algorithm for the location and determination of the number of compounds present in a data matrix. The standard procedure can be summarised in the following steps.

(a) Decomposition of the data matrix $\mathbf{X}(m \times n)$ by singular value decomposition (SVD)

$$\mathbf{X} = \mathbf{U} \times \mathbf{S} \times \mathbf{V}^{\mathrm{T}} \tag{4}$$

where each column of $\mathrm{U}(m \times m)$ contains a row-singular vector (or abstract chromatogram), each column of $\mathrm{V}(n \times n)$ consists of a column-singular vector (or abstract spectrum) and $\mathrm{S}(m \times n)$ is the diagonal matrix of the associated singular values. The superscript 'T' denotes transposition.

(b) Selection of the significant number of factors, i.e. the first '*p*' column and row singular vectors of **U** and **V**, that allow the reconstruction of the data matrix **X** within the noise. Theoretically, '*p*' is the number of eluting compounds.

(c) Creation of as many as possible ideal chromatographic profiles, known as needle targets, or fakes, which are *m* ideal chromatograms that sample each time of the data matrix. Each fake is a column vector where all the elements are zero except one, which has a value of 1. When all possible fake vectors are written together, sorted increasingly according to the position of the non-zero element, they form an identity matrix $m \times m$. The NS method selects *p* targets from a population of *m* in such a way that they correspond best with the elution times of the *p* compounds present. The targets are later refined by least squares for transformation into true elution profiles.

(d) Projection of each target into the space defined by the selected singular vectors. If *p* columns from **U** were selected, a given needle target ($\boldsymbol{tg}_{m,i}$) will be projected through a simple matrix product: $\mathbf{U}p\mathbf{U}p^{\mathrm{T}}\boldsymbol{tg}_{m,i}$. When the matrix **Tg** including all targets is projected, we have: $\mathbf{U}p\mathbf{U}p^{\mathrm{T}}\mathrm{Tg}$.

(e) Finally, the distance from the projection and each target is determined:

$$\mathrm{d} = \|\mathrm{Tg} - \mathrm{U}p\mathrm{U}p^{\mathrm{T}}\| \tag{5}$$

and represented versus retention time. The representation shows a multivalley curve, where ideally each minimum indicates the peak position of a given compound.

2.3.1. Modified NS

To study artefacts and complex mixtures, NS was slightly modified to achieve more informative results. Three modifications were carried out. The first concerns the way of projecting the targets: instead of a unique projection space for all factors, as in classical NS: $\mathbf{U}p\mathbf{U}p^{\mathrm{T}}$, *p* projection spaces (each associated with a given factor *j*) were considered: $\mathrm{u}_j\mathrm{u}_j^{\mathrm{T}}$. This makes it easier to visualise the contribution of each component.

The second modification introduced was the way of representation. Because targets and projections are very dissimilar, the cosine of the angle target–projection, instead of the distance target–projection (as in classical NS), is used to give more contrast to small anomalies. Due to targets and selected singular vectors having unit length, the projection of the target into each abstract chromatogram is the following:

$$p = \mathbf{u}_j\mathbf{u}_j^{\mathrm{T}}\mathbf{tg} \tag{6}$$

where $\mathbf{u}_j$ is the selected singular vector *j* and **tg** the target (i.e. an identity column vector, which examines each time point in the HPLC–DAD chromatogram). The representation of the cosine as a function of the analysis time is similar to the standard needle plot, but with positive peaks instead of valleys, where the differences with respect to the targets are inflated.

The third modification consists of the application of unimodality to each needle plot if the mixture is too complex. Because each factor is associated with a single compound, one must expect a unimodal needle plot if we are using individual projection spaces. Removing secondary peaks makes it easier to discover meaningless components, associated with noise, and to eliminate disturbing residual contributions under a given peak. Only the most meaningful feature will remain.

Displacements in peak positions were also described in classical NS. In the same way as the maxima in overlapped chromatographic elution profiles are displaced with respect to the individual signals for each compound, the addition of *p* repre-

sentations into a single NS multivalley plot produces displacements with respect to the true peak positions; this is, another reason for applying NS in individual projection spaces and removing small contributions from other peaks.

3. Data

The data were acquired on a Waters (Milford, MA, USA) Model 990 DAD liquid chromatographic system, equipped with a Hypersil C_{18} column (100× 0.46 mm I.D., 5 μm particle size). An acetonitrile–water (60:40, v/v) mobile phase under isocratic conditions was used. Injections were made through a Rheodyne six-port injection valve with a 20 μl sample loop. The flow-rate was 1 ml min^{-1}. Photometric detection was performed in the 200–280 nm range, with a spectral resolution of 1.4 nm. Data was obtained over an integration period of 1.4 s per spectrum.

Analytical-reagent-grade solvents from Merck (Darmstadt, Germany) and purified Milli-Q water (Bedford, MA, USA) were used. Pesticide standards (Pestanal quality) of iprodione (Ip), procymidone (Pr), chlorothalonil (Ct), chlorfenvinphos (Cf), fenamiphos (Fe), malathion (Ma), parathion-methyl (P-m), parathion-ethyl (P-e), tebuconazole (Te), triadimefon (Td), triazophos (Tz) and vinclozolin (Vi) were obtained from Riedel-de Haën (Seelze, Germany) and used without further purification.

4. Software

Analysis of the data was performed using in-house routines written in Matlab V. 4.0.

5. Results and discussion

Fig. 1 shows the superposed baseline-corrected chromatograms of the individual pure standards of the 12 pesticides. Baseline correction was performed by modelling selected regions without disturbances before and after the elution of the peaks. A linear model was established for modelling the drift of each diode. There is severe overlap between components,

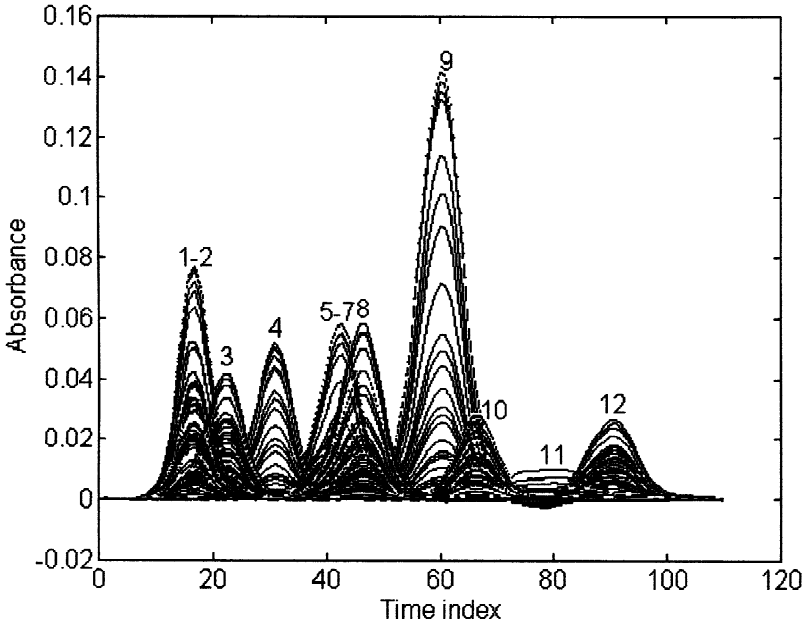

Fig. 1. Superposed baseline-corrected chromatograms of the 12 pesticide standards: (1) fenamiphos (Fe), (2) triadimefon (Td), (3) parathion-methyl (P-m), (4) iprodione (Ip), (5) malathion (Ma), (6) triazophos (Tr), (7) procymidone (Pr), (8) chlorothalonil (Ct), (9) vinclozolin (Vi), (10) chlorfenvinphos (Cf), (11) tebuconazole (Te) and (12) parathion-ethyl (P-e).

especially for the two first eluting compounds, and for the fifth, sixth and seventh eluting compounds. The absorption spectra of the compounds are shown in Fig. 2.

5.1. Application of the peak purity approaches using the whole data matrix

The performance of SIMPLISMA, OPA and NS for the assessment of peak purity in the 12-component mixture was investigated. The baseline-corrected chromatogram for the mixture is shown in Fig. 3. Baseline correction was performed in the same way as previously described for the standards.

Analysis of the data matrix by SIMPLISMA, using an offset of 3%, results in the selection of eight 'pure spectra' (Fig. 4). The first 'pure spectrum' was located at time index 31 (Fig. 4(a)). This time corresponds to the peak maximum of the Ip compound. In the second purity chromatogram, time 48 has the largest purity value and is selected (Fig. 4(b)). This corresponds to the Ct compound. In the third to sixth purity chromatograms, the following time indices are selected: 91 (P-e), 60 (Vi), 76 (Te) and 68 (Cf) (Fig. 4(c–f)). In the following purity chromatogram (Fig. 4(g)) time 17 has the highest purity value. This spectrum was not a pure spectrum but a mixture spectrum of the Fe and Td compounds which have very similar retention times. Time index 40 is selected in the eighth purity chromatogram (Fig. 4(h)). Around this time index three compounds, Ma, Tz and Pr, elute and the selected spectrum is a mixture of the three compounds. The random profile of the ninth purity chromatogram (Fig. 4(i)), together with the order of magnitude of the purity values (10^{-16}), indicates that all the variance in the data is explained by the eight spectra selected in the previous steps. The third eluting component (P-m) was not detected, probably due to the perturbing effect of the last compound (P-e). Both have similar absorp-

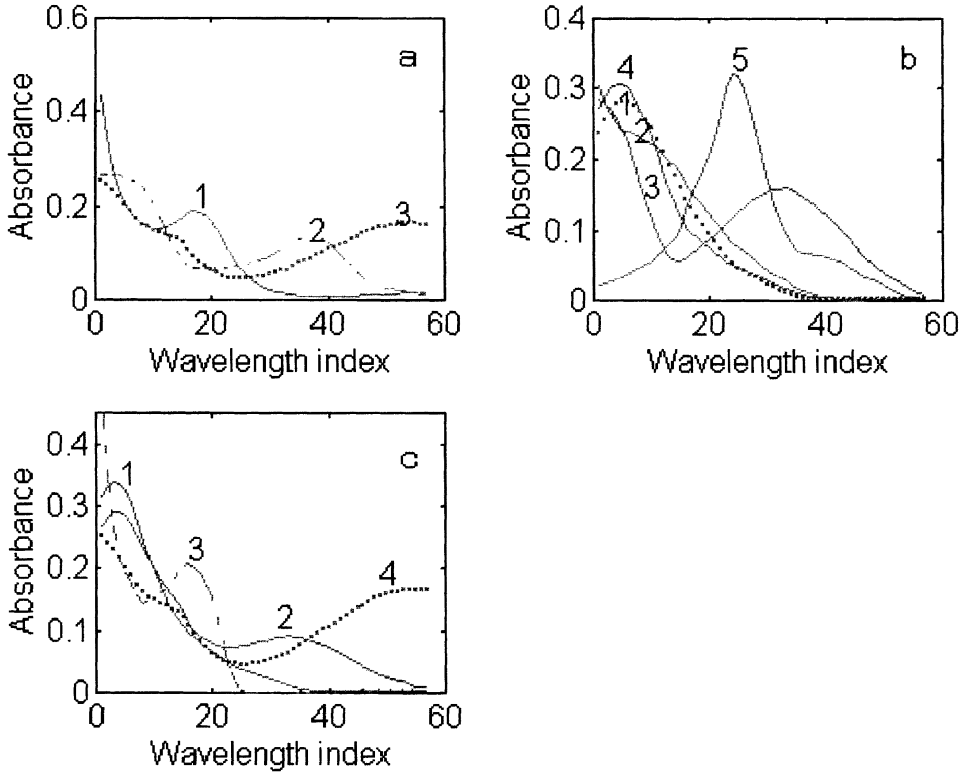

Fig. 2. Absorption spectra of: (a) (1) Td, (2) Fe and (3) P-m; (b) (1) Ip, (2) Ma, (3) Tz, (4) Pr and (5) Cf; (c) (1) Vi, (2) Cf, (3) Te and (4) P-e.

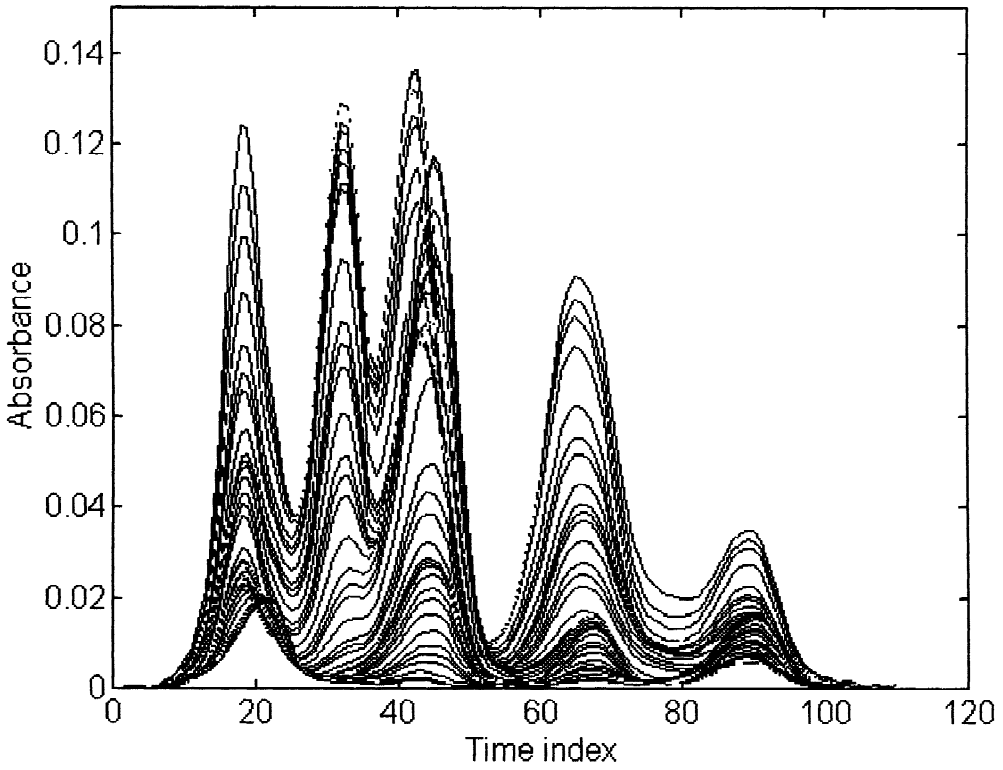

Fig. 3. Baseline-corrected chromatograms of the 12-component mixture.

tion spectra and, for that reason, the information corresponding to the P-m compound was removed when the P-e compound was selected.

OPA was applied to the same data set and the results are presented in Fig. 5, where the position of the selected spectra is marked on the chromatogram of the mixture. OPA results in the selection of 10 spectra around the time indices: 46, 32, 19, 42, 23, 62, 68, 78, 16 and 90. These indices correspond to the eluting compounds Ct, Ip, Td, Ma+Tz+Pr, Pm, Vi, Cf, Te, Fe and Pe, respectively. Resolution between the first two compounds was now possible, perhaps due to the presence of a selective small region at the beginning of the elution of the first compound. As for the SIMPLISMA approach, the spectrum selected at time index 42 has contributions from more than one compound. Thus, 10 spectra were selected, nine well correlated with the position of the standards, and the other represents three unresolved compounds.

With the aim of improving the performance of OPA, the modified OPA method was applied. The dissimilarity with the chromatogram of the mixture, after introducing the standard spectra of the 12 compounds, is shown in Fig. 6. There are some variations in the data that cannot be explained with the standard spectra, i.e. part of the variance in the experimental data matrix is not caused by changes in the concentration of the underlying compounds but by other factors. This effect is larger at time index 40, where three compounds elute.

The modified NS algorithm was also applied to the experimental data and the results are presented in Fig. 7. NS results in the selection of nine spectra with time indices 47, 43, 20, 32, 41, 62, 68, 78 and 23, which correspond to eluting compounds Ct, Tz+Pr, Fe+Td, Ip, Ma+Tz, Vi, Cf, Te and P-m, respectively. As for SIMPLISMA, the selected spectrum at time index 20 includes contributions from two compounds. However, NS shows a better resolution around time index 40, selecting two spectra with very close time indices.

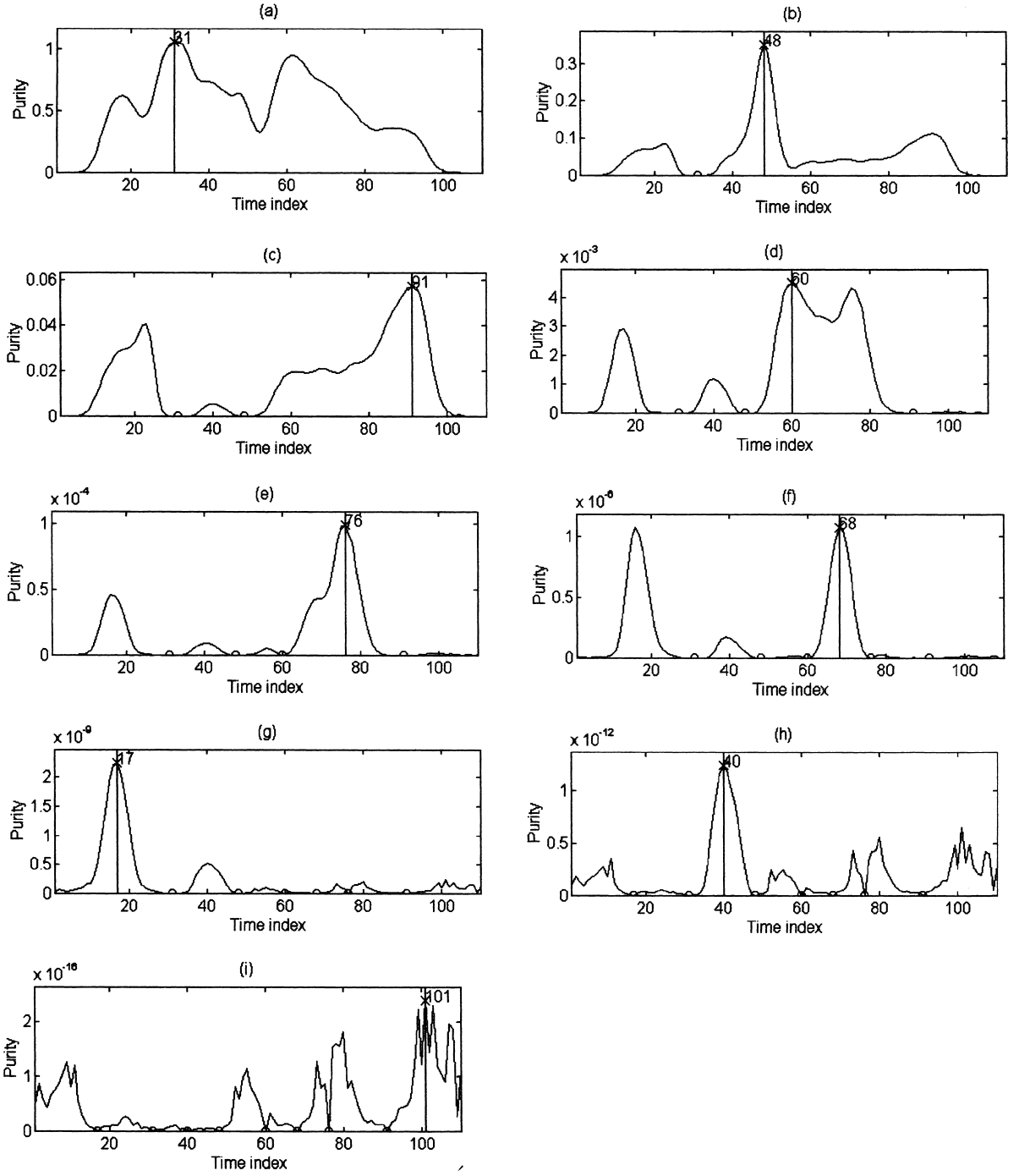

Fig. 4. Purity chromatograms obtained by SIMPLISMA for the 12-component mixture using the whole data matrix.

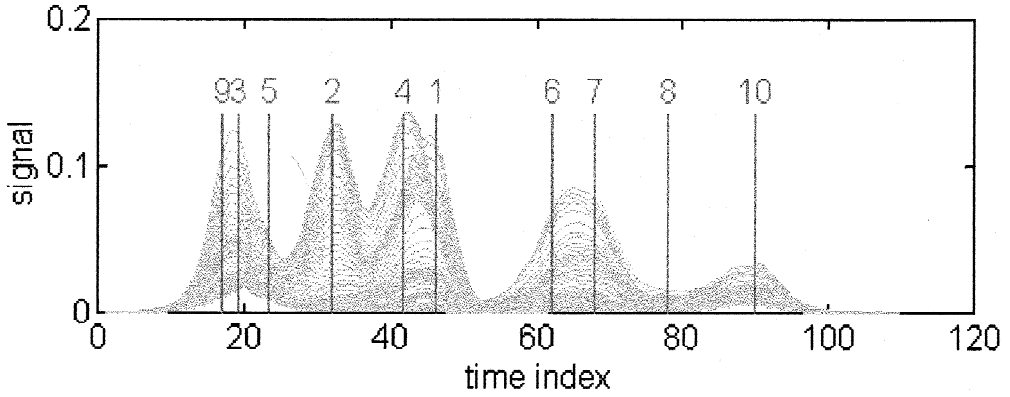

Fig. 5. Chromatograms of the mixture where the position of the selected spectra by OPA is indicated.

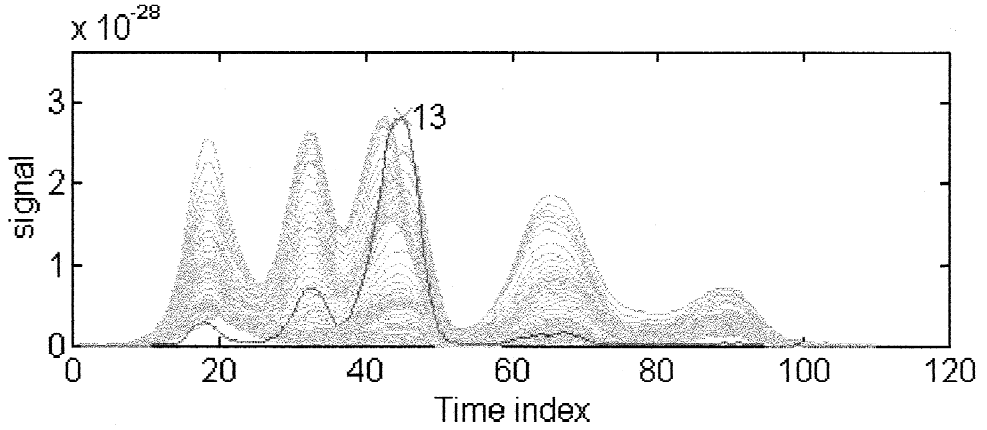

Fig. 6. Dissimilarity plot with the chromatogram of the 12-component mixture after introducing the standard spectra of the 12 components.

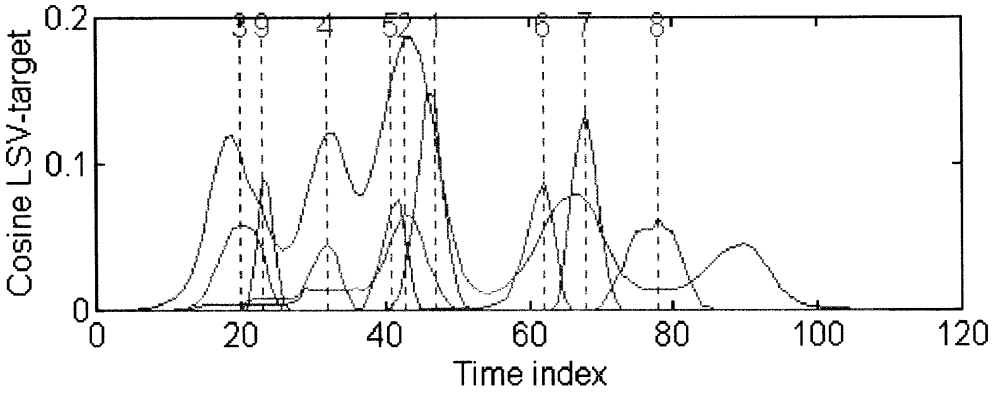

Fig. 7. Modified needle plot corresponding to the 12-component mixture, showing overlapping of the mean chromatogram, the projection of each factor and the position of the selected spectra.

Therefore, none of the three approaches, SIMPLISMA, OPA or NS, achieves the goal of finding the right number of compounds. In all cases the number of compounds was underestimated, mainly because compounds that have similar spectra are confounded and are selected as one single compound.

5.2. Application of OPA and NS approaches using submatrices

It is believed that some of the problems mentioned earlier arise because analysis of the total data matrix is attempted. Therefore, a logical procedure would be to extract a smaller matrix from the larger one and to apply multivariate purity approaches to these smaller matrices. With the aim of improving the performance of OPA and NS, the original data matrix was subdivided into three regions (Fig. 8).

Region 1. This includes the time indices 1 to 27 (Fig. 8(a)). Three compounds are present, Fe, Td and P-m, the first two having very similar retention times but also some selective regions in the spectra (Fig. 2(a)). The results obtained with OPA are given in Fig. 9(a). The three spectra with maxima at times 23, 18 and 20 indicate the presence of the three compounds P-m, Fe and Td, respectively. A fourth spectrum selected at time index 27 corresponds to the first compound of the second region (Ip).

NS was also applied in this time region and the results indicate the presence of three compounds with time indices 23, 19 and 20 (Fig. 9(b)). Only one time index differs slightly from those selected with OPA. The modified needle plot shows the mean chromatogram, the projection of each factor and the position of the selected times. In conclusion, both approaches give a good estimation of the number and location of the three compounds.

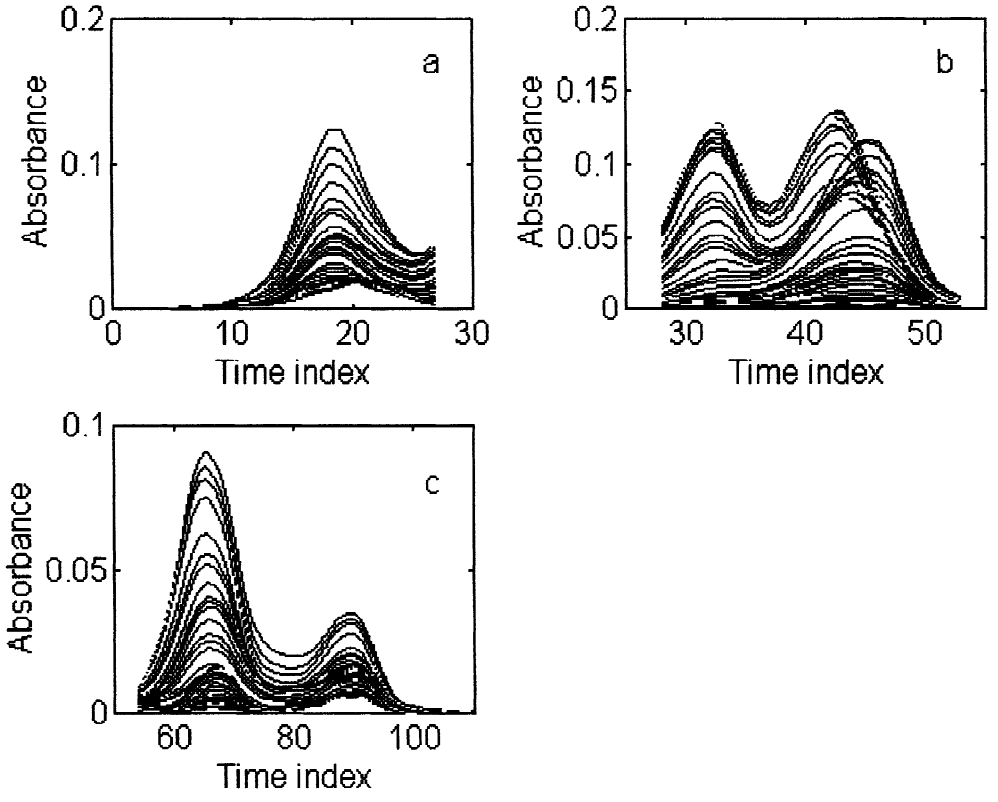

Fig. 8. Chromatograms corresponding to the three submatrices obtained from the whole data matrix: (a) region 1, (b) region 2 and (c) region 3.

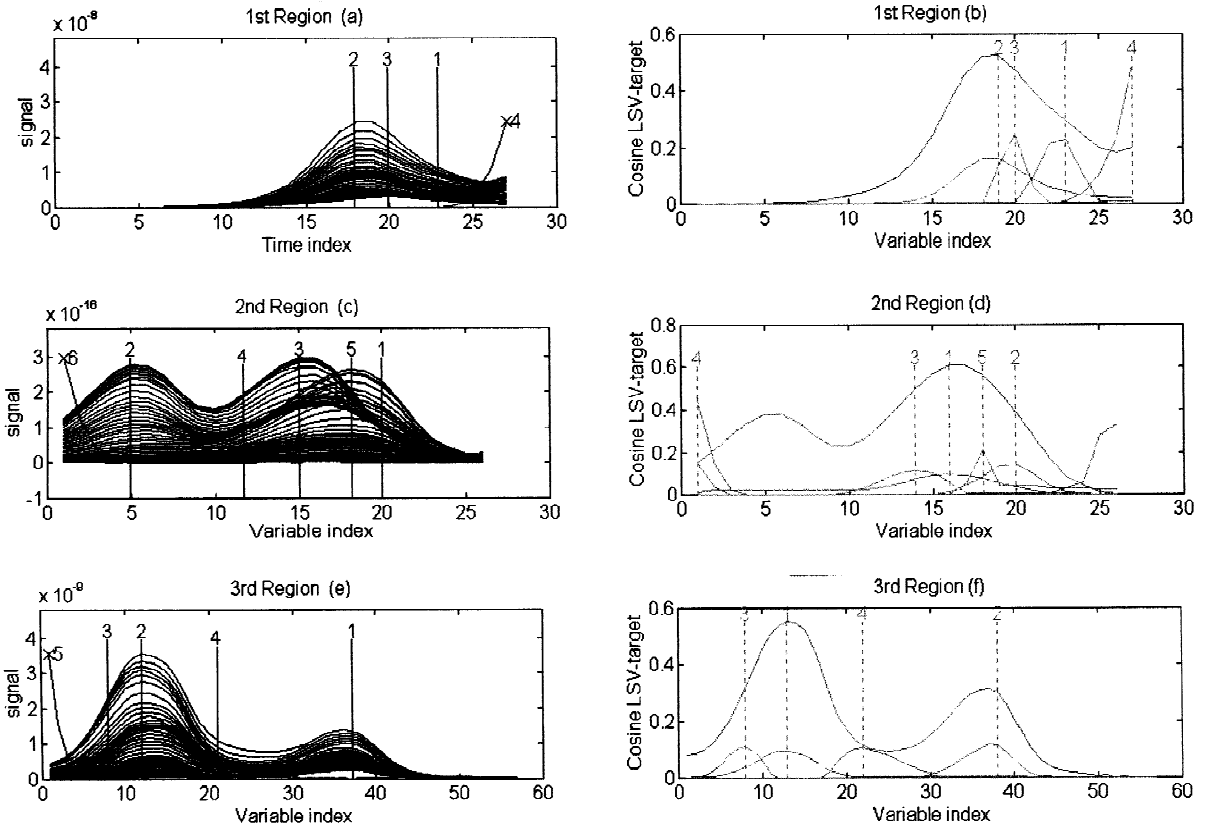

Fig. 9. Analysis of the 12-component mixture by OPA and modified NS, respectively: (a and b) region 1; (c and d) region 2; (e and f) region 3.

Region 2. This included time indices 28 to 53 (Fig. 7(b)). This is the more complex region; there are five compounds eluting, Ip, Ma, Tz, Pr and Ct. In fact, this region could be treated as a four-component system, since the first eluting compound (Ip) is reasonably separated from the other four compounds. The second, third and fourth compounds elute with low chromatographic resolution. The second compound, Ma, has a low signal relative to the other compounds, and its absorption spectrum does not present selectivity with respect to that of Pr.

Analysis of this region by OPA and NS indicates the presence of five compounds (Fig. 9(c and d)). The five spectra selected by OPA are those at time indices 47, 32, 42, 39 and 45. They correspond to the Ct, Ip, Tz, Ma and Pr compounds, respectively. The spectra selected by NS have indices 43, 47, 41, 28 and 45, which correspond to the Tz, Ct, Ma, Ip and Pr compounds, respectively. These time indices differ slightly from those selected by OPA, except for Ip which is selected at the beginning of its elution (time index 28). This compound was already selected in region 1 (time index 27, Fig. 9(b)). The selection order of the first, second and fourth compounds with OPA were different to that found with NS. With both approaches, Pr was the last selected compound, probably due to the perturbing effect of Ip, which has a similar absorption spectrum. However, eventually the OPA and NS approaches show similar results, giving the correct number of compounds as well as their correct location.

Region 3. This included the time indices 53 to 110 (Fig. 7(b)). Four compounds elute in the order Vi, Cf, Te and P-e. Tebuconazol (Te) has a poor chromatographic response with a wide and low peak. Analysis of this region by OPA and NS indicates the

presence of the four compounds (Fig. 9(e and f)). The four spectra selected by OPA are, in this order, those at time indices 90, 65, 61 and 74, which correspond to P-e, Cf, Vi and Te, respectively. The spectra selected by NS are those at time indices 68, 90, 61 and 74, which correspond to Cf, P-e, Vi and Te, respectively. With both approaches, the Te compound was selected last, probably due to the poor signal. In conclusion, both approaches give a good estimation of the number and location of the four compounds.

Comparing the time indices estimated with OPA and modified NS, using both the whole data matrix and the three submatrices, it can be concluded that similar results were obtained in the first two regions except for the Ip compound, which was selected at the beginning of its elution with NS in the second region. In the last region the Cf compound was selected with OPA at a slightly earlier time index (65) than with the whole data matrix (time index 68). The Te compound was also selected earlier (time index 74) with both OPA and NS than when using the whole data matrix (time index 78).

6. Conclusions

In this paper we demonstrate that by reducing a very complex multicomponent system to submatrices of peak clusters, extracted from the larger data matrix, the assessment of peak purity is improved. SIMPLISMA, OPA and the modified NS approaches were tested using the whole data matrix, but the results were not as good as one would hope. Due to the high complexity of the mixture the OPA algorithm was also slightly modified in order to detect smaller intensity peaks under the main compounds, but substantially better results were not obtained due to the concurrence of several sources of errors.

The application of OPA and the modified NS approaches in the three simpler submatrices allowed the correct determination and location of the number of compounds in each region. This strategy can be regarded as preferable to working with the whole matrix in the evaluation of peak purity for complex multicomponent systems. In this way the local complexity of the unresolved mixtures was much more favorable.

This strategy also appears interesting as a starting point for obtaining initial estimates for more sophisticated curve resolution methods, such as ALS (alternating least squares) [25,26], which allow recovery of the individual chromatograms and spectra of the individual compounds.

Acknowledgements

DLM thanks the Fonds voor Wetenschappelijk Onderzoek for financial assistance. AGF thanks the University of Almería (Spain) for financial support during the stay at FABI (Brussels).

References

[1] F. Cuesta Sánchez, B. van den Bogaert, S.C. Rutan, D.L. Massart, Chemom. Intell. Lab. Syst. 34 (1996) 139–171.
[2] F. Cuesta Sánchez, J. Toft, B. van den Bogaert, D.L. Massart, Anal. Chem. 68 (1996) 79–85.
[3] F. Cuesta Sánchez, S.C. Rutan, M.D. Gil García, D.L. Massart, Chemom. Intell. Lab. Syst. 36 (1997) 153–164.
[4] K. De Braekeleer, D.L. Massart, Chemom. Intell. Lab. Syst. 39 (1997) 127–141.
[5] K. De Braekeleer, F. Cuesta Sánchez, P.A. Hailey, D.C.A. Sharp, A.J. Pettman, D.L. Massart, J. Pharm. Biomed. Anal. 17 (1998) 141–152.
[6] W. Windig, J. Guilment, Anal. Chem. 63 (1991) 1425–1432.
[7] W. Windig, C.E. Heckler, Chem. Intell. Lab. Syst. 14 (1992) 195–207.
[8] F. Cuesta Sánchez, D.L. Massart, Anal. Chim. Acta 298 (1994) 331–339.
[9] R. Gargallo, F. Cuesta Sánchez, A. Izquierdo-Ridorsa, D.L. Massart, Anal. Chem. 68 (1996) 2241–2247.
[10] J. Toft, F. Cuesta Sánchez, B. van den Bogaert, F.O. Libnau, D.L. Massart, Vibrat. Spectrosc. 10 (1996) 125–138.
[11] O.M. Kvalheim, Y.-Z. Liang, Anal. Chem. 64 (1992) 936–946.
[12] Y.-Z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle, F. Erni, Anal. Chem. 64 (1992) 946–953.
[13] Y.-Z. Liang, O.M. Kvalheim, J. Chemom. 7 (1993) 15–43.
[14] H.R. Keller, D.L. Massart, Anal. Chim. Acta 246 (1991) 379–390.
[15] H.R. Keller, J.O. De Beer, Anal. Chem. 65 (1993) 471–475.
[16] J. Toft, O.M. Kvalheim, Chemom. Intell. Lab. Syst. 19 (1993) 65–73.
[17] F. Cuesta Sánchez, J. Toft, O.M. Kvalheim, D.L. Massart, Anal. Chim. Acta 314 (1995) 131–139.
[18] T.D. Jarvis, J.H. Kalivas, Anal. Chim. Acta 26 (1992) 13–24.

[19] G.A. Bakken, J.H. Kalivas, Anal. Chim. Acta 300 (1995) 173–181.
[20] B.G.M. Vandeginste, W. Derks, G. Kateman, Anal. Chim. Acta 173 (1985) 253–264.
[21] B.G.M. Vandeginste, G. Kateman, J.K. Strasters, H.A.H. Billiet, L. de Galan, Chromatographia 24 (1987) 127.
[22] G.G.R. Seaton, A.F. Fell, Chromatographia 24 (1987) 208.
[23] A. de Juan, B. van den Bogaert, F. Cuesta Sánchez, D.L. Massart, Chemom. Intell. Lab. Syst. 33 (1996) 133–145.
[24] P.J. Gemperline, J. Chem. Inf. Comput. Sci. 24 (1984) 206–212.
[25] R. Tauler, D. Barcelo, Trends Anal. Chem. 12 (1993) 319–327.
[26] R. Tauler, Chemom. Intell. Lab. Syst. 30 (1995) 133–146.