ELSEVIER

Chemometrics and Intelligent Laboratory Systems 29 (1995) 223–232

Chemometrics and intelligent laboratory systems

Rank mapping of three-way multicomponent profiles

Bjørn Grung, Olav M. Kvalheim *

Department of Chemistry, University of Bergen, N-5007 Bergen, Norway

Received 9 January 1995; accepted 18 April 1995

Abstract

In this work, fast graphic procedures for assessment of number of analytes (chemical rank) in local regions of three-way unresolved multicomponent profiles are presented. Collapsing the three-way profile along one direction by matrix summation and evolving rank analysis on the resulting matrix represents a first step for the approach developed here. This procedure is efficient for detecting analytes with low net analytical signal compared to the noise level. The effect of collapsing is a reduction of random noise due to cancellations, while structure from analytes is enhanced because of the additive effect of many similar contributions. In the second step, slicing of the three-way data array into complete two-way arrays (matrices) is performed. Projections are then used across these matrices to generate a rank map. The strategy developed in this work provides a rapid approach to rank analysis and permits the analyst to take into account that the analytical information is normally not uniformly distributed in the three directions.

Keywords: Rank mapping; Three-way multicomponent profiles

1. Introduction

Modern analytical instrumentation often leads to data arrays with more than two dimensions or directions. For instance, hyphenated techniques such as multidetection chromatography can lead to arrays of order higher than two when several samples of similar composition are analysed and compared.

A common feature of multiway analytical profiles is their evolving and continuous nature in time, wavelength or space (image analysis). Different regions may have contributions from different analytes. These features have generated an interest in procedures for local analysis of multicomponent profiles. For two-way analytical profiles, evolving factor analysis (EFA) [1,2] and heuristic evolving latent projections (HELP) [3,4] are examples of methods that include procedures for assessment of local chemical rank, i.e. number of analytes, as a crucial step prior to resolution. The systematic screening of the data matrix by means of a moving window, of either fixed [5,6] or evolving [1,2,7–10] size, is common to two-way evolutionary techniques for curve resolution. This produces a rank map from which the necessary information can be extracted to obtain resolved profiles of pure analytes.

Due to the size and structure of the data matrix, a three-way generalisation of the window technique would lead to a prohibitive number of possible sub matrices to be examined. Instead, we propose that a three-way data array is either collapsed by summing

* Corresponding author.

SSDI 0169-7439(95)00037-2

matrices in the least selective direction or sliced into two-way matrices in the most selective direction. This strategy takes into account that the analytical information may not be equally distributed in all directions. Time can thus be saved by concentrating the efforts on the most promising directions.

The aim of this work is to develop fast and accurate procedures for rank mapping of three-way multicomponent profiles with at least one continuous and evolving direction. These rank maps can, for example, be utilised for resolution of three-way analytical profiles [11].

2. Theory

2.1. Rank map determined by collapsing or slicing a three-way data array

Rank mapping of three-way profiles is performed for two reasons: (i) to determine how many analytes are present under the complete profile, and, (ii) to reveal information of rank trends in the evolving system. Both types of information are crucial for correct resolution of multicomponent profiles. A good strategy for rank mapping thus needs to incorporate both these aspects. As selective regions in three-way data contain crucial information for the subsequent resolution of the multicomponent profile, a first step is to decide which directions are least, respectively, most likely to contain selective information. For analytical techniques involving chromatographic separation, the retention time direction is usually the most selective direction. The chromatographic elution profiles are thus better suited for a rank analysis than the spectral profiles, as signal from an analyte in a spectrum often is distributed throughout the spectral profile. Furthermore, regions with no eluting analytes (zero-component regions) are rather well-defined in the chromatographic direction. These regions are used for estimation of the noise level. If we compare several similar samples with differing concentrations, the sample direction is usually the least selective direction.

When the decision about the least and most selective directions of the data array has been made, our

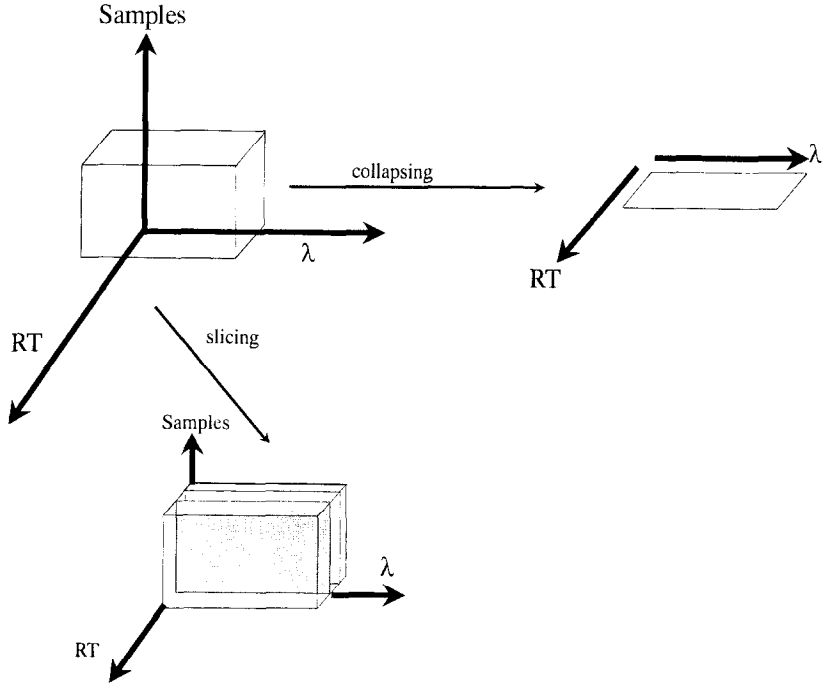

Fig. 1. Collapsing in the sample direction and slicing in the time direction for a three-way data array.

strategy for rank mapping of three-way multicomponent profiles goes in two steps (Fig. 1). First, we perform rank analysis on the matrix obtained by collapsing the three-way data array in the least selective direction. Collapsing is used to denote the operation of summing the matrices in that direction. Secondly, the three-way data array is sliced into its two-way constituents in the most selective direction. While collapsing leaves us with a single matrix for rank analysis, slicing through one of the directions in the data set provides several two-way matrices. Fortunately, complete rank analysis of all these matrices is not needed. Rather, shapes and magnitudes of various projections of these matrices are used to determine the local chemical rank, i.e. number of analytes present at a (time) point.

2.2. Procedures for rank mapping

In this paper, vectors are by definition column vectors. A transposed vector (superscript T) is thus a row vector. Indices i and j are used to denote retention time. The matrices $\mathbf{X}_i$ and $\mathbf{X}_j$ thus refer to the matrices acquired at retention time i and j, respectively.

Let us assume that we have collected a three-way multicomponent array containing a sample direction, a chromatographic direction and a spectral direction. The different samples all contain the same chemical components. We start our examination of the data by collapsing the three-way array into an ordinary matrix in the sample direction. Due to the additive effect of signal from analytes and cancellations of contributions from noise, this procedure enhances signal from analytes compared to noise in the collapsed matrix. This may be crucial for detecting analytes with small net analytical signal (NAS) [12]. The collapsed matrix can subsequently be analysed by means of the moving window techniques [5–10] and latent-projective graphs [3] to get an overview of the number of analytes in the full profile and in local regions. Eigenvalues and shape of score and loading vectors from the first few principal components extracted from the collapsed matrix can provide supplementary information for assessment of the total number of analytes under the collapsed multicomponent profile.

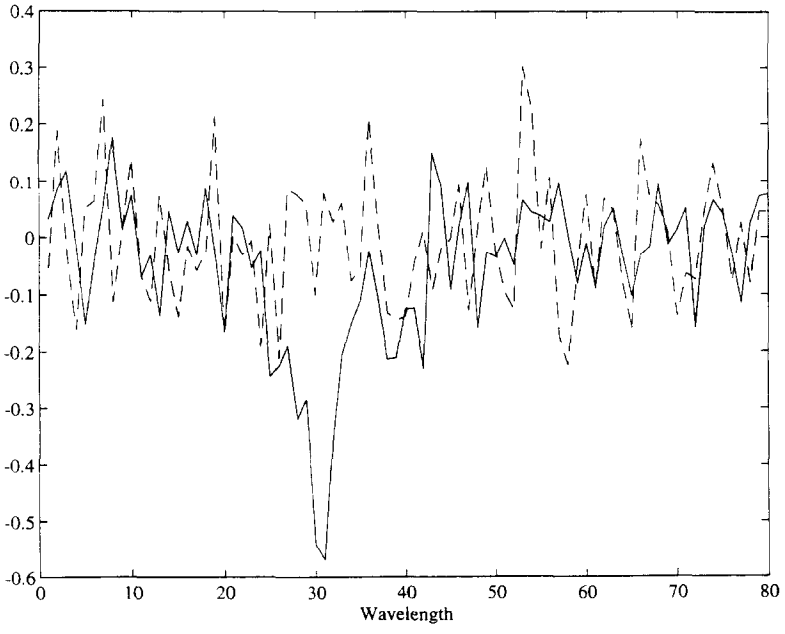

Fig. 2. Detection of the first analyte (solid line) by projecting onto a noise profile (dotted line). The solid line is noisy, but in the interval 25–35 structure is evident.

Retention time shifts between matrices will broaden the regions in which the analytes are detected, but they have no other effects on the rank maps.

We continue our examination of the data by slicing the three-way profile in the most selective direction. The first principal component is extracted from the matrix acquired at the first retention time. The shape of the first score and loading vector, $\mathbf{t}_i$ and $\mathbf{p}_i$, respectively, at the first retention time (i = 1) usually indicates that no analyte has yet eluted, i.e. that we are situated in a zero-component region [3]. Assuming the noise to be orthogonal to signal from analytes, the score and loading vectors at the first retention time can be used to locate the retention time at which the first analyte starts to elute. Projections of the matrices at retention time 2, 3,.. onto the profiles from the zero-component region show a noisy pattern until the concentration of the first analyte exceeds the limit of detection. Eq. (1a) shows this procedure for the spectral profile originating from a projection onto the score vector.

$$\mathbf{q}_j = \mathbf{t}_i^T \mathbf{X}_j / \|\mathbf{t}_i\| \quad (1a)$$

The index j represents the present retention time, $\mathbf{X}_j$ is the matrix at this retention time, $\mathbf{t}_i$ is a score vector extracted from the zero-component region at the retention time i and $\mathbf{q}_j$ the spectral profile obtained by projecting $\mathbf{X}_j$ on $\mathbf{t}_i$. Fig. 2 shows an example of the detection of an analyte by this procedure. Structure (peak) is observed around wavelengths corresponding to points 25–35 in the solid line. A similar operation produces the concentration profile $\mathbf{c}_j$ after projecting onto the loading vector extracted at retention time i:

$$\mathbf{c}_j = \mathbf{X}_j \mathbf{p}_i / \|\mathbf{p}_i\| \quad (1b)$$

The residual matrix obtained after subtraction of the projected profile calculated by Eq. (1a) represents another route to the location of start of elution:

$$\mathbf{E}_i = \mathbf{X}_i - \mathbf{t}_i \mathbf{q}_j^T \quad (2)$$

The matrix $\mathbf{E}$ contains only noise as long as the number of analytes remains unchanged. The norm of each column (or row) vector of $\mathbf{E}$ is subsequently calculated and collected in a residual norm vector. By plotting this vector with time, start of elution is detected as the time when this vector starts to show structure.

The two ways shown above for detecting the end of the zero-component region can actually be used throughout the analysis to detect rank changes. After the start of elution is located for the first analyte, two principal components are extracted from the matrix at this retention time. The second principal component represents noise, but information concerning the occurrence of a second analyte will appear in the projections [Eqs. (1a) and (1b)] onto this component. The matrices at the following retention times are thus projected onto these two principal components. Plots of the residual norm vector are also here used as a complementary way of detecting further analytes.

The procedures described above are repeated throughout the analysis. We always extract one principal component more than the current number of analytes from the slice where the rank changes. When the subsequent projections onto the noise profiles show structure, a new analyte has appeared. Since most of the operations are simple matrix multiplications (projections), this technique avoids large numbers of time-consuming PCA decompositions. However, experience has proved that further PCA decompositions are needed when the signal-to-noise-ratio (S/N-ratio) has changed substantially. The profiles obtained in the beginning of an analyte's elution period are significantly more noisy than the profiles obtained when decomposing a matrix originating from a retention time where the S/N-ratio is more favourable. Usually, two or three PCA decompositions are thus needed per analyte.

The techniques described above are sufficient to disclose the start of elution of every new analyte in a system. However, it is of course equally important to be able to correctly estimate the end of elution for an analyte. This is possible by the same procedures. It is sometimes useful to repeat the slicing procedure in the backward direction.

It is in fact not necessary to derive a rank map of the complete multicomponent profile. Rather the task is to obtain a clear picture of the *rank trend*. Thus, whether an interesting region starts exactly, e.g. at retention time 50 or 53 is usually of no importance. The crucial point is to make sure that we are able to pick slices representing unambiguously all the different rank situations in the multicomponent profiles. This is necessary for the subsequent resolution of the multicomponent system [11].

3. Experimental

Rank mapping was performed on four three-way data arrays representing three-component systems. The data were all simulating LC-DAD profiles of dimensions 80 × 100 × 10 (retention time × wavelength × samples) by means of MATLAB 4.2 for Windows on a 486 DX personal computer. Spectra and chromatograms were generated as sums of normalised Gaussian peaks and the matrices resulting for each analyte were scaled according to the concentrations and summed. The three-way data was made by stacking 10 ordinary two-way matrices of varying concentrations. Heteroscedastic noise (variance of noise depends on signal size) was added according to the following recipe: a three-way array of random homoscedastic noise (noise independent of signal size) with variance 0.0012 and zero mean was created. Heteroscedastic noise was calculated by element wise multiplication of the three-way data array and the homoscedastic noise matrix. The resulting noise matrix was added to the three-way multicomponent data arrays.

The four data arrays all contained one minor analyte coeluting with two major ones. Chromatographic resolution and spectral correlation were varied at two levels corresponding to fair and low chromatographic resolution (R_s) and net analytical signal (NAS) of the analytes. Figs. 3 and 4 show the chromatographic and spectral profiles, respectively, for fair and low levels of R_s and NAS. For each of the four systems, the concentrations of the three analytes in the 10 mixtures were varied according to Table 1.

4. Results and discussion

4.1. Rank anaiysis on collapsed matrices

Table 2 shows the first seven singular values of the collapsed matrix of the four simulated 3-component systems. The ratio between successive singular values suggests that 3 analytes coexist under all the examined profiles with the possible exception of the profile with both low chromatographic resolution (R_s) and low net analytical signal (NAS) for the minor analyte. For the latter system, the ratio between the singular values 3 and 4 is approximately 5.8. Fig. 5 shows the result of eigenstructure tracking analysis (ETA) [7,8] using a moving window of size 4 for the

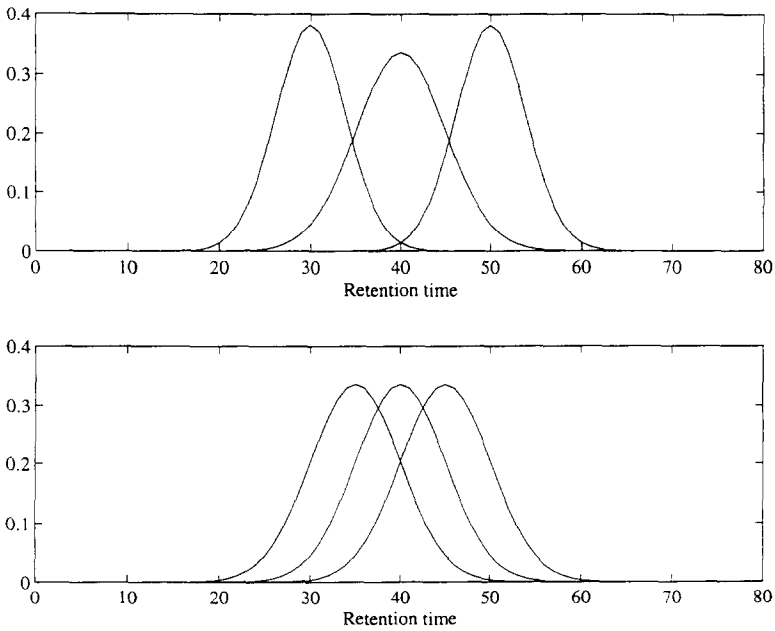

Fig. 3. Overlap of chromatograms of pure analytes for systems with fair (upper part) and low (lower part) chromatographic resolution.

Table 1
The relative concentrations of the three analytes in the 10 samples

Sample number	Analyte 1	Analyte 2	Analyte 3
1	49.07	1.96	48.97
2	2.42	2.91	94.66
3	48.01	3.85	48.14
4	59.89	4.76	35.35
5	19.21	5.66	75.13
6	59.89	6.54	33.57
7	55.49	7.41	37.10
8	58.56	8.26	33.18
9	41.94	9.09	48.97
10	24.93	9.91	65.16

Table 2
The first seven singular values from an SVD on the matrix collapsed in the sample direction for the four data sets. Data set 1 has both fair chromatographic resolution (R_s) and NAS. Data set 2 has a poor R_s and a fair NAS. Data set 3 has a fair R_s and a poor NAS. For data set 4, R_s and NAS are poor

Singular value	Data set 1	Data set 2	Data set 3	Data set 4
1	25.5938	27.2837	24.8124	26.4848
2	2.7346	1.9155	2.7004	1.8921
3	0.3656	0.1292	0.1139	0.0405
4	0.0074	0.0075	0.0074	0.0070
5	0.0071	0.0069	0.0072	0.0070
6	0.0068	0.0068	0.0068	0.0069
7	0.0066	0.0065	0.0067	0.0066
Ratio of fourth to third	49.4	17.2	15.4	5.79

systems with fair NAS and fair and poor R_s, respectively. In case of fair R_s (Fig. 5, upper part), it is obvious that fourth evolving eigenvalue corresponds to noise. For low R_s (Fig. 5, lower part), it is more difficult to decide whether the system contains three or four analytes due to the heteroscedastic noise in the fourth evolving eigenvalue. The singular values of the matrix obtained by collapsing the three-way array in the sample direction (Table 2) show, however, that there are only three analytes under the profile.

4.2. Rank analysis by projections and slicing

Tables 3 and 4 show the retention times of 'true' and detected rank changes using ETA after collapsing in the sample direction in combination with the projection procedure on the 3-way data sliced in the time direction. The overall impression is that the agreement between 'true' and detected is acceptable

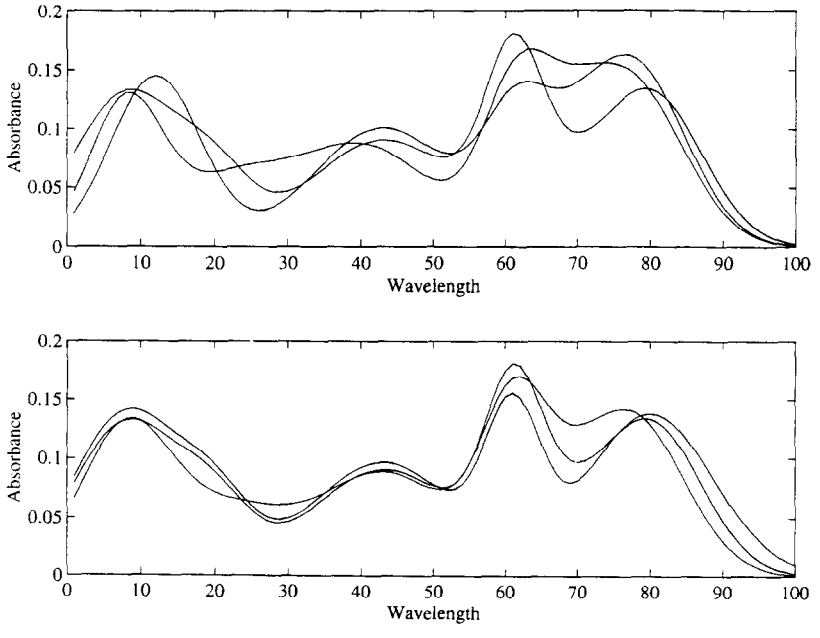

Fig. 4. Spectral overlap for systems with fair (upper part) and low (lower part) net analytical signal.

Table 3
The retention times of true and detected rank changes for the data sets with fair chromatographic resolution

Chemical rank	Detected RT		
	True RT	Fair NAS	Poor NAS
1	13	11	12
2	18	18	20
3	33	36	36
2	46	42	42
1	61	60	60
0	66	66	66

Table 4
The retention times of true and detected rank changes for the data sets with poor chromatographic resolution

Chemical rank	Detected RT		
	True RT	Fair NAS	Poor NAS
1	13	13	11
2	18	24	24
3	23	30	30
2	57	53	50
1	62	55	54
0	66	66	65

in most cases. The apparent disagreements between start and end of elution for some analytes can be ascribed to effects of noise and have no consequences for the success or failure of resolution. We shall now illustrate this and some other important points of the slicing procedures.

For the system with fair NAS but poor R_s, Fig. 6, upper part, shows the third spectral profile obtained by projecting the matrix at retention time 29 on the first three score vectors extracted by means of PCA of the matrix at retention time 27. The third loading vector extracted at time 27 (corresponding to noise) is shown together with the projected profile in order to enhance the interpretative power. Both vectors show the messy pattern characteristic of noise. Fig. 6, lower part, shows the same vectors for the projections of the matrix at the next retention time (RT = 30). Structure is now appearing in the third projected profile, in the wavelength interval 4–20. Proceeding to retention time 31, structure in the third projected

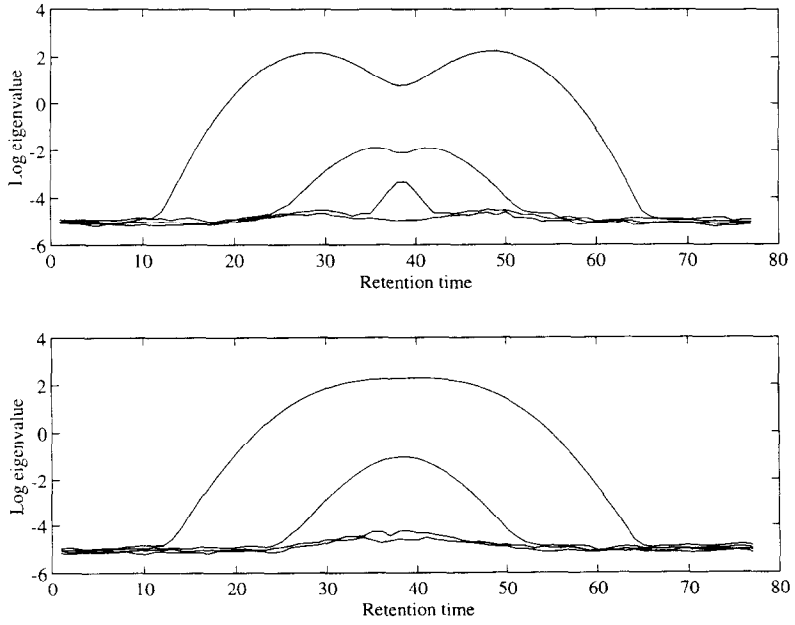

Fig. 5. (a) Result from eigenstructure tracking analysis [7,8] with window size 4 on the collapsed three-way matrix of the simulated system with fair chromatographic resolution and fair net analytical signal. The collapsing is performed in the sample direction. (b) Result from eigenstructure tracking analysis [7,8] with window size 4 on the collapsed three-way matrix of simulated system with low chromatographic resolution and fair net analytical signal. The collapsing is performed in the sample direction.

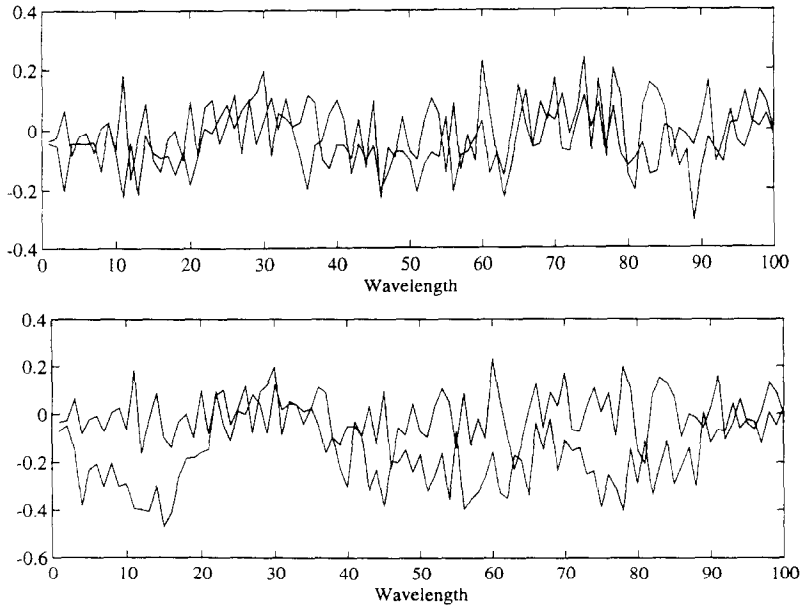

Fig. 6. (a) Projections according to Eq. (1a) of the matrix at retention time 29. The three score vectors used for projections (not shown) are extracted at retention time 27, where both analytes 1 and 2 have appreciable signal. The figure shows the third loading vector from PCA of the matrix at retention time 27 together with the third projected profile from the matrix at retention time 29. (b) Projections according to Eq. (1a) of the matrix at retention time 30. The three score vectors used for projections (not shown) are extracted at retention time 27, where both analytes 1 and 2 have appreciable signal. The figure shows the third loading vector from PCA of the matrix at retention time 27 together with the third projected profile from the matrix at retention time 30.

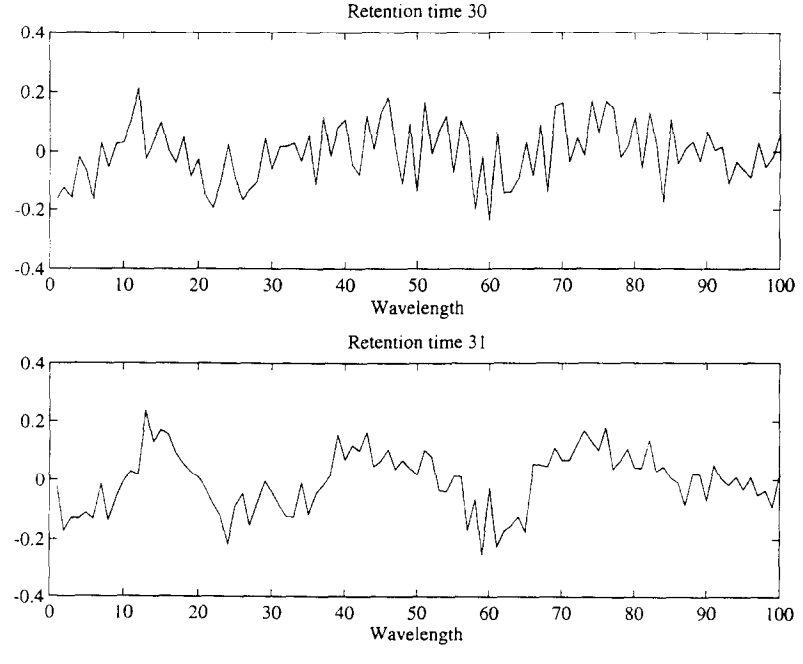

Fig. 7. The spectral loading vector of the third principal component extracted at retention time 30 and 31, respectively.

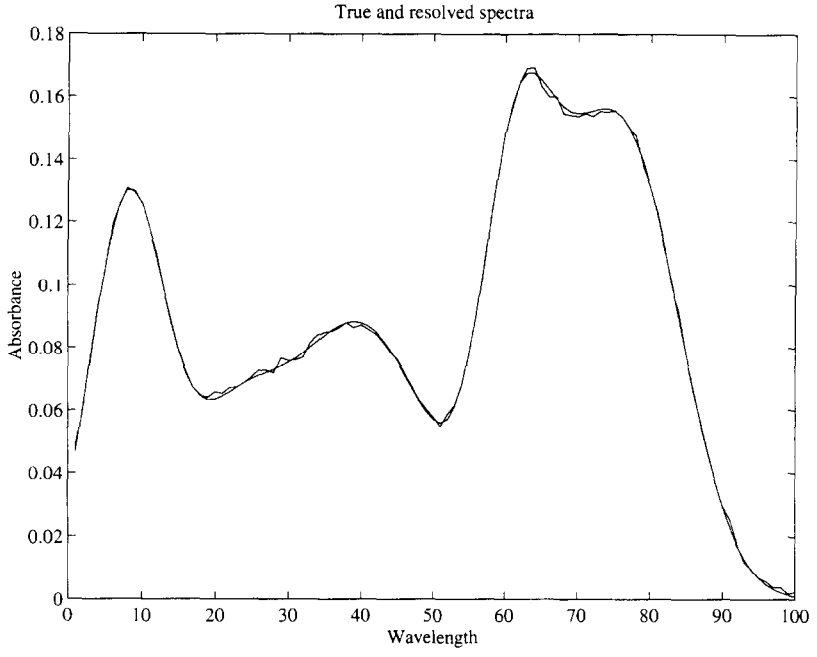

Fig. 8. Resolved and true spectrum of analyte 1 extracted by means of PCA of the matrix at retention time 23. We observe an almost perfect match between the two profiles.

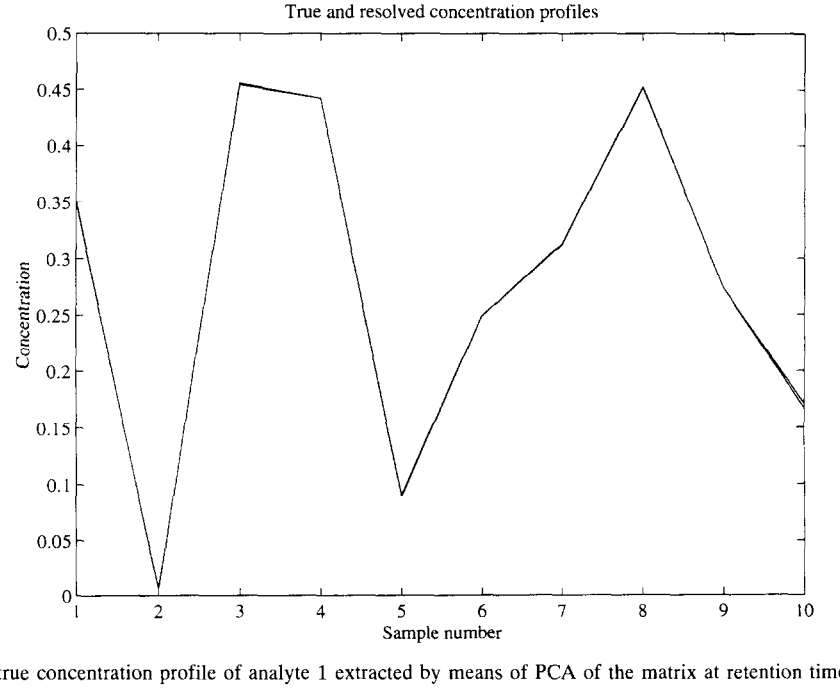

Fig. 9. Resolved and true concentration profile of analyte 1 extracted by means of PCA of the matrix at retention time 23. We observe an almost perfect match between the two profiles.

profile becomes evident (not shown). PCA of the matrices at retention times 30 and 31 gives the spectral patterns shown in Fig. 7. Only the third loading vector is shown. Structure is evident at retention time 31 (Fig. 7, lower part), and is also vaguely appearing at retention time 30 (Fig. 7, upper part). Thus, the projection procedure provides us with retention time 30 as start of elution for the third analyte. According to the simulation producing this system, this is too late. The third analyte should really start to show up at retention time 23. We observe the same discrepancy for the second analyte in this system. While the second analyte should appear at time 18 (Table 3), it is detected at time 24 (Table 3). Let us see how this can be explained.

At time 23, Table 4 shows that the estimated chemical rank is 1, while the 'true' chemical rank is 3. All analytes contribute to the signal, but noise and low NAS prevents detection of more than one analyte. Figs. 8 and 9 compare the spectrum and concentrations of the first analyte with the loadings and the scores, respectively, of the first principal component of the matrix at retention time 23. The agreement is excellent suggesting that the matrix at retention time 23 is truly a single-analyte region. Otherwise, contributions from the second and third analyte would mix with the first analyte in the first principal component. This shows that the added heteroscedastic noise is responsible for the delayed detection of analytes. For rank analysis of real systems, this effect is impossible to observe since there is no way to obtain noise-free data for rank analysis.

5. Conclusions

In this paper, fast and efficient procedures for rank mapping of three-way multicomponent profiles have been presented. Detailed information about local chemical rank from rank maps is crucial for resolution of a multicomponent data array into the profiles of the pure analytes. The strategy based on collapsing and slicing has been shown to represent a useful compromise between analysis time and information extracted. Furthermore, this strategy is not restricted to three-way multicomponent profiles. The principles of the procedures are just as valid for multicomponent profiles with more than three directions. The only difference is that collapsing needs to be performed in more than one direction before analysis and that collapsing may be necessary prior to slicing.

Acknowledgements

Bjørn Grung is grateful for a Ph.D. grant on a strategic technology program in chemometrics awarded by The Norwegian Research Council (NFR). An anonymous reviewer is thanked for useful comments.

References

[1] M. Maeder and A.D. Zuberbuhler, Anal. Chim. Acta, 181 (1986) 287.
[2] M. Maeder, Anal. Chem., 59 (1987) 527.
[3] O.M. Kvalheim and Y. Liang, Anal. Chem., 64 (1992) 936.
[4] Y.-Z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle and F. Erni, Anal. Chem., 64 (1992) 946.
[5] P. Geladi and S. Wold, Chemom. Intell. Lab. Syst., 2 (1987) 273.
[6] H.R. Keller and D.L. Massart, Anal. Chim. Acta, 246 (1991) 379.
[7] Y.-Z. Liang, O.M. Kvalheim, A. Rahmani and R.G. Brereton, J. Chemom., 7 (1993) 15.
[8] J. Toft and O.M. Kvalheim, Chemom. Intell. Lab. Syst., 19 (1993) 65.
[9] J. Toft and O.M. Kvalheim, Anal. Chem., 65 (1993) 2270.
[10] F. Cuesta Sánchez, J. Toft, O.M. Kvalheim and D.L. Massart, Anal. Chim. Acta, in press.
[11] B. Grung and O.M. Kvalheim, Chemom. Intell. Lab. Syst., 29 (1995) 213.
[12] A. Lorber, Anal. Chem., 58 (1986) 1167.