Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1096 (2005) 2–15
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Application of multivariate curve resolution methods to on-flow LC-NMR

Mohammad Wasim, Richard G. Brereton *

School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

Available online 29 June 2005

Abstract

The application of evolving window factor analysis (EFA), subwindow factor analysis (SFA), iterative target transformation factor analysis (ITTFA), alternating least squares (ALS), Gentle, automatic window factor analysis (AUTOWFA) and constrained key variable regression (CKVR) to resolve on-flow LC-NMR data of eight compounds into individual concentration and spectral profiles is described. CKVR has been reviewed critically and modifications are suggested to obtain improved results. A comparison is made between three single variable selection methods namely, orthogonal projection approach (OPA), simple-to-use interactive self-modelling mixture analysis approach (SIMPLISMA) and simplified Borgen method (SBM). It is demonstrated that LC-NMR data can be resolved if NMR peak cluster information is utilised.

Keywords: On-flow LC-NMR; Multivariate curve resolution; MCKVR; Factor analysis

1. Introduction

The analytical chromatographer frequently deals with the chromatography of mixtures, in most cases employing coupled chromatography. There are two fundamental approaches to the resolution of complex mixtures, the first is to improve physical separations, e.g. by optimising chromatography, and the second is to use computational methods for resolution, as discussed in this paper. There are numerous methods of coupled chromatography all with different characteristics both in terms of spectroscopy (e.g. sensitivity, selectivity) and chromatography (e.g. as a consequence of the detection method there may be limitations). Most chemometric or computational methods for resolution have been applied to liquid chromatography diode detection (LC-DAD), liquid chromatography mass spectrometry (LC–MS) and gas chromatography mass spectrometry. Liquid chromatography nuclear magnetic resonance (LC-NMR) poses specific challenges that many existing methods, designed for different systems, are unable to cope well with, especially in the limits of resolution: this paper tackles this problem.

On-flow LC-NMR has an important role in various fields [1–8] especially in the confirmation of chemical composition of mixtures. It is faster compared to the stop-flow LC-NMR. The main disadvantage of on-flow LC-NMR is its lower sensitivity compared to many other common methods and so it requires higher sample concentration, which can result in column overloading and cause severe overlap of chromatographic peaks. On-flow LC-NMR data is thus usually characterised by low signal-to-noise ratios and poorly resolved chromatographic peaks. LC-NMR data can be treated as evolutionary or two-way data. In two-way data each row represents a spectrum and each column represents a chromatographic or elution profile at a single variable. After acquiring data, different chemometric methods can be applied to extract the required information. In recent publications, chemometric analysis of on-flow LC-NMR has been used for retention time measurement [9], rank determination [10,11] and curve resolution [12–14].

Multivariate curve resolution (MCR) methods are a group of chemometric approaches suitable for multidimensional data and their purpose is the correct determination of response profiles of individual components in time as well as in the spectral dimension when mixtures cannot be resolved

* Corresponding author. Tel.: +44 117 9287658; fax: +44 117 9251295.
E-mail address: r.g.brereton@bris.ac.uk (R.G. Brereton).

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.05.101

simply by using the instrument. The methods have been classified in different ways [13–15] including both modeling and self-modeling curve resolution (SMCR) methods [16]. Modeling methods force a specific mathematical model for example the shape of elution profile [17] or the shape of a curve in kinetics [18]. Self-modeling methods do not demand a priori information about the spectral or concentration profiles but apply natural constraints [19] such as unimodality and non-negativity. SMCR can further be categorised as iterative, non-iterative and hybrid according to the algorithm used. Commonly used iterative methods include iterative target transformation factor analysis (ITTFA) [20,21], alternating least squares (ALS) [22,23], positive matrix factorization [24] and simplex-based methods [25]. Methods which take advantage of local rank information and are non-iterative in nature include evolving factor analysis (EFA) [26,27], window factor analysis (WFA) [28,29], heuristic evolving latent projections (HELP) [30,31], subwindow factor analysis (SFA) [32,33] and parallel vector analysis (PVA) [34]. A third category consists of hybrid methods like automatic window factor analysis (AUTOWFA) [35], and Gentle [36]. Two new methods have specifically been reported recently for LC-NMR, belonging to the last category, include canonical correlation analysis (CCA) [12] and constrained key variable regression (CKVR) [14]. Most multivariate methods for regression were first reported in the context of LC-DAD and infrared spectroscopy (IR) where noise level and chromatographic resolution are not such a serious problems; in those datasets these methods have usually yielded excellent results.

Curve resolution for LC-NMR data is a challenge to the chemometrician where most of the methods rely on a special kind of data structure called 'bilinear'. Bilinear structure in the data is perturbed by high levels of noise and other factors adding non-linearity in the data. Therefore, it becomes important to check the applicability of various curve resolution methods to data obtained with different types of instruments. In the present study we compare some of the curve resolution methods on LC-NMR and discuss CKVR in more detail.

The different approaches are tested on data containing eight compounds, among which are two sets of regioisomers. All the compounds elute closely thus provide an opportunity to check full potential of the curve resolution techniques.

2. Experimental

A mixture was prepared consisting of 2,6-dihydroxynaphthalene (I) (98%, Avocado, Research Chemicals Ltd., Heysham, UK), 2,3-dihydroxynaphthalene (II) (98%, Acros Organics, Geel, Belgium), diethyl maleate (III) (98%, Avocado), methyl *p*-toluenesulphonate (IV) (98%, Lancaster, Morecambe, UK), diethyl fumarate (V) (97%, Avocado), 1,2-diethoxybenzene (VI) (98%, Lancaster), 1,4-diethoxybenzene (VII) (98%, Lancaster) and 1,3-diethoxybenzene (VIII) (95%, Lancaster) each 50 mM. Acetonitrile (HPLC grade, Rathburn Chemicals, Walkburn, UK) and deuterated water (Goss Scientific Instruments Ltd. Great Baddow, Essex, UK) were used as solvent in 80:20 (v/v).

2.1. Chromatography

Chromatography was performed on a Waters (Milford, MA, USA) HPLC system, which consisted of a 717 plus autosampler, a 600s Controller, a model 996 diode array detector with a model 616 pump. DAD was used to visualise the appearance of the first component, which triggered the acquisition of NMR spectra. Acetonitrile (Rathburn) and deuterated water (Goss Scientific) were used as mobile phase in concentrations 80:20 (v/v) with flow rate 0.2 ml min^{-1} and injection volume 50 μl. A few drops of tetramethylsilane (TMS) (Goss Scientific) were added as a chemical shift reference. All the compounds were eluted within 10 min.

2.2. Spectroscopy

A 4 m PEEK tube with width of 0.005 in. was used to connect the eluent from the Waters HPLC instrument to a flow cell (300 μl) into the NMR probe on a 500 MHz NMR spectrometer (Jeol Alpha 500, Tokyo, Japan). For each spectrum, the spectral width was 7002.8 Hz, the pulse angle 90°, acquisition time 1.1698 s and pulse delay 2 s. The digital resolution was 0.855 Hz and chromatographic resolution was 3.1698 s. The acetonitrile singlet resulting from solvent was suppressed by pre-saturation using a DANTE sequence [37] A contour plot with sum of spectral and concentration profiles are presented in Fig. 1, where the solvent peak has been removed from the data and all pure LC-NMR spectra are shown in Fig. 2.

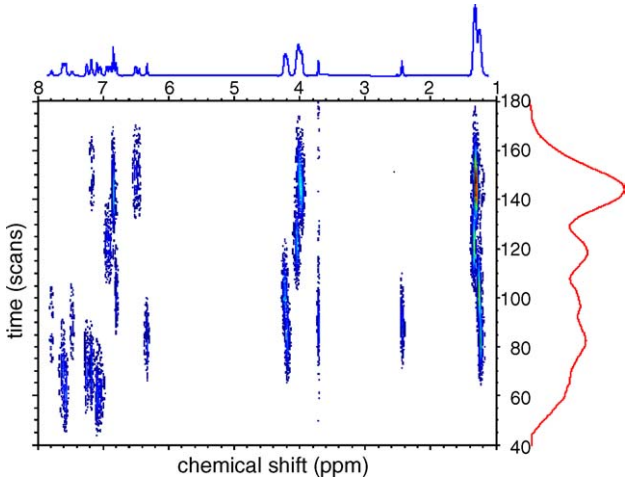

Fig. 1. Contour plot of whole data set with sum of spectral and concentration plots, the solvent peak was removed from the data.

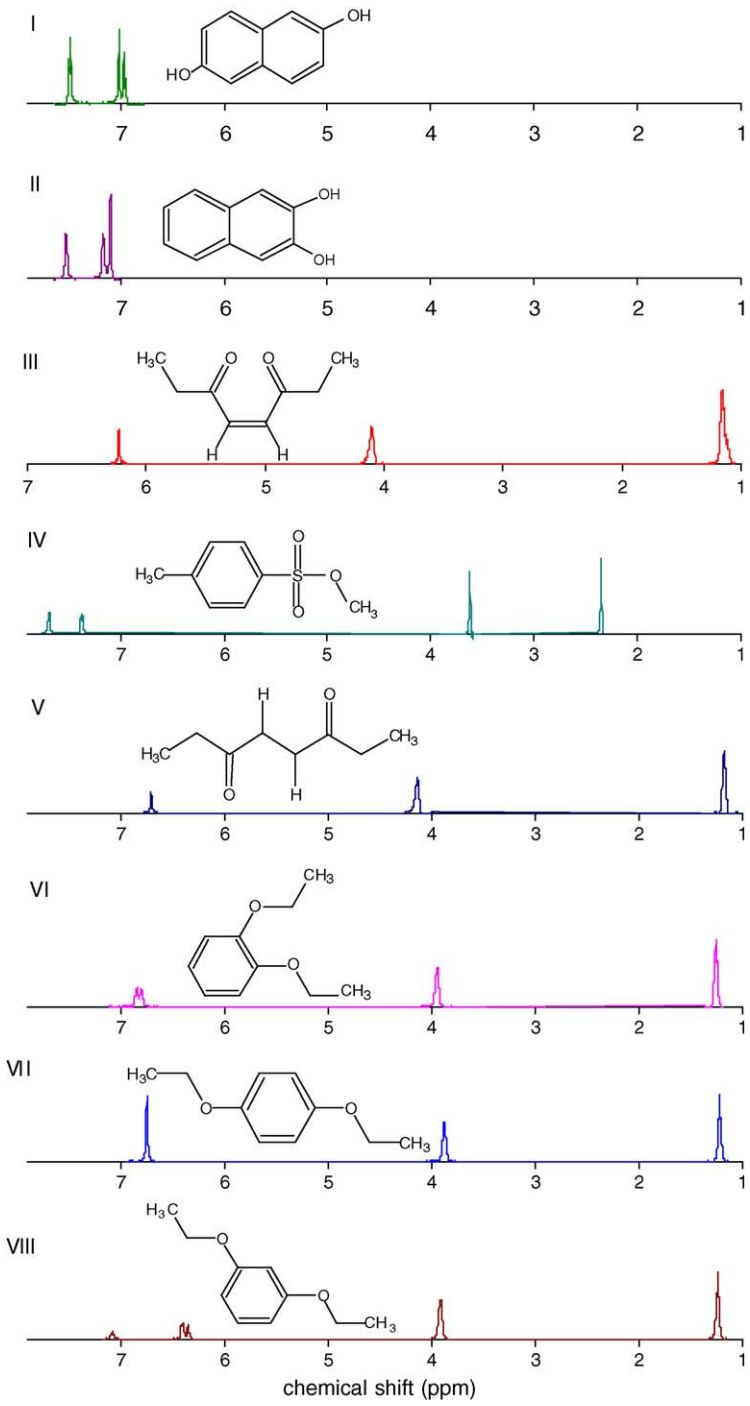

Fig. 2. Pure LC-NMR spectra of all compounds.

2.3. Software

The data analysis was performed by computer programs written in Matlab by the authors of this paper except the following: Fourier transformation and pre-processing of LC-NMR data was performed by in-house written software called LCNMR [14]. The ALS routines were obtained from the website [38] maintained by Tauler and co-workers and ITTFA and AUTOWFA from the website [39] maintained by Gemperline, which are part of a software called 'GUIPRO' [40]. All of the programs were used under Matlab 6.0.

3. Data analysis

In this paper seven methods are compared namely, iterative methods ALS and ITTFA, non-iterative methods EFA and SFA and hybrid methods Gentle, AUTOWFA and CKVR.

3.1. Preprocessing

Data was obtained in the time domain as free induction decays (FIDs), its size was 256 FIDs × 8192 spectral frequencies. The first FID was removed because it contained artefacts. The remaining FIDs were apodised, Fourier transformed and phase corrected [41]. In order to remove errors due to quadrature detection, which results in regular oscillation of intensity, a moving average over every four points in time was performed in the chromatographic direction [9]. The chromatographic regions where no compound elutes were truncated, which was performed by plotting a sum of all chromatographic profiles against time and deleting the regions by visual inspection. Similarly, the spectral region of the solvent peak and others where only noise exists were discarded (as described in Section 3.2.3.4.(b)). After preprocessing, the data size was reduced to 142 FIDs × 1410 spectral frequencies.

3.2. Curve resolution

The two-dimensional data matrix of intensity measurements is denoted by $\boldsymbol{X}$ ($M \times N$) with m as index of rows and n as index of columns. K is the total number of chemical components and k is the index of these components. Most of the multivariate data analysis methods require bilinear behaviour. Mathematically, a bilinear data matrix can be written as:

$$\boldsymbol{X} = \boldsymbol{CS} + \boldsymbol{E} \tag{1}$$

where $\boldsymbol{C}$ consists of the concentration profiles, $\boldsymbol{S}$ the spectra of the pure components and $\boldsymbol{E}$ is an error matrix. For a K component system, each measurement arises as a sum of individual measurements, as shown in Eq. (2)

$$x_{mn} = \sum_{k=1}^{K} c_{mk} s_{kn} + e_{mn} \tag{2}$$

where c is the concentration and s is the spectral intensity of each compound at a specific frequency and unit concentration.

In LC-NMR there are many factors, which affect bilinearity, for example noise, temperature induced shifts of the observed peaks and data preprocessing steps (e.g. baseline correction and smoothing). The performance of curve resolution methods will vary according to instrument, dataset, noise level, noise structure and chromatographic resolution.

3.2.1. Non-iterative methods

These methods, also called unique resolution methods [15], provide unique and true resolution when information arising from each component is uniquely defined mathematically. This information may be in the form of selective concentration regions, local rank or zero concentration windows. Although these methods appear robust mathematically in practice these do not necessarily yield high quality results due to the laborious process of locating selective chromatographic regions accurately, which is even more difficult in LC-NMR than in LC-DAD, which is due to the much broader and noisier chromatographic peakshape, often a consequence of high column loading required to obtain adequate spectral intensity.

3.2.1.1. Determination of local rank or elution windows. There are several methods to locate the regions where each component elutes, such as PC plots (score plots) [30,31], evolving principal component analysis (EPCA) [42], fixed size moving window-evolving factor analysis (FSW-EFA) [43] and plotting concentration profiles using key variables [11].

3.2.1.2. Resolution. Once the elution window is determined for each compound, non-iterative curve resolution methods can be applied. Below is a description of three of these approaches.

Evolving factor analysis (*EFA*): The method was developed by Gampp at al. [26] and Maeder [27]. We use only Maeder's approach in this paper, which is described below.

(a) Principal component analysis (PCA) [44,45] is performed on the data matrix $\boldsymbol{X}$ for K components.

$$\boldsymbol{X} = \boldsymbol{TP} + \boldsymbol{E} \tag{3}$$

where $\boldsymbol{T}$ is the scores matrix, $\boldsymbol{P}$ is the loadings matrix and $\boldsymbol{E}$ is the residual matrix. A rotation or transformation matrix $\boldsymbol{R}$ ($K \times K$) is calculated by locating zero concentration regions for each component.

(b) An individual concentration profile is calculated by using the rotation vector for each component

$$\boldsymbol{c}_k = \boldsymbol{T}\boldsymbol{r}_k \tag{4}$$

Steps (a) and (b) are repeated until all (K) analytes are resolved to give a matrix $\boldsymbol{C}$.

(c) Spectral profiles are obtained using least squares by:

$$\boldsymbol{S} = (\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{X} \tag{5}$$

Although EFA has an excellent theoretical background, in practice this method is often difficult to apply because it is hard to locate zero concentration windows. Most of the methods used to locate zero concentration regions, rely on using eigenvalue-based methods (e.g. EPCA), which in LC-NMR often predict windows that are narrower than the true windows [12,14].

Subwindow factor analysis (*SFA*): This method [32,33] extracts a spectral profile using two concentration regions (windows) where only the analyte of interest elutes. The regions are called the left and right subwindows. The determination of the left and right windows is made using EPCA or FSW-EFA. The method is useful when the resolution of all compounds cannot be attained and the main interest is to determine information about a specific component.

In addition to identifying one component without knowing the other components SFA has many other advantages. The largest eigenvalue produced during the calculation is an indicator of the quality of results, so all the results which have lower eigenvalues, can be discarded because small eigenvalues suggest low correlation between spectra from the left and right subwindows. When SFA is performed and results from the left and right subwindows are visualised some information can be obtained to improve results. When the boundaries of subwindows are not very clear and an impurity peak appears only from one subwindow then the size of that subwindow can be adjusted to remove that peak. In this way, SFA helps to define boundaries of each concentration profile with more accuracy with LC-NMR data. A major disadvantage of using SFA is that in situations when most of chromatographic profiles severely overlap, pure spectra cannot be obtained because left and right windows are very small.

3.2.2. Iterative methods

Iterative methods start with a set of starting profiles, one for each compound, which are either for the concentration or for the spectrum and then improves the initial profiles iteratively. In each iteration physical constraints are applied such as unimodality, non-negativity or closure. The methods stop when convergence is achieved. Convergence is set either by using error functions from the residual matrix or by limiting the number of iterations. There is a chance that iterative methods stop in local minima or diverge and do not provide correct solutions.

Since iterative methods start with estimated profiles, which in turn are obtained by determining key or pure variables, the next section will describe some common methods for finding the pure variables.

3.2.2.1. Selection of key variables. Variables which provide concentration or spectral profiles for single compounds from the original data, are called pure variables [46] or key variables [47,48]. Several methods for variable selection have been reported in the literature including SIMPLISMA [46], OPA [49–54], key set and iterative key set factor analysis (IKSFA) [47,48], latent projections [30] and the Simplified Borgen method (SBM) [55].

In hyphenated chromatography key variables can be found in both directions, i.e. time as well as spectral. In favourable cases, there are characteristic resonances (pure variables) for every compound in LC-NMR. In this paper we discuss only three single variable selection approaches, OPA, SIMPLISMA and SBM.

Orthogonal projection approach (*OPA*): OPA [49–54] is a stepwise approach and selects one pure or key variable in each step. The method calculates dissimilarity based on the mathematical concept of orthogonalisation [56]. The method compares each spectrum with one or more than one reference spectra. The first dissimilar plot represents a comparison of each spectrum with the average spectrum. The dissimilarity is plotted against the time, to give a dissimilarity plot, and a maximum is located. The first key variable corresponds to the maximum in the first dissimilarity plot. In the second step the average spectrum is replaced by the first reference spectrum corresponding to the first key variable. The second key variable is identified by locating a maximum point in the new dissimilarity plot, providing a second reference spectrum. In the third and later steps reference spectra identified in the previous steps are included in the calculations. The final dissimilarity plot exhibits only a random pattern. The total steps required by OPA are equal to one plus the number of compounds present in a dataset. The reference spectra are the purest possible spectra with significant signal-to-noise ratios for each compound.

SIMPLISMA: SIMPLISMA [46] is similar to OPA in operational details but it selects purest possible variables as compared to the most dissimilar variables selected by OPA. The calculations in SIMPLISMA are similar to OPA except it utilises a ratio of standard deviation to mean intensity of each spectrum. A factor called 'offset' is introduced to avoid those variables with very low mean intensity, i.e. noise. The method measures the purity of each spectrum, which is plotted against elution time and pure variables or spectra are selected by locating a maximum in the graph. In the first step, there is only one spectrum in the matrix, but in the later steps reference spectra are added to it as in OPA.

Simplified Borgen method (*SBM*): The SBM method selects pure variables with significant signal-to-noise ratios. The method assumes that the number of analytes or components in the data is known. It introduces an offset factor, which is similar to SIMPLISMA to reduce the effect of noise. SBM first decomposes the data matrix by PCA and then normalizes it so that every variable has a constant projection on the first PC. The key variables are obtained by locating maximum in the norm of the normalized data. A more detailed algorithm and Matlab code is available in [55].

The SBM selects most significant vectors as compared to the most pure vectors, which is also true for OPA, while SIMPLISMA selects pure spectra with more noise than the SBM or OPA selected spectra.

3.2.2.2. Resolution. Three methods are described for curve resolution from the iterative class of methods.

Alternating least squares (*ALS*): There are several ways of performing ALS, the procedure used in this paper is as follows.

(a) Perform PCA on $\boldsymbol{X}$ using Eq. (3).
(b) Retain the first K principal components (PCs), where K is the data rank assumed to be known.
(c) Reconstruct the data $\boldsymbol{X}_{\text{red}}$ using the scores and loadings matrices for K number of components.
(d) Since we are comparing the three variable selection methods: OPA, SIMPLISMA or SBM, therefore all were applied to obtain key variables and initial estimates of concentration or spectral profiles. In practice only one variable selection method is applied at a time. We will

suppose concentration profiles $\boldsymbol{C}$ are used as initial guess from one of the methods.

(e) Obtain spectral profiles $\boldsymbol{S}$ under non-negativity constraints using non-negative least squares (NNLS) [57].

(f) Obtain concentration profiles $\boldsymbol{C}$ under non-negativity and unimodality constraints [54].

Steps (e) and (f) are repeated until convergence, which is measured by the residual error between $\boldsymbol{X}_{\text{red}}$ and $\boldsymbol{CS}$ or when the number of iterations is increased to a predefined number. In our study we applied the unimodality constraint on concentration profiles and the non-negativity constraint on concentration and spectral profiles.

Although ALS calculations can be slow when the dataset is large, it is simple and widely used [58–62].

Iterative target transformation factor analysis (*ITTFA*): ITTFA iteratively resolves chromatographic profiles by finding a suitable transformation vector. When PCA is applied to $\boldsymbol{X}$, as in Eq. (3), a scores matrix $\boldsymbol{T}$ and a loadings matrix $\boldsymbol{P}$ are generated. The scores can be converted to real concentration profiles by finding a suitable rotation or transformation matrix $\boldsymbol{R}$

$$\boldsymbol{C} = \boldsymbol{TR} \tag{6}$$

The matrix $\boldsymbol{R}$ can be estimated by generating a test profile and then improving that profile iteratively. Different types of test vectors have been suggested, Gemperline [20] proposed selecting test vectors by needle search [63], while Vandeginste et al. [21] proposed the use of Varimax rotation [64]. We used Gemperline's approach of needle search which is described in [20]. The stopping criteria used in the iteration do not guarantee convergence to the optimal solution in all situations [65]. ITTFA has been used in many curve resolution applications as well as in kinetics [66,67].

3.2.3. Hybrid methods

This is a group containing some features of both iterative and unique resolution methods. Most methods require some parameters based on noise level, which are used to get better results. This feature makes these methods interactive and requires parameter adjustments many times.

3.2.3.1. Automatic window factor analysis (AUTOWFA). Window factor analysis and evolving window factor analysis, both self-modeling curve resolution methods, require information about the zero concentration windows of each component. Usually, evolving principal component analysis is used to collect this information. As this process involves a visual inspection of evolving eigenvalue plots, in forward and backward directions, its results can be influenced by personal judgement. AUTOWFA was developed to overcome this problem. In the original description the number of components was determined by PCA [15,16] but we applied needle search [63]. AUTOWFA makes use of iterative key set factor analysis [48] for finding pure concentration profiles along the time axis. The uniqueness test (UNIQ) [15,16] predicts the concentration profiles using target transformation [20] test vectors and measures the limits of each concentration window. In the final step window factor analysis [28,29] is applied and spectral profiles generated. The windows are improved by locating the edges of concentration profiles with respect to a predefined noise level and the process is repeated until no changes in the windows limits are found. A more detailed explanation of the method can be found in references [35,48].

3.2.3.2. Gentle. The method involves the following steps [36].

(a) Obtain key spectra $\boldsymbol{S}$ using a variable selection method.

(b) Obtain concentration profile $\boldsymbol{C}$

$$\boldsymbol{C} = \boldsymbol{XS}'(\boldsymbol{SS}')^{-1} \tag{7}$$

(c) Locate minima of each $\boldsymbol{C}$ profile by comparing with a preset intensity tolerance value. Remove negative parts in concentration so that the minimum of $\boldsymbol{c}_k$ after transformation equals the negative of the intensity tolerance value. The changes in concentration profiles are compensated in the spectral profiles.

(d) Remove side peaks (bimodality) in the concentration profiles by ordinary least squares and same correction is applied to correct spectral profiles.

Gentle is a quick procedure but there are not many applications in the literature [68,69].

3.2.3.3. Constrained key variable regression (CKVR). The original procedure presented in reference [14] is described as follows.

(a) SBM [55] is used in CKVR to determine the key variables equal to the rank of data, which are utilised to locate concentration profiles in data matrix $\boldsymbol{X}$ producing a matrix $\boldsymbol{C}_{\text{SBM}}$ ($M \times K$) and all concentration profiles are sorted according to peak maxima.

(b) Negative parts of concentration profiles are removed using as described in Section 3.2.3.2.

(c) Calculate the matrix of spectral profiles $\boldsymbol{S}$ by

$$\boldsymbol{S} = (\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{X} \tag{8}$$

(d) Locate regions of the NMR peak cluster. In LC-NMR data several contiguous regions in frequency appear where the signal-to-noise ratio is significant; we designate these regions "peak clusters". The regions of peak clusters are determined by the morphological score [70], which is calculated as follows.

i. Each column in the data is mean centred

$$x_{mn}^{mc} = x_{mn} - \bar{x}_n \tag{9}$$

where x_{mn}^{mc} is the mean centred intensity and $\bar{x}_n$ is the mean of each column.The Difference matrix is calculated by

$$\Delta x_{mn} = x_{(m+1)n} - x_{mn};$$
$$m = 1, 2, \ldots, M-1; \quad n = 1, 2, \ldots, N \tag{10}$$

ii. The morphological score [70] is calculated by

$$\mathrm{MS} = \frac{\left\| x_n^{mc} \right\|}{\| \Delta x_n \|} \tag{11}$$

iii. The morphological score of each column is then compared with the morphological score of the noise level which is calculated by

$$\mathrm{MS}_{nl} = \sqrt{\frac{\chi(M-1)F_{\mathrm{crit}}(M-1, M-2)}{2(M-2)}} \tag{12}$$

where χ is number of points used in the moving average $\chi = 4$ is used in this application because we employ a four point moving average in the smoothing step of the of LC-NMR data (Section 3.1). F_{crit} is calculated by an F-test at a given critical level (0.99 in our application) and $(M-1)$ and $(M-2)$ degrees of freedom.

(e) Find the rank of each NMR peak cluster by using the morphological score [70].

(f) Calculate the area of each profile in $\boldsymbol{S}$ within each peak cluster region and sort these in descending order.

(g) Least squares is performed on peak clusters by:

$$\bar{\boldsymbol{S}} = (\bar{\boldsymbol{C}}'\bar{\boldsymbol{C}})^{-1}\bar{\boldsymbol{C}}'\bar{\boldsymbol{X}} \tag{13}$$

where $\bar{\boldsymbol{X}}$ is a reduced data matrix and contains the peak cluster region, $\bar{\boldsymbol{C}}$ is a reduced matrix and contains concentration profiles equal to the rank of the peak cluster, containing first those profiles which are sorted in step (f) and $\bar{\boldsymbol{S}}$ is a reduced spectral profiles matrix and covers only those frequencies, which appear in a peak cluster region. If there are k compounds present in a peak cluster, where $k \leq K$ (total number of compounds), having signals at frequency n, the elution profiles are written as a linear combination of only k analytes in a mixture:

$$\boldsymbol{x}_n = \boldsymbol{c}_1 s_{n1} + \boldsymbol{c}_2 s_{n2} + \cdots + \boldsymbol{c}_k s_{nk} \tag{14}$$

In Eq. (14) every elution profile is expressed by a linear combination of k components. One may estimate spectral coefficients using Eq. (13) and the coefficients of remaining analytes are set equal to zero.

3.2.3.4. Modified constrained key variable regression (MCKVR). In the present work, it should be noted the steps described for CKVR can be performed by other methods as well. In the following section a modification of CKVR is described.

(a) A variable selection method such as OPA is applied on the whole data matrix to get key variables and corresponding concentration profiles.

(b) The peak cluster regions are determined by calculating the standard deviation at each frequency [10] given below

$$\sigma_n = \sqrt{\frac{\sum_{m=1}^{M}(x_{mn} - \bar{x}_n)^2}{M-1}} \tag{15}$$

where σ_n is the standard deviation, $\bar{x}_n$ is the average intensity of nth column and M is the total number of the rows in $\boldsymbol{X}$. This process not only selects the peak clusters but also reduces the number of variables. All variables where the signal-to-noise ratio is significant show higher standard deviation as compared to the regions where there is no significant signal. A cutoff value of standard deviation can be used to select peak clusters.

(c) The rank of each peak cluster is determined by the OPA selected concentration profiles [11] and OPA with Durbin-Watson statistics (DW) [11,71,72].

The Durbin-Watson test assumes that the observations and residuals follow a natural order. The residuals are the estimates for errors assumed to be independent. If they are not independent then the DW test checks for a sequential dependence in which each error is correlated with those before and after in the sequence. The DW test is applied on the dissimilarity values (d_m) generated by OPA. The statistic (dw) is defined as

$$\mathrm{dw} = \frac{\sum_{m=1}^{M}(d_m - d_{m-1})^2}{\sum_{m=1}^{M} d_m^2} \tag{16}$$

The dw values are plotted against the number of components, the rank is determined by locating a large increase in the dw.

(d) Pure concentration profiles are obtained by using a variable selection method on spectral peak clusters: we applied OPA in our study.

Steps (c) and (d) can be combined in a single step if only OPA is performed. When OPA is applied to individual peak clusters on the frequency direction, it provides key variables, which are used to locate concentration profiles in the data matrix $\boldsymbol{X}$. The maximum number of key variables selected in OPA is less than or equal to the total data rank. By plotting all concentration profiles together, it reveals not only the data rank but also information about the purity of each variable. As most of compounds have more than one characteristic resonance therefore, there is a good chance of finding a pure variable in LC-NMR data and hence a pure concentration profile ($\boldsymbol{C}_{\mathrm{pure}}$).

(e) Sometimes, two or more compounds resonate at similar frequencies, which in turn produce concentration profiles with two or more well-separated chromatographic peaks. In this situation, one can set zero intensity for every point in the second peak to obtain a pure concentration profile for the first compound and vice versa.

Spectral profiles are constructed by Eq. (8).

Table 1
A comparison of key variables selected by three methods on mixture containing eight compounds

Compound	OPA (ppm)	SIMPLISMA (ppm)	SBM (ppm)
I	7.099	7.099	7.099
II	7.182	7.182	7.184
III	1.253	1.241	1.255
IV	2.452	2.453	2.452
V	1.279	1.279	1.279
VI	1.357	6.940	1.359
VII	6.844	6.840	1.322
VIII	1.340	1.322	1.340

4. Results and discussions

The rank analysis of this data has already been reported in [11]. As EFA and SFA require information about local rank and CKVR and MCKVR requires all pure concentration profiles, therefore, we start by finding pure concentration profiles first, which will be used in EFA and SFA subsequently.

4.1. Determination of pure concentration profiles from the whole data matrix

A pure concentration profile can be defined as one that has positive intensity, is unimodal and has Gaussian peak shape with or without tailing. Pure concentration profiles were obtained by applying OPA ($\boldsymbol{C}_{\mathrm{OPA}}$), SIMPLISMA ($\boldsymbol{C}_{\mathrm{SIMP}}$), and SBM ($\boldsymbol{C}_{\mathrm{SBM}}$) to the whole data matrix. The results of these three variable selection methods are presented in Table 1 and the corresponding concentration profiles are presented in Fig. 3. Of these profiles, I and IV are pure by all methods, while profile VI was pure by SIMPLISMA only and profile VII was pure by OPA and SIMPLISMA. The rest of the profiles (II, III, V, VIII) have bimodality and V has badly defined peak shape in all cases. OPA provided pure concentration profiles for I, IV and VII, SIMPLISMA for I, IV, VI, VII and SBM for I and IV. In total four pure and four mixed profiles were generated by SIMPLISMA. It was observed (Table 1) that SIMPLISMA selected the same variable for component VIII as selected by SBM for compound VII. As SBM provided only two pure variables and the concentration profiles selected by SBM (VII and VIII) and SIMPLISMA (VIII) were not pure, OPA is selected as a method of choice for variable selection in the next sections. Since spectral peak clusters present data in the reduced factor space and there are more chances to obtain unimodal concentration profiles, therefore further exploratory analysis is performed on spectral peak clusters.

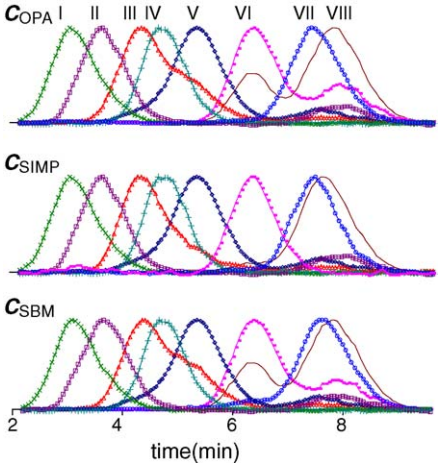

Fig. 3. Concentration profiles generated by three key variable selection methods.

4.1.1. Selection of NMR peak clusters

The determination of peak cluster regions can be automated either by calculating the standard deviation [10] or morphological score (MS) [13] at every frequency and only those variables, which have higher functional value than a preset lower limit, are accepted. Both of these methods produce unsatisfactory results when data contain several spectral peaks. These methods can be used as a first step for peak cluster location and then selected regions are tuned by visual inspection. The results of both methods are presented in Fig. 4. The standard deviation method produced peak cluster regions with good definition therefore this method was utilised for the selection of peak cluster regions. The result of regions selected for peak clusters is also presented in Fig. 4 where they are designated by a bar and a number over peak clusters.

4.1.2. Determination of pure concentration profiles and rank using NMR peak clusters

Rank determination of each peak cluster was performed by the following methods, concentration profiles by morphological score (MS) according to the CKVR method or by OPA and Durbin-Watson statistics (DW) according to the MCKVR method.

OPA was applied to the frequency dimension for eight components on each spectral peak cluster, which produced eight concentration profiles corresponding to each key vari-

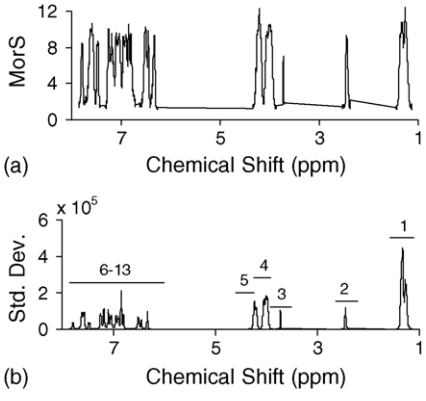

Fig. 4. A comparison of two peak cluster selection methods (a) morphological score and (b) standard deviation. The regions of separate peak clusters are more clearly defined by the standard deviation method as compared to the MS method. Peak cluster regions are indicated by a line and number.

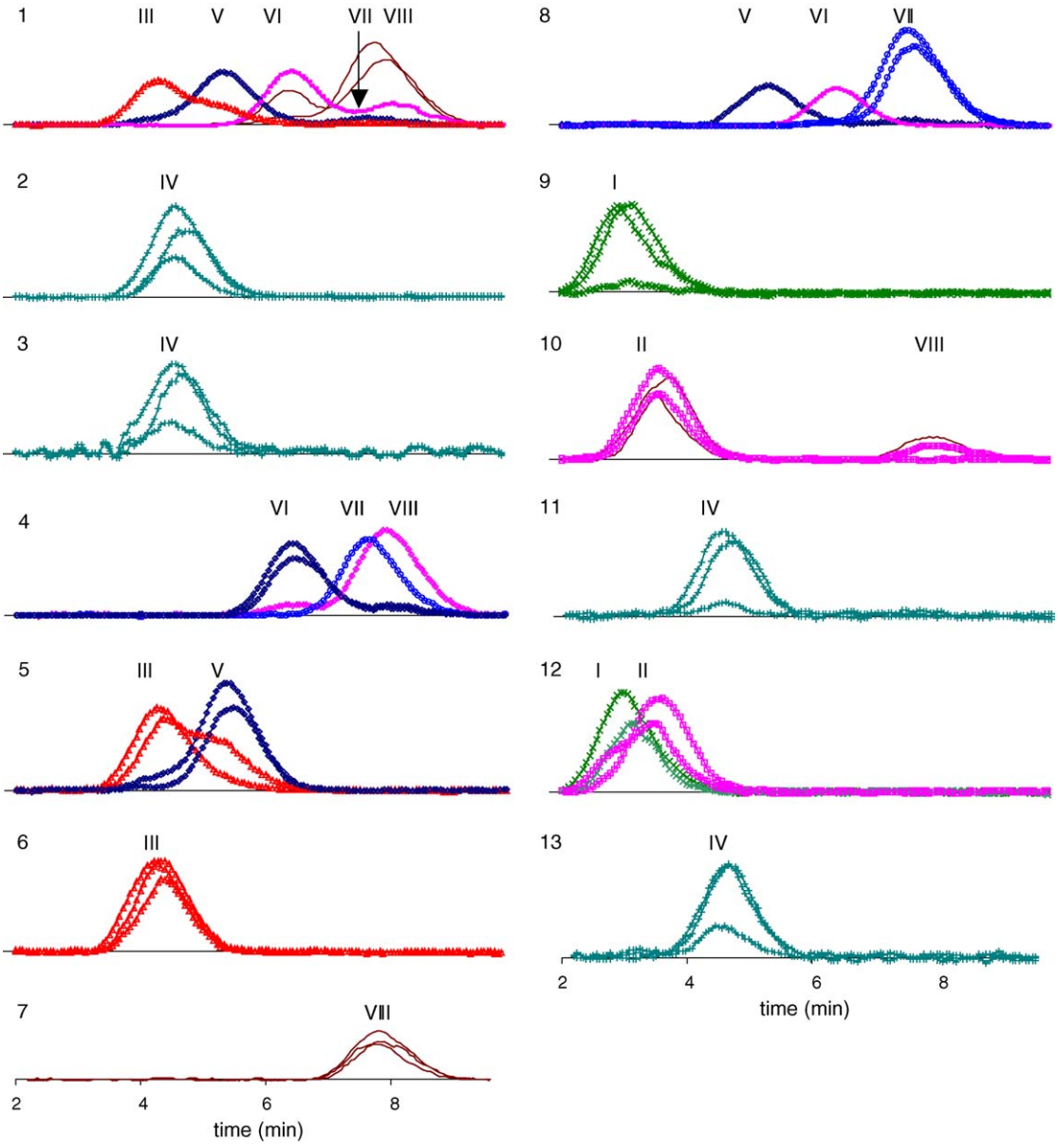

Fig. 5. Peak cluster analysis using OPA on 13 clusters. In cluster 1, compound VII is indicated by an arrow.

able. The results of OPA on the whole dataset are presented in Fig. 3, which provides the elution time of each compound, and on separate spectral clusters in Fig. 5, where fewer than eight concentration profiles are presented for brevity. A comparison of Fig. 3 and Fig. 5 reveals the true rank of each NMR peak cluster. The rank of each spectral peak cluster is determined by counting the number of concentration profiles showing different elution times. These plots also indicate the purity of each concentration profile. Results of OPA concentration profiles, DW statistics and MS are presented in Table 2, where it is clear that all methods perform well except MS. DW statistics found one more compound in cluster 10, which is due to a shift in chromatographic peak position in some of the profiles. A further analysis of Fig. 5 reveals that every compound has at least one pure profile except for compounds V and VI, which have bimodal profiles. This kind of bimodality can be removed simply by setting zeros in the region of the secondary peak. After removing bimodality from components V and VI, all remaining profiles were unimodal and $\boldsymbol{C}_{\text{pure}}$ is used to denote all pure concentration profiles.

4.2. Curve resolution methods

When curve resolution methods were applied, the quality of results was assessed by comparing the peak position and peak shapes with the pure NMR spectra of each compound.

4.2.1. Non-iterative methods

4.2.1.1. EFA. Concentration windows for each compound were located using EPCA. Reconstructed spectral profiles

Table 2
Rank analysis by three different methods: orthogonal projection approach (OPA), OPA-Durbin-Watson, morphological score (MS), the cluster numbers are illustrated Fig. 4(b)

Spectral cluster no.	True rank	OPA	OPA-DW	MS
1	5	5	5	6
2	1	1	1	2
3	1	1	1	1
4	3	3	3	3
5	2	2	2	3
6	1	1	1	2
7	1	1	1	1
8	3	3	3	4
9	1	1	1	2
10	2	2	3	2
11	1	1	1	1
12	2	2	2	2
13	1	1	1	1

using EFA are presented in Fig. 6, artefacts being indicated by a circle around the NMR peak. All profiles contain artefacts except compound VI, while V and VII exhibit only minor artefacts. EFA was applied again after improving the windows limits using $\boldsymbol{C}_{\text{pure}}$ but there was no improvement in the results. Because in LC-NMR chromatographic peaks often severely overlap, it is difficult to obtain good estimates of concentration regions.

4.2.1.2. SFA. In the original paper on SFA [32,33] the authors suggested using FSW-EFA for locating left and right subwindows for each component. We have already reported [11] that in LC-NMR, FSW-EFA does not produce good results because the data have low signal-to-noise ratio. In FSW-EFA a small window is created in the data matrix which moves across the data and eigenvalues are calculated at each step. A plot of eigenvalues against time provides evolving behaviour of different components in the data. The window size cannot be increased beyond a certain number of rows otherwise evolutionary information will be lost. As a small window in time in LC-NMR data contains a lot of noise, therefore, good results are not expected in this application. We utilised $\boldsymbol{C}_{\text{pure}}$ profiles to get information about window limits; the first singular values produced from the left and right subwindows are presented in Table 3 for every component and the predicted spectra are shown in Fig. 7. The spectra of components I, V, VI and VII are pure with the first

Table 3
The first singular values produced by SFA for eight compounds

Compound no.	First singular value
I	0.993
II	0.159
III	0.126
IV	0.649
V	0.798
VI	0.995
VII	0.892
VIII	0.892

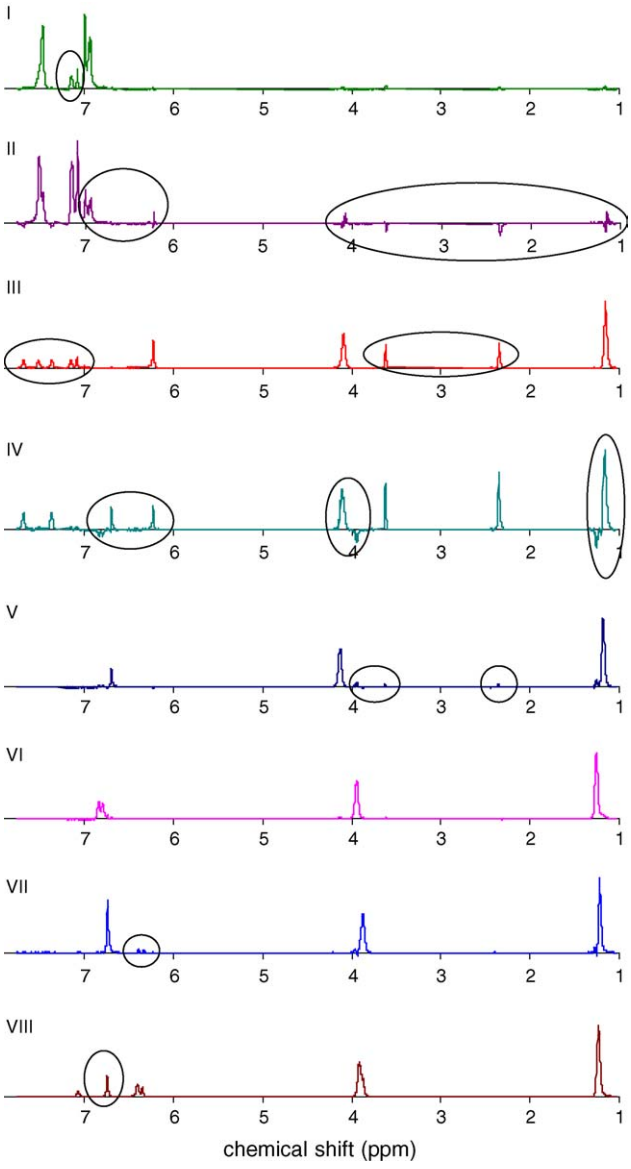

Fig. 6. Spectral profiles obtained by EFA, windows were located by EPCA; artefacts are indicated by circles around the peaks.

eigenvalue close to 1. Although the spectra of components II and III have a low first eigenvalue they appear pure. In the case of NMR data, a minimum in both plots can also be used to get the final spectrum with a risk of mixing impurity peaks from the left and right subwindows. A close look at components II and III reveals that these spectra consist of high noise and distorted peak shapes. Spectra of components IV and VIII also contain artefacts due to the very narrow subwindows. Although SFA is a useful method because it helps in the validation of results by looking at the first singular value, it works best when subwindows are reasonably large, which is not usual in LC-NMR. Nevertheless, five pure spectra were obtained by SFA, which represents an improved performance compared to EFA.

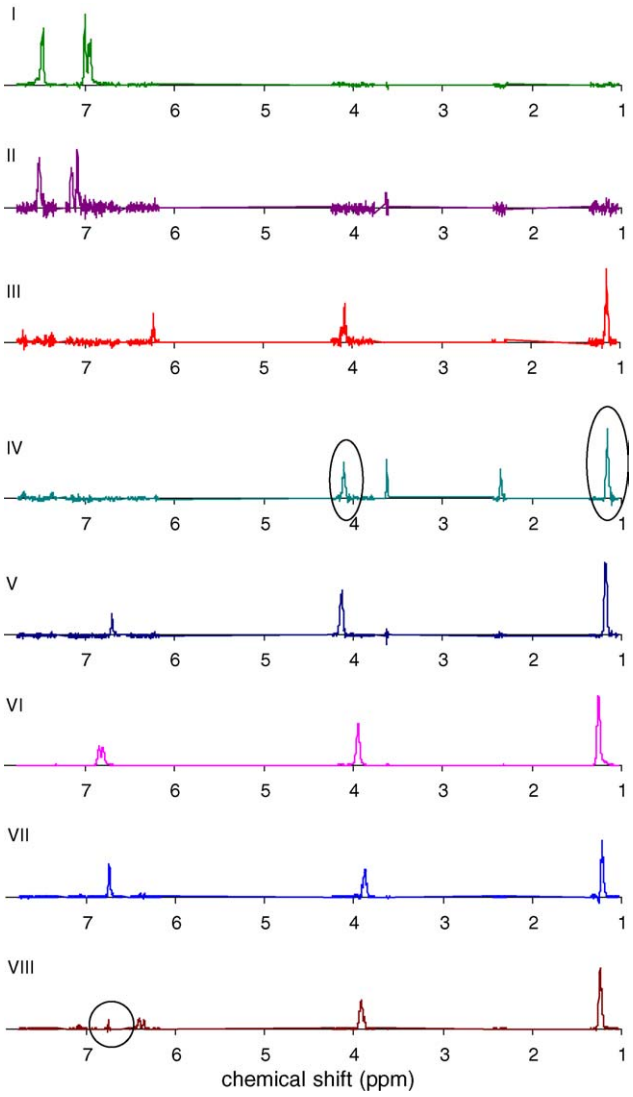

Fig. 7. Spectral profiles obtained by SFA, windows were located by $\boldsymbol{C}_{\text{pure}}$, impurity peaks are indicated by circles around the peaks.

4.2.2. Iterative methods

4.2.2.1. ALS. Initially, ALS was applied using concentration profiles obtained from OPA ($\boldsymbol{C}_{\text{OPA}}$), SIMPLISMA ($\boldsymbol{C}_{\text{SIMP}}$) and SBM ($\boldsymbol{C}_{\text{SBM}}$) using the 'average' unimodality implementation with 1.5 tolerance value. The method converged in 14 iterations for $\boldsymbol{C}_{\text{OPA}}$, in 32 for $\boldsymbol{C}_{\text{SBM}}$ but did not converge for $\boldsymbol{C}_{\text{SIMP}}$ after 50 iterations. Later, the method was applied again using all pure concentration profiles ($\boldsymbol{C}_{\text{pure}}$) as initial estimates. Results were not promising and were similar to the results obtained by EFA and are not illustrated for brevity.

4.2.2.2. ITTFA. ITTFA was performed using Gemperline's software GUIPRO, where the needle search is applied to obtain chromatographic peak location and constructing test vectors. The results of reconstructed spectral profiles were no better than the results calculated by EFA or ALS.

4.2.3. Hybrid methods

4.2.3.1. AUTOWFA. Results obtained by AUTOWFA were similar to the results obtained by other methods like EFA, ALS and ITTFA. There were no improvements; all artefacts were at the same positions.

4.2.3.2. Gentle. Results of Gentle were also similar to the previously described methods, except there were some negative peaks, which was the result of a correction applied to spectral profiles as produced.

4.2.3.3. CKVR. CKVR determines key variables using SBM and then creates unimodal concentration profiles applying Gentle routines. The pure concentration profiles generated by CKVR are similar to those produced by Gentle ($\boldsymbol{C}_{\text{Gentle}}$). The results of reconstructed spectra produced by CKVR are shown in Fig. 8, where spectra of compounds VII and VIII show artefacts. In the spectrum of compound VII there is an artefact at 7.178 ppm indicated by a circle, while in compound

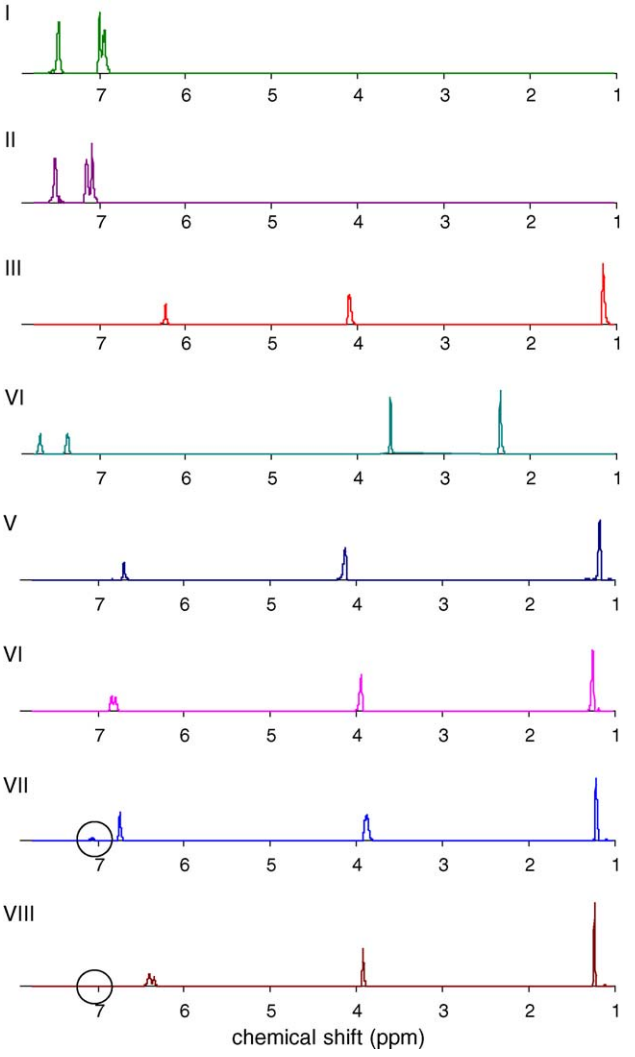

Fig. 8. Results of CKVR with original algorithm, circles indicate the regions of artefacts.

VIII the peak at 7.178 ppm is missing, moreover, the shape of the peak at 1.340 ppm is not correct. These artefacts are produced due to the poor reconstructions of concentration profiles generated by CKVR.

4.2.3.4. MCKVR. The results of MCKVR produced using $\boldsymbol{C}_{\text{pure}}$ are presented in Fig. 9, where all spectra are estimated correctly without any artefacts. All resonances are at the correct chemical shifts and have well defined peak shapes. In Fig. 10, a comparison is presented between the $\boldsymbol{C}_{\text{Gentle}}$ and $\boldsymbol{C}_{\text{pure}}$. Other curve resolution methods (e.g. EFA, ALS, ITTFA and AUTOWFA) also generated similar concentration profiles to those produced by Gentle. It can be observed that almost all profiles in $\boldsymbol{C}_{\text{Gentle}}$ are poor in comparison with the $\boldsymbol{C}_{\text{pure}}$, which caused the incorrect reconstructed spectra. It should be noted that constrained key variable regression

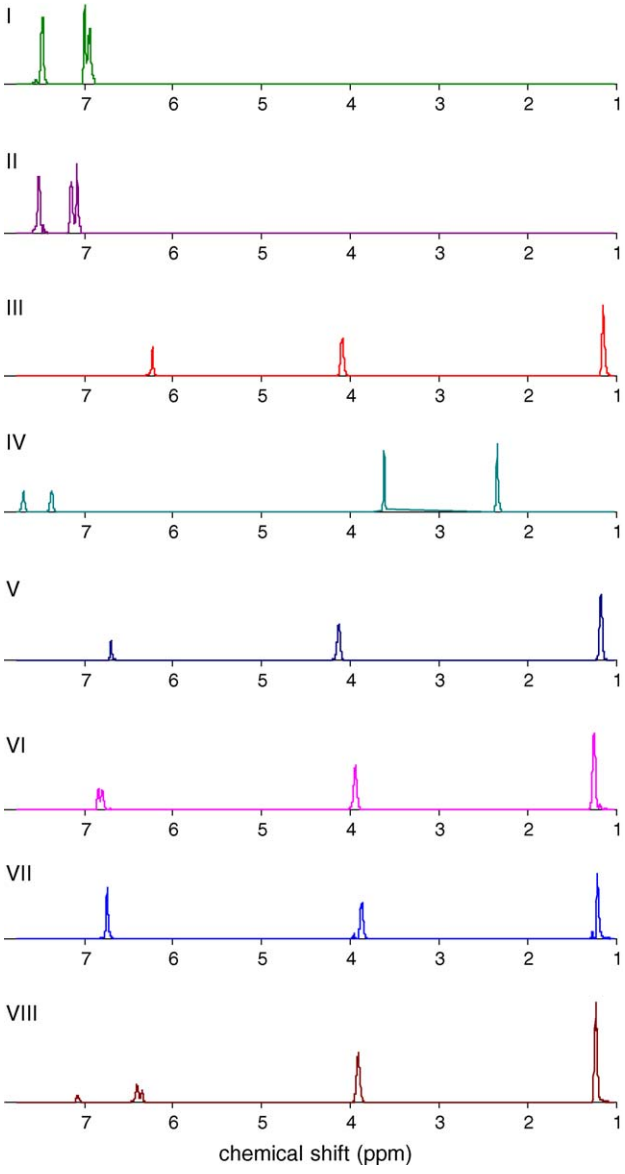

Fig. 9. Spectral profiles obtained by MCKVR; all profiles are pure.

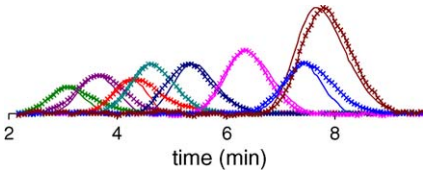

Fig. 10. A comparison of scaled concentration profiles generated by Gentle (×) and determined by OPA (−).

is a good approach only when all concentration profiles are pure with good confidence of the peak shapes.

Due to high noise levels most curve resolution methods converge to a point where incorrect concentration profiles are generated. The only way, which appears to work for LC-NMR data, is MCKVR, where only concentration profiles, which are unimodal and have well defined peak shape, are used in regression. Obviously, the use of the non-negativity constraint is not necessary but we use it to obtain positive spectral profiles.

5. Conclusions

The performance of CKVR for LC-NMR data requires many steps among which are peak cluster location and rank analysis of each peak cluster. In this paper, these procedures were performed in semi-automatic mode and results were confirmed by visual inspection at each stage. As selective resonances are common in NMR, most of the pure concentration profiles can be located easily in peak clusters. Pure concentration profiles can be located by searching each peak cluster using a suitable variable selection method. Because SIMPLISMA selects profiles containing high noise, the use of OPA or SBM is recommended. The few impure profiles can be modified by removing bimodality.

MCKVR involves many modifications, which includes rank analysis and creation of pure concentration profiles by using OPA. The new modifications are simple and more reliable than the original CKVR. Since LC-NMR data has low signal-to-noise ratio and highly overlapping peaks, the routine of Gentle does not produce correct reconstruction of pure concentration profiles, which are implemented in CKVR. MCKVR and CKVR both require estimates of the pure concentration profiles, which is the major limitation of these techniques, which depend on a good algorithm for determining elution profiles. However, it is not necessary to have any chromatographic information on the pure components in advance, and properties such as unimodality and non-negativity of chromatographic profiles can be used to obtain good guesses in cases where selective resonances are not available. In this paper, a successful application of MCKVR has been demonstrated for eight-compound mixture, which can be extended to more complex mixtures. For more complex mixtures the data can be sliced in different concentration windows, which will reduce the data complexity and provide more chances of finding pure spectral variables as compared to the whole data.

6. Nomenclature

$\boldsymbol{C}$	matrix of concentration profiles
d	dissimilarity value in OPA
dw	Durbin-Watson statistics
$\boldsymbol{E}$	residual error in a bilinear modal
K	number of compounds in X
M	number of rows in X
N	number of columns in X
$\boldsymbol{P}$	loadings matrix
$\boldsymbol{R}$	rotation matrix
$\boldsymbol{S}$	matrix of spectral profiles
$\boldsymbol{T}$	scores matrix
$\overline{x}_n$	mean intensity of the nth row
x^{mc}	mean centred intensity
Δx	difference between two intensity values in consecutive rows
$\boldsymbol{X}$	data matrix

Greek letters

χ	number of points used in moving average
σ	standard deviation of column or row indicated by subscript

Acknowledgements

M.W. wishes to thank the Ministry of Science and Technology, Government of Pakistan for providing Ph.D. grant and the support of PAEC, Pakistan. Both authors are thankful to R. Tauler and P. Gemperline for their software. We also thank Drs. C.Y. Airiau and Dr. M. Murray for their help with the experimental work.

References

[1] M.V.S. Elipe, Anal. Chim. Acta 497 (2003) 1.
[2] K. Albert (Ed.), On-line HPLC-NMR and Related Techniques, Wiley, Chichester, 2002.
[3] J.-L. Wolfender, K. Ndjoko, K. Hostettmann, J. Chromatogr. A 1000 (2003) 437.
[4] F.C. Stintzing, J. Conrad, I. Klaiber, U. Beifuss, R. Carle, Phytochemistry 65 (2004) 415.
[5] K. Iwasa, A. Kuribayashi, M. Sugiura, M. Moriyasu, D.U. Lee, W. Wiegrebe, Phytochemistry 64 (2003) 1229.
[6] J.-L. Wolfender, S. Rodriguez, K. Hostettmann, J. Chromatogr. A 794 (1998) 299.
[7] A.M. Gil, I.F. Duarte, M. Godejohann, U. Braumann, M. Maraschin, M. Spraul, Anal. Chim. Acta 488 (2003) 35.
[8] E. Bezemer, S. Rutan, Anal. Chim. Acta 459 (2002) 277.
[9] M. Wasim, M.S. Hassan, R.G. Brereton, Analyst 128 (2003) 1082.
[10] C.Y. Airiau, H. Shen, R.G. Brereton, Anal. Chim. Acta 447 (2001) 199.
[11] M. Wasim, R.G. Brereton, Chemom. Intell. Lab. Syst. 72 (2004) 133.
[12] H. Shen, C.Y. Airiau, R.G. Brereton, Chemom. Intell. Lab. Syst. 62 (2002) 61.
[13] H. Shen, C.Y. Airiau, R.G. Brereton, J. Chemom. 16 (2002) 165.
[14] H. Shen, C.Y. Airiau, R.G. Brereton, J. Chemom. 16 (2002) 469.
[15] J.-H. Jiang, Y. Liang, Y. Ozaki, Chemom. Intell. Lab. Syst. 71 (2004) 1.
[16] E.R. Malinowski, Factor Analysis in Chemistry, third ed., Wiley, New York, 2002.
[17] S.D. Frans, M.L. McConnel, J.M. Harris, Anal. Chem. 57 (1985) 1552.
[18] F.J. Knorr, J.M. Harris, Anal. Chem. 53 (1981) 272.
[19] A.K. Smilde, H.C.J. Hoefsloot, H.A.L. Kiers, S. Bijlsma, H.F.M. Boelens, J. Chemom. 15 (2001) 405.
[20] P.J. Gemperline, J. Chem. Inf. Comput. Sci. 24 (1984) 206.
[21] B.G.M. Vandeginste, W. Derks, G. Kateman, Anal. Chim. Acta 173 (1985) 253.
[22] E.J. Karjalainen, Chemom. Intell. Lab. Syst. 7 (1989) 31.
[23] R. Tauler, E. Casassas, Chemom. Intell. Lab. Syst. 14 (1992) 305.
[24] P. Paatero, Chemom. Intell. Lab. Syst. 37 (1997) 23.
[25] J.-H. Jiang, Y.-Z. Liang, Y. Ozaki, Chemom. Intell. Lab. Syst. 65 (2003) 51.
[26] H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbuhler, Talanta 32 (1985) 1133.
[27] M. Maeder, Anal. Chem. 59 (1987) 527.
[28] E.R. Malinowski, J. Chemom. 6 (1992) 29.
[29] W. Den, E.R. Malinowski, J. Chemom. 7 (1993) 89.
[30] O.M. Kvalheim, Y.-Z. Liang, Anal. Chem. 64 (1992) 936.
[31] Y.-Z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle, F. Erni, Anal. Chem. 64 (1992) 946.
[32] R. Manne, H. Shen, Y. Liang, Chemom. Intell. Lab. Syst. 45 (1999) 171.
[33] H. Shen, R. Manne, Q. Xu, D. Chen, Y. Liang, Chemom. Intell. Lab. Syst. 45 (1999) 323.
[34] J.-H. Jiang, S. Sasic, R.-Q. Yu, Y. Ozaki, J. Chemom. 17 (2003) 186.
[35] E.R. Malinowski, J. Chemom. 10 (1996) 273.
[36] R. Manne, B.-V. Grande, Chemom. Intell. Lab. Syst. 50 (2000) 35.
[37] G.A. Morris, R. Freeman, J. Magn. Reson. 29 (1978) 433.
[38] http://www.ub.es/gesq/mcr/ntheory.htm.
[39] http://personal.ecu.edu/gemperlinep/.
[40] P.J. Gemperline, E. Cash, Anal. Chem. 75 (2003) 4236.
[41] J.C. Hoch, A.S. Stern, NMR Data Processing, Wiley, New York, 1996.
[42] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S.D. Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part B, Elsevier, Amsterdam, 2003.
[43] H.R. Keller, D.L. Massart, Anal. Chim. Acta 246 (1991) 379.
[44] S. Wold, K. Esbensen, P. Geladi, Chemom. Intell. Lab. Syst. 2 (1987) 37.
[45] R.G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, Wiley, Chichester, 2003.
[46] W. Windig, J. Guilment, Anal. Chem. 63 (1991) 1425.
[47] E.R. Malinowski, Anal. Chim. Acta 134 (1982) 129.
[48] K.J. Schostack, E.R. Malinowski, Chemom. Intell. Lab. Syst. 6 (1989) 21.
[49] F.C. Sanchez, M.S. Khots, D.L. Massart, J.O. De Beer, Anal. Chim. Acta 285 (1994) 181.
[50] F.C. Sanchez, M.S. Khots, D.L. Massart, Anal. Chim. Acta 290 (1994) 249.
[51] F.C. Sanchez, J. Toft, B. van den Bogaert, D.L. Massart, Anal. Chem. 68 (1996) 79.
[52] F.C. Sanchez, B.G.M. Vandeginste, T.M. Hancewicz, D.L. Massart, Anal. Chem. 69 (1997) 1477.
[53] R. Tauler, D. Barcelo, Trends Anal. Chem. 12 (1993) 319.
[54] F.C. Sanchez, S.C. Rutan, M.D. Gil Garcia, D.L. Massart, Chemom. Intell. Lab. Syst. 36 (1997) 153.
[55] B.-V. Grande, R. Manne, Chemom. Intell. Lab. Syst. 50 (2000) 19.
[56] G. Strang, Linear Algebra and its Applications, third ed., Harcourt Brace Jovanovich, Orlando, 1998.
[57] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[58] S. Navea, A. de Juan, R. Tauler, Anal. Chem. 74 (2002) 6031.
[59] E. Teixido, L. Olivella, M. Figueras, A. Ginebreda, R. Tauler, J. Environ. Anal. Chem. 81 (2001) 295.
[60] J. Saurina, S.H. Cassou, A.I. Ridorsa, R. Tauler, Chemom. Intell. Lab. Syst. 50 (2000) 263.
[61] A.K. Smilde, R. Tauler, H.M. Henshaw, L.W. Burgess, B.R. Kowalski, Anal. Chem. 66 (1994) 3345.
[62] B.H. Cruz, J.M. Diaz, C. Arino, M. Estebam, R. Tauler, Analyst 127 (2002) 401.
[63] A. de Juan, B. Van den Bogaert, F.C. Sanchez, D.L. Massart, Chemom. Intell. Lab. Syst. 33 (1996) 133.
[64] H.H. Harman, Modern Factor Analysis, The University of Chicago Press, Chicago, 1968, pp. 304–313.
[65] J. Toft, O.M. Kvalheim, Chemom. Intell. Lab. Syst. 25 (1994) 61.
[66] M. Esteban, C. Arino, J.M. Diaz-Cruz, M.S. Diaz-Cruz, R. Tauler, Trends Anal. Chem. 19 (2000) 49.
[67] P.J. Gemperline, J.C. Hamilton, J. Chemom. 3 (1989) 455.
[68] H. Shen, B. Grung, O.M. Kvalheim, I. Eide, Anal. Chim. Acta 446 (2001) 313.
[69] S.A. Mjos, Anal. Chim. Acta 488 (2003) 231.
[70] H. Shen, L. Stordrange, R. Manne, O.M. Kvalheim, Y.-Z. Liang, Chemom. Intell. Lab. Syst. 50 (2000) 37.
[71] N.R. Draper, H. Smith, Applied Regression Analysis, third ed., John Wiley & Sons, New York, 1998.
[72] S. Gourvenec, D.L. Massart, D.N. Rutledge, Chemom. Intell. Lab. Syst. 61 (2002) 51.