ELSEVIER

Analytica Chimica Acta 446 (2001) 495–502

ANALYTICA CHIMICA ACTA

www.elsevier.com/locate/aca

Multivariate curve resolution applied to MS/MS data obtained from isomeric mixtures

C.G. Zampronio, L.A.B. Moraes, M.N. Eberlin, R.J. Poppi*

Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas, São Paulo, Brazil

Received 7 November 2000; accepted 25 June 2001

Abstract

Multivariate curve resolution (MCR) was applied to MS/MS data for modeling isomeric mixtures. A pentaquadrupole mass spectrometer consisting of three mass analyzing and two reaction quadrupoles was used to obtain a second-order data from mixtures of 2-, 3- and 4-ethylpyridine. MCR allowed recovering of the mass spectral profiles of the three isomers as well as the concentration distribution profiles. The results indicate that MCR is a suitable tool for MS/MS data analyses to distinguish and quantify the positional isomers in mixtures.

Keywords: Multivariate curve resolution; MS/MS; Isomeric distinction

1. Introduction

The distinction and quantification of isomeric compounds are very important in many fields of chemistry, and the analyses of positional isomers are an issue that has been studied by pentaquadrupole [1,2]. It is desirable that the mass spectra of the isomers contain unique structural diagnostic fragment ions. If not, then the collision-induced dissociation (CID) [3–6] can be used as an alternative to this approach because the daughter fragments normally retain the positional information of the isomeric ions from which they are derived. This procedure generates an MS/MS spectrum and a second-order dataset.

However, even using MS/MS techniques, the identification or quantification of the components in an isomeric mixture many times is not possible, because the fragments (parent fragments and daughter fragments) in the spectrum are similar or they have similar abundance. In these cases, the application of multivariate methods for these second-order data can be a hopeful alternative for isomeric analyses.

The application of multivariate methods to second-order data has been discussed in many papers, such as LC-DAD [7], UV-DAD [8,9], UV spectra recorded at different pH [10], fluorescence spectra recorded at different pH [11]. Among these chemometric methods, the rank annihilation factor analysis (RAFA) [12] and generalized rank annihilation (GRAM) [13] have been used in several applications. Both methods work well under certain conditions of the second-order data: the pseudo-rank of a pure analyte response is 1 and this pure analyte response has the same form in the standard as in the unknown mixture.

Pseudo-rank is defined as the rank of a measurement if it does not contain an experimental error. Since in all measurements noise is present, the pseudo-rank has to be estimated and it can be used to check the condition for applying RAFA or GRAM.

* Corresponding author. Tel.: +55-19-788-3126; fax: +55-19-788-3023.
E-mail address: ronei@iqm.unicamp.br (R.J. Poppi).

0003-2670/01/$ – see front matter
PII: S0003-2670(01)01272-7

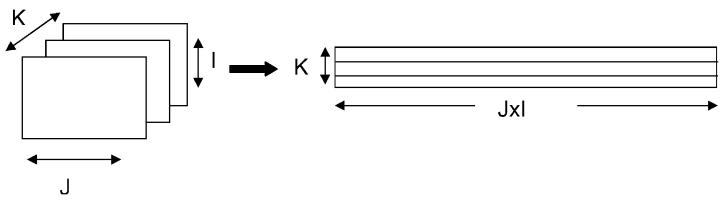

Fig. 1. Unfold of the $\boldsymbol{K}$ data matrices $\boldsymbol{J} \times \boldsymbol{I}$.

The MS/MS pure analyte response does not present rank equal to 1 and the application of second-order chemometric methods in situations where the pseudo-rank 1 restriction does not hold is considerably more complicated. Non-bilinear rank annihilation (NBRA) [14] is a direct generalization of GRAM for situations with pure analyte response and pseudo-rank higher than 1. The quantification is possible, but the resolution of individual profiles is not possible.

Several methods have been described for complicated second-order data [9], which are the cases where the rank 1 property and perhaps even rank linear additional do not hold. Among these methods, there is the multivariate curve resolution (MCR) with restrictions [15,16] which can be used in these situations.

The MCR is based on the analysis of unfolded augmented column-wise data matrices. In Fig. 1 is one example of unfold three-way data matrix with $\boldsymbol{K}$ samples or slabs, $\boldsymbol{I}$ rows and $\boldsymbol{J}$ columns.

The data matrices can be written as in Eq. (1), where $\boldsymbol{D}$ is the data matrix of mass spectra, and $\boldsymbol{C}$ and $\boldsymbol{S}$ are matrices, which usually are related with the concentration and spectra response profiles of the pure chemical species into the mixture.

$$\boldsymbol{D} = \boldsymbol{C}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E} \tag{1}$$

$\boldsymbol{E}$ is the matrix of the residuals, not explained by the pure chemical species in $\boldsymbol{C}$ and $\boldsymbol{S}$, which hopefully is close to the experimental error. The alternating least squares (ALS) is optimized for estimation of $\boldsymbol{C}$ and $\boldsymbol{S}$.

The matrix $\boldsymbol{C}$ has a number of rows equal to the number of samples experimentally measured and a number of columns equal to the number of proposed chemical contributions, describing how their concentrations change. On the other hand, $\boldsymbol{S}$ in this work is a matrix with a number of rows equal to the number of proposed chemical contributions and with a number of columns equal to the number of points related to the *m*/*z* value, described how pure individual mass spectra.

When an initial estimation of the individual mass spectra is initially available, the best least squares solution of the concentration profiles is estimated from:

$$\boldsymbol{C} = \boldsymbol{D}\boldsymbol{S}^{+} \tag{2}$$

where $\boldsymbol{S}^{+}$ is the pseudo-inverse of $\boldsymbol{S}$ matrix. If on the contrary, an initial estimation of the concentration profiles is available, the best least squares estimation of the mass spectra contributions is estimated from:

$$\boldsymbol{S} = \boldsymbol{C}^{+}\boldsymbol{D} \tag{3}$$

where now $\boldsymbol{C}^{+}$ is the pseudo-inverse of $\boldsymbol{C}$ matrix.

The least squares solutions obtained in this way are pure mathematical solutions. However, they probably will not be optimal from a chemical point of view. Therefore, an ALS optimization procedure is started, resolved by the iteration of the two equations previously given and some constraints are added to the model (non-negativity of concentrations and signals; unimodality of the concentrations and the signals, etc.).

In the case of MS/MS data as well as in the other examples described before, the final goal is the estimation, from a mixture, of the matrix containing the concentration profiles of the constituents and estimation of the pure spectra of those constituents. MCR can be used as a quantitative tool, where concentrations of news samples are estimated from the recovered individual mass spectra of each constituent.

2. Experimental

The MS/MS experiment was performed with an Extrel pentaquadrupole mass spectrometer [17], shown in

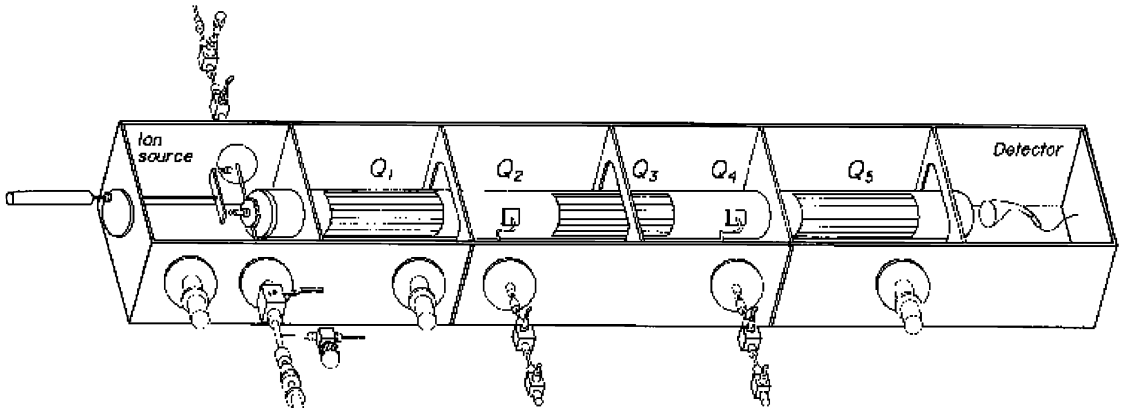

Fig. 2. Pentaquadrupole mass spectrometer and its arrangement of the three mass-analyzers (Q_1, Q_3 and Q_5) and two "rf-only" quadrupole reaction chambers (Q_2 and Q_4).

Fig. 2. The Q_1, Q_3 and Q_5 consist of mass-analyzing quadrupoles in which ion-mass selection and analysis are performed, and the two reaction quadrupoles Q_2 and Q_4 are used to perform either low-energy ion-molecule reactions or CID. For the MS/MS experiments applied in this study, parent ions were generated by 70 eV electron ionization (EI) and these ions were selected by Q_3. After dissociation by collision with argon in Q_4 of the selected parent ions, the daughter ion fragments were generated. Finally, Q_5 was scanned to record the sequential product spectrum.

An experimental design with 45 samples (Fig. 3) was performed from mixtures of 2-, 3- and 4-ethylpyridine with a molar ratio to range from 0.0 to 1.0. Equal quantity of deuterated pyridine (10 μl) was also added into all samples to be used as an internal standard and the total volume of each mixture was 110 μl.

The dataset was split in two groups: one group with nine samples to built the MCR model and to recover the pure spectrum profiles of the isomers. Another group with nine samples was used as validation set to verify the possibility of quantitative analysis of the isomers. The triangles presented in Fig. 3 were used in the modeling and the circles were used as validation set. Squares in this figure indicate samples not used in the modeling, because the results were not suitable.

Fig. 4 shows a typical MS/MS and the unfolded spectrum obtained from the measurement of a

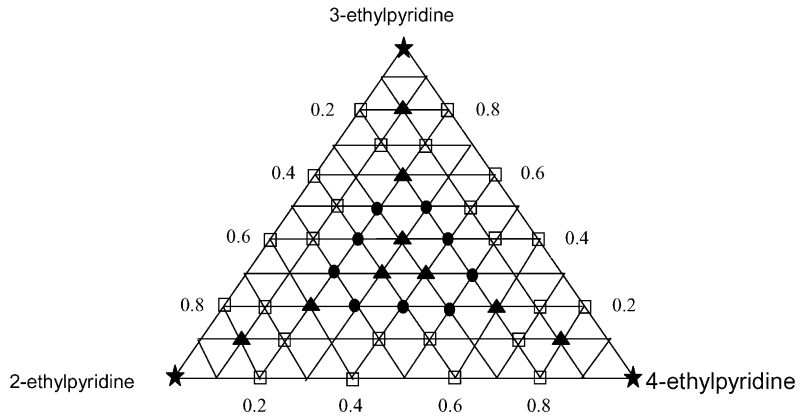

Fig. 3. Experimental design of isomeric mixtures: modeling set (▲); validation set (●); samples not used (□); and pure compounds (∗).

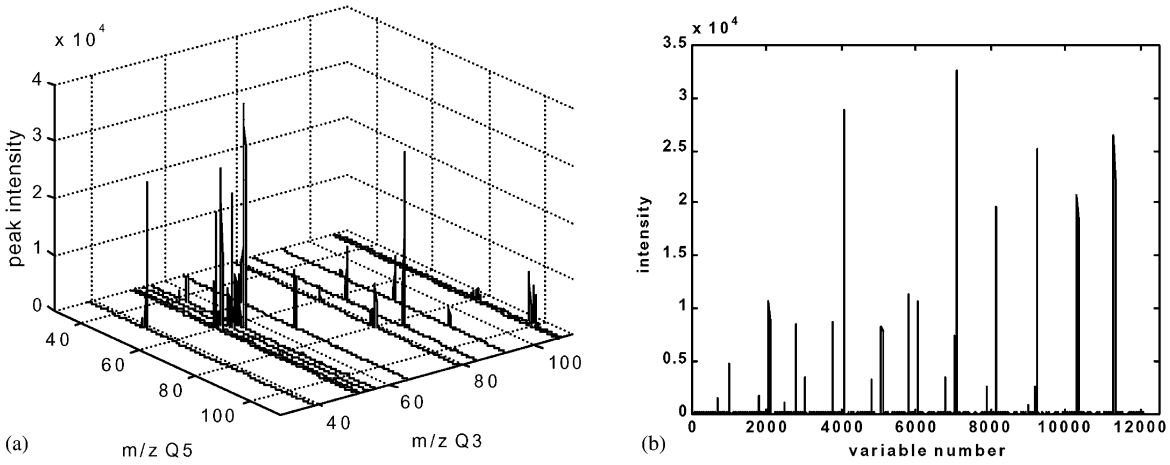

Fig. 4. (a) MS/MS spectrum from of 2-, 3- and 4-ethylpyridines (parent ion fragments *m*/*z* 39, 51, 65, 79, 92, 106 and 107), with molar fraction 0.3/0.3/0.4, in the presence of deuterated pyridine (parent ion fragments *m*/*z* 52, 54, 56 and 84); (b) unfolding spectrum, where in the abscissa is plotted the number of points.

mixture with a molar ratio of 0.3/0.3/0.4 of 2-, 3- and 4-ethylpyridine, respectively (parent ion fragments in Q_3 *m*/*z*: 39, 51, 65, 79, 92, 106 and 107). Also it is present the parent ion fragments in Q_3 from the deutered pyridine (*m*/*z*: 52, 53, 54, 56 and 84). The *m*/*z*: 30–110 range in the original spectra (MS/MS) is equivalent to 1101 recorded variables. As there are 11 parent ion fragments that generate 11 spectra with 1101 points, the total number of variables in the unfold spectrum is 12 111.

Triplicate spectra were obtained for each one of the 45 isomeric mixtures prepared and it was selected the replica with more quantity of the parent ion fragments detected with their respective daughter ion fragments. Besides, it was selected to be use in MCR modeling only the parent ion fragments common for all isomeric mixtures that are *m*/*z* 39, 51, 106 and 107.

For the data treatment, it was used MCR with restrictions, written in MATLAB [18]. The input conditions were the spectra from the mixture of the isomers. The restriction used was the non-negativity in spectral profiles and concentrations, and the convergence criterion in ALS algorithm was 1×10^{-3}.

3. Results and discussion

The replicated samples presented variation of about 10% of peak intensity, hence it was first necessary to find a pre-processing approach to minimize this variation. The first approach tested was baseline correction, where an average baseline was obtained from each sample.

The most intense peak usually normalizes the mass spectra. However, this procedure cannot be the best choice, because the mass spectrometry is a technique that shows heteroscedastic noise, and the way of normalization has a profound impact on the correlation structure. Kvalheim et al. [19] showed that the heteroscedasticity in data must be removed prior to normalization.

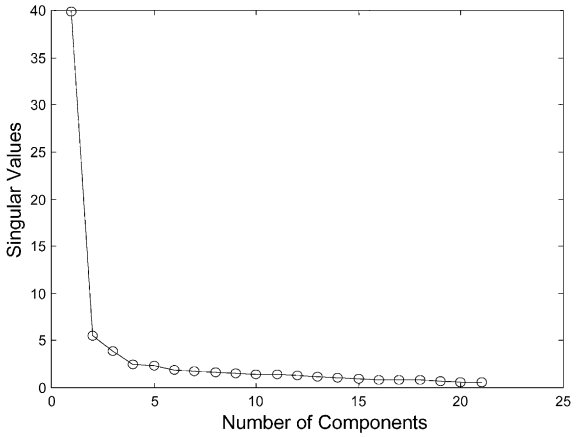

Fig. 5. Singular value decomposition of the argument data.

The homoscedastic noise occurs when the absolute noise is constant with increasing peak intensity and the heteroscedastic noise occurs when the relative standard deviation of the noise decreases with increasing peak intensity or the absolute noise increases with increasing peak intensity. For cases where the standard deviation of the noise is proportional to the peak intensity such as for MS/MS, the log transform followed by row centering is shown to be an effective procedure against the spurious correlation induced by constant-sum normalization [19]. In this work, Eq. (4) was used to remove the heteroscedastic noise of isomeric mixture spectra, where z is the intensity value and M is the total number of variables in the analytical profile.

$$x_i^{\mathrm{T}} = \ln z_i^{\mathrm{T}} - \frac{1}{M\sum_k \ln z_{ik}} \tag{4}$$

After using Eq. (4) for minimization of heteroscedasticity noise, the normalization by the internal standard was applied.

Performing these procedures described above, low signals become more intense and this result shows that the log transform damages the linear correlation between signal and concentration. However, it was verified that is important to remove the heterocedasticity noise to estimate the pure spectra.

Mixtures including a very low concentration of one isomer (or even in the absence of this) were used in the initial models using the MCR method to find the

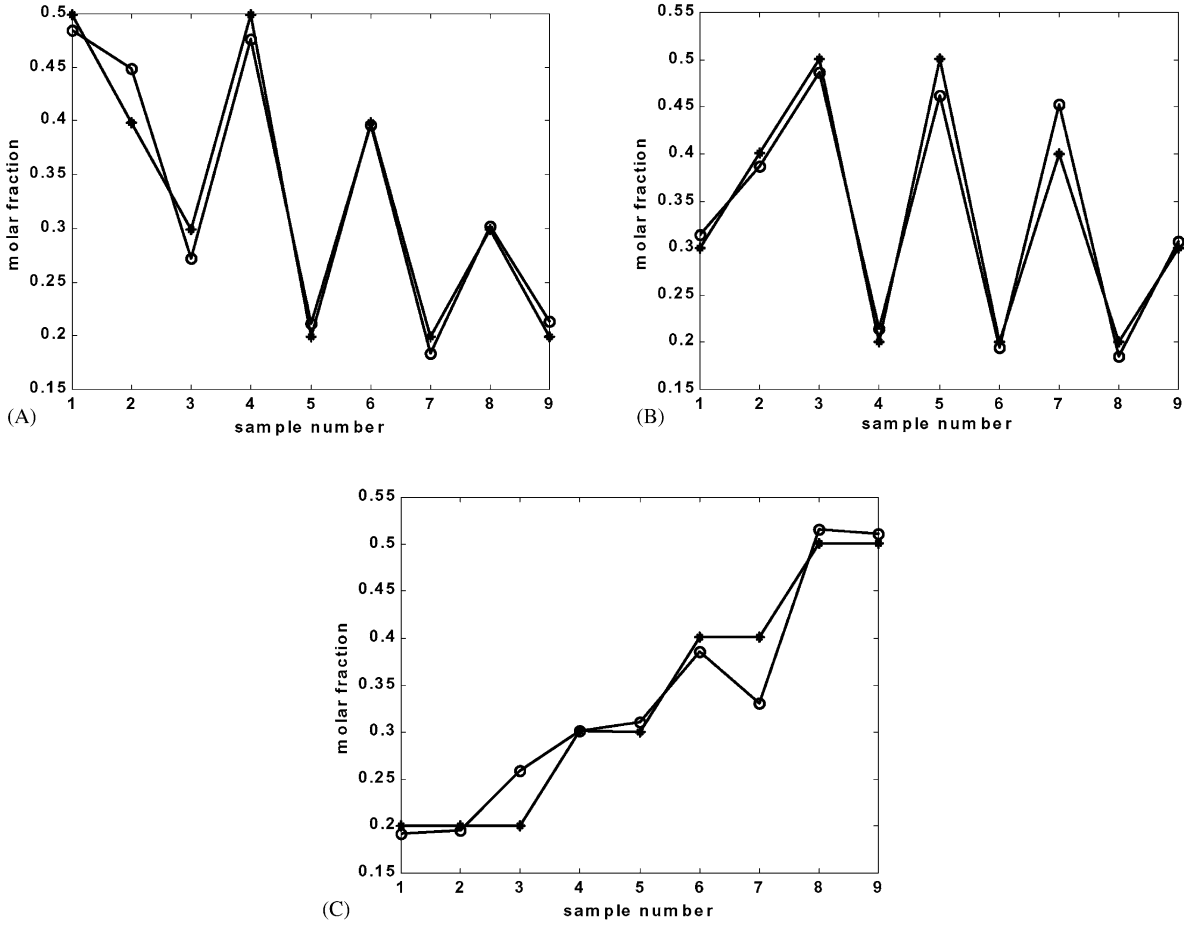

Fig. 6. Profile concentration for three isomers studied: (A) 2-ethylpyridine; (B) 3-ethylpyridine; (C) 4-ethylpyridine. True value (∗); recovered value (○).

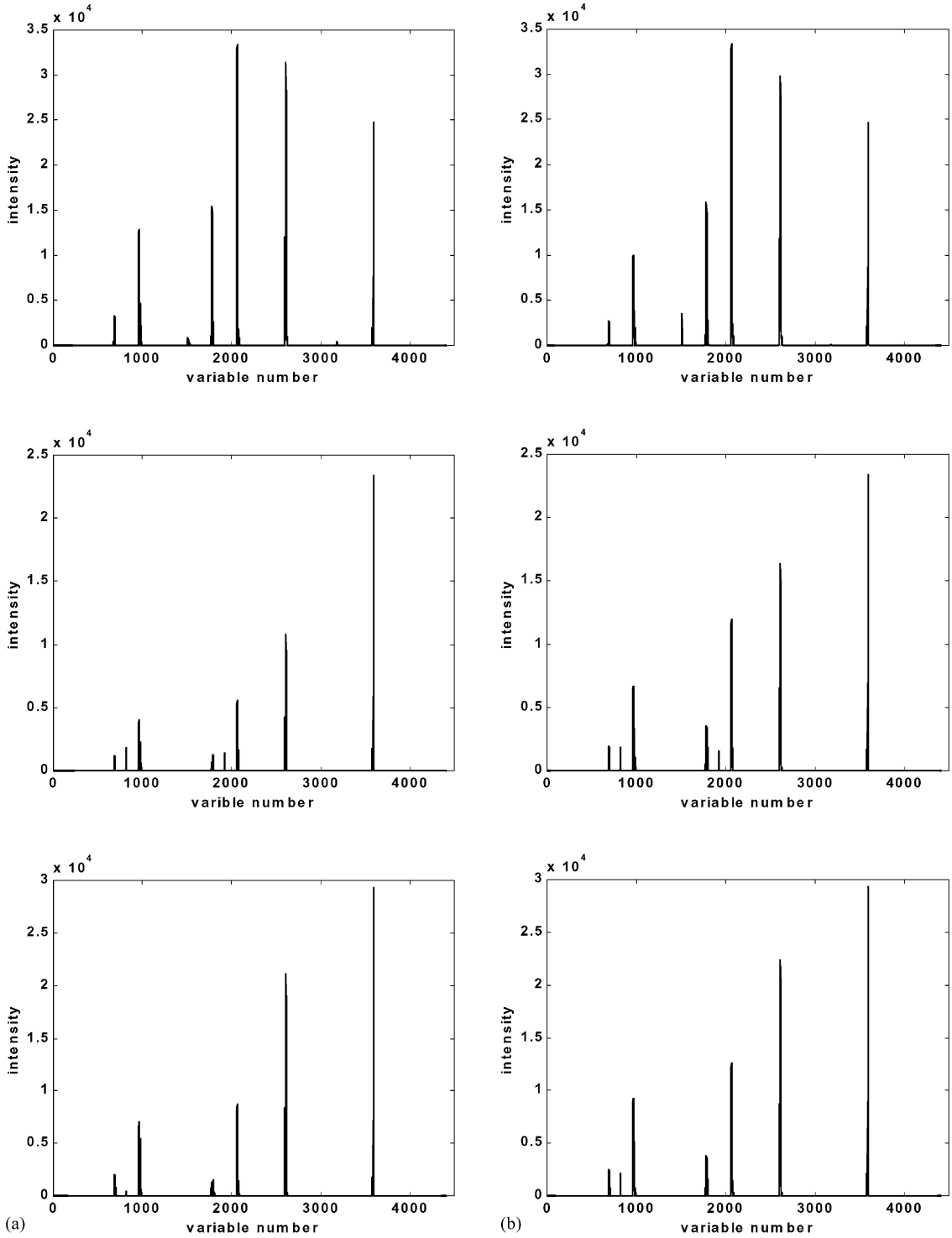

Fig. 7. (a) Experimental and (b) recovered unfold spectra for the isomers. (1) 2-Ethylpyridine; (2) 3-ethylpyridine; (3) 4-ethylpyridine.

concentration and spectral profiles. However, it was verified that the samples with low values of molar ratio presented high errors to recover the concentration profile. Therefore, only samples with intermediary molar ratio value were used in the modeling.

It was also possible to improve the results by taking out the parent ion fragments related to the deuterated pyridine (*m*/*z* 84, 56, 54, 52) because the internal standard became an interference component. Other parent ion fragments related to 2-, 3- and 4-ethylpyridines, as *m*/*z* 92, 79 and 65, were also removed because they were not common in all samples, and this fact introduce an additional augment into the matrix rank for some samples. The parent ion fragments *m*/*z* 92 and 65 are less intense for 2-ethylpyridine and the parent ion fragment *m*/*z* 79 is less intense for 3-ethylpyridine. Thus, sometimes these ions are not detected in Q_3 owing to pressure variations in the equipment and consequently the MS/MS showed different size for this dimension. Therefore, it was used in the modeling only the parent ion fragments presented in all samples.

The second-order dataset was unfolded where the parent ion fragments selected in Q_3 were *m*/*z* 39, 51, 106 and 107. These parent ion fragments with their daughter ion fragments were put side by side as shown in Fig. 1. In this work, nine samples were used with four parent ion fragments to build the MCR model. Each parent ion fragment was described for a spectrum with 1101 points that the sum of all intensity of the peaks in the spectrum (parent ion fragment and daughter ion fragments) is equivalent to intensity of the parent ion fragment in Q_3.

The rank value for the matrix of data unfolding for the pure analytes and mixtures was found before applying the MCR method. It was verified that the rank, calculated based on SVD value [20], is higher than one for all cases studied. For the pure analytes and mixtures, the rank obtained was 4. This value is owing to the fact that each parent ion fragment selected in Q_3 produces a new spectrum with the daughter ion fragments, which contributes to the total rank.

In the augmented data matrix with nine samples ($\boldsymbol{K} = 9$ and $\boldsymbol{J} \times \boldsymbol{I} = 4404$), the value of the rank was 3, as shown in Fig. 5 in the plot of number of components against singular value. This value can be related to the total number of species presented, because now there is a set of samples with different concentrations and isomer information.

Based on the results from the rank analysis and supposing a direct linear proportion between the fragment area and the concentration in the augmented MS data matrix, it is possible to explain the response as a product of the concentration profiles and the unit pure spectra. The MCR was applied to recover these three isomers information about the spectral and concentration profiles.

Fig. 6 presents the true and obtained concentrations for the three isomers in the mixtures into the nine samples of the validation set and it is possible to see that the values are closer. The root mean square error of prediction (RMSEP) of 0.0074, 0.0081 and 0.0106 molar fractions, were obtained for the 2-, 3- and 4-ethylpyridine isomers, respectively. In Fig. 7 is presented the experimental spectra of each isomer and the recovered spectra for each pure compound obtained by MCR method. A good agreement is observed between the profiles of experimental and the recovered spectra, which the correlation coefficient values between the true mass spectrum and the calculated are 0.997, 0.962, and 0.991 for the 2-, 3- and 4-ethylpyridine, respectively.

4. Conclusions

This paper demonstrated that the multivariate curve resolution is a suitable tool for extracting qualitative and quantitative information from complex second-order data, such as that of MS/MS. MCR recovered the pure mass spectral profiles of the three isomers with good accuracy, and low errors were obtained in determination of isomer concentrations in mixtures.

The combination of MS/MS with MCR is an alternative for the analysis of positional isomeric mixtures which are difficult to be accomplished by more traditional analytical techniques.

Acknowledgements

This work has been support by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), process number 98/11327-9.

References

[1] M.N. Eberlin, Mass Spectr. Rev. 16 (1997) 113.
[2] M. Carvalho, F.C. Gozzo, M.A. Mendes, R. Sparrapan, C. Kascheres, M.N. Eberlin, Chem. Eur. J. 4 (1998) 1161.
[3] F.W. McLafferty, Tandem Mass Spectrometry, Wiley, New York, 1983.
[4] K. Levsen, H. Schwartz, Mass Spectr. 2 (1983) 77.
[5] J.L. Holmes, Org. Mass Spectr. 20 (1985) 169.
[6] K.L. Busch, G.L. Glish, S.A. McLuckey, Mass Spectrometry/ Mass Spectrometry Techniques and Applications of Tandem Mass Spectrometry, VCH, New York, 1988.
[7] R. Tauler, D. Barceló, Trends Anal. Chem. 12 (1993) 319.
[8] R. Tauler, A.K. Smilde, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, Anal. Chem. 66 (1994) 3337.
[9] A.K. Smilde, R. Tauler, J. Saurina, R. Bro, Anal. Chim. Acta 398 (1999) 237.
[10] A.I. Ridorsa, J. Saurina, S.H. Cassou, R. Tauler, Chemom. Intell. Lab. Systems 38 (1997) 183.
[11] R. Tauler, I. Marqués, E. Casassas, J. Chemom. 12 (1998) 55.
[12] C.-N. Ho, E.R. Davidson, Anal. Chem. Soc. 111 (1978) 1108.
[13] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[14] B.E. Wilson, W. Lindberg, B.R. Kowalski, J. Am. Chem. Soc. 111 (1989) 3797.
[15] R. Tauler, A.K. Smilde, B. Kowalski, J. Chemom. 9 (1995) 31.
[16] R. Tauler, Chemom. Intell. Lab. Systems 30 (1995) 133.
[17] V.F. Juliano, F.C. Gosso, M.N. Eberlin, C. Kascheres, Anal. Chem. 68 (1996) 1328.
[18] B.M. Wise, N.B. Gallagher, PLS_Toolbox 2.0, for use with MATLAB.
[19] O.M. Kvalheim, F. Brakstad, Y. Liang, Anal. Chem. 66 (1994) 43.
[20] M. Esteban, C. Arino, J.M. Díaz-Cruz, M.S. Díaz-Cruz, R. Tauler, Trends Anal. Chem. 19 (2000) 49.