

Comparison between the direct trilinear decomposition and the multivariate curve resolution-alternating least squares methods for the resolution of three-way data sets

Anna de Juan ^{a,*}, Sarah C. Rutan ^b, Romà Tauler ^a, D. Luc Massart ^c

^a *Departament de Química Analítica, Universitat de Barcelona, Barcelona, Spain*

^b *Department of Chemistry, Virginia Commonwealth University, Richmond, USA*

^c *ChemoAC, Farmaceutisch Instituut, Vrije Universiteit Brussel, Brussel, Belgium*

Received 12 June 1997; revised 5 November 1997; accepted 18 December 1997

Abstract

Direct trilinear decomposition (DTD) and multivariate curve resolution-alternating least squares (MCR-ALS) methods are two of the most representative three-way resolution procedures. The former, non-iterative, is based on the resolution of the generalized eigenvector/eigenvalue problem and the latter, iterative, is focused on the optimization of initial estimates by using data structure and chemical constraints. DTD and MCR-ALS have been tested on a variety of three-way simulated data sets having common sources of variation in real response profiles, such as signal shift, broadening or shape distortions caused by noise. The effect of these factors on the resolution results has been evaluated through the analysis of several parameters related to the recovery of both qualitative and quantitative information and to the quality of the overall data description. Conclusions inferred from the simulated examples help to clarify the performance of both methods on a real example and to provide some general guidelines to understand better the potential of each method. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Three-way data sets; Direct trilinear decomposition; Multivariate curve resolution-alternating least squares

1. Introduction

Curve resolution methods are usually applied in the framework of Mixture Analysis. Extended work has been reported concerning the resolution of single matrices and the necessary conditions to recover the real response profiles have been recently stated by Manne in two theorems [1]. When these requirements are not fulfilled, the decomposition of a bilinear matrix is subject to ambiguities [2,3] and the true solutions are not obtained.

Many of the limitations associated with the resolution of a single matrix are partially or completely overcome when several matrices (the so-called three-way data sets) are treated together [4]. Three-way resolution methods always introduce a significant improvement in the recovery of the true response profiles and have the additional benefit of providing quantitative information.

Two tendencies prevail within the family of three-way resolution methods: the use of non-iterative procedures, the solutions of which are based on the resolution of a generalized eigenvalue/eigenvector problem [5,6]; and the application of iterative methods,

* Corresponding author.

focused on the optimization of initial estimates by using suitable data structure and chemical constraints [3,7–10].

Direct trilinear decomposition (DTD) [6] and multivariate curve resolution-alternating least squares (MCR-ALS) [3] methods are good representatives of the former and the latter tendencies, respectively. General advantages and drawbacks are recognised for both methods. DTD is fast, user-friendly and furnishes unique solutions. However, trilinear data structure is always assumed (i.e., common concentration profiles and spectra for the same species in the different appended matrices) and it is not possible to input data structure and chemical information to prevent meaningless solutions. ALS optionally assumes trilinearity and therefore, can work with non-trilinear matrices sharing only one order in common (either rows or columns). Input of external information in the resolution procedure is also allowed and use of data point weighting is possible [11]. Despite these advantages, unique solutions are not always guaranteed when the trilinearity constraint is not applied, the iterative optimization slows down the resolution process and more demanding user intervention is required.

Successful chemical applications of DTD [12,13] and MCR-ALS [14–17] have been reported, but there is a lack of comparative literature on the performance of these two methods. This work presents the results of both methods on a variety of simulated three-way data sets including common sources of variation in real chemical data, such as signal shifts, broadening or distortions caused by the presence of noise. Knowledge of these synthetic data sets allows a sound assessment of the performance of both methods in terms of recovery of the response profiles, quantification, and quality of the overall data description. Conclusions inferred from this theoretical study are used afterwards to interpret the DTD and MCR-ALS results when applied to a real example [12] and some general guidelines are suggested to take advantage of the capabilities of both procedures.

2. Theory

Both DTD and MCR-ALS work with the original tensor or stack of matrices (**D**) to obtain three smaller

matrices (**X**, **Y** and **Z** for DTD and **C**, **S** and **Z** for MCR-ALS) containing information associated with the evolution of the pure compounds profiles in each of the three directions of the initial three-way data set. Thus, the results coming from a typical data set formed by several high-performance liquid chromatography with UV diode array detection (HPLC-DAD) samples would be a first matrix having pure chromatographic profiles, a second, including pure spectra and a third one, containing information on the relative concentration of each compound in the different samples.

The three-way data set decomposition is carried out differently by DTD and MCR-ALS, as shown in Fig. 1. The picture shows how trilinearity is inherently assumed in DTD (**X**, **Y** and **Z** provide only one

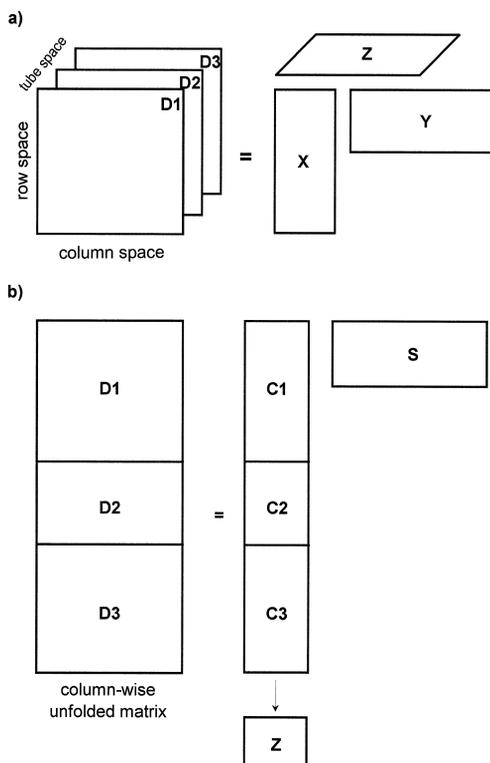


Fig. 1. Three-way data decomposition according to (a) direct trilinear decomposition (DTD) and (b) alternating least squares (ALS) resolution methods.

profile per compound for all the stacked matrices), whereas MCR-ALS allows independent modelling for the concentration profiles in $\mathbf{D}_1, \mathbf{D}_2, \dots$ unless the trilinear constraint is introduced in the resolution process. As can be seen in the picture, the pairs of matrices $(\mathbf{X}, \mathbf{C}), (\mathbf{Y}, \mathbf{S})$ and (\mathbf{Z}, \mathbf{Z}) contain the same kind of information (i.e., concentration profiles, spectra, and quantification of components), obtained by DTD and MCR-ALS, respectively. Detailed explanations about DTD [6,18], MCR-ALS [3,4] and their related algorithms are reported in the literature and will not be given in this section; only the main steps of each method will be mentioned for a better understanding of the comparative study.

The working procedure of DTD can be summarized as follows:

- (i) Application of Singular Value Decomposition to the row-wise, column-wise, and tube-wise unfolded matrices coming from the original tensor.
- (ii) Construction of the basis sets formed by the row space scores (\mathbf{U}), the column space scores (\mathbf{V}), and the first two vectors of the tube space scores (\mathbf{W}).
- (iii) Determination of two representative pseudosamples \mathbf{G}_1 and \mathbf{G}_2 by projection of the original tensor on the $(\mathbf{U}, \mathbf{V}, \mathbf{W})$ basis sets.
- (iv) Determination of \mathbf{X} and \mathbf{Y} matrices from the resolution of the generalized eigenvalue/eigenvector problem for matrices \mathbf{G}_1 and \mathbf{G}_2 .
- (v) Least-squares estimation of \mathbf{Z} , given \mathbf{X} and \mathbf{Y} . MCR-ALS operates following the sequence below:
 - (i) Determination of the number of compounds of the column-wise unfolded matrix.
 - (ii) Building of the initial estimates matrix (either concentration profiles (\mathbf{C}) or spectra (\mathbf{S})).
 - (iii) Selection of the constraints to be input in the iterative resolution process (e.g., non-negativity, unimodality, selectivity, trilinearity, ...).
 - (iv) Optimization of the initial estimates by using a constrained alternating least squares procedure until the convergence criterion is fulfilled. Each iterative cycle includes the calculation and constraint of the \mathbf{C} and \mathbf{S} matrices (see Fig. 1).
 - (v) Determination of the relative concentration of each particular compound in the different data matrices (\mathbf{Z} matrix) as the ratio between the area of its resolved concentration profile in a matrix \mathbf{C}_i

and the area of the same compound in another data matrix \mathbf{C}_j included in the same simultaneous analysis and taken as reference.

Both DTD and MCR-ALS algorithms have been implemented in a set of MATLAB routines [19]. For the sake of simplicity in the abbreviations in text and figures, the MCR-ALS method will be called in the rest of the text, ALS.

3. Data sets

3.1. Simulated data sets

All the three-way data sets are formed by two appended matrices (\mathbf{D}_A and \mathbf{D}_B) sized 110×44 representing a typical binary chromatographic system. In all the examples, the areas of the concentration profiles for compounds 1 and 2 are 25:75 in \mathbf{D}_A matrix and 70:40 in \mathbf{D}_B matrix. The chromatographic profiles have been simulated as slightly tailed Gaussian peaks ($\sigma = 5$ channel units) and a homoscedastic noise with a standard deviation of 0.01% \mathbf{A}_{\max} has been added to all data matrices, unless stated otherwise. This low noise level is introduced to see more clearly the effect of variations other than noise introduction (i.e., variations of signal position or signal shape) on the resolution results. When the effect of noise is assessed, more realistic noise levels are used.

Four basic trilinear systems are proposed to represent the variety of real situations, namely:

System 1: resolution between peaks, $R_s = 0.2$, similar spectra.

System 2: $R_s = 0.2$, dissimilar spectra.

System 3: $R_s = 0.8$, similar spectra.

System 4: $R_s = 0.8$, dissimilar spectra.

In all the simulated data, resolution has been defined as: $R_s = \Delta t_R / (2\sigma_1 + 2\sigma_2)$, where Δt_R is the difference between the maxima of the two simulated peaks and σ_1 and σ_2 are the standard deviations of their basic Gaussian profiles.

Fig. 2 shows the spectra used in the different simulations. Variations in real data like shift effects, signal broadening or noise addition are generated by modifying \mathbf{D}_B or both \mathbf{D}_B and \mathbf{D}_A in the chromatographic direction according to the effect to be studied as explained below. No more than one cause of variation is simulated in each three-way data set for

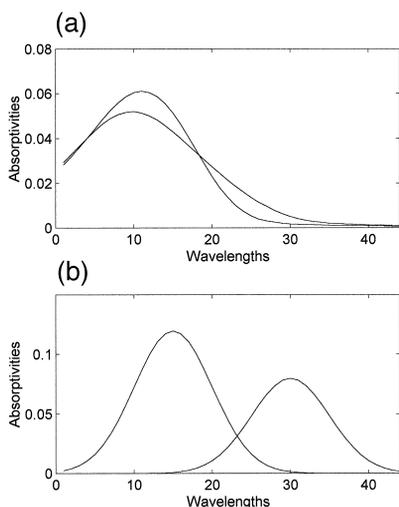


Fig. 2. Spectra used in the simulations of (a) systems 1 and 3, and (b) systems 2 and 4.

the sake of simplicity in the interpretation of the results. The spectra of components 1 and 2 do not change shape or position when going from matrix \mathbf{D}_A to \mathbf{D}_B in any of the simulated data sets.

3.2. Simulations related to signal shift

For each basic system, four three-way examples representing different signal shift levels are generated. In all of them, \mathbf{D}_A matrix is not modified, whereas the peaks related to compounds 1 and 2 in \mathbf{D}_B matrix are shifted apart from their initial position by 1, 2, 4 or 6 channels, depending on the shift intensity. The peak shapes are not modified in any of the two appended matrices.

3.3. Simulations related to changes in signal width

For each basic system, four three-way examples representing different signal width changes are generated. \mathbf{D}_A matrix remains invariant, whereas signal widths of compounds 1 and 2 are modified in matrix \mathbf{D}_B by changing the σ value used in the generation of the peaks. Thus, $\Delta\sigma = -1$ and -2 simulate situations of signal narrowing and $\Delta\sigma = +1$ and $+2$ account for signal broadening (please note that $\Delta\sigma$ is expressed in channel units). No changes in the peak

positions are introduced in any of the two appended data matrices.

3.4. Simulations related to signal distortions caused by noise

The effect of noise pattern and noise level is taken into account in the different simulations. Peak shapes and positions in matrices \mathbf{D}_A and \mathbf{D}_B are not modified. Two noise levels (standard deviation of noise equal to 0.01% A_{\max} and equal to 0.5% A_{\max}) and three noise patterns (homoscedastic, heteroscedastic \propto signal and heteroscedastic $\propto \sqrt{\text{signal}}$) are combined in the six simulations generated for each basic system, as shown below:

hol \rightarrow noise standard deviation: 0.01% A_{\max} .
Noise pattern: homoscedastic.

hsl \rightarrow noise standard deviation: 0.01% A_{\max} .
Noise pattern: heteroscedastic \propto signal.

hql \rightarrow noise standard deviation: 0.01% A_{\max} .
Noise pattern: heteroscedastic $\propto \sqrt{\text{signal}}$.

hoh \rightarrow noise standard deviation: 0.5% A_{\max} . Noise pattern: homoscedastic.

hsh \rightarrow noise standard deviation: 0.5% A_{\max} . Noise pattern: heteroscedastic \propto signal.

hqh \rightarrow noise standard deviation: 0.5% A_{\max} . Noise pattern: heteroscedastic $\propto \sqrt{\text{signal}}$.

4. Results and discussion

4.1. Comments on the simulated data

Both DTD and ALS have been run assuming the existence of two compounds in all the three-way data sets. DTD has been applied in the different unfolding directions and the results presented in this work are those related to the direction giving the best data fit. The Alternating Least Squares method has always been used introducing the non-negativity constraint in both chromatographic and spectral directions and the unimodal constraint in the chromatographic profiles. All the examples have adopted chromatographic profiles as initial estimates. The initial concentration profiles for systems 1, 3 and 4 have been built applying the needle search methodology [20]. The spectral selectivity detected by using local rank detection methods, such as FSMW-EFA [21], has been taken

into account to build the initial estimates of system 2. In this case, column vectors (chromatograms) placed in spectral selective zones of the original data matrices have been selected to be input as initial chromatographic profiles. The use of random estimates has been avoided. Owing to the special features of the chemical data sets that monitor the evolution of a process (e.g., a chromatographic elution, the development of a reaction, ...), initial estimates based on the internal structure of the data set can be easily obtained by using a big variety of data analysis methods (e.g., Evolving Factor Analysis [22], SIMPLISMA [23], or the ones mentioned above). These estimates, which were directly derived from the mathematical analysis of the data sets and which do not include information imposed externally by the chemist, are always better starting points than random estimates for this kind of structured data sets. The alternating least squares method has been run twice for each example: first, forcing trilinear structure in the model and second, without doing so. To make a difference between the results coming from these two different options, the acronyms ALSf and ALS will be used hereafter to refer to the presence or absence of the trilinear constraint in the iterative optimization procedure, respectively.

The performance of DTD and the Alternating least squares methods have been compared through the observation of several parameters related to the quality of the final results. The recovery of the qualitative information, i.e., chromatographic and spectral profiles, is assessed by looking at the dissimilarities between the true response profiles and the profiles obtained by using the resolution methods.

$$\text{dis}(x_{\text{calc}}, x_{\text{true}}) = \sqrt{1 - r^2(x_{\text{calc}}, x_{\text{true}})}$$

where x_{calc} and x_{true} are the response profile calculated with ALS or DTD and the true response profile, respectively, and r is the correlation coefficient between these two factors.

The quality of the overall data description is related to the percent of lack of fit, and is calculated as follows:

$$\% \text{ lack of fit} = 100 \sqrt{\frac{\sum_{i,j,k} e_{ijk}^2}{\sum_{i,j,k} d_{ijk}^2}}$$

where e_{ijk} is the residual of the ijk th element of the three-way data set, found as the difference between the element in the original tensor and the same element reproduced by using the results of the resolution method and d_{ijk} is the ijk th element of the original three-way data set. The quantification ability has been evaluated with a relative parameter of comparison between the true concentration ratios of a compound in the \mathbf{D}_A and \mathbf{D}_B matrices and the same concentration ratios obtained from the resolution results. The mathematical expression of this last parameter is shown below (the i th subscript indicates the different compounds in the data set).

$$\% \text{ rsdz} = 100 \sqrt{\frac{\sum_i \left(\left(\frac{c_{i,A}}{c_{i,B}} \right)_{\text{true}} - \left(\frac{c_{i,A}}{c_{i,B}} \right)_{\text{calc}} \right)^2}{\sum_i \left(\frac{c_{i,A}}{c_{i,B}} \right)_{\text{true}}^2}}$$

The conclusions inferred from the comparison of the DTD and alternating least squares methods change completely if the variation introduced in the three-way data sets is associated with a loss of the trilinear data structure (i.e., changes in width and signal position) or if it is not (i.e., signal distortions produced by noise). In order to clarify the understanding of the conclusions, two separated sections are devoted to explaining the results obtained for trilinear and non-trilinear data sets.

4.2. Comments on the resolution of non-trilinear data sets (effect of changes in width and signal position)

As a general rule, ALS works much better than DTD and ALSf in all aspects when trilinearity is broken, owing to its greater flexibility in the modelling of the response profiles. The best performance of ALS in these situations is specially clear when the analyzed three-way data sets have selectivity in the chromatographic direction (data sets coming from system 3), in the spectral direction (data sets coming from system 2) or in both of them (data sets coming from system 4), as shown in the example of Fig. 3. Indeed, when selectivity is present, there are no ambiguities in the decomposition of the augmented matrix by ALS [3] and the minimal dissimilarities between the true and the calculated profiles are due to

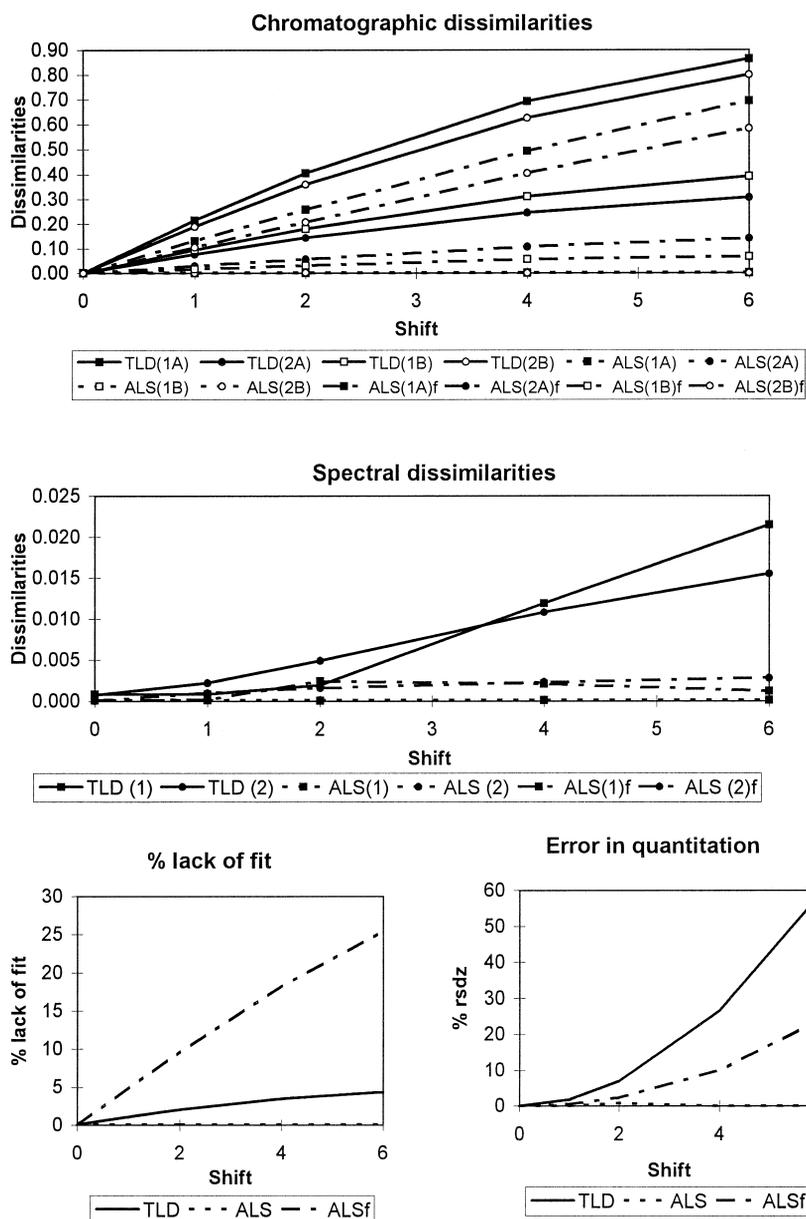


Fig. 3. Effect of the signal shift on the resolution of the three-way data sets generated by modifying the basic system 3. Solid lines: DTD; Pointed lines: ALS; Dashed lines: ALSf; Compound one: squares; Compound two: circles; Compounds in \mathbf{D}_A matrix: filled symbols; Compounds in \mathbf{D}_B matrix: empty symbols. Legends: e.g., DTD(1A), the number and the letter between parentheses when existing indicate the compound (1 or 2) and the data matrix (\mathbf{D}_A or \mathbf{D}_B) for which a certain parameter has been determined.

the noise added to all data sets (please note that for binary systems, the conditions required by Manne's theorems for the correct resolution of data matrices coincide with the presence of selectivity for both

compounds). When neither spectral nor chromatographic selectivity exists, as in system 1, ALS still performs better in most of the cases, as shown in Figs. 4 and 5. Only some reversals in this tendency can be

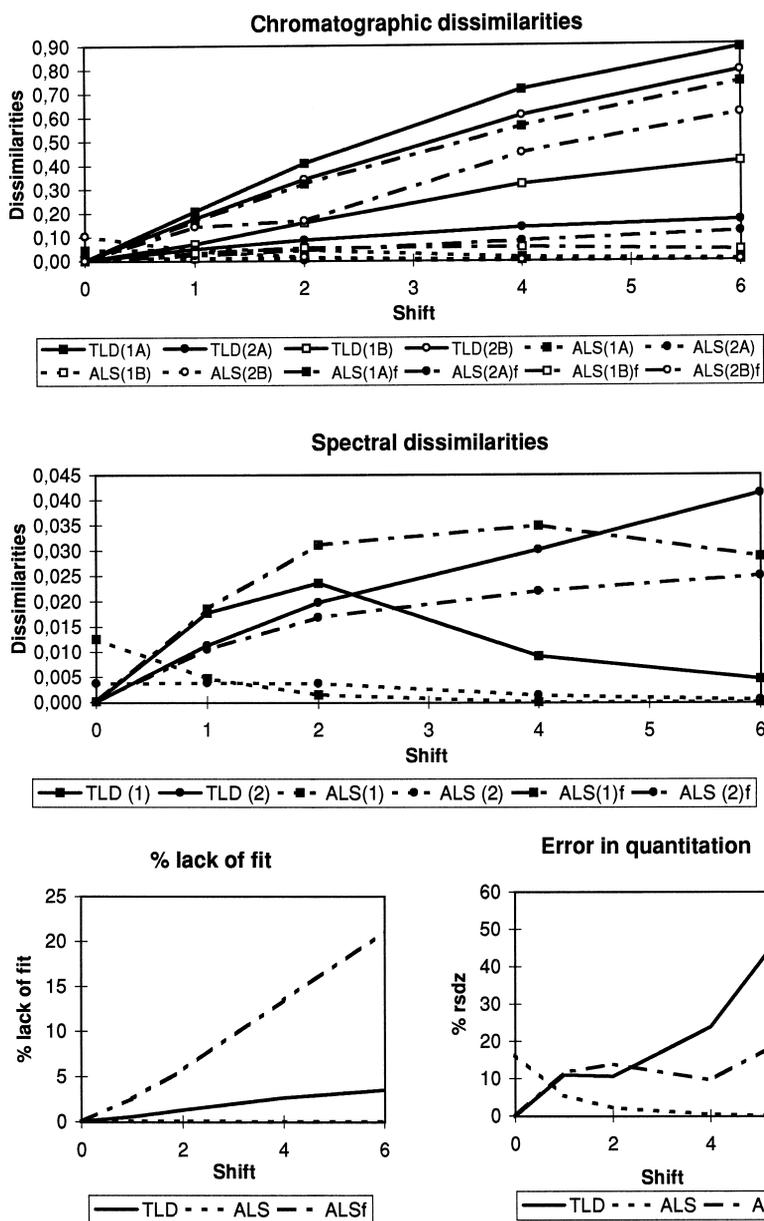


Fig. 4. Effect of the signal shift on the resolution of the three-way data sets generated by modifying the basic system 1. Legends, line and symbol patterns are the same as in Fig. 3.

found when the departures of trilinearity are extremely small or when these departures are associated with an increase in the overlapping between compounds (i.e., moderate positive signal broadening). In the latter case, there is a critical zone ($+2 >$

$\Delta\sigma > +1$) where the negative effects caused by the increase of correlation in the chromatographic direction on ALS and by the loss of trilinearity on DTD and ALSf are comparable. Such a behavior is specially noticed in the recovered response profiles,

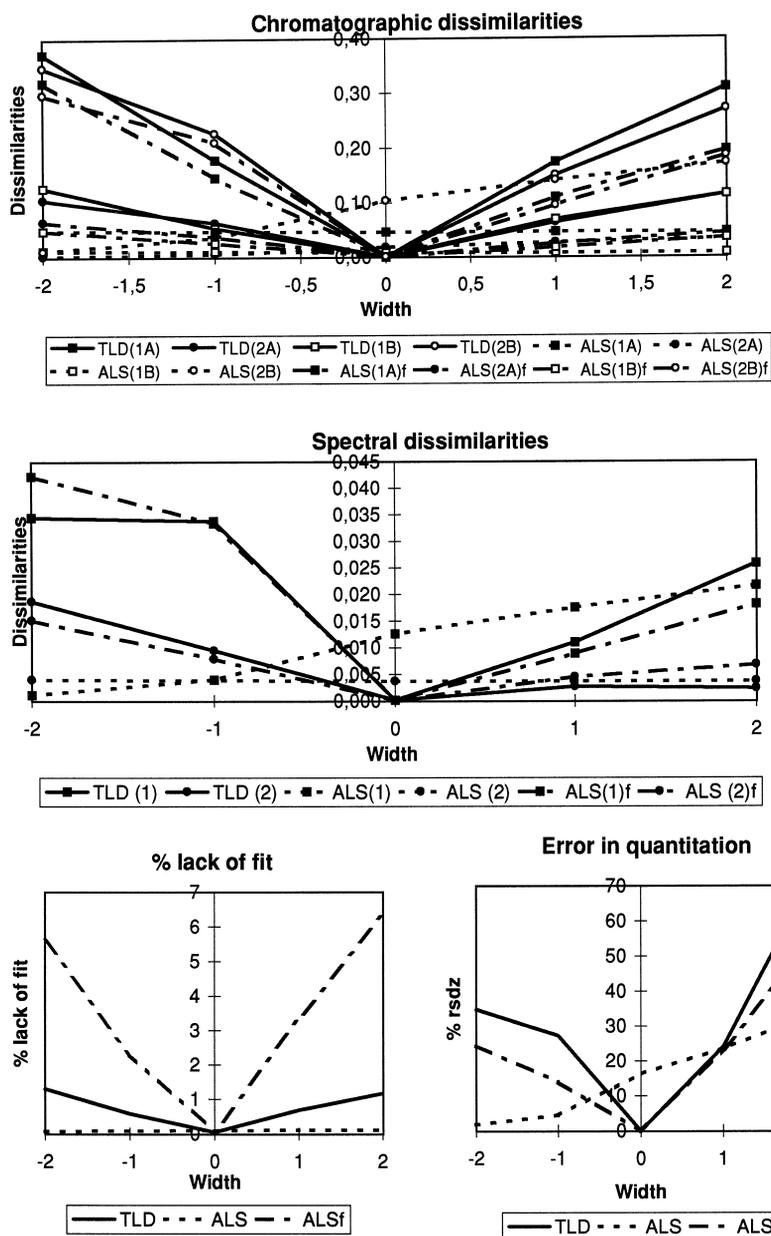


Fig. 5. Effect of width signal changes on the resolution of the three-way data sets generated by modifying the basic system 1. Legends, line and symbol patterns are the same as in Fig. 3.

which are actually linear combinations of the pure profiles. This distortion in the profiles shape also influences the quantification ability in the DTD and the alternating least squares procedures. On the one hand, variations in the form of the profiles (i.e., in matrices

\mathbf{X} and \mathbf{Y}) modify the least-squares calculation of the \mathbf{Z} matrix in DTD and, on the other hand, these variations are also associated with changes in the peak area of the different compounds, this parameter being used to obtain the quantitative information in the ALS and

ALSf. In contrast to the parameters above, the lack of fit is only severely affected by losses of trilinearity. In DTD and ALSf, the non-trilinear contributions present in the data and excluded from the model (which would account for the shifts or shape variations for the same compound in the two treated matrices), are responsible for the increase in the lack of fit. Owing to the greater flexibility in the modelling of the profiles, ALS always has more freedom in the data reproduction, though using linear combinations of the real solutions, when there is no selective information in the system and, therefore, the data fit is always better than for methods with trilinear structure imposed. This behavior and the possibility of calculating this parameter for real data sets provide an additional method to check the existence of trilinear structure in real three-way data sets since non-trilinear systems will always have a significantly greater lack of fit for DTD than for ALS when the same number of compounds is used in the resolution process.

Figs. 4 and 5 clearly show the factors that negatively affect the different resolution methods. Whereas DTD and ALSf are mainly influenced by trilinearity losses (i.e., by increases in the signal shift or in the changes of signal width, whatever their sign), ALS is more sensitive to variations in the correlation between the response profiles of the data set. Thus, improvements in the ALS results are associated with a decrease in the overlapping between compounds (i.e., big signal shifts or signal narrowing) and the opposite effect is promoted by increases in the correlation between profiles (i.e., signal broadening).

In all the comments above, DTD and ALSf have been put in the same group when compared with ALS. However, noticeable differences between them derive from the distinct way to include the trilinear condition in their resolution processes. DTD works on the whole data set to obtain the trilinear combination of response profiles that best fit the original data. This strategy gives priority to the data fit over the recovered shape for the response profiles. ALSf operates one by one on each column of the augmented concentration matrix to get the best common and constrained profile for each compound [24]. In contrast to DTD, ALSf is more focused on the quality of the individual response profiles than on the global fit. The

explanations above are in agreement with the general tendencies observed in Figs. 4 and 5, where ALSf provides smaller dissimilarities and errors in quantification, whereas DTD shows lower lack of fit. The larger lack of fit in the ALSf models is related to the removal of both the non-trilinear contributions and the trilinear information not in agreement with the constraints imposed in the optimization procedure (e.g., non-unimodal and negative parts of the concentration profile, ...). The larger distortions in the response profiles found when DTD is applied are due to the acceptance of any kind of profile shape, provided that it keeps the trilinear condition.

4.3. Comments on the resolution of trilinear data sets (effect of noise addition)

An unavoidable element in the real data sets is the presence of noise. This factor alone does not induce trilinearity losses unless it appears simultaneously with some of the data variation sources mentioned above. The simulated examples are used to analyze the stability of the resolution methods against the only introduction of different noise levels and patterns in the data sets.

All the tested three-way resolution methods are more affected by increases in the noise level than by variations in the noise pattern. Generally speaking, both DTD and ALSf provide more satisfactory results than ALS for the original noisy data sets.

The best performance of DTD and ALSf is more evident when there is no selectivity in the 3-way data set, as shown in Fig. 6. In this case, owing to the intrinsic trilinear structure of the data sets and the forced trilinear character of the resolution results, both methods are able to reach the real unique solutions. In contrast, ALS suffers from the ambiguity associated with the decomposition of the bilinear matrices lacking selective information and the profiles obtained, though correctly reproducing the original data, are linear combinations of the true response profiles. The results obtained with DTD and ALSf are fairly similar. For the sake of speed and for easy application, DTD is more recommendable to treat trilinear three-way data sets with no selective information.

The application of the tested resolution methods to data sets with selective information, i.e., to the ex-

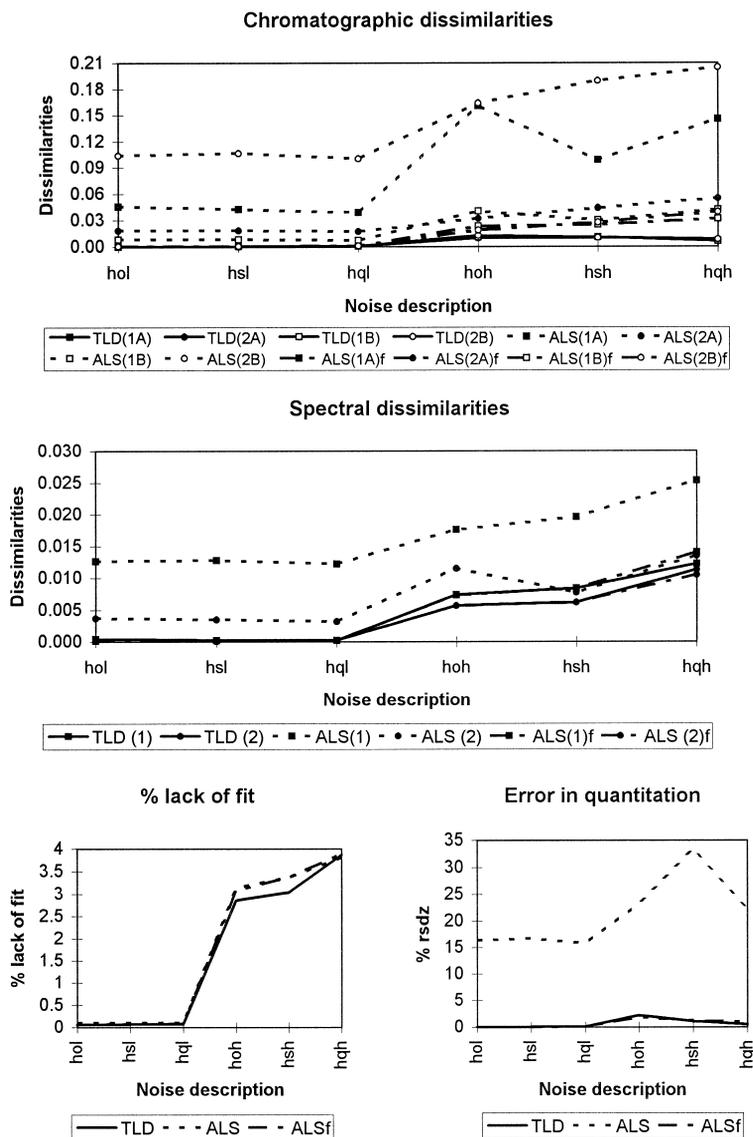


Fig. 6. Effect of noise on the resolution of the three-way data sets generated by modifying the basic system 1. Legends, line and symbol patterns are the same as in Fig. 3. Noise description of the data sets: **hol** \rightarrow noise level: 0.01% A_{\max} Noise pattern: homoscedastic. **hoh** \rightarrow 0.5% A_{\max} , homoscedastic. **hsl** \rightarrow 0.01% A_{\max} , heteroscedastic \propto signal. **hsh** \rightarrow 0.5% A_{\max} , heteroscedastic \propto signal. **hql** \rightarrow 0.01% A_{\max} , heteroscedastic $\propto \sqrt{\text{signal}}$. **hqh** \rightarrow 0.5% A_{\max} , heteroscedastic $\propto \sqrt{\text{signal}}$.

amples coming from systems 2, 3, and 4, offers a rather different situation (see Fig. 7). DTD, ALS and ALSf all get unique solutions in a very good agreement with the true profiles. Uniqueness in ALS is due to the presence of selectivity in the data sets. Nevertheless, DTD and ALSf usually show a subtle superi-

ority with respect to ALS, although some occasional reversals in this behavior can be found in some concrete simulations. Such a difference can likely come from the lesser introduction of noise in the profiles recovered by DTD and ALSf. Therefore, DTD, ALSf and all other resolution methods modelling the re-

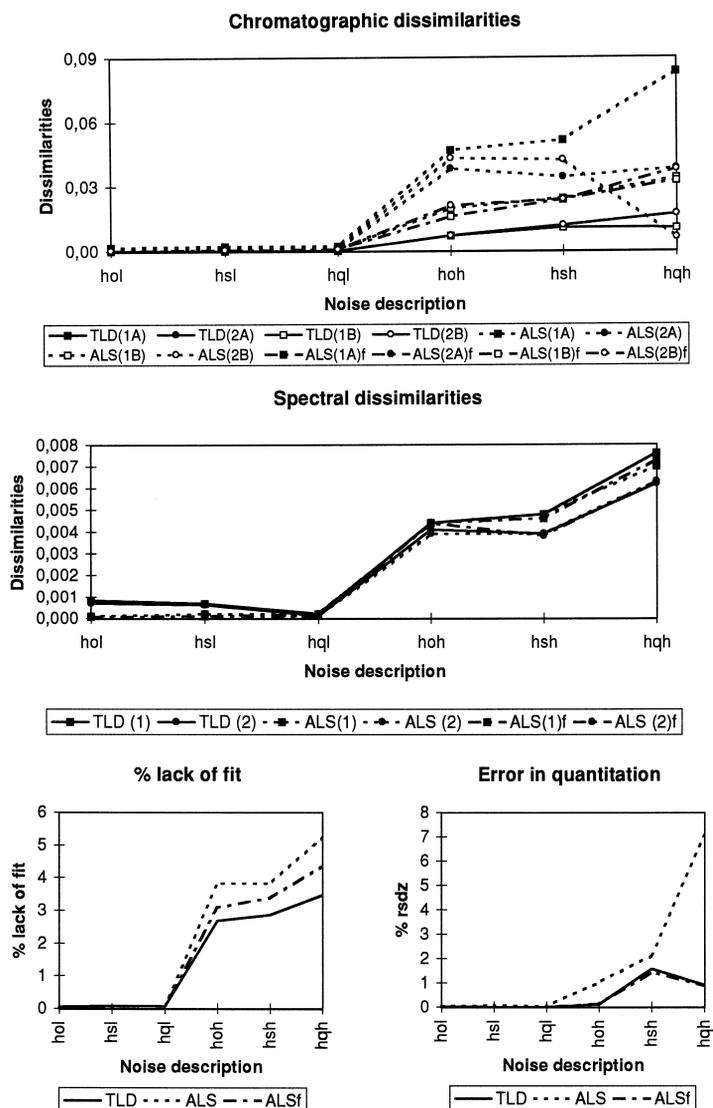


Fig. 7. Effect of noise on the resolution of the three-way data sets generated by modifying the basic system 3. Line and symbol patterns are the same as in Fig. 3. Noise description are the same as in Fig. 6.

response profiles by only taking the trilinear information of the data set, i.e., the common information shared by all the matrices treated together, perform a smoothing effect on the profiles obtained owing to the non-inclusion of the diverse noise contributions in the resolution process. In consequence, the profiles so recovered are more similar to the true noise-free signals than those obtained by using more flexible methods, such as ALS, where the presence of noise

in the final profiles can only be suppressed if it distorts the profiles in such a way that the constraints imposed in the optimization procedure are not respected.

5. Comments on a real example

The selected example consists of thin layer chromatographic data with fluorescence detection [8]. The

matrices contain glycine and/or glutamine as analytes and two contributions owing to background effects, which are also modelled as additional compounds. The three-way data sets are formed by three matrices of glycine standards, three of glutamine standards and three having mixtures of both analytes. The two background contributions are present in all matrices. To compare the performance of the resolution methods, the data sets called experiments 1a and 2, have been chosen. Both data sets have the structure mentioned above. Experimental details about these data can be found in the original work.

A prior exploratory analysis shows the presence of a high noise level in the data sets. The application of EFA and FSMW-EFA reveals the lack of selectivity in any of the directions of the 3-way data sets. Such a feature is due to the existence of background contributions during the whole elution process and to the big chromatographic and spectral overlap between the two analytes, glycine and glutamine.

The internal structure of the data sets can be known through the comparison of the lack of fit related to the ALS and DTD results. ALS has been run using the non-negativity constraint in both chromatographic and spectral directions. Unimodality has been applied to the concentration profiles of glycine and glutamine, whereas the background profiles have not been forced to respect this constraint. DTD results are those obtained in the unfolding direction giving the best data fit. ALS gives a lack of fit significantly lower than that of DTD for both experiment 1a (8.29% vs. 17.3%) and experiment 2 (9.58% vs. 15.46%) and therefore, the data sets are clearly non-trilinear.

Experiment 1a and experiment 2 are classified within the group of non-trilinear systems without selective information. According to the conclusions inferred from the simulated examples, the choice of ALS would be clear if there was complete or partial selectivity in the system. As this is not so, this real example can be placed in the critical zone where both DTD and ALS solutions, the first, because of the non-trilinear structure of the data and the second, because of the big overlap between compounds and the high noise level, are comparably damaged. In this situation, all the resolution methods can fail to find out the true solutions, but the application of both DTD and ALS can provide complementary information in

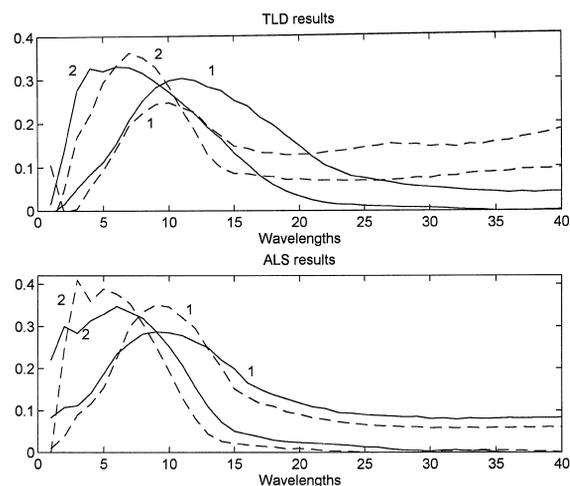


Fig. 8. Spectra recovered for glycine (1) and glutamine (2) in experiment 1a (solid lines) and experiment 2 (dashed lines) by using DTD and ALS.

order to obtain solutions as close as possible to the real ones.

A rather general observation is the best performance of ALS to recover the qualitative information. This is due to the inclusion of the non-trilinear contributions to build up the concentration profiles and the spectra and to the more reproducible procedure followed in the modelling of the response profiles by using the suitable chemical constraints. Fig. 8 shows the robustness of the ALS modelling procedure through the comparison of the glycine and glutamine spectra recovered for the two three-way data sets studied independently, experiment 1a and experiment 2. As DTD works to have the best trilinear combination of response profiles and no constraints are used to control the profiles shape, the variability of the spectra between different analogous data sets is much larger and depends on how far from the trilinear situation the different data sets are.

In contrast to the qualitative aspects of the resolution methods, there is no fix rule to know which method will provide the best quantification. Depending on how far the system from the trilinear condition is and on some other data features, such as the noise level, the quantification ability can turn out to be better either by applying ALS or using DTD. There is no objective way to know which method must be chosen to quantify a real data set unless, as in the

Table 1
Correlation coefficients from the calibration lines obtained comparing actual and predicted concentrations

Experiment	DTD ^a		ALS ^a	
	Glycine	Glutamine	Glycine	Glutamine
1a	0.930	0.993	0.836	0.860
2	0.362	0.693	0.871	0.920

^aSix samples (the three suitable standards (glycine or glutamine) and the three values from the mixture matrices) are used to obtain each correlation coefficient.

present example, some standards are available. In this case, the method having a better predictive ability for the known standards will consequently quantify more accurately the unknown samples. Table 1 shows the quantitative information found by DTD and ALS for the experiment 1a and experiment 2 data sets. DTD is the best method to quantify the first data set, that is likely close to be trilinear, and ALS behaves better to quantify the second. The poor quantitative performance of DTD in this last case must be attributed to the larger departure from trilinearity of the data set experiment 2.

6. Conclusions

The existence of selectivity and the internal structure of the three-way data set (trilinear or non-trilinear) are the key features in the selection of the most suitable resolution procedure. Therefore, the first step is to classify the real data set according to these two properties. The comparison of the lack of fit obtained with DTD and ALS is proposed as a simple method to confirm the presence or absence of trilinearity in a three-way data set, i.e., if the lack of fit for ALS is clearly lower than for DTD using models with the same number of components, the system is non-trilinear, and vice versa if the lack of fit for DTD is equal or lower than for ALS. The presence of selectivity can be easily found by applying local rank analysis, such as FSMW-EFA.

DTD and ALSf work much better than ALS with trilinear systems lacking selectivity. The presence of selectivity in the data sets makes the ALS results comparable to those coming from DTD and ALSf, though the smoother solutions of the latter are usu-

ally slightly better. Within DTD and ALSf, the first one is normally chosen because of the speed and easy applicability.

ALS is the best method to deal with non-trilinear data sets. Systems lacking selectivity and having a slight departure of trilinearity, as the real example presented, are the only ones showing solutions of similar quality with all the tested resolution methods. In these situations, ALS appears to be more robust in the modelling of the response profiles whereas either DTD or ALS can provide a best quantification.

7. Additional information

Graphical results completing those shown in the text about shift effect, change of signal width, and noise addition on the resolution results are available on request.

Acknowledgements

A.J. and R.T. acknowledge the financial support of the Spanish government through the DGICYT PB96-0377 project.

References

- [1] R. Manne, *Chemom. Intell. Lab. Syst.* 27 (1995) 89.
- [2] W.H. Lawton, E.A. Sylvester, *Technometrics* 13 (1971) 617.
- [3] R. Tauler, A.K. Smilde, B.R. Kowalski, *J. Chemometrics* 9 (1995) 31.
- [4] R. Tauler, *Chemom. Intell. Lab. Syst.* 30 (1995) 133.
- [5] E. Sánchez, B.R. Kowalski, *Anal. Chem.* 58 (1986) 496.
- [6] B.E. Wilson, E. Sánchez, B.R. Kowalski, *J. Chemometrics* 3 (1989) 493.
- [7] A.K. Smilde, R. Tauler, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, *Anal. Chem.* 66 (1994) 3345.
- [8] J.D. Carroll, J.J. Chang, *Psychometrika* 35 (1970) 283.
- [9] R.A. Harshman, *UCLA Work. Papers Phonet.*, 16 (1970) 1.
- [10] R.A. Harshman, *UCLA Work. Papers Phonet.*, 22 (1972) 111.
- [11] R. Ross, S. Leurgans, *Meth. Enzymol.* 246 (1995) 679.
- [12] M. Gui, S.C. Rutan, A. Agbodjan, *Anal. Chem.* 67 (1995) 3293.
- [13] K.S. Booksh, Z. Lin, Z. Wang, B.R. Kowalski, *Anal. Chem.* 66 (1994) 2561.
- [14] R. Tauler, A.K. Smilde, J.M. Henshaw, B.R. Kowalski, *Anal. Chem.* 66 (1994) 3337.

- [15] R. Tauler, A. Izquierdo-Ridorsa, R. Gargallo, E. Casassas, *Chemom. Intell. Lab. Syst.* 27 (1995) 163.
- [16] A. de Juan, A. Izquierdo-Ridorsa, R. Tauler, G. Fonrodona, E. Casassas, *Biophys. J.* 73 (1997) 2937.
- [17] X. Saurina, S. Hernández-Cassou, R. Tauler, *Anal. Chim. Acta* 67 (1995) 3722.
- [18] E. Sánchez, B.R. Kowalski, *J. Chemometrics* 4 (1990) 29.
- [19] MATLAB version 4.2. The Mathworks, 1994.
- [20] A. de Juan, B. van den Bogaert, F. Cuesta Sánchez, D.L. Massart, *Chemom. Intell. Lab. Syst.* 33 (1996) 133.
- [21] H.R. Keller, D.L. Massart, *Anal. Chim. Acta* 246 (1991) 379.
- [22] H. Gampp, M. Maeder, C. Meyer, A.D. Zuberbühler, *Anal. Chim. Acta* 193 (1987) 287.
- [23] W. Windig, J. Guilment, *Anal. Chem.* 63 (1991) 1425.
- [24] R. Tauler, D. Barceló, *Trends Anal. Chem.* 12 (1993) 319.