Evaluation of Parallel Factor Analysis for the Resolution of Kinetic Data by Diode-array High-performance Liquid Chromatography

Peter Hindmarch, Keyhandokht Kavianpour and Richard G. Brereton*

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

The PARAFAC algorithm for factor analysis of three or higher way datasets is summarised. A series of simulations of kinetic profiles of two-way diode-array HPLC data is described. A three-phase reaction system of reactant, intermediate and product is used to illustrate the method, each closely eluting and with similar spectra based on experimental HPLC with diode-array detection of chlorophyll degradation products. A kinetic parameter is varied to change the relative concentration of the intermediate in each series of simulations. Several indices of quality of reconstruction are introduced. It is concluded that the number of factors used to model the data is crucial to the quality of reconstruction. A good approach is first to use fewer factors than are expected, then increasing the number until each elution profile shows a single maximum.

Keywords: *Deconvolution; PARAFAC; high-performance liquid chromatography; kinetics; chlorophyll*

Three-way data are common in analytical chemistry.[1,2] An example is a series of chromatograms recorded in time. If these chromatograms consist, in turn, of two-way data such as in HPLC with diode-array detection (DAD) or GC–MS, the full series of chromatograms may be regarded as a three-way dataset. One mode is time or sample number, whereas the other modes are elution time and a spectroscopic parameter, such as wavelength or mass number. Conventionally, each chromatogram is analysed independently by factor analysis or multivariate calibration, but this ignores the fact that there are components common to the entire series of chromatograms with similar spectra and elution profiles. Treating the entire dataset as one three-dimensional block provides more information than treating each chromatogram separately.

There are several methods for three-dimensional factor analysis,[3–7] and it is the purpose of this paper to evaluate one of the most common, called PARAFAC. In this approach, the three-dimensional data are decomposed into a series of factors, each relating to one of the three physical variables.

Theory

PARAFAC (parallel factor analysis) is a method of decomposing a three-way data array, or tensor, into a series of two-way arrays. The original algorithms were developed by psychometricians for the decomposition of multiblock data.[8–11] Mathematically, PARAFAC can be seen as a simplification of the Tucker3 Model proposed by Tucker,[12] in which a three-way $I \times J \times K$ array is decomposed into three loadings matrices $\boldsymbol{A}(I \times L)$, $\boldsymbol{B}(J \times M)$ and $\boldsymbol{C}(K \times N)$, where L, M and N are the number of factors in the first, second and third modes, respectively. I, J, and K may be regarded as the number of samples, elution times and wavelengths, respectively. In most areas of chemistry, L, M and N will be equal and are the number of detectable components in a mixture, making the chemometric problem simpler than the psychometric problem.

In this case, a three-way array (or tensor) $\underline{\boldsymbol{X}}$, whose dimensions are sample number, elution time and spectral wavelength in the case of HPLC–DAD, is decomposed into three matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ such that,

$$x_{i,j,k} = \sum_{1}^{F} a_{i,f} b_{j,f} c_{k,f} + e_{i,j,k} \quad (1)$$

where F is the number of factors used in the model and e is the error term. $\boldsymbol{A}$ is a matrix of I rows consisting of sample numbers and L columns consisting of the number of detectable components in the mixture; $\boldsymbol{B}$ and $\boldsymbol{C}$ correspond to the elution profiles and spectra of these L components.
This model can also be written as

$$\underline{\boldsymbol{X}} = \sum_{1}^{F} \boldsymbol{a}_f \otimes \boldsymbol{b}_f \otimes \boldsymbol{c}_f + \underline{\boldsymbol{E}} \quad (2)$$

where $\otimes$ represents the ternary tensor product of the three vectors. The field of tensor algebra as applied to chemical data is discussed extensively elsewhere.[13] The definition and representation of tensor products varies depending on the context, but it is sufficient here to state that the tensor product of vectors ${}_{I,1}\boldsymbol{a}$, ${}_{J,1}\boldsymbol{b}$ and ${}_{K,1}\boldsymbol{c}$ is a vector with co-ordinates $x_i y_j z_k$. Combining all of these vectors over all factors gives the three-way data matrix. Graphically, the PARAFAC model for a three-way, two-component system is shown in Fig. 1.

The simplest way of implementing a PARAFAC model is by alternating least squares. Starting with a known three-way matrix, $\underline{\boldsymbol{X}}$, and two randomly initiated loadings matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, the third loadings matrix, $\boldsymbol{C}$, can be estimated. Then, from this new estimate of $\boldsymbol{C}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ then $\boldsymbol{C}$ can be successively estimated and so on until there is convergence in the model.

The advantage of this approach is that, apart from the number of factors, F, no prior knowledge of the system is required. Furthermore, apart from scaling considerations, PARAFAC produces a unique solution. However, as it is a numeric rather than an analytical method, care must be taken to ensure that the algorithm converges properly and operates at an acceptable speed. The theory of the PARAFAC algorithm is discussed further elsewhere,[14–16] as also are further applications to calibration.[17–20] Practical applications, however, have been limited.[21–23]

Method

Experimental

Spectra of three chlorophyll degradation products were obtained experimentally from a Waters (Milford, MA, USA)

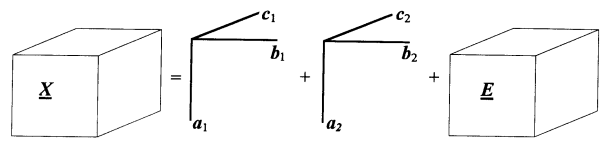

Fig. 1 Graphical representation of the PARAFAC method.

Model 990 HPLC–DAD system. For chlorophyll degradation mixtures, closely eluting compounds often possess very similar spectral characteristics, posing real problems in resolution by chemometric means. Spectra were recorded between 350 and 800 nm at 2 nm intervals.

A common problem involves detecting crucial intermediates that are present in low concentrations, *e.g.*, in the reaction A → B → C, where the first reaction is slow and the second is fast. A full kinetic model for the pathway requires the detection of B, which may be present in small amounts, dependent on the relative rates of the two reactions. If HPLC is employed to study such reactions, the intermediate, which may be a stereoisomer of the parent compound, could have very similar elution profiles and spectra to one of the other components. This situation is well established in the study of the degradation of chlorophyll by HPLC .

Simulation Design

The data were designed to simulate the degradation of chlorophyll-*a* products investigated elsewhere.[24] Each three-way dataset is constructed from modes representing elution profiles, spectra and degradation profiles of each component.

Three spectra obtained experimentally were used to represent three compounds with very similar spectral characteristics. The elution profiles were simulated and the degradation profiles represent a series of HPLC–DAD traces used to monitor a three-reactant system.

The three species formed are assumed to form part of a reaction series, where the reactant (R) is converted into an intermediate (I), which is then converted into a product (P).

Elution Profiles

The elution data matrix is represented by ${}_{20,3}\boldsymbol{A}$, representing 20 chromatographic points in time. The chromatographic profile of compound l (where l = 1, 2 and 3 for the reactant, intermediate and product, respectively) at time I is given by $a_{l,i.}$ Each column of $\boldsymbol{A}$ is an individual elution profile represented in this simulation by Gaussians centred at points 10, 14 and 6 in time and given by

$$a_{i,1} = e^{-\frac{(i-10)^2}{6}} \quad (5)$$

$$a_{i,2} = e^{-\frac{(i-14)^2}{6}} \quad (6)$$

$$a_{i,3} = e^{-\frac{(i-6)^2}{6}} \quad (7)$$

The data are designed so that the product elutes first, followed by the reactant and then the intermediate. The elution profiles are shown graphically in Fig. 2.

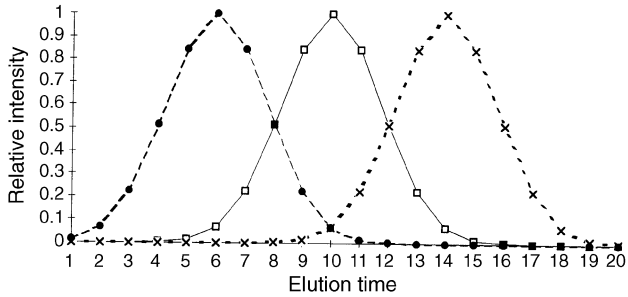

Fig. 2 Simulated elution profiles. □, Reactant; ×, intermediate; and ●, product.

Spectra

Spectra were chosen from previous work to represent each component. Brereton and co-workers[24,25] have shown that in chlorophyll degradation studies the ratio of absorbances between the two absorbance maxima, at approximately 430 and 665 nm, respectively, is very diagnostic. These spectra were chosen so that the reactant and the intermediate had spectra with similar features. Table 1 gives the spectral characteristics of each component. The spectra, shown in Fig. 3, were scaled to constant maximum absorbance and stored as ${}_{226,3}\boldsymbol{B}$, where each column represents a spectrum taken between 350 and 800 nm, digitised at a resolution of 2 nm, *i.e.*, 226 readings per spectrum.

Degradation profiles

The degradation profiles were designed to represent a reactant decreasing in concentration as the experiment proceeds, a minor intermediate increasing to a maximum, then decreasing and a product increasing in concentration. Concentrations are such that at any time the total concentration of all species is constant. The degradation data matrix ${}_{20,3}\boldsymbol{C}$ represents the concentration of the three components at 20 sampling points throughout the experiment where the three columns represent the reactant, intermediate and product respectively. The profiles for each component are given by

$$c_{k,1} = 5e^{-k/8} \quad (8)$$

$$c_{k,2} = (5 - 5e^{-k/8})e^{-k/d} \quad (9)$$

$$c_{k,3} = 5 - 5e^{-k/8} + e^{-k/d} - 5e^{-(k/8 + k/d)} \quad (10)$$

where k is the sample number arranged sequentially in time and d is a parameter varying according to the relative significance and kinetic stability of the intermediate. The greater the value of d the slower is the decomposition of the intermediate, and so the easier it is to detect. Different simulations were performed at different values of d and simulated degradation profiles are shown in Fig. 4.

Table 1 Position of maxima of designed and predicted elution profiles.

	Elution maximum		
	Reactant	Intermediate	Product
	10	14	6
Rate parameter, *d*	Factor 1	Factor 2	Factor 3
1	10	10	6
3	10	10	6
5	10	6	14
10	6	10	14
20	14	6	10

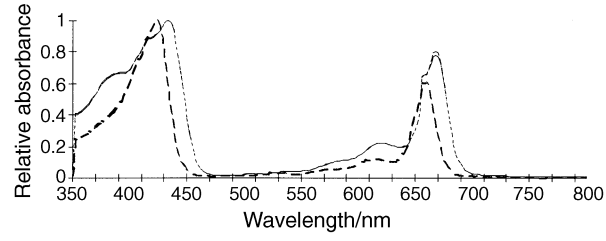

Fig. 3 Experimentally obtained spectra used in the simulations. Solid line, reactant; dotted line, intermediate; and dashed line, product.

Formation of three-way data set

The three-way, three-component model $_{20,226,20}\ \underline{\boldsymbol{X}}$ is formed by

$$x_{i,j,k} = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f} \tag{11}$$

where f is the component number and F is the number of components, which in this study is three. Five datasets were created in which the rate constant, d, was 1, 3, 5, 10 and 20.

Application of PARAFAC Algorithm

No pre-processing is performed on the data in this paper. The issue of pre-processing of three-way arrays is more complex than for the two-way case. Centring can be performed across either one, two or three of the modes and can distort the trilinear model. The order of any pre-processing is also critical. These issues have been discussed elsewhere.[9,15,26,27]

The datasets were decomposed by the PARAFAC written in Matlab 4.2 (Mathworks, Natick, MA, USA).

The algorithm was used to extract three factor matrices from each simulated dataset initialised using random vectors and a convergence limit of 1×10^{-6} between successive estimates of the sum of squares of the misfit.

Indicators of Quality of Reconstruction

Various functions can be used to compare the results from the simulations with the design data.

Component sum of squares

This gives a measure of the size of each component and factor, which aids in the identification of factors and gives an indication of their purity. The size of each predicted component, $\hat{S}_f$, is given by

$$\hat{S}_f = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} (\hat{a}_{i,f}\hat{b}_{j,f}\hat{c}_{k,f})^2 \tag{12}$$

The values obtained for the predicted model using eqn. (12) can be compared to the size of the true components, S_f, calculated in the same manner as above, but with the estimated vectors replaced by their true equivalents.

The square root of the ratio of the estimated to the true sum of squares, Q_f, is given by

$$Q_f = \sqrt{\hat{S}_f / S_f} \tag{13}$$

The closer this value is to unity, the better is the modelling of the factor f. The concentration of the reactant will decrease identically from sample to sample in each simulation. The relative concentration of the intermediate will increase with increasing d and the product will decrease. The total concentration of the reactant, intermediate and product at any one point will always be constant, but the sum of squares will not.

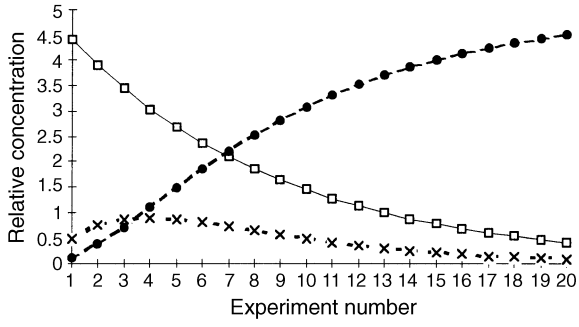

Fig. 4 Simulated degradation profiles. □, Reactant; ×, intermediate; and ●, product.

Regression

For each of the elution, spectral and degradation modes, the predicted data are regressed on to the real data. In each case a matrix $\boldsymbol{R}$ can be obtained, often called a rotation or transformation matrix. For example, for the elution data, if $\boldsymbol{A}$ is the true elution datum, $\hat{\boldsymbol{A}}$ is the predicted data and $\boldsymbol{R}_A$ is the rotation matrix, then

$$\hat{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{R}_A \tag{14}$$

$\boldsymbol{R}_A$ is found by the pseudo-inverse:

$$\boldsymbol{R}_A = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\hat{\boldsymbol{A}} \tag{15}$$

For a good model, each column and row of the rotation matrix contains only one value significantly greater than zero.

Using the rotation matrices obtained above and the rotation matrix, a 'predicted true' dataset can be obtained, denoted by a circle overscript, *e.g.* for the elution data

$$\mathring{\boldsymbol{A}} = \hat{\boldsymbol{A}}\boldsymbol{R}_{\bar{A}}^{-1} \tag{16}$$

Calculating a residual root mean sum of squares *RMSEP*($\boldsymbol{A}$) between the actual true and predicted true data, across each matrix, gives a further indication of the quality of regression:

$$\mathrm{RMSEP}(A) = \sqrt{\frac{\sum_{i=1}^{I}\sum_{f=1}^{F}(\mathring{a}_{i,f} - \hat{a}_{i,f})^2}{I \times F}} \tag{17}$$

Spectral characteristics and elution maxima

A simple measure of the success of the decomposition can obtained by comparing the predicted design parameters with those listed in Table 1. From the predicted elution data the position of the maximum of each component can be obtained as the maximum of each column of $\hat{\boldsymbol{A}}$. Similarly, the positions of the two absorbance maxima and their ratio can be computed from the estimated spectral data.

Results

Most methods for factor analysis depend first on determining the number of significant factors. This is particularly true when the aim is to model the entire dataset. The importance of detecting and modelling all significant components in two-way factor analysis has been discussed in the context of mid-infrared (MIR) spectrometry.[28,29] If a third significant factor is ignored, then the information from this compound is mixed with the other two compounds. In contrast, if a third factor is small it may become confused with the other two factors if a three-factor model is employed. PARAFAC depends crucially on a prior estimate of the number of significant factors as shown below. The following section reports the results assuming a three component mixture and the subsequent section a two component mixture.

Another important aspect is that the order in which the factors are extracted may differ according to how the algorithm is implemented, *e.g.*, the starting point of the iterations. This

means that, over a series of datasets, the first factor may correspond to physically different compounds in each run, so it is first necessary to reorder the factors according to presumed physical significance. In some cases, where the interpretation of each factor is in doubt, this can be difficult. In the tables, the factors are ordered according to the order in which they were extracted.

Three-factor Systems

The predicted maxima positions of the elution profiles, in terms of elution index for the three component, three-factor system, are given in Table 2. At the levels where *d*, the kinetic rate parameter, is low, *i.e.*, 1 and 3, the PARAFAC algorithm fails to position the three components correctly, whereas at higher levels of *d* all three components are correctly determined.

Table 2 lists the spectral parameters determined from the predicted data. In all cases the product (factor 3 at $d = 1$ and 3, factor 2 at $d = 5$ and 20 and factor 1 at $d = 10$) is predicted well, with a peak ratio of 1.65, and a low-wavelength absorbance maximum at 424 or 426 nm. There is a slight problem with predicting the high-wavelength absorbance maximum at higher values of *d*, presumably because the prediction ability decreases as the amount of intermediate increases. However, the high-wavelength absorption maximum is always ⩽660 nm. A 4 nm shift in position represents only two sampling points in the wavelength direction.

In this study, the data were designed with very similar spectral parameters and the PARAFAC algorithm has successively determined these, so it is trivial to establish the correspondence between components and factors. However, this will not always be true and in cases where there are several very similar components, a confident identification of the factors based on spectral parameters may not be possible.

The component sum of squares for the true and predicted data are given in Table 3. At all levels the size of the product is predicted remarkably well. The reactant and intermediate, however, are only closely estimated at the two higher levels of *d.* This can be understood by considering that the reactant and product had similar spectral characteristics. At the lower levels of *d*, the intermediate is relatively minor compared with the reactant, but as the intermediate increases in significance at higher levels of *d*, it is easier for the algorithm to distinguish between them.

The elution profiles for $d = 1$ and 20 are presented graphically in Fig. 5. It is obvious that the product and intermediate are not distinguished when *d* is low; these two

Table 2 Design and predicted spectral parameters for the three factor model.

Design	Reactant	Intermediate	Product
Spectral ratio	1.28	1.24	1.65
Absorbance max. 1/nm	434	434	424
Absorbance max. 2/nm	670	670	660
$d = 1$—	Factor 1	Factor 2	Factor 3
Spectrum ratio	1.29	1.28	1.65
Absorbance max. 1/nm	436	436	426
Absorbance max. 2/nm	670	670	660
$d = 3$—			
Spectrum ratio	1.30	1.27	1.65
Absorbance max. 1/nm	434	434	424
Absorbance max. 2/nm	670	670	660
$d = 5$—			
Spectrum ratio	1.28	1.65	1.24
Absorbance max. 1/nm	434	426	436
Absorbance max. 2/nm	670	660	670
$d = 10$—			
Spectrum ratio	1.65	1.28	1.24
Absorbance max. 1/nm	424	424	434
Absorbance max. 2/nm	658	668	668
$d = 20$—			
Spectrum ratio	1.24	1.65	1.28
Absorbance max. 1/nm	434	424	434
Absorbance max. 2/nm	668	656	668

Table 3 Size of each design component and factors for the two- and three-component systems

	Rate parameter, *d*				
	1	3	5	10	20
Reactant	9100	9100	9100	9100	9100
Intermediate	8	167	591	2570	7165
Product	14827	13416	11750	7637	3547
Three-component system—					
Factor 1	842	1825	8201	7640	6334
Factor 2	4741	3803	11751	8463	3547
Factor 3	14285	13412	572	2483	9461
Two-component system—					
Factor 1	9209	9872	11684	13111	5317
Factor 2	14312	13473	10795	7549	15294

Table 4 Prediction ratios, Q_f, for the two- and three-component models

	Ratio Q_f				
	Three-component model			Two-component model	
d	Reactant	Intermediate	Product	Reactant	Product
1	0.9999	10.2884	0.7218	1.0060	1.0009
3	0.9999	3.3054	0.6465	1.0416	1.0021
5	1.0000	0.9836	0.9493	0.9972	1.0892
10	1.0002	0.9829	0.9646	1.2003	0.9942
20	1.0197	0.9402	1.0001	n/a	n/a

Table 5 Root mean square error of prediction (RMSEP) for the three-factor models

	Data mode		
d	***A***	***B***	***C***
1	1.1×10^{-7}	4.19×10^{-7}	3.5×10^{-8}
3	4.2×10^{-8}	2.89×10^{-8}	2.3×10^{-8}
5	1.4×10^{-8}	2.89×10^{-7}	3.5×10^{-8}
10	3.7×10^{-8}	2.3×10^{-8}	3.6×10^{-8}
20	3.37×10^{-8}	3.55×10^{-7}	2.63×10^{-8}

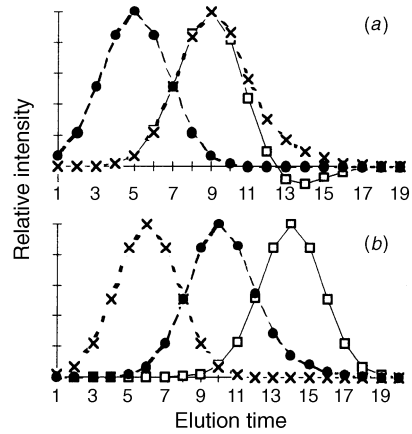

Fig. 5 Elution profiles obtained for the three factor models at (*a*) $d = 1$ and (*b*) $d = 20$.

species have similar spectral characteristics. Figs. 6(*a*) and (*b*) are representations of the corresponding spectra and it can be seen that they are recovered well.

It can be concluded that when the number of components is correctly known, the PARAFAC algorithm produces excellent decomposition results. These results are the best when all of the components are relatively significant, as shown by the square root of the ratios of predicted to true sum of squares, Q_f, given in Table 4, and the *RMSEP* in Table 5. For the elution data, ***A***, improves considerably from the $d = 1$ to the $d = 20$ level. There is also an improvement, but to a lesser extent, for the spectral data. The error in the kinetic profiles is reasonably constant at each level.

Table 6 Position of maxima of predicted elution profiles for the two factor model

	Elution maximum	
d	Factor 1	Factor 2
1	10	6
3	10	6
5	6	10
10	10	6
20	6	10

Two factor Systems

The PARAFAC algorithm was repeated on the datasets but with two rather than three factors used to model the data. The elution and spectral parameters found are given in Tables 6 and 7, respectively. As can be seen from Table 6, at each level of *d* the algorithm appears to detect successfully the reactant and product without any interference from the intermediate. Note that the product should elute at datapoint 6 and the reactant at datapoint 10.

Again, in Table 7, it appears that the two-component model produces good predictions of the spectrum ratios and absorbance maxima at each level of *d*, although the peak ratio for the product (1.56) is lower at $d = 20$.

However, when the sum of squares of the factors and components are computed (Table 3), the situation is not so straightforward. At the lower two levels of the intermediate the

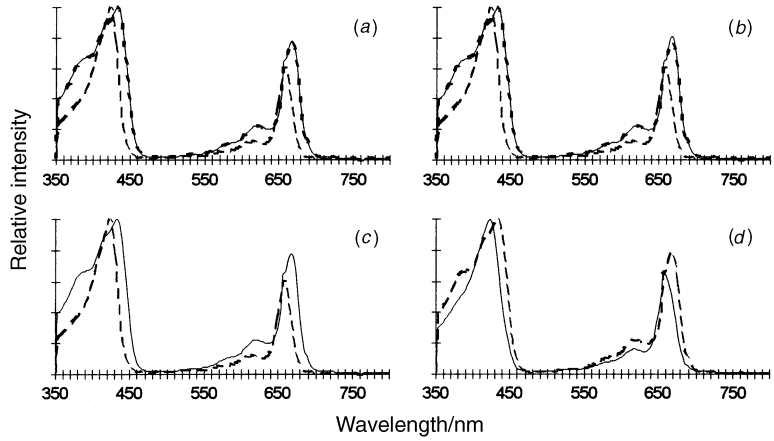

Fig. 6 Predicted spectra for (*a*) three factors at $d = 1$, (*b*) three factors at $d = 2$, (*c*) two factors at $d = 1$ and (*d*) two factors at $d = 20$.

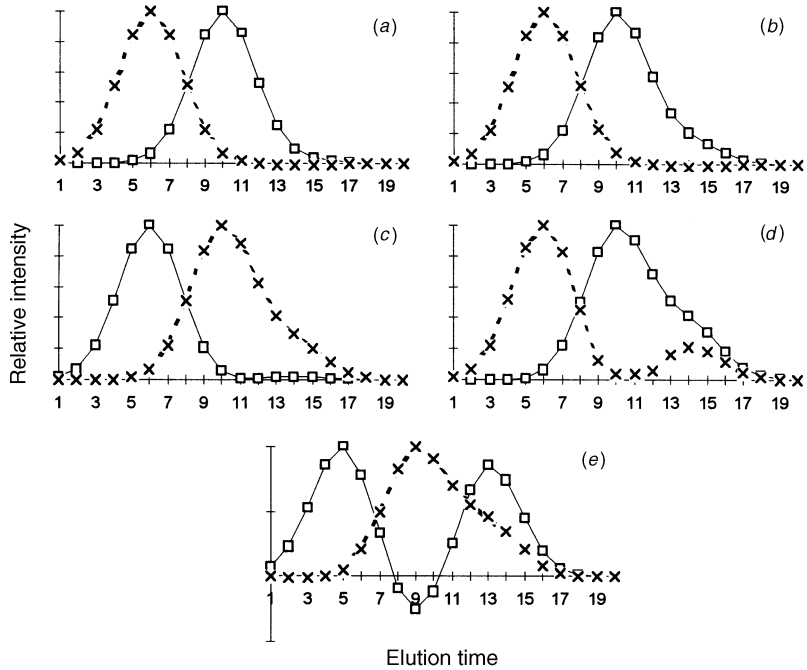

Fig. 7 Elution profiles obtained for the two factor models at (*a*) $d = 1$, (*b*) $d = 3$, (*c*) $d = 5$, (*d*) $d = 10$ and (*e*) $d = 20$.

factors predict the size of the component reasonably well, but at the higher levels it becomes more difficult to distinguish the factors. This result reinforces the observation from above that a univariate measure such as elution maximum is not a sophisticated measure of data quality and a multivariate method, utilising the data from all available modes, should always be used in preference. This is important as the PARAFAC algorithm distributes all of the observed systematic variance between the factors in the model, so that these are not necessarily pure factors.

Unlike the three-factor model above, predictions for the two factor model are better when the unmodelled component is relatively insignificant. As can be seen in Table 4, as the level of the intermediate increases the quality of reconstruction of the spectrally similar reactant decreases, but this is only observed when a multivariate measure such as the sum of squares or the rotation matrices is used. The quality of the product also falls but less significantly. Because at the $d = 20$ level a confident determination of the identification of the reactant and product cannot be made, the prediction ratio therefore cannot be calculated.

The five recovered elution profiles are shown in Fig. 7. These supplement the data in Table 6; it is obvious that for $d = 20$ the first factor has, in fact, two clear maxima. Interestingly, the intermediate is confused with the product and predicted as one factor, despite the difference in both spectral characteristics and elution profiles. This unexpected result can be explained in terms of kinetic profiles; the level of intermediate builds up rapidly and then decreases with time, and so the kinetics of the two compounds are fairly similar. Two components with identical kinetics but different spectra and elution profiles could be modelled as a single factor. Visual inspection of the predicted chromatograms in Fig. 7(*c*)–(*e*) should provide clues that the number of predicted components is too few, and so lead to rerunning the model including further components. The spectra, Fig. 6(*c*) and (*d*), are recovered well again.

Conclusions

PARAFAC is a powerful approach for resolving out series for two-way chromatograms recorded over a number of samples. The methods can be extended to three-way or higher data, *e.g.*, chromatograms could be recorded at different pH values and times; the change in chromatography with pH complements the change in intensity with time.

Table 7 Predicted spectral parameters for the two factor model. The design parameters are given in Table 1

$d = 1$—	Factor 1	Factor2
Spectrum ratio	1.28	1.65
Absorbance max. 1/nm	436	426
Absorbance max. 2/nm	670	660
$d = 3$—		
Spectrum ratio	1.28	1.65
Absorbance max. 1/nm	436	426
Absorbance max. 2/nm	670	660
$d = 5$—		
Spectrum ratio	1.65	1.28
Absorbance max. 1/nm	426	436
Absorbance max. 2/nm	660	670
$d = 10$—		
Spectrum ratio	1.27	1.65
Absorbance max. 1/nm	436	426
Absorbance max. 2/nm	670	660
$d = 20$—		
Spectrum ratio	1.56	1.27
Absorbance max. 1/nm	426	436
Absorbance max. 2/nm	660	670

The dataset in this paper is demanding, with the following properties. The middle chromatographic peak has no composition 1 or selective region, and most factor analysis methods find it difficult to resolve out unselective peaks. The spectra of the reactant and intermediate are almost identical, with similar spectra ratios and absorbance maxima, and also partially co-elute. Approaches such as windows factor analysis and evolutionary factor analysis will not resolve out neighbouring peaks with very similar spectra; these will simply be modelled by one principal component. Even two-dimensional peak purity methods such as derivatives depend on change in spectral composition over elution time and simply would not detect a difference between the reactant and intermediate. By using PARAFAC on a series of chromatograms, these peaks can be distinguished provided that the concentration of the intermediate is not too low. Hence PARAFAC has potential as a major technique for the resolution and quantification of a series of two-way of chromatograms, often in cases where normal factor analysis methods will fail.

The major drawback is that a good estimate of the number of components is required in advance for sensible models. If this is unknown, it is better to perform the models with fewer components first to see whether there are any elution profiles with more than one maximum. If so, the algorithm can be repeated, increasing the number of components until unimodal elution profiles are achieved.

The authors thank R. Bro for providing the Matlab PARAFAC algorithm and EPSRC for providing financial support for this project.

Appendix

List of Notation Used

i	Elution time index
I	Total number of elution points (20)
j	Spectral wavelength index
J	Number of points in each spectrum (226)
k	Sample number
K	Number of reaction times (20)
$\boldsymbol{A}$	Elution data matrix, with individual point $a_{l,i}$
$\boldsymbol{B}$	Spectra data matrix, with individual point $b_{l,j}$
$\boldsymbol{C}$	Concentration data matrix, with individual point $c_{l,k}$
l	Component number
L	Number of components
d	Kinetic rate parameter
$\underline{\boldsymbol{X}}$	Three-way data, with individual point $x_{i,j,k}$
f	Factor number
F	Total number of factors

Three-way matrices are represented by underlined upper-case bold italic characters, *e.g.*, $\underline{\boldsymbol{X}}$, two-way matrices by upper-case bold italic characters, *e.g.*, $\boldsymbol{A}$, vectors by lower case bold italic characters, *e.g.*, $\boldsymbol{a}_f$, and scalars by non-bold characters. Estimated variables are denoted by a 'hat,' *e.g.*, $\hat{\boldsymbol{A}}$, except in eqns. (16) and (17), where the 'estimated true' data are represented by a circle overscript. Dimensions of matrices are given as left-hand side subscripts, *e.g.*, ${}_{20,10}\boldsymbol{A}$ is a matrix of 20 rows by 10 columns.

References

1 Geladi, P., *Chemom. Intell. Lab. Syst.*, 1989, **7**, 11.
2 Ståhle, L., *Chemom. Intell. Lab. Syst.*, 1989, **7**, 95.
3 Mitchell, B. C., and Burdick, D. S., *Chemom. Intell. Lab. Syst.*, 1993, **20**, 149.

4 Wold, S., Geladi, P., Esbensen, K., and Öhman, J., *J. Chemom.*, 1987, **1**, 41.
5 Kvalheim, O. M., and Grung, B., *Chemom. Intell. Lab. Syst.*, 1995, **29**, 213.
6 Smilde, A. K., and Doornbos, D. A., *J. Chemom.*, 1991, **5**, 345.
7 Tauler, R., Smilde, A. K., Kowalski, B., *J. Chemom.*, 1995, **9**, 31.
8 Cattell, R., *Psychol. Bull.*, 1952, **49**, 499.
9 Kruskal, J. B., *Psychometrika*, 1976, **41**, 281.
10 Sands, R., and Young, F., *Psychometrika*, 1980, **45**, 39.
11 Kroonberg, P. M., and de Leeuw, J., *Psychometrika*, 1980, **45,** 69.
12 Tucker, L., in *Problems of Measuring Change*, ed. Harris, C., University of Wisconsin Press, Madison, WI, 1963, p. 122.
13 Burdick, D. S., *Chemom. Intell. Lab. Syst.*, 1995, **28**, 229.
14 Harshman, R. A., and Lundy, M. E., *Comput. Stat. Data Anal.*, 1994, **18,** 39.
15 Smilde, A. G., *Chemom. Intell. Lab. Syst.*, 1992, **15**, 143.
16 Henrion, R., *Chemom. Intell. Lab. Syst.*, 1994, **25**, 1.
17 Bro, R., and Heimdal, H., *Chemom. Intell. Lab. Syst.*, 1996, **34**, 85.
18 Smilde, A. K., *J. Chemom.*, 1992, **6**, 11.
19 Smilde, A. K., Van der Graaf, P. H., Doornbos, D. A., Steerneman, T., and Sleurink, A., *Anal. Chim. Acta*, 1990, **235**, 41.
20 Bro, R., *J. Chemom.*, 1996, **10**, 47.
21 Booksh, K. S., Muroski, A. R., and Myrick, M. L., *Anal. Chem.*, 1996, **68**, 3539.
22 Bro, R., *Chemom. Intell. Lab. Syst.*, 1996, **34**, 85.
23 Smilde, A. K., Tauler, R., Henshaw, J. M., Burgess, L. W., and Kowalski, B. R., *Anal. Chem.*, 1994, **66**, 3345.
24 Brereton, R. G., Rahmani, A., Liang, Y., Z., and Kvalheim, O. M., *Photochem. Photobiol*., 1993, **57**, 1048.
25 Elbergali, A. K., Brereton, R. G., and Rahmani A., *Analyst*, 1995, **120**, 2207.
26 Harshman, R. A., and Lundy, M. E., in *Research Methods for Multimode Data Analysis*, ed. Law, H. G., Snyder, C. W., Hattie, J. A., and McDonald, R. P., Praeger, New York, 1984, p. 216.
27 Ten Berge, J. M. F., in *Multiway Data Analyses*, ed. Coppi, R., and Bolasco, S., Elsevier, Amsterdam, 1989, p. 53.
28 Gurden, S. P., Brereton. R. G., and Groves, J. A., *Analyst*, 1996, **121**, 441.
29 Gurden, S. P., Brereton, R. G., and Groves, J. A., *Chemom. Intell. Lab. Syst.*, 1994, **23**, 123.

Paper 7/02232H
Received April 2, 1997
Accepted May 12, 1997