Chemometrics and Intelligent Laboratory Systems, 20 (1993) 149–161
Elsevier Science Publishers B.V., Amsterdam

An empirical comparison of resolution methods for three-way arrays

Ben C. Mitchell and Donald S. Burdick
Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708 (USA)

(Received 1 December 1992; accepted 10 May 1993)

Abstract

Mitchell, B.C. and Burdick, D.S., 1993. An empirical comparison of resolution methods for three-way arrays. *Chemometrics and Intelligent Laboratory Systems*, 20: 149–161.

In chemometrics applications it is common to resolve a trilinear array by solving a generalized eigenvalue problem, rather than by employing the iterative PARAFAC algorithm commonly used by psychometricians. Although an eigenanalysis-based procedure works perfectly in the absence of noise, it is not guaranteed to yield least squares resolutions when noise is present. The PARAFAC algorithm on the other hand is guaranteed to reduce the residual sum of squares at each iteration.

In this paper we propose synthesizing the two methods by using the resolution generated by eigenanalysis as starting values for the iterative PARAFAC algorithm. We find for simulated four-component data at moderate noise levels that following an eigenanalysis resolution with PARAFAC frequently leads to significant improvement in the quality of the resolution.

INTRODUCTION

There are two main approaches to the trilinear analysis of three-way data arrays. One approach is to resolve the array by performing a generalized eigenanalysis. Although they vary in details, several procedures based on this approach have been proposed in the chemometrics literature [1–4]. We will call such procedures EBPs for eigenanalysis-based procedures.

The other main approach utilizes alternating least squares in an iterative procedure which exploits the conditional linearity of the trilinear model. Its iterative nature means that starting values are required, but it is guaranteed to improve the least squares fit of the model to the data at each iteration. This approach is the one commonly used by psychometricians working in three-mode factor analysis [5,6]. Its prototype is the PARAFAC algorithm developed and popularized by Harshman. Chemists may be more familiar with this approach from the paper by Appellof and Davidson [7]. We will use the term PARAFAC for procedures following this approach because of its widespread use in the field of psychometrics.

Correspondence to: D.S. Burdick, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708 (USA).

In the absence of noise or when noise is low, both approaches work well [8,9]. If the noise is too high, on the other hand, neither method works very well: PARAFAC is sensitive to starting values and may encounter convergence problems, while an EBP resolution may be unattainable because of complex eigenvalues. Between these extremes, when the noise level is moderate, we can ask which of these approaches is better or if a synthesis is possible which combines features of both yielding a procedure superior to either. In this paper we undertake a simulation study in an attempt to answer this question.

To facilitate the investigation we have created a collection of data arrays briefly described below. The Section 'Procedures' reviews both PARAFAC and EBPs. Uncorrected correlation coefficients introduced in the Section 'Convergence and quality assessment criteria' will provide the basis for both assessing the quality of resolved factor matrices and the convergence criterion for PARAFAC. The Section 'Are EBPs enough?' addresses the main issue of this article: Are EBPs enough, or should resolutions obtained by an EBP be followed up with PARAFAC?

The empirical investigation utilizes 64 differing simulated three-way $I \times J \times K$ data arrays. Each array A has the trilinear form

$$A = S + E$$

where the ijkth element of S, which we refer to as the signal, is given by

$$s_{ijk} = \sum_{r=1}^{R} x_{ir} y_{jr} z_{kr} \qquad (1)$$

for $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$, and $k = 1, 2, \ldots, K$ and $\mathbf{E}$ is a commensurate array of discrepancies, or noise. An alternative notation for Eqn. 1 is the triple product [10] given by

$$\mathbf{S} = \otimes(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \qquad (2)$$

where $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are called the factor matrices comprising the signal. $\mathbf{X}$ is $I \times R$, $\mathbf{Y}$ is $J \times R$, and $\mathbf{Z}$ is $K \times R$. Each column of a factor matrix is referred to as a profile. So the second column of the factor matrix $\mathbf{Y}$ is called the second Mode $\mathbf{Y}$ profile.

The rank of S is defined to be the smallest integer R for which an expression in the form of Eqn. 1 holds. If each factor matrix in Eqn. 2 has R linearly independent columns, then R must be the rank of the signal.

For each signal S, a random array of noise was generated and factored in at the 25% level. For these particular arrays $\mathbf{X}$ is 40×4, $\mathbf{Y}$ is 40×4, and $\mathbf{Z}$ is 4×4. So A is a $40 \times 40 \times 4$ three-way array whose signal S has rank 4. The exact details concerning the construction of these arrays can be found in the Appendix.

A further distinctive characteristic of these arrays is the structure of the Mode $\mathbf{Z}$ profiles. An attempt has been made to mimic excitation emission frequency arrays encountered in phase-resolved fluorescence spectroscopy [11,12]. Theoretically, these profiles should be described by

$$z_{kr} = C_r \frac{\omega_k \tau_r}{1 + \omega_k^2 \tau_r^2}$$

where τ_r denotes the fluorescence lifetime for the rth component, ω_k denotes the kth modulated frequency, and C_r depends on, among other things, the concentration of the component [1].

PROCEDURES

The objective of the resolution procedures is to obtain an estimate of the factor matrices of the signal portion of the array. This is, of course, complicated by the presence of noise. In order to undertake this task, EBPs require two distinct $I \times J$ rank R matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ which in the absence of noise would be of the form

$$\mathbf{M}_1 = c_{11}\boldsymbol{x}_1 \otimes \boldsymbol{y}_1 + c_{12}\boldsymbol{x}_2 \otimes \boldsymbol{y}_2 + \cdots + c_{1R}\boldsymbol{x}_R \otimes \boldsymbol{y}_R$$

and

$$\mathbf{M}_2 = c_{21}\boldsymbol{x}_1 \otimes \boldsymbol{y}_1 + c_{22}\boldsymbol{x}_2 \otimes \boldsymbol{y}_2 + \cdots + c_{2R}\boldsymbol{x}_R \otimes \boldsymbol{y}_R$$

where the vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_R\}$ and $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_R\}$ are the linearly independent columns of the factor matrices $\mathbf{X}$ and $\mathbf{Y}$. These matrices can be written as

$$\mathbf{M}_1 = \mathbf{X}\mathbf{D}_1\mathbf{Y}' \text{ and } \mathbf{M}_2 = \mathbf{X}\mathbf{D}_2\mathbf{Y}' \qquad (3)$$

where the $\mathbf{D}_k$s are $R \times R$ diagonal matrices whose entries along the main diagonal equal z_{kr}. To obtain **X** and **Y** a generalized eigenanalysis is performed on $\mathbf{M}_1$ and $\mathbf{M}_2$ [1,4]. Given **X** and **Y**, **Z** can be readily found by linear least squares.

The two matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ arise as linear combinations of the Mode **Z** slabs of the array A. A Mode **Z** slab is obtained by fixing $k = k_0$, letting i and j vary, thereby obtaining

$$\mathbf{M}_{k_0} = [a_{ijk_0}]$$

There is nothing mathematically special about Mode **Z** slabs. But in phase-resolved fluorescence spectroscopy, the Mode **Z** direction is usually much smaller than the other two. It is therefore natural to use this direction when consolidating to two slabs.

PARAFAC exploits conditional linearity. Two of the factor matrices, say **X** and **Y**, are fixed and linear regression is used to obtain **Z**. Then **Y** and **Z** are held fixed and the same procedure is used to update **X**, and similarly for **Y**. This is done iteratively until the termination criterion is satisfied.

CONVERGENCE AND QUALITY ASSESSMENT CRITERIA

We begin by introducing the uncorrected correlation coefficient (UCC). UCCs will provide the basis for both the assessment of the resolution and the termination criterion for PARAFAC.

The uncorrected correlation coefficient for a pair of vectors is defined by

$$\mathrm{UCC}(x, y) = \frac{x \cdot y}{\sqrt{(x \cdot x)(y \cdot y)}} = \cos\theta \qquad (4)$$

where θ is the angle between the vectors x and y and $x \cdot y$ is the sum of the products of the corresponding elements of x and y. The UCC between a pair of arrays is obtained from Eqn. (4) by regarding the arrays as vectors. The UCC is a closeness measure for vectors that is unaffected by scalar multiplication.

The fit of the resolution to the data can be measured by first finding $\hat{A}$, the triple product of the resolved factor matrices $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$, and $\hat{\mathbf{Z}}$, and then calculating the UCC between $\hat{A}$ and the data array A. The resulting score is referred to as Ω.

$$\Omega = \mathrm{UCC}(\hat{A}, A)$$

where A is the data array and

$$\hat{A} = \otimes(\hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\mathbf{Z}})$$

We next define a criterion for comparing resolutions. Suppose we have the two sets of factor matrices each with R columns,

$$\{\mathbf{X}_A, \mathbf{Y}_A, \mathbf{Z}_A\} \text{ and } \{\mathbf{X}_B, \mathbf{Y}_B, \mathbf{Z}_B\}$$

The columns correspond to the components of the resolutions but the correspondence may not be the same for both sets. For example, the first columns of the matrices in factor set A may correspond to the third columns of the matrices in factor set B.

Define a matching to be one of the $R!$ possible ways to correspond the columns in factor set A with the columns in factor set B, and consider all possible matchings. Obtain a score for each matching by calculating the UCC between corresponding profiles, and averaging these $3R$ values. The criterion Υ is the highest of the $R!$ scores obtained by this process, i.e.

$$\Upsilon = \max[\mathrm{ave}(\mathrm{UCC})]$$

This procedure can be applied to any two sets of commensurate factor matrices. If one set is a resolution and the other the signal, Υ measures the quality of that resolution.

The criterion chosen for terminating PARAFAC, which we shall refer to as Δ, involves subtracting from 1 the value of Υ between the current resolution and the previous iteration:

$$\Delta = 1 - \Upsilon_{(n,n-1)}$$

A small value of Δ indicates that successive iterations are nearly identical.

The iterations are discontinued when Δ is less than some preset value Δ_0. An alternative that is often used is the scaled change in the residual sums of squares (RSS). This involves finding the RSS for both the current and previous resolution, taking the difference, and dividing the result by

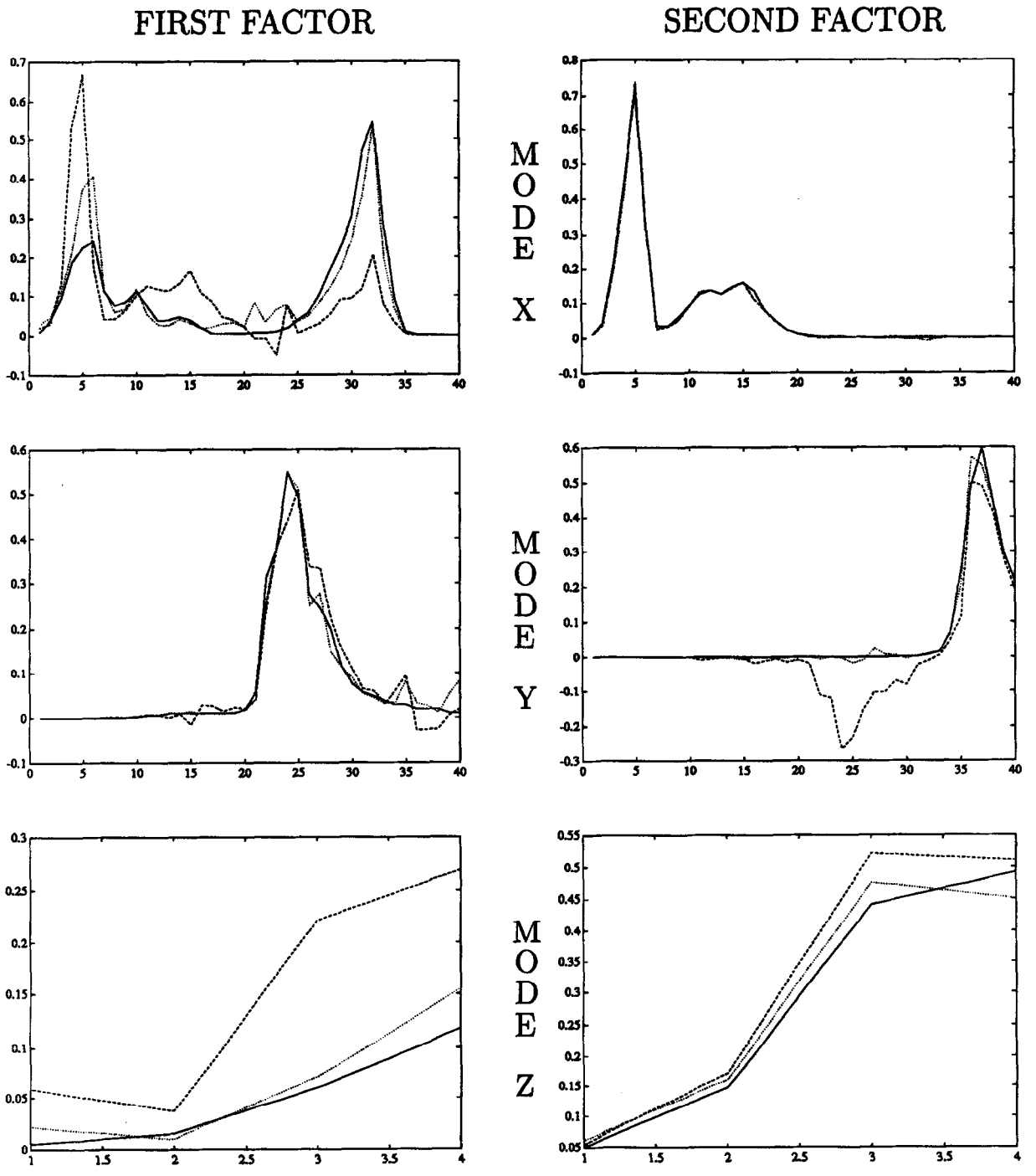

Fig. 1. The resolutions from the example obtained using LRA-1 (dashed curves) and PARAFAC (dotted curves), along with the signal (solid curves). The left column displays the profiles from the first factor while the right displays those from the second. The top row is Mode X, the middle Mode Y, and the bottom Mode Z.

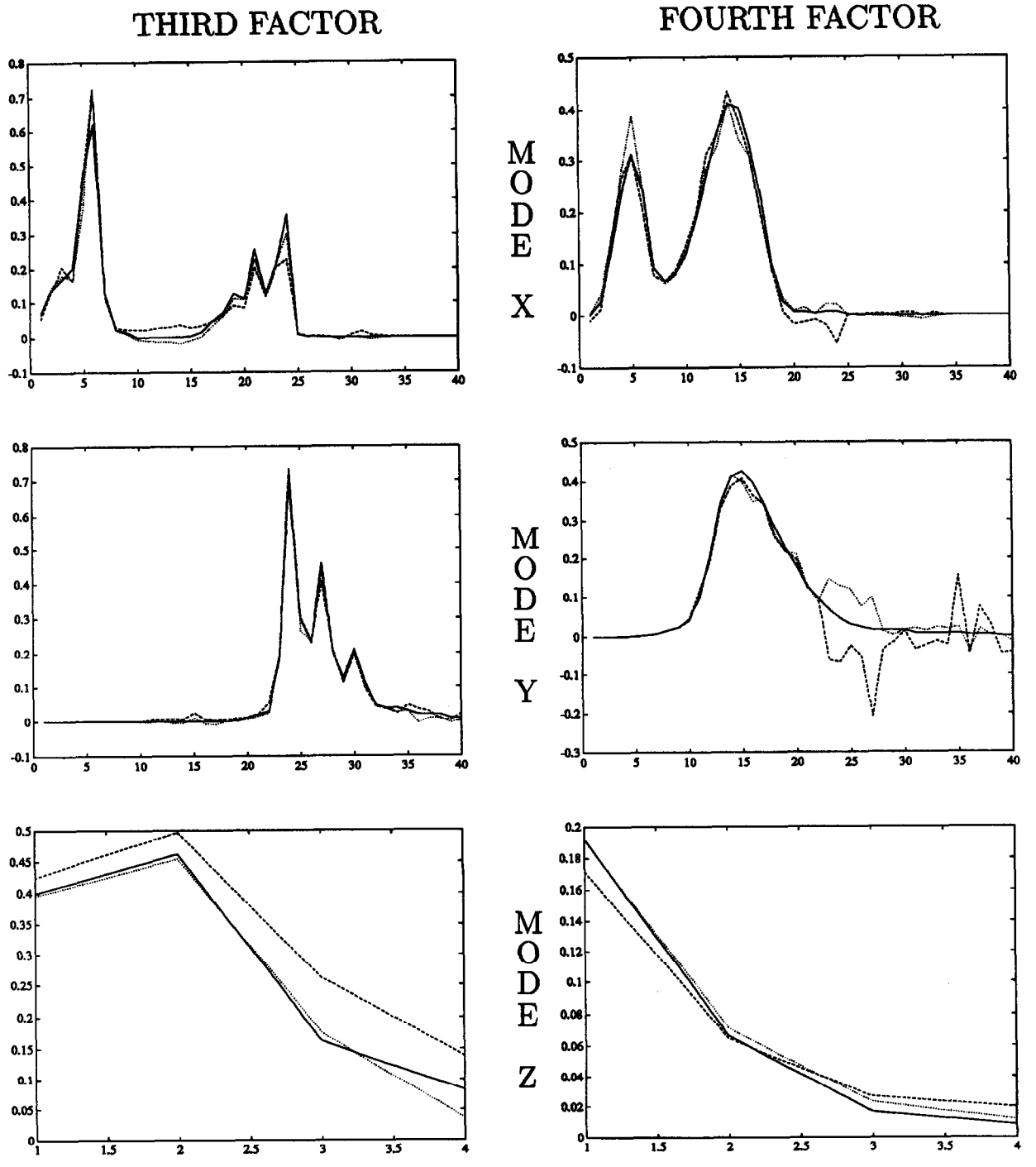

Fig. 2. The resolutions from the example obtained using LRA-1 (dashed curves) and PARAFAC (dotted curves), along with the signal (solid curves). The left column displays the profiles from the third factor while the right displays those from the fourth. The top row is Mode X, the middle Mode Y, and the bottom Mode Z.

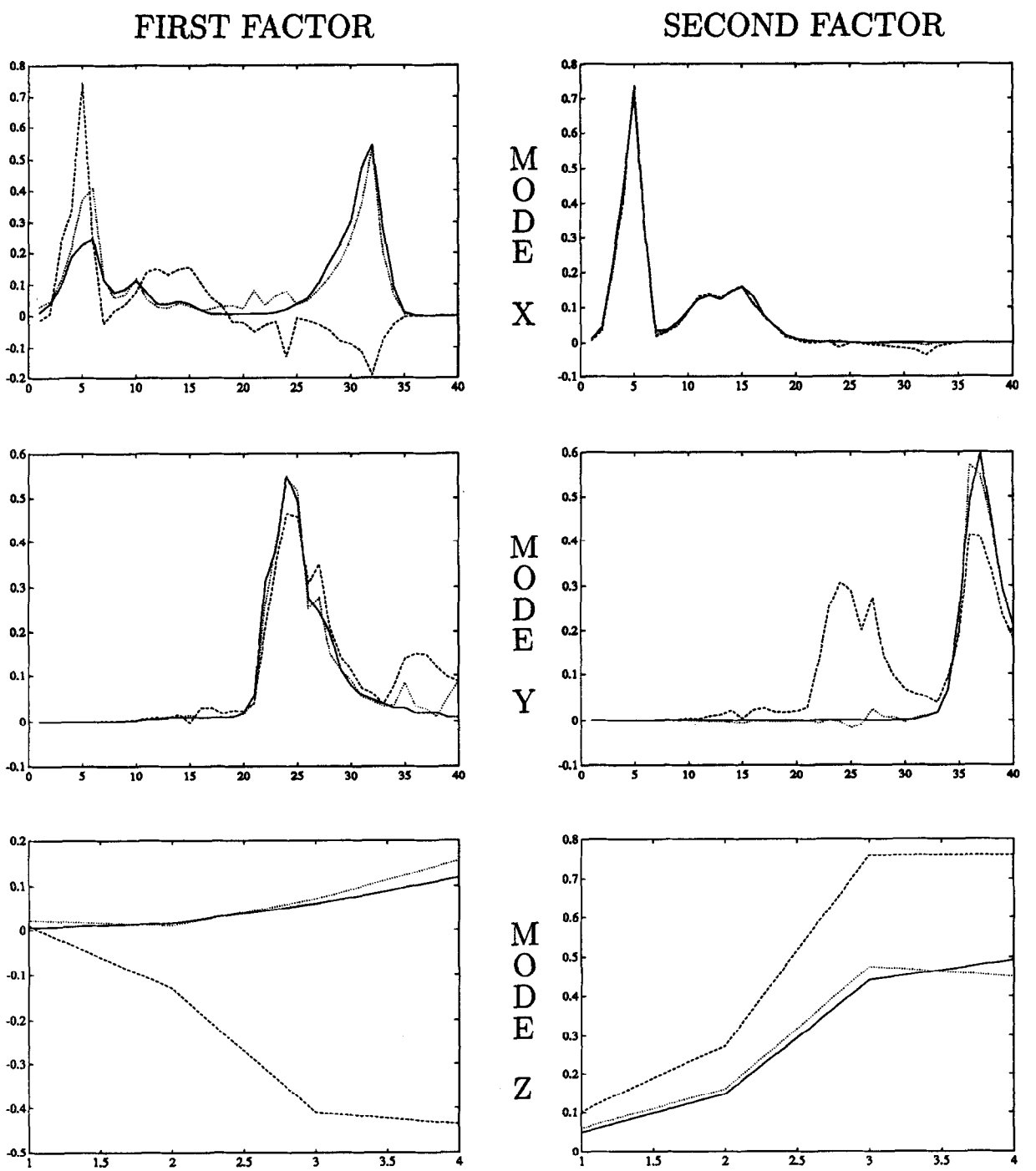

Fig. 3. The resolutions from the example obtained using LRA-4 (dashed curves) and PARAFAC (dotted curves), along with the signal (solid curves). The left column displays the profiles from the first factor while the right displays those from the second. The top row is Mode **X**, the middle Mode **Y**, and the bottom Mode **Z**.

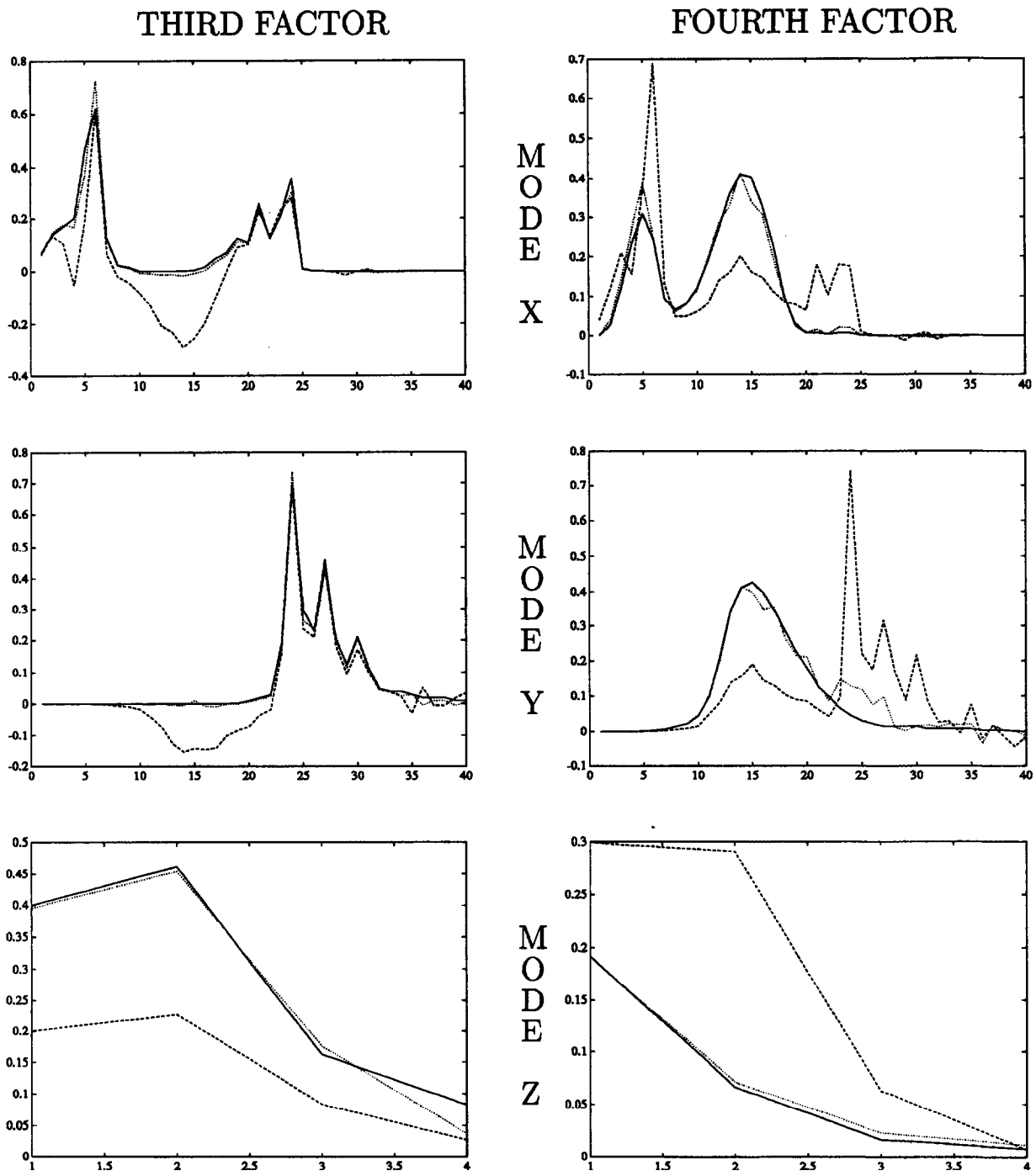

Fig. 4. The resolutions from the example obtained using LRA-4 (dashed curves) and PARAFAC (dotted curves), along with the signal (solid curves). The left column displays the profiles from the third factor while the right displays those from the fourth. The top row is Mode **X**, the middle Mode **Y**, and the bottom Mode **Z**.

the current RSS. Its behavior is similar to that of Δ but Δ requires roughly 1 to 2% of the computer time.

ARE EBPs ENOUGH?

Since PARAFAC is an iterative procedure, it must have starting values and may require many iterations to converge to a resolution. EBPs, however, are direct and relatively effortless. If the accuracy of the resolutions due to EBPs are sufficiently good, there is no need to look further. We can test this hypothesis by following an approach suggested by Sands and Young [13] in which the results of an EBP resolution are used as starting values for the iterative PARAFAC procedure.

If applying PARAFAC generally leads to substantially better resolutions, then perhaps the extra effort is warranted. The question to be addressed is: "Are EBPs enough?" In the context of our simulation the answer is "No". Before presenting the data which supports this answer, we illustrate the potential improvement from PARAFAC with the following example, taken from the simulation.

Example

Figs. 1–4 compare the resolutions obtained using an EBP and the EBP followed by PARAFAC on one of the data arrays. The EBP used is that suggested by Leurgans, Ross, and Abel (LRA) [3,4]. LRA suggest using a single Mode **Z** slab as $\mathbf{M}_1$ and the sum of all of the Mode **Z** slabs as $\mathbf{M}_2$. Figs. 1 and 2 display the results when the first Mode **Z** slab is used. We shall refer to this as LRA-1. Fig. 1 displays the first and second factors. The Mode **X** profiles can be seen in the top row, Mode **Y** in the middle row, and Mode **Z** in the bottom. Upon inspection we see that the first Mode **X** profile due to the EBP is not a very accurate reconstruction of the signal, displayed along with it. After PARAFACing, this profile, also displayed, becomes a much more accurate reconstruction. The same is true, to a lesser extent, of the second Mode **Y** profile. A quantitative corroboration of this observation can be found in Table 1, which lists the UCCs between resolution and signal for the various profiles.

TABLE 1

The UCCs between the indicated resolved profile from the figures and the corresponding signal

$\mathbf{X}_1$, $\mathbf{Y}_1$, and $\mathbf{Z}_1$ refer to the factor matrices obtained using LRA-1, similarly $\mathbf{X}_4$, $\mathbf{Y}_4$, and $\mathbf{Z}_4$ are from LRA-4. $\mathbf{X}_p$, $\mathbf{Y}_p$, and $\mathbf{Z}_p$ refer to the factor matrices obtained from PARAFAC using either of the LRA factor matrices as a start. $\mathbf{Z}_1$ was found by applying a single PARAFAC iteration to $\mathbf{X}_1$ and $\mathbf{Y}_1$, and similarly for $\mathbf{Z}_4$.

Factor matrix	Profile			
	1st	2nd	3rd	4th
$\mathbf{X}_1$	62.17	99.88	97.94	99.29
$\mathbf{Y}_1$	97.79	87.80	99.61	93.37
$\mathbf{Z}_1$	96.82	99.80	99.08	99.56
$\mathbf{X}_4$	11.51	99.69	74.75	72.24
$\mathbf{Y}_4$	94.57	76.10	92.30	45.89
$\mathbf{Z}_4$	−95.38	99.80	99.90	90.40
$\mathbf{X}_p$	95.21	99.82	98.72	99.21
$\mathbf{Y}_p$	99.09	99.39	99.70	98.07
$\mathbf{Z}_p$	99.40	99.65	99.72	99.91

In the table $\mathbf{X}_1$, $\mathbf{Y}_1$, and $\mathbf{Z}_1$ denote the factor matrices obtained using LRA-1, and similarly $\mathbf{X}_4$, $\mathbf{Y}_4$, and $\mathbf{Z}_4$ denote those obtained by LRA-4. The factor matrices returned by PARAFAC were the same for each start and denoted $\mathbf{X}_p$, $\mathbf{Y}_p$, and $\mathbf{Z}_p$. By inspection of the table, it can be seen that the UCC score for the first Mode **X** profile due to LRA-1 has been improved from 62.17 to 95.21% after PARAFACing. Similarly, the second Mode **Y** profile improved from 87.80 to 99.39%.

For this particular data array PARAFAC has markedly improved the accuracy of the reconstruction due to the EBP. This is evident by inspecting either the figures or the table. This example also shows that the selection of $\mathbf{M}_1$ and $\mathbf{M}_2$ can have a marked effect on the EBP resolution. For arrays with four Mode **Z** slabs, the four options for $\mathbf{M}_1$ allowed by LRA can produce different resolutions. Of the four in this example, the one displayed in Figs. 1 and 2 represents the most accurate reconstruction of the signal as measured by Υ (refer to Table 2). The least accurate is displayed in Figs. 3 and 4. In spite of

TABLE 2

Various values of Ω and Υ when LRA and PARAFAC are applied to the data array in the example

The subscripts used are: s, for signal; b, for beginning resolution prior to PARAFAC; and f, for final resolution subsequent to PARAFAC

EBP	Υ_{bs}	Υ_{fs}	Υ_{bf}	Ω_b	Ω_f	Ω_b/Ω_f
LRA-1	94.42	98.99	94.92	96.85	97.56	99.27
LRA-2	79.42	98.99	79.91	96.06	97.56	98.46
LRA-3	92.17	98.99	93.91	96.65	97.56	99.07
LRA-4	79.38	98.99	82.50	96.40	97.56	98.81

their differences taking either as the starting value for PARAFAC leads to the same resolution.

Various values of Υ and Ω for this example are listed in Table 2. The subscripts used are: s, for signal; b, for beginning resolution prior to PARAFAC; and f, for final resolution after applying PARAFAC. For example, Υ_{bs} is the value of Υ between the EBP factor matrices prior to PARAFAC and the signal matrices. The values in the first column under Υ_{bs} determine which EBP resolution was most accurate. Note that the values in the second column under Υ_{fs} are all the same because the PARAFAC algorithm converged to the same resolution for all four starting values.

We turn now from the examples to the full simulated data sets. An EBP requires two slabs or linear combinations of slabs as input for the eigenanalysis. Following LRA, we use the sum of the four Mode **Z** slabs and a single Mode **Z** slab as input to the eigenanalysis. Varying the selection of the single slab can yield as many as four starting values per data array for the PARAFAC algorithm. Occasionally the eigenanalysis produces complex roots, so that less than four starts may actually occur for some arrays. All 64 arrays did have at least one usable starting value. Of the 256 potential starts 32 were complex, so the analysis is based on the remaining 224 starts.

The results before and after PARAFAC for the 224 starts are compared in Fig. 5. The bar graph is that of the observed changes in Υ multiplied by 100.

$$\Delta\Upsilon = \Upsilon_{fs} - \Upsilon_{bs}$$

$\Delta\Upsilon$ is positive when PARAFAC improves the resolution. The predominance of positive values indicates that PARAFAC has a strong tendency to improve the resolutions due to EBPs. This

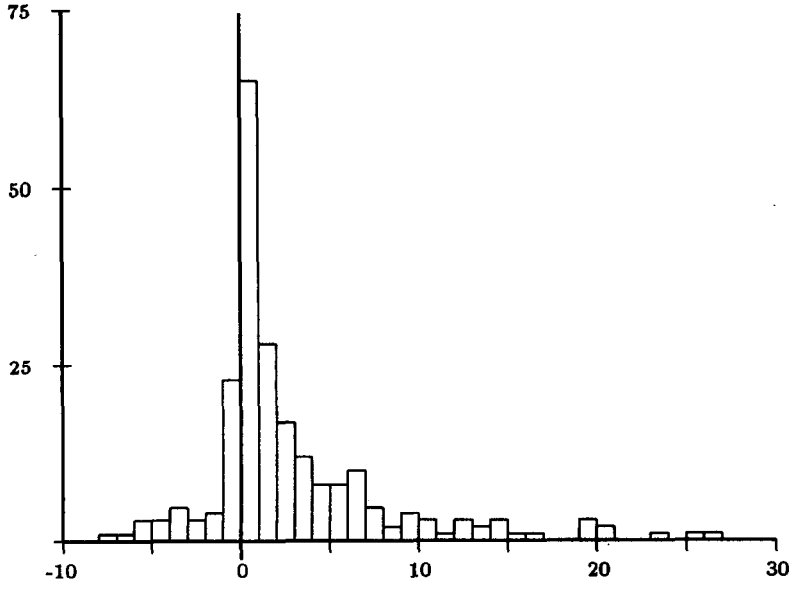

Fig. 5. A bar graph of observed frequencies of various values of $\Delta\Upsilon = \Upsilon_{fs} - \Upsilon_{bs}$ in percent.

conclusion is not limited to the LRA version of EBPs. Similar results were obtained when other EBPs were put through the same comparison.

Fig. 6 gives a statistical summary of the bf comparisons for the entire data set. A high value of Υ_{bf} means that the EBP resolution was close to the final resolution after applying PARAFAC. In other words if Υ_{bf} is high little movement occurs during the PARAFAC iterations. The median value of Υ_{bf} was 97.855, the lower and upper quartiles were 93.785 and 99.195, and the minimum and maximum were 71.02 and 99.93.

The example can be used as a base for interpreting these numbers. The first EBP resolution (LRA-1) depicted in Figs. 1 and 2 shows the amount of movement associated with an Υ_{bf} of 94.92, which is above the lower quartile. The bad case (LRA-4) has an $\Upsilon_{bf} = 82.50$ at the fourth percentile (8 out of 224 cases were lower).

If we limit our attention to the criterion Ω, we may be lulled into the belief that PARAFAC does not make much difference. When PARAFAC is applied to an EBP resolution, the fit to the data measured by Ω generally does not reflect the dramatic changes that may be occurring in the resolution. Consider the last three columns of Table 2. The fourth and fifth columns display Ω_b, the fit to the data prior to PARAFAC, and Ω_f, the fit subsequent to PARAFAC. The sixth column compares the two by taking the ratio. The ratios suggest that only modest changes are taking place in the fit, while in the meantime, the Υ_{bf} values suggest that substantial changes are occurring in the factor matrices, particularly for LRA-2 and LRA-4. This is reinforced by findings from the other arrays. Fig. 6 plots various values of Υ_{bf} along with Ω_b/Ω_f.

In conclusion we find that Υ is much more sensitive to changes brought about by PARAFAC than is Ω. The evidence also suggests that applying PARAFAC as a follow-up to an EBP may be the best strategy for obtaining a reliable resolution. This is so even in instances where PARAFAC has minimal effect on the fit.

APPENDIX: THE DATA ARRAYS

We now describe in detail how the data arrays used in the empirical study and for the example were constructed. The excitation and emission profiles were constructed to resemble actual observed excitation and emission profiles found in the Appendix of ref. 14. The profiles used can be found in Tables 3 and 4. The frequency profiles were generated by specifying lifetimes and modulation frequencies, then applying Eqn. 5:

$$z_{kr} = \frac{\omega_k \tau_r}{1 + \omega_k^2 \tau_r^2} \tag{5}$$

To construct the 64 data arrays used in the investigation, four Mode **X** factor matrices, four Mode **Y** factor matrices, and four Mode **Z** factor matrices were constructed. These were combined in a full 4^3 factorial design to generate 64 signal arrays. The Mode **X** and Mode **Y** factor matrices were constructed using the profiles listed in Table 5. The four Mode **Z** factor matrices were con-

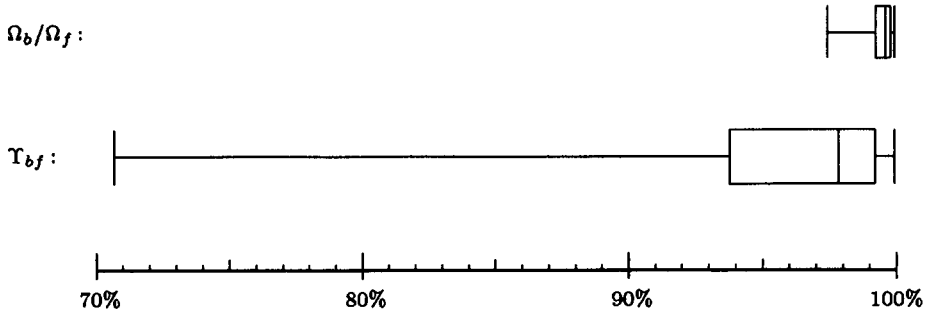

Fig. 6. A box plot comparing stand-alone EBP resolutions with EBP followed by PARAFAC. The upper plot is the ratios of the two fits Ω_b/Ω_f. The lower plot is of Ω_{bf}.

structed using the lifetimes and frequencies listed in Tables 6 and 7 and then applying Eqn. 5. As an example, the (2, 3) element of $\mathbf{Z}_2$ is given by

$$\mathbf{Z}_2(2, 3) = \frac{15 \times 0.100}{1 + (15 \times 0.100)^2}$$

and (4, 2) element of $\mathbf{Z}_4$ is

$$\mathbf{Z}_4(4, 2) = \frac{120 \times 0.070}{1 + (120 \times 0.070)^2}$$

The Mode **X** and Mode **Y** factor matrices were chosen somewhat arbitrarily; however, some effort was put forth to create factor matrices of varying degree of spectral overlap. It is evident from the lifetimes that this is indeed the case for the Mode **Z** factor matrices.

The three-digit variable ID is used to identify the combination taken. For instance, setting ID to 224, indicates the signal

$$S(224) = \otimes(\mathbf{X}_2, \mathbf{Y}_2, \mathbf{Z}_4)$$

TABLE 3

The Mode **X** signal profiles

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1	1	0	10	100	30	90	10	30	20	2	250
2	2	10	40	70	300	180	40	30	10	40	600
10	10	100	100	80	710	220	200	60	10	180	400
10	10	400	200	70	200	260	400	140	20	350	130
10	20	560	240	40	80	600	630	200	80	460	140
20	30	500	260	30	100	800	280	260	200	360	210
20	50	120	120	40	130	160	30	150	500	140	300
40	100	80	80	40	100	30	30	130	560	100	320
50	230	110	90	20	50	20	50	130	470	120	230
70	530	150	120	10	20	1	80	120	270	180	50
120	130	170	80	5	30	1	110	110	80	280	30
250	70	190	40	5	50	1	120	50	20	400	20
230	40	180	40	4	60	1	110	30	2	530	10
470	130	140	50	3	20	2	130	40	1	610	10
300	30	90	40	2	2	5	140	60	0	600	5
660	20	60	20	5	1	20	120	70	0	490	3
220	20	40	5	10	0	60	70	40	0	340	2
580	20	20	5	20	0	90	40	30	0	150	1
30	10	10	6	50	0	160	20	30	0	40	1
10	10	10	6	100	0	140	10	30	0	10	0
4	8	10	7	140	0	330	5	30	0	10	0
2	7	7	7	200	0	160	2	30	0	5	0
1	5	5	10	300	0	270	2	20	0	10	0
0	5	4	20	500	0	460	1	20	0	10	0
0	4	4	40	400	0	10	1	20	0	2	0
0	3	3	60	280	0	2	1	20	0	1	0
0	3	3	110	30	0	1	0	30	0	0	0
0	2	3	180	2	0	0	0	40	0	0	0
0	2	2	240	1	0	0	0	80	0	0	0
0	2	2	320	0	0	0	0	140	0	0	0
0	1	2	500	0	0	0	0	200	0	0	0
0	1	2	580	0	0	0	0	240	0	0	0
0	1	1	300	0	0	0	0	360	0	0	0
0	1	1	100	0	0	0	0	530	0	0	0
0	0	1	10	0	0	0	0	580	0	0	0
0	0	1	2	0	0	0	0	350	0	0	0
0	0	0	1	0	0	0	0	110	0	0	0
0	0	0	0	0	0	0	0	50	0	0	0
0	0	0	0	0	0	0	0	40	0	0	0
0	0	0	0	0	0	0	0	10	0	0	0

As another example, if ID is set to 312, then the signal indicated is

$$S(312) = \otimes(\mathbf{X}_3, \mathbf{Y}_1, \mathbf{Z}_2)$$

The Mode **X** and Mode **Y** profiles were normalized so that $\mathbf{X}'\mathbf{X}$ and $\mathbf{Y}'\mathbf{Y}$ had ones on the main diagonal and the Mode **Z** profiles were calculated from Eqn. 5. The data arrays are much to large to report explicitly. To reproduce them calculate S(ID) as indicated above, and generate noise arrays via the Matlab code:

```
rand('normal')
rand('seed',ID)
N=rand(2,2);
N=rand(40,160);
```

TABLE 4

The Mode Y signal profiles

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	2
0	0	0	0	0	1	0	0	0	0	0	5
0	0	0	0	0	2	0	0	0	0	0	10
0	0	0	0	0	10	0	0	0	0	0	20
0	1	0	0	0	20	0	0	0	1	0	30
0	2	0	0	0	20	0	0	0	2	0	60
1	5	0	0	0	10	0	0	0	40	0	130
2	5	0	0	0	20	0	0	0	400	0	280
10	10	0	0	0	110	0	0	0	460	0	460
10	10	0	0	0	510	0	0	0	550	1	550
10	10	1	0	0	550	0	0	0	350	2	570
10	10	2	0	0	490	0	1	0	250	5	530
200	10	5	0	0	310	0	2	0	150	10	460
640	10	10	0	0	190	1	10	0	80	30	380
320	10	20	0	0	100	2	50	0	50	60	300
710	20	40	0	0	60	10	120	0	30	160	240
270	60	100	0	0	40	20	230	0	20	280	170
360	330	180	0	1	20	30	380	0	10	390	130
200	400	280	0	2	10	200	530	0	5	500	90
110	580	370	0	80	5	730	620	0	2	550	60
100	520	420	0	270	4	310	650	0	1	550	40
60	290	440	0	490	4	240	560	0	0	540	30
40	260	450	1	460	3	480	540	0	0	510	20
30	210	440	2	350	3	210	500	0	0	460	20
20	120	430	10	250	2	130	520	0	0	380	20
30	80	400	20	200	2	220	430	1	0	330	20
40	60	380	50	160	2	120	340	2	0	270	10
30	50	370	200	110	1	50	270	10	0	220	10
20	40	350	470	70	1	40	220	20	0	180	10
20	30	330	560	50	1	40	180	80	0	140	10
10	30	300	430	40	1	30	150	300	0	110	10
10	20	280	300	30	0	20	120	600	0	80	5
10	202	250	230	20	0	20	90	730	0	70	5
2	20	230	190	10	0	20	70	550	0	60	5
1	10	200	160	2	0	10	60	360	0	50	2
0	10	180	120	0	0	10	50	260	0	40	1

TABLE 5

The Mode X and Mode Y factor matrices. Each profile should be normalized

Factor matrix	Mode X profiles				Factor matrix	Mode Y profiles			
$\mathbf{X}_1$	x_6	x_{10}	x_3	x_1	$\mathbf{Y}_1$	y_{10}	y_5	y_2	y_7
$\mathbf{X}_2$	x_3	x_2	x_{10}	x_5	$\mathbf{Y}_2$	y_2	y_9	y_7	y_{12}
$\mathbf{X}_3$	x_4	x_8	x_7	x_{11}	$\mathbf{Y}_3$	y_3	y_{12}	y_{10}	y_{11}
$\mathbf{X}_4$	x_9	x_{12}	x_{11}	x_3	$\mathbf{Y}_4$	y_6	y_9	y_4	y_8

TABLE 6

The lifetimes used to construct the four Mode Z factor matrices. The lifetime units are 10^{-6} s

T_1	0.001	0.010	0.100	1.000
T_2	0.005	0.010	0.100	0.500
T_3	0.010	0.005	0.050	0.100
T_4	0.050	0.070	0.200	0.500

TABLE 7

The frequencies in MHz used to construct each Mode Z profile

ω	5	15	60	120

Finally, the simulated data arrays are given by

$$A = S \odot (\mathbf{1} + 0.25N)$$

where $\mathbf{1}$ is an array of all ones and $\odot$ denotes term by term multiplication. The factor 0.25 is how we define 25% noise. The choice 0.25 was somewhat arbitrary, however, after preliminary investigation did appear to give rise to arrays that were neither too easy nor to difficult to resolve. Finally, actual discrepancies encountered in phase-resolved fluorescence spectroscopy are most likely a combination of additive and multiplicative noise. A fundamental assumption behind the PARAFAC routine is additive noise. As a result we thought it would be interesting to investigate the performance given multiplicative noise. One final note, the array used in the Example, which was one of the 64 simulated arrays, was generated by setting ID equal to 321.

REFERENCES

1 D.S. Burdick, X.M. Tu, L.B. McGown and D.W. Millican, Resolution of multicomponent fluorescent mixtures by analysis of the excitation–emission-frequency array, *Journal of Chemometrics*, 4 (1990) 15–28.

2 E. Sanchez and B.R. Kowalski, Tensorial resolution: a direct trilinear decomposition, *Journal of Chemometrics*, 4 (1990) 29–45.

3 S.E. Leurgans and R.T. Ross, Multilinear models: applications in spectroscopy, *Statistical Sciences*, 3 (1992) 289–319.

4 S.E. Leurgans, R.T. Ross and R.B. Abel, A decomposition for 3-way arrays, *SIAM Journal on Matrix Analysis and Application*, in press.

5 R.A. Harshman and M.E. Lundy, The PARAFAC model for three-way factor analysis and multidimensional scaling, in Law, Hattie and McDonald (Editors), *Research Methods for Multilinear Data Analysis*, Praeger, 1984, pp. 122–215.

6 R.A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an exploratory multi-mode factor analysis, *UCLA Working Papers in Phonetics*, 16 (1970) 1–84.

7 C.J. Appellof and E.R. Davidson, Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents, *Analytical Chemistry*, 53 (1981) 2053–2056.

8 D.W. Millican and L.B. McGown, Fluorescence lifetime selectivity in excitation–emission matrices for qualitative analysis of a two-component system, *Analytical Chemistry*, 61 (1989) 580–583.

9 D.W. Millican and L.B. McGown, Fluorescence lifetime resolution of spectra in the frequency domain using multiway analysis, *Analytical Chemistry*, 62 (1990) 2242–2247.

10 J.B. Kruskal, Rank, decomposition and uniqueness for 3-way and *N*-way arrays, in R. Coppi and S. Bolasco (Editors), *Multiway Data Analysis*, Elsevier, Amsterdam, 1989, pp. 7–18.

11 L.B. McGown and F.V. Bright, Phased-resolved fluorescence spectroscopy, *Analytical Chemistry*, 56 (1984) 1400–1408.

12 L.B. McGown and D.S. Kreiss, Spectral fingerprinting using phased-resolved fluorescence spectroscopy, *SPIE*, 910 (1988) 73–80.

13 R. Sands and F.W. Young, Component models for three-way data: an alternating least squares algorithm with optimal scaling features, *Psychometrika*, 45 (1980) 39–67.

14 J.N. Miller, *Standards in Fluorescence Spectrometry*, Chapman and Hall, New York, 1981.