ELSEVIER

Chemometrics and Intelligent Laboratory Systems 65 (2003) 35–49

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models

Jordi Riu [1], Rasmus Bro *

Department of Dairy and Food Science, Food Technology, The Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

Received 4 February 2002; accepted 14 August 2002

Abstract

In the last years, multi-way analysis has become increasingly important because it has proved to be a valuable tool, e.g. in interpreting data provided by instrumental methods that describe the multivariate and complex reality of a given problem. Parallel factor analysis (PARAFAC) is one of the most widely used multi-way models. Despite its usefulness in many applications, up to date there is no available tool in the literature to estimate the standard errors associated with the parameter estimates. In this study, we apply the so-called jack-knife technique to PARAFAC in order to find the associated standard errors to the parameter estimates from the PARAFAC model. The jack-knife technique is also shown to be useful for detecting outliers. An example of fluorescence data (emission/excitation landscapes) is used to show the applicability of the method.

Keywords: PARAFAC; Jack-knife; Errors; Multi-way; Resampling

1. Introduction

The current developments in instrumentation make it possible to obtain complex information that tries to adequately describe the multivariate reality of the problem under investigation. Such information should be analyzed according to the nature of both the data and the problem. During the last years, multi-way analysis has become increasingly important because it proved to be a valuable tool in interpreting some such complex data. Among the multi-way models, the most widely used in chemometrics are parallel factor analysis (PARAFAC) [1–3], Tucker3 [4–6] and multilinear partial least squares (*n*-PLS) [7]. PARAFAC, which is a generalization of principal component analysis (PCA) [8] to higher order arrays, has some very attractive features. One of them is the fact that there is no rotational indeterminacy in PARAFAC. For an application of PARAFAC to fluorescence emission/excitation data, this means that under mild assumptions, pure spectral profiles can be obtained provided that the multi-way spectral data follow the same structure as the PARAFAC model [9,10]. That is, the PARAFAC solution is unique. Despite its usefulness in some cases [11,12], to date there is no available tool in the literature to estimate the standard errors associated with the parameter estimates. In this

* Corresponding author. Tel.: +45-3528-3296; fax: +45-3528-3505.

E-mail address: rb@kvl.dk (R. Bro).

[1] On leave from: Department of Analytical and Organic Chemistry, Institute of Advanced Studies, Universitat Rovira i Virgili, Pl. Imperial Tarraco 1, 43005-Tarragona, Catalonia, Spain.

0169-7439/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0169-7439(02)00090-4

paper, we apply the so-called jack-knife technique to PARAFAC in order to find the associated standard errors in the parameter estimates from the PARAFAC model. The jack-knife technique is also shown to be useful for detecting outliers.

An example using the determination of four analytes from fluorescence data (emission/excitation landscapes) is used to show the usefulness of the method. In this application, the focus will be on outlier detection. Removing the samples (and variables) that do not stem from the same overall population as the bulk of the samples is mandatory for jack-knife standard error estimates to make sense. After removal of outliers, as discussed in this paper, the calculation of adequate standard error estimates is straightforward as will be shown.

2. Theory

2.1. PARAFAC model

PARAFAC [1–3] is a decomposition method that can be considered as one possible generalization of PCA to higher order arrays. For three-way data, PARAFAC decomposes the original data into trilinear components, each component consisting of one score vector and two loading vectors. For decomposition of fluorescence emission/excitation matrices, the scores correspond to the estimated relative concentration values and the loadings to the estimated pure emission and excitation profiles of each analyte in the fluorescence landscapes. A PARAFAC model for F components is defined by a score matrix **A**, and loading matrices **B** and **C** (also sometimes called first, second and third mode loadings, respectively) with respective elements a_{if} ($i=1\ldots I, f=1\ldots F$), b_{jf} ($j=1\ldots J$) and c_{kf} ($k=1\ldots K$). In this paper, **A** corresponds to the sample mode, **B** to the emission mode and **C** to the excitation mode. The model is found by minimizing the sum of squares of the residuals e_{ijk} in the model:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}, \qquad i = 1,\ldots,I; \quad j = 1,\ldots,J; \; k = 1,\ldots,K \tag{1}$$

where x_{ijk} is an element of the $I \times J \times K$ three-way array $\underline{\mathbf{X}}$. Unfolding or matricization [13] is a simple rearrangement of the three-way array into a two-way matrix by concatenating, for example, the horizontal slabs next to each other. The PARAFAC model can be expressed for the matricized array **X** (being **X** an $I \times JK$ matrix) [14]:

$$\mathbf{X} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})' + \mathbf{E} \tag{2}$$

where the sign $\odot$ stands for the Khatri–Rao product [2,15,16]. In the following, the first mode is assumed to be the sample mode and modes two and three are variable modes. It is also assumed that a three-way model is sought. Extensions to other scenarios are straightforward.

It is evident from Eq. (1) that the PARAFAC model, while not having rotational freedom, does suffer from some indeterminacies. The components of the model can be permuted without affecting the model fit. For example, allowing component 1 to change place with component 3 will not change the fit of the model. Additional to this permutation indeterminacy, the model also has an intrinsic scaling indeterminacy. Any column-vector of **A**, **B** or **C** may be scaled by any nonzero scalar as long as the corresponding column in any of the remaining component matrices is multiplied with the inverse value. Thus, it holds that if $s_a s_b s_c = 1$, then

$$\sum_{f=1}^{F} a_{if} b_{jf} c_{kf} = \sum_{f=1}^{F} (s_a a_{if})(s_b b_{jf})(s_c c_{kf}) \tag{3}$$

2.2. Jack-knife estimation of standard errors

The word jack-knife was suggested by Tukey [17] for use in statistics to describe a general approach for testing hypotheses and calculating confidence intervals in situations where apparently no better methods can be used. Given a set of values ($y_1, y_2, \ldots, y_I$) all sampled from the same population, and V being the statistic calculated from the I y-values, the standard error associated with the V statistic can be estimated by:

$$s_V = \sqrt{\frac{\sum_{m=1}^{I} \left(\overline{V} - V_{-m}\right)^2}{(I-1)/I}} \tag{4}$$

where V_{-m} is the calculated statistic using all the I y-values except y_m and $\overline{V}$ is the average of the I V_{-m} results. For instance, V can be the score value corresponding to one of the F components in one sample calculated with a PARAFAC model using the I fluorescence landscapes corresponding to the I samples. Among others, the jack-knife technique has been used for estimating the standard error associated with the individual model parameters in multivariate partial least squares regression (PLS) [18]. Since the jack-knife procedure (and all resampling methods) assumes random sampling from some distribution, only independent entities can be left out. Samples can often be considered as independent, but variables as emission or excitation wavelengths in fluorescence landscapes are almost always intercorrelated and can hence not be directly used in a jack-knife procedure.

It is also possible to use the jack-knife technique by leaving more than one sample out each time [19], but this is not pursued here.

Applying the jack-knife technique to the PARAFAC model for a set of I samples, I PARAFAC submodels are obtained, i.e. I sets of $\mathbf{A}$ matrices from which, using Eq. (4), the standard errors in the F scores in the I samples can be calculated. Each of the I submodels (given by parameters $\mathbf{A}_{-m}$, $\mathbf{B}_{-m}$ and $\mathbf{C}_{-m}$, where the $-m$ superscript means that the model is obtained leaving out sample m) has an associated scaling and permutation indeterminacy as has previously been pointed out and is thus not directly comparable to the other submodels. Hence, these indeterminacies have to be removed before jack-knife estimations of standard errors can be calculated.

2.3. Permutation indeterminacy

All the submodels have an intrinsic indeterminacy with respect to permutation. For example, from one model (or submodel) to another one, component 1 may be exchanged with another component. This also holds for bilinear models, but as PCA adopts the standard convention of selecting the first component as the one with largest variance and so on, this problem is then easily solved. For PARAFAC, setting the order of the components in each model such that the second and third modes have maximum correlation could be a solution for solving the permutation problem. However, it is important to note that the correlation matrix reflects the similarity of the scaled and centred matrices. In the particular PARAFAC case, the second and third modes are usually scaled (normalized) but they are not often centred. So instead of the maximum correlation, we will look for what is called the congruence coefficient [20].

For the mth PARAFAC submodel, the congruence between two matrices with normalized columns can be computed in the following way. For the second mode, the $F \times F$ matrix $(\mathbf{B}_{\text{overall}})'(\mathbf{B}_{-m})$ and for the third mode, the $F \times F$ matrix $(\mathbf{C}_{\text{overall}})'(\mathbf{C}_{-m})$ provide the congruence coefficients between the components in two models: the overall model and the mth submodel. The closer the elements of these products are to one, the more similar are the associated columns. Hence, if element (1, 2) of $(\mathbf{B}_{\text{overall}})'(\mathbf{B}_{-m})$ equals one, then the first column of $(\mathbf{B}_{\text{overall}})$ is equal to the second column of $(\mathbf{B}_{-m})$. In order to find the best matches for all the columns, all the $F!$ different permutations are calculated between the F components of the mth submodel ([component$_1$ component$_2$ ... component$_F$]). For each permutation, $\mathbf{M}=(\mathbf{B}_{\text{overall}})'(\mathbf{B}_{-m,\text{reordered}})+(\mathbf{C}_{\text{overall}})'(\mathbf{C}_{-m,\text{reordered}})$ is computed. Note that the same ordering is used in the two modes. The permutation with highest trace of $\mathbf{M}$ (ideally the trace should be $2F$) is chosen as the best ordering. As the first mode is usually not normalized and because the dimensions of $\mathbf{A}_{\text{overall}}$ and $\mathbf{A}_{-m}$ are not the same, this mode is not taken into account when finding the optimal matching. In most of the cases, the permutation indeterminacy could be solved using only one mode (the second or the third one), i.e. $(\mathbf{B}_{\text{overall}})'(\mathbf{B}_{-m,\text{reordered}})$ or $(\mathbf{C}_{\text{overall}})'(\mathbf{C}_{-m,\text{reordered}})$. But with only one mode, problems might arise if the selected mode had two components with very similar profiles so that they might be confused.

2.4. Scaling indeterminacy

The PARAFAC models also have scaling indeterminacy: the scale or the sign of a score or loading vector may change if another vector of the same component changes accordingly. To remove the scaling indeterminacy in order to have comparable parameters in all the submodels, the components of each submodel are scaled to have maximum agreement

with the overall model. For the mth submodel, we seek scaling matrices (**S**) that satisfy:

$$\min\left(\left\|\mathbf{A}_{\text{overall}-m}-\mathbf{A}_{-m}\cdot\mathbf{S}_{-m,\mathbf{A}}\right\|^2+\left\|\mathbf{B}_{\text{overall}}-\mathbf{B}_{-m}\cdot\mathbf{S}_{-m,\mathbf{B}}\right\|^2+\left\|\mathbf{C}_{\text{overall}}-\mathbf{C}_{-m}\cdot\mathbf{S}_{-m,\mathbf{C}}\right\|^2\right) \quad (5)$$

where $\mathbf{A}_{\text{overall}-m}$ is formed by all the scores of the overall model but the mth row. The **S** matrices are diagonal and satisfy: $\mathbf{S}_{-m,\mathbf{A}}\cdot\mathbf{S}_{-m,\mathbf{B}}\cdot\mathbf{S}_{-m,\mathbf{C}}=\mathbf{I}$ so that each element of the array approximated by the PARAFAC submodel given by $\mathbf{A}_{-m}$, $\mathbf{B}_{-m}$ and $\mathbf{C}_{-m}$ is not modified because it is simply multiplied by one. When the second and third mode loadings are normalized, the diagonal elements in the optimal scaling matrices are usually close to one.

The **S** scaling parameters are determined by a simple line-search procedure as a function of several variables (Eq. (5)) using Matlab 6.1 for Microsoft Windows [21], where all **S** matrices are initially set equal to **I** [22].

2.5. Detecting outliers

As in all modelling procedures, outliers should be detected and possibly removed before the final model is determined. There are mainly two kinds of outliers: outliers that have a high sum of squared residuals to the model (which may indicate that the sample is badly described by the model) or outliers that despite not having a high sum of squared residuals have a high influence on the estimated model. Of course, a combination of these two types is also possible because a high-residual sample tends to influence the estimated model parameters.

In order to detect outliers, a new plot based on the jack-knifing results is suggested. It is called Resample Influence Plot (RIP). In RIP, the sum of squared residuals of the mth sample to the model calculated in the mth jack-knife iteration is shown on one axis versus the sum of squares of the difference between the loadings (second or third mode) obtained with each jack-knife segment and the overall PARAFAC

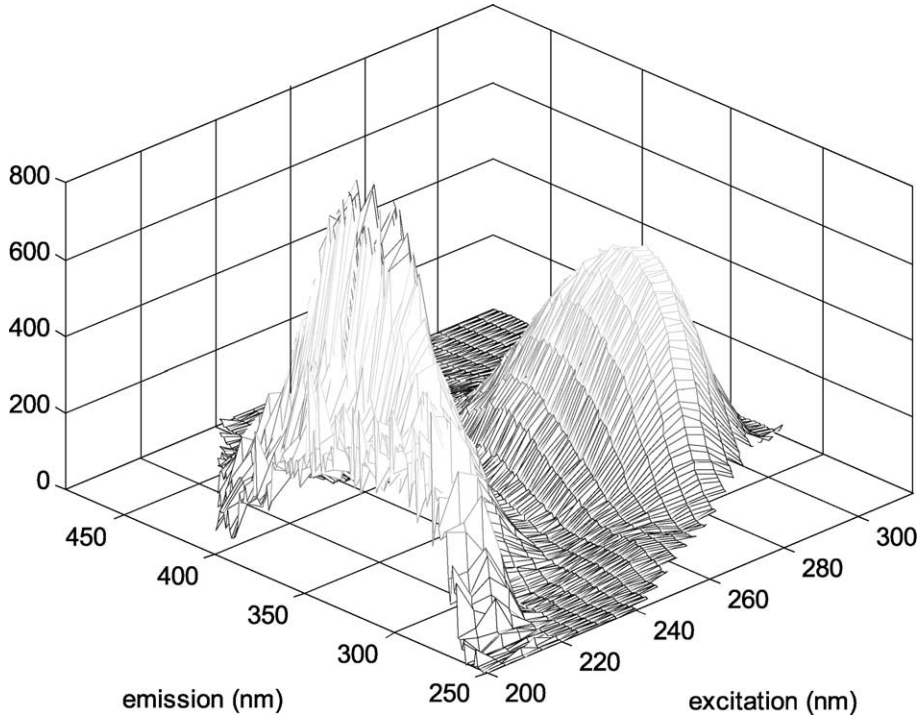

Fig. 1. Fluorescence landscape of one of the samples used in the text. The individual concentrations are 8×10^{-6} M for tryptophan, 20×10^{-6} M for dopa, 28×10^{-6} M for hydroquinone and 350×10^{-6} M for phenylalanine.

model. For instance, for the second mode and the mth jack-knife segment, the sum of squares of the difference between loadings is:

$$\sum_{f=1}^{F}\sum_{j=1}^{J}(b_{\text{overall},jf} - b_{-m,jf})^2 \quad (6)$$

It has to be pointed out that in order to handle possible different numbers of missing data in the samples, the sum of squared residuals will be scaled by the number of non-missing elements yielding the average squared residual per element. The RIP plot is viewed much in the same way as an influence plot in regression analysis [23] by looking for samples that have either high residuals and/or high differences between loadings calculated with or without the sample.

The RIP plot aids in understanding which samples differ the most from the overall population with respect to the PARAFAC model. The same plot can be obtained for the first (sample) mode, but it would provide little information because the differences among scores may be hidden due to differences in magnitude. In the same way, samples with high scores will tend to give higher differences from the scores of the overall model simply due to their scale.

In order to check for outliers specifically regarding to the score values, we will use another kind of plot called the Identity Match Plot (IMP), which is made by placing on one axis the scores obtained with the overall model, $\mathbf{A}_{\text{overall}}$, and on the other axis the scores predicted for the mth sample using the PARAFAC model obtained in the mth iteration (i.e. the model obtained leaving out the mth sample), $\mathbf{A}_{\text{predicted},m}$. The predicted scores for the mth sample are found using the following expression:

$$\mathbf{A}_{\text{predicted},m} = \mathbf{x}_m \cdot \left((\mathbf{C}_{-m} \odot \mathbf{B}_{-m})'\right)^{+} \quad (7)$$

where $\mathbf{x}_m$ corresponds to the vectorized experimental data for the mth sample, e.g. the mth fluorescence landscape, and the superscript '+' refers to the Moore–Penrose inverse. Possible missing data in the mth sample can be handled by using suitable PARAFAC algorithms [24]. In the IMP plot, one for each analyte, the outliers will be those samples that are far away from the ideal identity line.

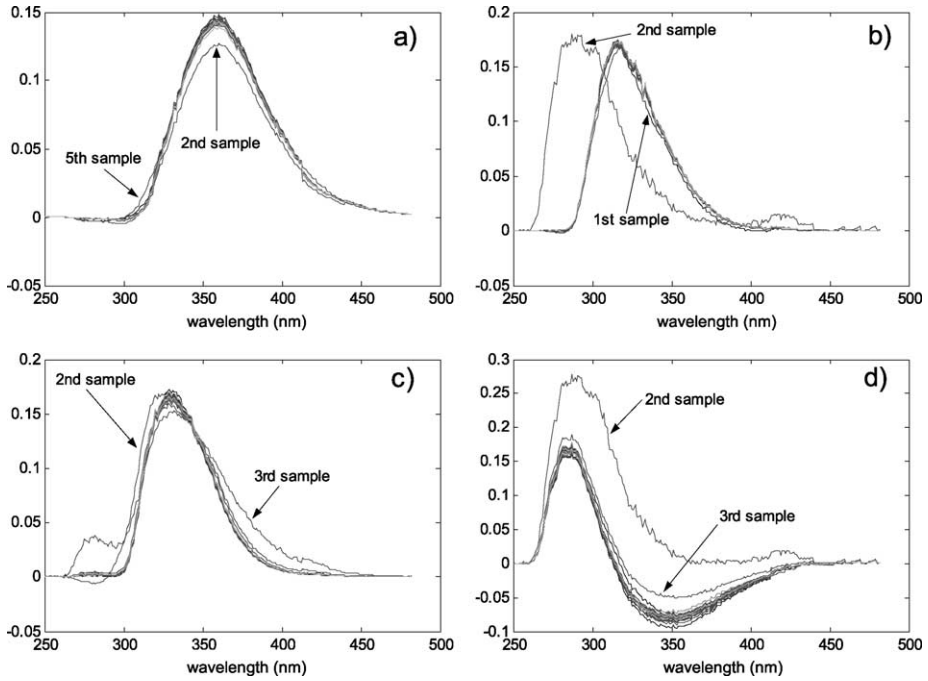

Fig. 2. Estimated pure emission spectral profiles of the 27 jack-knife segments for the four analytes in the first PARAFAC model. Corresponds to (a) tryptophan, (b) dopa, (c) hydroquinone and (d) phenylalanine.

3. Experimental part

The data set used consists of 27 fluorescence landscapes of 233 emission wavelengths (250–482 nm) and 24 excitation wavelengths (200–315 nm taken at 5 nm intervals) corresponding to 27 synthetic samples containing different concentrations of four analytes: hydroquinone, tryptophan, phenylalanine and dopa. A Perkin-Elmer LS50 B fluorescence spectrometer was used to measure the fluorescence landscapes. An example of a fluorescence landscape is shown in Fig. 1. More results can be found in Ref. [25].

4. Results and discussion

4.1. First model

With the application of jack-knife to the initial set of 27 samples, 27 PARAFAC submodels are obtained. Fig. 2 shows the estimated pure emission spectral profiles for the four analytes and Fig. 3 shows the pure excitation spectral profiles. In all the figures for all the models, the first component corresponds to tryptophan, the second component to dopa, the third component to hydroquinone and the fourth component to phenylalanine. There are 27 estimates of each of the pure spectra profiles, and ideally, they should all be the same. Looking at the figures, it is seen that the spectral profiles are unstable, especially the emission profiles. It is also seen that some spectral profiles have negative parts and that mainly jack-knife segments for the second and the third sample (i.e. when leaving out the second and the third sample) differ from the rest. Although it is difficult to visualize in the figure, the pure emission spectral profile for the first analyte (tryptophan) in the fifth sample also differs significantly from the robust average spectra of the rest. These results seem to indicate that the second and the third samples are outliers, and doubts remain about some other samples. Similar to jack-knife plots used in PLS regression [18], Fig. 4 shows the resampled estimates of the scores for analytes (components) 2 and 4: dopa and phenylalanine. It has to be pointed out that units are arbitrary and in order to find the correct concentration one needs to know the concen-

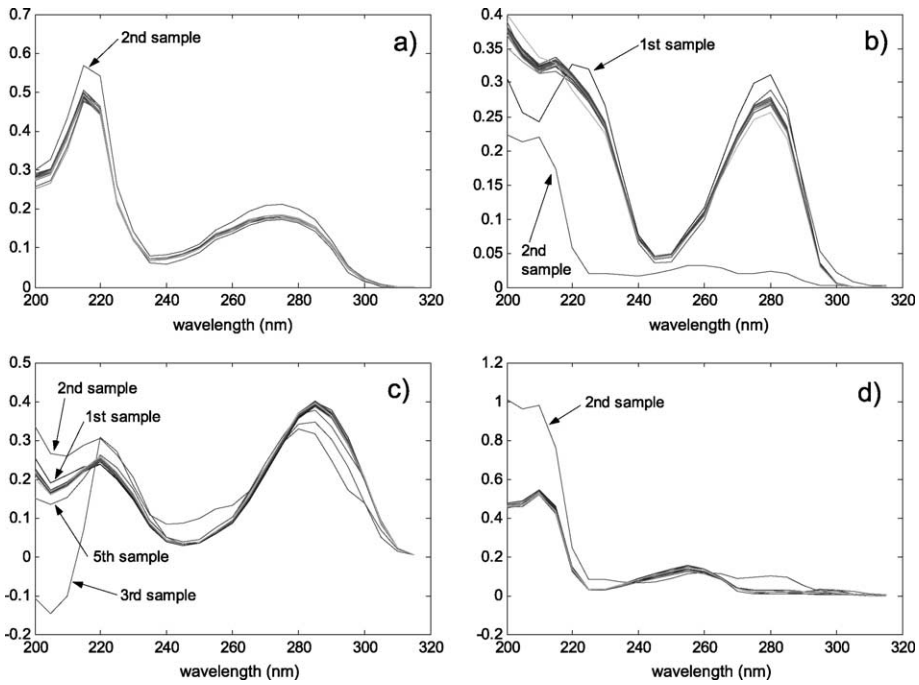

Fig. 3. Estimated pure excitation spectral profiles of the 27 jack-knife segments for the four analytes in the first PARAFAC model. Corresponds to (a) tryptophan, (b) dopa, (c) hydroquinone and (d) phenylalanine.

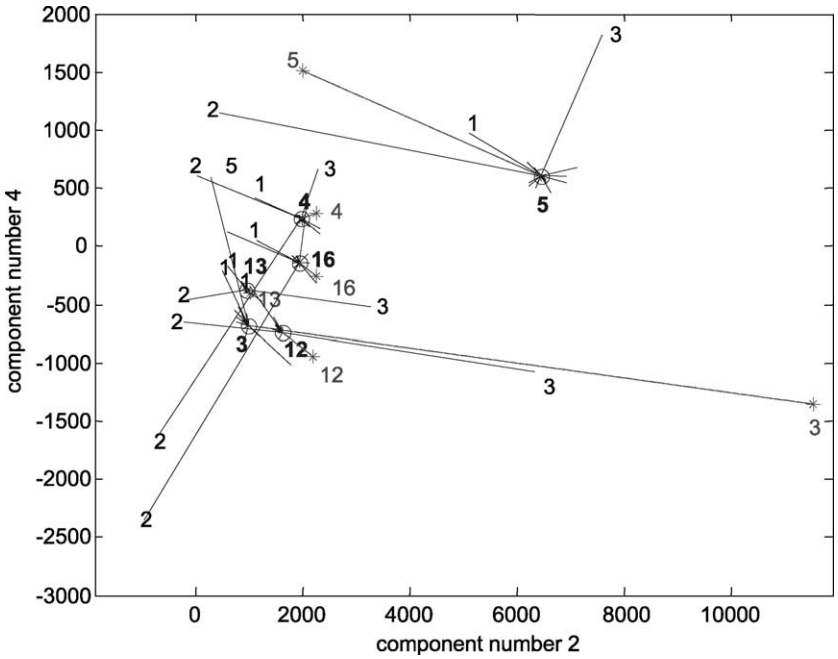

Fig. 4. Jack-knife estimation of the scores of six samples for the second (dopa) and fourth (phenylalanine) components in the first PARAFAC model. The circles represent the score values using the overall PARAFAC model; the asterisks correspond to the scores of the left-out samples calculated from the model with those samples left out; the bold figures indicate the sample number and the other numbers indicate the jack-knife segments.

tration value of at least one of the samples and then relate all the values to the known one. For the sake of clarity, only six samples are shown in Fig. 4, but the plot is representative of all the samples in all the components. For each of the six groups, the circles represent the score value found using the overall PARAFAC model (i.e. using the 27 samples). The asterisks (and the connecting dashed line) correspond to the score of the left-out sample calculated from the model with the sample left out. The 26 line segments for each group point to the corresponding score estimations when each of the 27 samples in turn was left out of the modelling. Only the lines corresponding to the first, second, third and fifth jack-knife submodels (the ones with highest deviations) are specified with a number. One can see that the deviations from the overall model when leaving out the second and the third sample are much higher than the rest in most of the cases. This can also be said for the first and the fifth samples, although the effect is not so clear. One can also see that the predicted scores for the third and fifth sample using the model with the third and fifth sample left out, respectively also differ considerably from the scores found using the overall model. Fig. 5 shows the RIP plots with the sum of squared residuals for each sample versus the sum-squared change in loadings. We can see that sample number 2 appears to be an outlier since its sum of squared residuals calculated using the model in the second jack-knife segment is the highest one. We can say the same about its sum of squares of the difference between loadings and hence it appears in the upper-right corner of the plot. Sample number 3 shows a similar behaviour, although not as dramatic. And although it is difficult to see in the figure, the sum of squared residuals for the fifth sample is almost twice than for the rest of the samples. Fig. 6 shows the IMP plots with the score values found using the overall model and the scores predicted for the *m*th sample using the model obtained in the *m*th jack-knife segment (Eq. (7)). Again, samples 2 and 3 show the most distinctive behaviour. Hence, it was decided to

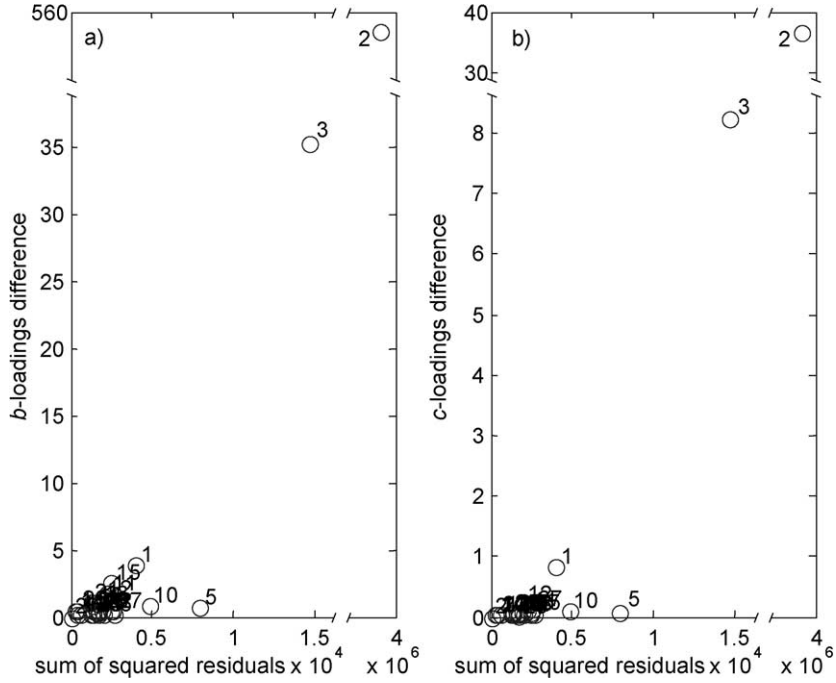

Fig. 5. RIP plots for (a) *b*-loadings and (b) *c*-loadings in the first PARAFAC model.

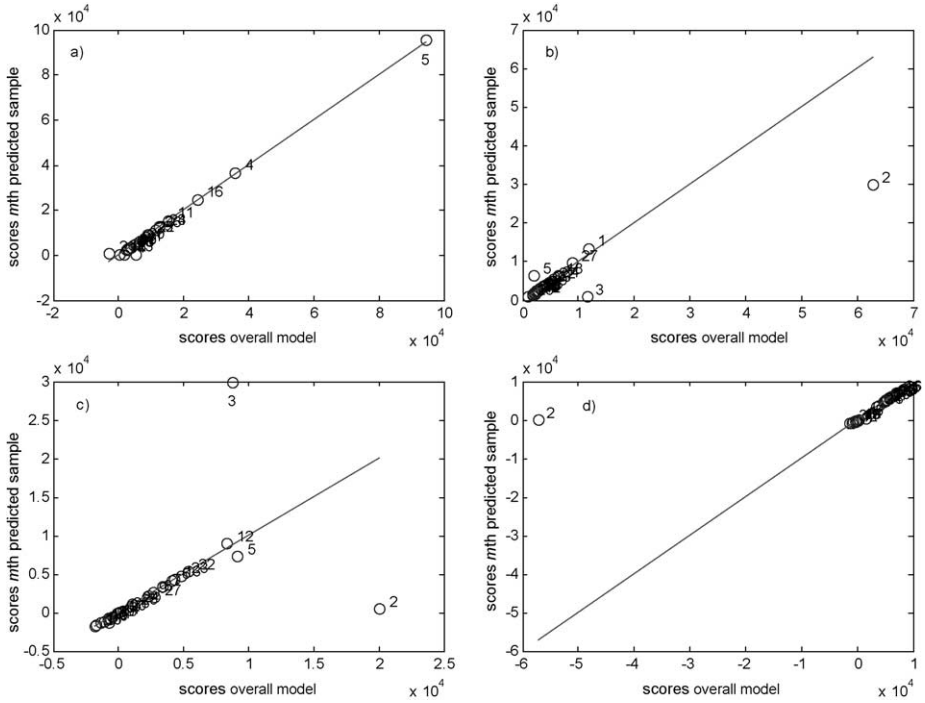

Fig. 6. IMP plots for the four analytes in the first PARAFAC model.

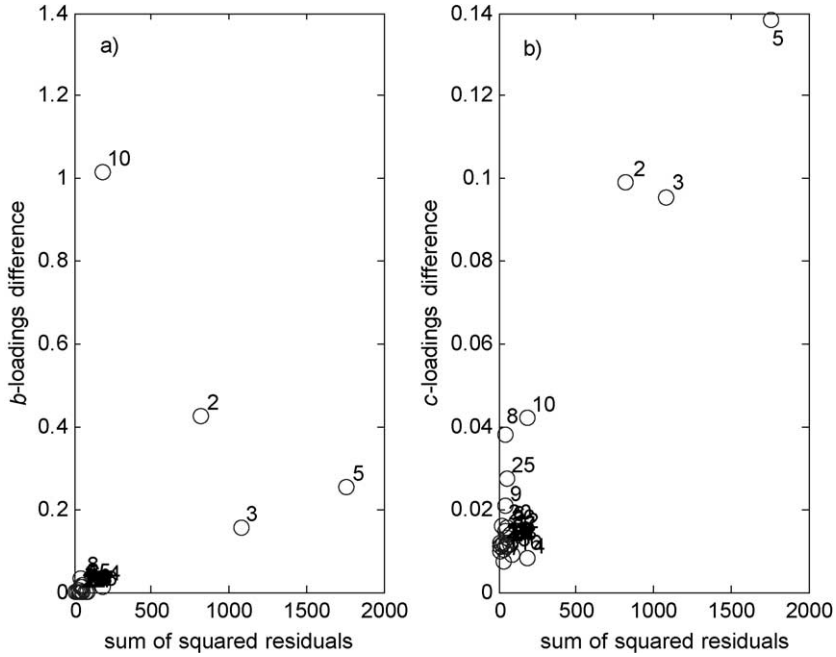

Fig. 7. RIP plots for (a) *b*-loadings and (b) *c*-loadings in the third PARAFAC model.

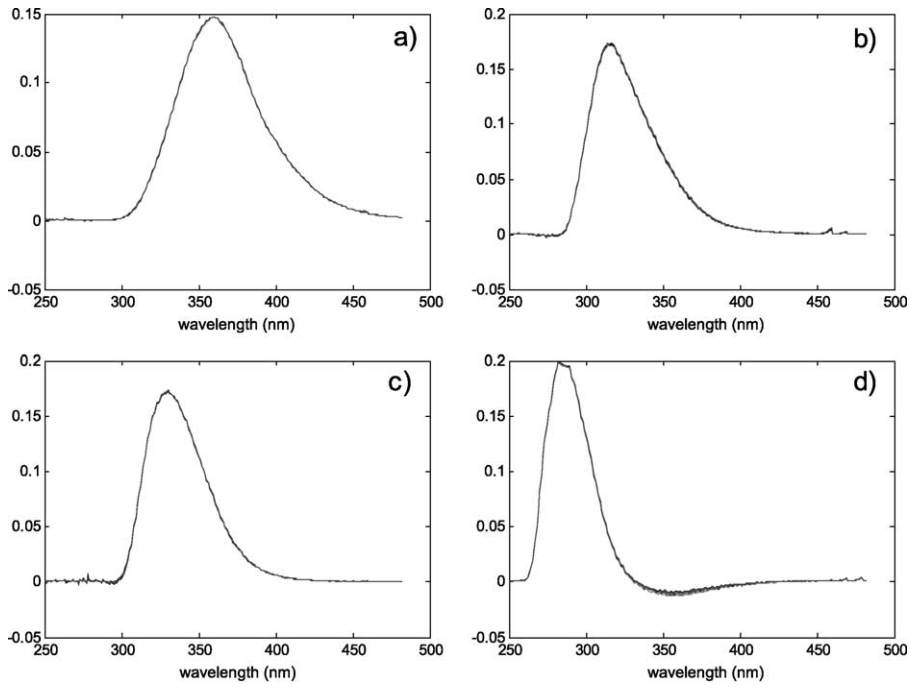

Fig. 8. Estimated pure emission spectral profiles of the 23 jack-knife segments for the four analytes in the fourth PARAFAC model. Corresponds to (a) tryptophan, (b) dopa, (c) hydroquinone and (d) phenylalanine.

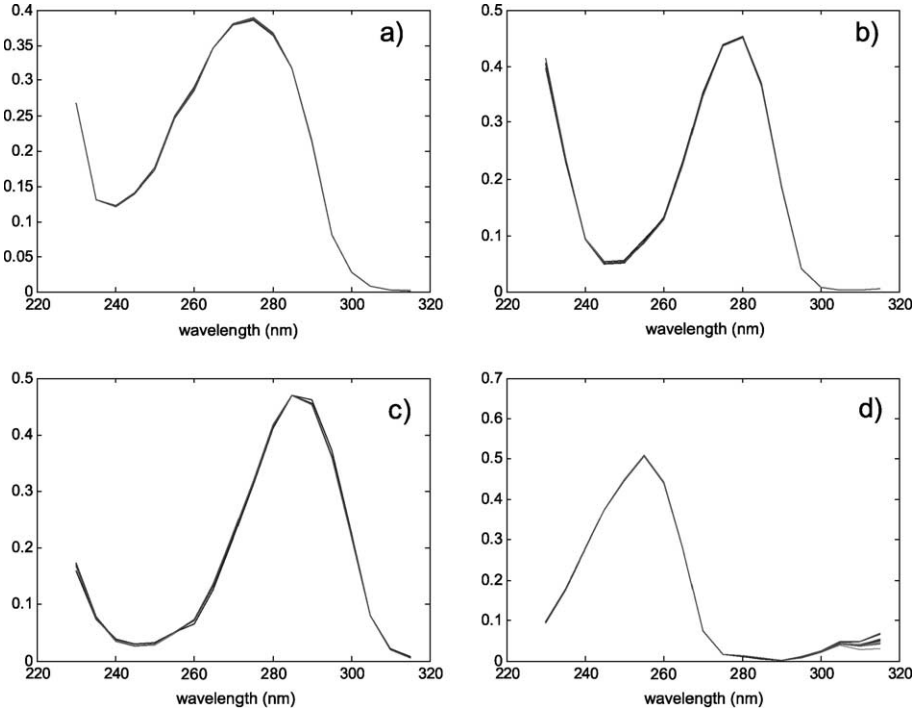

Fig. 9. Estimated pure excitation spectral profiles of the 23 jack-knife segments for the four analytes in the fourth PARAFAC model. Corresponds to (a) tryptophan, (b) dopa, (c) hydroquinone and (d) phenylalanine.

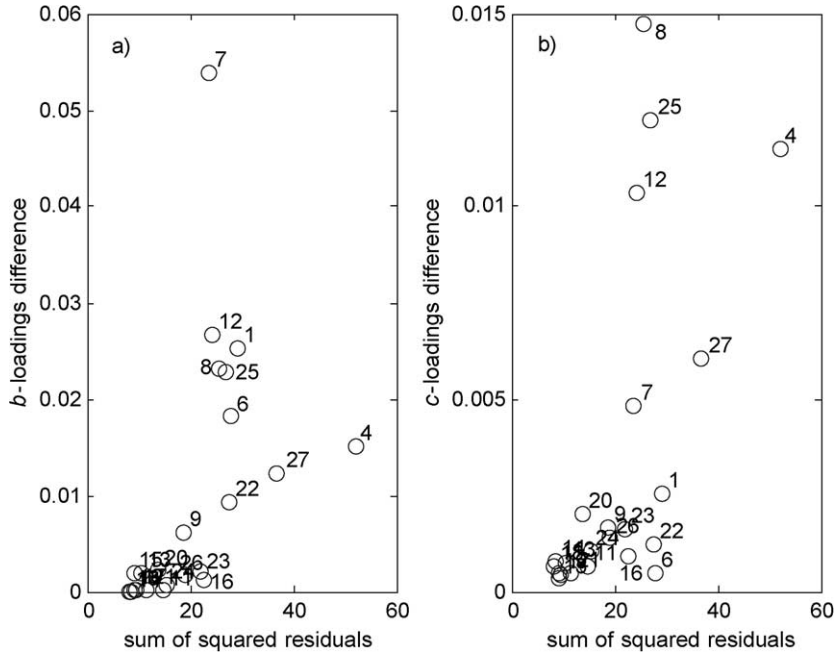

Fig. 10. RIP plots for (a) *b*-loadings and (b) *c*-loadings in the fourth PARAFAC model.

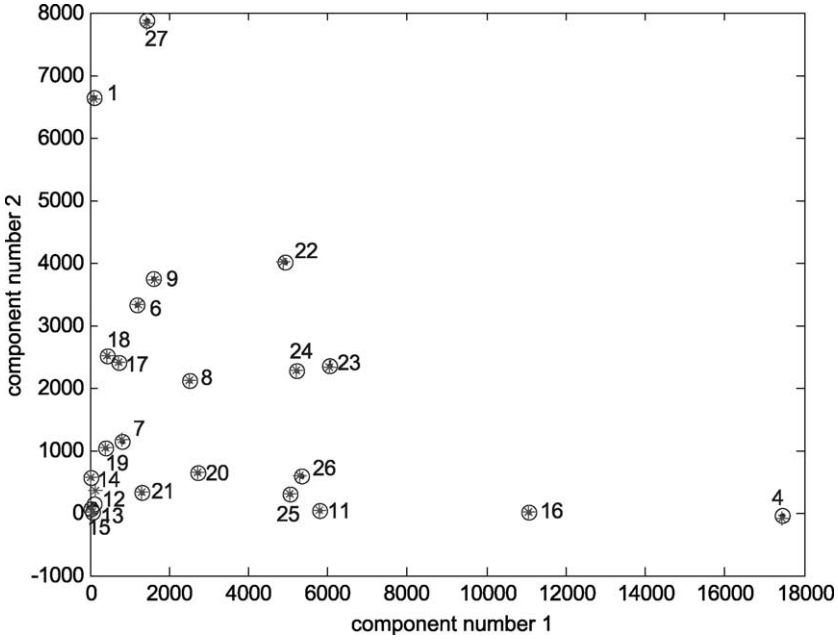

Fig. 11. Jack-knife estimation of the scores for the tryptophan and dopa in the fourth PARAFAC model. The circles represent the score values using the overall PARAFAC model; the asterisks correspond to the scores of the left-out samples calculated from the model with those samples left out.

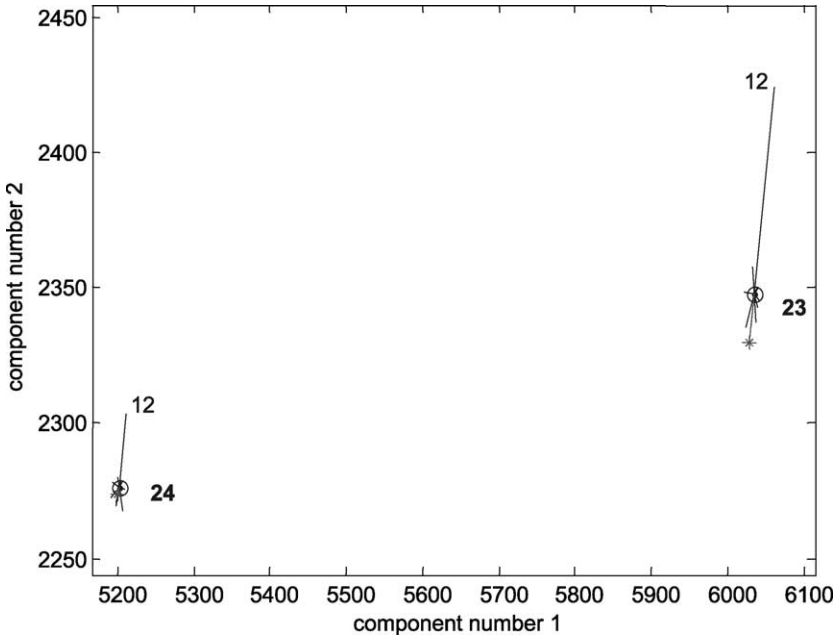

Fig. 12. Jack-knife estimation of the scores (samples number 23 and 24) for the first (tryptophan) and second (dopa) components in the fourth PARAFAC model. The circles represent the score values using the overall PARAFAC model; the asterisks correspond to the scores of the left-out samples calculated from the model with those samples left out; the bold figures indicate the sample number and the other numbers indicate the jack-knife segments.

Table 1
Number of samples in each component for which some jack-knife segments significantly differ from the rest

Jack-knife segment	Component			
	(1) Tryptophan	(2) Dopa	(3) Hydroquinone	(4) Phenylalanine
1	–	1	–	–
12	1	10	–	2
16	–	–	1	–
27	–	–	4	1

The PARAFAC model used is the fourth one in the text.

remove sample numbers 2 and 3 (doubts remain about sample number 5) and to carry out the jack-knife analysis again.

4.2. Second model

The 25 jack-knife pure emission and excitation spectral profiles (not shown) are again unstable (especially the emission ones), and the third and the fourth components have approximately the same profiles (in both emission and excitation). The unstable spectral profiles seem to indicate that the original data, or at least a part of it, are too noisy. In fact, the landscape in Fig. 1 shows that the area between 200 and 230 nm (excitation), corresponding to the extra band of the higher excited states, is noisy and distorted. The low ultraviolet wavelength area of the measured spectra is greatly influenced by the condition of the xenon lamp in the instrument and this is probably the reason for the noise in this area. Hence, it was decided to carry out the jack-knife analysis again, removing the excitation area between 200 and 230 nm and putting back the two samples that were previously detected as outliers, since maybe these were deemed outliers due to their behaviour in the low excitation area.

4.3. Third model

The third model was built using the whole set of 27 samples, the 233 initial emission wavelengths and the last 18 excitation wavelengths (230–315 nm). Fig. 7 shows the RIP plots. We can see that samples 2, 3, 5 and 10 are the ones for which their loadings differ the most from the overall model. Samples number 2, 3 and 5 are the ones with highest sum of squared residuals. In the IMP plots (not shown), no differences

Table 2
Jack-knife results for tryptophan

Sample	$\overline{V}$	s_V	RSD (%)	Deleted	$\overline{V}$	$s_{\overline{V}}^2$	RSD (%)	t_{cal}
1	83.74	59.33	70.84	–				
4	17419.15	43.35	0.25	–				
6	1192.43	77.03	6.46	–				
7	786.69	57.58	7.32	–				
8	2524.94	60.64	2.40	–				
9	1612.17	73.85	4.58	–				
11	5799.63	13.91	0.24	–				
12	83.17	19.75	23.75	–				
13	24.79	34.70	140.00	12	23.20	9.71	41.86	0.2126
14	10.52	6.15	58.51	–				
15	77.55	15.90	20.50	–				
16	11032.48	27.01	0.24	–				
17	727.53	33.53	4.61	–				
18	424.96	28.23	6.64	–				
19	385.75	20.95	5.43	–				
20	2708.09	24.56	0.91	–				
21	1286.35	13.89	1.08	–				
22	4900.23	44.26	0.90	–				
23	6035.80	33.42	0.55	–				
24	5201.42	18.96	0.36	–				
25	5039.56	15.99	0.32	–				
26	5310.14	7.34	0.14	–				
27	1452.20	85.63	5.90	–				

can be seen among scores, but samples 2, 3, 5 and 10 appear in the extreme of the calibration range. So a little variation in these samples may greatly affect the PARAFAC model, as it can be seen from the results shown in Fig. 7. Taking all of these observations into account, samples 2, 3, 5 and 10 were removed and the jack-knife analysis was carried out again.

4.4. Fourth model

Figs. 8 and 9 show the pure emission and excitation spectral profiles, respectively after removing samples 2, 3, 5 and 10. All the 23 profiles for each component are very similar, and they are much less unstable than the ones in Figs. 2 and 3. Fig. 10 shows the RIP plots. Now there are apparently no gross outliers. Sample number 7 is the one with the highest difference of loadings in the second mode, but as its difference in the third mode is not among the largest ones, and its sum of squared residuals is also not among the largest ones, it was decided not to remove this sample. The IMP plots (not shown) show a very good fit among the predicted score values and the scores of the overall model. Hence, it was concluded that there were no grounds for removing more samples.

4.5. Jack-knife estimation of standard errors

Fig. 11 shows the jack-knife estimations of the scores for components 1 and 2: tryptophan and dopa (this figure is also representative of the other components). The score values using the overall PARAFAC model (circles) are close to the score values predicting each sample when they are left out (asterisks). In the figure, it is difficult to distinguish the lines corresponding to the different jack-knife estimations for each sample. This figure seems to show that there are no more outliers, and that the standard errors associated to the score values are low, since they are calculated from jack-knife estimations for each sample in Eq. (4). But taking a closer look at a part of Fig. 11, Fig. 12 is obtained. In Fig. 12, jack-knife estimations for samples number 23 and 24 are seen. Clearly, jack-knife estimation for the 12th initial segment in the second component is much higher than the rest. This is also consistent with the fact that in Fig. 11, the score value for the 12th sample using the overall

Table 3
Jack-knife results for dopa

Sample	$\overline{V}$	s_V	RSD (%)	Deleted	$\overline{V}$	s_V^2	RSD (%)	t_{cal}
1	6654.64	51.99	0.78	12	6656.94	18.96	0.28	0.206450
4	−43.66	31.66	72.52	–				
6	3341.42	39.39	1.18	12	3339.66	14.00	0.42	0.206824
7	1158.38	65.54	5.66	12	1155.41	20.33	1.76	0.211288
8	2126.92	25.09	1.18	12	2125.86	11.65	0.55	0.194076
9	3742.22	7.62	0.20	–				
11	26.83	16.09	59.96	–				
12	147.34	54.11	36.73	–				
13	74.31	95.91	129.07	12	69.87	22.50	32.20	0.216567
14	573.50	5.15	0.90	12	573.72	2.46	0.43	0.179340
15	17.03	4.57	26.84	1	16.89	3.43	20.29	0.127501
16	13.31	31.10	233.65	–				
17	2397.91	6.13	0.26	–				
18	2521.58	9.31	0.37	–				
19	1052.49	3.17	0.30	–				
20	651.10	8.98	1.38	–				
21	319.90	4.63	1.45	–				
22	4017.22	55.58	1.38	12	4014.71	17.55	0.44	0.210673
23	2349.53	77.92	3.32	12	2345.97	22.20	0.95	0.213280
24	2276.86	29.91	1.31	12	2275.60	13.75	0.60	0.195224
25	302.18	15.03	4.97	–				
26	579.16	15.15	2.62	–				
27	7867.73	47.30	0.60	12	7869.91	12.20	0.16	0.214580

PARAFAC model is not very close to the predicted score of the 12th sample when it is left out. This phenomenon mainly occurs in the second component (in 11 of the 23 remaining samples), but also in the first component (in one sample), in the third (in five samples) and in the fourth component (in three samples). Table 1 summarises the number of samples in each component for which some jack-knife segment significantly differs from the rest (according to the test of Grubbs [26]). It has to be noted that 8 out of the 20 jack-knife segments in Table 1 that differ from the rest correspond to approximately zero score values. Table 2 shows jack-knife results for the first component, tryptophan (the one with fewer jack-knife segments that differ from the rest) and Table 3 shows jack-knife results for the second component, dopa (the one with the most jack-knife segments that differ from the rest). The last column in both Tables 2 and 3 shows the *t*-value found comparing the estimation of standard errors ($\overline{V}$ and s_V calculated from Eq. (4)), with and without the jack-knife segments specified in the 'deleted' column, using a *t*-test. Comparing all these *t*-calculated values with the tabulated α significant level of 5% ($t_{cal}=2.0796$), the conclusion in all the cases is that the results do not significantly differ among them.

The normality of the 22 jack-knife segments in every sample in every component was tested using the Kolmogorov–Smirnov test [27]; the distribution of the results did not differ from normality. In some cases, it was necessary to remove the jack-knife segments specified in Table 1, but as pointed out, the estimation of standard errors with or without these segments do not significantly change.

5. Conclusions

In this paper, we have applied the jack-knife technique in order to find the standard errors associated with the score values in a PARAFAC model. The jack-knife segments involved in finding these standard errors turned out to be normally distributed for most of the samples according to the Kolmogorov–Smirnov test. Only some samples did not have all the jack-knife segments normally distributed (and about half of these samples corresponded to approximate zero score values), but deleting some outlying jack-knife segments (at most one segment per sample), the distribution of the jack-knife segments turned out to be normal for all the samples. It is important to remark that the results (score averages and standard errors) with and without these outlying segments do not significantly differ.

The jack-knife technique has also been useful in order to detect outlying samples. In order to ease the outlier detection, two new tools, the Resample Influence Plot (RIP) and the Identity Match Plot (IMP) have been developed and applied.

Acknowledgements

Jordi Riu would like to thank the 'Secretaría de Estado de Educación y Universidades' of the Spanish Ministry of Education, Culture and Sports for providing his postdoctoral fellowship.

R. Bro acknowledges support provided by Frame program Advanced Quality Monitoring in the Food Production Chain (AQM), as well as the EU-project Project GRD1-1999-10337, NWAYQUAL. The data and m-file for performing jack-knifed PARAFAC are available from http://www.models.kvl.dk. The authors would also like to thank Giorgio Tomasi for providing some Matlab m-files.

References

[1] R.A. Harshman, UCLA Work. Pap. Phon. 16 (1970) 1–84.
[2] R.A. Harshman, J. Chemom. 15 (2001) 689–714.
[3] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[4] L. Tucker, Psychometrika 31 (1966) 279–311.
[5] P.M. Kroonenberg, J. de Leeuw, Psychometrika 45 (1980) 69–97.
[6] J.M.F. Ten Berge, J. de Leeuw, P.M. Kroonenberg, Psychometrika 52 (1987) 183–191.
[7] R. Bro, J. Chemom. 10 (1996) 47–61.
[8] S. Wold, K. Esbensen, P. Geladi, Chemom. Intell. Lab. Syst. 2 (1987) 37–52.
[9] J.B. Kruskal, in: R. Coppi, S. Bolasco (Eds.), Multiway Data Analysis, Elsevier, Amsterdam, 1989, pp. 8–18.
[10] N.D. Sidiropoulos, R. Bro, J. Chemom. 14 (2000) 229–239.
[11] L. Munck, L. Nørgaard, S.B. Engelsen, R. Bro, C.A. Andersson, Chemom. Intell. Lab. Syst. 44 (1998) 31–60.
[12] D. Baunsgaard, C.A. Andersson, A. Arndal, L. Munck, Food Chem. 70 (2000) 113–121.
[13] H.A.L. Kiers, J. Chemom. 14 (2000) 151–170.
[14] R. Bro, Doctoral Thesis: Multi-way Analysis in the Food In-

dustry: Models, Algorithms and Applications, University of Amsterdam, 1998.
[15] R.P. McDonald, Br. J. Math. Stat. Psychol. 33 (1980) 161–183.
[16] C.R. Rao, S. Mitra, Generalized Inverse of Matrices and its Applications, Wiley, New York, 1971.
[17] J.W. Tukey, Ann. Math. Stat. 29 (1958) 614.
[18] H. Martens, M. Martens, Food Qual. Prefer. 11 (2000) 5–16.
[19] S.D. Peddada, in: C.R. Rao (Ed.), Handbook of Statistics, vol. 9, North-Holland, Amsterdam, 1993, pp. 723–744.
[20] L. Tucker, A method for synthesis of factor analysis studies, Report number 984, Department of the Army, Washington, 1951.
[21] Matlab, The Mathworks, Natick, MA, 2001.
[22] P.E. Gill, W. Murray, M.H. Wright, Practical Optimization, Academic Press, London, 1981.
[23] H. Martens, T. Næs, Multivariate Calibration, Wiley, Chichester, 1989.
[24] C.A. Andersson, R. Bro, Chemom. Intell. Lab. Syst. 52 (2000) 1–4.
[25] D. Baunsgaard, Factors Affecting 3-way Modelling (PARAFAC) of Fluorescence Landscapes, Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Frederiksberg, Denmark, 1999.
[26] AOAC methods committee, J. AOAC Int. 78 (1995) 143A–160A.
[27] G. Kateman, F.W. Pijpers, Quality Control in Analytical Chemistry, Wiley, New York, 1981.