Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemometrics and Intelligent Laboratory Systems 71 (2004) 97–106
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area

Lisbeth Garbrecht Thygesen[a,1], Åsmund Rinnan[a,*], Søren Barsberg[b], Jens K.S. Møller[c]

[a] Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Technology, Rolighedsvej 30, DK-1958 Frederiksberg, Denmark
[b] Royal Veterinary and Agricultural University, Plant Fibre Laboratory, Agrovej 10, DK-2630 Taastrup, Denmark
[c] Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Chemistry, Rolighedsvej 30, DK-1958 Frederiksberg, Denmark

Received 15 June 2003; received in revised form 10 November 2003; accepted 7 December 2003

Available online 5 March 2004

Abstract

The use of fluorescence spectroscopy for recording multiple excitation and corresponding emission wavelengths and the subsequent technique of analyzing the resulting fluorescence landscapes is a rather new method as opposed to the use of just a single excitation wavelength. In a fluorescence landscape, several light-scatter effects are usually present, and often the part of the landscape containing information on the chemical and/or physical characteristics of the sample is surrounded by two Rayleigh scatter lines. When such landscapes are decomposed using parallel factor analysis (PARAFAC), the scatter effects may have detrimental effects on the resolved spectra, especially if the peaks from the analytes lie close to or on the Rayleigh scatter lines. Normally, all values close to and outside the Rayleigh scatter lines are set to missing values before decomposing the fluorescence landscapes by PARAFAC. In this paper, we introduce a novel pretreatment method applicable for two-dimensional fluorescence landscapes, where instead of inserting only missing values a mixture of zeros and missing values are inserted close to and outside the Rayleigh scatter lines. It is shown that, by the use of this technique, a physically and chemically meaningful decomposition is obtained, and furthermore the modeling converges faster. Constraining the PARAFAC solution to positive values in all modes gave results similar to those obtained for the unconstrained model, except that the loadings where less smooth and the number of iterations before convergence was smaller.

Keywords: PARAFAC; Fluorescence; Rayleigh scatter; Missing values; Inserting zeros into the EEM

1. Introduction

The role of fluorescence spectroscopy in the analysis of organic products has increased. Handling these data with parallel factor analysis (PARAFAC) [1] is a very powerful way of extracting information from an excitation–emission matrix (EEM), i.e. several samples where the emission intensity is depicted as a function of both excitation and emission wavelengths. Often these data also include light scattering effects, such as Rayleigh scatter. Since PARAFAC only decomposes trilinear structures and the scatter is on the diagonal (excitation = emission), this causes some mathematical difficulties in the decomposition. It is therefore of interest to remove this effect, or at least to reduce it as much as possible. Several ways of handling these scattering effects have been presented previously: weighting the scatter areas down (or areas containing information up) [2,3], inserting missing values [4] or plainly avoiding the part of the matrix that includes the scatter. The last method, however, can only be used in cases where the removed wavelengths contain little or no information. For some analysis, this is mostly the case, and thus there need not be any concern about the scattering effect. The method of inserting missing values, on the other hand, can lead to unacceptable decomposition of the spectra. Some of these models show photophysical impossible features, like a component emitting light at a higher energy level than the absorbed light.

In this paper, we describe a new way of handling these problems. The method can be seen as a pretreatment, where

* Corresponding author.
E-mail address: aar@kvl.dk (Å. Rinnan).
[1] Present affiliation: Department of Civil Engineering, The Danish Technical University, Building no. 118, DK-2800 Kgs. Lyngby, Denmark.

0169-7439/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2003.12.012

zeros are inserted into the matrix. PARAFAC will then converge faster and give better estimates of the spectra. This method can be used on any type of fluorescence landscapes, but it does not have a significant effect on the resolved spectra unless one or more of the peaks lie close to or on the Rayleigh scatter lines, as illustrated in the emission spectra in Fig. 1.

Andersen and Bro [5] explained that inserting zeros into the dataset instead of using missing values does not conform to the trilinear structure of the EEM. While this is true in general, this paper will show that in some practical situations, even though mathematically being untrue, inserting zeros in part of the spectra can be very handful for the decomposition of the underlying emission and excitation spectra.

1.1. Theory

1.1.1. Terms used

The area between the Rayleigh scatter lines will in this paper be called the ‘‘data area’’. The part of the landscape where the emission wavelength is smaller than the excitation wavelength will be termed below the first order Rayleigh scatter. Above the second order Rayleigh scatter is where the emission wavelength is larger than twice the excitation wavelength.

1.1.2. Information in the data

There is rarely any additional chemical information outside the data area. More specifically, the area below the first order Rayleigh scatter does not give any physical meaning, since a molecule cannot emit light of higher energy than what it absorbed, and can thus be deleted. However, removing this part is not a simple matter since it makes up a triangular area in the matrix. The easiest way of circumventing this problem would be to enlarge this area into the smallest possible rectangle and remove this part. By this, you might, however, remove some interesting information and, therefore, it is not a good method for solving the problem. A different, and much used method, is to insert missing values below the first order Rayleigh scatter. To ensure that all the Rayleigh scatter is removed, some extra data points around the first order Rayleigh scatter are also replaced with missing values. This often gives satisfying results, but the amount of missing values can affect the convergence of PARAFAC and the quality of the results. A third way of handling this type of data is to either weight down the part outside the data area, or weight up the data area. This is a powerful method, but it is computationally cumbersome, increasing the computational time before convergence by a factor of 10, or even as high as 100 compared to non-weighted PARAFAC.

The area above the second order Rayleigh scatter, on the other hand, is meaningful, but rarely holds any new chemical information. The information in this part is mostly an echo of the information in the data area. These values can be treated in the same way as for the values below the first order Rayleigh scatter, with the limitations described.

1.1.3. Constraints

PARAFAC modeling gives the least squares solution, and this solution is often also in accordance with the physical/chemical premises of the data. Sometimes, it may, however, be valuable to enforce some constraints on the PARAFAC solution, especially in situations where the number of missing values is relatively high. In these situations, small model-errors may strongly bias the estimated spectra. Thus, the estimated spectra may give no or little chemical meaning. By constraining the PARAFAC model with sound constraints in agreement with a priori knowledge about the system, such as non-negativity and maybe unimodality, the resolved spectrum will more clearly

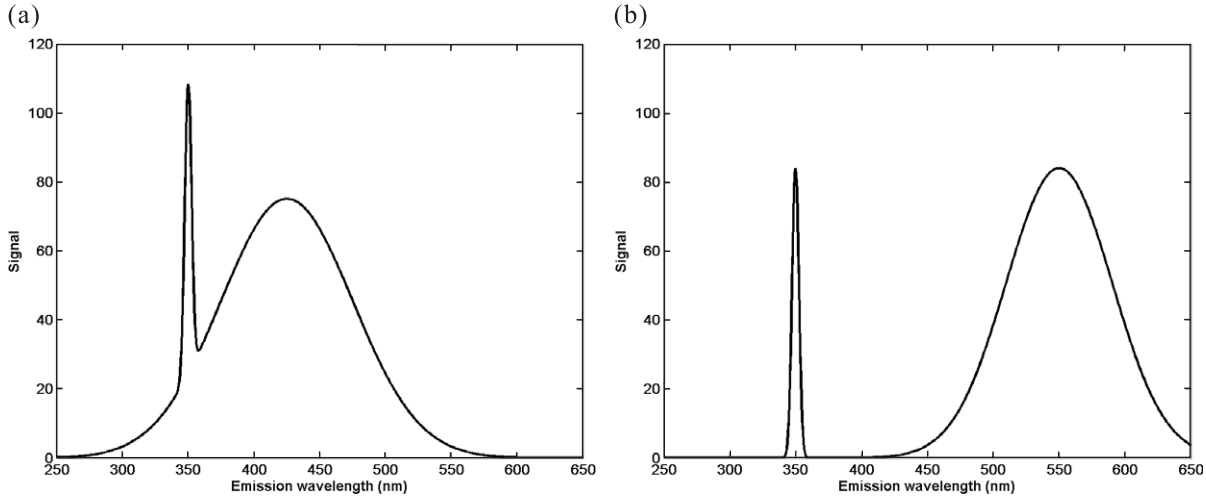

Fig. 1. Emission spectrum with a Rayleigh scatter line and a peak from a chemical component when excited at 350 nm (simulated data). (a) The problematic case, (b) the insertion of zeros does not affect the solution very much.

indicate the chemical attributes of the sample. Spectroscopic data should be strictly positive, and hence the use of non-negativity may be enforced upon the data to ensure this [6–8]. Fluorescence spectra of a single fluorophore without vibrational fine structure will only have one emission peak, making it adequate to use unimodality constraints. Normally, this constraint is not valid in the excitation mode, which—even in the absence of vibrational fine structure—may contain several excitation peaks (electronic transitions). In the case where only a limited range of excitation wavelengths is recorded, it may, however, be valid.

2. Experimental

2.1. Methods

A novel way of pretreating fluorescence spectra is to insert zero-values in parts or in the whole of the landscape outside the data area.

Four different pretreatment methods of the data are tested in this paper (see Fig. 2):

1. Only missing values—conventional method
2. Zeros below first order and missing values above second order—mixed method
3. Only zeros—all zeros method
4. Mostly zeros, but with a ribbon of missing values around the Rayleigh scatter lines—ribbon method

As Andersen and Bro clearly stated, methods 2–4 above do not conform to the trilinear structure of the EEM. However, a large amount of missing values may cause problems in the decomposition. Especially if the number of missing values at one emission/excitation wavelength is very large (typically 80% or more), small artifacts close to the missing values may lead to large artifacts in the extracted spectra (see Fig. 4). So, even though inserting zeros is not mathematically true, it can help PARAFAC to decompose into a solution that is more

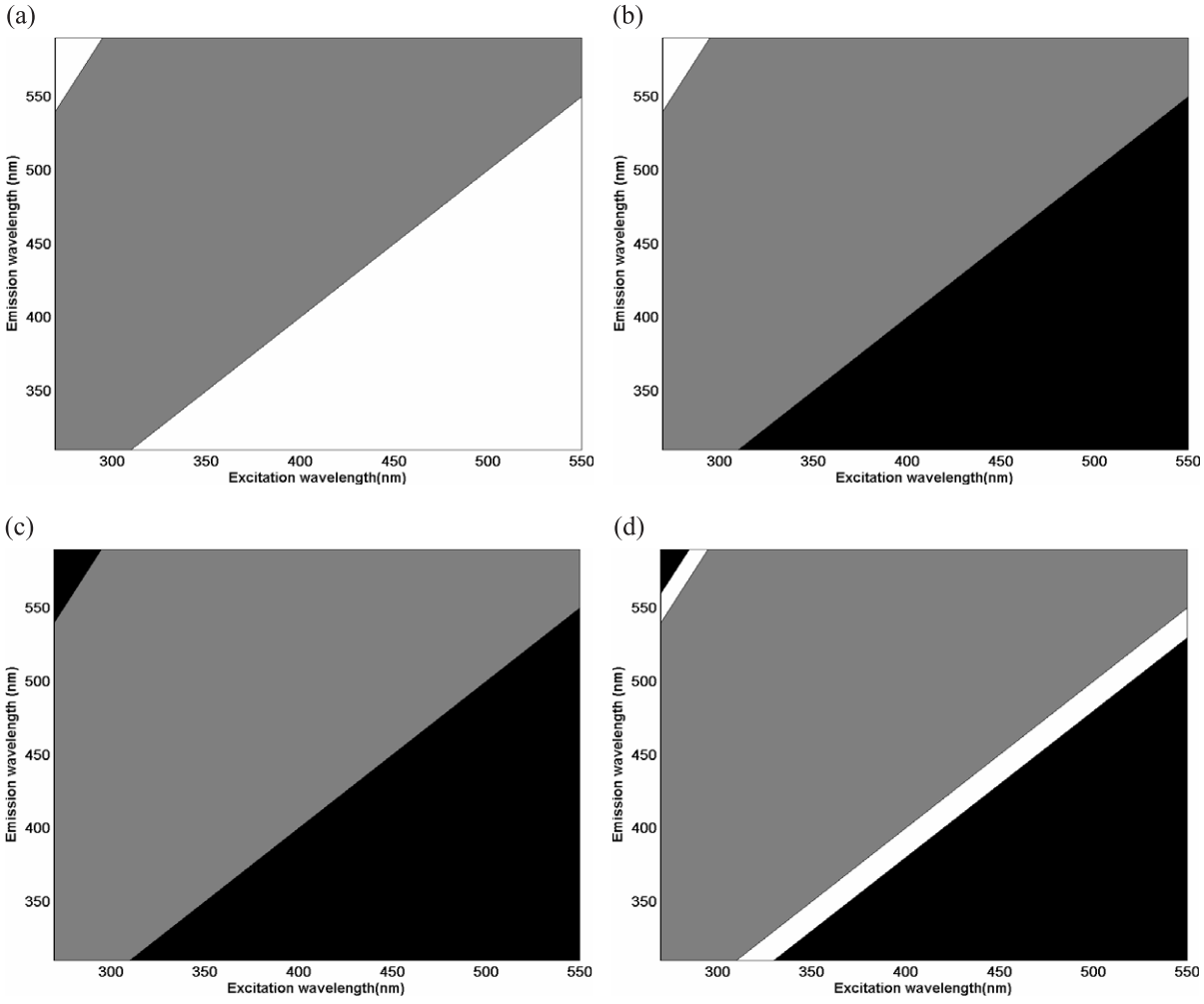

Fig. 2. Different strategies for inserting missing (white) and zeros (black) outside the data area of an excitation–emission matrix (data area: grey). (a) Only missing values, (b) zeros outside the first order, missing outside the second order, (c) only zeros and (d) mostly zeros with a ribbon of missing values around the Rayleigh scatter lines.

liable than upon keeping all the missing values (method 1). Inserting zeros may be seen as a weak form of non-negativity constraint, since the spectra are guided towards zero, but not set to strictly be non-negative. However, since one fluorophore's excitation and emission spectra might have some overlap (which is the case of, e.g., tyrosine), inserting zeros will destroy this possibility of overlap. Inserting zeros below the first order Rayleigh indicates that there should not be any overlap in the emission and excitation spectra.

2.1.1. Conventional method

This method has already been discussed and will therefore not be explained further.

2.1.2. Mixed method

Sometimes the areas with missing values especially below the first order Rayleigh scatter can be quite large. This can cause PARAFAC to convergence slower, as it has to estimate a large number of missing values. It can also give unstable or non-meaningful results. Therefore, inserting zeros outside the scatter line can stabilize the decomposition and make it converge faster. From a photophysical point of view this method correctly reflects the fact that no emission is expected below the first order Rayleigh scatter, whereas emission tails may extend into the part above the second order Rayleigh scatter, as is seen for some samples analyzed in the present work.

2.1.3. All zeros method

Instead of missing values above the second order Rayleigh scatter, zeros are inserted. This may cause PARAFAC to converge even faster and give better estimates of the spectra.

2.1.4. Ribbon method

In some of the above-mentioned methods, where zeros are inserted adjacent to the data area, PARAFAC may produce chemically unsound solutions due to this introduction of discontinuity in the data. One way to circumvent this problem is to insert a ribbon of missing values around the Rayleigh scatter lines, which also permits some overlap between one fluorophore's emission and excitation spectra. In this way, PARAFAC will be "free" to estimate a continuous shape of the peaks. Three different widths of the ribbon are chosen: 1, 5 and 10 units. A unit equals the width of a single step in the excitation or emission (wavelength) dimension, which is defined by the experimental conditions. It should be noted that, for the data analyzed in this paper, there is no difference in a width of 5 and 10 units for dataset II above the second order Rayleigh scatter, since the dimensions of the excitation–emission matrix are transgressed with the increase from 5 to 10 units. The ribbon was applied in two ways: only around the first order Rayleigh scatter (A) or both around the first and second order Rayleigh scatter (B). Ribbon method "1A" thus denotes a ribbon width of 1 unit in conjunction with using the ribbon only around the first order Rayleigh line.

2.2. Quality of a model

The pretreated dataset is decomposed into its pure chemical components by the use of PARAFAC (with or without constraints). This results in a number of factors for every model, and it is of vital importance to evaluate which models are adequate. In order to evaluate this, five different criteria were set (they are set in the order of importance):

- No emission spectrum of a component can have its peak placed at lower wavelength than the corresponding excitation spectrum peak
- Only small negative values in the resolved spectra are allowed
- The explained variance of the model should be high (above 97%)
- All factors should be smooth and not only describe noise
- Finally, the spectra were evaluated by visual inspection to make sure they were reasonable

All analyses were performed in Matlab 6.1 and 6.5 for Windows (Mathworks, Natick, MA, USA), with algorithms taken from the N-way toolbox version 2.10 [9], available at www.models.kvl.dk. The algorithms for the insertion of zeros were written in-house.

2.3. The data

2.3.1. Dataset I (wood fibers)

Thermo mechanical pulp (wood fibers) of spruce (*P. abies*) was supplied by Sunds Defibrator, Sundsvall, Sweden. The pulp was dried at ambient temperature and humidity before use. The pulp is autofluorescent due to its content of the fluorescent plant polymer lignin. Samples of different emission intensities were produced by the adsorption of (non-emissive) *p*-benzoquinone into the fiber cell walls. The control sample had no *p*-benzoquinone adsorbed, whereas three less emissive samples had three different quantities of the quinone adsorbed, the larger the adsorbed amount the smaller the emission. The experimental setup is described in more details elsewhere [10].

The samples were pressed into disks of 1 mm thickness by a Perkin-Elmer hydraulic press (at 5 bar). The disks were used for fluorescence measurements by a SPEX 1680 0.22-m double monochromator fluorescence spectrometer in front-face setup. The samples were measured at 35 excitation wavelengths (259–599 nm), and 30 emission wavelengths (355–645 nm), both with a step of 10 nm. For the control sample, as well as for the quinone-containing

samples, five independent excitation–emission intensity matrices were produced. However, for the quinone-containing sample with the highest content of quinone, only 3 of the 5 matrices have been included because the remaining 2 matrices were highly deviating from the other 18 matrices. The cause of this deviation was not further investigated and, because the aim of this study is to show the effects of inserting zeros into the EEM and not how to detect outliers, no further explanation as to why these were outliers will be given. Fluorescence data for sample number 1 of this dataset is shown in Fig. 3a and b.

2.3.2. Dataset II (dry-cured hams)

Lean raw pork from fresh hams was obtained from the local market, whereas Parma hams were from a processing plant in Parma, Italy. Parma ham ages ranges from salted (3 months) to matured (11 and 12 months) and further to aged (15 and 18 months). Samples were thermostated in a water bath at 25 °C before measuring. A total of 67 meat samples were submitted to duplicate measurements of surface autofluorescence spectroscopy.

The measurements were done using a BioView instrument (Delta Light and Optics, Lyngby, Denmark) equipped with a fiber optics measuring probe giving an open-end 180° excitation–emission geometry. The instrument used a pulsed xenon lamp for excitation, and a surface area of ~ 6 mm in diameter was sampled in each measurement. The samples were measured at 15 excitation wavelengths (270–550 nm), and 15 emission wavelengths (310–590 nm), both with a step of 20 nm. The emission wavelengths were shifted by 40 nm from each excitation wavelength applied. Before analysis of the data, the excitation wavelengths above 470 nm, and the emission wavelengths below 350 nm were removed. Thus, the dimension of the dataset was reduced from $67 \times 15 \times 15$ to $67 \times 11 \times 13$ (samples × excitation × emission). The areas deleted only contain very few measurement values and were thus removed to stabilize the model. The data and their chemical interpretation in relation to process control is described in detail elsewhere [11]. Fluorescence data for sample number 50 of this dataset is shown in Fig. 3c and d.

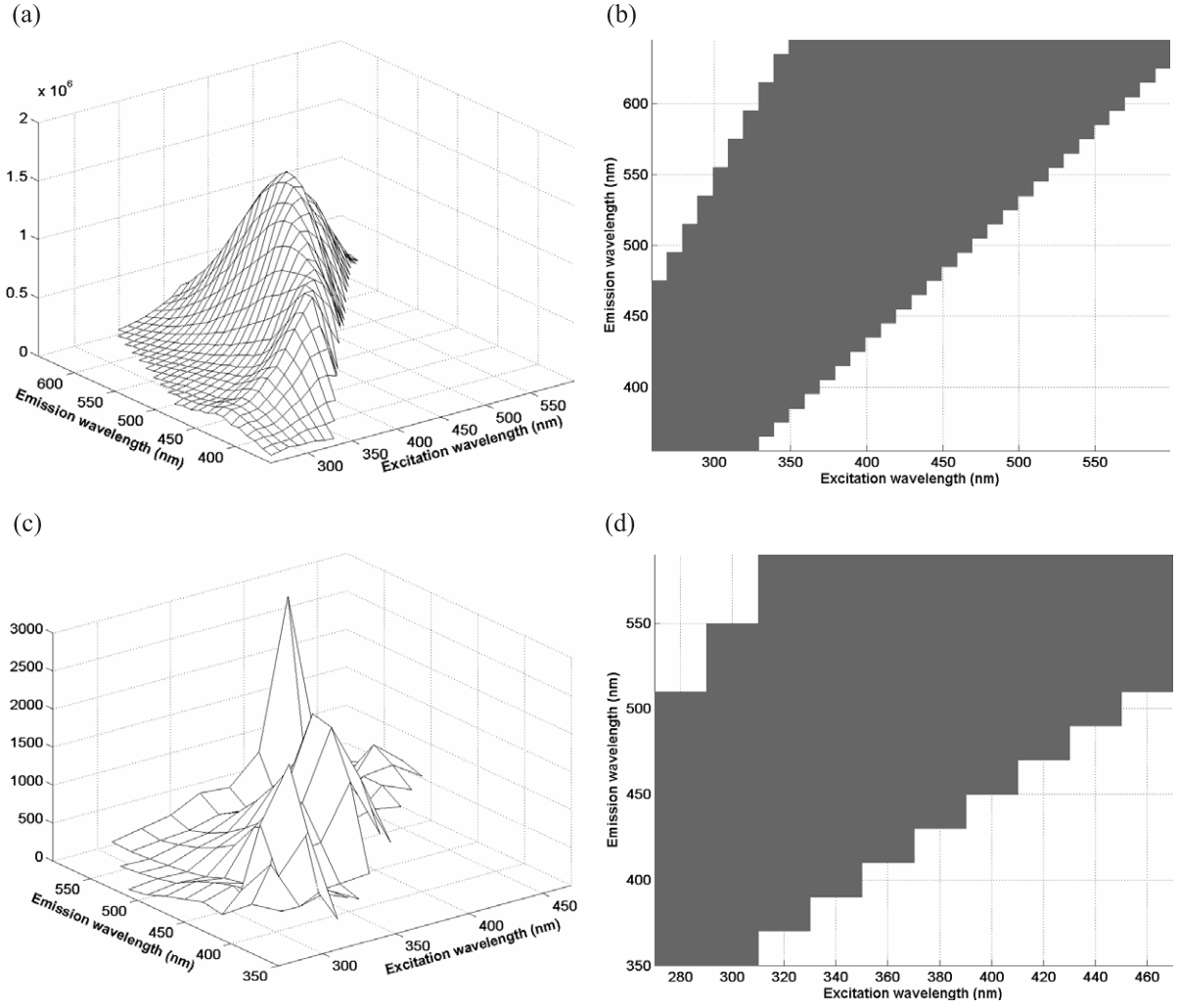

Fig. 3. Fluorescence landscape and data area for sample no. 1 from dataset I (a and b) and from sample no. 50 from dataset II (c and d).

2.4. Data analysis

Thirty-six PARAFAC models were calculated for each dataset, i.e. all combinations of the nine methods for insertion of zeros into the data matrices (of which six are the different ribbon methods), and four different ways of constraining the PARAFAC model: no constraints, non-negativity constraint on the two spectroscopic modes, non-negativity constraint on all three modes and unimodality constraint on the two spectroscopic modes. Apart for the constraints, PARAFAC was carried out using default input parameters. For dataset I, the number of components in the models was set to three based on an evaluation of model stability and the shape of the loadings. For dataset II, five component models were calculated in accordance with a previous publication on the dataset [11]. Normally, there is little sense in assessing model quality without discussing the number of components to include in the model. However, in the present study, focus is on the effect of inserting zeros into the data matrices, and thus a fixed number of components for each dataset have been chosen.

3. Results and discussion

Fig. 4 shows the scores and loadings for five different PARAFAC models obtained for dataset I. All models were calculated without constraints. The sample scores are very similar, both within each model and within the models, indicating that the fluorophores in the wood fiber samples co-vary. Furthermore, the sample score follow the intensity of the treatment (quinone-adsorption), i.e. in reality the dataset only contains four uniquely different sample types. These two characteristics are bound to make the PARAFAC model less stable for dataset I, and a Tucker3 model with dimensions [1 3 3] might have been more appropriate. However, a PARAFAC split-half analysis of dataset I gave the same sets of loadings for the two subsets, and thus it seems reliable to use PARAFAC with three components. In Fig. 4, the excitation and emission loadings are very different between models, even though the explained variance does not differ with more than 1–2%. From a physical point of view, the model for the conventional method is not acceptable, as the emission peak appears at a shorter wavelength than the excitation peak for at least one of the

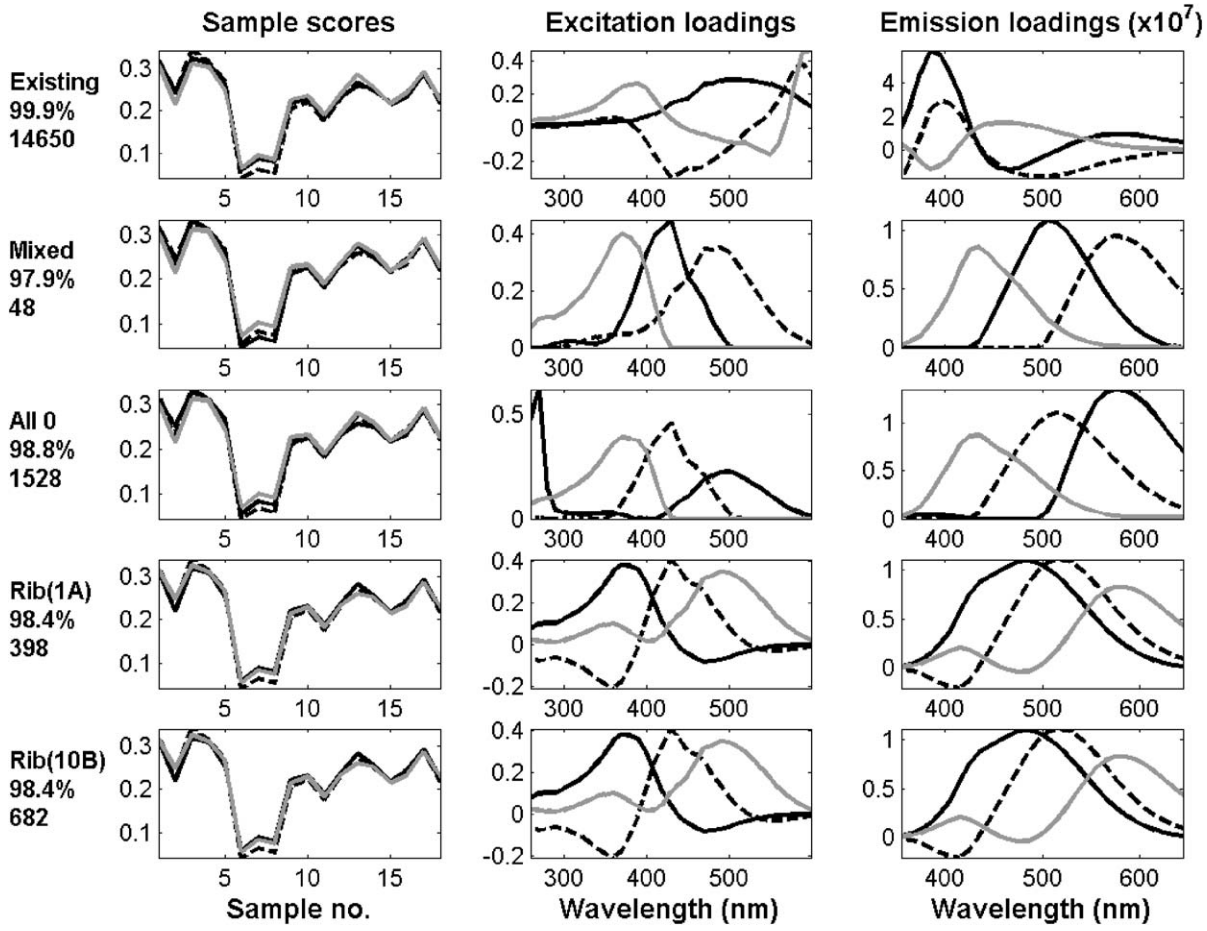

Fig. 4. Scores and loadings for five different unconstrained PARAFAC models for dataset I. The text to the left gives for each set of loadings the method used for insertion of zero values (line 1), the percent explained variance (line 2) and the number of iterations necessary before convergence was reached (line 3).

components, and more than half of at least one loading is negative. Furthermore, the emission loadings are not unimodal, as they should be if each component represented a single fluorophore. The model for the all zeros method, on the other hand, has three perfectly unimodal loadings with a reasonable spacing between each excitation peak and its corresponding emission peak. Another nice feature of the model for the all zeros method is that it converged in only 48 iterations, while the model for the conventional method took 14,650 iterations. The model for the mixed method gave emission loadings more or less identical to the model for the all zeros method, except that one of the excitation loadings has an extra peak below 300 nm. The models for the ribbon method were all rather similar (only some results shown), independent of whether a narrow or a broad ribbon of missing values was inserted and independent of whether ribbons of missing values were inserted both above and below the data area or only below. In Fig. 4, only two of the models for the ribbon method are included, namely methods 1A and 10B. For both models, the excitation and emission loadings are reasonably unimodal, but all have small negative values at the base of the main peaks. Using the criteria for an acceptable model set up in Section 1, the model for the conventional method is unacceptable. The other four models, however, are acceptable, with the model for the all zero method being the best from a physical point of view, albeit it explains slightly less of the variance compared to the other models.

Fig. 5 shows the model of the fluorescence landscape for sample no. 1 of dataset I for four different combinations of PARAFAC modeling constraints and zero insertion methods. Fig. 5a confirms what is already indicated in Fig. 4, namely that the unconstrained model for the conventional method has a spurious peak outside the data area. Fig. 5a shows that the spurious peak is an order of magnitude larger than the part modeling the data area. The non-negativity constrained model for the all zeros method (Fig. 5b) also shows a spurious peak outside the data area, but in contrast to what is the case for the model in Fig. 5a, the modeling of the data area is acceptable according to the criteria set up in the introduction. Fig. 5c and d illustrates the difference between "constraining" the data, i.e. the input of PAR-

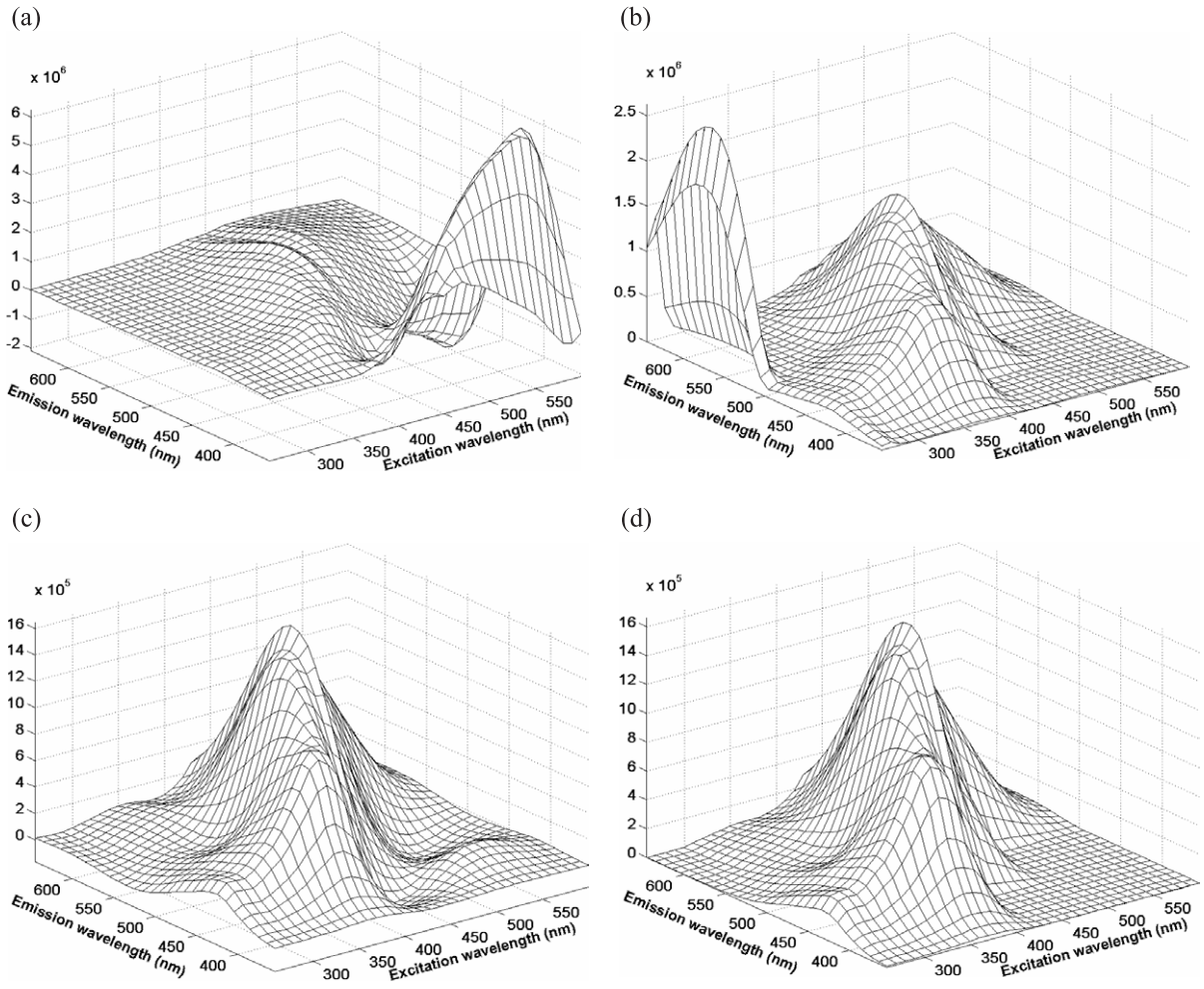

Fig. 5. Selected PARAFAC models of the fluorescence landscape of sample no. 1 from dataset I: (a) conventional method, unconstrained, (b) all zeros method, non-negativity constraint on all three modes, (c) ribbon (1A) method, unconstrained and (d) ribbon (10B) method, non-negativity constraint on all three modes.

AFAC, and constraining PARAFAC directly, for data subjected to pretreatment according to the ribbon method. In Fig. 5c, the missing values were "constrained" to zero, except for one row below the first order Rayleigh scatter, while PARAFAC was unconstrained. In Fig. 5d, zero values were only inserted far from the actual data area (10 measurement points from it at both sides), while PARAFAC was constrained to non-negativity in all three modes. The modeling of the core of the data area is similar. The difference is that the peaks of the model in Fig. 5c continue a bit below zero at the base, thus providing continuous, smooth loadings, while the peaks of the model in Fig. 5d are abruptly cut off at zero because of the non-negativity constraint. In turn, this may lead to spurious peaks outside the data area, as a form of compensation. Replacing missing values outside the data area with zeros, while keeping PARAFAC unconstrained, thus appears to be a more gentle way of guiding the model towards non-negativity.

Figs. 6 and 7 are similar to Figs. 4 and 5, but here dataset II is under investigation. A big difference between the two datasets is that the number of wavelengths in the excitation and emission modes for dataset II is only half the number of wavelengths included in dataset I. The resolved spectra for the components of dataset II therefore appear less smooth than those of dataset I. From Fig. 6, it can clearly be seen that the conventional method gives excitation and emission spectra that are unacceptable from a photochemical point of view, as more than one emission peak appears at a shorter wavelength than its corresponding excitation peak. The other four results are all very similar, giving one component with negative score values, but all the emission spectra are close to unimodal, and the excitation spectra have only small negative values. Both excitation and emission spectra are smooth. The three last methods result in identical decompositions, while the second from the top is very similar to these. From an analytical point of view, it can be argued that only four factors should be included in these models, since two of the spectra in both the excitation and the emission mode behave similarly. However, since five is the estimated number of components [11] for this dataset, it has also been used here. None of the models in Fig. 6 are good since one of the score factors is negative, but for all the four last models, the two other modes seem reasonable, and thus these models will be evaluated as acceptable. It should

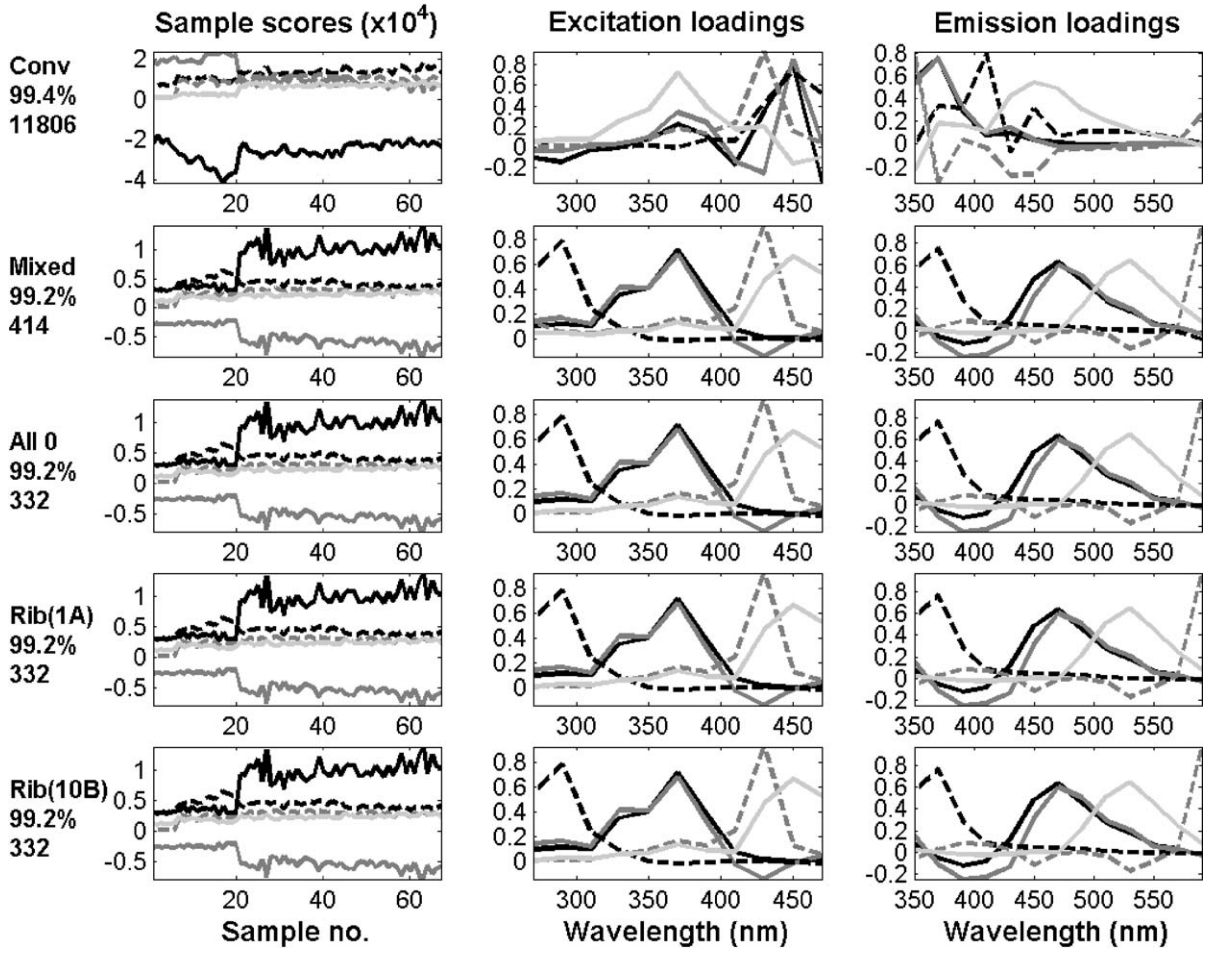

Fig. 6. Scores and loadings for five different unconstrained PARAFAC models for dataset II. The text to the left gives for each set of loadings the method used for insertion of zero values (line 1), the percent explained variance (line 2) and the number of iterations necessary before convergence was reached (line 3).

also be noted that modeling using the conventional method (missing values outside the data area) with three or four components also resulted in models where two components showed ‘‘complementary’’ or ‘‘mirrored’’ sample scores, i.e. each sample was modeled to ‘‘contain’’ either one or the other (i.e. to have a high score value for one and a low score value for the other). This is also seen in all five-factor models in Fig. 6, with one of the two ‘‘complementary’’ scores vectors being negative for all models, and thus this pattern appears to be unaffected by the replacement of missing values with zeros. We have also observed this pattern for PARAFAC modeling of fluorescence data from beet sugar processing juices (not published), and speculate that it may occur whenever a ‘‘discrete’’ PARAFAC model implying a few well-defined fluorophores is forced upon data of natural/biological origin, which in reality contains a continuum of only slightly different fluorophores, maybe because the same fluorescent groups are part of many different macromolecules. In other words, the negative scores vector is maybe a sign that PARAFAC modeling of dataset II is a simplification of a complex reality. Nevertheless, it may lead to something useful.

Fig. 7 shows that, for dataset II, the conventional method gives spurious peaks below the first order Rayleigh scatter line, where there should not be any peaks. The three other methods are all similar, showing all the same peaks. There are only minor differences in these predicted spectra, with the Ribbon (1A) method giving some small negative values in the spectra, and a small peak in the area below the first order Rayleigh scatter line.

Table 1 gives a condensed overview of the quality of the calculated models, evaluated from plots like those in Figs. 4–7. Each model was categorized as either ‘‘acceptable’’ or ‘‘unacceptable’’ based on the criteria mentioned in the introduction. No acceptable models where achieved at all for the datasets with missing values at all positions outside the Rayleigh scatter lines (the conventional method). For the pretreatments of the datasets with zeros inserted the PAR-

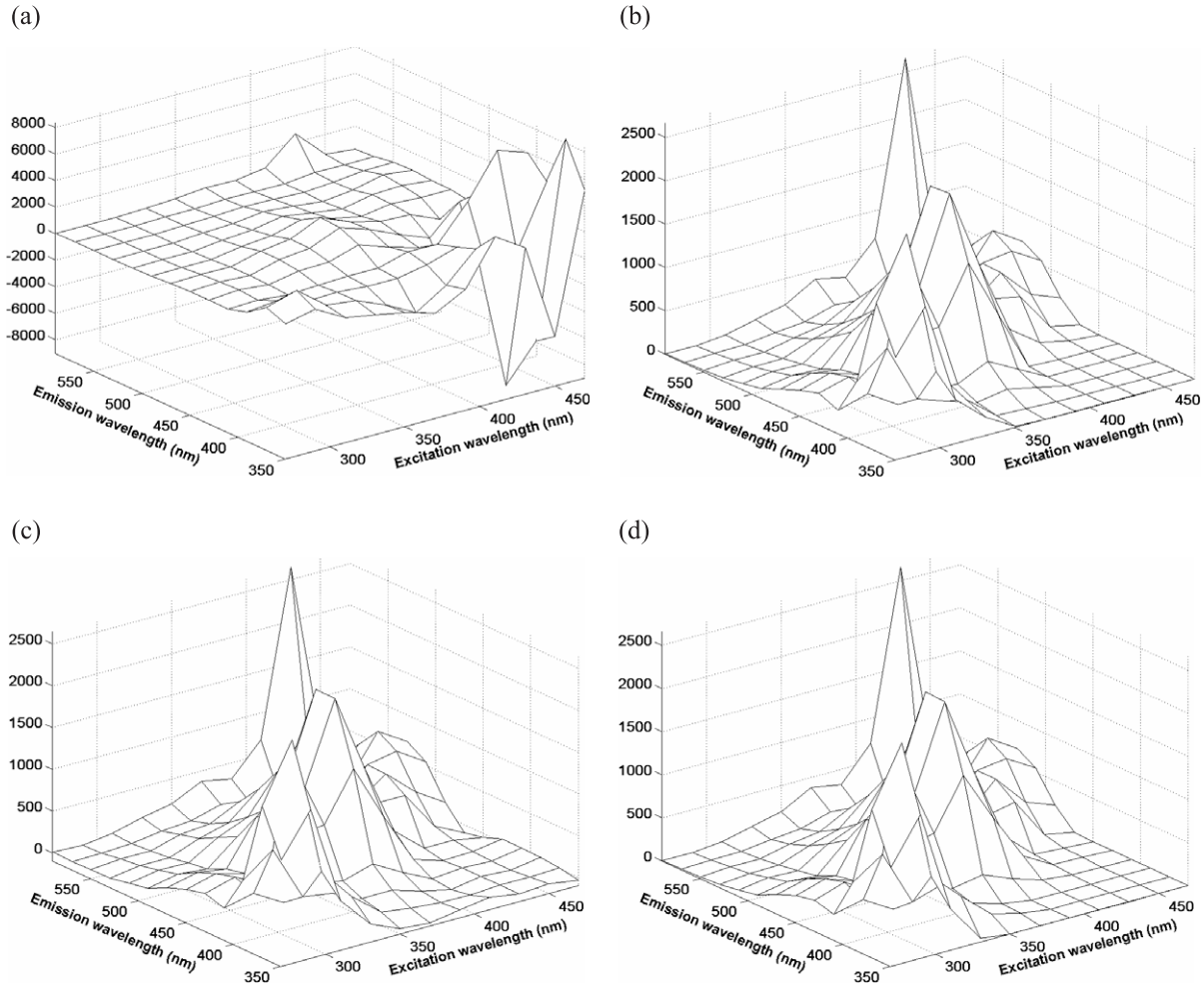

Fig. 7. Selected PARAFAC models of the fluorescence landscape of sample no. 50 from dataset II: (a) conventional method, unconstrained, (b) all zeros method, non-negativity constraint on all three modes, (c) ribbon (1A) method, unconstrained and (d) ribbon (10B) method, non-negativity constraint on all three modes.

Table 1
The results on the two datasets studied (I and II)

PARAFAC constraints	Conv		Mix		Zeros		Ribbon											
							1				5				10			
							A		B		A		B		A		B	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
None			x	x	x	x	x		x	x	x	x	x	x	x	x	x	x
Non-negativity, spectroscopic modes																		
Non-negativity, all modes			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Unimodality, spectroscopic modes																		

'x' means that the model was acceptable according to the criteria described in Section 1.1. An 'A' in the ribbon method means that the ribbon was inserted outside the first order Rayleigh scatter only and 'B' the ribbon was inserted on both sides of the data area.

AFAC constraints rather than the amount/placement of the zeros where decisive for the quality of the models. Unconstrained models and models with a non-negativity constraint on all three modes stand out as acceptable in contrast to models with a non-negativity or unimodality constraint on the spectroscopic modes only. The unconstrained models for all except the conventional method generally gave smoother loadings than the non-negativity constrained models for these methods, and fewer of them resulted in spurious peaks. The number of iterations before the model converged was only about 1/10 for the non-negativity constrained model relative to the corresponding unconstrained model—but the ratio in modeling time was more equal, as each iteration took more time for a non-negativity constrained modeling.

4. Conclusion

Inserting zeros instead of missing values in parts of the data area helps PARAFAC to converge faster, and leads to solutions that are physically and chemically meaningful. Insertion of zeros into the data matrices was also combined with constraining the PARAFAC model. Constraining only the spectroscopic modes had a detrimental effect on the model quality, while a non-negativity constraint on all three modes resulted in models that were essentially identical to the corresponding unconstrained models, but with less smooth loadings.

Future work lies in optimization of the method of inserting zeros. In this respect, several topics would be worth examining, such as evaluation of what kind of bandwidth of missing values to use for the ribbon method, and other, new and better methods of selecting the positions to insert zero values at. An investigation of different methods to initialize the PARAFAC model may also be appropriate.

Acknowledgements

Rinnan and Thygesen wish to thank the STVF (Danish Research Council) for financial support through project 1179. Barsberg wishes to thank the STVF (Danish Research Council) for financial support, and W. Kessler and R. Kessler, FH Reutlingen, Germany, for obtaining parts of the fluorescence data and for helpful discussions. The work on dry-cured hams was partly sponsored by an EU innovation program (OPUS, IN30905I), and J. Møller wishes to thank Dr. R. Bro for valuable advice during the preliminary data analysis.

References

[1] R. Bro, Chemometrics and Intelligent Laboratory Systems 38 (1997) 149–171.
[2] R. Bro, N.D. Sidiropoulos, A.K. Smilde, Journal of Chemometrics 16 (2002) 387–400.
[3] R.D. JiJi, K.S. Booksh, Analytical Chemistry 72 (2000) 718–725.
[4] R. Bro, Chemometrics and Intelligent Laboratory Systems 46 (1999) 133–147.
[5] C.M. Andersen, R. Bro, Journal of Chemometrics 17 (2003) 200–215.
[6] R. Bro, S. de Jong, Journal of Chemometrics 11 (1997) 393–401.
[7] R. Bro, N.D. Sidiropoulos, Journal of Chemometrics 12 (1998) 223–247.
[8] R. Bro, Multi-Way Analysis in the Food Industry-Models, Algorithms and Applications, PhD Thesis, Universiteit van Amsterdam, 1998, 135.
[9] C.A. Andersson, R. Bro, Chemometrics and Intelligent Laboratory Systems 52 (2000) 1–4.
[10] S. Barsberg, T. Elder, C. Felby, Chemistry of Materials 15 (2003) 649–655.
[11] J.K.S. Møller, G. Parolari, L. Gabba, J. Christensen, L.H. Skibsted, Journal of Agricultural and Food Chemistry 51 (2003) 1224–1230.