JOURNAL OF CHEMOMETRICS
J. Chemometrics 2001; **15**: 193–196

MEETING REPORT

TRICAP 2000: Three-way Methods in Chemistry and Psychology, Hvedholm Castle, Faaborg, Denmark, 16–21 July 2000

Johan A. Westerhuis[1]* and Marieke E. Timmerman[2]

[1]*Process Analysis and Chemometrics, University of Amsterdam, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, The Netherlands*
[2]*Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, NL-9712 TS Groningen, The Netherlands*

The TRICAP conferences focus on methods for three-way data analysis. Originating as a meeting for chemometricians and psychometricians (TRIC, Epe, The Netherlands, 1993) [1] and (TRICAP, Lake Chelan, WA, USA, 1997) [2], TRICAP 2000 was attended by researchers from signal processing and image analysis as well as chemical engineering. The conference aims at exchanging knowledge and findings in multi-way analysis and at stimulating international collaboration between researchers in multi-way methods. Since it is expected that three-way analysis will evolve even further, the next meeting (after several names had been discarded) was baptized TRIAD (ThRee-way analysis In All Disciplines). TRIAD will be organized in Kentucky, USA in the year 2003 by *Bill Rayens* and *Don Burdick*.

The local organizing committee of *Claus Andersson* and *Rasmus Bro* managed to organize a high-standard meeting in the beautiful Hvedholm Castle located in the southern part of the Danish island of Fyn. A total of 23 oral presentations were given, each of 45 min plus additional discussion time, on a broad range of topics such as curve resolution, classification, Boolean algebra, chemical process data and new notational schemes. Furthermore, three posters were presented. A small introduction on chemical terminology (for the non-chemists) and a presentation on the Matlab compiler were given. New methods were introduced and chemical as well as psychological applications were presented. Many discussions concerned the type of data and corresponding data preprocessing in the different fields. Different types of constraints were used to restrict the models. This seems to work especially well for chemical data where constraints represent hard chemical information. The conference program offered ample time for informal discussions. The attendees took their opportunities to exchange knowledge and ideas in a good atmosphere. The different viewpoints resulting from the different backgrounds of the researchers opened up interesting new perspectives. A brief overview of the presentations is given below.

Paul Geladi started by correcting his statement of the previous meeting by stating: 'Cross-validation does not suck'. He then presented the use of PARAFAC models for curve resolution

* Correspondence to: J. A. Westerhuis, Process Analysis and Chemometrics, University of Amsterdam, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, The Netherlands.

Copyright © 2001 John Wiley & Sons, Ltd.

although from the start it was known that these data did not have a trilinear structure. Even in such cases PARAFAC does a good job, since it gives reasonable estimates for pure spectra of the chemical compounds. If the proper preprocessing is used, PARAFAC models can give much better results than using two-way methods. His final conclusion was that using two-way methods for three-way data really sucks. *Roma Tauler* described the problem of rotational ambiguity in curve resolution. A rotation matrix is needed to rotate the solution found into the chemically 'true' solution. This rotation matrix is used to calculate maximum and minimum feasible boundaries around the estimated pure spectra. Different types of constraints, such as non-negativity, selectivity, unimodality or trilinearity constraints, or known profiles of some of the chemical compounds can be used to narrow the feasible bands. *Steve Gurden* showed the use of grey models in modeling spectroscopic batch process data. In the white part of the grey model, all external chemical information known of the process is incorporated, such as pure spectra and kinetic profiles, and the remaining systematic variation that is not completely understood is left for the black part. It became apparent that using chemical information by applying constraints really improves the interpretation of such chemical data sets. *Sungjin Hong* described models for shifted factor analysis (SFA), in which the factors are allowed to be shifted differently in (for example) time for different variables. He discussed quasi-alternating least squares algorithms to estimate the parameters of the models for SFA, in the two-way as well as the N-way case.

Nikos Sidiropoulos discussed a generalization of Kruskal's proof of the uniqueness of low-rank decomposition of three-way arrays to the case of low-rank decomposition of N-way arrays. Furthermore, he generalized a result of Carathéodory to N dimensions. The latter were related to the generalization of Kruskal's proof. *Bill Rayens* employed the theory of Kronecker canonical forms for matrix pencils to decompose the $I \times J \times 2$ array into a minimal sum of rank-one tensors. The decomposition produces a regular part, which is typically unique, and a singular part, which is arbitrary. Algorithms to separate the two parts were discussed. *Lieven De Lathauwer* introduced independent component analysis (ICA) as a fine-tuning of principal component analysis for higher-order data, but without the problem of rotational invariance. Using higher-order statistics, third- and fourth-order cumulants, which are tri- and quadrilinear for bilinear data, the rotational indeterminacy can be set off. Furthermore, these cumulants are zero for Gaussian noise. ICA is therefore able to give the exact underlying profiles hidden in the data.

Henk Kiers presented approaches for modeling three-way data that consist of offset terms, a three-way model (e.g. Tucker3 model) and error. He discussed conditions for which the offset terms can be uniquely determined and approaches to fit the offset terms and parameters of the model. On the basis of a simulation experiment he concluded that his procedure for simultaneous fitting of offset terms and the parameters of a three-way model often performed better than simple fitting or successive fitting. *Marieke Timmerman* considered a hierarchically ordered class of simultaneous component models for multivariate time series of more than one subject. Both intra-individual variability and inter-individual variability are covered in the models. Additionally, she treated the use of smoothness constraints and imposing latent curves in the Tucker3 model. *Jos ten Berge* presented results concerning the simplicity of transformed three-way arrays, which is useful to obtain simplified core arrays in the Tucker3 model. He extended previous results on transformations to simplicity by showing that the simplifying transformation of an $I \times J \times K$ array can also be used to simplify the complementary array of order $(JK - 1) \times J \times K$.

Maurizio Vichi introduced methods for the consensus classification of three-way dissimilarity data. The methods aim at simultaneously finding classifications for each of the elements in the third mode and finding a consensus between the several classifications. The methods differed in the way of classification. The multiple INDCLUS model, the partitions-of-partitions (POP) approach and the PARLSCLA method (hierarchical modeling using dendograms) were treated. *El Mostafa Qannari*

Copyright © 2001 John Wiley & Sons, Ltd.

proposed a clustering approach to analyze several two-way data sets that contain the scores of the same subjects. The method aims at forming clusters of homogeneous data sets. For each cluster the relationships between the individuals are depicted on the basis of a small number of components.

Eva Ceulemans presented the Tucker hierarchical class model (Tucker-HICLAS) for binary three-way three-mode data. The three-way data are decomposed into three bundle matrices that represent the equivalence and hierarchy relations among the elements of each mode, and a core array that represents the ternary relations between the bundle matrices. The algorithm to fit the model to data was tested by means of a simulation experiment, and an empirical application was presented. *Iven van Mechelen* discussed two classes of models for binary three-way three-mode data, namely the standard linear models and their Boolean counterparts. In particular, the PARAFAC and Tucker3 models and the INDCLAS and Tucker-HICLAS models were treated as representatives of the model classes. In both types of models the three-way data are decomposed. The Boolean models also include hierarchically organized classifications of the elements of each model. A theoretical and empirical comparison was offered.

Pieter Kroonenberg presented an overview of the diffusion of three-mode analysis in chemistry. This was studied by a citation analysis of several key publications.

A whole day was spent discussing chemical engineering process data. *Dora Kourti* started with an introduction on statistical process control and described the specific type of data one has to deal with in monitoring chemical batch processes and grade transitions. Such data usually have no trilinear structure. An interesting discussion started on how to preprocess batches that do not have equal duration. *Neal Gallagher* introduced the problem of end-point detection of batch reaction processes. The main problem, however, was that the process was not stationary and therefore the models had to be adjusted for each new batch. Various types of models were used and the PARAFAC model was sufficient in all cases, and was especially advantageous when the detector signal was weak.

In order to elucidate the crystalline structure of polymers, a series of small-angle X-ray scattering (SAXS) images are obtained while the polymer is stretched. *Barry Wise* presented the analysis of these images using co-ordinate transformations and morphing. Each image should theoretically be a summation of bilinear terms, and the series of images should be modeled using a PARAFAC model. However, owing to the change in chemistry of the pure components when the polymer is stretched, the PARAFAC loadings also change.

Multi-block methods were introduced by *Mike Piovoso*. These methods can give improved interpretation when the data are divided into conceptually meaningful blocks. An application of separating the batch process data in blocks of different variables or in blocks of different batch stages improved the detection of bad batches in a Nylon production reaction. Furthermore, information on which process variables are of importance in the various stages of the process becomes easier to detect. The ideas of multi-block models were extended by *Johan Westerhuis* who presented multi-way multi-block models based on covariate regression methods. Such models can be used to combine two-way and multi-way blocks for prediction as well as for exploratory analysis. *Agnar Höskuldsson* discussed the importance of variable selection by looking at the covariance between the predictor variables and the response and model building by a stepwise approach where the product of fit and precision is optimized.

It became apparent that a new notation scheme is needed for multi-way algebra, since writing multi-way expressions in matrix notation may complicate expressions of multi-way arrays. Two new types of notation were introduced. *Don Burdick* demonstrated the visually orientated ball-and-whisker diagrams where the ball represents any array and the whiskers represent the different modes of the array. Different operations in matrix algebra can be diagrammed by dot-knotting, star-braiding, m-joining, cloning, etc. The main advantage of such a notation is the visualization of complex matrix operations. The problem, however, is that it will be difficult to build such a notation scheme into a

Copyright © 2001 John Wiley & Sons, Ltd.

user-friendly software package. *Richard Harshman* introduced the Array Index Notation scheme where a_{IJKL} represents a four-way array with modes I, J, K and L. Such a notation is easier to use within a word-processor than the visual ball-and-whisker one. A very useful comment from the audience was that instead of trying to force all matrix operations in such a new notational framework, the notation should be used to discover new operations in multi-way algebra.

We conclude that the local organizers *Claus Andersson* and *Rasmus Bro* did a great job. The whole conference, including the nice trip to Egeskov Castle (which gave an impression of the beautiful landscape of Fyn island) and the conference dinner (where the waiters were probably educated by John Cleese, which led to hilarious situations for almost everybody), was very successful. We hope that the organizers of TRIAD 2003 will be able to maintain the high standard of these conferences.

REFERENCES

1. Geladi P. Meeting report TRIC 1993. *J. Chemometrics* 1994; **8**: 97–98.
2. Bro R, Kiers HAL. Meeting report TRICAP 1997. *J. Chemometrics* 1997; **11**: 463–466.

Copyright © 2001 John Wiley & Sons, Ltd.