JOURNAL OF CHEMOMETRICS
J. Chemometrics 2001; **15**: 19–27

Comments on three-way analyses used for batch process data

Age K. Smilde*

Process Analysis and Chemometrics, Department of Chemical Engineering, University of Amsterdam, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, The Netherlands

SUMMARY

Recently, several papers have appeared concerning the use of three-way models for batch process data. In these papers a number of points are raised. This paper discusses some of these points and illustrates some pitfalls. More specifically, some theoretical aspects of using different three-way models for batch process data and some practical consequences are discussed. The topics of cross-validation and data preprocessing are also discussed. These issues will be discussed using small simulated examples and theoretical arguments. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: batch process models; MSPC; three-way analysis; cross-validation

INTRODUCTION

Batch processes play an important role in the process industry. Owing to their non-steady state and (highly) non-linear behavior, batch processes are more difficult to control than continuous processes. The pioneering work of Nomikos and MacGregor provided a basis for multivariate statistical process control (MSPC) of batch processes [1,2]. This approach has been used successfully by others [3,4] and will be abbreviated here to the NM approach.

The NM approach consists of two steps. In the first step an empirical model is made of the batch process data using measurements of batch runs obtained under normal operating conditions. Control charts are built using this model. In the second step a new running batch is monitored and the on-line measurements are projected onto the model, i.e. these measurements are plotted in the control chart. Out-of-control situations can then be detected. One of the crucial factors in this procedure is the quality of the empirical model used for modeling the batch process data.

The data used to build a batch MSPC model are of a three-way nature. If J measurements are performed at K time points on I batches, then an $I \times J \times K$ array results. Several models are available to fit such data. In the NM approach a Tucker1 (or MPCA or unfold-PCA) model is used. There are alternatives, e.g. CANDECOMP/PARAFAC and Tucker models [5–8]. Recently, papers have appeared that discuss these alternative models [9–13]. For convenience, Dahl *et al* [9] will be referred to as the DPK paper and Westerhuis *et al* [11] as the WKM paper. These papers raise interesting

* Correspondence to: A. K. Smilde, Process Analysis and Chemometrics, Department of Chemical Engineering, University of Amsterdam, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, The Netherlands.
E-mail: asmilde@its.chem.uva.nl

Received 20 August 1999
Accepted 1 February 2000

points and some of these will be addressed. Specifically, some theoretical aspects of using different three-way models for batch process data and practical consequences will be discussed.

Some attention will be given to the preprocessing of batch process data and cross-validation of the three-way models, because these issues are not straightforward.

THEORETICAL ASPECTS OF THREE-WAY MODELS FOR BATCH PROCESS DATA

Philosophy of using three-way models for batch process data

In the first papers of Nomikos and MacGregor concerning building empirical models for batch process data, ample discussion is provided on alternative methods to build batch process models [1,2]. Such alternatives are e.g. fundamental models based on first principles (i.e. in the form of differential equations) or knowledge-based models (using expert systems).

The fundamental idea of using empirical models for batch process data is that with the use of the proper preprocessing method (see later) the non-linear behavior of the measured process variables is removed. The remaining variation is then modeled using an empirical model. This means that the part of the variation which can be modeled using differential equations describing the non-linear behavior is removed. This is analogous to the way in which fundamental models for continuous processes are often used. It is customary to linearize such models around the steady state value of the process to obtain a (local) linear model [14]. Exactly the same is done in batch process modeling with three-way models. The proper preprocessing acts as a 'steady state' and the deviations around this steady state are described with multilinear models.

Notation

In the following a scalar will be written as an italic lower-case character (x), a vector as a bold lower-case character ($\mathbf{x}$), a matrix as a bold upper-case character ($\mathbf{X}$) and a three-way array as an underlined bold upper-case character ($\underline{\mathbf{X}}$). This is consistent with the notation proposed by Kiers [15]. The matrix $\underline{\mathbf{X}}$ will be used to indicate a three-way array containing batch process data, with dimensions $I \times J \times K$, where I is the number of batches and J is the number of process variables measured at K points in time. The matrix $\mathbf{Y}$ ($I \times M$) will be used to indicate the M measured product qualities of the batch runs.

An operation often performed on a three-way $\underline{\mathbf{X}}$ is rearranging such an array into a matrix (see Figures 1 and 2). This is called unfolding in the chemometrics literature [16], but to avoid confusion with the psychometrics literature, the term matricizing will be used here [15]. For details concerning the matricizing operation, see References [15,17].

Available three-way models

There are different types of three-way models. The two main types are PARAFAC models and Tucker models. The three-way array $\underline{\mathbf{X}}$ can be modeled with a PARAFAC model if each element x_{ijk} of $\underline{\mathbf{X}}$ is written as

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk} \tag{1}$$

where a_{ir}, b_{jr} and c_{kr} are elements of the loading matrices $\mathbf{A}$ ($I \times R$), $\mathbf{B}$ ($J \times R$) and $\mathbf{C}$ ($K \times R$) respectively, e_{ijk} is the residual and R is the number of PARAFAC components. This will be denoted as a PARAFAC(R) model and is also sometimes called a trilinear model.

A Tucker3 model of $\underline{\mathbf{X}}$ can be written as

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{kr}g_{pqr} + e_{ijk} \tag{2}$$

where a_{ip}, b_{jq} and c_{kr} are elements of the loading matrices $\mathbf{A}$ ($I \times P$), $\mathbf{B}$ ($J \times Q$) and $\mathbf{C}$ ($K \times R$) respectively, g_{pqr} is an element of the core array $\underline{\mathbf{G}}$ ($P \times Q \times R$), e_{ijk} is the residual and P, Q and R are the numbers of Tucker components in the three modes. This will be denoted a Tucker (P,Q,R) model. Note that $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and e_{ijk} are not the same as in the PARAFAC model, but for convenience the same symbols are used. Equation (2) represents the Tucker3 model, since all the modes are reduced. It is also possible to reduce two modes:

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q} a_{ip}b_{jq}g_{pqk} + e_{ijk} \tag{3}$$

where a_{ip} and b_{jq} are elements of the loading matrices $\mathbf{A}$ ($I \times P$) and $\mathbf{B}$ ($J \times Q$) respectively, g_{pqk} is an element of the extended core array $\underline{\mathbf{G}}$ ($P \times Q \times K$), and e_{ijk} are the residuals. Model (3) is called the Tucker2 model, since only two modes (the first and the second) are reduced. Note that there are two alternative Tucker2 models, i.e. reducing modes 1 and 3 and reducing modes 2 and 3.

It is even possible to reduce only one mode. For example, if only the first mode is reduced, then a Tucker1 (P) model is obtained:

$$x_{ijk} = \sum_{p=1}^{P} a_{ip}g_{pjk} + e_{ijk} \tag{4}$$

where a_{ip} is an element of the loading matrix $\mathbf{A}$ ($I \times P$), g_{pjk} is an element of the extended core array $\underline{\mathbf{G}}$ ($P \times J \times K$), and e_{ijk} is the residual. The core array $\underline{\mathbf{G}}$ is often represented by the matrices $\mathbf{G}_p$ ($J \times K$), where $\mathbf{G}_p$ is the pth slice of $\underline{\mathbf{G}}$. The Tucker1 model is often used in chemometrics and is also used in the NM approach. An alternative name for the Tucker1 model is multiway PCA (MPCA) or unfold-PCA. Note that if each $\mathbf{G}_p$ is subjected to a PCA and only the first component (scores and loadings) is used, then (4) resembles a PARAFAC model. This is not the optimal way of fitting the PARAFAC model, which is done better by using proper algorithms for fitting (1) directly.

Hierarchical relations between three-way models

Some important observations can be made using Equations (1)–(4); see e.g. Reference [18]. Upon setting $P = Q = R$, $g_{rrr} = 1$ and all other elements of $\underline{\mathbf{G}}$ to zero, it is clear that a PARAFAC(R) model is a special case of a Tucker3 (R,R,R) model. Stated otherwise: a PARAFAC(R) model can be seen as a restricted Tucker3 (R,R,R) model. This means that for a given three-way array a PARAFAC(R) model will always fit worse than a Tucker3 (R,R,R) model. Likewise, by comparing a Tucker1 (P) model with a Tucker3 (P,Q,R) model, it is clear that the Tucker3 (P,Q,R) model forces extra structure on the three-way array $\underline{\mathbf{G}}$ ($P \times J \times K$) of the Tucker1 (P) model. Hence the Tucker1 (P) model always fits better than the Tucker3 (P,Q,R) model. Summarizing, in the order PARAFAC(P), Tucker3 (P,P,P), Tucker1 (P) the fit always improves. This is comparable to the fact that in the order first-degree polynomial, second-degree polynomial, third-degree polynomial the fit also always improves. Hence fit is not the only criterion for selecting between models, as was already remarked by WKM.

Table I. Singular value analysis of **G** matrices

	No correlation in **B** and **C**	Severe correlation in **B** and **C**
$\mathbf{G}_1$ (6 × 6)	0·9922; 0·1251; 0; 0; 0; 0	1·0000; 0·0001; 0; 0; 0; 0
$\mathbf{G}_2$ (6 × 6)	0·7991; 0·6012; 0; 0; 0; 0	0·7458; 0·6662; 0; 0; 0; 0

Modeling interactions

The main theoretical argument by WKM for using a Tucker1 model is that such a model allows for modeling the interactions between the time and variable dimensions of an array of batch process data. Such interactions are indeed present in many batch process data. WKM use the rank of $\mathbf{G}_p$ to show that interactions are present and that hence a Tucker1 model is appropriate. WKM state that if the rank of $\mathbf{G}_p$ is higher than one, then interactions are present. Whereas it might be true that interactions give rise to $\mathbf{G}_p$s of ranks higher than one, the reverse is not true: if a $\mathbf{G}_p$ has a rank higher than one, this does not necessarily mean that interactions are present. This is illustrated with a small example.

Suppose that $\underline{\mathbf{X}}$ has a PARAFAC(2) structure with component vectors $\mathbf{a}_1$, $\mathbf{a}_2$ (first mode), $\mathbf{b}_1$, $\mathbf{b}_2$ (second mode) and $\mathbf{c}_1$, $\mathbf{c}_2$ (third mode). Applying a Tucker1 model on those data would result in two loading matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ both of rank two. Each loading matrix in the Tucker1 model tries to model the subspace generated by the **B** and **C** loadings. A simple simulated example shows the problem. The vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{c}_1$ and $\mathbf{c}_2$ (all 6×1) were randomly chosen and a three-way array $\underline{\mathbf{X}}$ was formed using these loadings. No noise was added. A PARAFAC(2) model was fitted to $\underline{\mathbf{X}}$, and the loadings could be perfectly recovered (apart from trivial scaling and permutation freedom). A Tucker1 model was also applied to $\underline{\mathbf{X}}$, and this resulted in loading matrices $\mathbf{G}_1$ and $\mathbf{G}_2$. An SVD on $\mathbf{G}_1$ and $\mathbf{G}_2$ reveals the rank of those matrices. The singular values are shown in the second column of Table I. The conclusion is that the loading matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ from the Tucker1 model do not have rank one.

In order to understand the behavior of a Tucker1 model performed on trilinear data, the loadings in **B** and **C** were chosen to be very correlated. Note that in the random case above the correlation within all loading matrices was small. In the correlated simulation case the correlation coefficient between $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{c}_1$, $\mathbf{c}_2$ is 0·999, which is quite severe. A PARAFAC(2) model correctly reproduces the components, although at the cost of many iterations. A Tucker1 analysis gives again loading matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ which can be subjected to an SVD for rank analysis. The singular values of $\mathbf{G}_1$ and $\mathbf{G}_2$ are shown in the third column of Table I. Clearly, the first Tucker1 component captures the common variation in $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{c}_1$, $\mathbf{c}_2$, whereas the second component describes the (small) differences. Note that it is always possible to build a Tucker1 structure using a rank-one $\mathbf{G}_1$ and a rank-two $\mathbf{G}_2$. Analyzing this data set with the proper model (Tucker1) can give estimated loading matrices with the same singular values as in the third column of Table I.

Summarizing, if rank analyses of $\mathbf{G}_1$ and $\mathbf{G}_2$ show that these matrices do not have rank one, then a perfect trilinear structure can still be present. That is, Tucker1 loading matrices of rank higher than one do not indicate that complicated interactions between the time and variable modes are present in the batch process data. On the other hand, if $\mathbf{G}_1$ has rank one, then this does not prove that the underlying model is trilinear. The final conclusion is that rank analyses of loading matrices of Tucker1 models give very limited information as to what is the true underlying model of the batch process data set.

It is necessary to carefully analyze what kind of interactions are present in batch process data. WKM mention three types of interactions.

1. Process variables are not only correlated to each other at any given time, but also correlated to events that happen at earlier times in the process (autocorrelation).
2. The nature of the correlation of 1 can change during the batch run (change of autocorrelation).
3. The nature of the correlation among the variables may change over the course of the batch (changing cross-correlation).

All three-way models try to model these interactions. The PARAFAC model adds extra components, similar to PCA in two-way analysis. From two-way analysis it is known that this works for a small amount of interaction.

The Tucker1 model does not reduce the modes for the variables and the time points. Hence it uses all combinations of variables and time points and tries to model all those combinations simultaneously, including all the interactions. Stated differently, Tucker1 models the interactions on the level of the *manifest* (measured) variables. A Tucker1 model completely mixes interactions of types 1–3. This may or may not be desirable.

Tucker3 models the interactions by allowing the latent variables to interact via the core array. Hence interaction is modeled on the level of *latent* variables. If the variations in batch processes are driven by events that influence a set of measured variables, then modeling such an event via a latent variable might be sensible. This is the central idea of latent variables: to model underlying events that influence all or a set of manifest (measured) variables.

Summarizing, a clear understanding of what types of interactions are present in batch process data and diagnostics to detect such interactions are needed. It is not possible to give an *a priori* judgement about which three-way model is the best in fitting batch process data.

PRACTICAL ASPECTS OF THREE-WAY MODELS FOR BATCH PROCESS DATA

Degeneracies in PARAFAC models

In the DPK paper a batch polymerization data set is analyzed. The batch polymerization has five different stages. Temperatures (23) were measured for 39 batches at 109 time points. This results in a $39 \times 23 \times 109$ three-way array which can be analyzed. DKP compare Tucker1 and PARAFAC models for this data set. They encounter degeneracies for PARAFAC models with three or more components. The degeneracy consists of high and persistent correlation between two of the PARAFAC components which cancel each other in the overall fit of the model. This occurs e.g. if $\mathbf{a}_1 \approx -\mathbf{a}_2$, $\mathbf{b}_1 \approx \mathbf{b}_2$ and $\mathbf{c}_1 \approx \mathbf{c}_2$. Then the triple products of $(\mathbf{a}_1,\mathbf{b}_1,\mathbf{c}_1)$ and $(\mathbf{a}_2,\mathbf{b}_2,\mathbf{c}_2)$, which are the contributions of the first and second components respectively to the total PARAFAC model, cancel each other.

Degeneracies can be avoided by putting restrictions on the PARAFAC components, e.g. making the components in one of the modes orthogonal to each other. One of the possible causes of degeneracy can be that the three-way array at hand has a Tucker3 structure [19], as was already mentioned by DKP. A very similar data set was analyzed by Boqué and Smilde [12] where it was concluded that a Tucker3 (3,2,3) model was appropriate for these data. Hence the data of this batch process probably have a Tucker3 structure.

Cross-validation

One of the ways of establishing the number of significant components in a PCA of $\mathbf{X}$ is by using cross-validation [20,21]. It is tempting to use cross-validation to establish the model complexity of three-way models of batch process data, i.e. the number of components in a PARAFAC, Tucker1 or Tucker3 model and for comparing different three-way models on the same data. The number of

components in a Tucker1 model was established by cross-validation in the NM approach, and WKM use cross-validation to estimate the predictive abilities of the compared three-way models. There is, however, one serious problem with their implementation of cross-validation: the predicted and measured values which are compared to calculate the prediction error (PRESS) are not independent. This violates a fundamental assumption of cross-validation [22].

The NM approach for cross-validating three-way batch data is as follows (schematically).

(i) Leave out one batch.
(ii) Autoscale the data.
(iii) Calculate the model parameters (loadings).
(iv) Scale the left-out batch.
(v) Use the left-out batch and model parameters to calculate the new score of the batch.
(vi) Use the calculated score (from (v)) and the loadings (from (iii)) to calculate the predicted batch values.
(vii) Calculate the sum of squares of predicted values minus measured values (PRESS).

The problem with this procedure is that the predicted and measured values in step (vii) are not independent of each other, since in (v) the left-out batch is used to calculate the new score, which is then in turn used to calculate predicted batch values.

There are different generalizations of cross-validation to multiway models [23]. The Eastment and Krzanowski [20] approach can be generalized, and an approach based on handling missing data is also possible. In both approaches the predicted and measured values used to calculate PRESS are completely independent. Whereas the NM way of performing cross-validation might be good enough for establishing the number of components in a Tucker1 model, this cross-validation method has problems, and using the generalized Eastement and Krzanowski approach may lead to different results.

Establishing the number of components in the three modes of a Tucker3 model is not easy. There are different approaches, and one approach is by performing singular value decompositions on the three matricized arrays, making scree plots of these singular values and selecting per mode how many components to retain. This is the way WKM selected the Tucker3 model for their data. This approach works sometimes, but in the case of noisy three-way arrays it breaks down [23]. In some cases, cross-validation is indeed able to recover the correct number of Tucker components in noisy three-way arrays [23].

Comparing three-way methods for batch process data

Research on batch process modeling using three-way models has started only recently. Some comparisons between different three-way models have been made [9–13,24]. No clear conclusions can be drawn yet as to which three-way model is usually the best for batch process data. It is even difficult to define what 'best' is, because three-way models for batch data might serve different purposes:

1. *Process optimization*—building a regression model between $\underline{\mathbf{X}}$ and $\mathbf{Y}$ to investigate the type of variation present in the data. This may lead to changing the recipe of the batch.
2. *Process monitoring*—building control charts based on a three-way model of $\underline{\mathbf{X}}$.
3. *Product quality prediction*—building a regression model between $\underline{\mathbf{X}}$ and $\mathbf{Y}$ which can be used to predict $\mathbf{Y}$ from $\underline{\mathbf{X}}$ for future (running) batches.

This is not an exclusive set of goals; it is e.g. also possible to use an ($\underline{\mathbf{X}}$, $\mathbf{Y}$) regression model as the basis for monitoring charts [12,25]. If process optimization is the goal, then interpretation of model

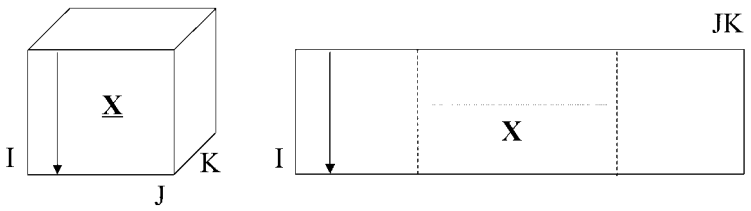

Figure 1. The three-way array X ($I \times J \times K$) and one of its matricized versions X ($I \times JK$). The arrow indicates the direction of the centering or scaling operation: across one mode (within a fiber of the array)

parameters is important. The PARAFAC model gives the highest degree of interpretability, because these models have the smallest number of parameters (for a given number of latent variables).

If process monitoring is the goal, then the ability to detect faults is important. One quality criterion to judge such a performance is the average run length, i.e. how long it takes to detect a certain disturbance [26]. The experience in this area with different three-way models is limited but shows that average run lengths are dependent on the type of fault and the three-way model used [13].

Preprocessing of batch process data

Preprocessing of batch process data is important because it can affect to a large extent the modeling of such data. Moreover, the basic philosophy of using empirical models for batch process data requires a proper kind of preprocessing. In this paper, only the centering and scaling operation of batch process data will be considered. Batch process data are a special case of three-way data, and therefore knowledge of preprocessing three-way data is valuable [27] (R. Bro and A. K. Smilde, submitted).

Assuming that a three-way array has a multilinear structure, it can be shown that centering performed across one mode (Figure 1) does not destroy this structure. Centering performed across several modes simultaneously (Figure 2) destroys the multilinear structure [27]. It is important to realize that this argument holds not only for trilinear models (PARAFAC) but also for all kinds of multilinear models (PARAFAC, Tucker2, Tucker3).

In the NM approach, centering is performed across the batch mode (Figure 1), and this centering operation retains the multilinear structure of the batch process data. This centering operation is sensible because it removes the non-linear process variable trajectories, thereby focusing the three-way model on deviations from such trajectories (see 'Philosophy of using three-way models for batch process data'). This is the recommended way of centering batch process data. An example of centering according to Figure 2 is given by Wold *et al* [28] and Wise *et al* [10], and it is not clear to what extent this influences the results. There are indications that centering according to Figure 2 can

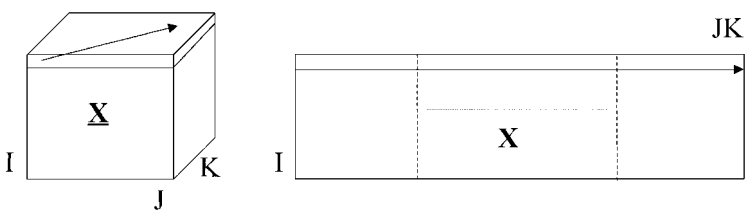

Figure 2. The three-way array $\underline{\mathbf{X}}$ ($I \times J \times K$) and one of its matricized versions X ($I \times JK$). The arrow indicates the direction of the centering or scaling operation: within one mode (across a slab of the array)

have a severe effect on the result of the three-way analysis, e.g. introducing spurious components in the multilinear models [27] (R. Bro and A. K. Smilde, submitted).

Scaling can also destroy the multilinear structure of three-way data. Scaling within one mode (Figure 2) does not destroy the multilinear structure, whereas scaling within two modes (Figure 1) does destroy such a structure. In the original approach of NM the scaling of Figure 1 was used, and this was followed by several others [4,12]. Although in some cases the type of scaling can make some difference [10], there are also indications that batch process data are not extremely sensitive to the type of scaling [2,11].

If three-way models are compared on a given set of batch process data, a wrong kind of preprocessing can favor one method over another. Destroying a multilinear structure (by using an improper kind of centering and scaling) favors the Tucker1 model of the data, because this model does not assume multilinear data.

CONCLUSIONS

In the area of modeling batch process data with three-way models, a lot of questions are still unanswered. Some pitfalls have been discussed and some theoretical arguments have been given. Practice has to show which three-way method is most suitable for the batch process data at hand. This depends also on the goal of the empirical batch process model.

REFERENCES

1. Nomikos P, MacGregor JF. Monitoring batch processes using multiway principal component analysis. *AIChE J.* 1994; **40**: 1361–1375.
2. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; **37**: 41–59.
3. Kosanovich KA, Dahl KS, Piovoso MJ. Improved process understanding using multiway principal component analysis. *Ind. Engng Chem. Res.* 1996; **35**: 138–146.
4. Neogi D, Schlags CE. Multivariate statistical analysis of an emulsion batch process. *Ind. Engng Chem. Res.* 1998; **37**: 3971–3979.
5. Harshman RA. Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multimodal factor analysis. *UCLA Working Papers Phonet.* 1970; **16**: 1–84.
6. Carroll JD, Chang J. Analysis of individual differences in multidimensional scaling via an *N*-way generalization of 'Eckart–Young' decomposition. *Psychometrika* 1970; **35**: 283–319.
7. Tucker LR. In *Contributions to Mathematical Psychology*, Frederiksen N, Gulliksen H, (eds). Holt, Rinehart and Winstor: New York, 1964; 110–182.
8. Smilde AK. Three-way analyses: problems and prospects. *Chemometrics Intell. Lab. Syst.* 1992; **15**: 143–157.
9. Dahl KS, Piovoso MJ, Kosanovich KA. Translating third-order data analysis methods to chemical batch processes. *Chemometrics Intell. Lab. Syst.* 1999; **46**: 161–180.
10. Wise BM, Gallagher NB, Watts Butler S, White DD, Barna GG. A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *J. Chemometrics* 1999; **13**: 379–396.
11. Westerhuis JA, Kourti T, MacGregor JF. Comparing alternative approaches for multivariate statistical analysis of batch process data. *J. Chemometrics* 1999; **13**: 397–413.
12. Boqué R, Smilde AK. Monitoring and diagnosing batch processes with multiway regression models. *AIChE J.* 1999; **45**: 1504–1520.
13. Louwerse DJ, Smilde AK. Multivariate statistical process control of batch processes using three-way models. *Chem. Engng Sci.* 2000; **55**: 1225–1235.
14. Stephanopoulos G. *Chemical Process Control: an Introduction to Theory and Practice*. Prentice-Hall: Englewood Cliffs, NJ, 1984.
15. Kiers HAL. Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics* 2000; **14**: 105–122.

16. Wold S, Geladi P, Esbensen KH, Öhman J. Multi-way principal components and PLS-analysis. *J. Chemometrics* 1987; **1**: 41–56.
17. Geladi P. Analysis of multi-way (multi-mode) data. *Chemometrics Intell. Lab. Syst.* 1989; **7**: 11–30.
18. Kiers HAL. Hierarchical relations among three-way methods. *Psychometrika* 1991; **56**: 449–470.
19. Kruskal JB, Harshman RA, Lundy ME. In *Multiway Data Analysis*, Coppi R, Bolasco S (eds). Elsevier: Amsterdam, 1989; 115–122.
20. Eastment HT, Krzanowski WJ. Cross-validatory choice of the number of components from a principal component analysis. *Technometrics* 1982; **24**: 73–77.
21. Wold S. Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics* 1978; **20**: 397–405.
22. Stone M. Cross-validatory choice and assessment of statistical predictions. *J. R. Statist. Soc B* 1974; **36**: 111–148.
23. Louwerse DJ, Kiers HAL, Smilde AK. Cross-validation of multiway component models. *J. Chemometrics* 1999; **13**: 491–510.
24. Smilde AK, Kiers HAL. Multiway covariates regression models. *J. Chemometrics* 1999; **13**: 31–48.
25. Nomikos P, MacGregor JF. Multiway partial least squares in monitoring batch processes. *Chemometrics Intell. Lab. Syst.* 1995; **30**: 97–108.
26. Wetherill GB, Brown DW. *Statistical Process Control: Theory and Practice*: Chapman and Hall: London, 1991.
27. Harshman RA, Lundy ME. In *Research Methods for Multimode Data Analysis*, Law HG, Snyder Jr CW, Hattie J, McDonald RP (eds). Praeger: New York, 1984; 216–284.
28. Wold S, Kettaneh-Wold N, Friden H, Holmberg A. Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics Intell. Lab. Syst.* 1998; **44**: 331–340.

