JOURNAL OF CHEMOMETRICS
J. Chemometrics 2001; **15**: 101–121

Modelling of spectroscopic batch process data using grey models to incorporate external information

Stephen P. Gurden, Johan A. Westerhuis, Sabina Bijlsma and Age K. Smilde*

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

SUMMARY

In both analytical and process chemistry, one common aim is to build models describing measured data. In cases where additional information about the chemical system is available, this can be incorporated into the model with the aim of improving model fit and interpretability. A model which consists of a 'hard' or 'white' part describing known sources of variation and a 'soft' or' black' part describing unknown sources of variation is called a 'grey' model. In this paper the use of a grey model is demonstrated using data from a first-order chemical batch reaction monitored by UV-vis spectroscopy. The resultant three-way data matrix is modelled using a Tucker3 structure, and external information about the spectroscopically active compounds is incorporated in the form of constraints on the model parameters with additional restrictions on the Tucker3 core matrix. The grey model is then used to analyse new batches. Different approaches to building grey models are described and some of their properties discussed. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: batch process monitoring; UV-vis spectroscopy; data modelling; multiway analysis; Tucker3

1. INTRODUCTION

In the fields of both analytical and process chemistry, one common aim is to build models describing measured data. This may be for different reasons: in data exploration the main aim is to understand more clearly the important factors underlying the process [1,2]; in process optimization the aim is to identify variables which most influence process efficiency [3,4]; in quality control one aim is to build predictive models between current process conditions and end-product quality; in multivariate statistical process control (MSPC) the aim is to build multivariate control charts for fault detection and diagnosis of new batches [5,6].

Whilst the methodology described in this paper is relevant to a wide variety of data-modelling applications, emphasis is placed on one particular field of interest: modelling of data from a chemical batch process. Batch reactions form a major part of the chemical process industry, and methods which allow increased understanding and control of these processes are therefore of importance. The usual

* Correspondence to: Age 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
Contract/grant sponsor: Netherlands Technology Foundation (STW).

Received 5 October 1999
Accepted 16 February 2000

procedure for a batch process is that a reactor is charged with raw materials corresponding to a specific recipe, and then conditions inside the reactor are controlled according to preset trajectories, with the aim of giving a consistent end-product. If J different on-line measurements are recorded at K time points for I separate batch runs, then a three-way array ($I \times J \times K$) of recorded process data is available.

Reasons for building a model of batch process data are twofold. Firstly, general understanding of the process as described by the measured data is sought. Investigation of the model parameters and residuals can confirm or increase knowledge of the process, depending on whether or not they can be interpreted to correspond to pre-existing physicochemical knowledge. Secondly, a model which describes the differences between batch runs is often required. In many process control applications a set of historical data corresponding to normal operating conditions (NOC) is available. Although nominally these data represent a series of replicate runs, in practice there are systematic differences between the batches owing to variation in process conditions beyond the control of the process operator. The aim of the batch process model is to describe this systematic experimental error (or 'common-cause' variation) with a view to statistical testing of future batch runs to ensure process control.

It should be noted that whilst general process investigation can be carried out on data from single batches using two-way techniques (such as curve resolution), the modelling of common-cause variation must be carried out on the batch process data as a whole.

1.1. External information

The common approach to data modelling of chemical data is purely data-driven, often based on empirical latent variable projection methods such as principal component analysis (PCA) [7–11] partial least squares (PLS) [12–14] or, more recently, three-way methods such as PARAFAC [15–21]. These models aim to explain significant variation in the data in terms of a much reduced number of latent factors (scores and loadings) which describe combinations of the process variables. Examination of these factors can yield information on relationships and key variables within the process, but this is not always the case owing to the difficulty sometimes found in interpreting 'soft' models.

In many cases, external information about the chemical system is available. This information may be available at different levels, e.g. the reaction kinetics or mass and energy balances may be partly or fully known; the set-points for controlled process variables are known; the spectra of some or all of the chemical species may be known. This 'hard' or 'white' knowledge constitutes information about the ideal behaviour of the process and, if incorporated into the data model, can thereby give model parameters which are directly related to known causes of process variation. A model that combines both hard (white) external information and soft (black) empirical information is called a grey model. This is equivalent to dividing the variation in the data into three parts:

- systematic variation due to known sources;
- systematic variation due to (at present) unknown sources;
- residual variation.

This categorization of sources of variation is shown in Figure 1. Some advantages of this division are as follows.

1. The interpretability of the model is improved. Some model components now directly correspond to known physicochemical phenomena, and the effect of these phenomena on the process is more easily understood. Furthermore, unknown sources of variation are modelled separately and can be examined in isolation.

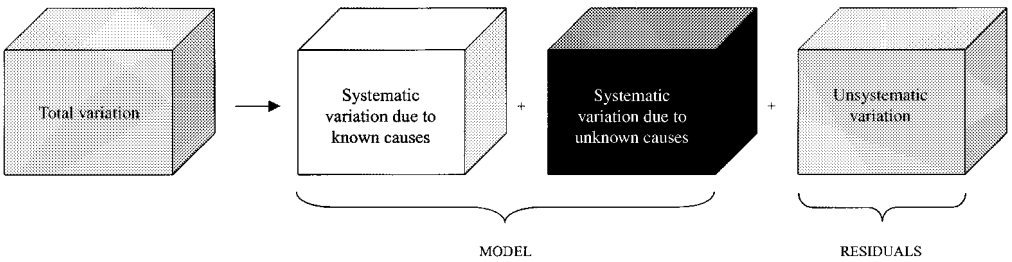

Figure 1. Schematic diagram for a grey model.

2. The relative influence of sources of variation on the process can be examined and used to determine the extent to which the process obeys ideal behaviour.
3. The numerical stability of the modelling algorithm is increased. In cases where the white part of the model accounts for a large part of the variance in the data, the remaining variation in the data is far better conditioned for empirical modelling, resulting in improved convergence speed and avoidance of 'degeneracy' problems [22,23].

1.2. Spectroscopic monitoring of a batch reaction

Spectroscopic techniques which can be implemented on-line, such as UV-vis, NIR and Raman, provide a rich source of information about conditions within a chemical system. They have the advantages of being non-destructive and able to generate data at frequent time intervals and for relatively low maintenance costs. Chemical information such as the concentration of spectroscopically-active chemical species present in the reaction, changes in solvent conditions, the presence of impurities, etc. is made available and can be used to follow whether the reaction is proceeding correctly. This information may not be manifest in the more standard process measurements of physical conditions within a chemical process, such as temperatures, pH, pressures, flow rates, etc., and is therefore of increasing interest for use in on-line process monitoring.

The use of a grey model is demonstrated in this paper using UV-vis spectroscopic data from a first-order batch reaction as an example. The subsequent use of the grey model for the analysis of new batches, some of which have known erroneous behaviour, is then described. Although the application described here is specifically concerned with spectroscopic batch data, the possibilities for other forms of process data are discussed in the conclusions.

1.3. Background

There are a number of reports in the engineering literature on the use of hybrid models, e.g. the mixing of first principles and 'black box' artificial neural network models for the modelling of non-linear processes. In general, hybrid model approaches have been shown to have significant advantages, since they allow the activation of a larger portion of the available *a priori* knowledge [24,25]. There have also been reports on approaches for incorporating external information into reduced factor models. For the case of two-way data, Takane and Shibayama [26] described a method of including exact external information about objects (rows) and/or variables (columns) in a bilinear model. Given $\mathbf{X}$ ($I \times J$) and external information on the objects, $\mathbf{M}$ ($I \times P$), and on the variables, $\mathbf{N}$ ($J \times Q$), the model is given by

$$\mathbf{X} = \mathbf{MAN}^{\mathrm{T}} + \mathbf{MB}^{\mathrm{T}} + \mathbf{CN}^{\mathrm{T}} + \mathbf{E} \quad (1)$$

where **A**, **B** and **C** are unknown and have to be estimated. One important idea of this approach is that the variation in **X** is partitioned into four parts: a part explained by **M** and **N** simultaneously, a part explained by **M** only, a part explained by **N** only and an unexplained part **E**. It is this separation of sources of variation which brings greater clarification as to the role of the external information within the measured data. This approach was extended to allow different sets of constraints on different bilinear components [27]. CANDELINC (canonical decomposition with linear constraints) [28] is related in that **X** is also decomposed into a linear model with certain subspace restrictions on the solution parameters, although there is no attempt to partition white and black parts of the model.

In this paper a different approach is described using a specific example from process analytical chemistry. A first-order batch reaction is monitored using UV-vis spectroscopy. The resultant three-way data matrix is modelled using a Tucker3 structure [29,30], and external information about the spectroscopically active compounds is incorporated in the form of constraints on the model parameters with additional restrictions on the Tucker3 core matrix [31,32]. The Tucker3 model describes three-way data in terms of the variation present with respect to each of the three dimensions, decomposing the data into three sets of latent components. Because of the ease with which external information about each dimension can be incorporated, in the form of restrictions upon these latent components, grey models can be constructed using this methodology.

2. EXPERIMENTAL

The reaction used in this paper has been described previously in the literature [33]. A two-step consecutive reaction of 3-chlorophenylhydrazonopropane dinitrile (A) with 2-mercaptoethanol (B) forms an intermediate adduct (C) which is hydrolysed to give 3-chlorophenylhydrazonocyanoacetamide (D) and ethylene sulphide (E):

$$\mathrm{A} + \mathrm{B} \xrightarrow{k_1} \mathrm{C} \xrightarrow{k_2} \mathrm{D} + \mathrm{E} \quad (2)$$

In this case, B is present in large excess (276:1) with respect to A and therefore pseudo-first-order kinetics can be assumed. Only A, C and D are spectroscopically active in the UV-vis window considered and the reaction is pH-dependent.

The reaction took place in a quartz cuvette using a reactant volume of 2·5 ml with an initial concentration of A of 54×10^{-6} mol l^{-1}. A water bath and thermocouple were used to maintain a temperature of 25 °C. A Hewlett Packard 8453 UV-vis spectrophotometer with diode array detection was used to record spectra with a path length of 1·00 cm. Spectra within the range 200–600 nm at a resolution of 1 nm were recorded every 10 s in a total run time of 45 min. Only the wavelength range 300–500 nm is analysed here. Spectra recorded from the start of the reaction up to time 2 min were discarded as they were not found to be reproducible. Finally, in order to reduce the size of the data set, every second spectrum was discarded, leaving a total of 130 spectra per experiment.

Thirty experiments were carried out under normal operating conditions (NOC). Three of these batches were retained for model validation, leaving a $27 \times 201 \times 130$ (batches × wavelengths × time points) three-way array of UV-vis absorbances. A typical set of spectra measured for one batch is shown in Figure 2.

In addition to the NOC data, nine extra experiments were run for which pH disturbances were introduced: after 20 min, 10 µl of NaOH with varying concentrations between 0·0251 and 0·4000 mol l^{-1} was added to the reaction. These nine disturbed batches give a $9 \times 201 \times 130$ three-way array.

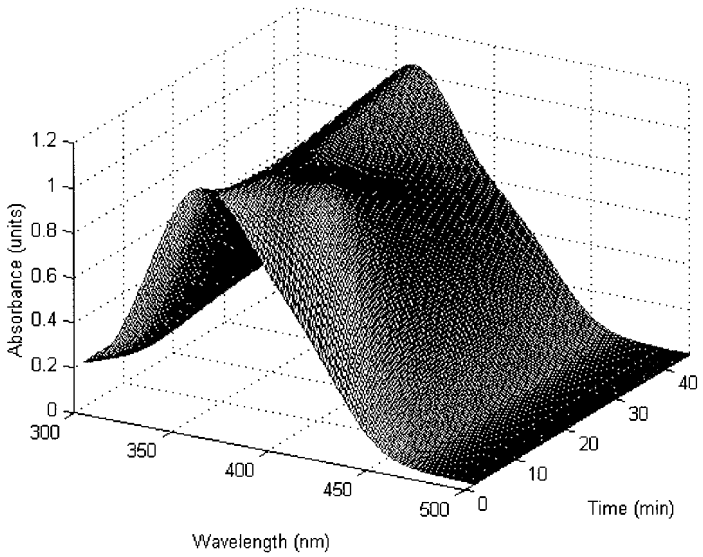

Figure 2. Typical set of spectra measured for one batch.

3. STRUCTURE OF THE DATA

For each batch experiment the set of recorded spectra can be expressed as a linear addition of contributions from the spectroscopically active compounds present in the reaction system, in accordance with the Beer–Lambert law [34]. For batch i this is given by

$$\mathbf{X}_i = \lambda_{i1}\mathbf{w}_1\mathbf{s}_1^{\mathrm{T}} + \lambda_{i2}\mathbf{w}_2\mathbf{s}_2^{\mathrm{T}} + \lambda_{i3}\mathbf{w}_3\mathbf{s}_3^{\mathrm{T}} + \mathbf{E}_i = \sum_{r=1}^{3} \lambda_{ir}\mathbf{w}_r\mathbf{s}_r^{\mathrm{T}} + \mathbf{E}_i = \mathbf{W}\mathbf{\Lambda}_i\mathbf{S}^{\mathrm{T}} + \mathbf{E}_i \tag{3}$$

where $\mathbf{X}_i$ is a K time points $\times$ J wavelengths matrix and $\mathbf{W}$ ($K \times 3$) and $\mathbf{S}$ ($J \times 3$) contain the normalized concentration and spectral profiles respectively of the spectroscopically active compounds. $\mathbf{\Lambda}_i$ (3×3) is a diagonal matrix giving concentration information relating to the overall presence of compound r (= 1,2,3) in batch i, and $\mathbf{E}_i$ is a residuals matrix.

The total set of I batch experiments form an $I \times J \times K$ array $\underline{\mathbf{X}}$. This array can be unfolded or 'matricized' [35] to give the $I \times JK$ matrix $\mathbf{X}$, allowing the convenient expression of the total batch data as

$$\mathbf{X} = \mathbf{\lambda}_1(\mathbf{w}_1 \otimes \mathbf{s}_1)^{\mathrm{T}} + \mathbf{\lambda}_2(\mathbf{w}_2 \otimes \mathbf{s}_2)^{\mathrm{T}} + \mathbf{\lambda}_3(\mathbf{w}_3 \otimes \mathbf{s}_3)^{\mathrm{T}} + \mathbf{E} = \sum_{r=1}^{3} \mathbf{\lambda}_r(\mathbf{w}_r \otimes \mathbf{s}_r)^{\mathrm{T}} + \mathbf{E} \tag{4}$$

where $\otimes$ is the Kronecker product [36,37]. This form expresses the data in terms of three independent, trilinear components, or 'triads', which each contain a vector describing the batch ($\mathbf{\lambda}_r$), spectral ($\mathbf{s}_r$) and time ($\mathbf{w}_r$) directions. For example, $\mathbf{\lambda}_1$ ($I \times 1$) gives the overall concentration of compound 1 in each batch and $\mathbf{s}_1$ ($J \times 1$) and $\mathbf{w}_1$ ($K \times 1$) give the spectral and concentration profiles of compound 1 respectively.

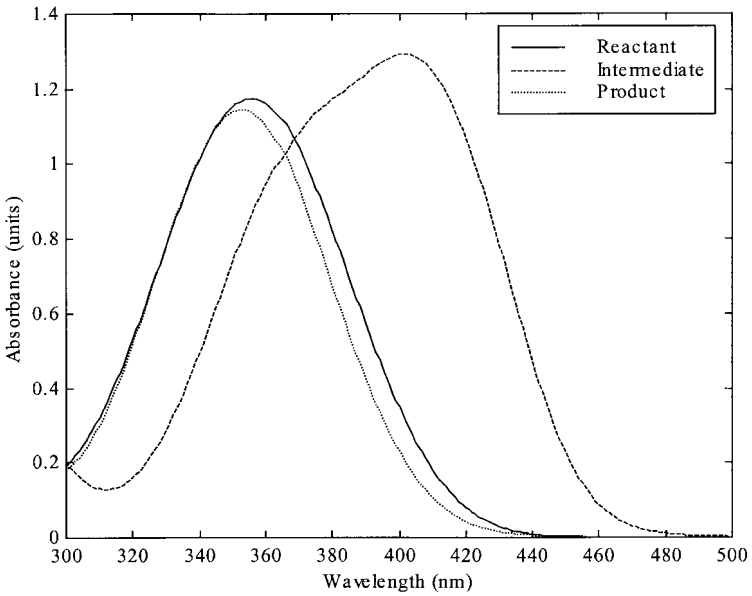

Figure 3. Estimated component spectral profiles.

3.1. Available external information

It has been shown previously [33] that applying a first-order kinetic reaction model to the measured spectra yields estimates of the rate constants (k_1 and k_2 in Equation (2)) for each batch. From these it is then possible to estimate both the compound spectra and the concentration time profiles for the reactant, intermediate and product. The averages of these estimated profiles taken over all batches are given in Figures 3 and 4. These profiles represent information about the ideal behaviour of the chemical process and should be incorporated into the white part of the model.

3.2. Sources of experimental variation

The external information used to define the white model describes process variation due to known causes for the average batch run. Deviation from this ideal behaviour will be present from different sources and at different levels. Some sources of variation will be specific to single batches (e.g. irreproducible spectral noise)—these will be termed 'batch-specific' effects. Other sources of variation will be common to all batches (e.g. reproducible temperature effects)—these will be termed 'batch-communal' effects. Note that in the latter case, although the size of the effect may differ from batch to batch, the effect itself is the same in all batches and thus can be described by a model of the batch data.

The assumption of linear additivity made by the Beer–Lambert law is held to be true at the relatively low concentrations (less than 55×10^{-6} mol l^{-1}) present in the reaction system. However, some other reasons for deviation from ideal behaviour and/or the structure given in Equation (4) are given below, along with an indication of whether they are expected to be batch-specific or batch-communal:

- pH and temperature changes which affect the rate constants and consequently the concentration profiles (batch-specific and/or batch-communal);

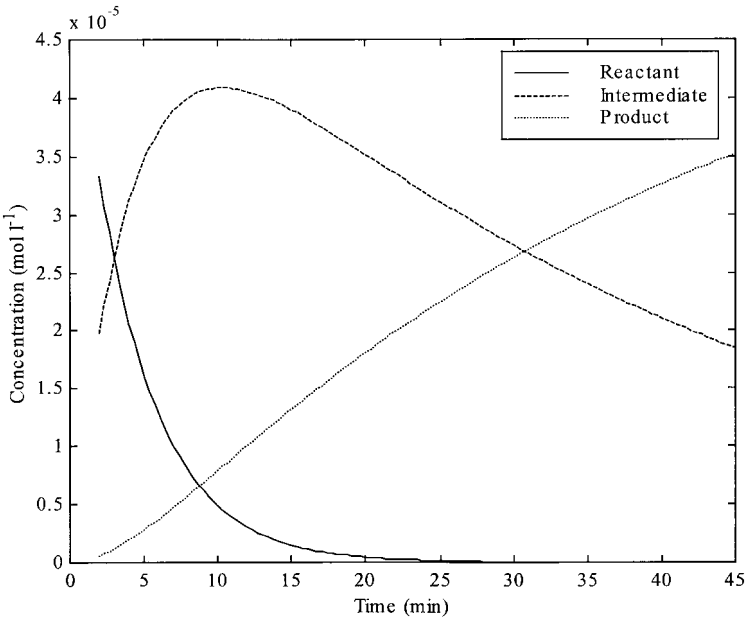

Figure 4. Estimated component kinetic time profiles.

- small differences in the initial concentration of the reactant (batch-communal);
- noise, 'spikes' and baseline drift in the UV-vis instrument (batch-specific and/or batch-communal);
- misspecification of the true compound spectra and reaction kinetics in the white model (batch-communal).

At present, no external information is available about the size and effect of these sources of common-cause variation and so they cannot be included in the white part of the model. If, however, there is a reproducible behaviour throughout the batches, then this may be captured by the black part of the model.

4. DATA MODELLING

In this section the application of different three-way models to the measured process data is described, beginning with a simple, empirical PARAFAC model and culminating in a grey model which is shown to provide an excellent fit to the data whilst retaining interpretable parameters. All models were fitted using ALS (alternating least squares) procedures [37–39] with multiple initiations and a percentage fit convergence criteria of 1×10^{-6}. The unfolded data matrix $\mathbf{X}$ ($I \times JK$) was column mean-centred prior to analysis.

4.1. The PARAFAC model

PARAFAC (parallel factor analysis) [15–18] can be considered a generalization of the bilinear PCA (principal component analysis) [7,9] model to data of higher orders. The model is given by

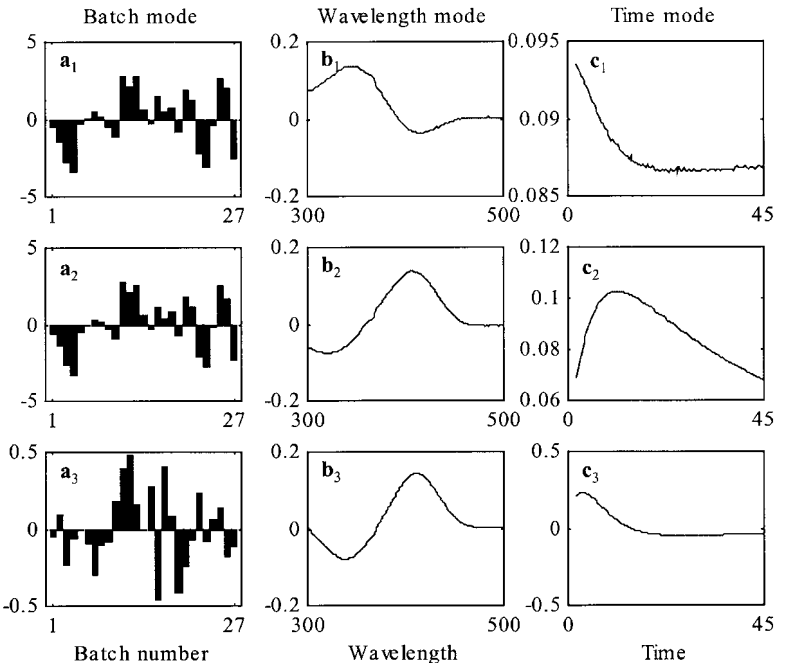

Figure 5. Loadings for the three-component PARAFAC model.

$$\mathbf{X} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathrm{T}} + \mathbf{E} = \sum_{r=1}^{R} \mathbf{a}_r(\mathbf{c}_r \otimes \mathbf{b}_r)^{\mathrm{T}} + \mathbf{E} \tag{5}$$

where $\odot$ is the Khatri-Rao product [37,40,41], $\mathbf{A}$ ($I \times R$), $\mathbf{B}$ ($J \times R$) and $\mathbf{C}$ ($K \times R$) are loadings (latent component) matrices describing variation in the I, J and K modes respectively and R is the number of components (triads) used for the model. Note that this model corresponds exactly to the postulated structure of the spectral batch data given in Equation (4).

A three-component PARAFAC model was fitted to the data, explaining 99·41% of the variation in the data. The loadings are plotted in Figure 5. Two observations are immediately apparent from these loadings. Firstly, the batch mode loadings for components 1 and 2 are seen to be highly correlated ($r^2 = 0{\cdot}9965$). This correlation is typical of degeneracy in the data [22] and implies that convergence of the ALS algorithm to stable parameters is not guaranteed. This is obviously undesirable if the aim is to construct robust chemical process models, and suggests that the PARAFAC model is not optimal for this data.

An explanation of the high correlation in the batch mode can be given by considering the structure of the data as described by Equation (4). For data which corresponded exactly to this postulated structure, the batch mode loadings would represent only the overall concentrations of the three compounds in the reactions. For the first-order model a change in the initial concentration of the reactant would produce a linear change in the concentration profiles for the intermediate and product and thus the matrix $[\boldsymbol{\lambda}_1\boldsymbol{\lambda}_2\boldsymbol{\lambda}_3]$ would be expected to have rank one. In fact, the batch mode loadings $\mathbf{A}$ are seen to have a rank of higher than one, indicating the presence of additional sources of variation in the real data analysed here.

A second observation is that the wavelength and time mode loadings are correlated and do not correspond to those estimated by the ideal kinetic model given in Figures 3 and 4. One reason for this

is the degeneracy in the model, but another is that the spectra for the reactant and product are very similar and so are very difficult to model separately using an unrestricted model. This reduces the interpretability of the model, as it is difficult to distinguish between the influence of the individual compounds on the model and the influence of deviations from predicted ideal behaviour.

4.2. Restricted Tucker3 model

The Tucker3 [29] model is related to the PARAFAC model and is given by

$$\mathbf{X} = \mathbf{AG}(\mathbf{C} \otimes \mathbf{B})^{\mathrm{T}} + \mathbf{E} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} \mathbf{a}_p (\mathbf{c}_q \otimes \mathbf{b}_r)^{\mathrm{T}} + \mathbf{E} \tag{6}$$

where $\mathbf{G}$ ($P \times QR$) is a matricized core array and P, Q and R are the numbers of latent components used to describe the I, J and K modes respectively (i.e. the number of columns in $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$). Thus a (2,4,3) Tucker3 model means that $P = 2$, $Q = 4$ and $R = 3$. From Equation (6) it can be seen that if a core element is zero ($g_{pqr} = 0$), this implies that the triad $\mathbf{a}_p(\mathbf{c}_q \otimes \mathbf{b}_r)^{\mathrm{T}}$ does not contribute to the model.

A Tucker3 structure was applied in order to resolve the problems of degeneracy and interpretability encountered with the PARAFAC model in the following way.

1. External information about the chemical process was introduced into the model. The wavelength loadings $\mathbf{B}$ were restricted to equal the spectra for the reactant, intermediate and product (see Figure 3) as estimated in advance from knowledge of the reaction kinetics. Similarly, the loadings $\mathbf{C}$ which represent variation in the time mode were constrained to equal the compound concentration profiles (see Figure 4).
2. A (1,3,3) structure was used, as there was only one source of variation in the batch mode known to correspond to the first-order model, i.e. variation due to changes in the initial concentration of the reactant. As in the PARAFAC model, no interactions between the three compounds were considered (in accordance with the assumption of linear additivity within the system). This was done by using a core matrix with the structure

$$\mathbf{G} = [\, g_{111} \quad 0 \quad 0 \mid 0 \quad g_{122} \quad 0 \mid 0 \quad 0 \quad g_{133} \,] \tag{7}$$

where only latent component pairs $(\mathbf{c}_1 \otimes \mathbf{b}_1)^{\mathrm{T}}$, $(\mathbf{c}_2 \otimes \mathbf{b}_2)^{\mathrm{T}}$ and $(\mathbf{c}_3 \otimes \mathbf{b}_3)^{\mathrm{T}}$ are allowed to contribute to the model, and all other component interactions are fixed to zero.[31,32].

The restricted (1,3,3) Tucker3 model explained 97·49% of the variation in the data. The loadings are given in Figure 6 and the model residuals are plotted with respect to each mode in Figure 7. In this case, as the batch mode loadings and the core elements are the only free parameters to be estimated, a stable estimate is obtained. The correlation between $\mathbf{b}_1$ and $\mathbf{b}_3$ is unproblematic, as these parameters are fixed, not estimated, and represent real phenomean.

The reduction in model fit compared to the PARAFAC model is significant (i.e. compared to the level of spectral noise). Although the amount of variation explained indicates that the data do follow ideal behaviour to a large extent, it is apparent from the model residuals that there is some systematic variation unaccounted for by the model, from the time of the first measurement up to around time 15 min. The reason for this is thought to be differences in the value of the rate constant k_1 which defines the concentration profile of the reactant (and consequently the intermediate and product) at the beginning of the reaction. The rate constant k_1 is approximately 10 times larger than k_2 and so temperature fluctuations associated with the start-up of the reaction will cause some deviation from ideal behaviour.

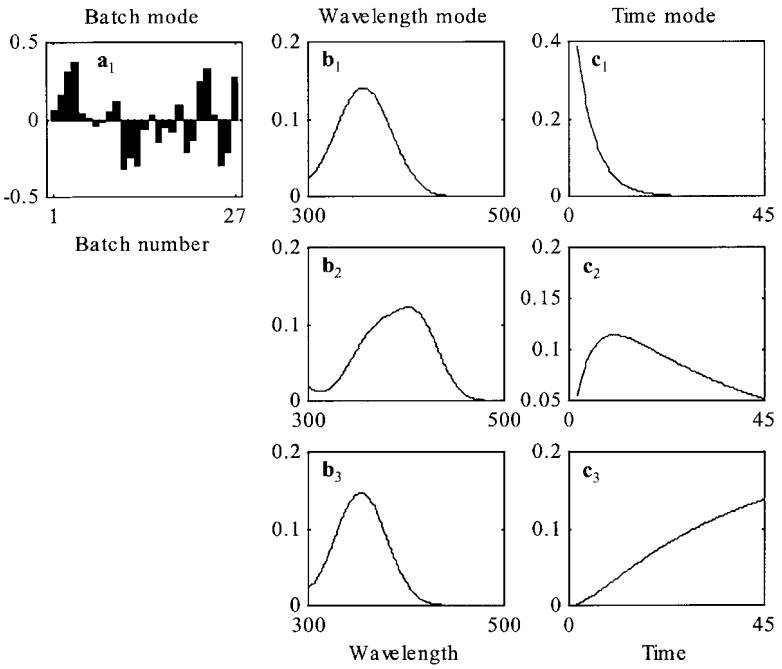

Figure 6. Loadings for the restricted (2,3,3) Tucker3 model.

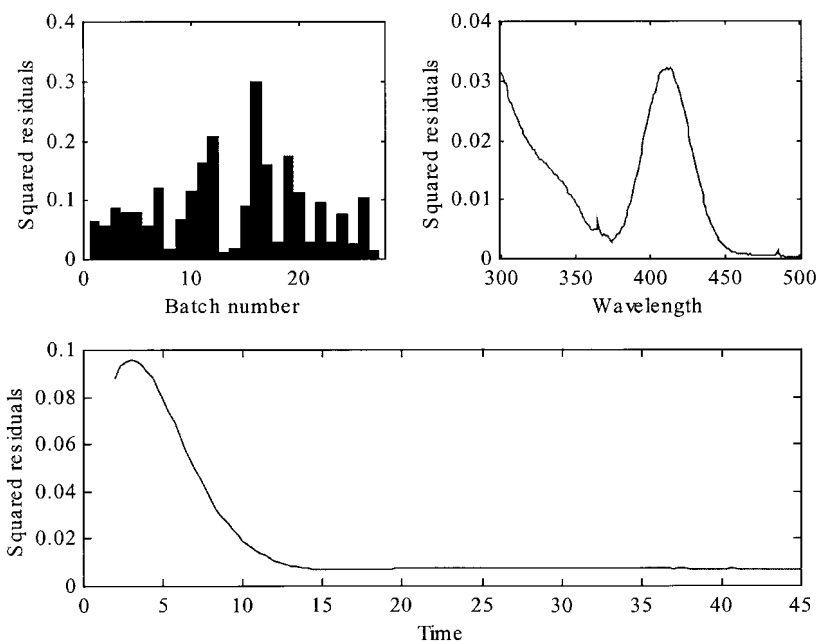

Figure 7. Residuals for the restricted (2,3,3) Tucker3 model.

4.3. The grey model

The restricted Tucker3 model described in the previous subsection represents the application of a white model to the data. The model parameters are restricted in such a way that the data are forced to describe variation corresponding only to pre-existing hard knowledge of the chemical system. However, the residuals from this model demonstrate that there is some structure in the data which is not fully understood, and therefore a soft, black model should be applied to these residuals in order to make the overall model as complete as possible.

One approach is to construct a soft model on the residuals from the restricted Tucker3 model. This new model is left free to describe any additional variation not accounted for by the white model. One advantage of this approach, where the white and black models are calculated sequentially, is that it is possible to use different types of models for the white and black parts. For example, it may be that the white model describes all the trilinear structure present in the data, in which case **E** may be best modelled using a two-way technique such as PCA[42], or it may be necessary to use a combination of linear and non-linear models. This sequential approach is discussed further in Section 6.

A second approach to constructing a grey model, which can be applied in this case, is to expand the Tucker3 core matrix by the addition of extra trilinear components. These additional components are not restricted and so are free to describe the additional variation in the data not corresponding to the hard model. In this way the white and black model components are calculated simultaneously and are found to give a better overall fit to the data than in the sequential approach.

For the data described here, models using from one to three additional components were calculated. It was found that two were optimal in modelling the remaining systematic variation in the data, the third component describing mostly spectral noise. The new core matrix is given as

$$\mathbf{G}^* = \left[\begin{array}{ccccc|ccccc|ccccc|ccccc|ccccc} g^*_{111} & 0 & 0 & 0 & 0 & 0 & g^*_{122} & 0 & 0 & 0 & 0 & 0 & g^*_{133} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g^*_{244} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g^*_{355} \end{array}\right] \tag{8}$$

The triads corresponding to g^*_{111}, g^*_{122} and g^*_{133} constitute the white part of the model (for which loadings **B** and **C** are fixed), and those corresponding to g^*_{244} and g^*_{355} constitute the black part.

The grey model explained 99·77% of the variance in the data. UV-vis data are relatively noise-free and so it is possible to explain such a high percentage of variance. The standard deviation of the model residuals was $5{\cdot}5 \times 10^{-4}$, which corresponds well to an independent estimation of the heteroscedastic spectral noise which was between 5×10^{-5} and 5×10^{-4} absorbance units. The grey model loadings are given in Figure 8 and the model residuals in Figure 9. Of the two additional components, the first (i.e. triad ($\mathbf{a}_2$,$\mathbf{b}_4$,$\mathbf{c}_4$)) is thought to describe a slight drift in the spectral baseline as the reaction progresses, caused by changes in temperature. Although relatively noisy, the time mode loadings may reflect the action of the water bath controller. Interestingly, the batch loadings ($\mathbf{a}_2$) show a clear cyclic behaviour (the batches were measured sequentially over a period of 2 weeks), indicating a dynamic effect. The second (triad ($\mathbf{a}_3$,$\mathbf{b}_5$,$\mathbf{c}_5$)) describes more significant variation thought to be due to differences in the rate constant k_1 at the start of the reaction due to temperature fluctuations, causing deviation from the ideal concentration profile of the reactant given by the white model. The model residuals show a large spike at approximately 360 nm and another at 485 nm in the wavelength mode (see Figure 9). These are artefacts of the instrument and could not be modelled systematically. There is also some residual spectral signal, probably due to error in the *a priori* estimation of the spectral profile for the intermediate, the precise nature of which is not available because it could not be physically isolated.

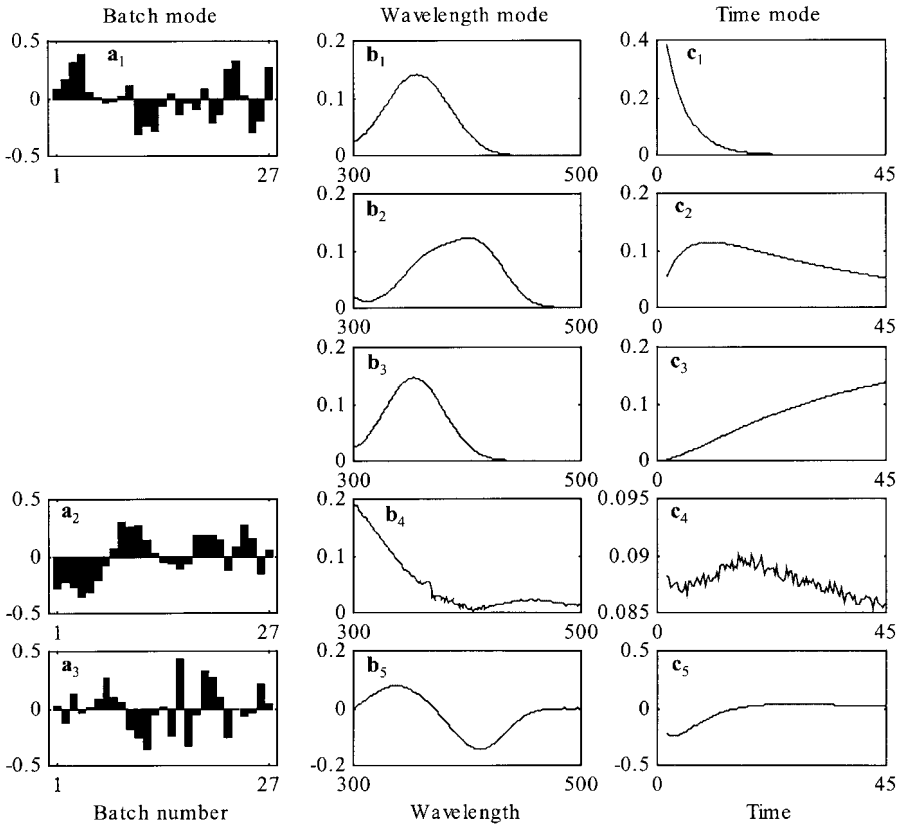

Figure 8. Loadings for the grey model.

5. APPLICATION OF THE GREY MODEL TO NEW BATCHES

One of the aims of process data modelling is to build a statistical basis for the monitoring of new batches from the same process, in order to detect whether any significant variation from normal operating conditions is present. In addition to the 27 NOC batches used to build the model, three additional NOC batches were available along with nine batches for which a known disturbance was introduced.

These new batches (numbered 28–30, NOC; 31–39, disturbed) were projected onto the grey model described in Section 4.3 as follows:

$$\mathbf{a}_{\text{new}}^{\text{T}} = \mathbf{x}_{\text{new}}^{\text{T}}\mathbf{P}(\mathbf{P}^{\text{T}}\mathbf{P})^{-\mathbf{1}} \tag{9}$$

where $\mathbf{a}_{\text{new}}$ ($P \times 1$) is the loadings for the new batch, $\mathbf{x}_{\text{new}}$ ($JK \times 1$) is the (vectorized) new batch data and $\mathbf{P}^{\text{T}} = \mathbf{G}^{*}(\mathbf{C} \otimes \mathbf{B})^{\text{T}}$ is the grey model loadings. From these new batch loadings a corresponding set of model residuals was then calculated:

$$\mathbf{e}_{\text{new}}^{\text{T}} = \mathbf{x}_{\text{new}}^{\text{T}} - \mathbf{a}_{\text{new}}^{\text{T}}\mathbf{P}^{\text{T}} \tag{10}$$

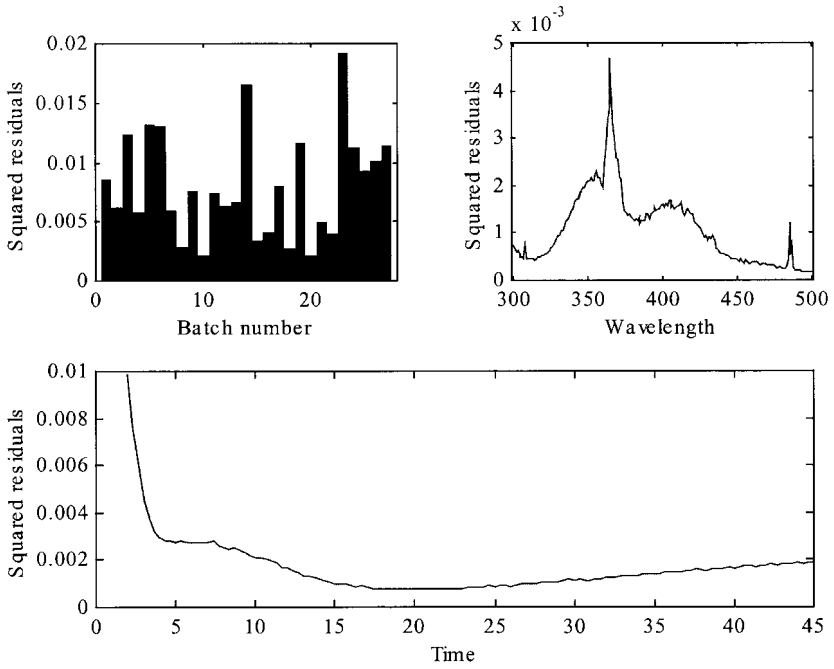

Figure 9. Residuals for the grey model.

and the sum of squares of these residuals was calculated and used to determine how well the new batches fitted the process model.

Figure 10 shows the sum of squared residuals for the 27 NOC batches used to build the model, the three additional NOC batches and the nine disturbed batches. 95% and 99% confidence limits are given in accordance with multivariate statistical process control methodology [6,43,44]. All the NOC data fall within the 95% control limits. The new NOC batches (30–32) are also within the confidence limits, indicating that despite the high percentage of variation described by the grey model, there is probably no significant degree of model overfit. Of the nine batches known to include process disturbances (33–41), all break the 95% confidence limit and only numbers 31 and 34 do not break the 99% confidence limit. The concentrations of NaOH used to disturb these batches were the lowest used (0·0251 and 0·1020 mol l^{-1} respectively), and so the effect of the process disturbance is not as large as for e.g. batch 37 for which the concentration of NaOH was higher (0·4000 mol l^{-1}).

This analysis of the batch residuals demonstrates only one use of the grey model, i.e. for off-line process monitoring. By way of comparison, using the PARAFAC model described in Section 4.1, none of the disturbed batches breaks the 99% limit and four do not break the 95% limit. It should be noted that, although it is not within the scope of this paper, it is possible to apply further detection and diagnostic tools [45,46] to investigate the nature of the detected process disturbances. This methodology can also be applied on-line for rapid detection of deviant process behaviour, as will be reported in a follow-up paper.

6. FURTHER DISCUSSION OF GREY MODELS

In Section 4.3, two approaches were discussed for the construction of grey models: a sequential and a simultaneous approach. These approaches result in grey models with different parameters and

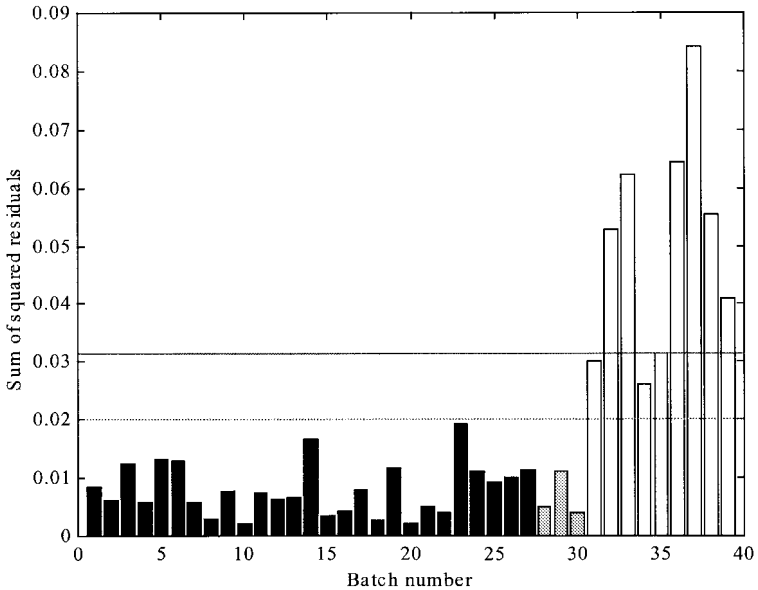

Figure 10. Sum of squared residuals for NOC batches (black), new NOC batches (grey) and new batches with known process disturbances (white). 99% (full line) and 95% (dotted line) confidence limits are shown.

properties which are discussed in this section. Table I summarizes the percentage variances explained by some different types of grey models built on the previously described data and is referred to later on in the text.

Sequential grey models are calculated by applying a white model to the data and then building a black model on the residual data. If, for example, the white part of the model took a restricted Tucker3 structure and the black part took a PARAFAC structure, the overall model could be expressed as

$$\begin{aligned}\mathbf{X} &= \mathbf{A}_w\mathbf{G}_w(\mathbf{C}_w \otimes \mathbf{B}_w)^T + \mathbf{E}_w = \mathbf{A}_w\mathbf{G}_w(\mathbf{C}_w \otimes \mathbf{B}_w)^T + \mathbf{A}_b(\mathbf{C}_b \odot \mathbf{B}_b)^T + \mathbf{E}_{sq} \\ &= \mathbf{A}_w\mathbf{P}_w^T + \mathbf{A}_b\mathbf{P}_b^T + \mathbf{E}_{sq} = [\mathbf{A}_w \mid \mathbf{A}_b][\mathbf{P}_w \mid \mathbf{P}_b]^T + \mathbf{E}_{sq} = \mathbf{A}_{sq}\mathbf{P}_{sq}^T + \mathbf{E}_{sq}\end{aligned} \tag{11}$$

where the 'w' and 'b' subscripts refer to the white and black parts of the model, **P** represents a generic loadings matrix whose structure depends upon the particular model applied, $\mathbf{E}_w$ is the residuals matrix for the white model only and $\mathbf{E}_{sq}$ is the residuals matrix for the sequential grey model.

Table I. Percentage variances explained by different types of grey models

	Type of grey model		
	Simultaneous	Sequential	Simultaneous with $\mathbf{A}_w^T\mathbf{A}_b = \mathbf{0}$
$\%VE_{total}$	99·77	99·74	99·71
$\%VE_{white}$	—[a]	97·49	95·82
$\%VE_{black}$	—[a]	2·25	3·89

[a] Partitioning of variance not possible for this model.

The simultaneous grey model can also be expressed generically as

$$\begin{aligned}\mathbf{X} &= \mathbf{A}_{sm}\mathbf{G}_{sm}(\mathbf{C}_{sm} \otimes \mathbf{B}_{sm})^{T} + \mathbf{E}_{sm} = \mathbf{A}_{w}\mathbf{G}_{w}(\mathbf{C}_{w} \otimes \mathbf{B}_{w})^{T} + \mathbf{A}_{b}\mathbf{G}_{b}(\mathbf{C}_{b} \otimes \mathbf{B}_{b})^{T} + \mathbf{E}_{sm} \\ &= \mathbf{A}_{w}\mathbf{P}_{w}^{T} + \mathbf{A}_{b}\mathbf{P}_{b}^{T} + \mathbf{E}_{sm} = [\mathbf{A}_{w} \mid \mathbf{A}_{b}][\mathbf{P}_{w} \mid \mathbf{P}_{b}]^{T} + \mathbf{E}_{sm} = \mathbf{A}_{sm}\mathbf{P}_{sm}^{T} + \mathbf{E}_{sm}\end{aligned} \tag{12}$$

where $\mathbf{E}_{sm}$ is the residuals matrix for the simultaneous grey model.

6.1. *Difference in model fit*

The total amount of variance explained by a model is defined by

$$\%\mathrm{VE}_{total} = \left(1 - \frac{\|\mathbf{E}\|^{2}}{\|\mathbf{X}\|^{2}}\right) \times 100\% \tag{13}$$

where $\mathbf{E}$ is the overall residuals matrix from the model. It is found that the simultaneous approach describes a higher (or equal) percentage of variance in $\mathbf{X}$ than the sequential approach, i.e. $\|\mathbf{E}_{sm}\|^{2} \leq \|\mathbf{E}_{sq}\|^{2}$. This can be explained by the fact that in the sequential approach, additional parameter constraints are indirectly imposed.

For the data discussed in this paper, the sequential model constructed using a restricted Tucker3 model for the white part and a two-component PARAFAC model for the black part described 99·74% and the simultaneous method with a comparable structure described 99·77% of the variation in the data (see Table I). Considered relative to the spectral noise present in the data, this difference is significant, although the sequential model was still found to have reasonable fault detection ability for the disturbed batches.

6.2. *Difference in estimated parameters*

Although visually similar, the batch loadings $\mathbf{a}_1$ for the white model given in Figure 6 are different from those for the simultaneous grey model given in Figure 8. This is despite the fact that the wavelength and time mode loadings are both fixed to equal the same ideal profiles. To understand this, it must be recognized that for both models the objective is to maximize the total amount of variation explained. For the white model, where $\mathbf{B}$ and $\mathbf{C}$ are completely fixed, variation in the data due to unknown sources can only manifest itself in $\mathbf{a}_1$, in the model residuals or, most likely, partially in both. Thus even variation which is not considered by the white model may be apparent in some of the unrestricted parameters.

When building a grey model, black model parameters are included which can describe more fully the variation due to unknown sources. Furthermore, when the white and black model parameters are calculated simultaneously, the model parameters are adjusted so as to maximize the amount of variation explained in the data, i.e. a loss of white model fit may be compensated for by a greater increase in fit for the black model. Thus the use of simultaneous grey models improves interpretability, as the black model is able to isolate those unknown sources of variation which may otherwise be partially included in the white model.

6.3. *Partitioning of variance*

A desirable property of a grey model is the ability to partition the amount of total variation explained into variation explained by the white and black parts separately. This gives a measure of the extent to which the process being modelled corresponds to the ideal behaviour postulated by the white model.

Both the sequential and simultaneous grey models can be written in the general form

$$\mathbf{X} = \mathbf{A}\mathbf{P}^{\mathrm{T}} + \mathbf{E} = \mathbf{A}_{\mathrm{w}}\mathbf{P}_{\mathrm{w}}^{\mathrm{T}} + \mathbf{A}_{\mathrm{b}}\mathbf{P}_{\mathrm{b}}^{\mathrm{T}} + \mathbf{E} = \hat{\mathbf{X}}_{\mathrm{w}} + \hat{\mathbf{X}}_{\mathrm{w}} + \mathbf{E} \tag{14}$$

Given that $\mathbf{A}_{\mathrm{w}}$ and $\mathbf{A}_{\mathrm{b}}$ are least squares estimates, it can be shown for the sequential grey model (see Appendices I and II) that the sums of squares can be partitioned as

$$\|\mathbf{X}\|^2 = \|\hat{\mathbf{X}}_{\mathrm{w}}\|^2 + \|\hat{\mathbf{X}}_{\mathrm{b}}\|^2 + \|\mathbf{E}_{\mathrm{sq}}\|^2 \tag{15}$$

and therefore the percentage variance explained by the white part of the model is given by

$$\%\mathrm{VE}_{\mathrm{white}} = \left(\frac{\|\hat{\mathbf{X}}_{\mathrm{w}}\|^2}{\|\mathbf{X}\|^2}\right) \times 100\% \tag{16}$$

and the black part by

$$\%\mathrm{VE}_{\mathrm{black}} = \left(\frac{\|\hat{\mathbf{X}}_{\mathrm{b}}\|^2}{\|\mathbf{X}\|^2}\right) \times 100\% \tag{17}$$

Substituting $\|\mathbf{E}_{\mathrm{sq}}\|^2$ from Equation (15) into Equation (13) will show that

$$\%\mathrm{VE}_{\mathrm{total}} = \%\mathrm{VE}_{\mathrm{black}} + \%\mathrm{VE}_{\mathrm{white}} \tag{18}$$

For the data described here, the sequential model explained 99·74% of the data. This can be partitioned into 97·49% for the white model and 2·25% for the black model (see Table I). Note that the amount of variance explained by the white model is exactly the same as that explained by the restricted Tucker3 model given in Section 4.2, because the same model is used.

A similar partitioning of variation is not possible for the simultaneous model as described. However, it can be shown (see Appendix III) that by introducing orthogonality between the white and black model parts, $\mathbf{A}_{\mathrm{w}}^{\mathrm{T}}\,\mathbf{A}_{\mathrm{b}} = \mathbf{0}$, the conditions for partitioning of variance are fulfilled. This can be done using a constrained component algorithm [47] and has the effect of reducing the model fit to 99·71%, with partitioning of 95·82% for the white model and 3·89% for the black model (see Table I). Thus, for the simultaneous model, partitioning of variance is possible but incurs a loss of model fit. It can be noted that the amount of variation described by the black model here (3·89%) is greater than that for the sequential model (2.25%) for reasons given in Section 6.2.

6.4. Soft constraints

Although not implemented here, it is also possible to take into account the fact that external information may be available only within a limited degree of certainty. If non-exact external information is available, it may still be possible to incorporate this by implementing soft constraints [48] on the model parameters, such as non-negativity. For example, if pure compound spectra are not known, the knowledge that they are unimodal and non-negative can still be used to restrict the solution parameter estimates. Another possibility is to specify 'target vectors' for the latent components, e.g. ideal time trajectories for controlled process variables, and to specify how closely the model parameters should correspond to these targets. This constraint can be applied by adding the penalty term $\|\mathbf{g} - \mathbf{a}\|^2$ to the general least squares regression step used to calculate one component

for one dimension:

$$\min_{\mathbf{a}} \left[(1-\alpha)\|\mathbf{X} - \mathbf{Z}\mathbf{a}^{\mathrm{T}}\|^2 + \alpha\|\mathbf{g} - \mathbf{a}\|^2 \right] \tag{19}$$

where **a** is the unconstrained least squares solution, **g** is the target vector, **Z** is a regressor matrix and α is the penalty coefficient, $0 \leq \alpha \leq 1$, which determines how tightly the constraint is applied.

6.5. Summary

The choice of which grey model to use depends mostly upon the nature of the data. If very different models are required for the white and black parts, then the sequential approach is appropriate and has the advantage of the partitioning of variance property. However, for spectroscopic batch data of the type described in this paper, where a three-way structure is appropriate for both white and black models, the simultaneous approach describes the maximum amount of variation using relatively few parameters and should be preferred. If partitioning between the white and black model parts is required, this can be imposed by using the appropriate orthogonality constraint $\left(\mathbf{A}_{\mathrm{w}}^{\mathrm{T}}\,\mathbf{A}_{\mathrm{b}} = 0\right)$, but with a slight loss of model fit. Another advantage of orthogonality would be in the area of MSPC, where control charts could be applied to the white and black models parts separately, immediately providing an extra diagnostic as to the reason for an alarm. Finally, if external information is to be included in the form of model restrictions, there are a variety of hard and soft ways to do this.

7. CONCLUSIONS

The incorporation of external information into a model of a spectroscopically monitored batch reaction has been found to improve both the fit of the three-way model, by removing degeneracy problems, and the interpretation of the model parameters. Using the methodology of a restricted Tucker3 model, the incorporation of exact external information into the model is made relatively easy, because, for three-way models, variation in each of the three modes is modelled explicitly. Careful consideration must be placed on the size and appropriate structure of the core matrix, as this should reflect known interactions within the process. It has been shown that although spectroscopic batch data may seem to have a PARAFAC structure, in practice a low level of variation in the batch mode may mean that other models must be considered. An advantage of the Tucker3 structure is that a different number of components can be used to model each mode of a three-way array.

This paper focuses on modelling of spectral process data, but the application of grey models to other type of batch process data, such as industrial measurements of temperatures, pressures and flow rates, is also of interest. Although industrial processes are often highly complex, there is a large amount of extra information, much of which is not currently being employed; for example:

- batch mode information such as feedstock quality, initial reactor conditions and product quality;
- process variable relationship information such as controller coupling, e.g. coolant flow rate depends upon measured reactor temperature;
- process variable time profile information such as set-point and controlled variable trajectories, and recipe information relating to multistage processes.

Incorporating this information into batch models could certainly help better identify the complex factors affecting industrial batch processes.

ACKNOWLEDGEMENTS

These investigations were supported by the Council for Chemical Sciences of the Netherlands

Organisation for Scientific Research (NWO-CW) with financial aid from the Netherlands Technology Foundation (STW). The authors would also like to thank Hans Boelens, also from our research group, for information about the UV-vis data.

APPENDIX I. PROOF THAT THE LEAST SQUARES SOLUTION TO A BILINEAR REGRESSION PROBLEM PRODUCES RESIDUALS ORTHOGONAL TO THE REGRESSOR MATRIX

The least squares solution to a bilinear regression problem produces residuals orthogonal to the regressor matrix.

The least squares solution for **A** to

$$\mathbf{X} = \mathbf{A}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{20}$$

is found by

$$\mathbf{A} = \mathbf{X}\mathbf{P}\left(\mathbf{P}^{\mathrm{T}}\mathbf{P}\right)^{-1} \tag{21}$$

If the residuals **E** are orthogonal to the regressor matrix **P**, then

$$\mathbf{E}\mathbf{P} = \mathbf{0} \tag{22}$$

This can be proved as follows:

$$\begin{aligned} \mathbf{E}\mathbf{P} &= \left(\mathbf{X} - \mathbf{A}\mathbf{P}^{\mathrm{T}}\right)\mathbf{P} = \mathbf{X}\mathbf{P} - \mathbf{A}\mathbf{P}^{\mathrm{T}}\mathbf{P} \\ &= \mathbf{X}\mathbf{P} - \mathbf{X}\mathbf{P}\left(\mathbf{P}^{\mathrm{T}}\mathbf{P}\right)^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{X}\mathbf{P} - \mathbf{X}\mathbf{P} = \mathbf{0} \end{aligned} \tag{23}$$

Note that, as a consequence of this, a bilinear model $\hat{\mathbf{X}} = \mathbf{A}\mathbf{P}^{\mathrm{T}}$ for which **A** is a least squares solution will also be orthogonal to its residuals, i.e.

$$\mathbf{E}\hat{\mathbf{X}}^{\mathrm{T}} = \mathbf{E}\mathbf{P}\mathbf{A}^{\mathrm{T}} = \mathbf{0} \tag{24}$$

APPENDIX II. PROOF THAT, FOR SEQUENTIAL GREY MODELS, TOTAL VARIATION EXPLAINED CAN BE PARTITIONED INTO THE WHITE AND BLACK MODEL PARTS

For grey models where white and black parts are calculated sequentially, variation can be partitioned provided that, for each model part, **A** is a least squares solution.

Given a sequentially calculated grey model

$$\mathbf{X} = \mathbf{A}_{\mathrm{w}}\mathbf{P}_{\mathrm{w}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{w}} = \mathbf{A}_{\mathrm{w}}\mathbf{P}_{\mathrm{w}}^{\mathrm{T}} + \mathbf{A}_{\mathrm{b}}\mathbf{P}_{\mathrm{b}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{sq}} = \hat{\mathbf{X}}_{\mathrm{w}} + \hat{\mathbf{X}}_{\mathrm{b}} + \mathbf{E}_{\mathrm{sq}} \tag{25}$$

the sums of squares may be partitioned by

$$\|\mathbf{X}\|^2 = \|\hat{\mathbf{X}}_{\mathrm{w}}\|^2 + \|\hat{\mathbf{X}}_{\mathrm{b}}\|^2 + \|\mathbf{E}_{\mathrm{sq}}\|^2 \tag{26}$$

or, written in the form of matrix traces,

$$\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) = \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right) + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right) + \mathrm{tr}\left(\mathbf{E}_{\mathrm{sq}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sq}}\right) \tag{27}$$

This can be proved on condition that $\mathbf{A}_{\mathrm{w}}$ and $\mathbf{A}_{\mathrm{b}}$ are least squares solutions and therefore the model parts are orthogonal to their corresponding residuals (see Appendix I):

$$\mathbf{E}_{\mathrm{w}}\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}} = \left(\hat{\mathbf{X}}_{\mathrm{b}} + \mathbf{E}_{\mathrm{sq}}\right)\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}} = \mathbf{0} \tag{28}$$

$$\mathbf{E}_{\mathrm{sq}}\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}} = \mathbf{0} \tag{29}$$

Using the matrix trace properties of equivalence of transpose, $\mathrm{tr}(\mathbf{M}) = \mathrm{tr}(\mathbf{M}^{\mathrm{T}})$, and permutation, $\mathrm{tr}(\mathbf{MN}) = \mathrm{tr}(\mathbf{NM})$ (if both products exist), Equation (27) can be proved as follows:

$$\begin{aligned}
\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) &= \mathrm{tr}\left[\left(\hat{\mathbf{X}}_{\mathrm{w}} + \hat{\mathbf{X}}_{\mathrm{b}} + \mathbf{E}_{\mathrm{sq}}\right)^{\mathrm{T}}\left(\hat{\mathbf{X}}_{\mathrm{w}} + \hat{\mathbf{X}}_{\mathrm{b}} + \mathbf{E}_{\mathrm{sq}}\right)\right] \\
&= \mathrm{tr}\left[\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}} + \hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{sq}}^{\mathrm{T}}\right)\left(\hat{\mathbf{X}}_{\mathrm{w}} + \hat{\mathbf{X}}_{\mathrm{b}} + \mathbf{E}_{\mathrm{sq}}\right)\right] \\
&= \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right) + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right) + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sq}}\right) + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right) \\
&\quad + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right) + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sq}}\right) + \mathrm{tr}\left(\mathbf{E}_{\mathrm{sq}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right) + \mathrm{tr}\left(\mathbf{E}_{\mathrm{sq}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right) + \mathrm{tr}\left(\mathbf{E}_{\mathrm{sq}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sq}}\right) \\
&= \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right) + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right) + \mathrm{tr}\left(\mathbf{E}_{\mathrm{sq}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sq}}\right) + 2 \times \mathrm{tr}\left(\mathbf{E}_{\mathrm{sq}}\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\right) + 2 \times \mathrm{tr}\left[\left(\hat{\mathbf{X}}_{\mathrm{b}} + \mathbf{E}_{\mathrm{sq}}\right)\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\right] \\
&= \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right) + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right) + \mathrm{tr}\left(\mathbf{E}_{\mathrm{sq}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sq}}\right)
\end{aligned} \tag{30}$$

APPENDIX III. PROOF THAT, FOR SIMULTANEOUS MODELS, TOTAL VARIATION EXPLAINED CAN BE PARTITIONED INTO THE WHITE AND BLACK MODEL PARTS PROVIDED THAT THEY ARE ORTHOGONAL

For grey models where white and black parts are calculated simultaneously, variation can be partitioned provided that $\mathbf{A}_{\mathrm{sm}}$ is a least squares solution and $\mathbf{A}_{\mathrm{w}}$ is orthogonal to $\mathbf{A}_{\mathrm{b}}$.

Given a simultaneously calculated grey model

$$\mathbf{X} = \mathbf{A}_{\mathrm{sm}}\mathbf{P}_{\mathrm{sm}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{sm}} = [\mathbf{A}_{\mathrm{w}} \mid \mathbf{A}_{\mathrm{b}}][\mathbf{P}_{\mathrm{w}} \mid \mathbf{P}_{\mathrm{b}}]^{\mathrm{T}} + \mathbf{E}_{\mathrm{sm}} = \mathbf{A}_{\mathrm{w}}\mathbf{P}_{\mathrm{w}}^{\mathrm{T}} + \mathbf{A}_{\mathrm{b}}\mathbf{P}_{\mathrm{b}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{sm}} = \hat{\mathbf{X}}_{\mathrm{w}} + \hat{\mathbf{X}}_{\mathrm{b}} + \mathbf{E}_{\mathrm{sm}} \tag{31}$$

the sums of squares may be partitioned by

$$\|\mathbf{X}\|^2 = \|\hat{\mathbf{X}}_{\mathrm{w}}\|^2 + \|\hat{\mathbf{X}}_{\mathrm{b}}\|^2 + \|\mathbf{E}_{\mathrm{sm}}\|^2 \tag{32}$$

or, written in the form of matrix traces,

$$\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) = \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right) + \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right) + \mathrm{tr}\left(\mathbf{E}_{\mathrm{sm}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sm}}\right) \tag{33}$$

This can be proved on condition that $\mathbf{A}_{\mathrm{sm}}$ is a least squares solution and therefore the simultaneous model is orthogonal to its residuals (see Appendix I):

$$\mathbf{E}_{\mathrm{sm}}\left(\hat{\mathbf{X}}_{\mathrm{w}}+\hat{\mathbf{X}}_{\mathrm{b}}\right)^{\mathrm{T}}=\mathbf{0} \tag{34}$$

and that $\mathbf{A}_{\mathrm{w}}$ is orthogonal to $\mathbf{A}_{\mathrm{b}}$ and therefore the white and black model parts are orthogonal:

$$\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}=\left(\mathbf{A}_{\mathrm{w}}\mathbf{P}_{\mathrm{w}}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{A}_{\mathrm{b}}\mathbf{P}_{\mathrm{b}}^{\mathrm{T}}\right)=\mathbf{P}_{\mathrm{w}}\mathbf{A}_{\mathrm{w}}^{\mathrm{T}}\mathbf{A}_{\mathrm{b}}\mathbf{P}_{\mathrm{b}}^{\mathrm{T}}=\mathbf{0} \tag{35}$$

Equation (33) can be proved as follows:

$$\begin{aligned}
\mathrm{tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) &= \mathrm{tr}\left[\left(\hat{\mathbf{X}}_{\mathrm{w}}+\hat{\mathbf{X}}_{\mathrm{b}}+\mathbf{E}_{\mathrm{sm}}\right)^{\mathrm{T}}\left(\hat{\mathbf{X}}_{\mathrm{w}}+\hat{\mathbf{X}}_{\mathrm{b}}+\mathbf{E}_{\mathrm{sm}}\right)\right] \\
&= \mathrm{tr}\left[\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}+\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}+\mathbf{E}_{\mathrm{sm}}^{\mathrm{T}}\right)\left(\hat{\mathbf{X}}_{\mathrm{w}}+\hat{\mathbf{X}}_{\mathrm{b}}+\mathbf{E}_{\mathrm{sm}}\right)\right] \\
&= \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right)+\mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right)+\mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sm}}\right)+\mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right)+\mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right) \\
&\quad +\mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sm}}\right)+\mathrm{tr}\left(\mathbf{E}_{\mathrm{sm}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right)+\mathrm{tr}\left(\mathbf{E}_{\mathrm{sm}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right)+\mathrm{tr}\left(\mathbf{E}_{\mathrm{sm}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sm}}\right) \\
&= \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right)+\mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right)+\mathrm{tr}\left(\mathbf{E}_{\mathrm{sm}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sm}}\right)+2\times\mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right)+2\times\mathrm{tr}\left[\mathbf{E}_{\mathrm{sm}}^{\mathrm{T}}\left(\hat{\mathbf{X}}_{\mathrm{w}}+\hat{\mathbf{X}}_{\mathrm{b}}\right)\right] \\
&= \mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{w}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{w}}\right)+\mathrm{tr}\left(\hat{\mathbf{X}}_{\mathrm{b}}^{\mathrm{T}}\hat{\mathbf{X}}_{\mathrm{b}}\right)+\mathrm{tr}\left(\mathbf{E}_{\mathrm{sm}}^{\mathrm{T}}\mathbf{E}_{\mathrm{sm}}\right)
\end{aligned} \tag{36}$$

REFERENCES

1. Hoaglin DC, Tukey JW. *Understanding Robust and Exploratory Data Analysis.* Wiley: New York, 1982.
2. Brereton RG (ed.). *Multivariate Pattern Recognition in Chemometrics, Illustrated by Case Studies*. Elsevier: Amsterdam, 1992.
3. Myers RH, Montgomery DC. *Response Surface Methodology: Process and Product in Optimization Using Designed Experiments*. Wiley: New York, 1995.
4. Box GEP, Draper NR. *Empirical Model-building and Response Surfaces*. Wiley: New York, 1987.
5. Kresta JV, MacGregor JF, Marlin TE. Multivariate statistical monitoring of process operating performance. *Can. J. Chem. Engng* 1991; **69**: 35–47.
6. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; **37**: 41–59.
7. Jolliffe IT. *Principal Component Analysis*. Springer: Berlin, 1986.
8. Gurden SP, Martin EB, Morris AJ. The introduction of process chemometrics into an industrial pilot plant laboratory. *Chemometrics Intell. Lab. Syst.* 1998; **44**: 319–330.
9. Wold S, Esbensen KH, Geladi P. Principal component analysis. *Chemometrics Intell. Lab. Syst.* 1987; **2**: 37–52.
10. Nomikos P, MacGregor JF. Monitoring batch processes using multiway principal component analysis. *AIChE J.* 1994; **40**: 1361–1375.
11. Wise BM, Gallagher NB. The process chemometrics approach to process monitoring and fault detection. *J. Process Control* 1996; **6**: 329–348.
12. Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* 1986; **185**: 1–17.
13. Höskuldsson A. PLS regression methods. *J. Chemometrics* 1988; **2**: 211–228.
14. Kourti T, Nomikos P, MacGregor JF. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *J. Process Control* 1995; **5**: 277–284.
15. Harshman RA. Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers Phonet*. 1970; **22**: 111–117.
16. 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.
17. Law HG, Snyder Jr CW, Hattie JA, McDonald RP (eds). *Research Methods for Multimode Data Analysis*. Praeger: New York, 1984.
18. Bro R. PARAFAC. Tutorial and applications. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 149–171.
19. Ross RT, Leurgans SE. Component resolution using multilinear models. *Biochem. Spectrosc.* 1995; **246**: 679–700.

20. Westerhuis JA, Kourti T, MacGregor JF. Comparing alternative approaches for multivariate statistical analysis of batch process data. *J. Chemometrics* 1999; **13**: 397–413.
21. Louwerse DJ, Smilde AK. Multivariate statistical process control of batch processes using three-way models. *Chem. Engng Sci.* 2000; **50**: 1225–1235.
22. Kruskal JB, Harshman RA, Lundy ME. How 3-MFA data can cause degenerate PARAFAC solutions, among other relationships. In *Analysis of Multiway Data Matrices*, Coppi R, Bolasco S (eds). North-Holland: Amsterdam, 1989; 115–122.
23. Rayens WS, Mitchell BC. Two-factor degeneracies and a stabilization of PARAFAC. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 173–181.
24. Braake HABT, van Can HJL, Verbruggen HB. Semi-mechanistic modeling of chemical processes with neural networks. *Engng Appl. Artific. Intell.* 1998; **11**: 507–515.
25. Simutis R, Oliveira R, Manikowski M, de Azevedo SF, Lubbert A. How to increase the performance of models for process optimization and control. *J. Biotechnol.* 1997; **59**: 73–89.
26. Takane Y, Shibayama T. Principal component analysis with external information on both subjects and variables. *Psychometrika* 1991; **56**: 97–120.
27. Takane Y, Kiers HAL, de Leeuw J. Component analysis with different sets of constraints on different dimensions. *Psychometrika* 1995; **60**: 259–280.
28. Carroll JD, Pruzansky S, Kruskal JB. CANDELINC: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika* 1980; **45**: 3–24.
29. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; **31**: 279–311.
30. Boqué R, Smilde AK. Monitoring and diagnosing batch processes with multiway covariates regression models. *AIChE J.* 1999; **45**: 1504–1520.
31. Smilde AK, Wang Y, Kowalski BR. Theory of medium-rank second-order calibration with restricted-Tucker models. *J. Chemometrics* 1994; **8**: 21–36.
32. Kiers HAL, Smilde AK. Constrained three-mode factor analysis as a tool for parameter estimation with second-order instrumental data. *J. Chemometrics* 1998; **12**: 125–147.
33. Bijlsma S, Louwerse DJ, Smilde AK. Estimating rate constants and pure UV-vis spectra of a two-step reaction using trilinear models. *J. Chemometrics* 1999; **13**: 311–329.
34. Cheng KL, Young VY. Ultraviolet and visible absorption spectroscopy. In *Instrumental Analysis*, Christian GD, O'Reilly JE, (eds). Prentice-Hall: Englewood Cliffs, NJ, 1986; 161–211.
35. Kiers HAL. Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics* (in press).
36. Magnus JR, Neudecker H. Kronecker products, the vec operator and the Moore–Penrose inverse. In *Matrix Differential Calculus with Application in Statistics and Econometrics*. Wiley: New York, 1988; 27–39.
37. Bro R. Multi-way analysis in the food industry. Models, algorithms and applications. *PhD Thesis*, University of Amsterdam, 1998.
38. Kroonenberg PM, De Leeuw J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 1980; **45**: 69–97.
39. De Leeuw J, Young FW, Takane Y. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika* 1976; **41**: 471–503.
40. Rao CR, Mitra SK. *Generalized Inverse of Matrices and Its Applications*. Wiley: New York, 1971.
41. McDonald RP. A simple comprehensive model for the analysis of covariance structures: some remarks on applications. *Br. J. Math. Statist. Psychol.* 1980; **33**: 161–183.
42. Hopke PK, Xie Y, Paatero P. Mixed multiway analysis of airborne particle composition data. *J. Chemometrics* 1999; **13**: 343–352.
43. Box GEP. Some theorems on quadratic forms applied in the study of analysis of variance problems: effect of inequality of variance in one-way classification. *Ann. Math. Statist.* 1954; **25**: 290–302.
44. Jackson JE, Muldholkar GS. Control procedures for residuals associated with principal component analysis. *Technometrics* 1979; **21**: 341–349.
45. Miller P, Swanson RE, Heckler ChE. Contribution plots: a missing link in multivariate quality control. *Appl. Math. Comput. Sci.* 1998; **8**: 775–792.
46. MacGregor JF, Jaeckle Ch, Kiparissides C, Koutoudi M. Process monitoring and diagnosis by multi-block PLS methods. *AIChE J.* 1994; **40**: 826–838.
47. Kiers HAL, Takane Y, ten Berge JMF. The analysis of multitrait–multimethod matrices via constrained components analysis. *Psychometrika* 1996; **61**: 601–628.
48. Bro R, Sidiropoulos ND. Least squares algorithms under unimodality and non-negativity constraints. *J. Chemometrics* 1998; **12**: 223–247.

