JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 301–331

Multiway multiblock component and covariates regression models

Age K. Smilde[1]*†, Johan A. Westerhuis[1] and Ricard Boqué[2]

[1]*Process Analysis and Chemometrics, Department of Chemical Engineering, University of Amsterdam, NL-1018 WV Amsterdam, The Netherlands*
[2]*Department of Analytical and Organic Chemistry, Rovira i Virgili University of Tarragona, E-43005 Tarragona, Catalonia, Spain*

SUMMARY

In this paper the general theory of multiway multiblock component and covariates regression models is explained. Unlike in existing methods such as multiblock PLS and multiblock PCA, in the new proposed method a different number of components can be selected for each block. Furthermore, the method can be generalized to incorporate multiway blocks to which any multiway model can be applied. The method is a direct extension of principal covariates regression and therefore works in a simultaneous fashion in which a clearly defined objective criterion is minimized. It can be tuned to fulfil the requirements of the user. Algorithms to calculate the components will be presented. The method will be illustrated with two three-block examples and compared to existing approaches. The first example is with two-way data and the second example is with a three-way array. It will be shown that predictions are as good as with the existing methods, but because for most blocks fewer components are required, diagnostic properties of the method are improved. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: multiblock regression models; multiblock PLS; three-way analysis; principal covariates regression

INTRODUCTION

General

Some problems in chemistry can be solved by analyzing a data table of measurements. To perform such an analysis, techniques such as principal component analysis are available [1,2]. Component analyses pertain to analyzing a single set of data with respect to its underlying structure. The purposes of such analyses can be very different, such as exploring the data, finding interesting patterns or estimating underlying pure species spectra and concentration profiles.

Other problems can be transformed into a regression problem where two blocks of data are involved, e.g. in multivariate calibration the connection between spectroscopic data and a property of interest is sought using regression techniques such as multiple linear regression, partial least squares and principal covariates regression [3–6].

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

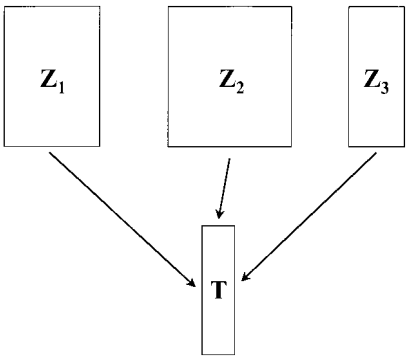

Figure 1. Example of a multiblock component problem.

Multiblock problems

Certain classes of chemical problems have to be solved by considering simultaneously more than two blocks of data. Such problems are called multiblock problems. There are two types of problems involved in multiblock data analysis. These types are visualized in Figures 1 and 2. The first type of problem in multiblock data analysis is a multiblock component problem (Figure 1). It is assumed that there is at least one mode in common between all the matrices involved, and component vectors are sought that summarize the information in all the matrices simultaneously as well as possible. Multiblock component problems can be encountered e.g. when the same set of samples is subjected to different instrumental analysis methods. If the same set of samples is analyzed in an IR and mass spectrometer, then a multiblock component model can give insight into what information about the samples is uniquely present in either the IR or the mass spectra, and what kind of information is common to the two instrumental methods.

The second type of problem is a multiblock regression problem. In Figure 2 an example is given in which two matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are used to predict $\mathbf{Y}$. Another regression problem is when multiple $\mathbf{Y}$s are predicted from a single $\mathbf{Z}$ (Figure 3). Multiblock regression problems arise if several parts of a reactor are modeled using process data to obtain multivariate control charts. Each part of the reactor generates a separate block of data [7]. Another example is the modeling of contaminants in a river when the different sources of contaminants are considered as different blocks of measurements [8]. A

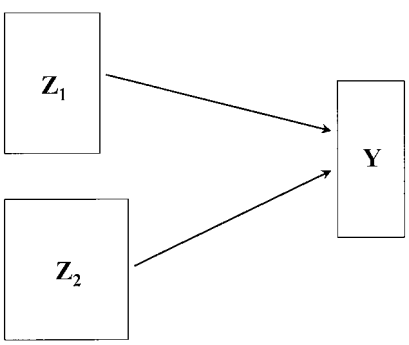

Figure 2. Example of a multiblock regression model with two predictor blocks and one response block.

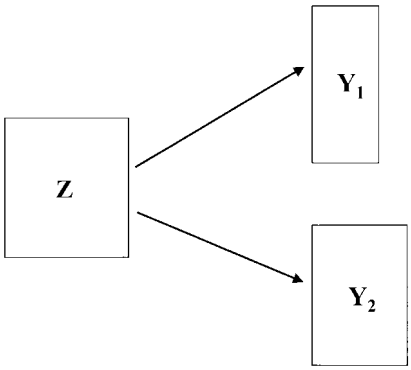

Figure 3. Example of a multiblock regression model with one predictor block and two response blocks.

third example is the manufacturing of tablets where granules are used as intermediate products. In this last example, three blocks of data are obtained: (1) the ingredients of the tablets, varied according to an experimental design; (2) the properties of the obtained granules; and (3) the properties of the obtained final tablets [9]. Obviously, in all the examples above, techniques should be used to model the data properly using the multiblock structure.

Summarizing, a component problem always has to do with finding common variation (or information) in multiple **Z** matrices, and a regression problem is always concerned with prediction of a **Y** matrix using a **Z** matrix.

There exist methods for tackling the multiblock component problem, such as multiblock PCA [10] and generalized Procrustes analysis [11]. Multiblock PLS is a method for multiblock regression problems [12,13]. Multiblock PCA and multiblock PLS work in a component-wise fashion, where one component at a time is calculated, in the philosophy of PLS [14]. Some theory is available regarding the behavior of different multiblock PCA and multiblock PLS methods [13].

Especially when dealing with multiblock problems, the purpose of the data analysis is an important issue. If the only purpose of the multiblock model is to predict a future value of **y** using predictors from several blocks, experience shows that it is usually better to put all the predictors in one superblock. For some multiblock PLS methods it can even be shown that the multiblock arrangement is mathematically equivalent to having one superblock [14].

Multiblock methods can serve other purposes than prediction. An appealing example of the usefulness of multiblock methods is in process monitoring [7,15] where the different blocks can be used to distinguish between process phases or operation units. Analyzing the data with a multiblock model then allows for detecting the location of the faults in a process. Hence, for plant-wide process monitoring, multiblock methods are useful tools. Exploring variation in complicated chemical systems where measurements have been performed with a certain ordering is another area where multiblock methods might be helpful [8].

Multiway problems

Another class of problems in chemistry generates three-way data. Three-way data are data that can be arranged meaningfully in a three-way table. Consider e.g. a batch process. If a batch process is running, measurements are made of the process variables at regular time points. The resulting data matrix has size (number of process variables $\times$ number of time points). If several batches have been finished, then it is meaningful to compare the performance of those batches, and this leads to

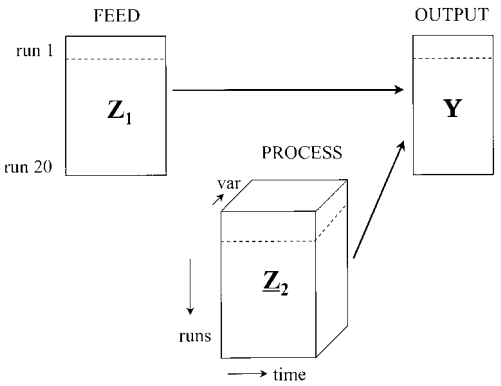

Figure 4. Example of a multiway multiblock regression problem.

analyzing a data set dimensioned (number of batches × number of process variables × number of time points), which is a three-way data array [16,17]. Another example is 3D-QSAR in which regression models are built between a multiway array and a measured activity of potential drugs [18]. Second-order calibration is an example in the field of analytical chemistry where three-way analysis tools are used for calibration [19]. In all these examples the use of specific multiway models helps in solving the problem.

Also in multiway analysis a distinction can be made between component and regression models. In component models the interest is in finding a simple component model for a three-way array. The purpose might again be e.g. exploration, finding patterns or estimating underlying spectra. In multiway regression the purpose is to build a model between a three-way array $\underline{\mathbf{Z}}$ and another array $\mathbf{Y}$ (possibly a vector, matrix or another three-way array). Such a model might be useful for predicting future y values from future z values [20]; to study the correlation between $\underline{\mathbf{Z}}$ and $\mathbf{Y}$; or to base control charts on [21,22]. There exist methods to deal with the multiway component problem [17,23–25]. Recently, methods to deal with the multiway regression problem have been developed [26–29].

In two-way analysis the structure imposed on a data matrix is usually that of a bilinear model. In three-way analysis, different structures can be imposed (see later). The choice of which structure to impose is problem- and data-dependent. In second-order calibration a trilinear structure should be imposed to have the second-order advantage [19]. Imposing a trilinear structure is also important in applications of fluorescence spectroscopy, owing to the properties of this kind of spectroscopy at low concentrations [20,30]. For batch process data the structure to impose on the three-way data is less clear [21,31,32].

Multiblock multiway problems

There are also multiway multiblock problems. An example of such a problem is visualized in Figure 4. Consider the same batch process as before. Not only are the measured process variables of the different batch runs available ($\underline{\mathbf{Z}}_2$), but also the initial quality properties of the feed ($\mathbf{Z}_1$) and the quality variables of the final product ($\mathbf{Y}$). Hence a natural question is whether there is a relation between $\mathbf{Z}_1$, $\underline{\mathbf{Z}}_2$ and $\mathbf{Y}$. If so, this relation should be modeled, and this model can then be used for multivariate statistical process control. An example of a multiway multiblock method is given by Kourti *et al*. [15] where the three-way arrays are rearranged to two-way arrays and treated as such. Hence in that approach a specific structure is imposed on the three-way array, which is in fact a Tucker1 structure as will become clear later in this paper.

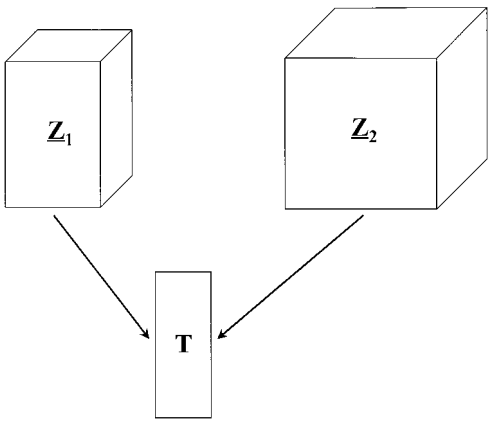

Figure 5. Example of a multiway multiblock component model.

Purpose of this paper

In this paper a general method for solving multiway multiblock component and regression problems is presented. The different types of multiway multiblock problems are visualized in Figures 5–7, which are the three-way counterparts of Figures 1–3. The general method is based on the simultaneous estimation of all components. The structure to impose on the three-way arrays involved can be chosen by the user. Hence a unified framework is given to tackle multiway multiblock component and regression problems. The theory will be presented, the algorithms will be outlined and two examples will be given.

THEORY

Multiway component models

Scalars will be written in italic lower-case characters; vectors in bold lower-case characters; matrices (two-way arrays) in bold upper-case characters; three-way and, in general, multiway arrays in

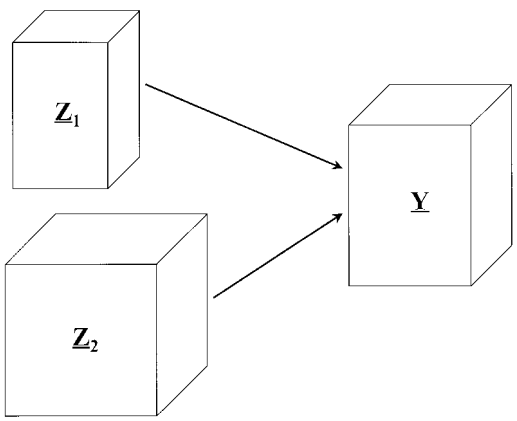

Figure 6. Example of a multiway multiblock regression model with two multiway predictor blocks and one multiway response block.

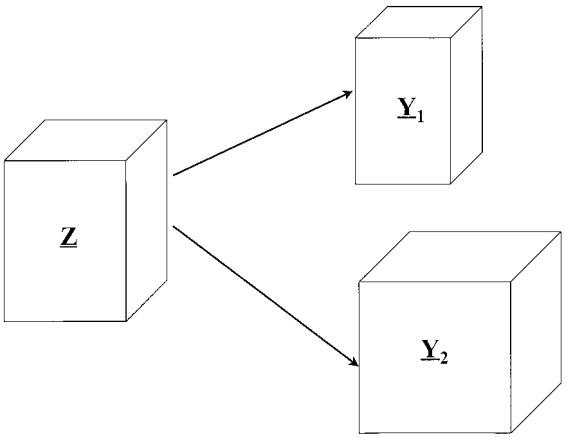

Figure 7. Example of a multiway multiblock regression model with one multiway predictor block and two multiway response blocks.

underlined bold upper-case characters. The notation for multiway analysis put forward by Kiers [33] will be used. For further information about notation and symbols, see Appendix I.

There are different ways to build multiway component models. These will be outlined briefly for building a model of the systematic variation in a three-way array $\underline{\mathbf{Z}}$ of size $I \times J \times K$.

One of the first models for a three-way array was developed by Tucker [34]. A good overview of these models is given by Kroonenberg [35]. There are different Tucker models, referred to as Tucker1, Tucker2 and Tucker3 models. For the three-way array $\underline{\mathbf{Z}}$ these models are

$$\mathbf{Z} = \mathbf{A}\mathbf{G}_{\mathrm{T1}} + \mathbf{E} \tag{1a}$$

$$\mathbf{Z} = \mathbf{A}\mathbf{G}_{\mathrm{T2}}(\mathbf{I} \otimes \mathbf{B})^{\mathrm{T}} + \mathbf{E} \tag{1b}$$

$$\mathbf{Z} = \mathbf{A}\mathbf{G}_{\mathrm{T3}}(\mathbf{C} \otimes \mathbf{B})^{\mathrm{T}} + \mathbf{E} \tag{1c}$$

where $\mathbf{Z} = [\mathbf{Z}_1|\ldots|\mathbf{Z}_K]$ is an $I \times JK$ matrix consisting of the K slices $\mathbf{Z}_k$ ($I \times J$) of the three-way array $\underline{\mathbf{Z}}$ next to each other. This will be referred to as '$\underline{\mathbf{Z}}$ properly matricized' if this arrangement is clear from the context [33]. The matrix $\mathbf{E}$ represents the unmodeled part of $\mathbf{Z}$.

Equation (1c) gives the Tucker3 model, where the loading matrices $\mathbf{A}$ ($I \times R_{\mathrm{A}}$), $\mathbf{B}$ ($J \times R_{\mathrm{B}}$) and $\mathbf{C}$ ($K \times R_{\mathrm{C}}$) serve as the summarizers of $\mathbf{Z}$, together with $\mathbf{G}_{\mathrm{T3}}$ ($R_{\mathrm{A}} \times R_{\mathrm{B}}R_{\mathrm{C}}$). The numbers of components in the three different modes are R_{A}, R_{B} and R_{C}. Usually $R_{\mathrm{A}} < I$, $R_{\mathrm{B}} < J$ and $R_{\mathrm{C}} < K$, thereby obtaining a dimension reduction of the original data array $\underline{\mathbf{Z}}$. The matrix $\mathbf{G}_{\mathrm{T3}}$ is the properly matricized three-way core array $\underline{\mathbf{G}}_{\mathrm{T3}}$ ($R_{\mathrm{A}} \times R_{\mathrm{B}} \times R_{\mathrm{C}}$), which is the generalization of the diagonal matrix containing the singular values in two-way analysis. Since a reduction is obtained in all three modes of $\underline{\mathbf{Z}}$, model (1c) is called the Tucker3 model. If only two modes are reduced, then a Tucker2 model appears (Equation (1b)), and if only one mode is reduced, then a Tucker1 model is obtained (Equation (1a)). Until recently, only Tucker1 models were used in chemometrics, since such models can be calculated easily with standard software. Yet, good algorithms exist for the Tucker2 and Tucker3 models, and these are gaining in popularity.

Another type of model for a three-way array was developed independently by Harshman [36] and Carroll and Chang [37]. This model is called a PARAFAC or CANDECOMP model and will be

referred to here as a PARAFAC model. The model for $\underline{\mathbf{Z}}$ ($I \times J \times K$) reads

$$\mathbf{Z} = \mathbf{A}\widetilde{\mathbf{I}}(\mathbf{C} \otimes \mathbf{B})^{\mathrm{T}} + \mathbf{E} \tag{2}$$

where $\mathbf{Z}$ ($I \times JK$) is the properly matricized $\underline{\mathbf{Z}}$, the same as above. The matrix $\widetilde{\mathbf{I}}$ is the matricized superidentity array $\boldsymbol{I}$, where $\boldsymbol{I}$ consists of ones on the superdiagonal and has all other elements equal to zero. The PARAFAC model is seen to be more restrictive than a Tucker3 model, and indeed, a PARAFAC model can be written as a special case of the Tucker3 model. The PARAFAC model is also sometimes referred to as the trilinear model, since it is trilinear in its parameters. Detailed discussion about differences between the Tucker models and the PARAFAC model can be found elsewhere [24,25].

Summarizing, all three-way models can be written as

$$\begin{aligned} \mathbf{Z} &= \mathbf{A}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \\ \text{(a)}\ \mathbf{P}^{\mathrm{T}} &= \mathbf{G}_{\mathrm{T1}}(\text{Tucker1}) \\ \text{(b)}\ \mathbf{P}^{\mathrm{T}} &= \mathbf{G}_{\mathrm{T2}}(\mathbf{I} \otimes \mathbf{B})^{\mathrm{T}}(\text{Tucker2}) \\ \text{(c)}\ \mathbf{P}^{\mathrm{T}} &= \mathbf{G}_{\mathrm{T3}}(\mathbf{C} \otimes \mathbf{B})^{\mathrm{T}}(\text{Tucker3}) \\ \text{(d)}\ \mathbf{P}^{\mathrm{T}} &= \widetilde{\mathbf{I}}(\mathbf{C} \otimes \mathbf{B})^{\mathrm{T}}(\text{PARAFAC}) \end{aligned} \tag{3}$$

where all matrices have been defined earlier [28,38]. Of course, the matrix $\mathbf{A}$ is different in all situations, but for convencience this is not reflected in the notation. The algorithms used to estimate the parameters in the three-way models are based on least squares. The estimates are obtained by minimizing the squared and summed entries of $\mathbf{E}$ with an alternating least squares algorithm (see Appendix IV).

Multiway multiblock component models

Two-way multiblock component problem. The general two-way multiblock component problem is visualized in Figure 1. In words, the problem is to find a common structure in $\mathbf{Z}_1,\ldots,\mathbf{Z}_M$. Obviously, there must be a connection between the data arrays: they should share a common mode, e.g. different measurements performed on the same samples. There are two types of methods for two-way multiblock component problems: component-wise methods and simultaneous methods.

Component-wise methods for two-way multiblock component problems. Component-wise methods solve the multiblock problem by calculating one component at a time. After calculating the first component, all matrices are deflated and the next component is calculated, much in the same way as in ordinary PCA with the NIPALS algorithm. Multiblock PCA is such a component-wise method. There are, however, different ways to deflate the matrices involved, leading to different multiblock solutions. Recently, Westerhuis *et al.* [13] published a paper clearing up some of the confusion.

Simultaneous method for two-way multiblock component problems. The multiblock component model presented in this paper is based on the ideas of principal covariates regression by De Jong and Kiers [5]. The general principal covariates regression model for a $\mathbf{Z}$ of $I \times J_{\mathrm{Z}}$ and a $\mathbf{Y}$ of $I \times J_{\mathrm{Y}}$ reads

$$
\begin{aligned}
&1.\ \mathbf{T} = \mathbf{ZW} \\
&2.\ \mathbf{Z} = \mathbf{TP}_{\mathrm{Z}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{Z}} \\
&3.\ \mathbf{Y} = \mathbf{TP}_{\mathrm{Y}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{Y}} \\
&4.\ \max_{\mathbf{W}}[\beta R_{\mathrm{Z}}^2 + (1-\beta)R_{\mathrm{Y}}^2] \text{ for a given } \beta \\
&5.\ R_{\mathrm{Z}}^2 = 1 - (\|\mathbf{Z} - \mathbf{ZWP}_{\mathrm{Z}}^{\mathrm{T}}\|^2 / \|\mathbf{Z}\|^2) \\
&6.\ R_{\mathrm{Y}}^2 = 1 - (\|\mathbf{Y} - \mathbf{ZWP}_{\mathrm{Y}}^{\mathrm{T}}\|^2 / \|\mathbf{Y}\|^2)
\end{aligned}
\tag{4}
$$

where β (between 0 and 1) is selected *a priori* and weights the importance of modeling **Z** or **Y**. The score matrix **T** has size $I \times R$, where R is usually much smaller than the minimum of I and J_{Z}. The sizes of the other matrices involved follow automatically. Problem (4) can be written in an equivalent and more convenient way as

$$
\begin{aligned}
&1.\ \mathbf{T} = \mathbf{ZW} \\
&2.\ \mathbf{Z} = \mathbf{TP}_{\mathrm{Z}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{Z}} \\
&3.\ \mathbf{Y} = \mathbf{TP}_{\mathrm{Y}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{Y}} \\
&4.\ \min_{\mathbf{W}}[\alpha\|\mathbf{Z} - \mathbf{ZWP}_{\mathrm{Z}}^{\mathrm{T}}\|^2 + (1-\alpha)\|\mathbf{Y} - \mathbf{ZWP}_{\mathrm{Y}}^{\mathrm{T}}\|^2] \text{ for a given } \alpha
\end{aligned}
\tag{5}
$$

where all matrices involved are as before in (4), and α is a weighting factor in between 0 and 1, serving the same goal as β in (4) (see Appendix II). From now on, formulation (5) will be used. Formulation (5) also shows a nice alternative way of interpreting the α weights. If $\alpha = 0$, all the emphasis is on using components **T** of **Z** for predicting **Y**. Whether or not these components are stable (i.e. good summarizers of **Z**) is not important. As soon as α is chosen to be non-zero, a penalty is put on the components **T**: they should, to some extent (controlled by the magnitude of α), be stable, i.e. a good summarizer of **Z**.

In the multiblock component model, scores of each block are defined, i.e. $\mathbf{T}_m = \mathbf{Z}_m \mathbf{W}_m$. The size of $\mathbf{T}_m$ is $I \times R_m$; hence the number of components in $\mathbf{T}_m$ is R_m, and this number can be different for each block m. These scores $\mathbf{T}_m$ should serve two goals: (i) explain the variation in their associated block m and (ii) convey similarities between the blocks. The latter goal is achieved by combining the scores in an overall score matrix **T** ($\mathbf{T} = [\mathbf{T}_1|\ldots|\mathbf{T}_M]$) from which the final R superscores $\mathbf{T}_{\mathrm{sup}}$ are obtained ($\mathbf{T}_{\mathrm{sup}} = \mathbf{TW}$). Then the multiblock component model is

$$
\begin{aligned}
&\mathbf{Z}_m = \mathbf{Z}_m\mathbf{W}_m\mathbf{P}_m^{\mathrm{T}} + \mathbf{E}_m = \mathbf{T}_m\mathbf{P}_m^{\mathrm{T}} + \mathbf{E}_m, \quad m = 1, \ldots, M \\
&\mathbf{T} = [\mathbf{Z}_1\mathbf{W}_1| \cdots |\mathbf{Z}_M\mathbf{W}_M] \\
&\mathbf{T} = \mathbf{TWP}_{\mathrm{T}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{T}} = \mathbf{T}_{\mathrm{sup}}\mathbf{P}_{\mathrm{T}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{T}} \\
&\min_{\mathbf{W}_{\mathrm{S}}} \left(\sum_{m=1}^{M} \alpha_m \|\mathbf{Z}_m - \mathbf{Z}_m\mathbf{W}_m\mathbf{P}_m^{\mathrm{T}}\|^2 + \alpha_{\mathrm{T}} \|\mathbf{T} - \mathbf{TWP}_{\mathrm{T}}^{\mathrm{T}}\|^2 \right)
\end{aligned}
\tag{6}
$$

where all α_m and α_{T} values are non-negative and these values sum to unity. These α_m values weight

the importance of modeling the individual matrices $\mathbf{Z}_m$. The $\left\|\mathbf{T} - \mathbf{TWP}_{\mathrm{T}}^{\mathrm{T}}\right\|^2$ part of the loss function minimizes the difference between $\mathbf{T}$ and $\mathbf{TWP}_{\mathrm{T}}^{\mathrm{T}}$ and thus maximizes the variance of $\mathbf{T}$ described by the superscores $\mathbf{T}_{\mathrm{sup}}$. This variation explained increases if $\mathbf{T}_1,\ldots,\mathbf{T}_M$ are more similar. Therefore increasing α_{T} forces the block scores to become more similar.

The restriction of $\mathbf{T}_{\mathrm{sup}}$ to be in the column space of $\mathbf{T}$, forced by using $\mathbf{T}_{\mathrm{sup}} = \mathbf{TW}$, is not active. This is shown in Appendix III. The components $\mathbf{T}_{\mathrm{sup}}$ will automatically lie in the column space of $\mathbf{T}$. For clarity, the multiblock model in Equation (6) is given in terms of the weight matrix $\mathbf{W}$.

In this multiblock component model the number of block scores, R_m, of each block can be selected differently. This is a distinct advantage when the dimensions of the blocks are rather different from each other, as is usually the case when two-way blocks are combined with three-way or multiway blocks. The number of final superscores, R, can also be selected independently from the number of block scores. Furthermore, each block can be given extra weight in the modeling by adjusting the appropriate α_m values. For some blocks it might be more difficult to obtain stable components than for other blocks, and in that case the corresponding α_m value should be increased.

For analysis of the multiblock component model the superscores give an idea of the common variation present in all blocks. The block scores give information corresponding to the specific block. Such a model could be used for monitoring a process where it is reasonable to divide the process variables into separate blocks. In the case of a process disturbance the superscores signal that the process is out of control. The block scores then give information on the specific part of the process that caused the process to be out of control. An algorithm for the multiway version of this multiblock component model can be found in Appendix IV.

Component-wise methods for multiway multiblock component problems. Component-wise methods for multiway multiblock component problems rely on the matricization of the three-way matrices into proper two-way matrices. Then the two-way methods dealt with above can be applied [15]. This is illustrated in Figure 8. A three-way multiblock component problem (Figure 5) is solved by matricizing the three-way arrays. This assumes a Tucker1 structure for all the three-way matrices involved.

Simultaneous methods for multiway multiblock component problems. Equation (6) can easily be extended to the multiway case. The problem is to find block score vectors $\mathbf{T}_m$ $(I \times R_m)$ in $\underline{\mathbf{Z}}_m$, where $\underline{\mathbf{Z}}_m$ has dimensions $I \times J_m \times K_m$ and the object modes (I) of all the three-way arrays are in common. The block scores are combined into a large score matrix $\mathbf{T}$ from which superscores $\mathbf{T}_{\mathrm{sup}}$ $(I \times R)$ are obtained. One way to solve this problem is to formulate it as (6), but with $\mathbf{P}_m$ now having a structure that can be Tucker1 (T1), Tucker2 (T2), Tucker3 (T3) or PARAFAC (PAR):

$$
\begin{aligned}
&\mathbf{Z}_m = \mathbf{T}_m\mathbf{P}_m^{\mathrm{T}} + \mathbf{E}_m = \mathbf{Z}_m\mathbf{W}_m\mathbf{P}_m^{\mathrm{T}} + \mathbf{E}_m, \quad m = 1,\ldots,M\\
&\mathbf{P}_m^{\mathrm{T}} = \text{T1,T2,T3 or PAR}\\
&\mathbf{T} = [\mathbf{Z}_1\mathbf{W}_1|\cdots|\mathbf{Z}_M\mathbf{W}_M]\\
&\mathbf{T} = \mathbf{TWP}_{\mathrm{T}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{T}} = \mathbf{T}_{\mathrm{sup}}\mathbf{P}_{\mathrm{T}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{T}}\\
&\min_{\mathbf{W}\mathrm{s}}\left(\sum_{m=1}^{M}\alpha_m\|\mathbf{Z}_m - \mathbf{Z}_m\mathbf{W}_m\mathbf{P}_m^{\mathrm{T}}\|^2 + \alpha_{\mathrm{T}}\|\mathbf{T} - \mathbf{TWP}_{\mathrm{T}}^{\mathrm{T}}\|^2\right)
\end{aligned}
\tag{7}
$$

where the minimization has to be performed over all $\mathbf{W}$s. The matrices $\mathbf{Z}_1,\ldots,\mathbf{Z}_M$ are the properly matricized three-way arrays $\mathbf{Z}_1,\ldots,\mathbf{Z}_M$. Several remarks are appropriate.

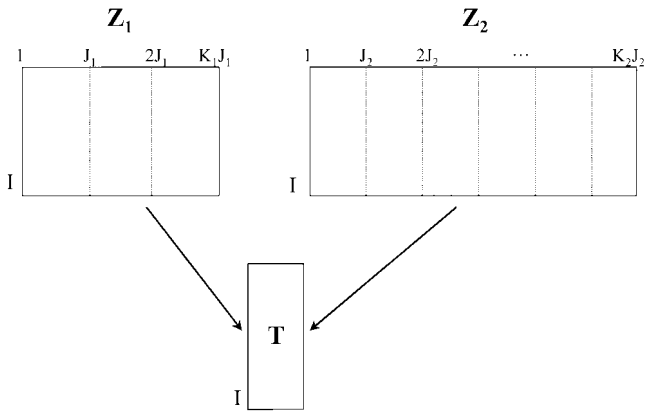

Figure 8. Three-way multiblock component problem: the matricizing solution.

1. For each three-way array $\underline{\mathbf{Z}}_m$ it can be decided to model it with a Tucker1 (T1) model ($\mathbf{P}_m^{\mathrm{T}} = \mathbf{G}_{\mathrm{T1},m}$), a Tucker2 (T2) model ($\mathbf{P}_m^{\mathrm{T}} = \mathbf{G}_{\mathrm{T2},m}(\mathbf{I} \otimes \mathbf{B})^{\mathrm{T}}$), a Tucker3 (T3) model ($\mathbf{P}_m^{\mathrm{T}} = \mathbf{G}_{\mathrm{T3},m}(\mathbf{C}_m \otimes \mathbf{B}_m)^{\mathrm{T}}$) or a PARAFAC (PAR) model ($\mathbf{P}_m^{\mathrm{T}} = \tilde{\mathbf{I}}(\mathbf{C}_m \otimes \mathbf{B}_m)^{\mathrm{T}}$).
2. The number of components in the batch direction, $R_{\mathrm{A},m}$, can be selected differently for each block.
3. The values $\alpha_1,\ldots,\alpha_M$ and α_{T} are non-negative and sum to unity. These values have to be chosen *a priori* and weight the importance of modeling the associated three-way array and the similarity of the block scores respectively.
4. Equation (7) can easily be extended to four-way arrays and higher.
5. The $\underline{\mathbf{Z}}_1,\ldots,\underline{\mathbf{Z}}_M$ arrays can have different number of modes, e.g. $\mathbf{Z}_1$ ($I \times J_1$), $\underline{\mathbf{Z}}_2$ ($I \times J_2 \times K_2$) and $\underline{\mathbf{Z}}_3$ ($I \times J_3 \times K_3 \times L_3$). The only requirement is that all arrays have at least one mode in common. A two-way $\mathbf{Z}_1$ can only be modeled with a Tucker1 structure.
6. It can be shown that $\mathbf{T}_{\mathrm{sup}}$ is automatically in the range of $\mathbf{T}$ (see Appendix III).
7. If T1, T2 or T3 models are chosen for all $\underline{\mathbf{Z}}_m$, then all $\mathbf{T}_m$ can be chosen orthogonal without changing the solution, owing to the rotational freedom of Tucker models. Imposing the constraint of orthogonality on $\mathbf{T}_m$ if $\underline{\mathbf{Z}}_m$ is modeled with PARAFAC makes the constraint active, because PARAFAC has no rotational freedom. $\mathbf{T}_{\mathrm{sup}}$ can always be chosen orthogonal because of rotational freedom in the $\mathbf{T} = \mathbf{TWP}_{\mathrm{T}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{T}}$ part of the model.
8. An alternating least squares (ALS) algorithm to solve this problem is given in Appendix IV.

Recommendations for selecting R_m, R, α_m and α_T. The number of metaparameters in the multiway multiblock component model is rather large. Therefore the *a priori* selection of all parameters should not be based on a cross-validation strategy but on common sense. Some rules of selecting the parameters are given below.

The number of components, $R_{m,\mathrm{A}}$, $R_{m,\mathrm{B}}$ and $R_{m,\mathrm{C}}$, can be obtained by building separate multiway component models of $\underline{\mathbf{Z}}_m$. The number of components will be the same in the multiblock model. The α_m values in the multiblock model can be used to give extra weight to some of the blocks based on previous information from the separate models. The number of final components, R, and α_{T} should be optimized using cross-validation of the multiblock model.

Multiway regression models

General problem. The general problem of multiway regression models is to find a quantitative relationship between a multiway array $\underline{\mathbf{Z}}$ (the predictor) and a multiway array $\underline{\mathbf{Y}}$ (the predictand). If $\mathbf{Z}$ and $\mathbf{Y}$ are both matrices, then an abundance of techniques are available to estimate linear relationships between $\mathbf{Z}$ and $\mathbf{Y}$, e.g. partial least squares regression, ridge regression, principal component regression and principal covariates regression [4–6,39].

For the case in which $\underline{\mathbf{Z}}$ and $\underline{\mathbf{Y}}$ are multiway (higher than two), only four methods exist. Three of these methods work in a component-wise fashion and the other works in a simultaneous fashion. The methods will be explained briefly using the example of a three-way $\underline{\mathbf{Z}}$ and a vector $\mathbf{y}$.

Component-wise solutions to the multiway regression problem. One of the earliest solutions to the problem of finding a relation between $\underline{\mathbf{Z}}$ ($I \times J \times K$) and $\mathbf{y}$ ($I \times 1$), which have the same I objects in common, is to matricize $\underline{\mathbf{Z}}$ to form a two-way array $\mathbf{Z}$ ($I \times JK$) and perform a regular PCR or PLS analysis. This amounts to assuming a Tucker1 model for $\mathbf{Z}$ and continuing in a component-wise PCR or PLS analysis.

Other component-wise solutions of the problem above were developed by Ståhle [26] and Bro [27,28]. Both decompose the $\underline{\mathbf{Z}}$ block into components (Bro into trilinear components, Ståhle into bilinear components), with the restriction that each component has the highest covariance with $\mathbf{y}$. After having found the first component, $\underline{\mathbf{Z}}$ and $\mathbf{y}$ are deflated. This is performed differently by Ståhle and Bro, resulting in orthogonal (Ståhle) or non-orthogonal (Bro) score vectors of the $\underline{\mathbf{Z}}$ block. Yet, essentially, both methods are the same, i.e. they give equal predictions [40]. This is analogous to the differences between the PLS methods of Wold (orthogonal $\mathbf{t}$ scores) and Martens (non-orthogonal $\mathbf{t}$ scores) [6].

Simultaneous solution for the multiway regression problem: multiway covariates regression. A solution called multiway covariates regression is suggested by Smilde [28] and worked out by Smilde and Kiers [29] as a direct generalization of principal covariates regression [5]. For a three-way $\underline{\mathbf{Z}}$ ($I \times J \times K$) and a univariate $\mathbf{y}$ ($I \times 1$) the problem is cast as follows:

$$\begin{aligned}
&\mathbf{T} = \mathbf{ZW} \\
&\mathbf{Z} = \mathbf{TP}_{\mathrm{Z}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{Z}} \\
&\mathbf{y} = \mathbf{Tp}_{\mathrm{y}} + \mathbf{E}_{\mathrm{y}} \\
&\mathbf{P}_{\mathrm{Z}}^{\mathrm{T}} = \text{T1, T2, T3 or PAR} \\
&\min_{\mathbf{W}} [\alpha \|\mathbf{Z} - \mathbf{ZWP}_{\mathrm{Z}}^{\mathrm{T}}\|^2 + (1-\alpha)\|\mathbf{y} - \mathbf{ZWp}_{\mathrm{y}}\|^2] \text{for a given } \alpha
\end{aligned} \tag{8}$$

where again α is between 0 and 1, has to be chosen *a priori* and has the same interpretation as earlier. The abbreviations T1, T2, T3 and PAR are the same as before. The matrix $\mathbf{Z}$ is the properly matricized three-way array $\underline{\mathbf{Z}}$, and $\mathbf{T}$ has size $I \times R$, where R is a small number. An ALS algorithm to solve (8) is given by Smilde and Kiers [29]. Note that $\mathbf{P}_{\mathrm{Z}}$ is allowed to follow different three-way structures, and generalizations to four-way and higher-order arrays for both $\underline{\mathbf{Z}}$ and $\underline{\mathbf{Y}}$ are straightforward. An application of the use of multiway covariates regression on batch process modeling is given elsewhere [21].

Multiway multiblock regression models

General problem. There are two general multiway multiblock regression problems. The first is predicting multiple $\underline{\mathbf{Y}}$ blocks with a single $\underline{\mathbf{Z}}$ block, and the second is predicting a single $\underline{\mathbf{Y}}$ block with multiple $\underline{\mathbf{Z}}$ blocks. These situations are visualized in Figures 7 and 6 respectively. An example of a problem which fits in this framework is batch MSPC; see Figure 4.

Component-wise methods for the multiway multiblock regression problem. Different methods to deal with the two-way multiblock regression problem (see Figures 2 and 3) have been published [7–9,12,14]. All these methods are extensions of PLS to the multiblock case and work in a component-wise fashion.

Multiway multiblock regression problems can be dealt with by matricizing the multiway matrices and using the two-way approaches. This approach assumes a Tucker1 structure for all three-way arrays involved. As stated in the Introduction (see subsection on 'Multiway problems'), it is sometimes preferable to have specific three-way models imposed on the three-way arrays involved. Consider multiway multiblock applications where one of the multiway arrays consists of fluorescence measurements or measurements from a hyphenated instrument (chromatography–spectroscopy). In that case it would make sense from a chemical point of view to give that array a trilinear (PARAFAC) structure [19,20]. In batch process monitoring applications there are indications that a Tucker3 or PARAFAC structure for the three-way batch process data is a reasonable alternative to a Tucker1 model [21,31,32].

Summarizing, it is worthwhile to relax the restriction of always having Tucker1 models for the multiway arrays involved and allow for other types of structures of the three-way arrays. This is one of the main goals of this paper.

Simultaneous methods for the multiway multiblock regression problem: multiway multiblock covariates regression models. The multiway multiblock covariates regression model, which is the main topic of this paper, will be explained first using a two-way example. This example is used in the Results section to illustrate the method. The extension toward multiway arrays is then straightforward, as will be shown at the end of this subsection.

Suppose two data sets $\mathbf{Z}_1$ ($I \times J_{Z1}$) and $\mathbf{Z}_2$ ($I \times J_{Z2}$) are available to predict $\mathbf{Y}$ ($I \times J_Y$). The idea is to take R_1 components $\mathbf{T}_{Z1}$ ($I \times R_1$) from $\mathbf{Z}_1$ by using $\mathbf{T}_{Z1} = \mathbf{Z}_1\mathbf{W}_{Z1}$, and similarly to take R_2 components $\mathbf{T}_{Z2}$ ($I \times R_2$) from $\mathbf{Z}_2$ by using $\mathbf{T}_{Z2} = \mathbf{Z}_2\mathbf{W}_{Z2}$. These two component matrices are combined in a matrix $\mathbf{T} = [\mathbf{T}_{Z1}|\mathbf{T}_{Z2}]$ ($I \times (R_1+R_2)$). From this matrix of scores, superscores are determined by using a weight matrix $\mathbf{W}$ ($(R_1+R_2) \times R$): $\mathbf{TW} = \mathbf{T}_{sup}$. Here R is the number of final superscores used to predict $\mathbf{Y}$. The superscores $\mathbf{T}_{sup}$ should predict $\mathbf{Y}$ well, but they should also be stable components from $\mathbf{T}$, because scores $\mathbf{T}_{Z1}$ and $\mathbf{T}_{Z2}$ may be rather correlated as they are both determined to predict $\mathbf{Y}$. Thus four regression problems should be solved simultaneously:

$$\begin{aligned}
\mathbf{Z}_1 &= \mathbf{Z}_1\mathbf{W}_{Z1}\mathbf{P}_{Z1}^{T} + \mathbf{E}_{Z1} \\
\mathbf{Z}_2 &= \mathbf{Z}_2\mathbf{W}_{Z2}\mathbf{P}_{Z2}^{T} + \mathbf{E}_{Z2} \\
\mathbf{T} &= \mathbf{TWP}_{T}^{T} + \mathbf{E}_{T} \\
\mathbf{Y} &= \mathbf{TWP}_{Y}^{T} + \mathbf{E}_{Y}
\end{aligned} \tag{9}$$

where $\mathbf{T} = [\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}]$ and the matrices $\mathbf{P}_{Z1}$, $\mathbf{P}_{Z2}$, $\mathbf{P}_T$ and $\mathbf{P}_Y$ are loading matrices for $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{T}$ and $\mathbf{Y}$ respectively. The regression problems in (9) are combined in one problem using α weights:

$$\begin{aligned}
&\min_{\mathbf{W}s} \left(\alpha_{Z1}\|\mathbf{Z}_1 - \mathbf{Z}_1\mathbf{W}_{Z1}\mathbf{P}_{Z1}^{T}\|^2 + \alpha_{Z2}\|\mathbf{Z}_2 - \mathbf{Z}_2\mathbf{W}_{Z2}\mathbf{P}_{Z2}^{T}\|^2\right.\\
&\qquad \left. + \alpha_{T}\|\mathbf{T} - \mathbf{T}\mathbf{W}\mathbf{P}_{T}^{T}\|^2 + \alpha_{Y}\|\mathbf{Y} - \mathbf{T}\mathbf{W}\mathbf{P}_{Y}^{T}\|^2\right)\\
&= \min_{\mathbf{W}s} \left(\alpha_{Z1}\|\mathbf{Z}_1 - \mathbf{Z}_1\mathbf{W}_{Z1}\mathbf{P}_{Z1}^{T}\|^2 + \alpha_{Z2}\|\mathbf{Z}_2 - \mathbf{Z}_2\mathbf{W}_{Z2}\mathbf{P}_{Z2}^{T}\|^2 + \alpha_{T}\|[\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}]\right.\\
&\qquad \left. - [\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}]\mathbf{W}\mathbf{P}_{T}^{T}\|^2 + \alpha_{Y}\|\mathbf{Y} - [\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}]\mathbf{W}\mathbf{P}_{Y}^{T}\|^2\right) \qquad (10)
\end{aligned}$$

where the αs are all non-negative and sum to unity. These α weights have to be chosen *a priori* or estimated with e.g. cross-validation. The selection of the α weights will be discussed below. Note that the **W**s are the only free parameters in (10), since the **T**s are determined by the **W**s, and the **P**s are determined by the **T**s and $\mathbf{Z}_1$, $\mathbf{Z}_2$ and **Y**. Using an alternating least squares approach, $\mathbf{W}_{Z1}$, $\mathbf{W}_{Z2}$ and **W** are updated sequentially until convergence is reached. Appendix IV shows the algorithm to solve Equation (10).

The model in Equation (10) can easily be extended to multiway arrays. An example of a two-way $\mathbf{Z}_1$ ($I \times J_{Z1}$), a three-way $\underline{\mathbf{Z}}_2$ ($I \times J_{Z2} \times K_{Z2}$) and a two-way array **Y** ($I \times J_Y$) will be discussed in the Results section of this paper. First, $\underline{\mathbf{Z}}_2$ is matricized to $\mathbf{Z}_2$ ($I \times J_{Z2}K_{Z2}$). Again, R_{Z1} and R_{Z2} components of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ respectively are derived using $\mathbf{T}_{Z1} = \mathbf{Z}_1\mathbf{W}_{Z1}$ and $\mathbf{T}_{Z2} = \mathbf{Z}_2\mathbf{W}_{Z2}$. These components should fit $\mathbf{Z}_1$ and $\mathbf{Z}_2$. The components are combined into the score matrix **T** from which superscores $\mathbf{T}_{sup}$ are obtained that describe both **T** and **Y**:

$$\begin{aligned}
\mathbf{Z}_1 &= \mathbf{Z}_1\mathbf{W}_{Z1}\mathbf{P}_{Z1}^{T} + \mathbf{E}_{Z1}\\
\mathbf{Z}_2 &= \mathbf{Z}_2\mathbf{W}_{Z2}\mathbf{P}_{Z2}^{T} + \mathbf{E}_{Z2}\\
\mathbf{P}_{Z2} &= \text{T1, T2, T3 or PAR}\\
\mathbf{T} &= [\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}]\\
\mathbf{T} &= \mathbf{T}\mathbf{W}\mathbf{P}_{T}^{T} + \mathbf{E}_{T}\\
\mathbf{Y} &= \mathbf{T}\mathbf{W}\mathbf{P}_{Y}^{T} + \mathbf{E}_{Y}
\end{aligned} \qquad (11)$$

where the abbreviations T1, T2, T3 and PAR refer to the structure of the matrix $\mathbf{P}_{Z2}$, having a Tucker1, Tucker2, Tucker3 or PARAFAC structure (see Equation (3)). The model parameters (**W**s) of (11) are estimated in the same way as presented for the two-way case.

If $\mathbf{P}_{Z2}$ has a Tucker structure, then the model has rotational freedom. For a $\mathbf{P}_{Z2}$ with a PARAFAC structure, no rotational freedom exists, because this would destroy the minimum sum of squares solution of the PARAFAC model. However, the superscores $\mathbf{T}_{sup}$ can always be rotated to orthogonality.

The problem formulation as shown in Equations (10) and (11) is generic. It can be used for any number of blocks, where each block can have a different number of ways and different multiway structures imposed.

Predictions can be made using the **W**s and **P**s. This will be outlined for the two-way example, following the model in Equation (10). Suppose new measurements are available for $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively $\mathbf{z}_{1new}$ ($J_{Z1} \times 1$) and $\mathbf{z}_{2new}$ ($J_{Z2} \times 1$); then

$$\begin{aligned}
\mathbf{t}_{\mathrm{Z1,new}}^{\mathrm{T}} &= \mathbf{z}_{\mathrm{1,new}}^{\mathrm{T}}\mathbf{W}_{\mathrm{Z1}} \\
\mathbf{t}_{\mathrm{Z2,new}}^{\mathrm{T}} &= \mathbf{z}_{\mathrm{2,new}}^{\mathrm{T}}\mathbf{W}_{\mathrm{Z2}} \\
\mathbf{t}_{\mathrm{sup,new}}^{\mathrm{T}} &= [\mathbf{t}_{\mathrm{Z1,new}}^{\mathrm{T}}|\mathbf{t}_{\mathrm{Z2,new}}^{\mathrm{T}}]\mathbf{W} \\
\hat{\mathbf{y}}_{\mathrm{new}}^{\mathrm{T}} &= \mathbf{t}_{\mathrm{sup,new}}^{\mathrm{T}}\mathbf{P}_{\mathrm{Y}}^{\mathrm{T}}
\end{aligned} \tag{12}$$

and residuals in all blocks can easily be obtained and used for diagnostic purposes:

$$\begin{aligned}
\mathbf{e}_{\mathrm{Z1,new}}^{\mathrm{T}} &= \mathbf{z}_{\mathrm{1,new}}^{\mathrm{T}} - \mathbf{t}_{\mathrm{Z1,new}}^{\mathrm{T}}\mathbf{P}_{\mathrm{Z1}}^{\mathrm{T}} \\
\mathbf{e}_{\mathrm{Z2,new}}^{\mathrm{T}} &= \mathbf{z}_{\mathrm{2,new}}^{\mathrm{T}} - \mathbf{t}_{\mathrm{Z2,new}}^{\mathrm{T}}\mathbf{P}_{\mathrm{Z2}}^{\mathrm{T}}
\end{aligned} \tag{13}$$

and if a new measurement is available for y ($\mathbf{y}_{\mathrm{new}}$), then a y residual can be calculated as $\mathbf{y}_{\mathrm{new}} - \hat{\mathbf{y}}_{\mathrm{new}}$. These predictions hold for both the two-way and the multiway case, the only difference being in the structure of $\mathbf{P}_{\mathrm{Z2}}$.

Recommendations for choosing the metaparameters in the multiway multiblock covariates regression models. In the multiway multiblock covariates regression models there are a number of metaparameters to select. These are: the number of components for each block, R_{Z1} and R_{Z2}; the number of final superscores, R; the α values for blocks $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{T}$ and $\mathbf{Y}$; and the scaling of the blocks. Although it is hard to give general rules for setting these parameters (as is the case for all procedures using metaparameters), it is possible to give some guidelines based on practical experience. More detailed information will be presented in the examples, but a general outline is given below.

For the two-way problems, all matrices are autoscaled and normalized to unit sum of squares prior to the analysis. Assuming two predictor blocks $\mathbf{Z}_1$ and $\mathbf{Z}_2$ and a single $\mathbf{Y}$ block, separate principal covariates regression models are built between each predictor block $\mathbf{Z}$ and the $\mathbf{Y}$ block. Cross-validation over α and the number of components is used to establish α_{Z1} and α_{Z2} and the predictive ability of each separate block. These results give also an indication of the number of components to use in the multiblock model.

Using the α values of the separate analyses, a good initial guess for the α values of multiblock covariates regression models can be obtained by keeping the ratios $\alpha_{\mathrm{Z1}}/\alpha_{\mathrm{Y}} = k_1$ and $\alpha_{\mathrm{Z2}}/\alpha_{\mathrm{Y}} = k_2$ the same in the multiblock model as in the separate models. The α values should obey the following rules:

$$\begin{aligned}
&1.\ \alpha_{\mathrm{Z1}} + \alpha_{\mathrm{Z2}} + \alpha_{\mathrm{T}} + \alpha_{\mathrm{Y}} = 1 \\
&2.\ \alpha_{\mathrm{Z1}}/\alpha_{\mathrm{Y}} = k_1 \\
&3.\ \alpha_{\mathrm{Z2}}/\alpha_{\mathrm{Y}} = k_2 \\
&4.\ \alpha_{\mathrm{T}} = \gamma\alpha_{\mathrm{Y}}
\end{aligned} \tag{14}$$

where the constants k_1 and k_2 are obtained from the separate single-block models. α_{T} is only related to α_{Y} by the parameter γ, which has to be optimized in the multiblock model. In the Results section, k_1, k_2 and γ will be optimized for the final multiblock model.

If there are R_1 components needed to predict $\mathbf{Y}$ from $\mathbf{Z}_1$, and R_2 for predicting $\mathbf{Y}$ from $\mathbf{Z}_2$, then the

final number of components in the multiblock model, R, will be in between $\max(R_1,R_2)$ and R_1+R_2. Finally, the difference in predictive abilities of the $\mathbf{Z}_1$ and $\mathbf{Z}_2$ blocks in the separate principal covariates regression models gives an indication of the importance of both blocks $\mathbf{Z}_1$, and $\mathbf{Z}_2$. These blocks are then block-scaled according to this importance.

The procedure for a multiway multiblock covariates regression model is similar, only the multiway block is treated differently, mainly in its preprocessing. In the example presented above with a three-way $\underline{\mathbf{Z}}_2$, this array is first centered across the mode of the objects (batches) and then scaled to unit sum of squares within the J mode (process variable mode). That is, each slab in the three-way array associated with variable j is given unit sum of squares. This is called slab scaling, and this type of scaling does not destroy the multilinear structure of $\underline{\mathbf{Z}}_2$ [R Bro, A K Smilde, submitted manuscript, 41]. After this scaling, the sum of squares of $\underline{\mathbf{Z}}_2$ equals J. Separate covariates regression models are built, where for the model connecting $\underline{\mathbf{Z}}_2$ with $\mathbf{Y}$ a multiway covariates regression model is used. This results in α values, indications of predictive abilities and number of components.

The multiway multiblock covariates regression model is now built using the information of the separate covariates regression models in the same way as for the two-way case. Hence α values can be obtained and a realistic range of the number of components. Block scaling is again applied according to the results with respect to the predictive abilities. A detailed example will be given in the second case study.

It is a good idea to cross-validate the multiway multiblock covariates regression model to some extent using a realistic range of the number of components and a range of α values. The initial settings of these metaparameters, however, prevent the cross-validation procedure from becoming impractical.

RESULTS

To illustrate the previous ideas, two applications of multiblock covariates regression will be presented. The first application deals with the manufacturing of pharmaceutical tablets by wet granulation. Here two blocks with predictor variables are present to predict tablet properties. This example only deals with two-way matrices. In the second example, process data from a batch process are used to predict the quality of the polymers. Here the history of the batch reactor and the batch process data themselves are divided into separate blocks. The process data in this case have a three-way structure, where for several different batches, eight process variables are measured over 116 consecutive time intervals.

Application 1. Wet granulation and tableting process

A multiblock covariates regression model was applied to process data of the manufacturing of pharmaceutical tablets by wet granulation. This is a two-step process: the first step consists of the granulation of the powder mixture to improve the tableting properties of the blend; the second step consists of the compression of the granules into tablets.

A total number of 53 batches are used to build the multiblock covariates regression model. Four process variables and two composition variables were varied according to a Box–Behnken experimental design [42]. In such a design, each process variable is varied on three levels, but no corner points are used, since the corner points are often the ones that turn out to be infeasible. Furthermore, the center point of the design was repeated seven times. The six design variables were combined into block $\mathbf{Z}_1$. Additionally, the quadratic terms of these six design variables were also included in block $\mathbf{Z}_1$. Previous work showed that these quadratic terms were important for the modeling of the tablet quality variables [43]. Block $\mathbf{Z}_2$ corresponds to 14 measured granule properties, such as flow properties and particle size distribution. Finally, two quality properties,

crushing strength and disintegration time of the tablets, were measured, logarithmically transformed and included in **Y**. The logarithmic transformation was used to remove the heteroscedastic noise structure of the two tablet properties.

Preliminary results. Before the application of multiblock covariates regression to the pharmaceutical process data, preliminary PLS2 and principal covariates regression models were developed and compared. Both response variables are modeled simultaneously. This may not be the optimal way for minimizing prediction errors. However, the main goal for these models is monitoring, and therefore a single model is preferred. The results of the preliminary principal covariates regression models are used to find the optimal number of components for each predictor block and to find the optimal scaling and initial guess of the α values for each block in the multiblock covariates regression model, as discussed in the strategy for setting the metaparameters in the multiblock covariates regression problem. The optimal number of components was determined by leave-one-batch-out cross-validation. For the principal covariates regression model the optimal α value was also determined in the cross-validation, where it was varied from 0·1 to 1 in steps of 0·1. The fraction of explained variance by the model, both in **Z** and **Y**, as well as the results of the cross-validation are shown in Table I for each of the optimal models. The $Q^2\mathbf{Y}$ values of the cross-validation are calculated as

$$Q^2\mathbf{Y} = 1 - \frac{\sum_{i=1}^{I}\sum_{j_Y=1}^{J_Y}\left(\mathbf{Y}_{i,j_Y}^{CV} - \hat{\mathbf{Y}}_{i,j_Y}^{CV}\right)^2}{\sum_{i=1}^{I}\sum_{j_Y=1}^{J_Y}\left(\mathbf{Y}_{i,j_Y}^{CV}\right)^2} \tag{15}$$

where $\mathbf{Y}_{i,j_Y}^{CV}(I \times J_Y)$ and $\hat{\mathbf{Y}}_{i,j_Y}^{CV}$ are matrices that contain the measured and predicted values of the cross-validation samples in scaled units.

Table I shows that PLS2 and principal covariates regression give almost equal results for both blocks. The optimal number of components and the percentage explained and cross-validation results are rather similar. Block $\mathbf{Z}_1$ describes the main part of **Y** (86%), while $\mathbf{Z}_2$ describes only 32%. With three components, only 45% of $\mathbf{Z}_1$ is described, whereas with only two components, 75% of $\mathbf{Z}_2$ could be explained. The optimal models could easily be determined from the cross-validation results.

For the principal covariates regression model the optimal α value is somewhat different for the two blocks. For $\mathbf{Z}_1$ an α of 0·6 was sufficient, indicating that it is relatively easy to draw stable components from this block, whereas an α of 0·95 for $\mathbf{Z}_2$ was necessary to find stable components. This could be expected, because it is easier to draw stable components from design variables with low correlations than it is to draw components from highly correlated granulation properties. If it is easy to find stable components, then these components can be focused to minimize **Y**, because the $(1-\alpha)\left\|\mathbf{Y} - \mathbf{T}\mathbf{P}_Y^T\right\|^2$ part of the minimization will be dominant when $1-\alpha$ is high. Figures 9 and 10 show the cross-validation results of the principal covariates regression models of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ respectively. In Figures 9(a) and 10(a), $R^2\mathbf{Y}$ and $Q^2\mathbf{Y}$ are given for a varying number of components with α set at the optimal value given in Table I. These results are similar to the ones obtained using PLS. In Figures 9(b) and 10(b), $R^2\mathbf{Y}$ and $Q^2\mathbf{Y}$ are given for varying α value at the optimal number of components. The effect of α is clearly shown in the plots. For the $\mathbf{Z}_1$ model an α higher than 0·8 leads to a large decrease in both $R^2\mathbf{Y}$ and $Q^2\mathbf{Y}$, whereas for the $\mathbf{Z}_2$ model, $Q^2\mathbf{Y}$ decreases when α_{Z2} decreases. Thus α_{Z2} must be chosen rather high in order not to lose prediction power.

The subspace described by the scores of the PLS and principal covariates regression models is almost equal for both $\mathbf{Z}_1$ and $\mathbf{Z}_2$ blocks. The variation in the PLS scores could be described for 99%

Table I. Fraction of explained variance for descriptor blocks $\mathbf{Z}_1$ and $\mathbf{Z}_2$ and quality variable $\mathbf{Y}$ for PLS2 and principal covariates regression (PCovR) models

Method	Block	Factors	α	$R^2\mathbf{Z}$	$R^2\mathbf{Y}$	$Q^2\mathbf{Y}$
PLS2	$\mathbf{Z}_1$	3		0·45	0·86	0·80
	$\mathbf{Z}_2$	2		0·75	0·32	0·25
PCovR	$\mathbf{Z}_1$	3	0·6	0·45	0·87	0·80
	$\mathbf{Z}_2$	2	0·95	0·76	0·31	0·26

using the principal covariates scores for both blocks. These results corroborate the previous results that for this specific example both methods give similar single-block models.

Comparison between multiblock covariates regression and multiblock PLS. The performance of the multiblock covariates regression model was compared to multiblock PLS (superscore method version [9]). Both quality properties of the tablets were included in the $\mathbf{Y}$ block just as was done in the preliminary results using PLS2 and principal covariates regression models. For both methods, equal scaling was applied to better study the similarities and differences between the methods. Each block was first autoscaled to zero mean and unit variance for all variables. Then each block was scaled to unit sum of squares. From the single-block regression results it was observed that the design variables block $\mathbf{Z}_1$ is clearly more predictive for the response than the granulation properties block $\mathbf{Z}_2$. As already stated in the strategy for setting the metaparameters of the multiblock covariates regression model, the weight of each block was set according to the validated explained variance of $\mathbf{Y}$. Therefore for the multiblock models the weight of $\mathbf{Z}_1$ was set to three times the weight of $\mathbf{Z}_2$. The weight of $\mathbf{Y}$ was chosen to be in the middle of the weights of the two predictor blocks. To give the blocks the weights as mentioned above, the following scaling was applied:

$$\left\|\mathbf{Z}_1\right\|^2 = 3, \qquad \left\|\mathbf{Z}_2\right\|^2 = 1, \qquad \left\|\mathbf{Y}\right\|^2 = 2 \tag{16}$$

Optimal α values for the principal covariates regression models of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ with $\mathbf{Y}$ were found to be around 0·6 and 0·95 respectively. The α values for the multiblock model were optimized around the values found for the single-block models. Figures 11(a) and 11(b) show the $R^2\mathbf{Y}$ and $Q^2\mathbf{Y}$ values for different settings of the α values. In Figure 11(a), k_1 was varied from 0·25 to 9, while k_2 was fixed at 19. In Figure 11(b), k_2 was varied from 1·5 to 99, while k_1 was fixed at 1·5. In all cases γ was maintained at 1, because no effect of γ was found. This means that any combination of components from block $\mathbf{T}$ will lead to stable superscores $\mathbf{T}_{\text{sup}}$ that predict both $\mathbf{T}$ and $\mathbf{Y}$ very well. The models are quite robust, i.e. a small change in k_1 or k_2 value hardly affects the predictive properties of the model. The optimal values of k_1 and k_2 were selected to be 1 and 4 respectively.

Table II shows the final results of the multiblock modeling with both the multiblock PLS and multiblock covariates regression models. The multiblock PLS model uses four components for each block, which was determined using cross-validation. The multiblock covariates regression model uses three components for block $\mathbf{Z}_1$ and two components for block $\mathbf{Z}_2$. The number of four final components was determined using cross-validation. The weight factors k_1, k_2 and γ were determined using cross-validation as shown in Figure 11. For both models, more of $\mathbf{Y}$ is explained and $Q^2\mathbf{Y}$ increased by combining $\mathbf{Z}_2$ and $\mathbf{Z}_1$. For multiblock PLS, four components are used to describe 93% of

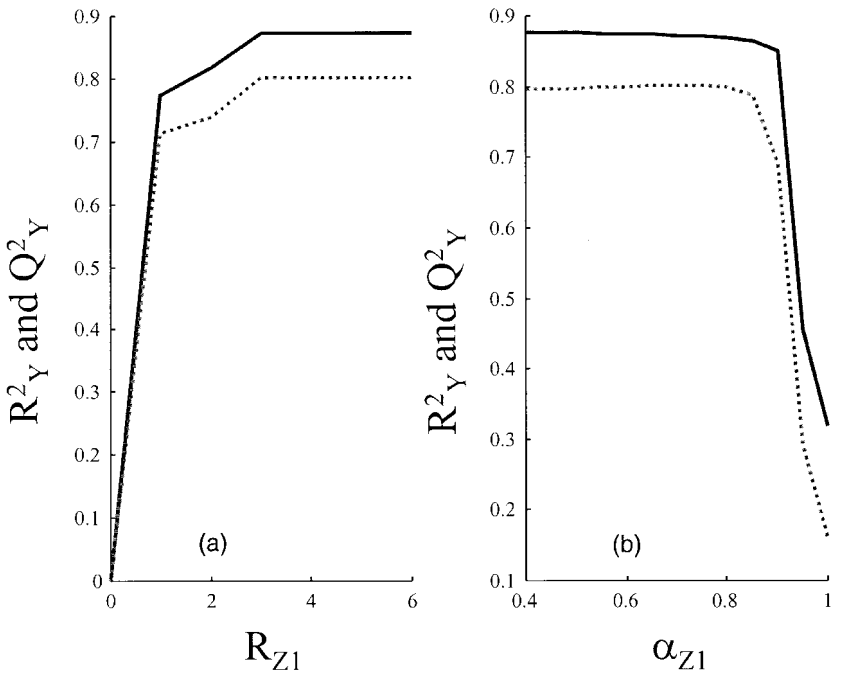

Figure 9. Results of principal covariates regression of **Y** on $\mathbf{Z}_1$ of the two-way example (—, $R^2\mathbf{Y}$; ·····, $Q^2\mathbf{Y}$; see text).

the variance in **Y**. For the multiblock covariates regression method, four components were also used to describe 94% of the variance in **Y**. In both cases $Q^2\mathbf{Y}$ equals 0·89, which means that both models have similar predictive properties.

In Table II the percentage explained of the predictor blocks is given for both the superscores ($R^2\mathbf{Z}_1$ and $R^2\mathbf{Z}_2$) and the block scores ($R^2\mathbf{Z}_1$(b) and $R^2\mathbf{Z}_2$(b)). The superscores are linear combinations of the combined block scores that are used to predict **Y**. In both methods the superscores are optimized to describe **Y**, and the block scores are optimized to describe the variation of the predictor blocks. In general, the block scores usually describe more variation in the predictor blocks than the superscores. However, in multiblock covariates regression the number of superscores can be higher than the number of block scores and therefore describe more of the variation of the predictor blocks. To study the models, both variations explained should be considered. The variation described by the block scores gives an idea of the rank of the block and the stability of the model. The variation described by the superscores gives information on what proportion of the components actually has predictive power for **Y**. In the multiblock covariates regression model, three block scores describe 45% of the variance in $\mathbf{Z}_1$, while four superscores describe only 34% of the variance of $\mathbf{Z}_1$. This indicates that

Table II. Results of final multiblock models using multiblock PLS (MB-PLS) and multiblock covariates regression (MBCovR) models

Method	R_{Z1}, R_{Z2}, R	k_1, k_2, γ	$R^2\mathbf{Z}_1$	$R^2\mathbf{Z}_1$(b)	$R^2\mathbf{Z}_2$	$R^2\mathbf{Z}_2$(b)	$R^2\mathbf{Y}$	$Q^2\mathbf{Y}$
MB-PLS	4, 4, 4		0·48	0·51	0·62	0·82	0·93	0·89
MBCovR	3 2, 4	1, 4, 1	0·34	0·45	0·80	0·76	0·94	0·89

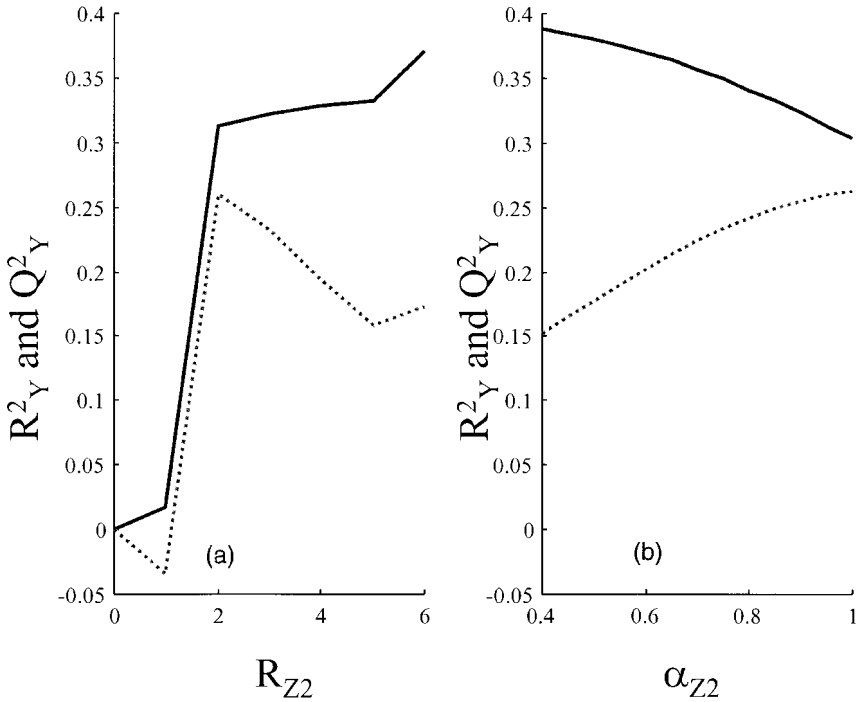

Figure 10. Results of principal covariates regression of **Y** on $\mathbf{Z}_2$ of the two-way example (—, $R^2\mathbf{Y}$; ·····, $Q^2\mathbf{Y}$; see text).

part of the subspace selected by the block scores is not predictive for **Y**, but stabilizes the model. For the multiblock PLS model the variance explained is almost similar for both sets of scores. Block $\mathbf{Z}_2$ is well described by only two components of the multiblock covariates model (76%); however, the multiblock PLS model needs four components because of the restriction of an equal number of components for each block.

Because the optimization criteria of the two methods differ, the space described by the several scores is also different. The subspace described by the superscores of multiblock PLS is described for 75% by the superscores of multiblock covariates regression, although both methods describe 93% of the variation in **Y**.

The multiblock PLS and multiblock covariates regression models both describe the variation in the tablet properties rather well. Predictive properties of the two models are comparable. The main advantage of the multiblock covariates regression model is in the lower number of block scores needed from each block to come to such a model. This lower number of block scores is an advantage for diagnostic purposes when a specific block needs to be examined. However, this advantage comes with a disadvantage of the *a priori* selection of a larger number of metaparameters.

Application 2. Batch process data

The second application describes a real polymerization batch process. Data were supplied by Dupont and have also been referred to previously by Nomikos and MacGregor [16] and Kosanovich *et al.* [44] The autoclave in this chemical batch process converts an aqueous effluent from an upstream evaporator into a polymer product. The recipe specifies reactor and heat source pressure trajectories

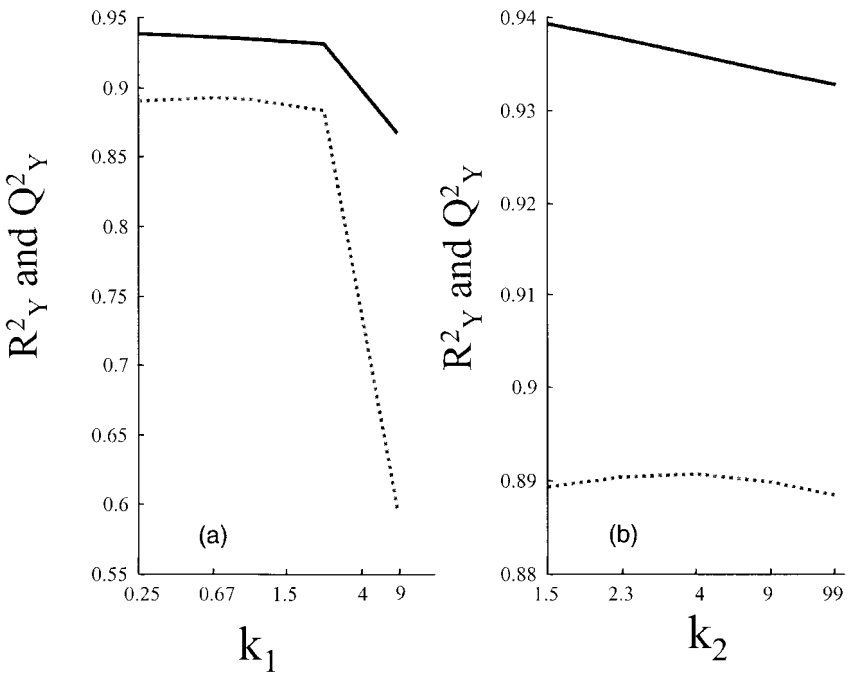

Figure 11. Results of multiblock principal covariates regression of the two-way example (—, $R^2\mathbf{Y}$; ·····, $Q^2\mathbf{Y}$; see text).

through five stages. Figure 12 shows the operating profiles for some of the variables measured in the process. The autoclave body pressure (P_{ab}) profile clearly shows the five different stages into which the process is divided. A multiblock model can serve as a starting point for a multivariate statistical process control strategy of this batch process.

Data are available from the autoclave process (**X**), from the autoclave process history (**Xh**) and from the polymer properties (**Y**). The variables in the respective blocks of data are listed in Table III.

Some pretreatment of the data was necessary. As the number of process measurements at each stage varies through the different batches, the raw data set was linearized to get, for each batch, the same number of measurements for each process variable at each stage. In this way a total of 116 time intervals resulted for every batch (9, 43, 22, 20 and 22 time intervals at each stage respectively). Variable 9 from **X**, the heat source pressure control setpoint, was skipped since it was found to be redundant. After screening of outliers, there were 42 batches found to be in normal operating conditions. Hence the sizes of the respective data sets were **X** ($42 \times 8 \times 116$), **Xh** (42×4) and **Y** (42×2). Measurements available from the upstream evaporator did not contain information to predict the polymer properties.

Initial calculations. Just as for the two-way case with the granulation data, also in this case initial calculations gave information on optimal α values and number of components. In addition, the block scaling can be applied according to the predictive abilities of the two prediction blocks. For the initial calculations both PLS2 and principal covariates regression between $\mathbf{Z}_1$ and **Y** are performed. For regressing **Y** on **X**, unfold PLS and multiway covariates regression are compared. Each column of the two-way matrices **Xh** and **Y** was autoscaled to zero mean and unit variance, and then the matrices were scaled to unit sum of squares. For the three-way array **X**, slab scaling

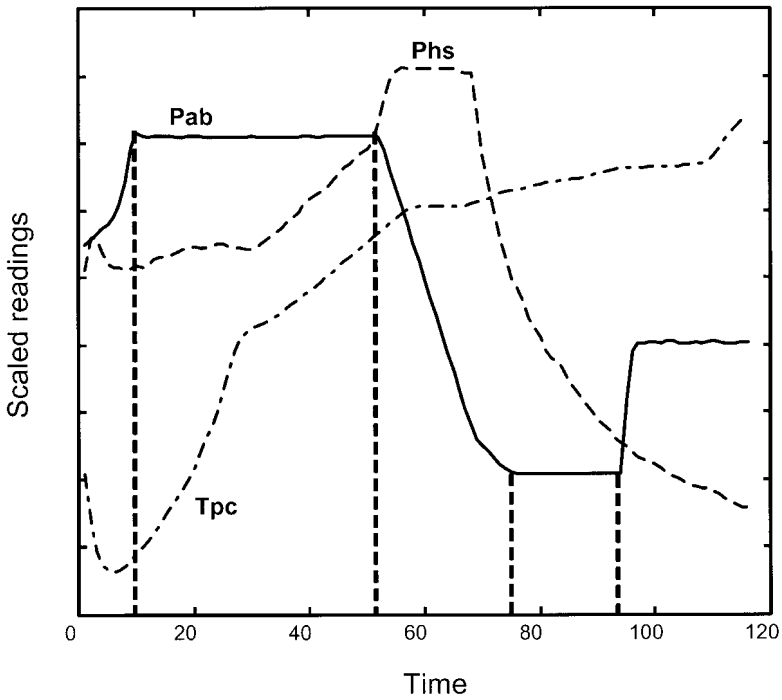

Figure 12. Typical profiles of batch process variables (see Table III(a)) during a batch run.

in the process variables mode is used after mean centering of each column. In this slab scaling, each process variable is scaled to have unit sum of squares. Thus the total sum of squares of $\underline{\mathbf{X}}$ after scaling equals the number of process variables, i.e. 8:

$$\left\|\mathbf{Xh}\right\|^2 = 1, \qquad \left\|\underline{\mathbf{X}}\right\|^2 = 8, \qquad \left\|\mathbf{Y}\right\|^2 = 1 \tag{17}$$

The results for both single-block models are presented in Table IV. The $Q^2\mathbf{Y}$ values are calculated using leave-three-batches-out cross-validation. Block **Xh** with the history of the autoclave is described by two factors in the principal covariates regression model, and a low value for α_{Xh} was found to be optimal. A low α value is usually connected to rather uncorrelated process variables. If it is easy to find stable components from a block, the loss function should be focused to minimize **Y** as well as possible. This was also noticed for block $\mathbf{Z}_1$ in the two-way example of the granulation data. A $Q^2\mathbf{Y}$ value of 0·22 is rather low. This means that not much information is available to predict the product quality. The results of the PLS and principal covariates regression models are almost equal.

For the three-way array $\underline{\mathbf{X}}$ with process data a Tucker3 model with three components in the batch direction, two in the process variable direction and three in the time direction was found to be optimal. A very high α was needed to find scores that had predictive power. High α values are usually needed for three-way arrays to give stable components that have to predict **Y** [29]. The multiway covariates regression model gives a higher $Q^2\mathbf{Y}$ value than the unfold PLS model. The PLS model seems to overfit the Y data, because the difference between $R^2\mathbf{Y}$ and $Q^2\mathbf{Y}$ is rather large. A Tucker3 (323) structure seems to better fit the X data than the unfold structure used by the PLS model. $Q^2\mathbf{Y}$ keeps

Table III. (a) Autoclave process measurements ($\underline{\mathbf{X}}$), (b) autoclave process history (**Xh**) and (c) polymer properties (**Y**)

Variable	Description
(a)	
Tpc	Polymer center temperature
Tps	Polymer side temperature
Tva	Vapor temperature
Pab	Autoclave body pressure
Ths	Heat source supply temperature
Thj	Heat source jacket vent temperature
Thc	Heat source coil vent temperature
Phs	Heat source supply pressure
Phs,sp	Heat source pressure control setpoint
(b)	
Timecyc	Autoclave cycle time, this batch
Timetot	Autoclave time since last cleaning
Times12	Autoclave time in stages 1 and 2
Tpc	Polymer temperature at end of maximum Phs
(c)	
MW	Molecular weight—aim
Ends	Titrated ends—aim

increasing if $\alpha_{\underline{\mathbf{X}}}$ increases, with $\alpha_{\underline{\mathbf{X}}} = 1$ giving the highest $Q^2\mathbf{Y}$ value. This indicates that a principal component regression type of model fits the data equally well [45]. However, in a multiblock covariates regression model this would mean that k_x equals infinity, which of course is not possible.

Multiway multiblock covariates regression. Multiway multiblock covariates regression was compared to multiblock unfold PLS regression. For both models the same scaling of the blocks was used, i.e. the two-way blocks **Xh** and **Y** were centered and scaled to unit sum of squares, and the three-way array was centered and each process variable slab was scaled to unit sum of squares. This means that the total sum of squares of **X** equals the number of process variables, i.e. 8 (see Equation (17)). It was decided not to use block scaling, because the $Q^2\mathbf{Y}$ values of the separate single-block regression models were similar for both blocks. In order to find the optimal k_1, k_2 and γ values for the multiblock, some models were cross-validated with different settings of k_1, k_2 and γ which were close to the results obtained from the single-block models. In this example there was some effect of increasing γ, and increasing γ gave an increased $Q^2\mathbf{Y}$. However, the model is quite robust, because small changes in k_1, k_2 and γ hardly affect the model.

Table V shows the final results of the modeling of the batch process data using multiway multiblock unfold PLS regression and multiway multiblock covariates regression with a Tucker3 (323) structure of the three-way array $\underline{\mathbf{X}}$. For the PLS model, two components were found to be optimal. However, $Q^2\mathbf{Y}$ was not improved by combining the two blocks. The results are similar to the results where only $\underline{\mathbf{X}}$ is used in a single-block regression model.

The multiway multiblock covariates regression model also uses two components for block **Xh**, but a Tucker3 (323) structure is used for the multiway array $\underline{\mathbf{X}}$, and three final components are selected to predict **Y**. The fit of **Y** ($R^2\mathbf{Y}$) is less than for the PLS multiblock model, but the predictions are somewhat better. This again shows that the Tucker3 (323) structure better fits the three-way batch

Table IV. Results of single-block regression models between **Xh** and **Y** using PLS and principal covariates regression (PCovR) and between $\underline{\mathbf{X}}$ and **Y** using unfold PLS (U-PLS) and multiway covariates regression (MCovR)

Method	Block	Factors	α	$R^2\mathbf{X}$	$R^2\mathbf{Y}$	$Q^2\mathbf{Y}$
PLS	**Xh**	2		0·61	0·34	0·22
PCovR	**Xh**	2	0·10	0·65	0·35	0·22
U-PLS	$\underline{\mathbf{X}}$	2		0·28	0·46	0·16
MCovR	$\underline{\mathbf{X}}$	323	0·99	0·41	0·36	0·23

data than the unfold structure used in the PLS model. The results for both multiblock models are similar to the multiway regression models between $\underline{\mathbf{X}}$ and **Y**. The influence of block **Xh** in the multiblock model is only small.

Figure 13 shows the effect of both k_{Xh} and $k_{\underline{\mathbf{X}}}$ on $R^2\mathbf{Y}$ and $Q^2\mathbf{Y}$. In Figure 13(a) it is shown that $Q^2\mathbf{Y}$ does not increase if k_{Xh} increases. This is caused by the high value of $k_{\underline{\mathbf{X}}}$. However, if the influence of $\underline{\mathbf{X}}$ is decreased by lowering $k_{\underline{\mathbf{X}}}$, then $Q^2\mathbf{Y}$ decreases as shown in Figure 13(b). $k_{\underline{\mathbf{X}}}$ should be very high to have a positive $Q^2\mathbf{Y}$ value. In fact, a $k_{\underline{\mathbf{X}}}$ value of infinity gives the highest $Q^2\mathbf{Y}$ value, which indicates that a component model predicts the polymer quality equally well. Yet, as was mentioned earlier, $k_{\underline{\mathbf{X}}} = \infty$ cannot be selected.

Summarizing, the two-way granulation data example showed that combining several blocks of information can improve the prediction of some product quality measurement. Multiblock PLS and multiblock covariates regression give similar results, but have different optimization criteria, resulting in different components. The multiblock covariates regression model needs some extra initial calculations, but can be fine-tuned to better fit the data.

In the three-way application of the batch process data, combining the different sets of information did not give an improvement in prediction quality. However, here the use of multiblock models can be focused on monitoring the different blocks simultaneously as described by MacGregor *et al.* [7] In this case a multiblock component model might be more appropriate than a regression model. This will be one of the topics in a follow-up paper on multiway multiblock covariates regression models.

CONCLUSIONS

A general methodology is presented for multiblock problems. Both component and regression models are considered. In both cases, multilinear models for each block of data are combined in a single optimization criterion. The method can accommodate any number of blocks, where each block can have a different number of modes and a different multiway structure. Furthermore, each block can be modeled using a different number of components, which usually means that using this method most blocks can be modeled with fewer components than with the existing methods. This is specifically helpful for diagnostic purposes.

Table V. Results of multiway multiblock PLS (MWMB-PLS) and multiway multiblock covariate regression (MWMBCovR) models for modeling of batch process data

Method	$r_{\mathrm{Xh}}, r_{\underline{\mathbf{X}}}, r$	$k_{\mathrm{Xh}}, k_{\underline{\mathbf{X}}}, \gamma$	$R^2\mathbf{Xh}$	$R^2\mathbf{Xh}$(b)	$R^2\underline{\mathbf{X}}$	$R^2\underline{\mathbf{X}}$(b)	$R^2\mathbf{Y}$	$Q^2\mathbf{Y}$
MWMB-PLS	2		0·33	0·54	0·27	0·27	0·47	0·18
MWMBCovR	2, 323, 3	0·05, 150, 5	0·37	0·71	0·41	0·41	0·36	0·23

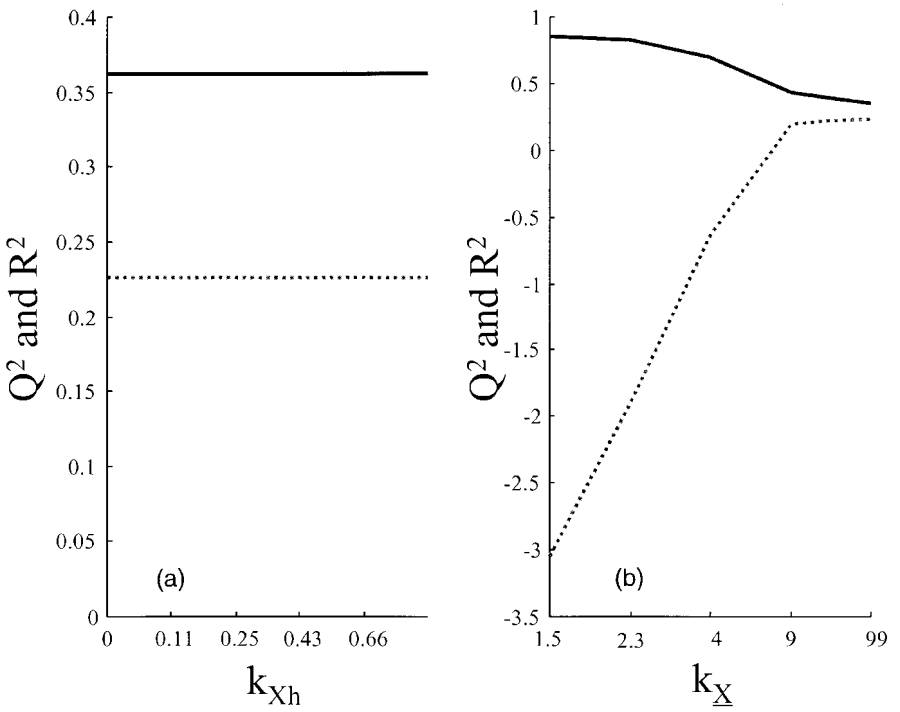

Figure 13. Results of multiway multiblock covariates regression of **Y** on **Xh** and $\underline{\mathbf{X}}$ of the three-way example (—, $R^2\mathbf{Y}$; ·····, $Q^2\mathbf{Y}$; see text).

The method is illustrated with two practical examples and compared to other multiblock methods as far as these exist. These examples show that sometimes an advantage is obtained in prediction ability when using multiblock methods. This is not always the case, as the second example shows. There are several reasons for performing multiblock regressions, e.g. prediction, building monitoring charts, getting insight into the structure of the data. Since multiblock methods are still very new, the relative merits of the proposed methodology have to be proven in practice. More research is needed to investigate properties of the presented methodology.

ACKNOWLEDGEMENTS

Thanks go to Henk Kiers (University of Groningen, The Netherlands) for making comments on an earlier version of this paper that considerably improved the quality of this version, and to Ken Dahl (Dupont, USA) for providing the batch polymerization data. The software used in this paper to calculate multiblock multiway component and covariates regression models is available from www-its.chem.uva.nl/research/pac/.

APPENDIX I

$\mathbf{Z}$	two-way data matrix (predictor)
$\underline{\mathbf{Z}}$	three-way data array (predictor)
$\mathbf{Z}_m$	*m*th two-way matrix in multiblock situation

$\underline{\mathbf{Z}}_m$	*m*th three-way array in multiblock situation
$\mathbf{Y}$	two-way data matrix (predictand)
$\underline{\mathbf{Y}}$	three-way data array (predictand)
$\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$	loading matrices in a three-way model
$\underline{\mathbf{G}}$	core array in a Tucker3 model
$\mathbf{T}$	component matrix
$\mathbf{W}$	weight matrix of T
$\mathbf{P}_Z$, P_{Zm}, $\mathbf{P}_Y$	loading matrices of $\mathbf{Z}$, $\mathbf{Z}_m$ and $\mathbf{Y}$
$\mathbf{E}_Z$, $\mathbf{E}_Y$	matrix of residuals for $\mathbf{Z}$ and $\mathbf{Y}$
i, j, k	running indices for the different modes of $\mathbf{Z}$ and $\mathbf{Y}$
m	running index for the $\mathbf{Z}$ blocks
r	running index for the components
I, J, J_Z, J_Y, K, J_m, K_m	sizes of the arrays
M	number of blocks
R_A, R_B, R_C	number of components in a Tucker model
R	number of final components in a model
α, β	weighting factors
k_1, k_2, γ	weighting factors for multiblock model
W_m, $\mathrm{W}_{\mathrm{Z}_\mathrm{m}}$	weight matrix of Z_m
$\mathrm{W}_{\#\mathrm{m}}$	part of W that corresponds to Z_m

APPENDIX II

In this appendix it is shown that problem (4) can be written as problem (5). Consider the maximization of problem (4):

$$\begin{aligned}&\max_{\mathbf{W}}[\beta R_\mathrm{Z}^2 + (1-\beta)R_\mathrm{Y}^2]\\ &= \max_{\mathbf{W}}\{\beta[1-(\|\mathbf{Z}-\mathbf{ZWP}_\mathrm{Z}^\mathrm{T}\|^2/\|\mathbf{Z}\|^2)] + (1-\beta)[1-(\|\mathbf{Y}-\mathbf{ZWP}_\mathrm{Y}^\mathrm{T}\|^2/\|\mathbf{Y}\|^2)]\}\end{aligned} \tag{18}$$

and, in terms of $\mathbf{W}$, problem (18) is equivalent to

$$\begin{aligned}&\min_{\mathbf{W}}\left[\left(\frac{\beta}{\|\mathbf{Z}\|^2}\right)\|\mathbf{Z}-\mathbf{ZWP}_\mathrm{Z}^\mathrm{T}\|^2 + \left(\frac{1-\beta}{\|\mathbf{Y}\|^2}\right)\|\mathbf{Y}-\mathbf{ZWP}_\mathrm{Y}^\mathrm{T}\|^2\right]\\ &= \min_{\mathbf{W}}\ (\alpha_1\|\mathbf{Z}-\mathbf{ZWP}_\mathrm{Z}^\mathrm{T}\|^2 + \alpha_2\|\mathbf{Y}-\mathbf{ZWP}_\mathrm{Y}^\mathrm{T}\|^2)\end{aligned} \tag{19}$$

where the non-negative weighing factors α_1 and α_2 are implicitly defined. Since only the relative sizes of α_1 and α_2 are of importance for finding $\mathbf{W}$ in problem (19), the values of α_1 and α_2 can be chosen to sum to unity, for convenience, which is done in the formulation of Equation (5). Hence problem (18) for a certain β can always be transformed to the more convenient problem (19).

APPENDIX III

Proof that $\mathbf{T}_\mathrm{sup}$ *is in the range of* $\mathbf{T}$ *in Equation (6)*

For the estimation of $\mathbf{W}$ in Equation (6), only the last part of the minimization function needs to be considered, and the solution to this problem (see Appendix IV) is

$$\begin{aligned}\mathbf{W} &= \mathbf{T}^{+}\mathbf{T}(\mathbf{P}_{\mathrm{T}}^{\mathrm{T}})^{+} = (\mathbf{P}_{\mathrm{T}}^{\mathrm{T}})^{+}\\ \mathbf{T}_{\mathrm{sup}} &= \mathbf{T}\mathbf{W} = \mathbf{T}(\mathbf{P}_{\mathrm{T}}^{\mathrm{T}})^{+}\end{aligned} \tag{20}$$

Thus $\mathbf{T}_{\mathrm{sup}}$ is always in the range of $\mathbf{T}$, for it is a projection of $\mathbf{T}(\mathbf{P}_{\mathrm{T}}^{\mathrm{T}})^{+}$ on the subspace generated by the columns of $\mathbf{T}$.

APPENDIX IV

In this appendix the different algorithms will be outlined. The algorithm to calculate Tucker3 models (TUCKALS) can be found elsewhere [46,47]. Likewise, ALS algorithms to estimate parameters in the PARAFAC model can be found elsewhere [36,37,48].

ALS algorithm to solve (7)

Here an ALS algorithm will be outlined for the multiway multiblock component problem (7), where only two blocks are present and Tucker3 models are assumed for the three-way arrays $\underline{\mathbf{Z}}$, which are both of size $I \times J \times K$, for convenience. Hence problem (7) becomes

$$\begin{aligned}&1.\ \mathbf{Z}_m = \mathbf{Z}_m\mathbf{W}_m\mathbf{P}_m^{\mathrm{T}} + \mathbf{E}_m, \quad m = 1,2\\ &2.\ \mathbf{P}_m^{\mathrm{T}} = \mathbf{G}_m(\mathbf{C}_m \otimes \mathbf{B}_m)^{\mathrm{T}}, \quad m = 1,2\\ &3.\ \mathbf{T} = [\mathbf{Z}_1\mathbf{W}_1|\mathbf{Z}_2\mathbf{W}_2]\\ &4.\ \mathbf{T} = \mathbf{T}\mathbf{W}\mathbf{P}_{\mathrm{T}}^{\mathrm{T}} + \mathbf{E}_{\mathrm{T}}\\ &5.\ \min_{\mathbf{W}\mathrm{s}}\left(\sum_{m=1}^{2}\alpha_m\|\mathbf{Z}_m - \mathbf{T}\mathbf{P}_m^{\mathrm{T}}\|^2 + \alpha_{\mathrm{T}}\|\mathbf{T} - \mathbf{T}\mathbf{W}\mathbf{P}_{\mathrm{T}}^{\mathrm{T}}\|^2\right)\end{aligned} \tag{21}$$

First the objective function (loss function) in (21) is rewritten as

$$\begin{aligned}&\min_{\mathbf{W}\mathrm{s}}(\alpha_1\|\mathbf{Z}_1 - \mathbf{Z}_1\mathbf{W}_1\mathbf{P}_1^{\mathrm{T}}\|^2 + \alpha_2\|\mathbf{Z}_2 - \mathbf{Z}_2\mathbf{W}_2\mathbf{P}_2^{\mathrm{T}}\|^2 + \alpha_T\|\mathbf{T} - \mathbf{T}\mathbf{W}\mathbf{P}_{\mathrm{T}}^{\mathrm{T}}\|^2)\\ &= \min_{\mathbf{W}\mathrm{s}}(\alpha_1\|\mathbf{Z}_1 - \mathbf{Z}_1\mathbf{W}_1\mathbf{P}_1^{\mathrm{T}}\|^2 + \alpha_2\|\mathbf{Z}_2 - \mathbf{Z}_2\mathbf{W}_2\mathbf{P}_2^{\mathrm{T}}\|^2\\ &\quad + \alpha_{\mathrm{T}}\|[\mathbf{Z}_1\mathbf{W}_1|\mathbf{Z}_2\mathbf{W}_2] - [\mathbf{Z}_1\mathbf{W}_1|\mathbf{Z}_2\mathbf{W}_2]\mathbf{W}\mathbf{P}_{\mathrm{T}}^{\mathrm{T}}\|^2)\end{aligned} \tag{22}$$

For the estimation of $\mathbf{W}_1$, only the terms including $\mathbf{W}_1$ need to be considered. Therefore the second term in the loss function can be neglected. The third term of the loss function can be rewritten as follows:

$$\begin{aligned}
&\min_{\mathbf{W}_1} \alpha_T \|[\mathbf{Z}_1\mathbf{W}_1|\mathbf{Z}_2\mathbf{W}_2] - [\mathbf{Z}_1\mathbf{W}_1|\mathbf{Z}_2\mathbf{W}_2]\mathbf{W}\mathbf{P}_T^T\|^2 \\
&= \min_{\mathbf{W}_1} \|\sqrt{\alpha_T}[\mathbf{Z}_1\mathbf{W}_1|\mathbf{Z}_2\mathbf{W}_2](\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)\|^2 \\
&= \min_{\mathbf{W}_1} \|\sqrt{\alpha_T}[\mathbf{Z}_1\mathbf{W}_1(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_1 + \mathbf{Z}_2\mathbf{W}_2(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_2]\|^2 \\
&= \min_{\mathbf{W}_1} \|\{-\sqrt{\alpha_T}[\mathbf{Z}_2\mathbf{W}_2(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_2]\} - \mathbf{Z}_1\mathbf{W}_1\sqrt{\alpha_T}(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_1\|^2 \qquad (23)
\end{aligned}$$

The last step is used to arrive at the form of $\mathbf{F} - \mathbf{Z}_1\mathbf{W}_1\mathbf{H}$ (see later). In (23), $\mathbf{I} - \mathbf{W}\mathbf{P}_T^T$ is partitioned into two parts: $(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_1$ corresponds to the first R_1 rows of $\mathbf{I} - \mathbf{W}\mathbf{P}_T^T$, and $(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_2$ corresponds to the last R_2 rows of $\mathbf{I} - \mathbf{W}\mathbf{P}_T^T$. The part $\mathbf{Z}_2\mathbf{W}_2(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_2$ is not affected by $\mathbf{W}_1$ and can be removed from the minimization.

The total minimization over $\mathbf{W}_1$ is then as follows:

$$\begin{aligned}
&\min_{\mathbf{W}_1}[\|\sqrt{\alpha_1}\mathbf{Z}_1 - \mathbf{Z}_1\mathbf{W}_1\sqrt{\alpha_1}\mathbf{P}_1^T\|^2 + \|\{-\sqrt{\alpha_T}[\mathbf{Z}_2\mathbf{W}_2(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_2]\} - \mathbf{Z}_1\mathbf{W}_1\sqrt{\alpha_T}(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_1\|^2] \\
&= \min_{\mathbf{W}_1} \|[\sqrt{\alpha_1}\mathbf{Z}_1|\{-\sqrt{\alpha_T}[\mathbf{Z}_2\mathbf{W}_2(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_2]\}] - \mathbf{Z}_1\mathbf{W}_1[\sqrt{\alpha_1}\mathbf{P}_1^T|\sqrt{\alpha_T}(\mathbf{I} - \mathbf{W}\mathbf{P}_T^T)_1]\|^2 \\
&= \min_{\mathbf{W}_1} \|\mathbf{F} - \mathbf{Z}_1\mathbf{W}_1\mathbf{H}\|^2 \qquad (24)
\end{aligned}$$

This is a Penrose problem which has the following known solution [49]:

$$\mathbf{W}_1 = \mathbf{Z}_1^+\mathbf{F}\mathbf{H}^+ \qquad (25)$$

where $\mathbf{H}^+$ is the pseudoinverse of $\mathbf{H}$. Once $\mathbf{W}_1$ is estimated and $\mathbf{T}_1$ is calculated ($\mathbf{T}_1 = \mathbf{Z}_1\mathbf{W}_1$), then $\mathbf{T}$ ($\mathbf{T} = [\mathbf{T}_1|\mathbf{T}_2]$) and $\mathbf{T}_{sup}$ ($\mathbf{T}_{sup} = \mathbf{T}\mathbf{W}$) should be updated and also $\mathbf{P}_1$ ($\mathbf{P}_1 = (\mathbf{T}_1^+\mathbf{Z}_1)^T$) and $\mathbf{P}_T$ ($\mathbf{P}_T = (\mathbf{T}_{sup}^+\mathbf{T})^T$) should be updated. In the next step, $\mathbf{W}_2$ can be estimated in the same way as $\mathbf{W}_1$. After $\mathbf{W}_2$ has been calculated, again $\mathbf{T}$, $\mathbf{T}_{sup}$, $\mathbf{P}_2$ ($\mathbf{P}_2 = (\mathbf{T}_2^+\mathbf{Z}_2)^T$) and $\mathbf{P}_T$ have to be updated. In the last step, $\mathbf{W}$ should be estimated. $\mathbf{W}$ is only present in the last part of the loss function:

$$\min_{\mathbf{W}}(\alpha_T\|\mathbf{T} - \mathbf{T}\mathbf{W}\mathbf{P}_T^T\|^2)$$

In the same way as in (25), $\mathbf{W} = \mathbf{T}^+\mathbf{T}(\mathbf{P}_T^T)^+ = (\mathbf{P}_T^T)^+$. After $\mathbf{W}$ is estimated, $\mathbf{T}_{sup}$ should be updated.

The ALS algorithm to solve (22) is as follows.

1: Initialize $\mathbf{T}_1$, $\mathbf{B}_1$,$\mathbf{C}_1$,…,$\mathbf{T}_M$, $\mathbf{B}_M$, $\mathbf{C}_M$, $\mathbf{T}_{sup}$
2: Update $\mathbf{G}_1$,…,$\mathbf{G}_M$
3: Calculate current error sum of squares
Until convergence of fit error

For $m = 1,\ldots,M$

4: Estimate $\mathbf{W}_m$
5: Update $\mathbf{T}_m$, $\mathbf{B}_m$, $\mathbf{C}_m$, $\mathbf{G}_m$, $\mathbf{T}$, $\mathbf{P}_T$
End

6: Estimate $\mathbf{W}$

7: Update $\mathbf{T}_{\text{sup}}$
8: Calculate fit error
End

Some comments are appropriate:

1. Initialization of $\mathbf{T}_m$ can be done by using the left singular vectors of $\mathbf{Z}_m$. $\mathbf{W}_m$ can be determined by $\mathbf{Z}_m^+\mathbf{T}_m$. $\mathbf{T}_{\text{sup}}$ can be initialized by using the left singular vectors of $[\mathbf{T}_1|\ldots|\mathbf{T}_M]$, and $\mathbf{W}$ equals $[\mathbf{T}_1|\ldots|\mathbf{T}_M]^+\mathbf{T}_{\text{sup}}$. Likewise, initial values of $\mathbf{B}_m$ and $\mathbf{C}_m$ can be found by taking singular vectors of a properly matricized matrix $\mathbf{Z}_m$.
2. Updating of the core array $\mathbf{G}_m$ is performed in the usual TUCKALS fashion by solving a Penrose problem [49].
3. The loading matrices $\mathbf{B}$ and $\mathbf{C}$ can be updated with the usual TUCKALS steps.
4. If all the matrices involved are two-way matrices, essentially the same algorithm can be used, but all $\mathbf{P}_m$ matrices can be updated with simple least squares steps.

Because each step in the above algorithm reduces the sum of squared errors, and since this sum of squared errors is bounded from below to zero, the ALS algorithm always converges. Like all iterative algorithms, it might converge to a local minimum. This can be checked by starting the algorithm with different starting values.

ALS algorithm to solve the multiblock regression problem in (10)

The multiblock regression problem as presented in Equation 10 is repeated here for convenience:

$$\begin{aligned}\min_{\mathbf{W}\text{s}}(&\alpha_{\text{Z}1}\|\mathbf{Z}_1-\mathbf{Z}_1\mathbf{W}_{\text{Z}1}\mathbf{P}_{\text{Z}1}^{\text{T}}\|^2+\alpha_{\text{Z}2}\|\mathbf{Z}_2-\mathbf{Z}_2\mathbf{W}_{\text{Z}2}\mathbf{P}_{\text{Z}2}^{\text{T}}\|^2+\alpha_{\text{T}}\|[\mathbf{Z}_1\mathbf{W}_{\text{Z}1}|\mathbf{Z}_2\mathbf{W}_{\text{Z}2}]\\&-[\mathbf{Z}_1\mathbf{W}_{\text{Z}1}|\mathbf{Z}_2\mathbf{W}_{\text{Z}2}]\mathbf{W}\mathbf{P}_{\text{T}}^{\text{T}}\|^2+\alpha_{\text{Y}}\|\mathbf{Y}-[\mathbf{Z}_1\mathbf{W}_{\text{Z}1}|\mathbf{Z}_2\mathbf{W}_{\text{Z}2}]\mathbf{W}\mathbf{P}_{\text{Y}}^{\text{T}}\|^2)\end{aligned}\tag{26}$$

For the solution, each weight matrix $\mathbf{W}_{\text{Z}1}$, $\mathbf{W}_{\text{Z}2}$ and $\mathbf{W}$ will be estimated sequentially in an iterative way until convergence. $\mathbf{W}_{\text{Z}1}$ only affects the first, third and fourth terms of the loss function. The third term in this loss function is the same as in the component model and was rewritten already in Equation (23). The total loss function for the minimization of $\mathbf{W}_{\text{Z}1}$ then equals

$$\begin{aligned}&\min_{\mathbf{W}_{\text{Z}1}}[\|\sqrt{\alpha_{\text{Z}1}}\mathbf{Z}_1-\mathbf{Z}_1\mathbf{W}_{\text{Z}1}\sqrt{\alpha_{\text{Z}1}}\mathbf{P}_{\text{Z}1}^{\text{T}}\|^2+\|\{-\sqrt{\alpha_{\text{T}}}[\mathbf{Z}_2\mathbf{W}_{\text{Z}2}(\mathbf{I}-\mathbf{W}\mathbf{P}_{\text{T}}^{\text{T}})_2]\}.\\&-\mathbf{Z}_1\mathbf{W}_{\text{Z}1}\sqrt{\alpha_{\text{T}}}(\mathbf{I}-\mathbf{W}\mathbf{P}_{\text{T}}^{\text{T}})_1\|^2+\|\sqrt{\alpha_{\text{Y}}}\mathbf{Y}-\mathbf{Z}_1\mathbf{W}_{\text{Z}1}\sqrt{\alpha_{\text{Y}}}\mathbf{W}_{\,1}\mathbf{P}_{\text{Y}}^{\text{T}}-\sqrt{\alpha_{\text{Y}}}\mathbf{Z}_2\mathbf{W}_{\text{Z}2}\mathbf{W}_{\,2}\mathbf{P}_{\text{Y}}^{\text{T}}\|^2]\\=&\min_{\mathbf{W}_{\text{Z}1}}\|[\sqrt{\alpha_{\text{Z}1}}\mathbf{Z}_1|\{-\sqrt{\alpha_{\text{T}}}[\mathbf{Z}_2\mathbf{W}_{\text{Z}2}(\mathbf{I}-\mathbf{W}\mathbf{P}_{\text{T}}^{\text{T}})_2]\}|\sqrt{\alpha_{\text{Y}}}(\mathbf{Y}-\mathbf{Z}_2\mathbf{W}_{\text{Z}2}\mathbf{W}_{\,2}\mathbf{P}_{\text{Y}}^{\text{T}})]\\&-\mathbf{Z}_1\mathbf{W}_{\text{Z}1}[\sqrt{\alpha_{\text{Z}1}}\mathbf{P}_{\text{Z}1}^{\text{T}}|\sqrt{\alpha_{\text{T}}}(\mathbf{I}-\mathbf{W}\mathbf{P}_{\text{T}}^{\text{T}})_1|\sqrt{\alpha_{\text{Y}}}\mathbf{W}_{\,1}\mathbf{P}_{\text{Y}}^{\text{T}}]\|^2\\=&\min_{\mathbf{W}_{\text{Z}1}}\|\mathbf{F}-\mathbf{Z}_1\mathbf{W}_{\text{Z}1}\mathbf{H}\|^2\end{aligned}\tag{27}$$

This again is a Penrose problem which has the following known solution:

$$\mathbf{W}_{\text{Z}1}=\mathbf{Z}_1^+\mathbf{F}\mathbf{H}^+\tag{28}$$

where $\mathbf{H}^+$ is the pseudoinverse of $\mathbf{H}$. $\mathbf{W}_{\,1}$ and $\mathbf{W}_{\,2}$ are those parts of $\mathbf{W}$ that correspond to $\mathbf{Z}_1$ and $\mathbf{Z}_2$

respectively. Once $\mathbf{W}_{Z1}$ is estimated and $\mathbf{T}_{Z1}$ is calculated ($\mathbf{T}_{Z1} = \mathbf{Z}_1\mathbf{W}_{Z1}$), then $\mathbf{T}$ ($\mathbf{T} = [\mathbf{T}_{Z1}|\mathbf{T}_{Z2}]$) and $\mathbf{T}_{sup}$ ($\mathbf{T}_{sup} = \mathbf{TW}$) should be updated and also $\mathbf{P}_{Z1}$ ($\mathbf{P}_{Z1} = (\mathbf{T}_{Z1}^{+}\mathbf{Z}_1)^{T}$), $\mathbf{P}_T$ ($\mathbf{P}_T = (\mathbf{T}_{sup}^{+}\mathbf{T})^{T}$) and $\mathbf{P}_Y$ ($\mathbf{P}_Y = (\mathbf{T}_{sup}^{+}\mathbf{Y})^{T}$) should be updated. In the next step, $\mathbf{W}_{Z2}$ can be estimated in the same way as $\mathbf{W}_{Z1}$ was updated in Equation (27). After $\mathbf{W}_{Z2}$ has been calculated, again $\mathbf{T}$, $\mathbf{T}_{sup}$, $\mathbf{P}_{Z2}$ ($\mathbf{P}_{Z2} = (\mathbf{T}_{Z2}^{+}\mathbf{Z}_2)^{T}$), $\mathbf{P}_T$ and $\mathbf{P}_Y$ have to be updated.

In the last step, $\mathbf{W}$ should be estimated. This can be done as follows:

$$\begin{aligned}
&\min_{\mathbf{W}}(\alpha_T\|\mathbf{T} - \mathbf{TWP}_T^T\|^2 + \alpha_Y\|\mathbf{Y} - \mathbf{TWP}_Y^T\|^2)\\
&\min_{\mathbf{W}}\|[\sqrt{\alpha_T}\mathbf{T}|\sqrt{\alpha_Y}\mathbf{Y}] - \mathbf{TW}[\sqrt{\alpha_T}\mathbf{P}_T^T|\sqrt{\alpha_Y}\mathbf{P}_Y^T]\|^2 \qquad (29)\\
&\min_{\mathbf{W}}\|\mathbf{F} - \mathbf{TWG}\|^2\\
&\mathbf{W} = \mathbf{T}^{+}\mathbf{FG}^{+}
\end{aligned}$$

After $\mathbf{W}$ is estimated, $\mathbf{T}_{sup}$, $\mathbf{P}_T$ and $\mathbf{P}_Y$ have to be updated.

In the alternating least squares algorithm, $\mathbf{W}_{Z1}$, $\mathbf{W}_{Z2}$ and $\mathbf{W}$ are estimated sequentially, and between the steps the corresponding scores and loadings have to be updated as described in the text above. Then each step, including the necessary updates, leads to a decrease in the total loss function, and thus the algorithm will converge because it has a lower bound of zero.

After convergence, both sets of block scores $\mathbf{Z}_1\mathbf{W}_{Z1}$ and $\mathbf{Z}_2\mathbf{W}_{Z2}$ and also the superscores $[\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}]\mathbf{W}$ can be rotated to orthogonality:

$$\begin{aligned}
\min_{\mathbf{Ws}}(&\alpha_{Z1}\|\mathbf{Z}_1 - \mathbf{Z}_1\mathbf{W}_{Z1}\mathbf{Q}_1\mathbf{Q}_1^{-1}\mathbf{P}_{Z1}^T\|^2 + \alpha_{Z2}\|\mathbf{Z}_2 - \mathbf{Z}_2\mathbf{W}_{Z2}\mathbf{Q}_2\mathbf{Q}_2^{-1}\mathbf{P}_{Z2}^T\|^2\\
&+ \alpha_T\|[\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}][\mathbf{Q}_1|\mathbf{Q}_2]^T - [\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}][\mathbf{Q}_1|\mathbf{Q}_2]^T\mathbf{WQ}_T\mathbf{Q}_T^{-1}\mathbf{P}_T^T\|^2\\
&+ \alpha_Y\|\mathbf{Y} - [\mathbf{Z}_1\mathbf{W}_{Z1}|\mathbf{Z}_2\mathbf{W}_{Z2}][\mathbf{Q}_1|\mathbf{Q}_2]^T\mathbf{WQ}_T\mathbf{Q}_T^{-1}\mathbf{P}_Y^T\|^2) \qquad (30)
\end{aligned}$$

Here $\mathbf{Q}_1$, $\mathbf{Q}_2$ and $\mathbf{Q}_T$ are non-singular rotation matrices. The block scores now become $\mathbf{Z}_1\mathbf{W}_{Z1}\mathbf{Q}_1$ and $\mathbf{Z}_2\mathbf{W}_{Z2}\mathbf{Q}_2$ and the corresponding loadings are $\mathbf{Q}_1^{-1}\mathbf{P}_{Z1}^T$ and $\mathbf{Q}_2^{-1}\mathbf{P}_{Z2}^T$. The superscores become $[\mathbf{Z}_1\mathbf{W}_{Z1}\mathbf{Q}_1|\mathbf{Z}_2\mathbf{W}_{Z2}\mathbf{Q}_2]\mathbf{WQ}_T$ and the loading for the response $\mathbf{Y}$ becomes $\mathbf{Q}_T^{-1}\mathbf{P}_Y^T$.

An algorithm to solve Equation (26) runs in the same way as presented above for the component model. The single addition to be made is between step 6 and step 7; the loadings $\mathbf{P}_T$ and $\mathbf{P}_Y$ should be updated.

REFERENCES

1. Jolliffe IT. *Principal Component Analysis*. Springer: Berlin, 1986.
2. Jackson JE. *A User's Guide to Principal Components*. Wiley: New York, 1991.
3. Draper NR, Smith H. *Applied Regression Analysis*. Wiley: New York, 1998.
4. Wold S, Ruhe A, Wold H, Dunn WJ. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Statist. Comput.* 1984; **5**: 735–743.
5. De Jong S, Kiers HAL. Principal covariatess regression. Part 1. Theory. *Chemometrics Intell. Lab. Syst.* 1992; **14**: 155–164.
6. Martens H, Næs T. *Multivariate Calibration*. Wiley: Chichester, 1989.
7. MacGregor JF, Jaeckle Ch, Kiparissides C, Koutoudi M. Process monitoring and diagnosis by multiblock PLS methods. *AIChE J.* 1994; **40**: 826–838.

8. Frank IE, Kowalski BR. A multivariate method for relating groups of measurements connected by a causal pathway. *Anal. Chim. Acta* 1985; **167**: 51–63.
9. Westerhuis JA, Coenegracht PMJ. Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares. *J. Chemometrics* 1997; **11**: 367–377.
10. Wold S, Hellberg S, Lundstedt T, Sjostrom M, Wold H. PLS modeling with latent variables in two or more dimensions. *Proc. Symp. on PLS Model Building*, Frankfurt am Main, 1987; 1–21.
11. Gower JC. Generalized Procrustes analysis. *Psychometrika* 1975; **40**: 33–51.
12. Wangen LE, Kowalski BR. A multiblock partial least squares algorithm for investigating complex chemical systems. *J. Chemometrics* 1988; **3**: 3–20.
13. Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemometrics* 1998; **12**: 301–321.
14. Wold S, Kettaneh-Wold N, Tjessem K. Hierarchical multiblock PLS and PC models, for easier model interpretation, and as an alternative to variable selection. *J. Chemometrics* 1996; **10**: 463–482.
15. 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.
16. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; **37**: 41–59.
17. Smilde AK. Three-way analyses: problems and prospects. *Chemometrics Intell. Lab. Syst.* 1992; **15**: 143–157.
18. Nilsson J, De Jong S, Smilde AK. Multiway calibration in 3D QSAR. *J. Chemometrics* 1997; **11**: 511–524.
19. Sanchez E, Kowalski BR. Tensorial calibration: 2. Second-order calibration. *J. Chemometrics* 1988; **2**: 265–280.
20. Bro R. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemometrics Intell. Lab. Syst.* 1998; **46**: 133–147.
21. Boqué R, Smilde AK. Monitoring and diagnosing batch processes with multiway regression models. *AIChE J.* 1999; **45**: 1504–1520.
22. Nomikos P, MacGregor JF. Multiway partial least squares in monitoring batch processes. *Chemometrics Intell. Lab. Syst.* 1995; **30**: 97–108.
23. Bro R. PARAFAC. Tutorial and applications. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 149–171.
24. Law HG, Snyder CW, Hattie JA, McDonald RP. *Research Methods for Multimode Data Analysis*. Praeger: New York, 1984.
25. Coppi R, Bolasco S. *Multiway Data Analysis*. North-Holland: Amsterdam, 1989.
26. Ståhle L. Aspects of the analysis of three-way data. *Chemometrics Intell. Lab. Syst.* 1989; **7**: 95–100.
27. Bro R. Multiway calibration. Multi-linear PLS. *J. Chemometrics* 1996; **10**: 47–61.
28. Smilde AK. Comments on multilinear PLS. *J. Chemometrics* 1997; **11**: 367–377.
29. Smilde AK, Kiers HAL. Multiway covariatess regression models. *J. Chemometrics* 1999; **13**: 31–48.
30. Ross RT, Lee CH, Davis CM, Ezzedine BH, Fayyad EA, Leurgans SE. Resolution of the fluorescence spectra of plant pigment-complexes using trilinear models. *Biochim. Biophys. Acta* 1991; **1056**: 317–320.
31. Westerhuis JA, Kourti T, MacGregor JF. Comparing alternative approaches for multivariate statistical analysis of batch process data. *J. Chemometrics* 1999; **13**: 397–413.
32. Louwerse DJ, Smilde AK. Multivariate statistical process control of batch processes using three-way models. *Chem. Engng. Sci.* 2000; **55**: 1225–1235.
33. Kiers HAL. Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics* 2000; **14**: 105–122.
34. Tucker LR. In *Contributions to Mathematical Psychology*, Frederiksen N, Gulliksen H (eds). Holt, Rinehart and Winston: New York, 1964; 110–182.
35. Kroonenberg PM. *Three-mode Principal Component Analysis Theory and Applications*. DSWO Press: Leiden, 1983.
36. Harshman RA. Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multi-modal factor analysis. *UCLA Working Papers Phonet*. 1970; **16**.
37. 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.
38. Kiers HAL. Hierarchical relations among three-way methods. *Psychometrika* 1991; **56**: 449–470.
39. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 1990; **12**: 55–67.
40. De Jong S. Regression coefficients in multilinear PLS. *J. Chemometrics* 1998; **12**: 77–81.

41. Harshman RA, Lundy ME. In *Research Methods for Multimode Data Analysis*, Law HG, Snyder CW, Hattie JA, McDonald RP (eds). Praeger: New York, 1984; 216–284.
42. Box GEP, Behnken DW. Some new three-level designs for the study of quantitative variables. *Technometrics* 1960; **2**: 455–475.
43. Westerhuis JA, de Haan P, Zwinkels J, Jansen WT, Coenegrach CF, Lerk CF. Optimisation of the composition and production of mannitol/microcrystalline cellulose tablets. *Int. J. Pharmaceut.* 1996; **143**: 151–162.
44. Kosanovich KA, Dahl KS, Piovoso MJ. Improved process understanding using multiway principal component analysis. *Ind. Engng. Chem. Res.* 1996; **35**: 138–146.
45. Geladi P, Xie YL, Polissar A, Hopke PK. Regression on parameters from three-way decomposition. *J. Chemometrics* 1998; **12**: 337–354.
46. Kroonenberg PM, De Leeuw J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 1980; **45**: 69–97.
47. Kiers HAL, Kroonenberg PM, Ten Berge JMF. An efficient algorithm for TUCKALS3 on data with large numbers of observation units. *Psychometrika* 1992; **57**: 415–422.
48. Kiers HAL, Krijnen WP. An efficient algorithm for PARAFAC of three-way data with large numbers of observation units. *Psychometrika* 1991; **56**: 147–152.
49. Penrose R. On the best approximate solutions of linear matrix equations. *Proc. Camb. Philos. Soc.* 1956; **52**: 17–19.

