

Multiway methods to explore and model NIR data from a batch process

Laila Stordrange^{a,*}, Tarja Rajalahti^{a,1}, Fred Olav Libnau^b

^aDepartment of Chemistry, University of Bergen, Allégt. 41, N-5007, Bergen, Norway

^bAmersham Health AS, N-4521 Spangereid, Norway

Received 15 June 2003; received in revised form 25 October 2003; accepted 30 October 2003

Abstract

Multiway methods are tested for their ability to explore and model near-infrared (NIR) spectra from a pharmaceutical batch process. The study reveals that blocking of data having a nonlinear behaviour into higher-order array can improve the predictive ability. The variation in each control point is independently modelled and N -way techniques overcome the nonlinearity problem. Important issues as variable selection and how to fill in for missing values have been discussed. Variable selection was shown to be essential to be able to perform multiway modelling. For spectra not yet monitored, use of mean spectra from calibration set gave close to the best results. Decomposing the spectra by N -way techniques gave additional information about the chemical system. To support the results simulated data sets were used. © 2004 Elsevier B.V. All rights reserved.

Keywords: Near-infrared spectroscopy (NIR); N -way methods; PARAFAC; Tucker3; N-PLS; Missing data

1. Introduction

Near-infrared spectroscopy (NIR) is a fast and non-destructive technique that has obtained a strong position, among others, in the pharmaceutical, polymers, wood and food industry. Generally, NIR spectroscopy is used in combination with multivariate techniques for qualitative [14,21] or quantitative analysis [4,10,23]. Many chemical processes are batch productions, i.e., a vessel is fed according to a recipe and after a certain time the product has obtained a specified quality and the reaction is terminated. To obtain a product of high quality and to know when to terminate a process the NIR technique can be used to monitor one or several process parameters. Spectra are acquired of several samples (objects) comprising expected variation in process. The resulting matrix (objects \times variables) combined with multivariate techniques gives a model used to monitor process variables in future batches. Batch process data can be organized into higher-order arrays, e.g., blocking of batch process data gives a three-way data set of batch \times time \times variable (see Fig. 1). Organizing data in a three-way array is expected to give

additional knowledge of variation in time and batch direction. For modelling data of higher-order array, N -way techniques are necessary.

The first N -way techniques were proposed by psychometricians and date back to the 1960s. Application of N -way techniques for solving chemical problems is on the other hand relatively new. Some methods from psychometrics like parallel factor analysis (PARAFAC) [6] and Tucker3 [15] have been used to increase understanding of chemical reactions or for process monitoring [11,27,28]. PARAFAC has been shown to be especially powerful in combination with fluorescence, where spectra can be decomposed into unique profiles [6,9]. Three-way PLS was developed by Ståhle [22] in 1989. Later, Bro [7] developed a general multiway PLS (N -PLS) for three- or higher-order arrays.

Little research has been carried out combining multiway techniques with NIR spectroscopy. There can be several reasons for this. As mentioned earlier, PARAFAC can be used to resolve spectra into pure profiles directly due to its uniqueness property [6]. However, obtaining pure profiles from NIR spectra is difficult due to broad overlapping bands and baseline variations. This is shown in two papers of Geladi et al. [11,12]. They used PARAFAC in combination with NIR spectra from a batch organic synthesis. The spectral loadings obtained were not pure and spectra of pure chemicals were needed for interpretation. Another problem in using N -way techniques is the requirement of

* Corresponding author. Tel.: +47-55-58-98-51; fax: +47-55-58-94-90.

E-mail address: Laila.Stordrange@kj.uib.no (L. Stordrange).

¹ Visiting scientist from Research group for Chemometrics, Umeå University, Sweden.

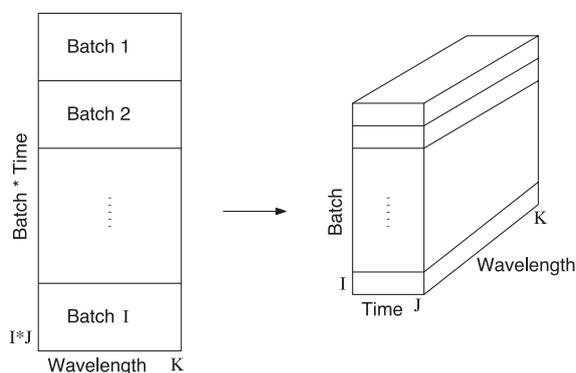


Fig. 1. Spectra acquired at specific time intervals of a batch process can be stacked as a three-dimensional array (batch \times time \times wavelength).

data to be available for the entire duration of the batch for the model to be applied in real-time. In practice, information on the complete batch trajectory is only available once the production has terminated. To control the process using the empirical formula, quick and reliable estimates of process variables are needed to fill in missing data.

Even if there are difficulties combining N -way techniques with NIR spectroscopy, beneficial results have been obtained. In the examples of Geladi et al. [11,12], loadings in time direction explained the reaction pattern well and increased understanding of the synthesis was obtained combining the loadings with spectral knowledge. PARAFAC has also been shown to be a useful tool for classification and to study time series of malting of barley using NIR spectroscopy [1]. No applications of NIR spectra in combination with N -way regression techniques are found in literature. However, three-way PLS have been tested on fluorescence and UV/visible spectroscopy with better results than two-way PLS [5,8]. Other results have indicated that the type of data and scaling is more important than the type of regression model [13]. Since the NIR technique has problems with non-linear behaviour, broad overlapping bands and scatter effects among others, N -way regression techniques should be tested on NIR data to see if regression models of better predictive ability could be obtained. Successful studies have been performed where missing values were filled in using different techniques in the multivariate statistical process control (MSPC) field of process variables, as flow rates, temperatures and density [16].

There are three objectives in this paper. The first is to test N -way decomposition methods for exploration of NIR data. PARAFAC and Tucker3 are tested for this purpose. Many organic synthesis are controlled empirically, i.e., based on knowledge of how the process evolves, reagents are added at certain time points to obtain a product within product specification and of high yield. To employ an empirical formula, knowledge of process parameters is needed. Our second aim is to test the performance of N -way calibration methods to model a process parameter used for surveillance of an organic synthesis. A great advantage of using NIR spectroscopy is the speed that makes the technique suitable

for on-line monitoring of processes. However, in order to use N -way calibration models on-line, all data from a batch need to be available. The third and last objective in this paper is to test different ways to fill in missing values. NIR data acquired from a synthesis of contrast agent is used as a case study.

2. Theory

2.1. Notation

Scalars are indicated by lower-case italics, e.g., a , and vectors by bold lower-case letters, e.g., \mathbf{x} . Bold-face capitals, e.g., \mathbf{X} , are used for matrices, and underlined bold-face capitals, e.g., $\underline{\mathbf{X}}$, for three-way arrays. The size of three-way data is given as $I \times J \times K$, i.e., I batches, J objects and K variables. Kiers and Van Mechelen [15] have presented a standardized notation and terminology to be used for three- and higher-order arrays. This paper follows these directives.

2.2. Multiway decomposition methods

PARAFAC [6] decomposes the three-way array, $\underline{\mathbf{X}}$, into a score matrix, \mathbf{A} ($I \times R$), and two loading matrices, \mathbf{B} ($J \times R$) and \mathbf{C} ($K \times R$), by alternating least squares. R is number of components. First, the three-way array, $\underline{\mathbf{X}}$ ($I \times J \times K$), is matricized to a two-way matrix, \mathbf{X} ($I \times JK$). The loadings in \mathbf{B} and \mathbf{C} are set as, e.g., random numbers;

$$\mathbf{X}^{I \times JK} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{E}^{I \times JK} \quad (1)$$

where \odot is the columnwise Kronecker product (also denoted the Khatri–Rao product) [15]. Putting $(\mathbf{C} \odot \mathbf{B})^T$ equal to \mathbf{D} gives;

$$\mathbf{X}^{I \times JK} = \mathbf{A}\mathbf{D} + \mathbf{E}^{I \times JK} \quad (2)$$

The columnwise loadings in \mathbf{A} are estimated by alternating least squares;

$$\mathbf{A} = \mathbf{X}^{I \times JK} \mathbf{D}^T (\mathbf{D}\mathbf{D}^T)^{-1} \quad (3)$$

The loadings \mathbf{B} and \mathbf{C} can now be estimated likewise, starting by unfolding the three-way array to a matrix of size $(J \times IK)$ and $(K \times IJ)$, respectively. The loadings are calculated until convergence, that is little change in fit or loadings.

In Tucker3 [15,3], a core array $\underline{\mathbf{G}}$, of size $(P \times Q \times R)$, is employed;

$$\mathbf{X}^{I \times JK} = \mathbf{A}\mathbf{G}^{P \times QR}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}^{I \times JK} \quad (4)$$

The \otimes is the Kronecker product [15]. The function of this core array is to give a summary description of the three-way information, in terms of summarizing components for the three different loadings. Several Tucker3 algorithms exist [3]. The SVD-based algorithm is used here, i.e., the columns

in \mathbf{A} , \mathbf{B} and \mathbf{C} are constrained to be orthogonal. \mathbf{A} is initialized by;

$$[\mathbf{A}, \mathbf{S}, \mathbf{V}] = \text{SVD}(\mathbf{X}^{I \times JK}(\mathbf{C} \otimes \mathbf{B}), \mathbf{P}) \quad (5)$$

\mathbf{P} is the number of components to be extracted in mode \mathbf{A} , \mathbf{B} and \mathbf{C} are determined in the same way. Iteration is performed until relative change in fit is small. The core array $\underline{\mathbf{G}}$ can be determined by a simple regression of $\underline{\mathbf{X}}$ onto \mathbf{A} , \mathbf{B} and \mathbf{C} [3];

$$\mathbf{G}^{P \times QR} = \mathbf{A}^T \mathbf{X}^{I \times JK}(\mathbf{B} \otimes \mathbf{C}) \quad (6)$$

The core gives a summary of all interactions present in the three-way data.

A distinction from PARAFAC is that Tucker3 allows for extraction of different number of factors in each mode. In addition the Tucker3 model has rotational freedom, and is hence not structurally unique as the PARAFAC model. PARAFAC can be considered as a restricted Tucker3 model where the core array is a three-way array with zeros in all positions except for the superdiagonal, which contains only ones.

2.3. Multiway calibration

The score matrix, \mathbf{A} , from PARAFAC or Tucker3 decomposition (Eqs. (1) and (4), respectively) can be related to a response, \mathbf{y} :

$$\mathbf{y} = \mathbf{A}\mathbf{b} + \mathbf{f} \quad (7)$$

The regression coefficient, \mathbf{b} , is calculated by ordinary least squares. Predictions of new samples can be made, inserting scores estimated using the loading vectors in mode two and three (and core array for Tucker3) from calibration set, into Eq. (7).

In N -PLS [7,20], $\underline{\mathbf{X}}$ ($I \times J \times K$), is matricized to a two-way matrix, \mathbf{X} ($I \times JK$). The weights of the second and the third mode are calculated by;

$$\mathbf{z} = \mathbf{X}^T \mathbf{y} \quad (8)$$

\mathbf{X} and \mathbf{y} are column-centered. $\mathbf{z}^{1 \times JK}$ is matricized and \mathbf{w}^J and \mathbf{w}^K are equal to the first left and right singular vectors of \mathbf{Z} ($J \times K$), respectively. Putting $\mathbf{w}_1 = (\mathbf{w}^K \otimes \mathbf{w}^J)$, the N -PLS algorithm can be written as;

$$\begin{aligned} \mathbf{t}_A &= \mathbf{X}_{A-1} \mathbf{w}_A \\ \mathbf{b}_A &= (\mathbf{T}_A^T \mathbf{T}_A)^{-1} \mathbf{T}_A \mathbf{y} \\ \mathbf{X}_A &= \mathbf{X}_{A-1} - \mathbf{t}_A \mathbf{w}_A \end{aligned} \quad (9)$$

$$\mathbf{y}_A = \mathbf{y} - \mathbf{T}_A \mathbf{b}_A$$

These steps are repeated for $A=1, 2, \dots$, up to significant number of factors. Note that the calculation of the \mathbf{b} vector and the deflation of \mathbf{y} has to be carried out using the nondeflated \mathbf{y} and all previous \mathbf{t} vectors [20], $[\mathbf{T}_A =$

$\mathbf{t}_1, \dots, \mathbf{t}_A]$. If the response is a matrix the loadings of \mathbf{Y} are equal to $\mathbf{q}_A = \mathbf{t}_A^T \mathbf{Y} / \mathbf{t}_A^T \mathbf{t}_A$ and prediction of new samples are estimated as; $\hat{\mathbf{Y}} = \mathbf{T}\mathbf{B}\mathbf{Q}^T$.

2.4. Missing data

Meng et al. [16] used PARAFAC for MSPC of a batch process. In the paper, they presented a methodology where the unknown observations are calculated as a weighted combination of the scores up to the current time point in the new batch and those previously computed from a calibration data set. The method was shown to give smoother scores avoiding spurious alarms in control charts. In this paper, the scores are not used for MSPC. The weighted scores are instead used to calculate estimates of missing spectra. The mean of scores from calibration set, \mathbf{a}_{mean} , and scores for a new batch up to the current time point j , \mathbf{a}_j , are calculated as;

$$\mathbf{a}_{mean} = \text{mean}(\mathbf{A}) \quad (10)$$

$$\mathbf{a}_j = \mathbf{x}_{new,1:j} \mathbf{D}_{1:j}^T (\mathbf{D}_{1:j} \mathbf{D}_{1:j}^T)^{-1} \quad (11)$$

$\mathbf{x}_{new,1:j}$ is the vectorized $\mathbf{X}^{j \times K}$, and $\mathbf{D}_{1:j}$ is the reduced Khatri–Rao product of $\mathbf{B}_{1:j} \odot \mathbf{C}^T$. Weighted scores, $\mathbf{a}_{weight,j}$, are calculated from the scores \mathbf{a}_{mean} and \mathbf{a}_j using the weight λ_j ;

$$\mathbf{a}_{weight,j} = \lambda_j \mathbf{a}_{mean} + (1 - \lambda_j) \mathbf{a}_j, \quad \lambda_j = \frac{J-j}{J-1} \quad (12)$$

To estimate the process measurements from $j+1:J$ the product of the weighted scores and the loading matrix corresponding to the time period $j+1:J$ is used;

$$\mathbf{x}_{new,j+1:J} = \mathbf{a}_{weight,j} \mathbf{D}_{j+1:J} \quad (13)$$

A combination of the already known observations up to j and the estimated values from $j+1:J$ can be combined;

$$\mathbf{x}_{new} = [\mathbf{x}_{new,1:j} \mid \mathbf{x}_{new,j+1:J}] \quad (14)$$

$\mathbf{x}_{new}^{1 \times JK}$ is matricized to a matrix of size ($J \times K$) with no missing values.

For Tucker3, the weighted infilling can be performed putting \mathbf{D}_j equal to $\mathbf{G}(\mathbf{C} \otimes \mathbf{B}_{1:j})^T$ (see Eq. (4)).

The reduced scores for an N -PLS model can be estimated by;

$$\begin{aligned} t_{1,j} &= \mathbf{x}_{new,1:j} \mathbf{w}_{1,j} \\ \mathbf{x}_{defl1,1:j} &= \mathbf{x}_{new,1:j} - t_{1,j} \mathbf{w}_{1,j} \end{aligned} \quad (15)$$

$$t_{2,j} = \mathbf{x}_{defl1,1:j} \mathbf{w}_{2,j}$$

and so on, $\mathbf{w}_{1,j} = (\mathbf{w}_1^K \otimes \mathbf{w}_{1,1:j}^J)$ and $\mathbf{w}_{2,j} = (\mathbf{w}_2^K \otimes \mathbf{w}_{2,1:j}^J)$. The scores can be used to find an estimate of the missing process data;

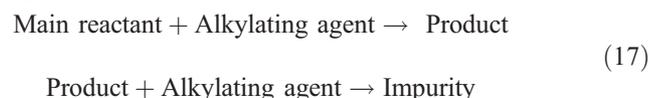
$$\mathbf{x}_{new,j+1:J} = \mathbf{t}_{weight,j} \mathbf{W}_{j+1:J}^T \quad (16)$$

For a two-factor model, $\mathbf{t}_{weight,j}$ is the weighted scores of $\mathbf{t}_j=[t_{1,j} \ t_{2,j}]$ and \mathbf{t}_{mean} from *N*-PLS, $\mathbf{W}_{j+1:j}$ is equal to $[\mathbf{w}_{j1} \ \mathbf{w}_{j2}]$.

Other methods to infill missing values have been proposed [17,18]. Two of the methods, i.e., (1) employ mean process data and (2) scores from the reduced data set, are tested in this paper. The first method puts missing data equal to mean process data in calibration set. The method assumes that future observations are in perfect accordance with mean values obtained for data set used to make the model. However, small changes from the recipe is expected from batch to batch. This behaviour can be expected to be observed in the non-weighted scores, \mathbf{a}_j and \mathbf{t}_j , and employing these scores to find missing data might give better estimates and thereby better predictive ability of the model.

3. Case study: NIR spectra from a batch production

To investigate the ability of multiway techniques in order to explore and model NIR spectra, a case study from a batch production is chosen. The process is a synthesis of contrast agent:



The main reactant reacts with the alkylating agent and the desired product is formed. When the product concentration gets high compared to the concentration of the main reactant, the product tends to react with the alkylating agent and impurity forms. The synthesis takes about 26 h and is controlled empirically by measuring the concentration of main reactant and impurity by HPLC at three fixed time points. If the synthesis is evolving too fast or too slow, reagents are added to slow down or speed up the reaction, respectively. The data set has been used in two other studies [23,24] and additional information about the system can be found in these papers. The requirement from the factory is to find a proper way to measure the concentration of main reactant and impurity by NIR spectroscopy. A good model for impurity is easily obtained, the problem is main reactant and especially measuring this compound at control point III.

3.1. Experimental

Near-infrared spectra were collected offline using a FOSS NIRSystems 6500 spectrophotometer. Each spectrum was recorded using a probe attached by fibre optics to the spectrophotometer. The wavelength region from 1100 to 1900 nm, with a resolution of 2 nm, was used in this study. The data set comprises 17 batches where a spectrum has been collected at each control point. This gives a three-way array of size 17 batches \times 3 time points \times 401 wavelengths. Additional 16 batches were used to test the performance of

the regression model. Prior to data analysis the data were preprocessed by second-order differentiation (Savitzky–Golay \pm 5 points). Data analysis was carried out using MATLAB (The MathWorks, version 6.5) and the *N*-way Toolbox [2] for MATLAB version 2.10.

4. Results and discussion

4.1. Inspection of data

4.1.1. Wavelength selection

Inspecting the wavelength region from 1100 to 1900 nm by interval PLS [19] revealed that the region from 1622 to 1660 nm to model the response best, while principal variable [19] indicated 1644 nm to be the variable of highest covariance with the response. Table 1 gives the predictive ability of *N*-PLS and two-way PLS for interesting wavelength regions. For *N*-PLS the region from 1638 to 1650 nm gives the lowest predictive ability. It is also observed that *N*-PLS has severe problems when using the wavelength region from 1100 to 1900 nm. This is different from two-way PLS which is not affected by if a smaller or larger region is employed as long as the informative region is incorporated.

Decomposing the region from 1100 to 1900 nm by *N*-PLS shows that small weights are given to the informative region around 1644 nm (see Fig. 2). This is different from two-way PLS, where large regression coefficients are obtained for the same region. Why *N*-PLS, in contrast PLS1, performs poorly when a large region is used becomes comprehensible when inspecting the centered data. In performing *N*-PLS, each control point and the accompanying response is centered separately, and the concentration differences between control points are lost. When calculating the covariance between the spectra and the response, *N*-PLS is more vulnerable to wavelengths region changing in intensity when the process evolves, and the response is erroneously registered as being dependent on these variables, making future predictions terrible. This is not obtained for PLS1, where all the control points are centered simultaneously, giving three distinct bands. When modelling data the wavelength region of three distinct bands, i.e., the region around 1644 nm, is weighted the most. Since *N*-PLS blocks data the distinct concentration difference between control points is absent when modelling, and the result is that variable selection is more important for *N*-PLS than for two-way PLS.

Table 1
Predictive ability of main reactant for different wavelength regions using *N*-PLS and PLS1, full cross-validation

Wavelength (nm)	<i>N</i> -PLS		PLS1	
	No. LVs	RMSEP	No. LVs	RMSEP
1100–1900	9	1.63	6	0.72
1622–1660	3	0.71	3	0.76
1638–1650	3	0.59	1	0.78

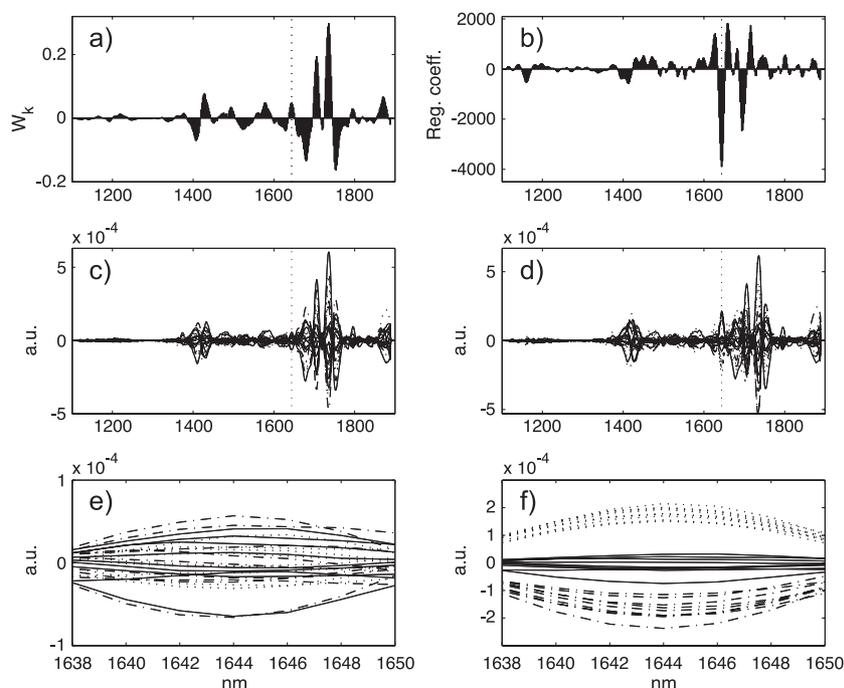


Fig. 2. Performance of *N*-PLS and PLS1 for the region from 1100 to 1900 nm: (a) first weight *N*-PLS, (b) first regression coefficient PLS1, (c) centred spectra of *N*-PLS and (d) PLS1, (e) centred spectra for the region from 1638 to 1650 nm of *N*-PLS, and (f) PLS1. Spectra acquired in control point I is denoted (---), control point II (—), and control point III (· · ·).

The wavelength region from 1638 to 1650 nm gives the best predictive ability for *N*-PLS and close-to-best predictive ability for PLS1. This region was chosen for further analysis.

4.1.2. Signal variation

To determine if there is enough signal variation in each control point to model the main reactant, the correlation coefficients between the scores and the response were calculated (Table 2). The first score gave the highest correlation for each control point. There seems to be a reasonable correlation between the predictor and the response in each block, so the data should be appropriate for multiway analysis.

For impurity the signal variation is small in the different control points and *N*-way techniques are expected to fail for modelling this compound.

4.2. PARAFAC and Tucker3 for exploration of data

The aim of using PARAFAC is to test whether it is possible to resolve pure spectral profiles. This will enable us

Table 2
The correlation coefficients between first score and response for the three control points

Control point	Correlation coefficient
I	0.93
II	0.83
III	0.60

to obtain more information about the chemical system and to resolve concentration profiles directly. Calculation of residual sum of squares and core consistency [29] for different factors revealed two factors to be significant. A split-half test [6] was carried out. That is, the batches were divided into two groups that span the variation and if the correct number of factors is used the split half test should give identical profiles in modes 2 and 3. This was not obtained here.

While PARAFAC requires the same number of factors to be extracted in all modes, Tucker3 does not have this limitation. A dimensionality test was performed for Tucker3 (see Fig. 3). The plot indicates [2 2 1] factors to be optimal, i.e., high explained variance and low dimensionality. Note that reducing the number of loadings in mode 3 from two to one does not give any reduction in explained variance, i.e., no information is lost.

Earlier investigations of the system indicated that the signal used to model the main reactant and the impurity probably stems from the alkylating agent [24]. This explains one significant factor in the spectral direction. Loadings in the concentration mode of Tucker3 model seem to resemble the absolute concentration profiles of main reactant and impurity. To check if the two significant factors in time direction correspond to the main reactant and impurity, a dimensionality test of simulated data was performed. The simulated data set was a symbiosis of the real NIR spectra (observed bands, offset, multiplicative and additive effects, and homoscedastic and heteroscedastic noise [25,26]), and the assumption that the band at 1644 nm originates from the

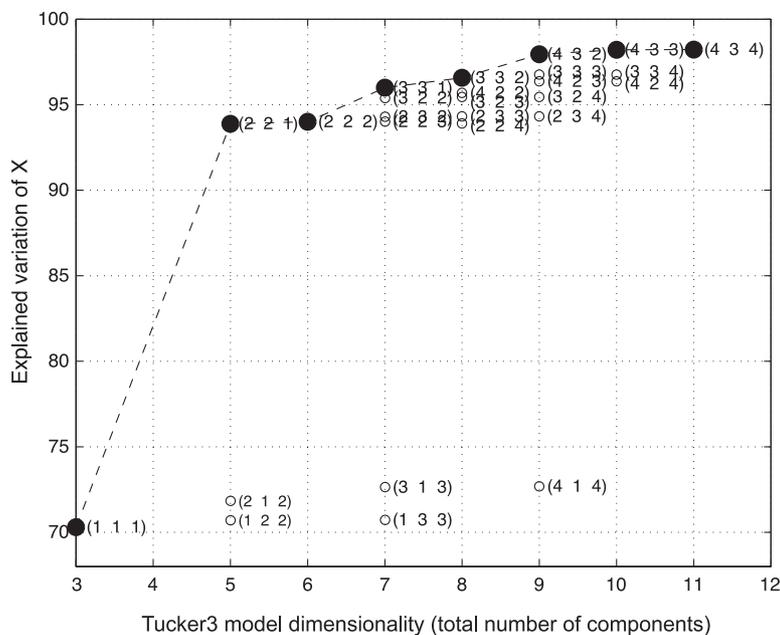


Fig. 3. Sum of square error as a function of Tucker3 model dimensionality. From [1 1 1] up to [4 3 4] factors have been tested.

alkylating agent. A real and a simulated spectrum are shown in Fig. 4.

Three simulated data sets were made using different concentration profiles, i.e., the sum of the responses for main reactant and impurity, the response for the main reactant and the response for the impurity. A dimensionality test using PARAFAC and Tucker3 was carried out on the simulated data sets (see Table 3).

The results strongly indicate that the contribution from the impurity is too small to add any observable differences and that only the alkylating agent reacting with main reactant gives measurable results.

So why do we observe a [2 2 1] dimensionality in the process data? The Tucker3 decomposition can be viewed as a three-way PCA analysis, i.e., the method detects major variation in data. The reason for two significant

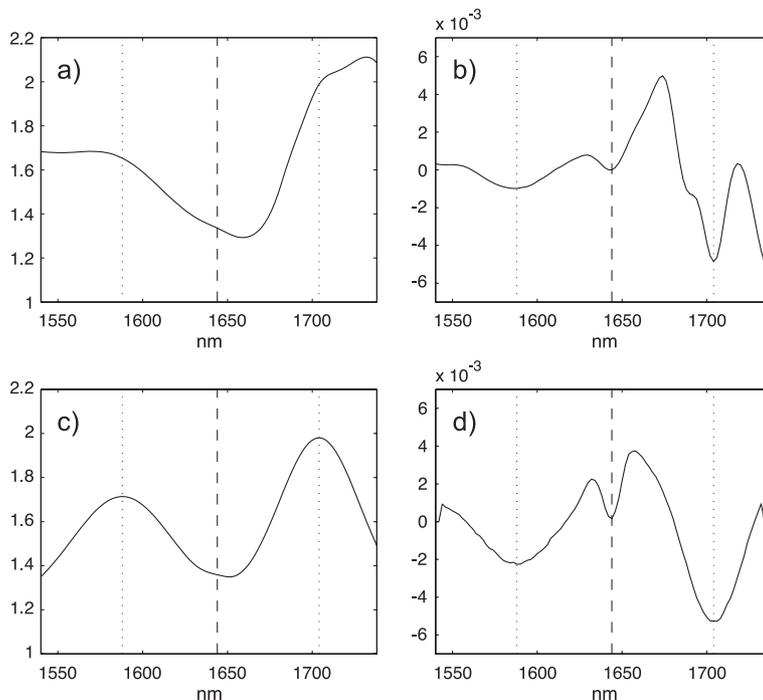


Fig. 4. A NIR spectrum and a simulated spectrum. (a) NIR spectrum, (b) second-order differentiated NIR spectrum, (c) simulated spectrum, and (d) second-order differentiated simulated spectrum. Peaks from solvent (· · ·) and of the alkylating agent (— —) utilised to make a simulated data set.

Table 3
Dimensionality test of simulated data set

Response multiplied with “alkylating band”	PARAFAC		Tucker3
	SSQ	Core consistency	
Main reactant + impurity	2 or 4	2 or 4	[2 2 1]
Main reactant	2	2	[2 2 1]
Impurity	0	2 or 3	[0 0 0]

components in the batch mode and in the concentration mode is probably because two components are necessary to describe the variation in the alkylating band. From these result the cause seems to be twofold: (i) composition of solvent and reagent and their quality differs somewhat from batch to batch, and (ii) addition of reagent during the runtime of the synthesis.

Note that the system is complex and that the rank is probably higher than two. The fit plot in Fig. 3 seems to suggest that there could be additional components of minor variance. The bands in the near-infrared region are broad and overlapped, and therefore the chance of masking minor components is high.

4.3. Multiway calibration

The scores from PARAFAC and Tucker3 were used to model the main reactant. The batches were divided into a calibration set and a validation set. The predictive ability is given in Table 4 together with the result from *N*-PLS and PLS1. PARAFAC, Tucker3 and *N*-PLS perform equally. The *N*-way methods perform better than two-way PLS. A difference is observed in control point III (low concentration) where multiway methods are able to model the main reactant, while two-way PLS fails to do so.

Why do *N*-way regression methods model the main reactant better than two-way regression techniques? Inspection of the residuals (Fig. 5) reveals that they locate close to the perfect fit line for *N*-PLS, while for two-way PLS the residuals are large and have a systematic pattern. This is especially observed for control points I and III, where all except one object plot beneath the perfect fit-line. In addition, it is revealed that the slope of fit line at each control point differs, and mostly for control point III. Predictions of new samples for two-way regression are given by;

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} \quad (18)$$

i.e., one regression coefficient, \mathbf{b} , is estimated and used for spectra acquired at all control points. If the absorbance is not linearly related to the response, \mathbf{y} , poor predictive ability can be expected.

N-way techniques decompose data by projecting them onto the loadings in modes 2 and 3. The result is scores that explain the behaviour of the batch. The scores obtained for one batch are used to model the response

at all control points for the given batch. For *N*-PLS, even the same regression coefficients are used at all control points;

$$\hat{\mathbf{Y}} = \mathbf{T}\mathbf{B}\mathbf{Q}^T \quad (19)$$

The only difference is that a weight at each control point is given by the loadings for the response, \mathbf{Q} .

The main difference between the two-way and the *N*-way methods is that two-way methods treat data points from all control points simultaneously, while *N*-way methods operate on each control point separately. While the concentration needs to be linearly correlated at all control points for two-way PLS, the concentration for *N*-way techniques needs only to be linear correlated within each control point. The result is that *N*-way regression may be a good alternative handling data with nonlinear behaviour.

Since the slope differ at all the control points the predictive ability of making one model per control point was tested (see Table 4). Better predictive ability was obtained at control point I. However, the predictive ability at control points II and III is the same as modelling all control points simultaneously with two-way PLS. Employing *N*-way regression does not only give beneficial results due to handling nonlinearities, but since the technique finds common scores for the behaviour of all control points in a batch it makes the model more robust against spectral fluctuations.

4.4. Missing values in multiway calibration

An improvement in predictive ability is obtained performing multiway regression compared to two-way PLS. However, a problem using multiway calibration for on-line monitoring is how to make predictions in early stages of a batch production when the NIR spectra are missing in the control points not yet monitored.

Filling in random values for missing values gave poor predictions (see Table 5). Predictions of a spectrum acquired in one control point are not independent of the spectra acquired at the other control points in a batch. The easiest way to overcome the dependency problem is to insert the mean spectrum from the calibration samples for the missing spectra. The predictive ability obtained is better than for two-way PLS, but it is far from the low prediction errors obtained using the true spectra (see Table 4).

The composition of a large batch is expected to be unique due to small differences from the recipe from batch to batch,

Table 4
Results off-line monitoring

Control point	PARAFAC	Tucker3	<i>N</i> -PLS	PLS1	PLS1 per control point
I	0.32	0.33	0.33	0.90	0.74
II	0.47	0.47	0.47	0.77	0.78
III	0.23	0.24	0.23	1.15	1.19

RMSEP of validation set using two- and three-way regression methods.

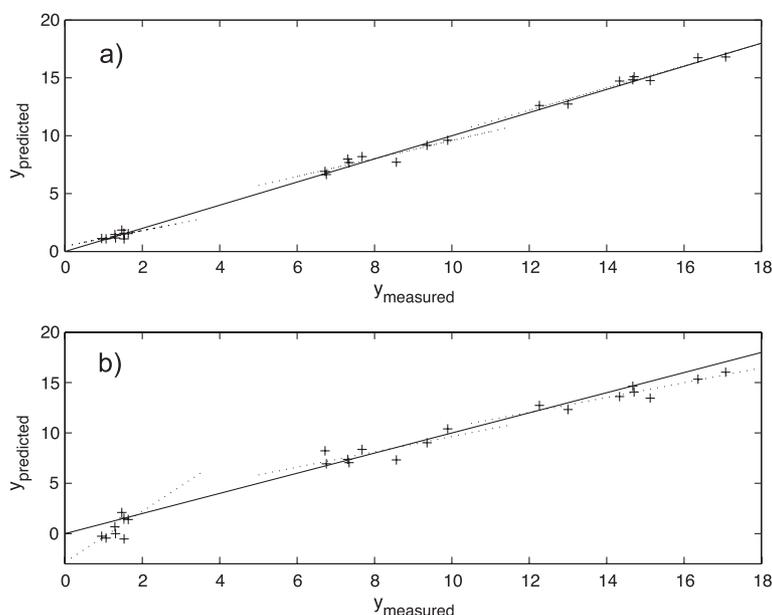


Fig. 5. Predicted vs. measured for validation samples calculated using (a) *N*-PLS and (b) two-way PLS. Stipled line is first-order polynomial fit of residuals at each control point.

slight changes in quality of added compounds and environmental changes in the surrounding of the vessel. Altogether, a batch score calculated using the spectra acquired up to present time of process monitoring could be expected to reflect the uniqueness of the given batch monitored. Improvements in the predictive ability of models were obtained using scores (see Table 5). However, the improvements are small and the reason may be that reagents are

added during reaction time making the score less unique for the missing spectra of the batch.

The weighted infilling performed equally to the use of mean spectra, and even if the scores give a better result, use of weighting with mean score suppresses the contribution from the scores.

Having NIR spectrophotometer equipped on-line, scanning spectra more frequently is expected to give more reliable estimates of the spectra. The unweighted scores from PARAFAC seem to give the most reliable estimates and should be used for this purpose.

Table 5
Results on-line monitoring

Control point	Validation set			Test set		
	I	II	III ^a	I	II	III ^a
<i>PARAFAC</i>						
Random	8.64	5.23	0.23	8.63	5.27	0.26
Mean	0.77	0.59	0.23	0.93	0.73	0.26
Scores	0.74	0.58	0.23	0.88	0.72	0.26
Weighted	0.76	0.59	0.23	0.92	0.72	0.26
<i>Tucker</i>						
Random	9.05	4.60	0.24	10.18	5.49	0.26
Mean	0.77	0.59	0.24	0.92	0.73	0.26
Scores	0.74	0.61	0.24	0.93	0.87	0.26
Weighted	0.77	0.58	0.24	0.92	0.76	0.26
<i>N-PLS</i>						
Random	9.56	5.87	0.23	9.86	5.97	0.26
Mean	0.78	0.60	0.23	0.93	0.73	0.26
Scores	0.73	0.59	0.23	0.92	0.73	0.26
Weighted	0.78	0.59	0.23	0.93	0.73	0.26

Data for the entire duration of the batch process is missing in control points I and II. The predictive ability (RMSEP) is given for different ways to in-fill yet unknown data.

^a No data missing.

5. Conclusion

Variable selection has been shown to be Alpha and Omega to succeed with *N*-way modelling when many irrelevant variables are measured. Applying the wavelength region from 1100 to 1900 nm gave no comprehensible result. In comparison, when modelling using a narrow wavelength region of high covariance with the dependent variable, calibration models of better predictive ability than for two-way PLS were obtained. Inspecting the residuals revealed the residuals from two-way PLS to be nonlinear, while for *N*-PLS they were random and small. While two-way PLS treats data from all control points at the same time, *N*-way methods stack each control point in a block. The *N*-way technique has here been shown to be beneficial if the data set has a nonlinear behaviour. PARAFAC, Tucker3 and *N*-PLS performed equally for regression purposes.

For infilling of missing spectra, mean spectra from calibration set, scores from spectra acquired up to present time and a weighted infilling technique combining these

two were tested. Applying mean spectra from calibration set gave close to the same predictive ability as the scores. However, minor improvements in the predictive ability were obtained using scores and if spectra are acquired more frequent the results indicate that employing PARAFAC scores, which do not require a dependent value, to fill in for missing values will be a better option.

Decomposing the interesting region by *N*-way methods increased the knowledge of the system. Simulated data sets were decomposed in the same manner. Comparing the results revealed one compound to contribute to the wavelength band used for modelling and that two factors are needed to describe the variation in batch and time direction. The variation is expected to be caused by differences in composition of compounds from batch to batch and changes in process matrix over time due to addition of reagent as the process evolves.

Acknowledgements

This project has been given financial support through a PhD grant to Laila Stordrange from the Norwegian Research Council, project No. 138494/432. Tarja Rajalahti is grateful to NORFA for a visiting scientist grant at University of Bergen.

References

- [1] N. Allosio, P. Boivin, D. Bertrand, P. Courcoux, Characterisation of barley transformation into malt by three-way factor analysis of near infrared spectra, *J. Near Infrared Spectr.* 5 (1997) 157–166.
- [2] C.A. Andersson, R. Bro, The N-way toolbox for MATLAB, *Chemom. Intell. Lab. Syst.* 52 (2000) 1–4.
- [3] C.A. Andersson, R. Bro, Improving the speed of multi-way algorithms, *Chemom. Intell. Lab. Syst.* 42 (1998) 93–103.
- [4] H. Antti, M. Sjöström, L. Wällback, Multivariate calibration models using NIR spectroscopy on pulp and paper industrial applications, *J. Chemom.* 10 (1996) 591–603.
- [5] J.L. Béltran, R. Ferrer, J. Guiteras, Multivariate calibration of polycyclic aromatic hydrocarbon mixtures from excitation–emission fluorescence spectra, *Anal. Chim. Acta* 373 (1998) 311–319.
- [6] R. Bro, PARAFAC. Tutorials and applications, *Chemom. Intell. Lab. Syst.* 38 (1997) 149–170.
- [7] R. Bro, Multiway calibration. Multilinear PLS, *J. Chemom.* 10 (1996) 47–61.
- [8] R. Bro, H. Heimdal, Enzymatic browning of vegetables. Calibration and analysis of variance by multiway methods, *Chemom. Intell. Lab. Syst.* 34 (1996) 85–102.
- [9] R. Bro, Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis, *Chemom. Int. Lab. Syst.* 46 (1999) 133–147.
- [10] C. Coffey, B.E. Cooley, D.S. Walker, Real time quantitation of a chemical reaction by fiber optic near-infrared spectroscopy, *Anal. Chim. Acta* 395 (1999) 335–341.
- [11] P. Geladi, J. Forsström, Monitoring of a batch organic synthesis by near-infrared spectroscopy: modeling and interpretation of three-way data, *J. Chemom.* 16 (2002) 329–338.
- [12] P. Geladi, P. Åberg, Three-way modelling of a batch organic synthesis process monitored by near infrared spectroscopy, *J. Near Infrared Spectr.* 9 (2001) 1–9.
- [13] S.P. Gurden, J.A. Westerhuis, R. Bro, A.K. Smilde, A comparison of multiway regression and scaling methods, *Chemom. Intell. Lab. Syst.* 59 (2001) 121–136.
- [14] U.G. Indahl, N.S. Sahni, B. Kirkhus, T. Næs, Multivariate strategies for classification based on NIR-spectra—with application to mayonnaise, *Chemom. Intell. Lab. Syst.* 49 (1999) 19–31.
- [15] H.A.L. Kiers, I. Van Mechelen, Three-way component analysis: principles and illustrative application, *Psychological Methods* 6 (2001) 84–110.
- [16] X. Meng, A.J. Morris, E.B. Martin, On-line monitoring of a batch process using a PARAFAC representation, *J. Chemom.* 17 (2003) 65–81.
- [17] P. Nomikos, J.F. MacGregor, Multivariate SPC charts for batch processes, *Technometrics* 37 (1995) 41–59.
- [18] P. Nomikos, J.F. MacGregor, Multi-way partial least squares in monitoring batch processes, *Chemom. Int. Lab. Syst.* 30 (1995) 97–108.
- [19] L. Nørgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, S.B. Engelsen, Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy, *Appl. Spectr.* 54 (2000) 413–419.
- [20] A.K. Smilde, Comments on multilinear PLS, *J. Chemom.* 11 (1997) 367–377.
- [21] N. Smola, U. Urleb, Qualitative and quantitative analysis of oxytetracycline by near-infrared spectroscopy, *Anal. Chim. Acta* 410 (2000) 203–210.
- [22] L. Stähle, Aspects of the analysis of three-way data, *Chemom. Intell. Lab. Syst.* 7 (1989) 95–100.
- [23] L. Stordrange, F.O. Libnau, D. Malthe-Sørenssen, O.M. Kvalheim, Feasibility study of NIR for surveillance of a pharmaceutical process, including a study of different preprocessing techniques, *J. Chemom.* 16 (2002) 529–541.
- [24] L. Stordrange, O.M. Kvalheim, P.A. Hassel, D. Malthe-Sørenssen, F.O. Libnau, A comparison of techniques for modelling data with non-linear structure, *J. Near Infrared Spectr.* 11 (2003) 55–70.
- [25] J. Toft, O.M. Kvalheim, Eigenstructure tracking analysis for revealing noise pattern and local rank in instrumental profiles: application to transmittance and absorbance IR spectroscopy, *Chemom. Intell. Lab. Syst.* 19 (1993) 65–73.
- [26] J. Toft, O.M. Kvalheim, Erratum: Eigenstructure tracking analysis for revealing noise pattern and local rank in instrumental profiles: application to transmittance and absorbance IR spectroscopy, *Chemom. Intell. Lab. Syst.* 22 (1994) 295.
- [27] E.N.M. van Sprang, H.-J. Ramaker, J.A. Westerhuis, S.P. Gurden, A.K. Smilde, Critical evaluation of approaches for on-line batch process monitoring, *Chem. Eng. Sci.* 57 (2002) 3979–3991.
- [28] J.A. Westerhuis, T. Kourti, J.F. MacGregor, Comparing alternative approaches for multivariate statistical analysis of batch process data, *J. Chemom.* 13 (1999) 397–413.
- [29] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, *J. Chemom.* 17 (2003) 274–286.