PERGAMON

Chemical Engineering Science 55 (2000) 1225–1235

Chemical Engineering Science

Multivariate statistical process control of batch processes based on three-way models

D. J. Louwerse, A. K. Smilde*

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

Received 19 November 1998; accepted 7 May 1999

Abstract

The theory of batch MSPC control charts is extended and improved control charts are developed. Unfold-PCA, PARAFAC and Tucker3 models are discussed and used as a basis for these charts. The results of the different models are compared and the performance of the control charts based on these models is investigated. It is found that this performance depends on the type of fault occuring in the batch process. A strategy is provided to partition reference data describing the normal operating conditions, in order to be able to monitor a new incomplete batch on-line. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Batch MSPC; Tucker; PARAFAC; PCA; Control charts; Process monitoring

1. Introduction

Multivariate statistical process control (MSPC) for monitoring batch and semi-batch processes is still a relatively new technique. Conventional (univariate) statistical process control (SPC) techniques are already used in the process industry. It is expected that batch MSPC will also become more important since batch production processes play an important role in chemical industry. Pharmaceuticals, biochemicals and a large number of polymers, for instance, are often produced batch wise. Batch MSPC can play an important role when it proves to be cost effective, or when a safe operation or a better insight in the process variability can be achieved. Due to the non steady-state behaviour of batch processes, however, batch MSPC is more complicated than MSPC of continuous processes. Nevertheless, the number of applications is rising. The first comprehensive paper in this field appeared by Nomikos and MacGregor (1994). Since then batch MSPC is getting more attention, see Nomikos and MacGregor (1995), Kourti and MacGregor (1995), Martin, Morris, Papazoglou and Kiparissides (1996), Dong and McAvoy (1996).

In MSPC of continuous processes the normal variation of variables around their steady state is modelled, where in batch MSPC the normal variation of variables around an optimal time trajectory has to be modelled. An additional time dimension is present in batch process data which makes it three dimensional in nature.

Three-way batch data can be analysed with various methods. Up to now unfold principal component analysis (unfold-PCA), also referred to as multiway-PCA, is often used for batch MSPC (Nomikos & MacGregor, 1995). For unfold-PCA the three-way array is unfolded to a two-way array. Then the data is modelled with PCA and finally the result is folded back again to obtain a three-way representation of the model and the modelled data. Using unfold-PCA is straightforward because PCA is a well-known and well understood tool for analysing data. PCA can also be used for MSPC of continuous processes.

Due to the unfolding process, shown in the next section, the number of parameters of an unfold-PCA model is very large. The number of parameters of a PARAFAC and of a Tucker3 model is, compared to unfold-PCA, considerably reduced since the data is compressed in three directions instead of one as in unfold-PCA. Hence it might be expected that PARAFAC and Tucker3 models are more stable. For PARAFAC and Tucker3 models, batch MSPC charts are constructed and compared to the presently used unfold-PCA control charts. For this

* Corresponding author. Tel: +31-20-5255062; fax: +31-20-5256638.

E-mail address: asmilde@anal.chem.uva.nl (A. K. Smilde)

PII: S0009-2509(99)00408-X

comparison it is important to look at the model fit and the residuals, and especially to look at the detecting and diagnostic power of the control charts in case of erroneous batches. MSPC serves three goals: detecting, locating and diagnosing erroneous variation. Hence, control charts should be judged in this perspective.

The purpose of this paper is threefold. First it is shown how unfold-PCA, PARAFAC and Tucker3 models can be used to model batch process data and also how a new batch is projected on each model. Secondly, the theory of batch MSPC charts is described and improved. Results are compared to existing batch MSPC charts. Thirdly, a method for dealing with unfinished batches in on-line monitoring is introduced. Monitoring results of MSPC charts based on unfold-PCA, PARAFAC and Tucker3 models are presented and compared.

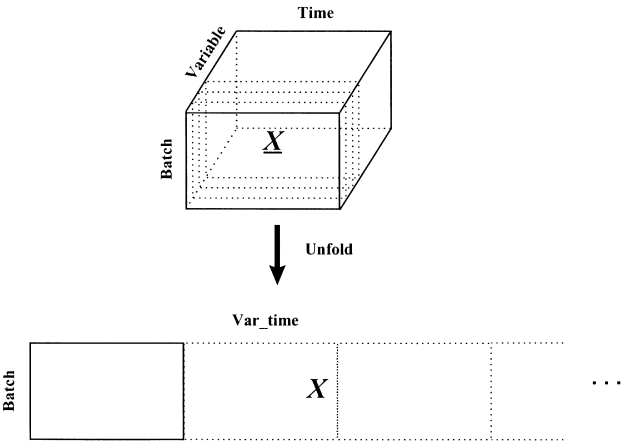


Fig. 1. Three-way data $\underline{\boldsymbol{X}}$, with modes "batch", "variable" and "time", unfolded to a two-way matrix $\boldsymbol{X}$ with mode "batch" and a combined mode "var_time".

2. Three-way data models

Normal characters, including capitals, are scalars and refer to single data elements, or can have a special meaning as will be explained in the text. Bold lower case characters refer to vectors, bold capitals refer to two-way arrays or matrices and bold underlined capitals refer to three-way arrays. The superscript$^{\mathrm{T}}$ attached to a matrix, like in $\boldsymbol{X}^{\mathrm{T}}$, refers to the transpose. Batch data, $\underline{\boldsymbol{X}}(I \times J \times K)$ with I batches, J variables and K time points, will be used for illustration.

For unfold-PCA, $\underline{\boldsymbol{X}}$ is first unfolded in a two-way array $\boldsymbol{X}$. There are three possible ways to unfold $\underline{\boldsymbol{X}}$, leaving one mode intact and unfolding the other two modes into one combined mode. In case of batch MSPC it is important to determine differences between batches and to project new batches on the model. This is possible by maintaining the batch (first) mode and unfolding the variable (second) and the time (third) modes in a newly formed second mode, here called the var_time mode. Fig. 1 shows this process by taking frontal slices and placing them next to each other. The number of variables in the newly formed var_time mode is $M = J \times K$. Columns of $\boldsymbol{X}(I \times M)$ represent the batches for a single variable at a single point in time and rows represent all possible combinations of variable and time for a single batch. Using PCA $\boldsymbol{X}$ can be modelled with R principal components

$$x_{im} = \sum_{r=1}^{R} a_{ir} b_{mr} + e_{im}, \tag{1}$$

or in matrix notation

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{E}, \tag{2}$$

where x_{im}, a_{ir}, b_{mr} and e_{im}, respectively, are elements of $\boldsymbol{X}(I \times M)$, $\boldsymbol{A}(I \times R)$, $\boldsymbol{B}(M \times R)$ and $\boldsymbol{E}(I \times M)$. Without changing the model, $\boldsymbol{B}$ can be made column-orthogonal. Then $\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{I}$ and the explained variance of the data is accounted for in $\boldsymbol{A}$. Vectors of $\boldsymbol{A}$ are called score vectors. $\hat{\boldsymbol{X}} = \boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}$ is the modelled part of $\boldsymbol{X}$, and $\boldsymbol{E}$ is the residual matrix. A three-dimensional model can be constructed by folding back the two-dimensional model. The score matrix $\boldsymbol{A}$ remains unchanged, but $\hat{\boldsymbol{X}}$, $\boldsymbol{B}$ and $\boldsymbol{E}$ are converted into their three-dimensional representations $\underline{\hat{\boldsymbol{X}}}$, $\underline{\boldsymbol{B}}$ and $\underline{\boldsymbol{E}}$. Fig. 2a visualises the unfold-PCA model, and the unfolding and back folding process.

The PARAFAC model of $\underline{\boldsymbol{X}}$ with R components (Harshman, 1970; Carroll & Chang, 1970) can be described as

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk}, \tag{3}$$

where: i, j and k are running indices for the units of the three different modes I, J and K; r is an index of the R components; x_{ijk}, a_{ir}, b_{jr}, c_{kr} and e_{ijk} are elements of $\underline{\boldsymbol{X}}(I \times J \times K)$, $\boldsymbol{A}(I \times R)$, $\boldsymbol{B}(J \times R)$, $\boldsymbol{C}(K \times R)$ and $\underline{\boldsymbol{E}}(I \times J \times K)$; $\hat{x}_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$ is the modelled part of x_{ijk}. Vectors of $\boldsymbol{A}$ are called score vectors and vectors of $\boldsymbol{B}$ and $\boldsymbol{C}$ are called loading vectors. A two-way matrix representation of the PARAFAC model is given by

$$\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{C} \circ \boldsymbol{B})^{\mathrm{T}} + \boldsymbol{E}, \tag{4}$$

where $\circ$ indicates the Khatri–Rao product (Rao & Mitra, 1971) of $\boldsymbol{C}$ and $\boldsymbol{B}$ partitioned in columns (see appendix). Without changing the model, the vectors of $\boldsymbol{B}$ and $\boldsymbol{C}$ can be normalised to unit length when simultaneously the matching vectors of $\boldsymbol{A}$ are compensated. Fig. 2b visualises the PARAFAC model.

The Tucker3 model of $\underline{\boldsymbol{X}}$, with R, S and T components for the first, second and third mode, respectively (Tucker, 1966; Kroonenberg & De Leeuw, 1980), can be described as

$$x_{ijk} = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} a_{ir} b_{js} c_{kt} h_{rst} + e_{ijk}, \tag{5}$$

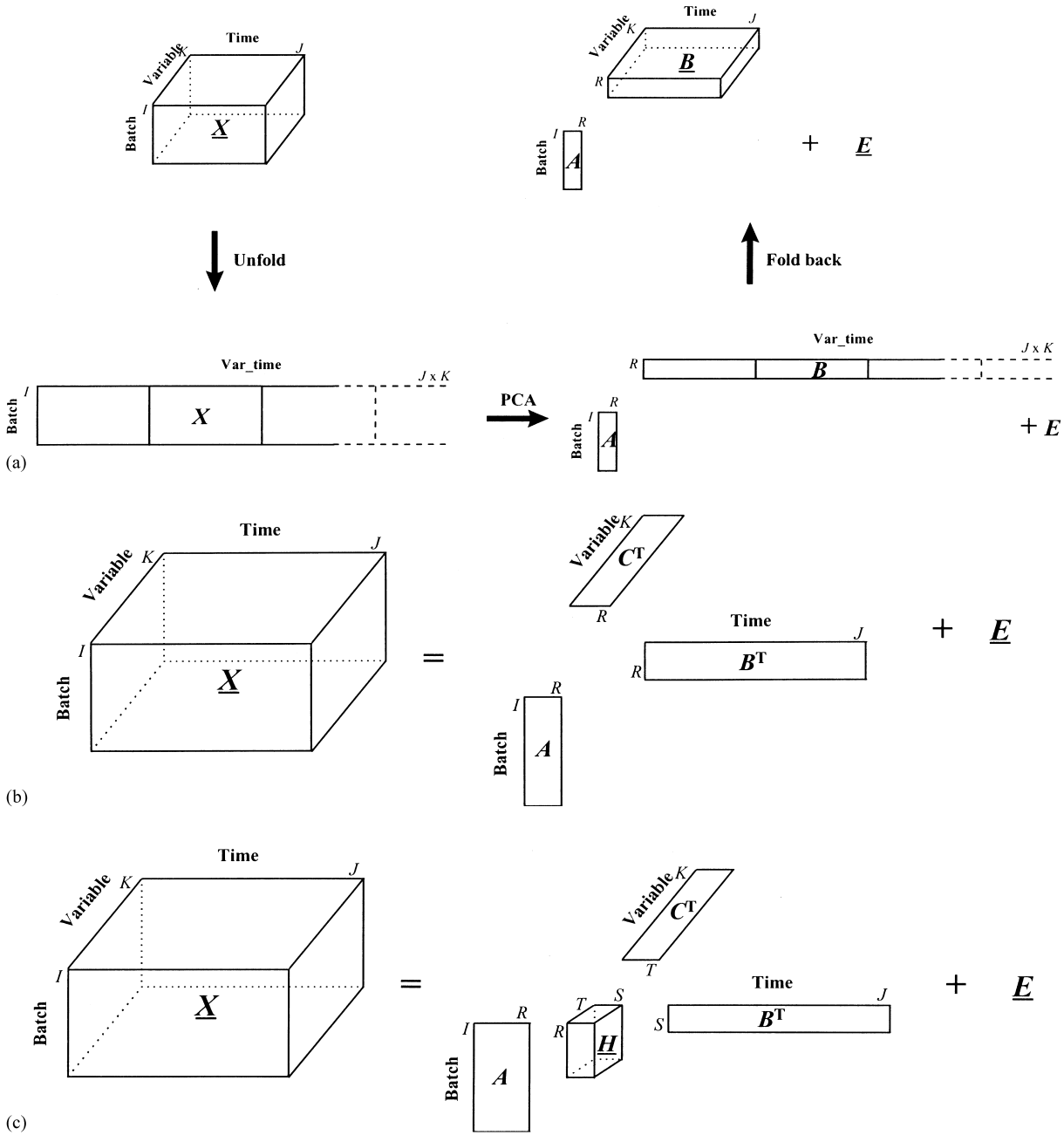


Fig. 2. Three dimensional representation of (a) the unfold-PCA model, (b) the PARAFAC model and (c) the Tucker3 model.

where: r, s and t are indices for the components for the different modes; x_{ijk}, a_{ir}, b_{js}, c_{kt}, h_{rst} and e_{ijk} are elements of $\underline{\boldsymbol{X}}(I \times J \times K)$, $\boldsymbol{A}(I \times R)$, $\boldsymbol{B}(J \times S)$, $\boldsymbol{C}(K \times T)$, $\underline{\boldsymbol{H}}(R \times S \times T)$ and $\underline{\boldsymbol{E}}(I \times J \times K)$; $\hat{x}_{ijk} = \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T} a_{ir}b_{js}c_{kt}h_{rst}$ is the modelled part of x_{ijk}. Vectors of $\boldsymbol{A}$ are called score vectors and vectors of $\boldsymbol{B}$ and $\boldsymbol{C}$ are called loading vectors. Elements of $\underline{\boldsymbol{H}}$ are weights for all possible component interactions. A two-way matrix representation of the Tucker3 model often is given by

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{H}(\boldsymbol{C} \otimes \boldsymbol{B})^{\mathrm{T}} + \boldsymbol{E}, \tag{6}$$

where $\otimes$ indicates the Kronecker product and $\boldsymbol{H}(R \times ST)$ is a two-way representation of $\underline{\boldsymbol{H}}(R \times S \times T)$. Without changing the model the matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ can be made column-orthogonal. The variance described by the model is then accounted for in $\underline{\boldsymbol{H}}$. Fig. 2c visualises the Tucker3 model.

It can be shown that for a given number of components R the unfold-PCA model always fits better than the Tucker3 model, which in turn fits always better than the PARAFAC model (Kiers, 1991). This can be explained by the difference in the amount of parameters for each

model. The unfold-PCA model estimates for every component a parameter for each time point of every variable. The high number of parameters can describe a relative large amount of variation. The PARAFAC model only estimates a parameter for each time point over the variables and a parameter for each variable over the time points. For the time mode the effect of the variables are averaged and vice versa. Due to this averaging there is less variation explained by the PARAFAC model compared to the unfold-PCA model. The PARAFAC model only allows interactions between the same component of all modes. The Tucker3 model also allows interactions between a particular component of a certain mode with all components of the other modes. For this reason, more variation is explained by the Tucker3 model compared to the PARAFAC model with a similar model size. The reasoning above is comparable to the case that a relationship between an x and y can increasingly be fitted better with higher-order polynomials (more complex models). The question is, of course, how complex the model should be.

3. Projecting a new batch on an existing model

To construct MSPC charts it is necessary to model batch data representing the normal operating conditions (NOC) with unfold-PCA, PARAFAC or Tucker3. Then data of new batches can be compared with the NOC data by projecting the new batches on one of the models. A data vector of a new batch, $\boldsymbol{x}_{\text{new}}(M \times 1)$, can be projected on either model to calculate a new score vector, $\boldsymbol{a}_{\text{new}}(R \times 1)$ and a new residual vector $\boldsymbol{e}_{\text{new}}(M \times 1)$. For unfold-PCA with column orthogonal $\boldsymbol{B}$, Eq. (2) is rearranged and the score and residuals of the new batch are (Nomikos & MacGregor, 1994)

$$\boldsymbol{a}_{\text{new}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}_{\text{new}},$$

$$\boldsymbol{e}_{\text{new}} = \boldsymbol{x}_{\text{new}} - \boldsymbol{B}\boldsymbol{a}_{\text{new}}. \tag{7}$$

For PARAFAC Eq. (4) is similarly rearranged and the score and residuals are

$$\boldsymbol{a}_{\text{new}} = [(\boldsymbol{C} \circ \boldsymbol{B})^{\mathrm{T}}(\boldsymbol{C} \circ \boldsymbol{B})]^{-1}(\boldsymbol{C} \circ \boldsymbol{B})^{\mathrm{T}}\boldsymbol{x}_{\text{new}},$$

$$\boldsymbol{e}_{\text{new}} = \boldsymbol{x}_{\text{new}} - (\boldsymbol{C} \circ \boldsymbol{B})\boldsymbol{a}_{\text{new}}. \tag{8}$$

For Tucker3 rearranging Eq. (6) results in

$$\boldsymbol{a}_{\text{new}} = (\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{H}[(\boldsymbol{C}\otimes\boldsymbol{B})^{\mathrm{T}}(\boldsymbol{C}\otimes\boldsymbol{B})]^{-1}(\boldsymbol{C}\otimes\boldsymbol{B})^{\mathrm{T}}\boldsymbol{x}_{\text{new}},$$

$$\boldsymbol{e}_{\text{new}} = \boldsymbol{x}_{\text{new}} - (\boldsymbol{C}\otimes\boldsymbol{B})\boldsymbol{H}^{\mathrm{T}}\boldsymbol{a}_{\text{new}}. \tag{9}$$

MSPC charts can be used to compare the scores and residuals of the new batch with the scores and residuals of the NOC data.

4. Batch MSPC charts

Control charts that are used in batch MSPC generally are based on the Hotelling statistic (D-chart) and on the residuals statistic (Q-chart).

4.1. Hotelling statistic: D-chart

This statistic assesses the statistical significance of the difference between two sets of samples, drawn from a multivariate normal distribution. The definition of the Hotelling statistic is explained in the appendix. If instead of $\boldsymbol{x}_i$ with J variables, a reduced (principal) component space with R components and PCA-, PARAFAC- or Tucker3 score vectors is used, the T^2-statistic is then called the D-statistic (Nomikos & MacGregor, 1994) and becomes

$$(\boldsymbol{a}_{\text{new}} - \bar{\boldsymbol{a}}_I)^{\mathrm{T}}\boldsymbol{S}_R^{-1}(\boldsymbol{a}_{\text{new}} - \bar{\boldsymbol{a}}_I)\frac{I(I-R)}{R(I^2-1)} \sim F(R, I-R), \tag{10}$$

where $\boldsymbol{a}_{\text{new}}$ is the score of the new batch, $\bar{\boldsymbol{a}}_I$ the average score of the initial I batches and $\boldsymbol{S}_R$ the covariance matrix of the I initial scores with R components. The $\bar{\boldsymbol{a}}_I$ contain the column averages of $\boldsymbol{A}(I \times R)$ and $\boldsymbol{S}_R$ is calculated as $(1/(I-1))\,\boldsymbol{A}_c^{\mathrm{T}}\boldsymbol{A}_c$ where $\boldsymbol{A}_c$ is the column centered $\boldsymbol{A}$. This is the approach often used in batch MSPC.

The problem with this approach, however, is that the I initial scores of the reference set are constructed differently from the future scores of the new batches. The future scores are not drawn from the same distribution as the initial scores. The initial scores are maximised in order to optimally model the variance of the data, whereas the future scores are calculated by fitting new data on this model.

A better approach is schematically given hereafter:

a. For all batches $i = 1, \ldots, I$.
b. Model the data from all initial batches except batch i $(1, \ldots, i-1, i+1, \ldots, I)$.
c. Project batch i on the model, calculate the score vector $\boldsymbol{a}_i$ and the residual vector $\boldsymbol{e}_i$.
d. End i-loop.
e. Calculate the average score vector $\bar{\boldsymbol{a}}_I$ and the covariance matrix of the projected scores $\boldsymbol{S}_R$ as mentioned above.
f. Develop the D- and Q-charts.
g. Calculate a model with all initial batches I, project a new batch on this model and calculate $\boldsymbol{a}_{\text{new}}$ and $\boldsymbol{e}_{\text{new}}$.
h. Use $\boldsymbol{a}_{\text{new}}$, $\bar{\boldsymbol{a}}_I$ and $\boldsymbol{S}_R$ to calculate the D-statistic with Eq. (10) and use $\boldsymbol{e}_{\text{new}}$ to calculate the Q-statistic.

In this way all initial and future scores are projected and can be assumed coming from the same distribution and Eq. (10) holds. A D-chart can be constructed based on the F-distribution and a desired error limit. A small error is made since the score of the future batch is

projected on a model based on one extra batch compared to the score of the initial batches.

4.2. Residuals statistic: Q-chart

The sum of squared residuals of batch i can be calculated as

$$Q_i = \sum_{j=1}^{J} \sum_{k=1}^{K} e_{ijk}^2, \tag{11}$$

where e_{ijk} is the typical element of $\boldsymbol{E}_i$ $(J \times K)$, the matrix of residuals of the ith batch which is the back-folded $\boldsymbol{e}_i$ $(JK \times 1)$ vector (see Eqs. (7)–(9)). This Q_i value should remain within certain limits. If the residuals are independent and normally distributed Q_i follows a χ^2-distribution (Nomikos & MacGregor, 1994,1995),

$$Q_i \sim g\chi_h^2, \tag{12}$$

with h degrees of freedom and a weight g to account for the magnitude of Q_i. The residuals, however, still can be partly correlated. Therefore, the h degrees of freedom of Eq. (12) are not known in advance. However, it can still be assumed that the sum of squares, Q_i $(i = 1, \ldots, I)$, of the NOC batches follow this distribution, $Q_i \sim g\chi_h^2$ (Nomikos & MacGregor, 1995). The weight g and the degrees of freedom h of this distribution can be estimated with the first $(\mu_{g\chi_h^2} = g \cdot h)$ and second $(\sigma_{g\chi_h^2}^2 = 2g^2 \cdot h)$ moment of this distribution. Q-charts can be constructed for Q_i based on this distribution and a desired error limit. An alternative procedure to find control limits for Q_i is given by Jackson and Mudholkar (1979), which is similar to the approach presented above.

In traditional Q-charts the parameters g and h, and hence the upper 95% confidence limit, Q_{95}, is constant for the whole batch run (Nomikos & MacGregor, 1994). A new method is proposed in this paper, that uses several models for on-line monitoring, this will be explained later. As a consequence, different values of g and h are obtained, and consequently, the upper 95% confidence limit Q_{95} can vary as a function of time. This makes the interpretation more difficult, a straight line upper limit is common practice and for operators easier to deal with. This problem can be avoided easily by plotting the ratio Q_i/Q_{95} or Q_i/Q_{99}, the upper limit line is then constant like in case of the D-chart.

4.3. Detecting power

Usually the upper limits of the charts are set on a 95 or 99% significance level. When the results of the D-charts and Q-charts have to be compared for several models and several sizes of the models, actually various distributions are being compared. For instance the D-statistic value based on a model with $R = 2$ is being compared with a value based on $R = 3$, or Q-statistic values with several degrees of freedom h are being compared. In order to obtain comparable results, p-values will be reported. This p-value is the probability that a value at least as large as the statistic obtained comes from normal operating conditions. The 95 or 99% significance level corresponds to p-values of 0.05 (5%) or 0.01 (1%). The D- and Q-statistic of an erroneous batch calculated with one of the models, ideally has a p-value much lower then 0.05 or 0.01. Stated otherwise, if a p-value lower than 0.05 or 0.01 is found, the control charts signals an error. By reporting the p-values, an idea is given of the average run length (ARL) of the control chart; that is, how long does it take for a control chart to detect an error. ARLs are performance measures of control charts; the longer it takes to detect an error (high ARL), the poorer the performance of the control chart (Wetherill & Brown, 1991).

5. On-line monitoring

In order to be able to monitor a batch process on-line, the total run time K of a batch process is subdivided in several time periods. One way to subdivide the total run time of a batch is according to scheduling points, when distinct stages are detectable due to the chemistry or physics of the process, or for specific chemical or physical reasons. The total run time can be subdivided accordingly and for every stage, or for every time period between two scheduling points, the corresponding NOC data is modelled and a D- and Q-chart is constructed. If there are no specific scheduling points or distinct stages the run time can be subdivided in expanding time periods, like $0\text{-}K/n, 0\text{-}2K/n, \ldots, 0\text{-}K$ time points.

It is also possible to use a more complex approach, introduced by Ränner, MacGregor and Wold (1998), based on a recursive multi-block PCA method. This method processes data in a sequential and adaptive manner with a controlled rate of adaptation. Another alternative is given by Nomikos and MacGregor (1994,1995). These approaches can also be used for PARAFAC and Tucker3 models.

The D-chart measures the deviation between a new batch and the normal operating batches in terms of variation which can be handled by the model, whereas the Q-chart shows the variation which cannot be handled by the model. Hence, both charts are complementary (Nomikos & MacGregor, 1994; Nomikos & MacGregor, 1995). When a new batch has a large D-statistic and a moderate Q-statistic, then it means that variation is present in this new batch which is already to some extent present in the training (NOC) data, but is extreme compared to the normal operating batches, e.g. a slow drift in temperatures and pressure might show such behavior. When a new batch has a large Q-statistic, then completely new variation is encountered, not present in the model, e.g. sudden upsets and sensor failure might show

such behaviour. Hence, using the two types of control charts simultaneously is important from a chemical engineering point of view: they can both signal events, but they have different diagnostics values.

When a chart signals an out-of-control situation, contribution plots can help in locating which variables cause the out-of-control situation and diagnose the possible cause of the erroneous behaviour (Miller, Swanson & Heckler, 1998; Kourti, Nomikos & MacGregor, 1995; Boqué & Smilde, 1999).

6. Experimental

The use of batch MSPC charts for the several mentioned three-way models is illustrated with a benchmark data set of a simulated semi-batch emulsion polymerisation of styrene–butadiene (Broadhead, Hamielec and MacGregor, 1985). Meaningful disturbances like impurities in the initial charge of the organic phase and in the butadiene feed to the reactor were added. Measurements were taken from flow rates, temperatures, density, estimates of the conversion and energy release. A detailed description can be found in literature (Nomikos & MacGregor, 1994; Broadhead et al., 1985). Fifty batches were simulated to construct the NOC data, by introducing typical variations.

Three additional batches were simulated, one with normal conditions and two with product that was out of the specification region. One erroneous batch had an initial organic impurity contamination in the butadiene feed. The other erroneous batch had the same problem, but the contamination was higher and started halfway through the batch operation.

The NOC data is arranged in a three-way array $\underline{\boldsymbol{X}}(I \times J \times K)$ of $I = 50$ batches, $J = 9$ variables and $K = 200$ time points. To describe the variation of the variables about their average trajectory, for every variable at each time point the 50 values were centred and scaled to unit variance. The three additional batches are scaled with the NOC parameters.

On-line monitoring was achieved by subdividing the total run time K in 10 time periods with 0–20, 0–40, 0–60, …, 0–200 time points. For every time period the corresponding NOC data is modelled, and a D-chart and Q-chart is constructed.

6.1. PCA, PARAFAC and Tucker3 batch MSPC charts

Batch MSPC charts based on an unfold-PCA model with two and with three principal components were constructed as described earlier. Results of using unfold-PCA for modelling the D- and Q-charts are presented with two methods. The method that uses the Hotelling statistic for PCA as described in literature, called the uncorrected model; and the new approach presented in this paper, called the corrected model. The new (corrected) approach is also used for charts based on the PARAFAC and the Tucker3 model. To compare the performance of PARAFAC and unfold-PCA, batch MSPC charts based on the PARAFAC model are also constructed with two and three components. Since it is the aim to model differences between batches and also for computational convenience, the PARAFAC score vectors of the batch mode were constrained to be orthogonal (Harshman, 1970). For Tucker3, batch MSPC charts are constructed based on a model with $R = 4$, $S = 2$, $T = 3$ components. R, S and T, respectively, represent the number of components for the batch, the variable and the time mode. The choice of the Tucker3 model size is indicated by Tucker3 cross-validation results (Louwerse, Smilde & Kiers, 1998).

The p-values of the D- and Q-statistic for the three future batches are calculated for all models and all time periods. For instance, in case of the Tucker3 model with $R = 4$, $S = 2$ and $T = 3$ components the D-statistic, according to Eq. (10), is compared with the $F(4,45)$-distribution to calculate a p-value. The four-degrees of freedom are from the number of components of the batch mode, and the 45 degrees of freedom are from the 50 reference batch scores minus the four components and one lost degree because of the centring operation. The Q-statistic is compared to the χ^2-distribution of Eq. (12) to calculate a p-value; g and h are calculated with the NOC values and are known when future batches are projected.

7. Results and discussion

The results will be discussed keeping in mind the goals of MSPC: detecting, locating and diagnosing erroneous variation. Hence, it is important that at least one of the control charts gives an out-of-control signal (detecting), but it is also important which chart signals (locating and diagnosing).

Table 1 shows the amount of variance that is explained for all described models. Besides that models with more components explain more variation, unfold-PCA explains the largest part of the variation for a certain number of components, as already mentioned in the theory section. The $4 \times 2 \times 3$ component Tucker3 model explains more variation than the two component unfold-PCA model. It seems that the Tucker3 model is considerably larger than the unfold-PCA model with respect to the number of components. However, the number of estimated parameters for the NOC data modelled with Tucker3 ($4 \times 50 + 2 \times 9 + 3 \times 200 + 4 \times 2 \times 3 = 842$) is small when it is compared to the unfold-PCA model ($2 \times 50 + 2 \times (9 \times 200) = 3700$). The amount of variation explained by the models might seem low, but this is not an exception in batch models (Nomikos

Table 1
Percentage explained variation of the NOC data as a function of the model size and the modelled time periods

Model size	PCA		PARAFAC		Tucker3
Model time period	2	3	2	3	$4 \times 2 \times 3$
0–20	49.3	63.0	38.1	47.9	55.3
0–40	36.3	49.6	25.1	34.2	40.9
0–60	28.4	39.6	21.0	27.6	33.6
0–80	25.7	35.2	19.7	25.4	29.6
0–100	24.6	32.7	18.7	23.6	26.8
0–120	24.2	31.5	17.8	22.5	25.9
0–140	23.5	30.6	17.3	21.6	25.1
0–160	23.4	30.0	17.8	21.6	24.9
0–180	23.9	30.1	18.8	22.3	25.6
0–200	24.0	29.9	19.1	22.4	25.6

Table 2
(a) *D*-statistic of a future batch, erroneous from the start

Model	Comp. no.	Modelled time periods 0–20	0–40	0–60	0–80	0–100	0–120	0–140	0–160	0–180	0–200
UNFOLD-PCA$_{unc}$	2	14	2.4	**0.049**	**0.024**	**0.038**	**0.030**	**0.019**	**0.007**	**0.005**	**0.004**
	3	6.0	4.9	**0.16**	**0.089**	**0.11**	**0.11**	**0.071**	**0.027**	**0.019**	**0.014**
UNFOLD-PCA$_{corr}$	2	10	**0.70**	**0.001**	**0.001**	**0.002**	**0.003**	**0.002**	**< 0.001**	**< 0.001**	**< 0.001**
	3	7.3	1.8	**0.004**	**0.005**	**< 0.001**	**0.011**	**0.007**	**0.001**	**0.001**	**< 0.001**
PARAFAC	2	6.1	**0.12**	**0.010**	**0.026**	**0.008**	**0.003**	**0.004**	**0.002**	**< 0.001**	**0.001**
	3	16	**0.43**	**0.048**	**0.062**	**0.004**	**0.014**	**0.007**	**0.004**	**0.004**	**0.002**
Tucker3	4, 2, 3	9.0	1.6	**0.21**	**0.69**	**0.075**	**0.44**	**0.37**	**0.085**	**0.097**	**0.050**

P-values $*$ 100% are tabulated as a function of the model, the model size and the used time periods. The boldface numbers indicate a detection signal ($\alpha = 0.01$)

(b) *Q*-statistic of a future batch, erroneous from the start

Model	Comp. no.	Modelled time periods 0–20	0–40	0–60	0–80	0–100	0–120	0–140	0–160	0–180	0–200
UNFOLD-PCA$_{unc}$	2	**0.23**	**0.007**	3.1	19	2.7	4.4	6.8	**0.028**	**0.073**	**0.24**
	3	**0.13**	**< 0.001**	**0.34**	8.7	**0.85**	1.4	1.3	**< 0.001**	**0.002**	**0.026**
UNFOLD-PCA$_{corr}$	2	2.5	**0.31**	7.5	35	7.7	10	17	**0.13**	**0.30**	1.2
	3	2.4	**0.016**	6.0	29	11	9.5	12	**0.05**8	**0.16**	**0.85**
PARAFAC	2	11	11	8.8	5.1	6.9	13	20	5.4	3.3	4.5
	3	2.4	4.2	3.5	3.3	12	14	24	3.8	3.2	4.8
Tucker3	4, 2, 3	24	16	8.2	3.0	5.8	11	17	3.5	3.8	6.0

P-values $*$ 100% are tabulated as a function of the model, the model size and the used time periods. The boldface numbers indicate a detection signal ($\alpha = 0.01$).

& MacGregor, 1994,1995). Note that the variation not captured by the model is summarized in the Q-charts.

Results of the *D*- and *Q*-statistic are shown in Tables 2 and 3. The *p*-values of the erroneous future batches are tabulated as a function of the used models and the modelled time periods. The *p*-value of an erroneous batch ideally is as low as possible, at least below 0.05 (or 5%), corresponding to a significance level of 95%. For convenience all *p*-values lower than 0.01 (1%) are made bold in the tables; this represents cases where the control

Table 3
(a) *D*-statistic of a future batch, erroneous from halfway

Model	Comp. no.	Modelled time periods 0–120	0–140	0–160	0–180	0–200
UNFOLD-PCA$_{unc}$	2	36	24	2.6	**0.65**	**0.098**
	3	55	42	5.4	1.9	0.33
UNFOLD-PCA$_{corr}$	2	19	4.5	**0.10**	**0.048**	**0.001**
	3	31	10	**0.23**	**0.17**	**0.005**
PARAFAC	2	**0.87**	**0.37**	**< 0.001**	**< 0.001**	**< 0.001**
	3	1.2	9.0	**< 0.001**	**0.005**	**< 0.001**
Tucker3	4, 2, 3	2.8	**0.054**	**0.019**	**0.040**	**0.006**

P-values * 100% are tabulated as a function of the model, the model size and the used time periods. The boldface numbers indicate a detection signal ($\alpha = 0.01$).

(b) *Q*-statistic of a future batch, erroneous from halfway

Model	Comp. no.	Modelled time periods 0–120	0–140	0–160	0–180	0–200
UNFOLD-PCA$_{unc}$	2	**0.060**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
	3	**0.006**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
UNFOLD-PCA$_{corr}$	2	**0.27**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
	3	**0.24**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
PARAFAC	2	1.1	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
	3	**0.39**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
Tucker3	4, 2, 3	1.9	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**

P-values * 100% are tabulated as a function of the model, the model size and the used time periods. The boldface numbers indicate a detection signal ($\alpha = 0.01$).

chart clearly detects abnormal behaviour. All *p*-values of the normal behaving future batch were larger than 0.05 (or 5%), corresponding to a significance level below 95%, they are not individually reported.

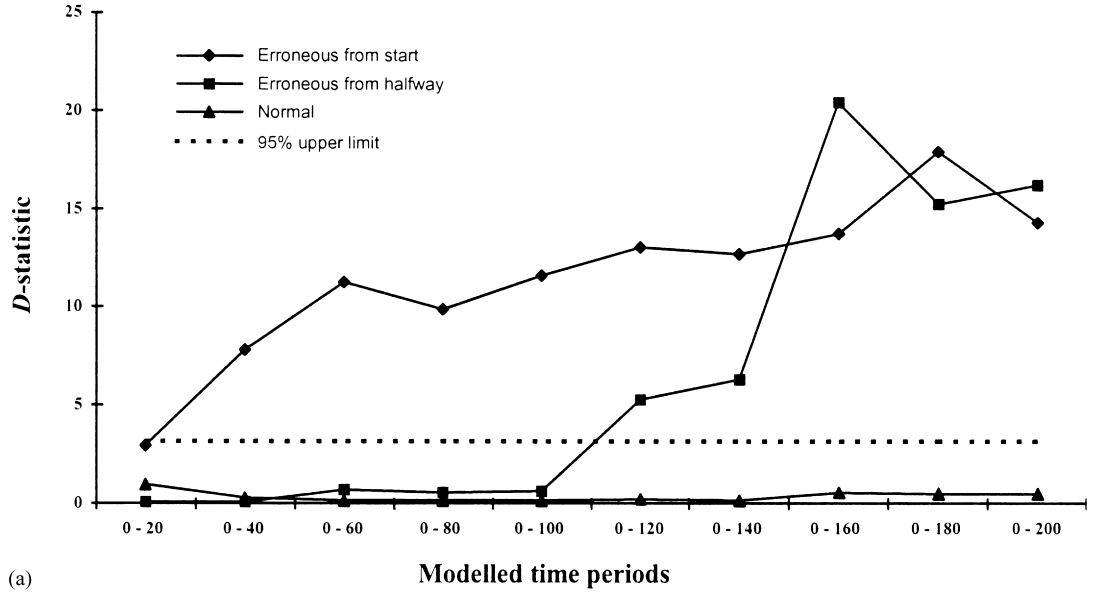
The *p*-values of the future batch erroneous from halfway, are reported after the introduction of the error. Control charts of the PARAFAC model with two components are shown as an example in Fig. 3. The *D*- and *Q*-statistic is shown for all three future batches.

The results will be discussed in three parts: (i) comparing corrected and uncorrected unfold-PCA, (ii) comparing performance of the different models in terms of detecting erroneous behaviour in *either* one of the charts and (iii) comparing performance of the different models in terms of the separate *D*- and *Q*-charts.

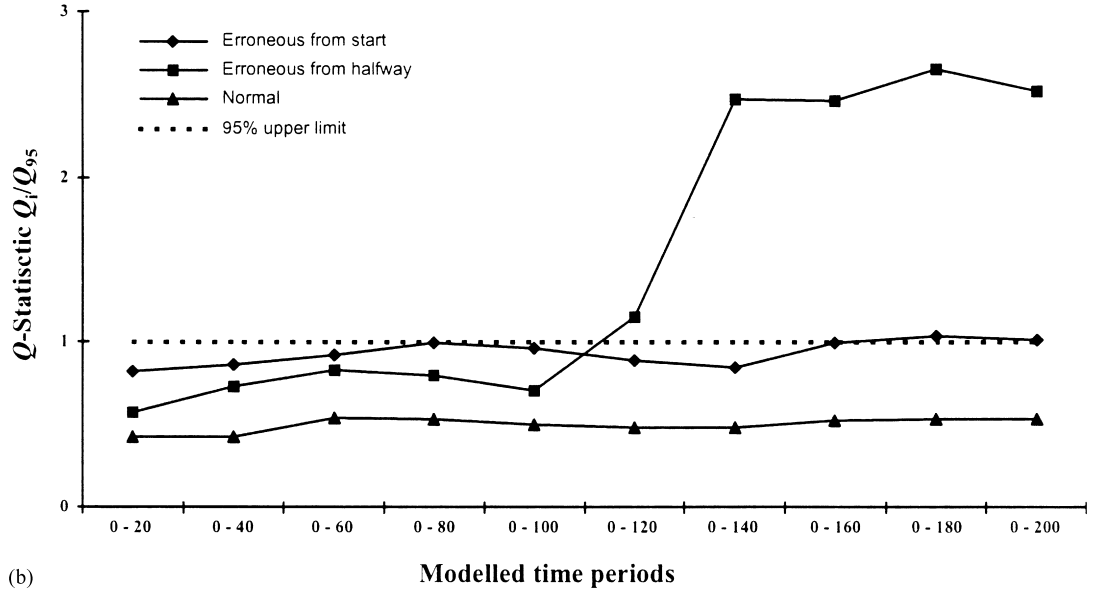
7.1. Corrected versus uncorrected unfold-PCA

When the uncorrected unfold-PCA (unfold-PCA$_{unc}$) method is compared to the corrected unfold-PCA (unfold-PCA$_{corr}$) method there is a clear difference. When an error is introduced in the operating conditions of a running batch, the *D*-statistic signals earlier (see Table 2a and Table 3a) and the *Q*-statistic signals later (see Table 2b and Table 3b) when using unfold-PCA$_{corr}$. Stated otherwise, for a given significance level α, the *D*-limits become smaller and the *Q*-limits become larger in the corrected approach.

In chemical engineering practice, an α of 99% is chosen for the *D*-chart and an α of 99.9% is chosen for the *Q*-chart to avoid the pitfall of the uncorrected approach. In practice, choosing a 99% level for the *Q*-chart in the uncorrected approach leads to more false positive signals, that is, an actual type I error larger than 1% (Nomikos, 1996). To reduce this actual type I error, a limit of 99.9% for the *Q*-chart is chosen in chemical engineering practice. Hence, using the corrected approach, this problem is solved since the control limits are better approximates of the true limits. The results of the corrected unfold-PCA support and motivate the chemical engineering practice. From now on, only the results of unfold-PCA$_{corr}$ will be considered.



(a)



(b)

Fig. 3. (a) *D*-statistic of future batches as a function of the modelled time periods for a PARAFAC model with 2 components. (b) Quotient of the *Q*-statistic of future batches and the 95% upper limit as a function of the modelled time periods for a PARAFAC model with 2 components.

7.2. Comparing performance of the different models in terms of detecting erroneous behaviour in either one of the charts

For the batch erroneous from the start, both PARAFAC and unfold-PCA_{corr} give a signal after 40 time points. Tucker3 detects slightly later; at 60 time points. For the batch which is erroneous halfway the batch run, the PARAFAC and Unfold-PCA_{corr} charts detect this directly (time point 120). The Tucker3 charts lag again slightly behind (after 140 time points).

Summarizing, in overall detecting capabilities the PARAFAC and unfold-PCA_{corr} charts seem to perform similarly. The Tucker3 based charts detect slightly later. The Tucker3 charts were not optimized in terms of number of Tucker3 components. This might have improved the performance of the Tucker3 chart.

7.3. Comparing performance of the different models in terms of the separate D- and Q-charts

D-charts based on the PARAFAC model perform slightly better than the unfold-PCA_{corr} ones, especially for the batch with an error halfway (Table 3a): the PARAFAC model with two components detects the error in a very early stage. The unfold-PCA_{corr} charts are

comparable in performance to the Tucker3 based charts (see Table 2a and Table 3a), although the Tucker3 based chart shows a slightly better performance for the batch with an error halfway (Table 3a). It seems to be favourable, both for PARAFAC and unfold-PCA$_{corr}$, to have a parsimonious model, i.e., to use a low number of components. The performance of the Tucker3 model might be enhanced by lowering the number of components in the batch mode. This was not pursued further.

The Q-chart based on unfold-PCA$_{corr}$ performs slightly better than the PARAFAC and Tucker3 based charts for the batch with an error in the start (Table 2b). Note that also the unfold-PCA$_{corr}$ chart has problems in detecting this error once the batch has run for 60 time points or more. Apparently this error is hard to detect in a Q-chart. The initial organic impurity in the feed at the start of the batch causes extreme variation which can be modelled to some extent, since D-charts detect the abnormality and the Q-charts do not show consistent high-residual errors.

For the batch with an error halfway, the conclusions are less clear (Table 3b). The unfold-PCA$_{corr}$ and PARAFAC based Q-charts perform comparable. The Tucker3 based chart performs perhaps slightly worse, but it still gives a warning if a level of 5% is defined as warning limit.

Summarizing, the PARAFAC and Tucker3 based charts seem to perform slightly better in the D-chart than unfold-PCA$_{corr}$. In the Q-chart, unfold-PCA$_{corr}$ seems to perform slightly better. Hence, there is no king method. All depends on what types of faults are to be expected and detected in the batch process being monitored. Chemical engineering experience has to show what methods are the most useful for a given application.

8. Conclusions

The correction of unfold-PCA proposed in this paper is theoretically justified and improves the performance of the Q-chart based on unfold-PCA. The correction removes the necessity of using a too high limit for the Q-chart, which was the engineering solution of the problem with the Q-chart.

For overall detection, that is, receiving a signal from either the D- or Q-chart, there is no king method. PARAFAC and unfold-PCA$_{corr}$ perform equally well. The D-charts perform best for parsimonious models, i.e. PARAFAC models with a low number of components. Hence, errors like slow drift in the batch process tend to be detected earlier with, e.g., a PARAFAC based D-chart. In the Q-chart, unfold-PCA$_{corr}$ seems to perform slightly better.

The strategy of using a sequence of models, based on a growing number of time points seems to work satisfactory. Of course, it is possible to divide the training data in a finer grid resulting in, e.g., 200 models. Since all calculations can be done off-line, once these models are known they can be used on-line in a straightforward manner.

Appendix A

A.1. Khatri–Rao product

If the matrices $\boldsymbol{C}(K \times R)$ and $\boldsymbol{B}(J \times R)$, are partitioned into R column vectors, $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_R], \boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_R]$, then the Khatri–Rao product is

$$\boldsymbol{C} \circ \boldsymbol{B} = [\boldsymbol{c}_1 \otimes \boldsymbol{b}_1, \ldots, \boldsymbol{c}_R \otimes \boldsymbol{b}_R],$$

where $\otimes$ is the Kronecker product.

A.2. Hotelling statistic

According to Seber (1984), T^2 is defined as

$$T^2 = m\,\boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y} \quad \text{with } \boldsymbol{y} \sim N_d(0, \Sigma), \quad \boldsymbol{V} \sim W_d(m, \Sigma), \tag{A.1}$$

where $\boldsymbol{y}(d \times 1)$ is a multivariate observation; $\boldsymbol{V}(d \times d)$ is the covariance matrix of m multivariate observations; N_d and W_d denote the d-dimensional normal and Wishart distribution, respectively. The vector $\boldsymbol{y}$ and the matrix $\boldsymbol{V}$ are statistically independent. For the true multivariate population, the mean is zero and the covariance matrix is Σ. If this holds then T^2 follows an F distribution

$$\frac{(m-d+1)}{d}\frac{T^2}{m} \sim F(d, m-d+1). \tag{A.2}$$

Consider a set of initial multivariate observations $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_I$, and a future observation $\boldsymbol{x}_{new}$ all characterised by J variables. For this case, Tracy, Young and Mason (1992) derived that if

$$\boldsymbol{x}_i \sim N_J(\mu, \Sigma), \tag{A.3}$$

then

$$\bar{\boldsymbol{x}}_I \sim N_J(\mu, \Sigma/I),$$

$$(I-1)\,\boldsymbol{S}_I \sim W_J(I-1, \Sigma), \tag{A.4}$$

where $\bar{\boldsymbol{x}}_I$ and $\boldsymbol{S}_I$, respectively, are the average vector and the covariance matrix of the initial observations. In the case that $\boldsymbol{x}_{new}, \bar{\boldsymbol{x}}_I$, and $\boldsymbol{S}_I$ are independent it follows that

$$(\boldsymbol{x}_{new} - \bar{\boldsymbol{x}}_I)^{\mathrm{T}} S_I^{-1} (\boldsymbol{x}_{new} - \bar{\boldsymbol{x}}_I) \frac{I(I-J)}{J(I^2-1)} \sim F(J, I-J). \tag{A.5}$$

References

Boqué, R., & Smilde, A. K. (1999). Monitoring and diagnosing batch processes with multiway regression models, *AIChE Journal*, *45*(7), 1504–1520.

Broadhead, T. O., Hamielec, A. E., & MacGregor, J. F. (1985). Dynamic modeling of the batch, semi-batch and continuous production of styrene-butadiene copolymers by emulsion polymerisation. *Makromolekulare Chemie Supplement*, *10*, 105–128.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via a n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*, 283–319.

Dong, D., & McAvoy, T. J. (1996). Batch tracking via nonlinear principal component analysis. *AIChE Journal*, *42*, 2199–2208.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: model and conditions for an explanatory multi-mode factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.

Jackson, J. E., & Mudholkar, G. S. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, *21*, 341–349.

Kiers, H. A. L. (1991). Hierarchical relations among three-way methods. *Psychometrika*, *56*, 449–470.

Kourti, T., & MacGregor, J. F. (1995). Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Intelligent Laboratory Systems*, *28*, 3–21.

Kourti, T., Nomikos, P., & MacGregor, J. F. (1995). Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *Journal Proc. Control*, *5*, 277.

Kroonenberg, P. M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, *45*, 69–97.

Louwerse, D. J., Smilde, A. K., & Kiers, H. A. L. (1998). Cross-validation of multiway component models, *Journal of Chemometrics*, to appear.

Martin, E. B., Morris, A. J., Papazoglou, M. C., & Kiparissides, C. (1996). Batch process monitoring for consistent production. *Computers Chemical Engineering*, *20*, S599–S604.

Miller, P., Swanson, R. E., & Heckler, C. E. (1998). Contribution plots: a missing link in multivariate quality control. *Applied Mathematics and Computer Science*, *8*, 775–792.

Nomikos, P. (1996). Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis. *ISA Transactions*, *35*, 259–266.

Nomikos, P., & MacGregor, J. F. (1994). Monitoring batch processes using multiway principal component analysis. *AIChE Journal*, *40*, 1361–1375.

Nomikos, P., & MacGregor, J. F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics*, *37*, 41–59.

Rao, C. R., & Mitra, S. (1971). *Generalized inverse of matrices and its applications* (pp. 12–13), New York: Wiley.

Ränner, S., MacGregor, J. F., & Wold, S. (1998). Adaptive batch monitoring using hierarchical PCA. *Chemometrics and Intelligent Laboratory Systems*, *41*, 73–81.

Seber, G. A. F. (1984). *Multivariate observations*. New York: Wiley.

Tracy, N. D., Young, J. C., & Mason, R. L. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, *24*, 88–95.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279–311.

Wetherill, G. B., & Brown, D. W. (1991). *Statistical process control, theory and practice*. London: Chapman & Hall.