PERGAMON

Chemical Engineering Science 57 (2002) 3979–3991

Chemical Engineering Science

www.elsevier.com/locate/ces

Critical evaluation of approaches for on-line batch process monitoring

Eric N. M. van Sprang, Henk-Jan Ramaker, Johan A. Westerhuis, Stephen P. Gurden, Age K. Smilde*

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

Received 4 October 2000; received in revised form 26 February 2002; accepted 11 June 2002

Abstract

Since the introduction of batch process monitoring using component models in 1992, different approaches for statistical batch process monitoring have been suggested in the literature. This is the first evaluation of five proposed approaches so far. The differences and similarities between the approaches are highlighted. The derivation of control charts for these approaches are discussed. A control chart should give a fast and reliable detection of disturbances in the process. These features are evaluated for each approach by means of two performance indices. First, the action signal time for various disturbed batches is tested. Secondly, the probability of a false warning in a control chart is computed. In order to evaluate the five approaches, five different data sets are studied: one simulation of a batch process, three batch processes obtained from industry and one laboratory spectral data set. The obtained results for the performance indices are summarised and discussed. Recommendations helpful for practical use are given.

Keywords: Statistical process control; Batch process monitoring; On-line monitoring; Control chart; Action signal time; Type I error

1. Introduction

Batch processes are widespread in the chemical, pharmaceutical, food and bio-technical industries. Batch processes are characterised by finite duration, non-steady-state behaviour, high conversions and, most importantly, the recipe-driven approach. If during the development phase a completed batch run is known to have produced an on-specification product, the most important process variables from that batch are stored in a recipe. This recipe is then used for future batch runs, with an aim to consistent production of the on-specification product. However, due to the complexity of most industrial batch reactions the recipe-driven approach is not always sufficient. Monitoring of batch processes is advantageous for several reasons: safety, cost reduction, quality control, fault diagnosis, fault detection, process improvement or a better process understanding.

Since the pioneering work of multivariate statistical process control (MSPC) for batch processes by MacGregor and Nomikos (1992), Nomikos and MacGregor (1994) batch process monitoring (BPM) has been applied to several different processes. Furthermore, commercial companies have developed BPM software and incorporated some of the proposed methods. There is a clear trend in the ascent of supervisory control and data acquisition systems (SCADA) and distributed control systems (DCS) in process plants. The combination of these systems and BPM seems to be very promising.

The main concept of BPM is to model the common-cause variation present in batch runs obtained under normal operating conditions (NOC). This model is subsequently used to determine whether a new batch corresponds to normal operating behaviour or not. Therefore, the monitoring performance depends heavily upon this NOC data.

Since the introduction of BPM in 1992, various researchers have proposed several variants and extensions to the original methodology proposed. Wold, Kettaneh, Fridén, and Holmberg (1998) introduced a variant of multivariate batch modelling for monitoring and diagnosis purposes. An alternative way of arranging the data is the most important feature of this work. Rännar, MacGregor, and Wold (1998) used an approach based on a recursive multiblock (hierarchical) PCA/PLS method. Louwerse and Smilde (1999) extended BPM by the use of PARAllel FACtor analysis

* Corresponding author. Tel: +31-20-525-5062; fax: +31-20-525-6638.
E-mail address: asmilde@its.chem.uva.nl (A. K. Smilde).

0009-2509/02/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0009-2509(02)00338-X

(PARAFAC) and Tucker three-way models. Additionally, improved control charts were developed using a modified PCA method. Boqué and Smilde (1999) used multivariate statistical procedures based on multiway covariates regression models. Martin and Morris (1996); Martin, Morris, Papazoglou, and Kiparisassides (1996) introduced a control chart based on a non-parametric method. Wise, Gallagher, Butler, White, and Barna (1999), Westerhuis, Kourti, and MacGregor (1999) and Dahl, Piovoso, and Kosanovich (1999) applied and compared several alternatives for multivariate statistical analysis of batch process data.

The aim of this paper is to summarise and critically evaluate the different approaches proposed to statistical batch process monitoring in the literature so far. From the proposed approaches a representative selection is made. Only approaches which build component models of process data are considered, that is, no regression models between process conditions and product quality data are compared. These are chosen to focus on the component models since the subject of regression models is a field of its own. The particular properties, advantages and disadvantages of each proposed approach are investigated. This is the first systematic treatment of type I error and type II error for a variety of methods, which is very useful for practice. The evaluation is performed using five data sets covering a range of industrial and laboratory processes. The different approaches are compared by the use of two performance indices: the overall type I error and the action signal time (AST).

Since the different approaches were originally proposed by different authors using different notations, each approach is summarised in Section 2 using a uniform notation. The similarities and differences in terms of how the data is modelled and used for on-line monitoring are highlighted. Furthermore, this section describes the derivation of the control charts used for the D- and SPE-statistics. Section 3 introduces and defines the two performance indices, the overall type I error and AST, used to evaluate and compare the approaches. Section 4 gives a short description of the data sets studied. The results of the evaluation are given in Section 5. Finally, conclusions are drawn and recommendations are given in Section 6.

2. Theory

2.1. Notation

Boldface underlined capital characters denote three-way arrays, $\underline{\mathbf{X}}$; boldface capital characters denote matrices, $\mathbf{X}$; boldface lower-case characters represent vectors, $\mathbf{x}$; italic lower-case characters denote scalars, x; and $\mathbf{X}'$ denotes the matrix transpose of $\mathbf{X}$. A list of abbreviations and notations is given in Appendix A.

2.2. Modelling of normal operating conditions (NOC)

The J process variables measured over K time points from a single NOC batch run are stored in matrix $\mathbf{X}_i$ $(J \times K)$. A common way of storing several NOC batch runs is to stack each run in a three-way data array $\underline{\mathbf{X}}$ $(I \times J \times K)$. The different batches are denoted by $i = 1, \ldots, I$ and form the first mode of $\underline{\mathbf{X}}$. The process variables are denoted by $j = 1, \ldots, J$ and form the second mode of $\underline{\mathbf{X}}$. The time points at which measurements are taken are denoted by $k = 1, \ldots, K$ and form the third mode of $\underline{\mathbf{X}}$. Note that it is assumed here that each batch is of the same length although, despite the recipe-driven approach, this is not always the case. One solution to the problem of unequal batch length is to use a 'maturity' variable as an alternative to time for the third mode axis, such as % conversion (Nomikos & MacGregor, 1995; Tates, Louwerse, & Smilde, 1999). Other possibilities are the use of dynamic time warping (Kassidas, MacGregor, & Taylor, 1998) or the PARAFAC2 model (Kiers, Ten Berge, & Bro, 1999; Wise, Gallagher, & Martin, 2001).

Various kinds of process variables are monitored during a batch run, for example temperatures, pressures, flow-rates or spectral data such as infrared absorbances. The automation of modern batch processes facilitates the storing of process variables within a database. The process variables are usually highly correlated and, therefore, a multivariate approach is required. There are often only a few underlying phenomena driving the process. Latent variable methods, such as principal component analysis (PCA) or partial least squares (PLS), aim to extract these underlying factors from the data to form a reduced-dimension subspace, thus making the large amounts of highly correlated data more comprehensible. The number of components, R, to use can be selected using different criteria such as cross-validation (Wold, 1978; Nomikos & MacGregor, 1995).

Prior to building a latent variable model, preprocessing of $\underline{\mathbf{X}}$ is performed, usually in the form of centering and scaling. Centering is performed so that the actual model describes deviations from a mean. The process variables are typically measured in different units and scaling is required to give each process variable equal influence prior to modelling.

2.2.1. Nomikos and MacGregor (NM)

In order to derive a component model, $\underline{\mathbf{X}}$ is first matricised (Kiers, 2000). Nomikos and MacGregor (1994) matricise $\underline{\mathbf{X}}$ in the batch direction resulting in a matrix $\mathbf{X}^{I \times JK}$ $(I \times JK)$ where the second mode is nested within the third mode. Each frontal slice $\mathbf{X}_k$ $(I \times J)$ is placed next to $\mathbf{X}_{k+1}$ $(I \times J)$ as shown in Fig. 1a.

The purpose of monitoring using the NM approach, is to detect deviations from the desired level of operation, which is the mean trajectory of the batches in time. Therefore, in order to model the deviations from the mean trajectory, the mean trajectory is removed by centering the data across the columns of $\mathbf{X}^{I \times JK}$. To account for the difference in measurement units, slab-scaling is used to give each process variable equal variance. That is, all measurements for each process variable, $\mathbf{X}_j$ $(I \times K)$, are scaled to unit sum of squares (Harsman & Lundy, 1984). An alternative to this form of

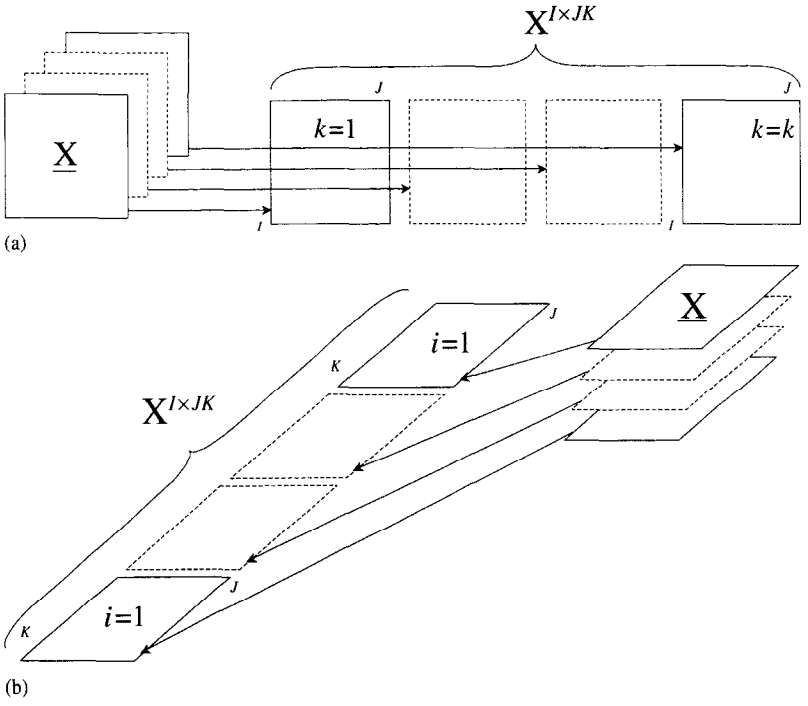

Fig. 1. Matricising of $\underline{\mathbf{X}}$: (a) according to NM; (b) according to WKFH.

scaling is tube-scaling better known as autoscaling, whereby each column of $\mathbf{X}^{I\times JK}$ is divided by its standard deviation. Although this way of scaling is not preferred because it can disturb the multilinear nature of the data (Harsman & Lundy, 1984), there are indications that the difference between slab-scaling and tube-scaling is not very crucial for batch process data (Westerhuis et al., 1999).

The preprocessed matrix $\mathbf{X}^{I\times JK}$ is used to build a PCA model. For the NM approach the model is given by

$$\mathbf{X}^{I\times JK} = \mathbf{TP}' + \mathbf{E}, \tag{1}$$

where $\mathbf{T}$ $(I \times R)$ are the scores, $\mathbf{P}$ $(JK \times R)$ the loadings and $\mathbf{E}$ $(I \times JK)$ the residuals.

2.2.2. Wold, Kettaneh, Fridén, and Holmberg (WKFH)

Wold et al. (1998) matricise $\underline{\mathbf{X}}$ in the process variable direction giving $\mathbf{X}^{KI\times J}$ as shown in Fig. 1b.

Each horizontal slice $\mathbf{X}_{i+1}$ $(J \times K)$ is placed beneath the horizontal slice $\mathbf{X}_i$ $(J \times K)$. Here the third mode is nested within the first mode.

Each column (process variable) of $\mathbf{X}^{KI\times J}$ is centered by subtracting its column mean and scaled by its standard deviation. Note that the centering operation does not remove the mean trajectory of the batches in time. This approach models the correlation between the process variables and uses the score trajectories to monitor the process. The PCA model for WKFH becomes

$$\mathbf{X}^{KI\times J} = \mathbf{TP}' + \mathbf{E}, \tag{2}$$

where $\mathbf{T}$ $(KI \times R)$ are the scores, $\mathbf{P}$ $(J \times R)$ the loadings and $\mathbf{E}$ $(KI \times J)$ the residuals.

2.2.3. Rännar, MacGregor and Wold (RMW)

Rännar et al. (1998), describe an algorithm for adaptive batch process monitoring using hierarchical PCA.

The concept of adaptive process monitoring is summarised in Fig. 2. Here, $\mathbf{X}_k$ $(I \times J)$ represents a frontal slice of $\underline{\mathbf{X}}$. First, a block scores vector, $\mathbf{b}_k$, which represents the local variation at time point k, is computed. Then this block scores vector is placed in the consensus matrix, $\mathbf{B}_k$, along with a super scores vector, $\mathbf{t}_{k-1}$, which summarises previous process variation up to time point $k-1$. The block scores can be weighted by the adaptive parameter, d. Using the consensus matrix, $\mathbf{B}_k$, a new super scores vector, $\mathbf{t}_k$, is computed which now represents the total variation up to time point k. The same procedure is repeated for time point $k+1$ until the end of the batch run.

The adaptive nature of this approach is determined by the adaptive parameter d. If d is high, much weight is given to the current process measurement relative to all previous variation. For $d = \infty$ this algorithm is equivalent to performing a PCA on each separate block, $\mathbf{X}_k$. In general, increasing d will lead to more variance being explained by the model. If d is chosen to be zero, all super scores, $\mathbf{t}_k$, will be the same as the first super score, $\mathbf{t}_1$, and the algorithm is not adaptive at all. An advantage of tuning the adaptive parameter for each block would be that the model can then adapt to different stages in the process. Note that, as yet, there exists no theory on how to tune d. Therefore, in this evaluation two

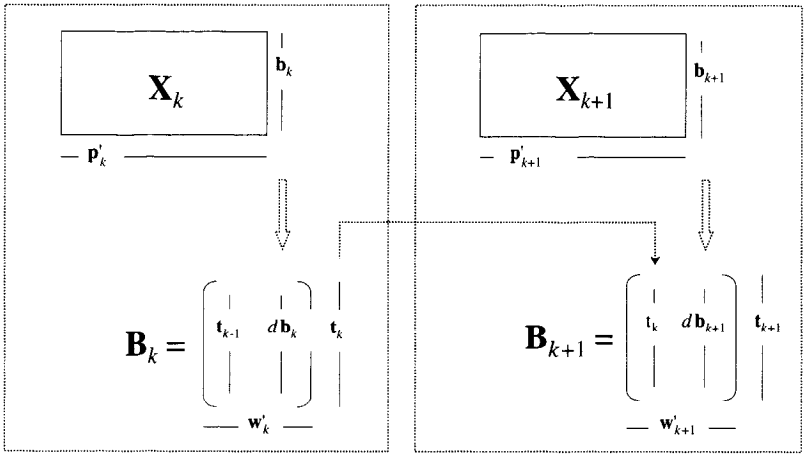

Fig. 2. Adaptive batch monitoring according to RWM.

values of d are chosen, 0.3 and 1.0, which are the same as those used in the example by Rännar et al. (1998).

2.3. On-line monitoring

This section describes how to monitor the evolution of a new batch using the different approaches. All the approaches studied use the concept of projecting an observation vector $\mathbf{x}_k$ onto a space defined by $\mathbf{P}$. This projection results in scores $\mathbf{t}$ and residuals $\mathbf{e}$ which are computed by

$$\mathbf{t}'_k = \mathbf{x}'_k\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}, \tag{3}$$

$$\hat{\mathbf{x}}'_k = \mathbf{t}'_k\mathbf{P}', \tag{4}$$

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k. \tag{5}$$

The content of the observation vector $\mathbf{x}_k$ and the size of the matrix $\mathbf{P}$ depend on the approach chosen as will be explained. This can be seen as an important distinctive feature between the different approaches. Each observation vector is preprocessed before projection using the means and scaling parameters obtained from the NOC data. The features of the different methods will be discussed below.

2.3.1. Nomikos and MacGregor (NM) approach

As stated previously, the matrix $\mathbf{P}$ of the NM approach has dimensions $(JK \times R)$. To compute the scores $\mathbf{t}_k$ and the residuals $\mathbf{e}_k$ according to Eqs. (3)–(5), the observation vector $\mathbf{x}_k$ must have the dimensions of $(JK \times 1)$. This is depicted in Fig. 3.

The observation vector consists of three parts: the past measurements, a current measurement and the unknown future measurements. Take for example a measurement at time interval $k = 10$. The vector $\mathbf{x}_k$ exists of the previous measurements $k = 1,\ldots,9$, the current measurement $(k = 10)$ with size $(J \times 1)$ and the future measurements $k = 11,\ldots,K$. The latter is the basic problem of this approach, that is, the future behaviour of the new batch must be inferred somehow because the batch run is not completed.

Nomikos and MacGregor (1994) suggested two approaches to fill in the unknown part of $\mathbf{x}_k$: the zero deviations approach and the current deviations approach.

2.3.2. Zero deviations approach

The unknown future measurements are assumed to behave at the desired level of operation as defined from the NOC data. The remainder of the batch is assumed to have zero deviations from the mean trajectory.

For a process disturbance, this can be seen as an optimistic scenario. If, for example, a temperature is currently too high, its future evolution is assumed to return directly to the desired level of operation.

2.3.3. Current deviations approach

The current deviations approach assumes that the future measurements will continue to deviate from the desired level of operation at the same level as present at time interval k.

For a process disturbance, this can be seen as a pessimistic scenario, because it is assumed that a process variable which is too high/low will remain so for the rest of the batch run.

2.3.4. Missing data approach

Nomikos and MacGregor (1994) suggested using the missing data approach to overcome the need for assuming the future evolution of a batch.

Using the missing data approach, the future process measurements are regarded as missing values as shown in Fig. 4. Only the part of the loadings $\mathbf{P}$ up until time point k are used to compute the scores and residuals. This part is denoted as $\mathbf{P}_k$. Thus the loadings $\mathbf{P}_k$ are now 'growing' in time and have dimensions $(kJ \times R)$. The vector $\mathbf{x}_k$ necessarily has to be of the size $(kJ \times 1)$. Thus a current measurement

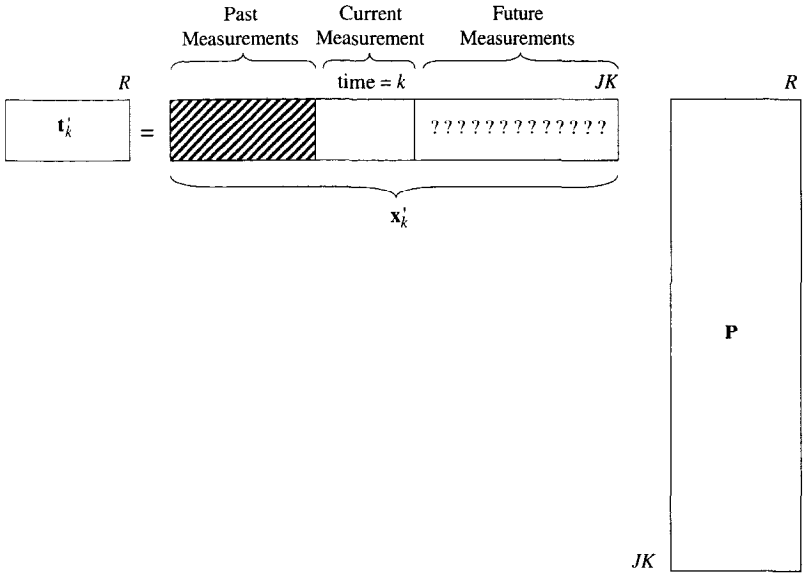

Fig. 3. On-line monitoring according to NM.

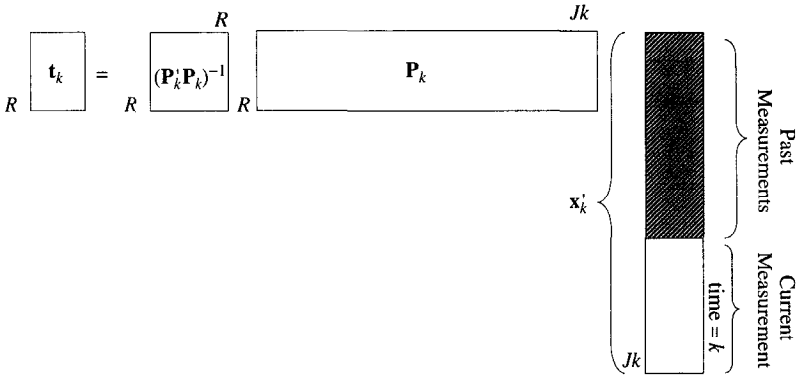

Fig. 4. The missing data approach.

is simply added to the past measurements in $\mathbf{x}_k$. The least squares solution is used to calculate the scores $\mathbf{t}_k$.

$$\mathbf{t}_k = (\mathbf{P}_k'\mathbf{P}_k)^{-1}\mathbf{P}_k'\mathbf{x}_k. \tag{6}$$

Note that only when $k = K$ the term $(\mathbf{P}_k'\mathbf{P}_k)^{-1}$ is equal to identity (because $\mathbf{P}_k$ is then the same as $\mathbf{P}$ which is an orthogonal matrix). A poor monitoring performance can be expected in the early stage of the batch run, because the reduced space on which $\mathbf{x}_k$ is projected is not well defined.

2.3.5. *Rännar, MacGregor and Wold (RMW) approach*

The RMW approach builds local models on each frontal slice $\mathbf{X}_k$, giving K local models. As can be seen from Fig. 2, this results in local loadings $\mathbf{P}_k$ ($J \times R$). The observation $\mathbf{x}_k$ ($J \times 1$) projected on the local reduced space spanned by $\mathbf{P}_k$ corresponds to the measurement at time k. The super scores are computed in a slightly different way to that described in the paper by Rännar et al. (1998):

$$b_{rk} = \mathbf{x}_k^{'(r)}\mathbf{p}_{rk}(\mathbf{p}_{rk}'\mathbf{p}_{rk})^{-1}, \tag{7}$$

$$\mathbf{t}_{rk} = [t_{r(k-1)}d \cdot b_{rk}]\mathbf{w}_{rk}(\mathbf{w}_{rk}'\mathbf{w}_{rk})^{-1}. \tag{8}$$

The difference between both approaches is that in this study the scores are rescaled so as to carry the variance captured by the model, this being necessary for monitoring purposes. Note that the scores for each component must be computed separately. Therefore the observation vector is deflated to compute the score for the rth component by

$$\mathbf{x}_k^{(r+1)} = \mathbf{x}_k^{(r)} - t_{rk}\mathbf{p}_{rk}'. \tag{9}$$

A clear advantage of this approach is the fact that no filling in procedure is needed because only local models are used. However, the use of many different models does make this adaptive monitoring algorithm computationally more complex.

2.3.6. Wold, Kettaneh, Fridén and Holmberg (WKFH) approach

With this approach, PCA is performed on $\mathbf{X}^{KI\times J}$ and to compute the scores $\mathbf{t}$ and residuals $\mathbf{e}$ according to Eqs. (3)–(5), the observation vector $\mathbf{x}_k$ has dimensions $(J \times 1)$. Thus $\mathbf{x}_k$ is the measurement at time interval k which can be projected on $\mathbf{P}$ $(J \times R)$. The advantage of this approach is that no assumptions about future measurements are necessary.

2.4. Deriving the control limits

In general, on-line monitoring of batch processes is done using two types of control-charts: the D-chart to monitor deviation within the model relative to the centerpoint and the squared prediction error (SPE) chart to monitor deviations from the model. Note that it is also possible to construct control charts for individual scores. In this paper attention is paid only to the D- and SPE-charts to avoid an abundance of results.

Statistical limits for the D and SPE statistics are derived from the NOC data. The I batches from the NOC data are projected on the model resulting in I values of the D and SPE at each time interval. Statistical distributions for these test statistics are derived and used to compute control limits using a certain α value.

2.4.1. D-statistic

The Hotelling T^2 statistic is called the D-statistic when a reduced space with R components is used instead of $\mathbf{x}_k$ with J or JK variables. In general, the D-statistic is given by

$$D_k = (\mathbf{t}_{new,k} - \bar{\mathbf{t}}_k)'\mathbf{S}_k^{-1}(\mathbf{t}_{new,k} - \bar{\mathbf{t}}_k)\frac{I(I-R)}{R(I^2-1)} \sim F(R, I-R), \tag{10}$$

where $\mathbf{t}_{new,k}$ $(R \times 1)$ are the scores of the new batch at time interval k and $\bar{\mathbf{t}}_k$ $(R \times 1)$ contains the means of the columns of the score matrix $\mathbf{T}_k$ $(I \times R)$. $\mathbf{S}_k$ $(R \times R)$ is the variance–covariance matrix of $\mathbf{T}_k$ and is an estimation of the variation across the batches. Since the scores are linear combinations of a large number of variables, according to the central limit theorem the scores are independently normally distributed across the batch direction. The test statistic D_k follows an F-distribution with R and I–R degrees of freedom (Tracy, Young, & Mason, 1992).

Notice that for the NM approach, $\bar{\mathbf{t}}_k = 0$ since the data is column mean centered. Furthermore, it is assumed that $\mathbf{S}_k$ is constant in time and is equal to the variance–covariance matrix of the scores calculated when all data is available, $\mathbf{T}_{k=K}$ (Nomikos & MacGregor, 1994).

For the RMW approach, $\bar{\mathbf{t}}_k = 0$ also, but the variance–covariance matrix $\mathbf{S}_k$ is now calculated for each time interval k using the super scores $\mathbf{T}_k$ $(I \times R)$ of the NOC batch runs.

For the WKFH approach, the matrix $\mathbf{T}(KI \times R)$ from Eq. (2) is rearranged to obtain a matrix $\mathbf{T}_k$ $(I \times R)$ for each time interval k, i.e. the I rows corresponding to time interval k are selected from $\mathbf{T}$ and used to form $\mathbf{T}_k$. Here $\bar{\mathbf{t}}_k$ is unequal to zero and $\mathbf{S}_k$ is the variance–covariance matrix of the score matrix $\mathbf{T}_k$.

2.4.2. Squared prediction error (SPE)

For all approaches the SPE_k statistic is computed by

$$\mathrm{SPE}_k = \sum_{j=1}^{J} e_{jk}^2 \sim g\chi_h^2, \tag{11}$$

where e_{jk} is the prediction error of the process variable j at time interval k. The control limits for the SPE are obtained by fitting a weighted χ^2-distribution to the reference distribution obtained from the NOC data at each time point. In Eq. (11), parameter h represents the degrees of freedom and parameter g the weight to account for the magnitude of SPE_k. These two parameters can be estimated in different ways (Box, 1954; Nomikos & MacGregor, 1995; Jackson & Mudholkar, 1979). In this study the estimation is done according to Jackson and Mudholkar (1979), whereby g and h are functions of the eigenvalues of the residual variance–covariance matrix at each time interval k. In order to increase the number of observations available for each estimation, Nomikos and MacGregor (1995) used a time window from $k-2$ to $k+2$. This approach is also used here.

Note that the use of a χ^2-distribution implicitly assumes normality of the errors, which may not always be true in practice. However, as the parameters of the χ^2-distribution used to calculate the SPE limits are obtained directly from the moments of the sampling distribution of the actual NOC data, this approximating distribution is found to work well even in cases where the errors are not normal.

2.5. Summary

To evaluate the different approaches, it is important to realise that the approaches are conceptually different. In the NM approach, a model is constructed which captures the correlation between the process variables and the correlation of the process variables in time. In this approach, each batch run is regarded as an object and has by construction the disadvantage that the future behaviour of the batch run is not known during on-line monitoring.

The RMW approach captures the correlation of the process variables at a certain time interval. Since the local models are constructed in a recursive way, the time behaviour is included implicitly, the extent to which can be tuned using the parameter d.

The WKFH approach models the correlation patterns between the process variables that occur during the NOC batch runs. By rearranging the scores, the dynamic behaviour of the process is implicitly captured by the average score trajectories.

3. Performance indices

This evaluation uses two performance indices to evaluate how well each approach performs in terms of fault detection. The two indices are the overall type I error and the action signal time (AST).

3.1. Overall type I error (false warning)

The first performance index used is the type I error. The type I error is related to the null hypothesis (H_0) that the process is in-control. If H_0 is rejected when it is true, a type I error has occurred. Thus,

$$\alpha_{\text{imposed}} = P(\text{type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true}). \quad (12)$$

To calculate the actual value of α, Nomikos and MacGregor (1994) used Eq. (13) to calculate the probability of an overall type I error.

$$\alpha_{\text{actual}} = P(\text{overall type I error}) = \frac{\sum \text{false warnings}}{IK}. \quad (13)$$

In the present paper the following procedure is used to calculate α_{actual}. One batch is removed from the NOC data set and a model is built on the remaining $I-1$ batches. Control limits for the D-chart and SPE-chart are derived from this model. Then the left out NOC batch is monitored. As this batch is from the NOC data, it is assumed to be in control over the entire trajectory and, therefore, a crossing of a control limit is considered to be a false warning. The procedure is repeated for every NOC batch and the total number of false warnings for the NOC data is calculated. This leave-one-out approach is used so as to maintain independence between the monitored batch and the NOC model.

3.2. Action signal time (AST) and type II error

To asses the type II error, it is common in statistical process control to use average run length (ARL) curves to study the efficiency of a control chart. However, such a rigorous analysis requires a large amount of data or a theoretical treatment. Both alternatives are not feasible in batch monitoring approaches. Therefore, in this study the action signal time (AST) is used to evaluate the detection power of a method and serves as an indication of the type II error. The AST is defined as the time between the introduction of an error and the out-of-control signal, as can be seen in Fig. 5. The signal is said to be out-of-control only if at least three consecutive points are outside the control limit. The choice of the number of observations that define an out-of-control signal is according to practice, when action is only undertaken after some consecutive points have been out of control. In this study, different batch processes are studied. Each process has a different batch run length. To compare the AST of the different processes, the relative AST is used. The relative AST (RAST) is the ratio between the AST and the total erroneous time trajectory (ET) of the batch run. The ET is the time period starting from the introduction of a disturbance until the end of the batch run (see Fig. 5). There are several possibilities of defining the RAST with each definition having its advantages and disadvantages. This will be the topic of a subsequent paper. The current alternative appeared to be reasonable.

The times of introduction of the errors for the different processes are given in Table 1. Out-of-control signals before the fault occurred are ignored. For data sets 1 and 5 prior knowledge of the exact time of disturbance is available. For the remaining data sets the time of introduction of the error is obtained after a thorough data analysis, using process knowledge, and discussion with the process engineers.

Note that there is distinction between an out-of-control signal, defined as three points above the control limit, and a warning, defined as one point above the control limit.

In ordinary hypothesis testing there is a clear relationship between the type II error and the type I error: by choosing a lower α, the type I error decreases and the type II error increases. Due to the distinction between an out-of-control signal and a warning, such a direct relationship is not present anymore between the type I error and the AST. There will remain a tendency, however, for the AST to become larger when choosing a lower α, which also decreases the type I error.

3.3. Summary

The emphasis of the comparison lies on the detection of erroneous batches. Therefore, the AST is the most relevant index of the two. The overall type I error is mainly a tool for confirmation if the control limits are at an acceptable level. If the type I error is not satisfactory, the user can raise or lower the limits until an acceptable level of α is reached.

4. Description of the data

In order to evaluate the different approaches for on-line monitoring, the approaches are applied to five data sets. Each data set has different specific features as will be explained in the following. The NOC data sets are used to compute the overall type I error and disturbed batches are used to compute the AST. The time of disturbance occurrence and the dimensions of the three-way data array for the studied data sets are given in Table 1.

4.1. Data set 1

Data set 1 consists of a simulated emulsion co-polymerisation of styrene–butadiene. This process is well described in Nomikos and MacGregor (1994). This data set consists of 50 NOC batches for which 9 process variables are measured during 200 time intervals. The following process variables are measured: two flow-rates, four temperatures, density, conversion and rate of energy release. Two disturbed batches

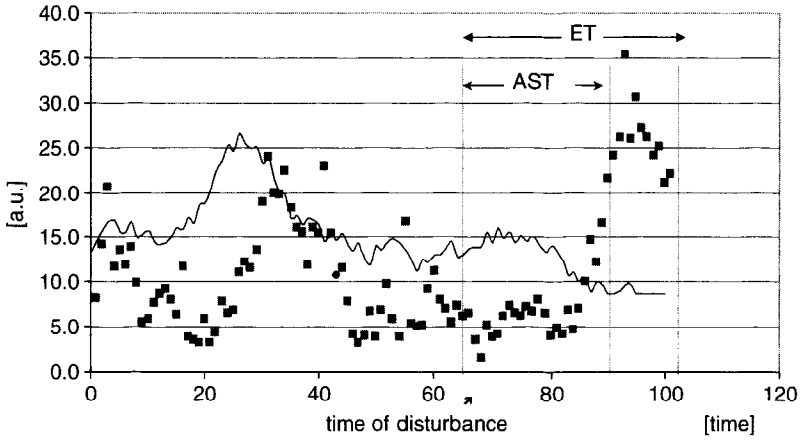

Fig. 5. The action signal time (AST).

Table 1
Description of the investigated data sets and disturbed batches

	I	J	K	Disturbed batches	Time of disturbance
Data set 1	50	9	200	Batch 106[a]	1
				Batch 99[a]	100
Data set 2	36	10	100	Batch 50	56
Data set 3	46	8	116	Batch 41	58
				Batch 46	94
				Batch 47	94
Data set 4	67	15	58	Batch 14	1
				Batch 17	1
				Batch 24	1
				Batch 49	46
Data set 5	27	201	271	Batch 2[a]	138

[a]Exact time of disturbance is known.

are available to calculate the AST. The first has an initial organic impurity contamination in the butadiene feed. The second has a similar, but larger, contamination halfway the batch run.

4.2. Data set 2

Data set 2 is an industrial (Du Pont) two-stage polymerisation process with 36 selected NOC batches for which 10 process variables are measured during 100 time intervals. One disturbed batch is available to calculate the performance indices, which was known to have a poor end-product quality. This data set is described in Nomikos and MacGregor 1995.

4.3. Data set 3

Data set 3 consists of an industrial (DuPont) polymerisation in an autoclave with 47 NOC batch runs. This data set is also referred to by Kosanovich, Dahl, and Piovoso (1996). The polymer is produced through five stages during these stages the process is monitored by 8 process variables (six temperatures and two pressures) during 116 time intervals. Prior to model building, the raw data is linearly interpolated to align the NOC batches, this is described in Boqué and Smilde (1999). Three disturbed batches are available to calculate the AST. A univariate analysis of batch 41 clearly shows disturbances for two pressure and four temperature measurements halfway the batch run. This is probably due to an upset in the first stage of the process. For batch 46 and batch 47, univariate analysis shows disturbances at the end of the batch run for one pressure and five temperature measurements.

4.4. Data set 4

This data set is from an industrial (Shell) suspension polymerisation of polyvinylchloride (PVC) in a batch reactor. The data set consists of 67 batch runs obtained under NOC and 15 process variables were measured for a period of 58 time intervals. The NOC batches are aligned using the conversion as a maturity variable. A more detailed

description is given by Tates, Louwerse, Smilde, Koot, and Berndt (1999). The process variables consist of: eight temperatures, condenser duty, agitator speed and power supply, two mass streams of cooling water, batch reactor level and pressure. Three disturbed batches are available to calculate the AST. Analysis of univariate plots of the process variables is used to study the disturbed batches. Analysis of batch 17 shows that the temperature and the amount of refrigerant water to the jacket are disturbed at the end of the batch run. Batch 24 shows disturbances for the batch level and the power supply to the agitator. Batch 49 reveals disturbances for the temperature and amount of cooling water, and a temperature disturbance within the reactor.

4.5. Data set 5

Data set 5 consists of a laboratory spectroscopic batch process of a two-step bio-chemical conversion reaction described in Bijlsma, Louwerse, and Smilde (1999) and can be obtained from the Process Analysis & Chemometrics website of the University of Amsterdam. The data set consists of 27 NOC batch runs measured at wavelengths 300–500 nm during 271 time intervals. One batch run with a pH disturbance at time interval 138 is available to compute the AST.

5. Results and discussion

Prior to analysis, the number of principal components for each method is determined using cross validation (Wold, 1978; Nomikos & MacGregor, 1995). It is found that a three-component model is enough to explain common-cause variation for all approaches and also for all data sets. This is in agreement with the number chosen by the authors for data sets 1 and 5.

5.1. AST

The results for the average AST for the control charts with a 95% confidence limit are given in Fig. 6. These results were obtained as follows. First, each of the disturbed batches (see Table 1) is monitored using the SPE-chart and the D-chart for each approach. This results in 2 AST values for each disturbed batch and 20 AST values for each approach. Although the fault could appear first either in the SPE-chart or in the D-chart, from a practical point of view the control chart, which gives the shortest AST, is of interest. Therefore using the shortest AST of the two charts will result in 10 AST values (one for each disturbed batch) for each approach. As stated earlier, the relative AST is used for comparing the AST of different processes. To obtain an overall picture of the RAST, the RAST is averaged out over the data sets for each monitoring approach. A similar procedure is followed for the control charts with 99% confidence limits.

In Fig. 6, the results are divided in three groups. The first group is formed by the missing data approach and the RMW approach ($d = 0.3$), for which the shortest average AST is seen. For data sets 1 and 2, Nomikos and MacGregor (1995) also reported a superior performance for the missing data approach in terms of AST. As described earlier, however, for the first 10% of the batch run, the D-chart for the missing data approach is not found to be useful as not enough data is available to obtain a reliable score **t** and, hence, a reliable D. This leads to many false warnings in the beginning of the batch run. There are three batches with a disturbance immediately from the beginning. As will become apparent in the next section, both the missing data and the RMW ($d = 0.3$) approaches give high overall type I errors and so although a fault seems to be detected quickly by the D-chart, this could be due to the shortcomings of these approaches.

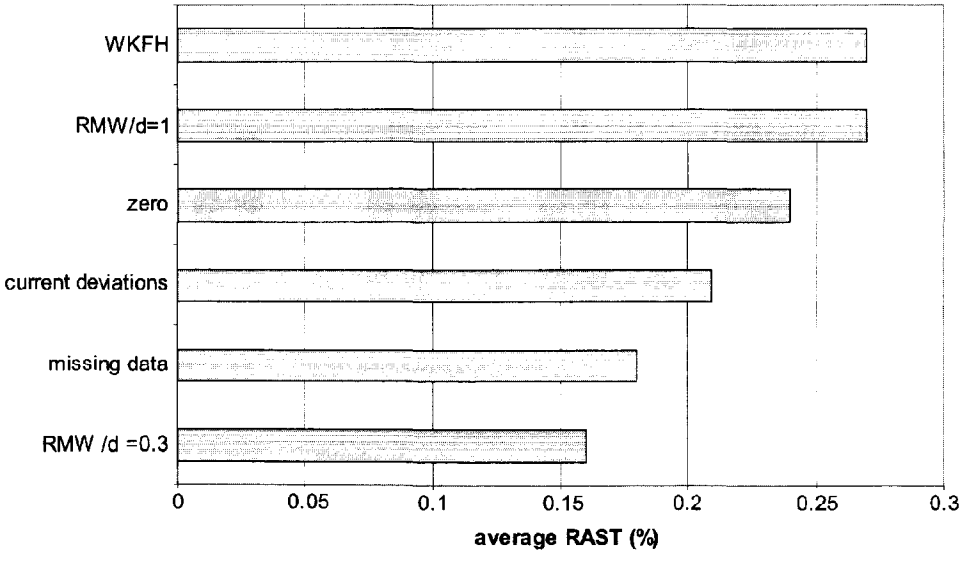

Fig. 6. The overall AST found for 95% confidence limits.

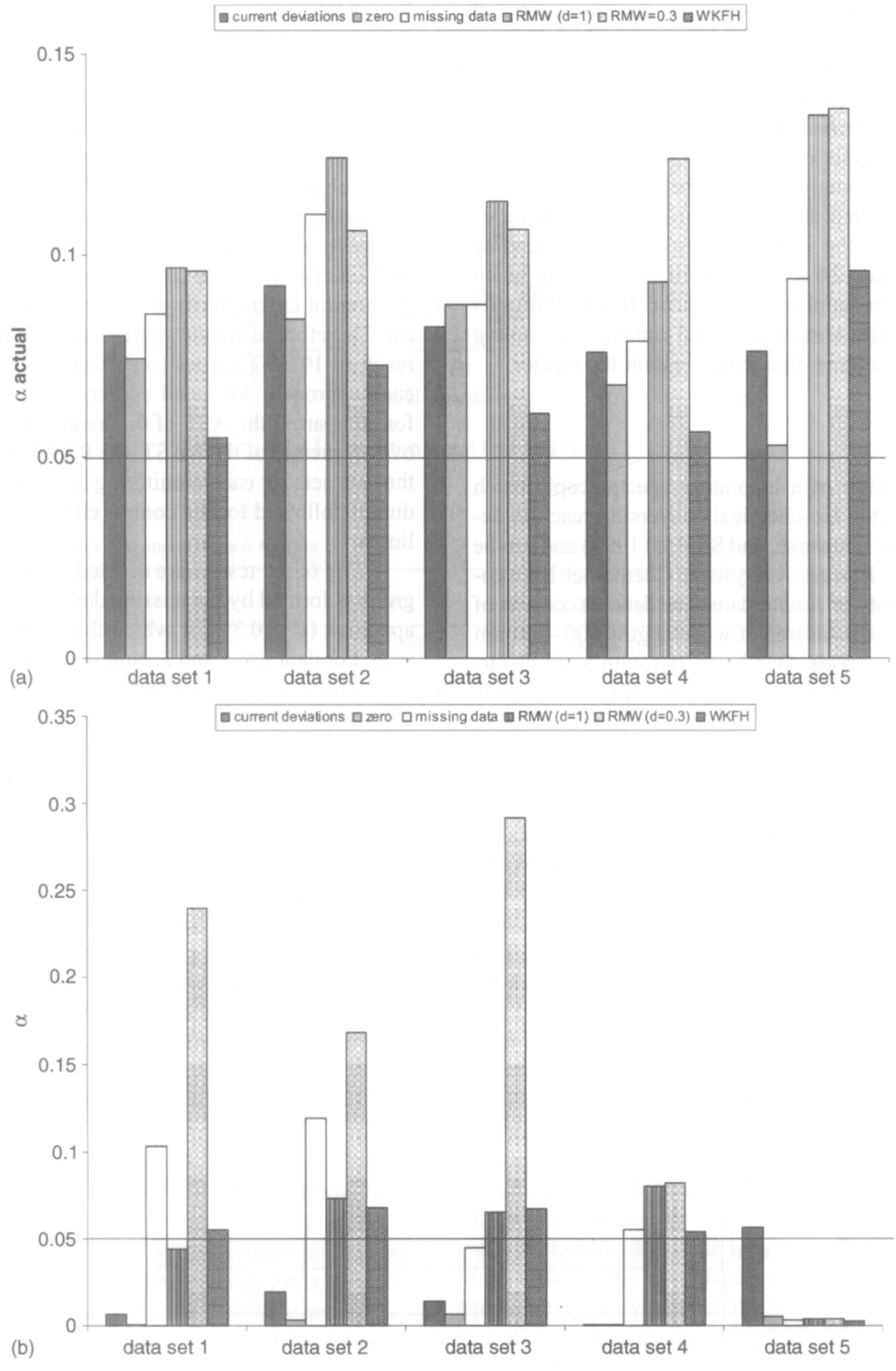

Fig. 7. Results for the overall type I error for: (a) the SPE using 95% confidence limits; (b) the D-statistic using 95% confidence limits.

The second group is formed by the NM current deviations and NM zero approach. The third group is formed by the WKFH and RMW ($d = 1$). Both groups are generally found to perform well with a small difference in the AST. From Fig. 6 can be seen that the third group has a slightly larger AST.

A similar pattern to that seen in Fig. 6 is found for the control charts with 99% confidence limits and the results are not given here. The differences between the average AST for the 95% and 99% limits are small. This implies that the test statistic crosses both the 95% and 99% confidence limit at the moment of detection. Therefore, the detection of a fault

is not very sensitive to the exact confidence limit as long as reasonable values for these limits are chosen.

Note that using the NM approach, faults are generally detected in the SPE-chart. Only if the disturbance occurred at the beginning of a batch, then the D-chart sometimes signals faster than the SPE-chart. Contrary, in the WKFH approach most of the faults are detected in the D-chart. This is due to the different construction of the models.

5.2. Overall type I error

In general, it is expected that α_{actual} is slightly higher than $\alpha_{imposed}$ for the SPE. This is because the SPE limits are approximated from the NOC residuals and it is assumed that the residuals from a new batch come from the same population. In fact, the residuals from a new batch will generally be slightly higher than the NOC residuals since the NOC data is modelled in a least-squares way in which the residuals are minimised. This is reflected in the overall type I error. The D-statistic is approximated by a F-distribution which takes into account the fact that the new batch is independent from the NOC data (Tracy et al., 1992).

The α_{actual} values for the overall type I error for the SPE using 95% confidence limits are given in Fig. 7a.

The solid line represents the value for $\alpha_{imposed}$. Comments on Fig. 7a. are as follows:

- It is clear that the value for α_{actual} is indeed slightly higher than $\alpha_{imposed}$. This is also in agreement with the results found by Nomikos and MacGregor (1995). It seems to be a general conclusion for these approaches, the data sets used being dissimilar in nature.
- The RMW approach ($d=0.3$ and 1) has the highest value for α_{actual} for all data sets. For this approach, where local models at each time interval k are used, the NOC data is modelled relatively well and the NOC residuals are correspondingly low. Thus, the difference between the NOC residuals and those for a new batch are higher than for other approaches.
- The patterns obtained using 99% confidence limits for the SPE lead to the same conclusions as when 95% limits were used.

The results for the overall type I error for the D-statistic using 95% confidence limits are given in Fig. 7b. Comments on Fig. 7b. are as follows:

- The spectral data set (data set 5) gives different results from the engineering data sets.
- The α_{actual} values of the RMW approach with $d=0.3$ are far too high. The low value for the adaptive parameter d implies that a relatively high importance is given to the process history for that batch. This may mean that a false warning, such as a spike in the data, can continue to influence the D-chart over a length of time and, therefore, high values for the overall type I error are seen. The weighting parameter d can be tuned in such a way that $\alpha_{imposed}$ equals α_{actual}, which may provide a method for determining the correct value to use for d.
- The WKFH approach and the RMW approach with $d=1$ give reasonable values for α_{actual}.
- In general, the filling-in approaches, NM current deviations and NM zero approach have values for α_{actual} which are too low.
- The results found for the missing data approach must be interpreted with care. It has already been mentioned that the D-chart at the beginning of the batch is unreliable, resulting in too many false warnings and, therefore, a high overall type I error.
- The patterns obtained using 99% confidence limits for the D-statistic lead to the same conclusions as when 95% limits were used.

6. Conclusions and recommendations

In this paper an evaluation is made of five different approaches for monitoring batch processes. The theory behind each approach is explained together with the specific features. The most important distinction between the approaches is the way in which the correlation structures in the data, both between process variables and in time, are modelled, sometimes leading to a need to assume the future behaviour of a new batch during on-line monitoring. To evaluate the different approaches, performance indices were used and applied to five different data sets. Some recommendations on using the methods are given as follows:

- The results of the evaluation presented here suggested that the NM current deviations, NM zero- and the WKF approach on average give the lowest reliable AST.
- The control charts for the NM missing data approach are found to be particularly unreliable at the start of the batch and should, perhaps, be ignored until data from enough time points are available.
- The method presented in this paper for estimating the overall type I error could be used in practice to correctly set the confidence limits depending upon the monitoring approach being used. This also provides a method for determining the correct value to use for the adaptive parameter d in the RMW approach, i.e. d can be tuned in such a way that $\alpha_{imposed}$ equals α_{actual}.
- The detection of a fault in the control charts is not very sensitive to the exact confidence limit as long as reasonable values for these limits are chosen.

Acknowledgements

This study is supported by the Council for Chemical Sciences of the Netherlands Organization for Scientific

Research (NWO-CW) with financial aid from the Netherlands Technology Foundation (STW). The companies DuPont and Shell are thanked for making their process data available for this study.

Appendix A.

NM	approach by Nomikos and MacGregor (1994)
NOC	normal operation conditions
RMW	approach by Rännar et al. (1998)
WKFH	approach by Wold et al. (1998)
SPE	squared prediction error
AST	action signal time
$\mathbf{B}$	consensus matrix
$\mathbf{B}_k$	consensus matrix at time interval k
$\mathbf{b}$	block scores vector
$\mathbf{b}_k$	block scores vector at time interval k
d	adaptive parameter for RMW approach
g	weighting factor for the χ^2 distribution
h	degrees of freedom for the χ^2 distribution
i	ith batch
I	number of batches
j	jth process variable
J	number of process variables
k	kth time interval
K	number of time intervals
$\mathbf{P}$	loading matrix
$\mathbf{P}_k$	loading matrix at time interval k
$\mathbf{p}$	loading vector
$\mathbf{p}_k$	loading vector at time interval k
R	number of components
r	rth component
$\mathbf{S}$	variance–covariance matrix
$\mathbf{S}_k$	variance–covariance matrix at time interval k
$\mathbf{T}$	score matrix
$\mathbf{t}$	score vector
$\mathbf{T}_k$	score matrix at time interval k
$\mathbf{t}_k$	score vector at time interval k
$\mathbf{t}_{new,k}$	score vector of a new batch at time interval k
$\mathbf{X}$	two-way data matrix
$\underline{\mathbf{X}}$	three-way data array
$\mathbf{X}_k$	frontal slice
$\mathbf{x}_k$	observation vector at time interval k
α	confidence level

References

Bijlsma, S., Louwerse, D. J., & Smilde, A. K. (1999). Estimating rate constants and pure UV-vis spectra of a two-step reaction using trilinear models. *Journal of Chemometrics, 13*, 311–329.

Boqué, R., & Smilde, A. K. (1999). Monitoring and diagnosing batch processes with multiway covariates regression models. *A.I.Ch.E. Journal, 45*, 1503–1520.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics, 25*, 290–302.

Dahl, S. K., Piovoso, M. J., & Kosanovich, K. A. (1999). Translating third-order data analysis methods to chemical batch processes. *Chemometrics and Intelligent Laboratory Systems, 46*, 161–180.

Harsman, R. A., & Lundy, M. E. (1984). Data preprocessing and the extended PARAFAC model. In H.G. Law, C.W. Snyder, J.A. Hatti, & R.P. McDonald, (Eds.), *Research methods for multimode data analysis* (p. 216). New York: Praeger Publishers.

Jackson, J. E., & Mudholkar, G. S. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics, 21*(3), 341–349.

Kassidas, A., MacGregor, J. F., & Taylor, P. A. (1998). Synchronization of batch trajectories using dynamic time warping. *A.I.Ch.E. Journal, 44*, 864–875.

Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics, 14*, 105–122.

Kiers, H. A. L., Ten Berge, J. M. F., & Bro, R. (1999). PARAFAC2 Part I. A direct fitting algorithm for the PARAFAC2 model. *Journal of Chemometrics, 13*, 275–294.

Kosanovich, K. A., Dahl, K. S., & Piovoso, M. J. (1996). Improved process understanding using multiway principal component analysis. *Industrial Engineering Chemistry Research, 35*, 138–146.

Louwerse, D. J., & Smilde, A. K. (1999). Mutivariate statistical process control of batch processes based on three-way models. *Chemical Engineering Science, 55*, 1225–1235.

MacGregor, J. F., & Nomikos, P. (1992). Monitoring batch processes. In Reklaitis, Rippin, Hortasco, & Sunol (Eds.), *Batch processing systems engineering: Current status and future directions, NATO ASI Series F (vol.* 143) (pp. 242–258). Heidelberg: Springer.

Martin, E. B., & Morris, A. J. (1996). Non-parametric confidence bounds for process performance monitoring charts. *Journal of Process Control, 6*, 349–358.

Martin, E. B., Morris, A. J., Papazoglou, M. C., & Kiparisassides, C. (1996). Batch process monitoring for consistent production. *Computers & Chemical Engineering, 20*, 599–604.

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

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

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

Tates, A. A., Louwerse, D. J., Smilde, A. K., Koot, G. L. M., & Berndt, H. (1999). Monitoring a PVC batch process with multivariate statistical process control charts. *Industrial & Engineering Research, 38*, 4769–4776.

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

Westerhuis, J. A., Kourti, T., & MacGregor, J. F. (1999). Comparing alternative approaches for multivariate statistical analysis of batch process data. *Journal of Chemometrics, 13*, 397–413.

Wise, B. M., Gallagher, N. B., Butler, S. W., White Jr., D. D., & Barna, G. G. (1999). A comparison of principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *Journal of Chemometrics, 13*, 379–396.

Wise, B. M., Gallagher, N. B., & Martin, E. B. (2001). Application of PARAFAC2 to fault detection and diagnosis in semiconductor etch. *Journal of Chemometrics, 15*, 285–298.

Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics, 4*, 397–405.

Wold, S., Kettaneh, N., Friden, H., & Holmberg, A. (1998). Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics and Intelligent Laboratory Systems, 44*, 331–340.