Statistical Batch Process Monitoring Using Gray Models

E. N. M. Van sprang, H.-J. Ramaker, J. A. Westerhuis, and A. K. Smilde
Process Analysis & Chemometrics, Dept. of Chemical Engineering, University of Amsterdam, 1018 WV Amsterdam, The Netherlands

D. Wienke
DSM Research, Analysis for Products & Processes, 6160 MD Geleen, The Netherlands

DOI 10.1002/aic.10348
Published online in Wiley InterScience (www.interscience.wiley.com).

A complete strategy for monitoring industrial batches processes using gray models is presented including fault detection and fault diagnosis tools. The use of gray models is a novel concept in batch process modeling and monitoring. A gray model is a hybrid model, intermediate between hard (white) process models and soft (black) models, combining the advantages of both approaches. The principles of gray models are explained and it is shown how these models can be constructed. For this purpose an industrial batch process is available that is spectroscopically monitored, and an explanation is provided as to how the spectroscopic measurements are combined with prior process knowledge. To show the versatility of the strategy, two types of gray models are constructed and used for statistical batch process monitoring. The two models are compared and validated for both on-line monitoring and post-batch analysis. For the latter, the batch consistency number (BCN) is introduced to have a fast and simple post-batch analysis. The results show how these models help to detect and diagnose process upsets. The use of gray models for batch process monitoring results in a fast detection of process upsets and a good fault diagnosis.
© 2005 American Institute of Chemical Engineers *AIChE J,* 51: 931–945, 2005
Keywords: batch processes, statistical batch process monitoring, MSPC, control charts, gray models

Introduction

Batch processes are a significant part of today's production in the food and (bio-)chemical industries. Batch processes are recipe driven and produce low-volume and high-value products. Because of control actions and for safety reasons, it is common to carry out process measurements such as temperature, pressure, or flow measurements. A disadvantage of these process measurements is that they do not give direct insight in the process chemistry. To have chemical insight during the batch run, spectroscopic measurements can be performed.[1,2] For a quick detection of process upsets or unusual process behavior it is common to use the process measurements for monitoring purposes. In the 1990s Nomikos and MacGregor[3] introduced the concept of multivariate statistical monitoring of batch processes. Since the introduction of the method, many extensions and applications have been reported in the literature.[4,5]

In principle, batch process monitoring is divided into three phases according to the ITA trajectory:

- *Initial phase*. This phase consists of collecting a set of historical batch runs, which are monitored on-line and operated under normal operating conditions (NOC).
- *Training phase*. In this phase, suspicious batch runs or batch runs that are not considered to be obtained under normal operating conditions (NOC) are removed from the historical data set. When a representative historical data set of NOC

Correspondence concerning this article should be addressed to A. K. Smilde at asmilde@science.uva.nl.

© 2005 American Institute of Chemical Engineers

batches is obtained, a model is built to demonstrate the common cause variation. With the obtained process model, test statistics are computed and statistical confidence limits are derived to construct control charts.

- *Application phase*. In this phase a new batch is monitored by means of control charts derived in the *Training* phase.

A batch process can be monitored either on-line or post-batch. With on-line monitoring, the aim is to have a fast detection of process upsets. Then, if an alarm is given in the control charts, the model is used to locate and diagnose the cause of the alarm. In post-batch analysis, the aim is to study the general behavior of batch runs and obtain insight in the differences and similarities between batches.

There are a variety of possible models to capture the process behavior, such as principal-component analysis (PCA), multivariate linear regression (MLR), or projection to latent structures (PLS). Such models are known as empirical or black box models. These models fit the data properly, but cannot be generalized to different situations and do not always generate a good process insight. Another type of models are white models that, although they are based on first principles and do provide process insight, do not fit the process data that well and are not able to explain batch-to-batch differences.

A combination of the previous black box and white models are gray (hybrid) models. Gray models try to combine the advantages of the black box and white models—that is, a good model interpretation and a good data fit. A gray model is constructed by including known external information as white knowledge, such as fundamental knowledge of the process at laboratory scale or trajectories of process variables, into a black model. The gray model fits the process data using the external information as constraints on the model parameters. In this manner fundamental knowledge and empirical models are combined. A more detailed discussion of gray models can be found in Gurden et al.[6] The concept of combining white information with black models has been reported in several publications.[7,8] The gray concept implicitly divides the data into a part consisting of variation arising from known causes (white part) and into a part consisting of systematic variation stemming from unknown causes (black part). This makes the model very suited to search for unknown phenomena in the data.[9] Other examples are the estimation of kinetic constants using spectroscopic process data and the knowledge of reaction kinetics or improving curve resolution models by combining hard kinetic models with soft models.[8-10]

In the following, two types of gray models are presented and validated for batch process monitoring. The models are validated for both on-line monitoring and post-batch analysis. The first type is a gray Tucker1 model and the second type is a gray Tucker3 model. The *gray Tucker1* model is a relatively simple model and can directly be used for on-line batch monitoring. If the gray Tucker1 model is used in post-batch analysis a rearrangement of the data is needed. The second type, a *gray Tucker3* model, is more complex compared to the gray Tucker1 model and cannot be used directly for on-line batch process monitoring. However, this model is directly applicable for post-batch analysis.

Moreover, the models are used to capture the process behavior and to derive control charts for process monitoring. The performance of the models is validated for on-line detection and diagnosis and also for post-batch analysis. For this purpose, the batch consistency number (*BCN*) is introduced as a simple analysis to study the differences and similarities between batches.

The outline of this article is as follows: after the introduction, the concepts and construction of gray models are discussed together with the derivation of the monitoring statistics in the theory section. Next, in the experimental section, the structure of the process data is discussed and the construction of gray models is explained. Furthermore, it is shown how the models are used for statistical batch process monitoring. After that, the results are presented and discussed and, finally, conclusions are proposed.

Theory

Notation

Scalars are written as lowercase italics (x) and vectors as lowercase bold characters ($\mathbf{x}$). Uppercase bold characters ($\mathbf{X}$) represent matrices and three-way arrays are given as underlined uppercase bold characters ($\underline{\mathbf{X}}$). The symbol "$^{\mathbf{T}}$" refers to the transpose.

Structure of the data

A single batch run that is monitored on-line at K intervals is represented by a matrix $\mathbf{X}$ ($K \times J$), where J is the number of process variables. In the case of a series of multiple batch runs, the individual batch runs can be stacked on top of each other and form the three-way array $\underline{\mathbf{X}}$ ($I \times J \times K$), where I is the number of batch runs.

Black models

There are two sensible directions to matricize a three-way array.[11] The first direction is in the batch direction; that is, the three-way array will be matricized in a matrix $\mathbf{X}_{I \times JK}$. The second direction is in the process variable direction, which results in a matrix $\mathbf{X}_{KI \times J}$. The matricizing direction will have an influence on the interpretation of the model parameters as will become clear in the following. Both matrices can be modeled using a Tucker1 model. Assuming that the three-way array is matricized in the process variable direction, a single element in $\mathbf{X}_{KI \times J}$ is modeled by

$$x_{ki,j} = \sum_{r=1}^{R} c_{ki,r} b_{jr} + e_{ki,j} \tag{1}$$

or in matrix notation,

$$\mathbf{X}_{KI \times J} = \mathbf{C}\mathbf{B}^{\mathbf{T}} + \mathbf{E} \tag{2}$$

where matrix $\mathbf{C}$ ($KI \times R$) consists of time-related profiles, matrix $\mathbf{B}$ ($J \times R$) consists of information regarding the correlation between the process variables, and matrix $\mathbf{E}$ is the residual matrix. If the three-way array is matricized in the batch direction, a single element in $\mathbf{X}_{I \times JK}$ is modeled by

$$x_{i,jk} = \sum_{r=1}^{R} a_{i,r} p_{r,jk} + e_{i,jk} \tag{3}$$

or in matrix notation,

$$\mathbf{X}_{I \times JK} = \mathbf{A}\mathbf{P}^{\mathbf{T}} + \mathbf{E} \tag{4}$$

where matrix $\mathbf{A}$ ($I \times R$) describes the differences between the batches, matrix $\mathbf{P}$ ($JK \times R$) consists of information regarding the systematic variation and correlation between the process variables, and matrix $\mathbf{E}$ is the residual matrix. As shown, the matricizing direction determines the content of the model parameters and therefore also its interpretation.

The second option is to model the three-way array with a Tucker3 model. The Tucker3 model decomposes the three-way array $\underline{\mathbf{X}}$ into the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}$. A single element in the three-way array is given by

$$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \tag{5}$$

or in matrix notation,

$$\mathbf{X}_{I \times JK} = \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^{\mathbf{T}} + \mathbf{E} \tag{6}$$

where matrix $\mathbf{A}$ ($I \times P$) captures the systematic differences between the batches. Matrix $\mathbf{G}$ ($P \times QR$) consists of the interaction parameters between the three reduced components. Matrix $\mathbf{C}$ ($K \times R$) consists of time-related profiles and matrix $\mathbf{B}$ ($J \times Q$) consists of information concerning the correlation between the process variables.

The relation between the Tucker3 and the Tucker1 models is explained in the following. A Tucker3 model tries to compress the data in each of the three directions ($I \times J \times K$) into a reduced number of latent variables P, Q, and R. P is the number of latent variables in the batch direction (I), Q is the number of latent variables in the process variable direction (J), and R is the number of latent variables in the time direction (K). It is also possible to compress only two directions and then the model becomes

$$\mathbf{X}_{I \times JK} = \mathbf{A}\mathbf{G}(\mathbf{I} \otimes \mathbf{B})^{\mathbf{T}} + \mathbf{E} \tag{7}$$

where $\mathbf{G}$ ($P \times QJK$) and $\mathbf{I}$ ($Q \times Q$) constitute the identity matrix. This is called a Tucker2 model.[12] Of course, it is also possible to compress in only one direction and the model becomes

$$\mathbf{X}_{I \times JK} = \mathbf{A}\mathbf{G} + \mathbf{E} \tag{8}$$

with $\mathbf{A}$ ($I \times P$) and $\mathbf{G}$ ($P \times JK$). A comparison of Eqs. 8 and 4 shows that both models are identical. Both models model the matricized three-way array with a bilinear model. The model in Eq. 4 is called a Tucker1 model. In fact, a Tucker1 model with the proper restrictions is equal to a Tucker3 model.

In short, both models are well able to describe the variation in the data. The difference between the two models is that the Tucker3 model is able to reduce the data with a different number of latent variables in each of the three directions (batch, variable, time). Interactions between the three reduced directions are given in the core array $\mathbf{G}$. In contrast, the Tucker1 model reduces the data in only one direction.

Because batch processes are recipe driven, it is assumed that the batches are very similar. However, in practice, batches will have slightly different run lengths, which are reflected in a different number (K_i) of sampled spectra. There are several ways of synchronizing the batches. Batches can be synchronized using maturity variables,[13] interpolation,[14] or dynamic time warping (DTW).[15] In this study the differences in batch run length are dealt with using a DTW algorithm, which is adapted to handle spectroscopic data.[16] For simplicity, the synchronization is done in a post-batch manner, although it is also possible to apply this technique on-line. More suggestions can be found in Kourti.[17]

Gray models

Gray models are hybrid models. That is, a part of the model consists of known systematic variation and is combined with an unknown empirical part. The known or "white" part of the model is constrained to reflect available external process information such as process kinetics. The empirical or "black" part of the model is unconstrained and fits further systematic variation that is not explained by the white part. Together, the white and black parts form the gray model.

Unsystematic variation in the data is not modeled and forms the residuals. The model parameters are calculated simultaneously with the aim of maximizing the total amount of variation explained in the data. The separation between the white and black parts of the model makes it possible to isolate the unknown sources of systematic variation and to study the cause of this unknown variation.

Creating gray models

The black models are made gray by estimating the models in a single least-squares optimization with multiple constraints applied according to prior external process information ("white" knowledge). First, a general discussion is given of the Tucker1 model in Eq. 1. In Figure 1, assume near-infrared (NIR) process spectra in matrix $\mathbf{X}$. Matrix $\mathbf{X}$ is decomposed in the matrices $\mathbf{C}_w$, $\mathbf{B}_w$, $\mathbf{C}_b$, $\mathbf{B}_b$, and the residual matrix $\mathbf{E}$. Data related to time and concentration profiles are represented in the matrices $\mathbf{C}_w$ and $\mathbf{C}_b$ and spectral information is captured in the matrices $\mathbf{B}_w$ and $\mathbf{B}_b$. The first step in creating a gray model is collecting the available white knowledge. Having NIR process spectra the following white information can be conceived as follows: The spectral response is a linear sum of the contributions of the individual compounds according to the Lambert–Beer law.

Sometimes the reaction kinetics are known and can be imposed on matrix $\mathbf{C}_w$. If the reaction kinetics are not known, it is possible to impose physically sensible restrictions such as nonnegativity or unimodality on the profiles in $\mathbf{C}$, given that concentrations can never be negative. If pure compound spectra are available they can be imposed on matrix $\mathbf{B}_w$. If no pure compound spectra are available, restrictions can be imposed such as nonnegativity.

The model implicitly imposes the Lambert–Beer law if a Tucker1 model is used because this model assumes linear

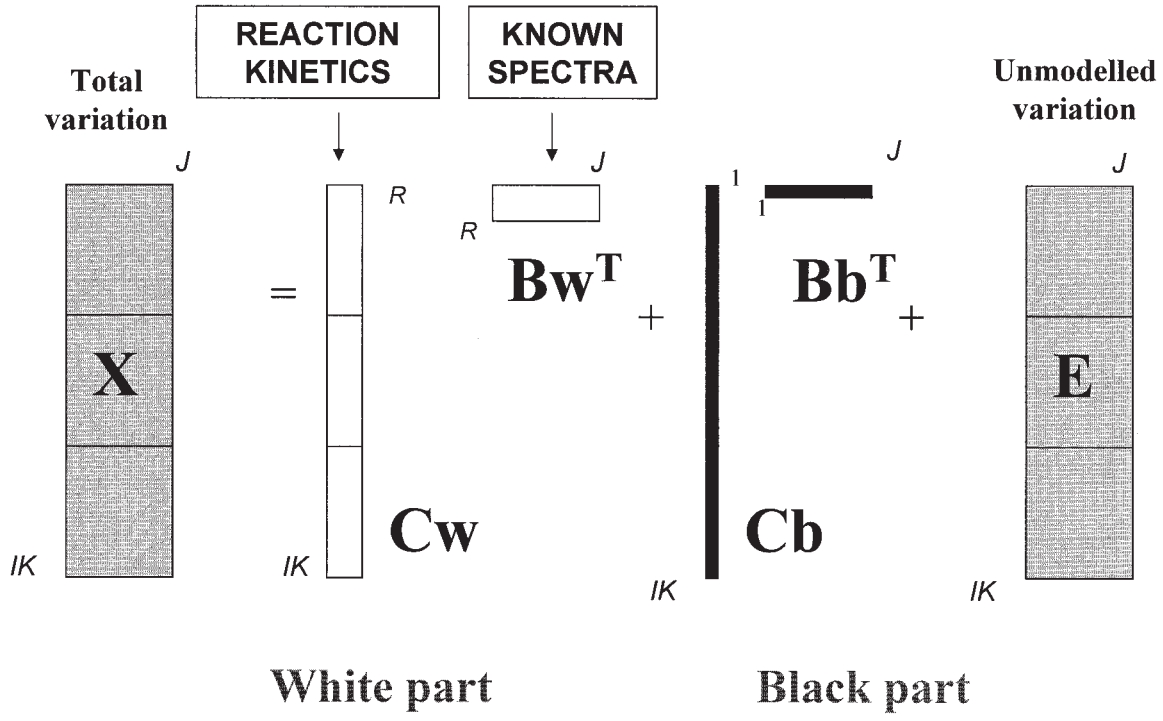

Figure 1. View of a gray Tucker1 model.

addition of the individual components and no interaction. After imposing all white information on the model, additional unrestricted components can be added to the model to capture unknown systematic variation in the data.

A similar reasoning holds for the Tucker3 model. The NIR process spectra are stored in the three-way array $\underline{\mathbf{X}}$ as shown in Figure 2. The model decomposes the three-way array $\underline{\mathbf{X}}$ into a white part and a black part. The white part is defined by the matrices $\mathbf{A}_w$, $\mathbf{B}_w$, $\mathbf{C}_w$, and $\mathbf{G}_w$, which consist of prior process information. With the Tucker3 model, the Lambert–Beer law is incorporated by restricting elements in the core array $\mathbf{G}_w$ to equal zero.[6] If available, pure compound spectra are imposed

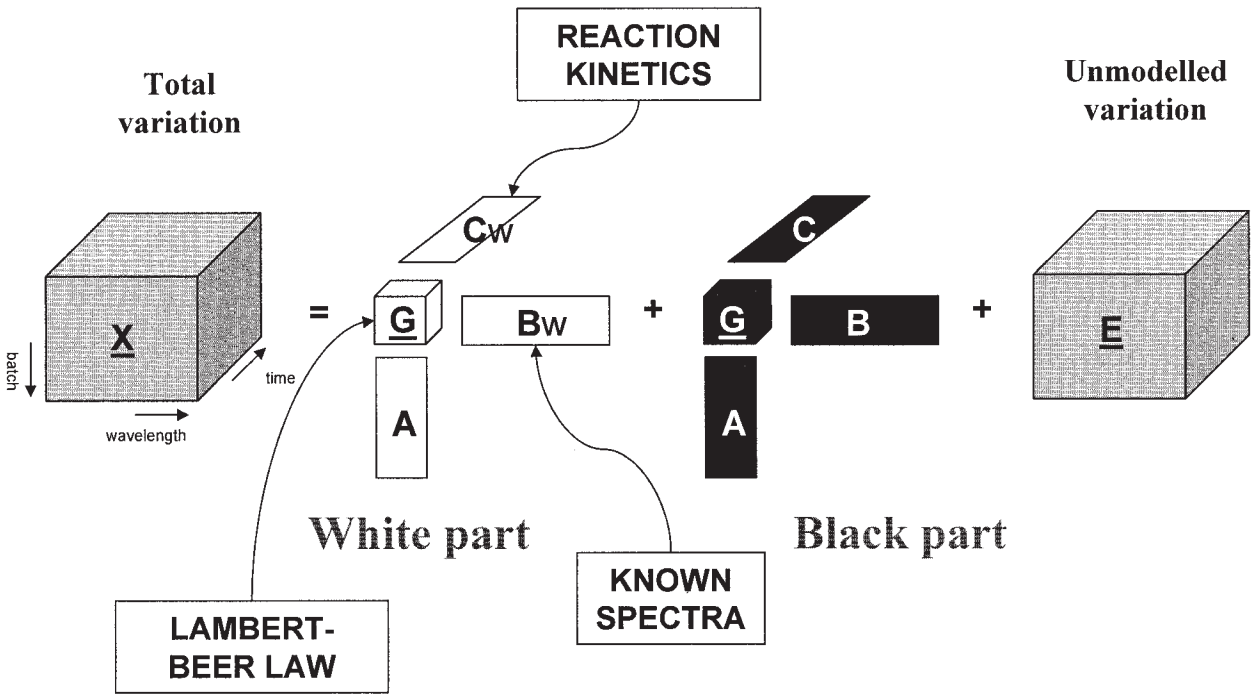

Figure 2. View of a gray Tucker3 model.

on matrix $\mathbf{B}_w$ and process kinetics on matrix $\mathbf{C}_w$. Just as in the gray Tucker1 model, adding new components to the model will define the black part. The black part of the model is defined by the matrices $\mathbf{A}_b$, $\mathbf{B}_b$, $\mathbf{C}_b$, and $\mathbf{G}_b$. Note that the scores $\mathbf{A}_w$ are not white but black because there are no assumptions made about $\mathbf{A}$ or restrictions placed on $\mathbf{A}$.

Batch process monitoring

In spite of the close relationship between the two models, they are substantially different. Both models have a different principle of modeling. Gray Tucker1 is applied with the aim of modeling a time interval $\mathbf{X}_k$ ($I \times K$) and the gray Tucker3 is applied with the aim of modeling the entire three-way array $\underline{\mathbf{X}}$ ($I \times J \times K$). Therefore, both models generate a different test statistic, as will become clear in the following. Furthermore, the test statistic of the gray Tucker1 model has a direct physical interpretation (such as the concentration behavior), whereas the test statistic of the gray Tucker3 model does not have a direct physical interpretation.

On-line monitoring of a new batch using the gray Tucker1 model

For on-line monitoring with a gray Tucker1 model, the local observation vector $\mathbf{x}_k^*$ ($J \times 1$) is used. A new batch is monitored by projecting the local observation vector $\mathbf{x}_k^*$ on $\mathbf{B}$ according to the following equation:

$$\hat{\mathbf{c}}_k = [\mathbf{x}_k^{*\mathbf{T}}\mathbf{B}(\mathbf{B}^\mathbf{T}\mathbf{B})^{-1}]^\mathbf{T} \tag{9}$$

However, the unconstrained solution may result in negative values, which is not wanted, because $\hat{\mathbf{c}}_k$ are relative concentrations and can never be negative. Therefore, the new batch is monitored by regressing the observation $\mathbf{x}_k^*$ onto $\mathbf{B}$ using nonnegativity constraints[18] just as in modeling the historical batches in the trainings phase according to

$$\min_{\mathbf{c}_k} \|\mathbf{x}_k^* - \mathbf{B}\mathbf{c}_k\|^2 \qquad \text{with } \mathbf{c} \geq 0 \tag{10}$$

The local residuals are computed by

$$\mathbf{e}_k^* = \mathbf{x}_k^* - \mathbf{B}\hat{\mathbf{c}}_k \tag{11}$$

The obtained estimates $\hat{\mathbf{c}}_k$ and the local residuals $\mathbf{e}_k^*$ will be used as test statistics for process monitoring.

On-line monitoring of a new batch using the gray Tucker3 model

Monitoring a new batch using the gray Tucker3 model is more complicated than batch process monitoring using a Tucker1 model. Consider the Tucker3 model with

$$\mathbf{V} = (\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^\mathbf{T})^\mathbf{T} \tag{12}$$

A new batch is monitored by projecting the observations $\mathbf{x}$ ($JK \times 1$) onto the model $\mathbf{V}$ ($JK \times P$).

$$\mathbf{a}_{new} = [\mathbf{x}^\mathbf{T}\mathbf{V}(\mathbf{V}^\mathbf{T}\mathbf{V})^{-1}]^\mathbf{T} \tag{13}$$

As can be seen, the monitoring procedure requires an observation vector $\mathbf{x}$ ($JK \times 1$) of full-length JK. This is not available in on-line monitoring because at time k only the observations are known until k. Therefore, assumptions have to be made about the future unknown observations ($k + 1$) to K. For this purpose the method of current deviations is used.[3] That is, it is assumed that the batch remains at the same level (current deviation) for the remainder of the batch run. The residuals are computed according to

$$\mathbf{e}_k = \mathbf{x}_k - \mathbf{V}\mathbf{a}_{new} \tag{14}$$

Statistics

The statistical confidence limits are derived using a leave-one-out principle.[5] In this approach, every batch is left out from the population and treated as if it were a new batch. The remaining batches are used to build a model. The left-out batch is projected onto the model and the obtained scores and residuals are stored for deriving the confidence limits of the control charts.

D-statistic and squared prediction error (SPE) statistic

The two multivariate statistics are the D-statistic and the *SPE*-statistic.[22] The D-statistic monitors deviations *within* the model relative to the center point and the *SPE*-statistic monitors deviations *from* the model. The statistical limits for the D and *SPE* statistics are derived from the NOC data. The I batches from the NOC data are monitored, 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.

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 the case of the gray Tucker1 model, the D-statistic is given by

$$D_k = (\hat{\mathbf{c}}_{new,k} - \bar{\hat{\mathbf{c}}}_k)^{*\mathbf{T}}\tilde{\mathbf{S}}_k^{-1}(\hat{\mathbf{c}}_{new,k} - \bar{\hat{\mathbf{c}}}_k)\frac{I(I-R)}{R(I^2-1)} \sim F(R, I-R) \tag{15}$$

where $\mathbf{c}_{new,k}$ ($R \times 1$) represents the estimates of the relative concentrations of the new batch at time interval k and $\bar{\mathbf{c}}_k$ ($R \times 1$) contains the means of the columns of matrix $\mathbf{C}_k$ ($I \times R$) that are the estimated relative concentrations of the NOC data. The matrix $\tilde{\mathbf{S}}_k$ ($R \times R$) is the variance–covariance matrix of $\mathbf{C}_k$. D_k follows an F-distribution with (R, $I - R$) degrees of freedom. In the case of the gray Tucker3 model, the D-statistic is given by

$$D_k = (\mathbf{a}_{new,k} - \bar{\mathbf{a}}_k)^{*\mathbf{T}}\mathbf{S}_k^{-1}(\mathbf{a}_{new,k} - \bar{\mathbf{a}}_k)\frac{I(I-P)}{P(I^2-1)} \sim F(P, I-P) \tag{16}$$

where $\mathbf{a}_{new,k}$ ($P \times 1$) represents the scores of the new batch at time interval k and $\bar{\mathbf{a}}_k$ ($P \times 1$) contains the means of the columns of the score matrix $\mathbf{A}_k$ ($I \times P$). $\mathbf{S}_k$ ($P \times P$) is the variance–covariance matrix of $\mathbf{A}_k$, which consists of the scores obtained from the NOC data.

SPE-Statistic. For both models, the SPE_k statistic is computed by

$$SPE_k = \sum_{j=1}^{J} e_{jk}^2 \sim g\chi_h^2 \tag{17}$$

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. 17, parameter h represents the degrees of freedom and parameter g represents the weight to account for the magnitude of SPE_k. These two parameters can be estimated in different ways.[3]

In this study the estimation is done according to Jackson and Mudholkar,[19] whereby g and h are functions of the eigenvalues of the residual variance covariance matrix at each time interval k.

Univariate statistics

The individual test statistics $a_{k,r}$ and $c_{k,r}$ can be monitored using univariate control charts. The upper and lower control limits for the Tucker1 model are computed by

$$UCL_k = \bar{c}_{k,r} + t_{(I-1,\alpha/2)}\zeta_k\left(1 + \frac{1}{I}\right)^{1/2} \quad \text{and}$$

$$LCL_k = \bar{c}_{k,r} - t_{(I-1,\alpha/2)}\zeta_k\left(1 + \frac{1}{I}\right)^{1/2} \tag{18}$$

where $\bar{c}_{k,r}$ is the mean computed from matrix $\mathbf{C}_k$ and ζ_k is the corresponding standard deviation computed from $\mathbf{C}_k$. The t is the Student's t-value with $I - 1$ degrees of freedom with an α significance level and I is the number of historical batches. The control limits for the Tucker3 model are derived in a similar way according to

$$UCL_k = \bar{a}_{k,r} + t_{(I-1,\alpha/2)}s_k\left(1 + \frac{1}{I}\right)^{1/2} \quad \text{and}$$

$$LCL_k = \bar{a}_{k,r} - t_{(I-1,\alpha/2)}s_k\left(1 + \frac{1}{I}\right)^{1/2} \tag{19}$$

where $\bar{a}_{k,r}$ is the mean computed from matrix $\mathbf{A}_k$ and s_k is the corresponding variance computed from $\mathbf{A}_k$.

Monitoring levels

To simplify the operation of statistical process monitoring in terms of detection and diagnostics, different monitoring levels are defined. The first level is the process or batch level. This includes monitoring the process using multivariate statistics such as the *D*-statistic and the *SPE*-statistic. Then, if an out of control signal is given in the control charts, a second monitoring level is used. That is, the process variables or components are studied individually using contribution plots[20,21] and univariate statistics to identify and locate the cause of the out of control signal.

Post-batch process monitoring

The aim of post-batch monitoring is to find similarities and differences between batches. For this purpose, both models can be used. The batches are monitored in a similar way as with on-line monitoring. For the gray Tucker3 model an observation vector $\mathbf{x}$ ($JK \times 1$) is available. This vector is projected onto the model and from the obtained scores and residuals *D*- and *Q*-statistics are computed. The derivation of the *D*-statistic is identical to the on-line situation and is computed according to

$$D_i = (\mathbf{a} - \bar{\mathbf{a}})\mathbf{S}^{-1}(\mathbf{a} - \bar{\mathbf{a}})\frac{I(I-P)}{P(I^2-1)} \sim F(P, I-P) \tag{20}$$

where $\mathbf{a}$ ($P \times 1$) is the score vector found for the observation vector and $\mathbf{S}$ ($P \times P$) is the variance covariance matrix of $\mathbf{A}$ ($I \times P$). D_i follows an *F*-distribution with ($P, I - P$) degrees of freedom. The residual statistic is computed by means of the *Q*-statistic[22] given by

$$Q_i = \sum_{jk=1}^{JK} e_{jk}^2 \sim g\chi_h^2 \tag{21}$$

where e_{jk} is the residual at interval jk and Q_i follows a weighted chi-square distribution with h degrees of freedom.

Post-batch monitoring using the gray Tucker1 model is more complex because the data have to be rearranged before analysis. The test statistics are based on the local observation vector $\mathbf{x}_k^*$ ($J \times 1$) and reflect the behavior of the process at time interval k. Thus, each batch generates a vector $\mathbf{c}_k$ ($R \times 1$) and a vector $\mathbf{e}_k$ ($J \times 1$) for each time interval. The generated vectors are stored in the matrices $\mathbf{C}_{total}$ ($I \times RK$) and $\mathbf{E}_{total}$ ($I \times JK$), respectively. By constructing a PCA model on the matrix $\mathbf{C}_{total}$, a post-batch analysis is performed. This is analoguous to the "top level" proposed by Wold et al.[4] Once a PCA model is built, the statistics are similar to the statistics as described for the gray Tucker3 model. As a result two residual matrices have to be analyzed. That is, the constructed matrix $\mathbf{E}_{total}$ ($I \times JK$) and the residual matrix from the PCA model.

Batch consistency number

As stated previously, performing a post-batch analysis with gray models results in multiple control charts: that is, for the gray Tucker1 model, one *D*-chart for the scores and two *Q*-charts, because there are two residual matrices to be studied; the gray Tucker3 model is studied with a *D*-chart and a *Q*-chart. It is much more convenient to have a single index summarizing the behavior of a batch because it is not very easy to express the behavior of a single batch run with the approach described above. For this purpose, the batch consistency number (*BCN*) is introduced. The *BCN* is a single index defined according to the concepts proposed by Raich and Cinar,[23] which were further elaborated by Yue and Qin.[24] The idea of the *BCN* is to combine both test statistics into a single index that summarizes the behavior of a batch. A derivation of the *BCN* is given in Appendix C.

For the grayTucker3 model the *BCN* is computed according to

$$BCN = \mathbf{x}^{\mathbf{T}}\theta\mathbf{x} \sim g\chi_h^2 \tag{22}$$

where $\mathbf{x}$ ($JK \times 1$) is the observation vector used in the gray Tucker3 model and

$$\theta = \frac{\mathbf{V}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}\mathbf{S}^{-1}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}\mathbf{V}^{\mathbf{T}}}{F(I, I-R)}\frac{I(I-R)}{R(I^2-1)} + \frac{(\mathbf{I}-\tilde{\mathbf{V}})^{\mathbf{T}}(\mathbf{I}-\tilde{\mathbf{V}})}{g\chi_h^2} \tag{23}$$

where $\tilde{\mathbf{V}} = \mathbf{V}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}\mathbf{V}^{\mathbf{T}}$ and $\mathbf{I}$ is the identity matrix. The *BCN* follows a weighted chi-square distribution with h degrees of freedom.[24]

Computing the batch consistency number for the gray Tucker1 model is done in a slightly different manner because this model has a local observation vector. The *BCN* is found by first computing the local batch consistency numbers (BCN_k) for each time interval. After that, the local batch consistency numbers are summarized and divided by the number of time intervals. The local BCN_k is computed according to

$$BCN_k = \mathbf{x}_k^{*\mathbf{T}}\Theta\mathbf{x}_k^* \sim g\chi_h^2 \tag{24}$$

where $\mathbf{x}_k^*$ ($J \times 1$) is the local observation vector and

$$\Theta = \frac{\mathbf{B}(\mathbf{B}^{\mathbf{T}}\mathbf{B})^{-1}\tilde{\mathbf{S}}_k^{-1}(\mathbf{B}^{\mathbf{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathbf{T}}}{F(I, I-R)}\frac{I(I-R)}{R(I^2-1)} + \frac{(\mathbf{I}-\tilde{\mathbf{B}})^{\mathbf{T}}(\mathbf{I}-\tilde{\mathbf{B}})}{g\chi_h^2} \tag{25}$$

where $\tilde{\mathbf{B}} = \mathbf{B}(\mathbf{B}^{\mathbf{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathbf{T}}$. From the expression in Eq. 25, it follows that the observation vector is regressed on matrix $\mathbf{B}$ in an ordinary least-squares manner. However, this regression does not take into account the imposed nonnegativity restrictions. This is done with an extra regression step as described previously. Therefore, $\mathbf{x}_k^*\mathbf{B}(\mathbf{B}^{\mathbf{T}}\mathbf{B})^{-1}$ is replaced with $\hat{\mathbf{c}}_k$. The same reasoning holds for the regression of the observation vector onto $(\mathbf{I}-\tilde{\mathbf{B}})$. This regression step is now replaced by the residuals $\mathbf{e}_k$. Therefore, the local BCN_k is computed for every interval k according to

$$BCN_k = \frac{\hat{\mathbf{c}}_k\tilde{\mathbf{S}}_k^{-1}\hat{\mathbf{c}}_k^{\mathbf{T}}}{F(I, I-R))}\frac{I(I-P))}{P(I^2-1)} + \frac{\mathbf{e}^{\mathbf{T}}\mathbf{e}}{g\chi_h^2} \tag{26}$$

After that, the *BCN* is found by

$$BCN = \frac{1}{K}\sum_{k=1}^{K} BCN_k \sim g\chi_h^2 \tag{27}$$

Experimental

Process description

The process of making a urethane resin is a multistage fed-batch process. In this study, only the first stage is of interest. The main reaction in the first stage is given by

where a di-isocyanate reacts with an alcohol to a urethane resin with an extra NCO group for further OH addition in subsequent steps.

The operation of this first stage is divided into three steps. The first step, after cleaning and purging of the reactor with N_2, is filling the reactor with a di-isocyanate. In the second step a second reactant (alcohol) is added dropwise and the reaction starts. The third step is the part of the process where all the di-isocyanate has reacted and reaction mixture is kept at a constant temperature. This process is operated according to a predefined change in temperature. All reactants are added in precisely known stoichiometric ratios. Furthermore, it is known that a side reaction occurs. The alcohol reacts with the urethane to an undesired side product coupling to the second NCO group of the di-isocyanate.

Instrumentation

The full-scale process spectra are obtained using an in-line Foss NIR Systems 6500 spectrometer with a 2-nm resolution and a transflection probe with a 2 × 2-mm path length in the range of 1100 to 2500 nm. The laboratory-scale compound spectra are obtained using a model FT-NIR 9x-III spectrometer (Perkin Elmer Cetus Instruments, Norwalk, CT) with a 2-nm resolution and a 2-mm path length in the range of 1200 to 2500 nm.

Data pretreatment

The collected spectra have a baseline offset, which is removed by the use of second-derivative spectra estimated using a Savitzky–Golay filter.[25] A visual inspection showed that the best results were obtained using a window of 21 channels and a second-order polynomial. There are 18 batches available ($I = 18$) for this study. It is known that four of the available batch runs are not operated under the defined normal operating conditions and are therefore excluded from the model. The remaining 14 batches are used for modeling and validation. For every batch run i ($i = 1 \cdots I$), the selected J wavelengths are stacked in a three-way array $\underline{\mathbf{X}}$ ($I \times J \times K$).

Band assignment and band selection

From the literature[26] it is known that the di-isocyanate (NCO ~ 1920 nm), the alcohol (OH ~ 1398–1421 nm), and the product (NH ~ 1485–1503 nm) are spectroscopically active in the near-infrared range (1100–2500 nm). Furthermore, it is known that the monitored reaction is exothermal and that a predefined temperature gradient exists. The functional groups present in the compounds are shape sensitive for temperature changes. This is not desired because the pure compound spectra are considered as time-invariant factors. Therefore, a selection of spectral bands is made in such a way that the bands are only slightly affected by temperature changes and consist of sufficient chemical process information. The selected range is validated by laboratory experiments. That is, pure compound spectra of di-isocyanate and

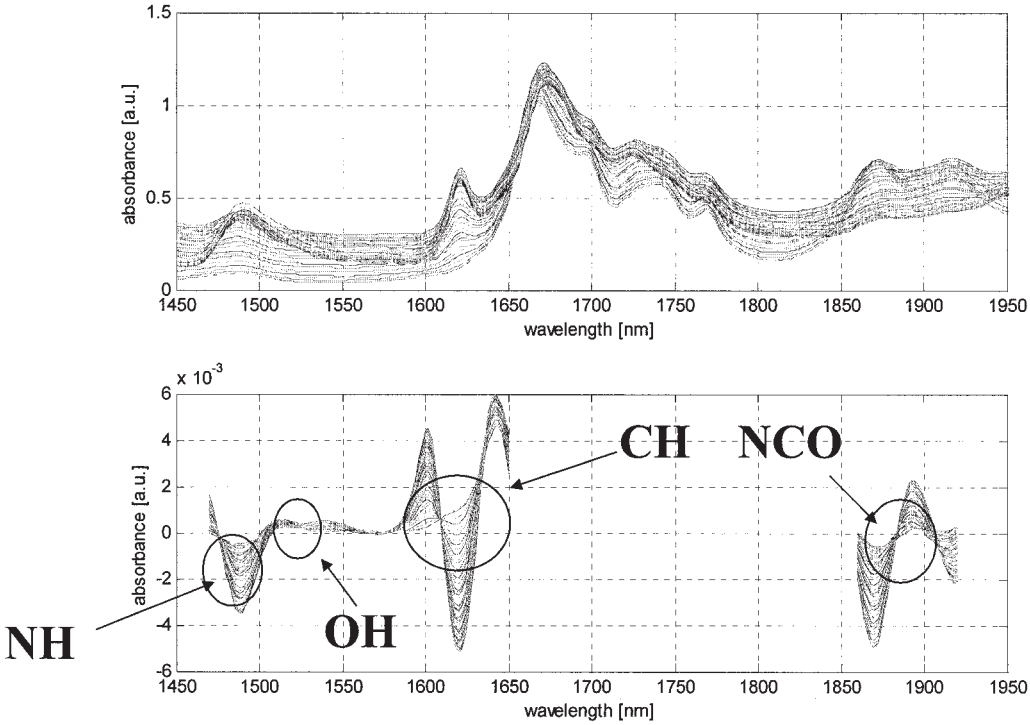

Figure 3. Obtained spectra for a single batch run.

The raw spectra are given in the upper graph and the second derivative spectra of the selected wavelength channels are given in the lower graph.

alcohol were measured at different temperatures. Based on the observed changes in the measured spectra, a wavelength selection was made compromising between temperature effects and chemical compound information.[7]

The top graph in Figure 3 shows the spectra obtained during a single batch run. In the bottom graph of Figure 3, the second-derivative spectra and the selected wavelength range are given. The circles highlight the wavelength ranges corresponding to the functional groups (NCO, OH, CH, and NH) present in the system.

Mean centering

To study the differences between batches it is common to column center the process data across the batch mode in the situation where the three-way array is modeled using a Tucker3 model. This operation removes the common variation in the batches and the remainder of the data consists of the differences between the batches around a zero mean.

To understand the effect of the centering operation, the data are assumed to have systematic variation present in all the batches and unsystematic variation specific for each batch. The systematic variation is process specific with linear differences between the batches. The unsystematic variation is caused by nonlinear deviations and experimental error and thus is typically batch specific. If the data consist of only systematic variation such as well-behaved spectra according to the Lambert–Beer law, there is no difference between the model loadings **B** and **C** obtained from centered process data and model loadings obtained from noncentered process data (Appendix C).

However, process data usually consist of additional unsystematic variation. If after centering the systematic variation is relatively small compared to the unsystematic variation such as nonlinear behavior, this will cause estimation problems of the model loadings. Therefore, in this study, the models are constructed using noncentered data.

Results and Discussion

Construction of the models

The model parameters are estimated using restricted alternating least squares based on prior knowledge of the polyurethane process. The following is known of the urethane process: The law of Lambert–Beer holds, that is, linear addition and no interaction between the individual compounds is allowed. Furthermore, there are estimations of the pure compound spectra available from a previous study.[7] Also, it is known that the concentration can never be negative; therefore, nonnegativity constraints are imposed on the concentration profiles. An overview of the prior knowledge is summarized in Table 1. Furthermore, it is pointed out where this knowledge is incorporated into the models.

In the gray Tucker1 model, the model implicitly gives the Lambert–Beer law. The pure compound spectra are imposed as restrictions on the columns of **B** and nonnegativity constraints are

Table 1. Prior Information Available of the Polyurethane Process

Prior Knowledge	Tucker1	Tucker3
Lambert–Beer	Implicit	**G**
Spectra	**B**	**B**
Nonnegative constraint	**C**	**C**

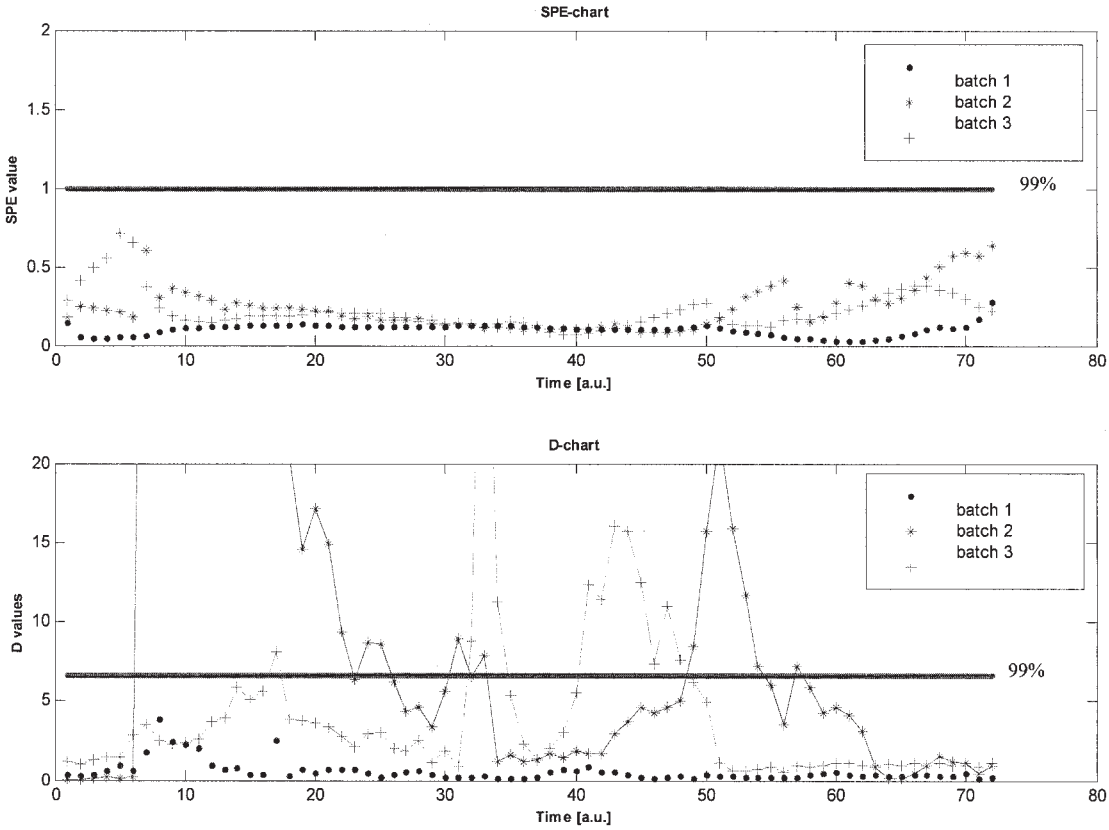

Figure 4. Monitoring level 1 for a gray Tucker1 model.

Monitoring batch 1, batch 2, and batch 3.

imposed on the profiles of **C**. In the gray Tucker3 model, restricting relevant elements of the core array **G** to equal zero, according to ($\mathbf{G} = [g_{111}\ 0\ 0\ 0\ g_{122}\ 0\ 0\ 0\ g_{133}]$), imposes the Lambert–Beer law. The pure compound spectra are incorporated in matrix **B** and nonnegativity constraints are imposed on **C**.

A priori, the number of black components is unknown. Therefore, a series of models was constructed starting with the simplest model consisting of white knowledge only. Each successive model had an additional black component. Next, the magnitude of the root mean square error (RMSE) of experimental variation was used as a reference to make a decision about the number of black components. The RMSE of experimental variation was estimated with spectra obtained from the period before the addition of the alcohol because in this period there is only di-isocyanate present in the reactor. From the magnitude of the RMSE of both white models it was concluded that additional black components do not consist of any significant systematic variation and thus are not included in the models. The first model in this study is a gray Tucker1 model with three components ($R = 3$). The second model is a gray Tucker3 model with $P, Q, R = [1\ 3\ 3]$.

On-line monitoring

The monitoring performance of both models is validated. From prior analysis, it is known that four batches of the historical data set are operated in a different manner. Only two of the four faulty batches are used for testing the models and control charts because there are only two distinct types of faults. The remaining two batches have a similar type of fault and are therefore not used for testing. Furthermore, one normal batch is selected as a reference and is excluded from the model. This results in three batches for performance testing (batches 1, 2, and 3) and 13 batches for modeling the NOC behavior. Batch 1 is a normal batch operated under NOC; batch 2 has a clear different dosing profile of the alcohol; and batch 3 had a manual dosing of the alcohol instead of the normal automatic dosing, which results in slightly different compound profiles.

First monitoring level with the gray Tucker1 model

The three batches are monitored using the *D*-chart and *SPE*-chart. Both charts are presented with 99% confidence limits in Figure 4. The top graph in Figure 4 corresponds to the normalized SPE-chart. That is, the *SPE*-statistic is normalized for the 99% confidence limit. There, it is shown that not one of the monitored batches crosses the confidence limit. This implies that there is no significant breaking of the correlation between the channels and the batches behave within the model plane. Deviations in the model plane are observed in the *D*-chart. In the *D*-chart, the bottom graph of Figure 4, the differences between the batches become apparent. Batch 1 behaves as expected and no alarm is observed. For batch 2 there is a clear out of control signal observed starting at time = 7. This excessive signal in the *D*-chart indicates that there is a significant change in the model plane. The batch continues to be out of control until time = 35, where it operates below the confidence limits until the batch crosses the limits again at time = 44. The operation of batch 3 shows multiple crossings of the confidence limits. The first major crossing at time = 31 to 33

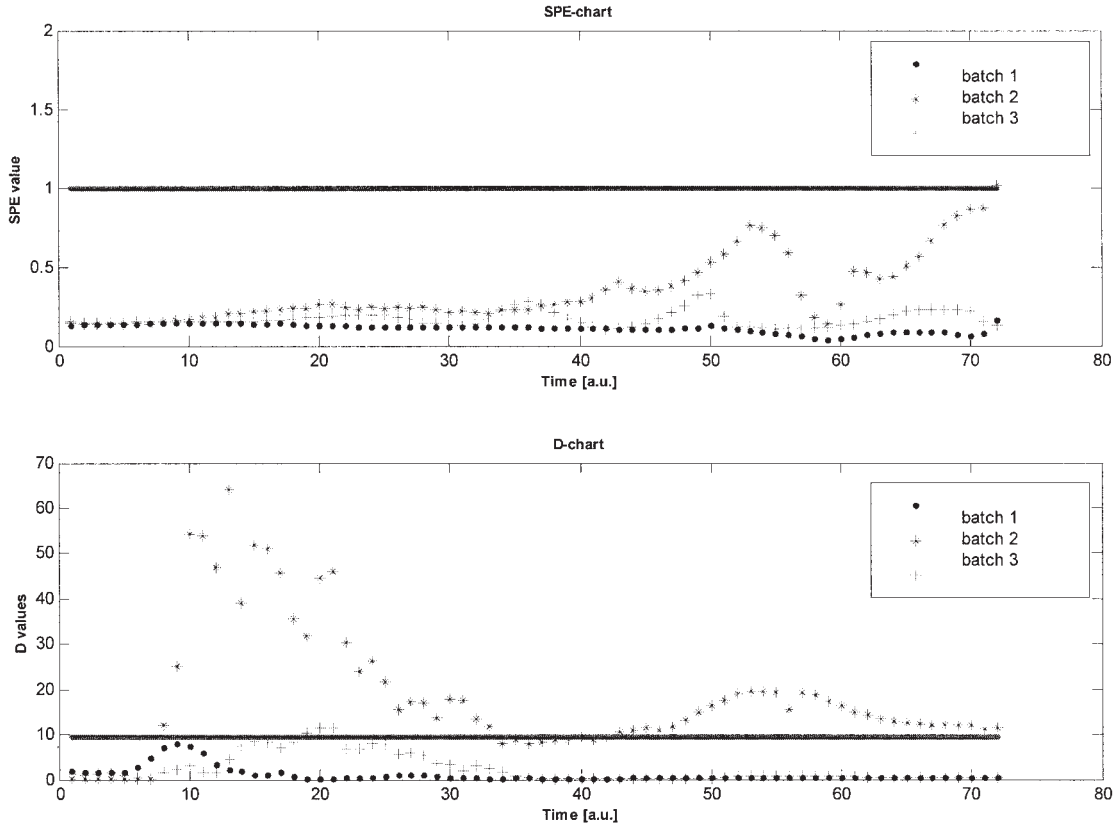

Figure 5. Monitoring level 1 for a gray Tucker3 model.

Monitoring batch 1, batch 2, and batch 3.

and the second at time = 40 to 50, implying there are two events during the process.

Because of the nature of the gray Tucker1 model, the concentration behavior of the three chemical compounds in the reactor is captured by the *D*-chart. Deviations and disturbances such as contaminations are likely to be captured in the *SPE*-chart; however, this will occur of course only if the disturbances are spectroscopically active.

First monitoring level with the gray Tucker3 model

The same batches are monitored using the gray Tucker3 model. The normalized *SPE*-chart and *D*-chart with 99% confidence limits are given in Figure 5. The *SPE*-chart shows that none of the batches crosses the confidence limit, implying that there are no significant breakings of the correlation. In the *D*-chart is shown that the normal batch behaves as expected. Batch 2 gives a clear out of control signal at time = 8 and stays above the confidence limits during the entire batch run. Observing the behavior of batch 3, it can be seen that this batch touches the confidence limit around time = 13 to 28. This implies that this batch operates on the borders of specified normal conditions. For the remainder of the run the batch behaves within the specified control limit.

In general, both models are detecting the same deviations, although the time of observation may differ. Furthermore, the gray Tucker1 model is more sensitive in detecting deviations. This can be explained by the nature of the models.

The gray Tucker3 model is a global model in which the behavior of a batch is captured by modeling the autocorrelation and cross-correlation between the process variables with the matrices **G**, **B**, and **C**. Linear differences from the auto- and cross-correlation are captured by the scores in matrix **A**. Unsystematic variation and nonlinear differences are captured by the residuals **E**.

It is shown that if a process is monitored with a global model, the process history is taken into account up to time k and will lead to graduated changes if an event occurs.

With the gray Tucker1 model it is easy to understand that this model will take into account only the correlation between the process variables and not the auto- and cross-correlation over time. However, the auto- and cross-correlations of the process are implicitly captured by the control limits. Here, it is shown that this model is well able to detect deviations in the process. In this specific situation, a semibatch process is monitored by means of NIR spectra. The nature of the process data is defined in such a way that the autocorrelation is captured by the concentration profiles and the model captures the correlation between the spectra. In principle there is no need for taking into account the entire process history because this is implicitly captured by the current measurement. Concluding, for on-line monitoring there is no need for the more complex gray Tucker3 model.

Second monitoring level gray Tucker1

It is clear from the control charts that batch 2 is not operating under NOC conditions and that batch 3 is operating on the border of the specified confidence limits. In the second monitoring level the aim is to diagnose the cause of

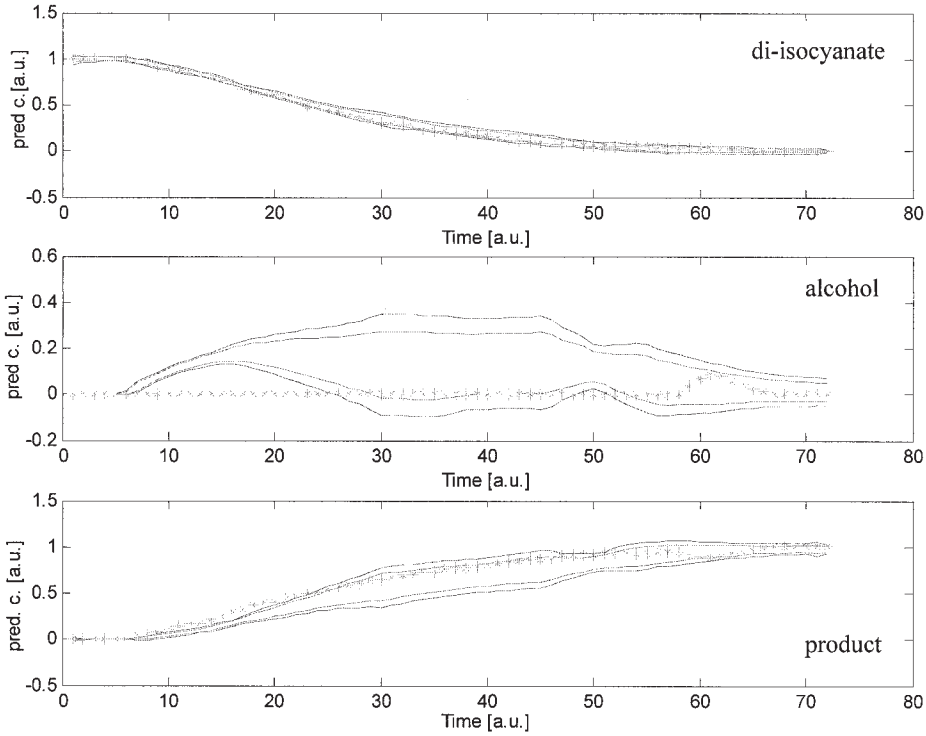

Figure 6. Second monitoring level using a gray Tucker1 model for batch process monitoring.

Monitoring batch 2.

the observed deviations in the control charts. First, the gray Tucker1 model will be discussed.

With the gray Tucker1 model it is possible to study the behavior of the individual chemical compounds using univariate control charts. The behavior of the individual compounds for batch 2 is given in Figure 6 with 95 and 99% confidence limits. The top graph corresponds to the di-isocyanate, the middle graph corresponds to the alcohol, and the bottom graph corresponds to the product. For the alcohol profile, it is obvious that its behavior is different from that under normal conditions and is assigned as the cause for the out-of-control signal in the *D*-chart. The observed profiles suggest that there is no alcohol present in the reactor. However, the reaction takes place, implying there is a reactant added to the system. Because it is known that there are no out-of-control signals observed in the *SPE*-chart, it is assumed that the added alcohol is similar to normal conditions. Furthermore, it is known that the NIR probe is situated in the lower part of the reactor. Because the alcohol is added dropwise in the top of the reactor, it is entirely possible that the alcohol has reacted before it can reach the NIR probe. This implies that either the reaction kinetics are faster than usual or a lower amount of alcohol is added to the system.

The compound profiles for batch 3 are given in Figure 7. Looking at the di-isocyanate profiles and the product profiles it seems that the reaction behaves according to normal operating conditions. A closer look at the profiles shows the apparent differences. At the start of the reaction the process behaves as expected until time = 10 where the alcohol concentration is higher than expected. It seems that the reaction is inhibited. After time = 20, it seems that the reaction is accelerating until there is no alcohol present in the reactor. This is marked as an out-of-control signal in the *D*-chart at time = 30. Next, the alcohol is increasing, which is expressed as a signal in the *D*-chart at time = 40.

Second monitoring level gray Tucker3

In the case of the gray Tucker3 model it is not straightforward to compute the individual component profiles. However, in this study it is possible to compute the individual contribution of the chemical compound to the *D*-statistic because no interactions between the chemical compounds are allowed. The derivation of the individual contributions is given in Appendix B.

The individual contributions of the three components (di-isocyanate, alcohol, and product) for batch 2 are given in Figures 8a, b, and c, respectively. It is easy to see that from time = 9 the alcohol has the largest contribution to the *D*-statistic and thus the alcohol can be assigned as the main cause for the out-of-control signal in the *D*-chart. For batch 3, the individual contributions of the di-isocyanate, alcohol, and product are given in Figures 9a, b, and c, respectively. It is known that it behaves on the border of the specified confidence limits. For this batch, the individual contributions do not give a clear indication as to which compound has the largest contribution to the *D*-statistic. In general, the contributions of the gray Tucker3 model are less clear compared to the compound profiles found with the gray Tucker1 model. An advantage of the gray Tucker1 model is the direct interpretation of chemistry in the reactor, whereas the gray Tucker3 model gives only an indication which component is deviating.

Batch consistency number

The batch consistency number (*BCN*) is computed for all batches and presented with a 95-percentile limit (that is, the

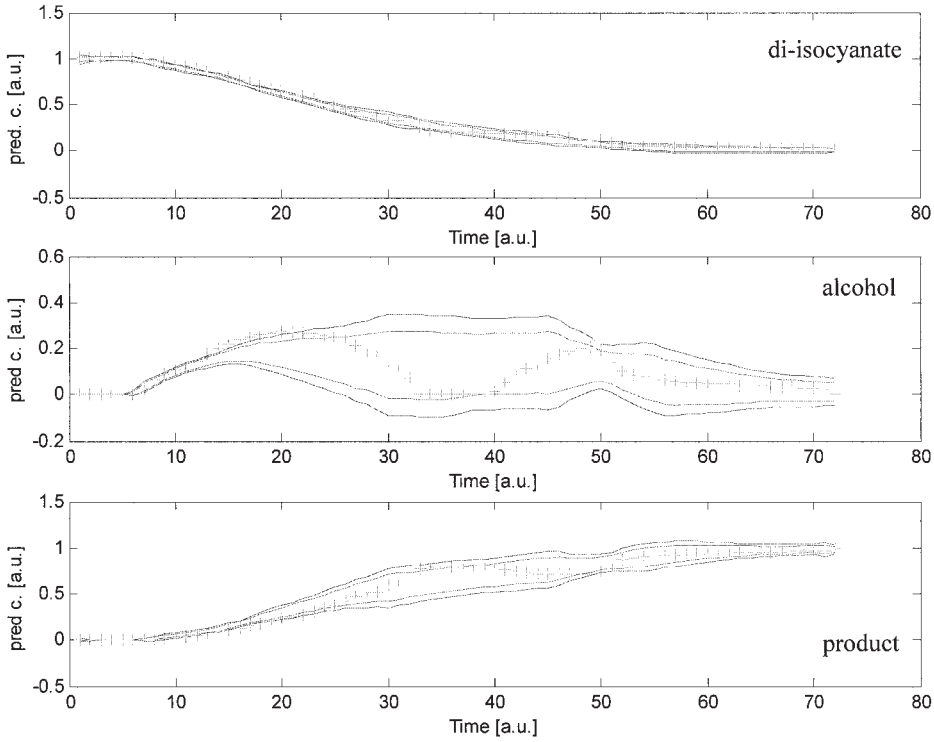

Figure 7. Second monitoring level using a gray Tucker1 model for batch process monitoring.

Monitoring batch 3.

value below which 95 percent of the reference *BCN* will fall). The *BCN* is computed in a leave-one-out manner similar to the derivation of the control charts. The results for the *BCN* obtained with a gray Tucker1 model are given in Figure 10. Numbers 15 to 17 in Figure 10 are the indices of the three monitored batches. The results found for the post-batch analysis of the three monitored batches correspond with the results found for the on-line situation because it can be expected that if a batch is out of control with on-line monitoring, this batch is also out of control with the post-batch analysis.

The *BCN* of the normal batch (number 15) is of such a magnitude that it falls within the magnitude of the NOC

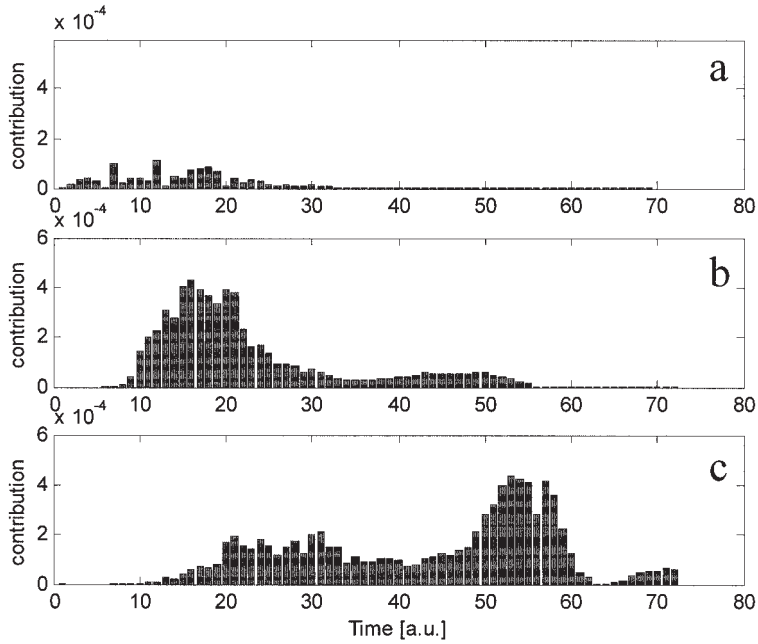

Figure 8. Contribution plots for a gray Tucker3 model.

Monitoring batch 2: (a) di-isocyanate, (b) alcohol, (c) product.

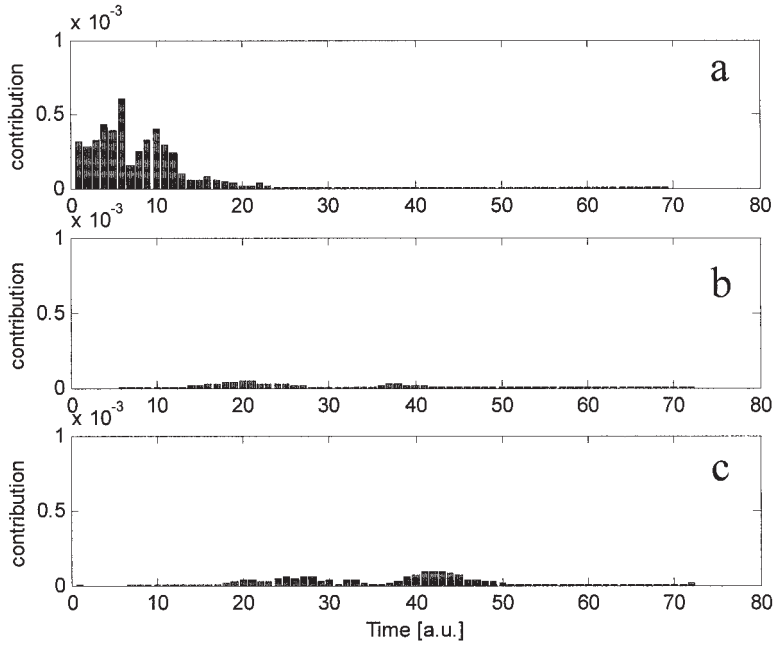

Figure 9. Contribution plots for a gray Tucker3 model.

Monitoring batch 3: (a) di-isocyanate, (b) alcohol, (c) product.

batches. For batch 2 (number 16) a large *BCN* is observed, indicating that this batch is not operated in a consistent manner. For batch 3 (number 17) it is known that it behaves on the border of normal operating conditions and the *BCN* shows that the batch behaved in a consistent manner compared to the NOC batches. Furthermore, two NOC batches (numbers 2 and 4) have a relatively large *BCN*. Both batches are considered as NOC batches according to the operator's log. An individual inspection revealed that these batches had a deviating dosing profile of the alcohol compared to the average alcohol profile.

The results for the gray Tucker3 model are given in Figure 11. In this figure, the numbers 15 to 17 are the indices of the three monitored batches. As can be seen, the *BCN* of the normal batch lies in the same order of magnitude as the NOC batches. Batch 2 (number 16) has a large *BCN*, indicating that batch 2 is not operated under NOC. The magnitude for batch 3 (number 17) lies at the same level as that of most of the NOC batches, indicating that the overall consistency is within NOC variation. It is interesting to observe that compared to the previous figure, the batch at number 4 has a relatively large *BCN* and the batch at number 2 lies within NOC variation. Because both models are different by nature there is no clear explanation for this behavior.

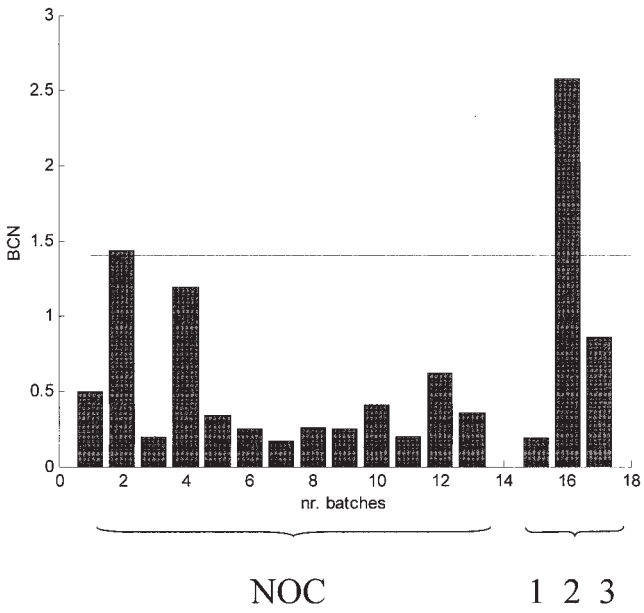

Figure 10. BCN for a gray Tucker 1 model.

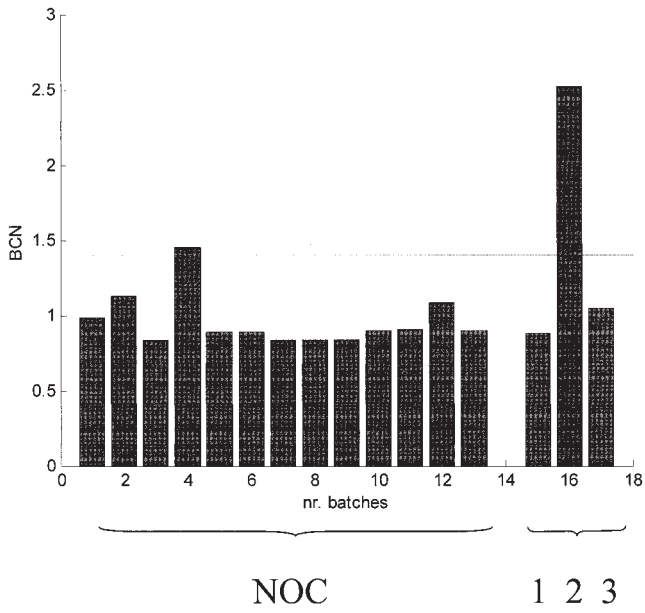

Figure 11. BCN for a gray Tucker3 model.

Table 2. Summary of the Findings: Similarities and Differences between the Gray Models

Gray Tucker1	Gray Tucker3
Unique model parameters	Unique model parameters
Simple on-line monitoring	Simple post-batch monitoring
Within-batch variation	Between-batch variation

Conclusions

A strategy is presented to monitor batch process with gray models. To this extent, the batch process monitoring performances of a gray Tucker1 model and a gray Tucker3 model are studied using an industrial case study. Both models are used to capture the NOC behavior of the process of interest and to derive statistical control charts to detect deviations from NOC and to diagnose the cause of these deviations. The differences between the two models are summarized in Table 2.

The gray Tucker1 model has unique model parameters because of the enforced restrictions. The gray Tucker1 model has a better fit compared to that of the gray Tucker3 model, which is also reflected in the sensitivity of detecting deviations from normal operating conditions. The most robust part of the gray Tucker1 model is its on-line monitoring capability, especially in assigning causes of abnormal behavior once an on-line control charts detects a fault. This is attributed to an easy interpretation of its model parameters in terms of within-batch variation. The gray Tucker1 model is relatively poor in post-batch monitoring and judging batch-to-batch consistency.

The gray Tucker3 model also has unique model parameters because of the enforced restrictions. However, this model is much more rigid compared to the gray Tucker1 model, which is reflected in the poorer model fit. The use of a gray Tucker3 model for post-batch analysis is straightforward. A *BCN* can be defined and interpreted. The model parameters have an easy interpretation in terms of between-batch variation. This makes the gray Tucker3 model very useful for post-batch monitoring and analysis. The gray Tucker3 model has a relatively poor performance in on-line monitoring, especially in terms of diagnosing faults.

Summarizing, a gray Tucker1 model is more suited for on-line monitoring and a gray Tucker3 model for post-batch monitoring. The gray Tucker1 model focuses more on within-batch variation, whereas the gray Tucker3 model focuses more on between-batch variation, thus making the models complementary.

Literature Cited

1. Workman J Jr, Veltkamp DJ, Doherty S, Anderson BB, Creasy KE, Koch M, Tatera JF, Robinson AL, Bond L, Burgess LW, Bokerman GN, Ullman AH, Darsey GP, Mozayeni F, Bamberger JA, Greenwood MS. Process analytical chemistry. *Anal Chem*. 1999;71:121R-180R.
2. Westerhuis JA, Gurden SP, Smilde AK. Spectroscopic monitoring of batch reactions for on-line fault detection and diagnosis. *Anal Chem.* 2000;72:5322-5330.
3. Nomikos P MacGregor JF. Monitoring batch processes using multiway principal component analysis. *AIChE J.* 1994;40:1361-1375.
4. Rannar S, MacGregor JF, Wold S. Adaptive batch monitoring using hierarchical PCA. *Chemomet Intell Lab Syst.* 1998;41:73-81.
5. Louwerse DJ, Smilde AK. Multivariate statistical process control of batch processes using three-way models. *Chem Eng Sci.* 2000;50:1225-1235.
6. Gurden SP, Westerhuis JA, Bijlsma S Smilde AK. Modelling of spectroscopic batch process data using grey models to incorporate external information. *J Chemomet.* 2001;15:101-121.
7. Van Sprang ENM, Ramaker H-J, Westerhuis JA, Smilde AK, Gurden SP, Wienke D. NIR spectroscopic monitoring of a series of industrial batch processes using a bilinear grey model. *Appl Spectrosc*. 2003;57:1007-1019.
8. de Juan A, Maeder M, Martinez M Tauler R. Combining hard- and soft-modelling to solve kinetic problems. *Chemomet Intell Lab Syst.* 2000;54:123-141.
9. Ramaker H-J, Van Sprang ENM, Gurden SP, Westerhuis JA, Smilde AK. Improved monitoring of batch processes by incorporating external information. *J Process Control.* 2002;12:569-576.
10. Bijlsma S, Louwerse DJ, Smilde AK. Estimating rate constants and pure UV-vis spectra of a two-step reaction using trilinear models. *J Chemomet.* 1999;13:311-329.
11. Kiers HAL. Towards a standardized notation and terminology in multiway analysis. *J Chemomet.* 2000;14:105-122.
12. Kroonenberg PM. *Three-Mode Principal Component Analysis: Theory and Applications*. Leiden, The Netherlands: DWO Press; 1983.
13. Tates AA, Louwerse DJ, Smilde AK, Koot GLM, Berndt H. Monitoring a PVC batch process with multivariate statistical process control charts. *Ind Eng Chem Res.* 1999;38:4769-4776.
14. Boqué R, Smilde AK. Monitoring and diagnosing batch processes with multiway covariates regression models. *AIChE J.* 1999;45:1504-1520.
15. Kassidas A, MacGregor JF, Taylor PA. Synchronization of batch trajectories using dynamic time warping. *AIChE J.* 1998;44:864-875.
16. Ramaker H-J, Van Sprang ENM, Boelens HFM, Westerhuis JA, Smilde AK. Dynamic time warping for spectroscopic data. *Anal Chim Acta*. 2003;498:133-153.
17. Kourti T. Multivariate dynamic data modelling and statistical process control of batch processes, start-ups and grade transitions. *J Chemomet.* 2003;17:93-109.
18. Bro R, De Jong S. A fast non-negativity-constrained least squares algorithm. *J Chemomet.* 1997;11:393-401.
19. Jackson JE, Mudholkar GS. Control procedures for residuals associated with principal component analysis. *Technometrics.* 1979;21:341-349.
20. Miller P, Swanson RE, Heckler ChE. Contribution plots: A missing link in multivariate quality control. *Appl Math Comput Sci.* 1998;8:775-792.
21. Westerhuis JA, Gurden SP, Smilde AK. Generalized contribution plots in multivariate statistical process monitoring. *Chemomet Intell Lab Syst.* 2000;51:95-114.
22. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics.* 1995;37:41-59.
23. Raich A, Cinar A. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. *AIChE J.* 1996;42:995-1009.
24. Yue HH, Qin SJ. Reconstruction-based fault identification using a combined index. *Ind Eng Chem Res.* 2001;40:4403-4414.
25. Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem.* 1964;36:1627-1639.
26. Crandall EW, Ashvin NJ. The near-infrared spectra of polymers. *J Appl Polym Sci.* 1977;21:449-454.

Appendix A

In the following it is shown that column mean centering for the Tucker3 model does not affect the model loadings in the noiseless case. The column centered matrix $\mathbf{X}_c$ is a projection of $\mathbf{X}$ that removes the constant terms in the data.

$$\mathbf{X}_c = \mathbf{P}^{\perp}\mathbf{X} \qquad \text{and} \qquad \mathbf{P}^{\perp} = (\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathbf{T}}/I) \tag{A1}$$

where $\mathbf{P}^{\perp}$ is an orthogonal projection matrix. Take the noiseless case where $\mathbf{X}$ ($I \times JK$) is modeled using a Tucker3 model given in the following equation:

$$\mathbf{X} = \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^T \tag{A2}$$

and the orthogonal projection of $\mathbf{X}$ is given by

$$\mathbf{P}^{\perp}\mathbf{X} = \mathbf{X}_c = \mathbf{P}^{\perp}\mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^T \tag{A3}$$

From Eq. A3 it can be seen that model loadings $\mathbf{B}$ and $\mathbf{C}$ are not affected by the centering operation.

Appendix B

In general contribution plots are computed to study which process variable has the largest contribution to the *D*-statistic. However, in the current monitoring problem it is not so interesting to know which process variable has the largest contribution, because the process variables are the wave channels, but rather to know which chemical component contributes to the *D*-statistic. The current model is a gray Tucker3 model with $P, Q, R = [1\ 3\ 3]$.

The score a_k is computed by

$$a_k = \mathbf{x}_k^{\mathbf{T}}\mathbf{v}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1} \qquad \text{and} \qquad \mathbf{v} = [\mathbf{G}(\mathbf{C}^{\mathbf{T}} \otimes \mathbf{B}^{\mathbf{T}})]^{\mathbf{T}} \quad \text{(B1)}$$

where $\mathbf{x}_k$ ($JK \times 1$) is the observation vector of the whole batch. Furthermore, the "current deviations" approach is used to anticipate for the missing values. The *D*-statistic for time interval k is computed by

$$D_k = (a_k - \bar{a}_k)s_k^{-1}(a_k - \bar{a}_k) \quad \text{(B2)}$$

where s_k is the variance of score a_k on time interval k.

$$D_k = [\mathbf{x}_k^{\mathbf{T}}\mathbf{v}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1} - \bar{\mathbf{x}}_k^{\mathbf{T}}\mathbf{v}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1}]s_k^{-1}[\mathbf{x}_k^{\mathbf{T}}\mathbf{v}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1} - \bar{\mathbf{x}}_k^{\mathbf{T}}\mathbf{v}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1}] \quad \text{(B3)}$$

$$D_k = \|s_k^{-1/2}[(\mathbf{x}_k^{\mathbf{T}} - \bar{\mathbf{x}}_k^{\mathbf{T}})\mathbf{v}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1}]\|^2 \quad \text{(B4)}$$

$$D_k = \left\{ s_k^{-1/2} \sum_{jl=1}^{JL} [(x_{jl} - \bar{x}_{jl})v_{jl}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1}] \right\}^2 \quad \text{with } l = 1 \cdots L \quad \text{and} \quad L = K \quad \text{(B5)}$$

$$D_k = \left\{ s_k^{-1/2} \sum_{j=1}^{J} \sum_{l=1}^{L} \left[(x_{jl} - \bar{x}_{jl}) \sum_{r=1}^{R} \sum_{q=1}^{Q} g_{pqr}b_{j,r}c_{l,q}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1} \right] \right\}^2 \quad \text{(B6)}$$

Take $h_k = s_k^{-1/2}(\mathbf{v}^{\mathbf{T}}\mathbf{v})^{-1}$

$$D_k = \left\{ h_k \sum_{j=1}^{J} \sum_{l=1}^{L} [(x_{jl} - \bar{x}_{jl}) g_{pqr}c_{l,q}b_{j,r}] \right\}^2 \quad \text{(B7)}$$

With the method of current deviations an observation at time interval k will be assumed to be valid throughout the rest of the batch run. Because of this, an unrealistic contribution is found for individual compounds. Therefore, the local contribution is computed. The local contribution is computed by means of the local observation vector $\mathbf{x}_{i,k}$ ($J \times 1$).

The first component

$$contribution_k^{111} = [h_k g_{111}\mathbf{b}_1^{\mathbf{T}}(\mathbf{x}_{i,k} - \bar{\mathbf{x}}_{i,k})c_{k,1}]^2 \quad \text{(B8)}$$

The second component

$$contribution_k^{122} = [h_k g_{122}\mathbf{b}_2^{\mathbf{T}}(\mathbf{x}_{i,k} - \bar{\mathbf{x}}_{i,k})c_{k,2}]^2 \quad \text{(B9)}$$

The third component

$$contribution_k^{133} = [h_k g_{133}\mathbf{b}_3^{\mathbf{T}}(\mathbf{x}_{i,k} - \bar{\mathbf{x}}_{i,k})c_{k,3}]^2 \quad \text{(B10)}$$

Appendix C

The batch consistency number (*BCN*) is a single index summarizing the batch behavior compared to historical batches obtained under normal operating conditions. Because the *BCN* is a summary for the behavior of a single batch run, the *BCN* should be function of the *D* and *Q* statistics or $BCN = f(D, Q)$. The definition of the *BCN* is given below.

$\mathbf{x}$ is the observation vector and can be written as

$$\mathbf{x} = \hat{\mathbf{x}} + \hat{\mathbf{x}}^{\perp} \quad \text{(C1)}$$

where $\hat{\mathbf{x}} = \tilde{\mathbf{V}}\mathbf{x}$ and $\hat{\mathbf{x}}^{\perp} = (\mathbf{I} - \tilde{\mathbf{V}})\mathbf{x}$ with $\tilde{\mathbf{V}} = \mathbf{V}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}\mathbf{V}^{\mathbf{T}}$.

The *D*-statistic is given by

$$D_i = (\mathbf{a} - \bar{\mathbf{a}})^{\mathbf{T}}\mathbf{S}^{-1}(\mathbf{a} - \bar{\mathbf{a}})\frac{I(I-P)}{P(I^2-1)} \sim F(I, I-P) \quad \text{(C2)}$$

where $\mathbf{a} = [\mathbf{x}^{\mathbf{T}}\mathbf{V}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}]^{\mathbf{T}}$ and D is approximated by an *F*-distribution with $(I, I - P)$ degrees of freedom.

The *Q*-statistic is given by

$$Q_i = \|(\mathbf{I} - \tilde{\mathbf{V}})\mathbf{x}\|^2 = \mathbf{x}^{\mathbf{T}}(\mathbf{I} - \tilde{\mathbf{V}})^{\mathbf{T}}(\mathbf{I} - \tilde{\mathbf{V}})\mathbf{x} \sim g\chi_h^2 \quad \text{(C3)}$$

which is approximated by a weighted chi-square distribution. Combining both the D_i and Q_i results in

$$BCN = \frac{\mathbf{x}^{\mathbf{T}}\mathbf{V}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}\mathbf{S}^{-1}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}\mathbf{V}^{\mathbf{T}}\mathbf{x}}{F(I, I-R)}\frac{I(I-P)}{P(I^2-1)} + \frac{\mathbf{x}^{\mathbf{T}}(\mathbf{I} - \tilde{\mathbf{V}})^{\mathbf{T}}(\mathbf{I} - \tilde{\mathbf{V}})\mathbf{x}}{g\chi_h^2} \quad \text{(C4)}$$

$$BCN = \mathbf{x}^{\mathbf{T}}\theta\mathbf{x} \sim g\chi_h^2 \quad \text{(C5)}$$

where

$$\theta = \frac{\mathbf{V}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}\mathbf{S}^{-1}(\mathbf{V}^{\mathbf{T}}\mathbf{V})^{-1}\mathbf{V}^{\mathbf{T}}}{F(I, I-R)}\frac{I(I-P)}{P(I^2-1)} + \frac{(\mathbf{I} - \tilde{\mathbf{V}})^{\mathbf{T}}(\mathbf{I} - \tilde{\mathbf{V}})}{g\chi_h^2} \quad \text{(C6)}$$

The *BCN* is a quadratic function of $\mathbf{x}$; assuming $\mathbf{x}$ follows a multinormal distribution the *BCN* can be approximated by a weighted chi-square distribution.[24]

Manuscript received Oct. 22, 2003, and revision received July 15, 2004.