Chemometrics and Intelligent Laboratory Systems 28 (1995) 3-21

Chemometrics and intelligent laboratory systems

Tutorial

Process analysis, monitoring and diagnosis, using multivariate projection methods

Theodora Kourti *, John F. MacGregor

McMaster Advanced Control Consortium, Department of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada

Received 5 August 1994; accepted 15 November 1994

Abstract

Multivariate statistical methods for the analysis, monitoring and diagnosis of process operating performance are becoming more important because of the availability of on-line process computers which routinely collect measurements on large numbers of process variables. Traditional univariate control charts have been extended to multivariate quality control situations using the Hotelling T^2 statistic. Recent approaches to multivariate statistical process control which utilize not only product quality data (**Y**), but also all of the available process variable data (**X**) are based on multivariate statistical projection methods (principal component analysis, (PCA), partial least squares, (PLS), multi-block PLS and multi-way PCA). An overview of these methods and their use in the statistical process control of multivariate continuous and batch processes is presented. Applications are provided on the analysis of historical data from the catalytic cracking section of a large petroleum refinery, on the monitoring and diagnosis of a continuous polymerization process and on the monitoring of an industrial batch process.

Contents

* Corresponding author.

SSDI 0169-7439(94)00079-4

1. Introduction

The objective of statistical process control (SPC) is to monitor the performance of a process over time to verify that it is remaining in a 'state of statistical control'. Such a state of control is said to exist if certain process or product variables remain close to their desired values and the only source of variation is 'common-cause' variation, that is, variation which affects the process all the time and is essentially unavoidable within the current process.

Traditionally, SPC charts (Shewhart, CUSUM and EWMA) are used to monitor a small number of key product variables (Y) in order to detect the occurrence of any event having a 'special' or 'assignable' cause. By finding assignable causes, long term improvements in the process and in product quality can be achieved by eliminating the causes or improving the process or its operating procedures. However, monitoring only a few quality variables is totally inadequate for most modern process industries. The traditional SPC approaches ignore the fact that with computers hooked up to nearly every industrial process, massive amounts of data are being collected routinely every few seconds on many process variables (X), such as temperatures, pressure, flow rates, etc. Final product quality variables (Y), such as polymer properties, gasoline octane numbers, etc., are available on a much less frequency basis, usually from off-line laboratory analysis. All such data should be used to extract information in any effective scheme for monitoring and diagnosing operating performance. However, all these variables are not independent of one another. Only a few underlying events are driving a process at any time, and all these measurements are simply different reflections of these same underlying events. Therefore, examining them one variable at a time as though they were independent, makes interpretation and diagnosis extremely difficult. Such methods only look at the magnitude of the deviation in each variable independently of all others. Only multivariate methods that treat all the data simultaneously can also extract information on the directionality of the process variations, that is on how all the variables are behaving relative to one another. Furthermore, when important events occur in processes they are often difficult to detect because the signal to noise ratio is very low in each variable. But multivariate methods can extract confirming information from observations on many variables and can reduce the noise levels through averaging.

The application of multivariate projection methods, such as principal component analysis (PCA), partial least squares (PLS), multi-block PLS and multi-way PCA to process monitoring and fault diagnosis is reported here. Similarities and differences with the traditional methods are discussed. The use of the projection methods for analyzing and interpreting historical plant operating records available in computer data bases is illustrated with an example from a large petroleum refinery. On-line monitoring and diagnosis of process operating performance in continuous processes (using PLS and multi-block

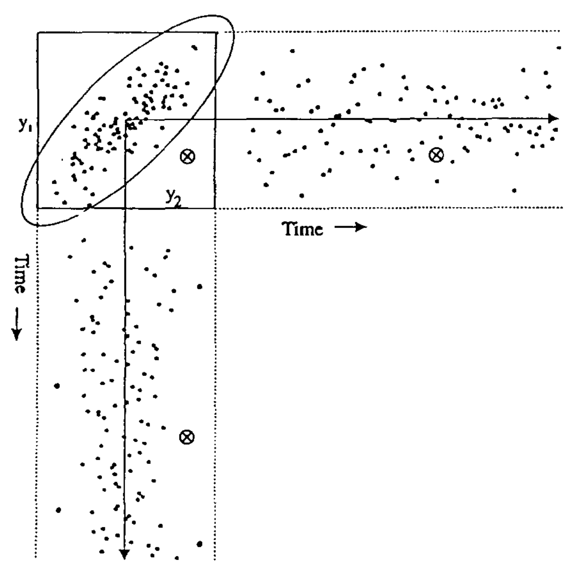

Fig. 1. Quality control of two variables — the misleading nature of univariate charts.

PLS) and batch processes (using multi-way PCA) is presented and illustrated with a continuous polymerization process and an industrial batch process.

2. Multivariate methods for monitoring product quality

Statistical process control charts such as the Shewhart chart [1], the CUSUM plot [2] and the EWMA chart [3], are well established statistical procedures for monitoring stable univariate processes. A Shewhart chart consists of plotting the observations sequentially on a graph which also contains the target value and upper and lower control limits. If an observation exceeds the control limits a statistically significant deviation from normal operation is deemed to have occurred, which triggers the search for an assignable cause. The control limits are usually determined by analyzing the variability in a reference set of process data collected when only normal or 'common cause' variability is present and acceptable operation is achieved. The limits are then usually set at plus and minus three standard deviations about the target.

In most industries, traditional univariate control charts (Shewhart, CUSUM and EWMA) are used to separately monitor key measurements on the final product which in some way define the quality of that product. The difficulty with this approach is that these quality variables are not independent of one another, nor does any one of them adequately define product quality by itself. Product quality is only defined by the correct simultaneous values of all the measured properties, that is, it is a multivariate property.

The difficulty with using independent univariate control charts can be illustrated by reference to Fig. 1. Here only two quality variables (y_1, y_2) are considered for ease of illustration. Suppose that, when the process is in a state of statistical control where only common cause variation is present, y_1 and y_2 follow a multivariate normal distribution and are correlated ($\rho_{y_1 y_2} = 0.8$) as illustrated in the joint plot of y_1 vs. y_2 in Fig. 1. The ellipse represents a contour for the in-control process, and the dots represent a set of observations from this distribution. The same observations are also plotted in Fig. 1 as individual Shewhart charts on y_1 and y_2 vs. time with their corresponding control limits. Note that by inspection of each of the individual Shewhart charts the process appears to be clearly in a state of statistical control, and none of the individual observations give any indication of a problem. The only indication of any difficulty is that a customer has complained about the performance of the product corresponding to the ⊗ in Fig. 1. If only univariate charts were used, one would clearly be confused. The same customer apparently liked all the other lots of product sent to him, many of them with values of y_1 and y_2 much further from target. The true situation is only revealed in the multivariate y_1 vs. y_2 plot where it is seen that the lot of product indicated by the ⊗ is clearly outside the joint confidence region, and is clearly different from the normal 'in-control' population of product.

In spite of the misleading nature of univariate quality control charts they continue to be almost the only form of monitoring used by industry. However, several multivariate extensions of the Shewhart, CUSUM and EWMA based on Hotelling's T^2 statistic have been proposed in the literature (see review articles by Wierda [4] and Sparks [5]).

2.1. Traditional multivariate quality control charts

Natural extensions of the Shewhart chart to situations where one observes a vector of k variables $y_{k \times 1}$ at each time period are the multivariate χ^2 and T^2 charts. The T^2 chart has its origins in the work of Hotelling [6], and several references [7–12] discuss the charts in more detail.

Given a $(k \times 1)$ vector of measurements $\boldsymbol{y}$ on k normally distributed variables with an in-control covariance matrix $\boldsymbol{\Sigma}$ one can test whether the mean $\boldsymbol{\mu}$ of these variables is at its desired target $\boldsymbol{\tau}$ by computing the statistic

$$\chi^2 = (\boldsymbol{y} - \boldsymbol{\tau})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\tau}) \quad (1)$$

This statistic will be distributed as a central χ^2 distribution with k degrees of freedom if $\boldsymbol{\mu} = \boldsymbol{\tau}$. A multivariate χ^2 control chart can be constructed by plotting χ^2 vs. time with an upper control limit (UCL) given by $\chi^2_\alpha(k)$ where α is an appropriate level of significance for performing the test (e.g. $\alpha = 0.01$).

Note that this multivariate test overcomes the difficulty illustrated in the example of Fig. 1, where univariate charts were incapable of detecting the special event denoted by ⊗. The χ^2 statistic in Eq. (1) represents the directed or weighted distance (Mahalanobis distance) of any point from the target $\boldsymbol{\tau}$. All points lying on the ellipse in Fig. 1 would have the same value of χ^2. (The ellipse is the solution to Eq. (1) for $\chi^2 = \chi_\alpha(k)$, for two variables). Hence, a χ^2 chart would detect as a special event any point lying outside of the ellipse.

When the in-control covariance matrix $\boldsymbol{\Sigma}$ is not known, it must be estimated from a sample of n past multivariate observations as

$$\mathbf{S} = (n-1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})^{\mathrm{T}} \quad (2)$$

When new multivariate observations (y) are obtained, then Hotelling's T^2 statistic given by

$$T^2 = (y - \boldsymbol{\tau})^{\mathrm{T}} \mathbf{S}^{-1} (y - \boldsymbol{\tau}) \quad (3)$$

can be plotted against time. An upper control limit (UCL) on this chart is given by:

$$T^2_{\mathrm{UCL}} = \frac{(n-1)(n+1)k}{n(n-k)} F_\alpha(k, n-k) \quad (4)$$

where $F_\alpha(k,\ n-k)$ is the upper $100\alpha\%$ critical point of the F distribution with k and $n-k$ degrees of freedom [13].

The above charts are for a single new multivariate observation vector at each time. If an average of m new multivariate observations are to be used at each time or if the estimate of the variance $\mathbf{S}$ is based on pooling estimates from rational subgroups, then the above definitions of the χ^2 and T^2 charts and their UCLs must be correspondingly redefined [4]. Furthermore, if the charts are utilized to examine past data that are also used in computing $\mathbf{S}$, then the distributional properties of T^2 are different from the above [4,13].

Alternatively, other types of multivariate charts, such as multivariate CUSUM and multivariate EWMA charts may be used [4,5].

The above ideas are illustrated here by monitoring the properties of low-density polyethylene produced in a multi-zone tubular reactor (we consider here the first two zones). Details on this simulated process can be found in MacGregor et al. [14]. The operating conditions in the reactor influence the molecular properties of the polymer produced (weight and

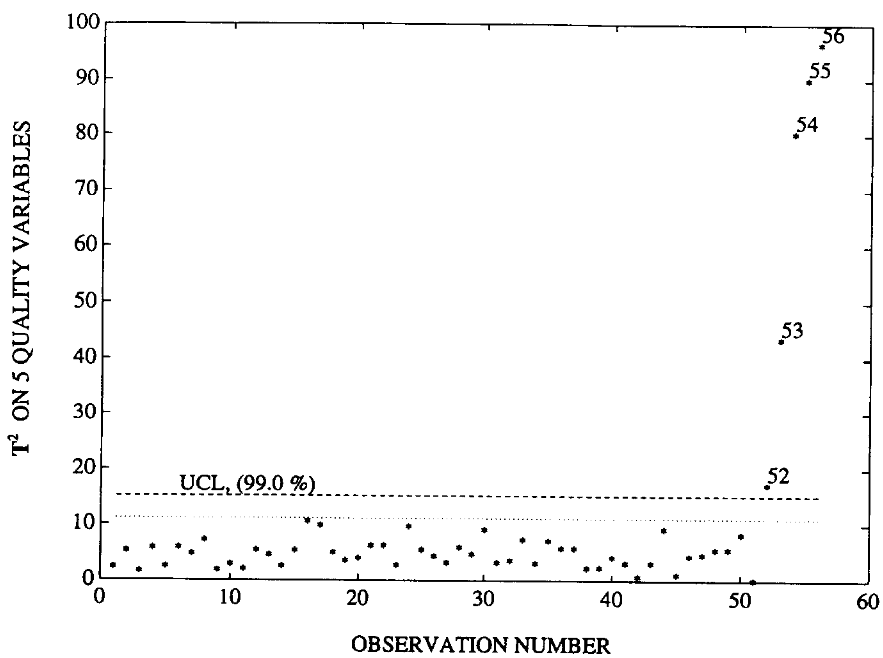

Fig. 2. T^2 chart on five product properties of polyethylene.

number average molecular weights, MW_w and MW_n, and long and short chain branching, LCB and SCB) and these in turn affect the behaviour of the polymer in its final application. The productivity variable of interest is the conversion per pass, CONV. The five product variables (Y) of interest (MW_w, MW_n, LCB, SCB and CONV) are monitored with a T^2 chart (Fig. 2). The unnumbered observations in this chart were obtained by simulating normal operating conditions; the numbered observations correspond to simulated problematic operation caused by increasing levels of fouling in the first zone of the reactor. The dashed line (– – –) corresponds to a 99% limit and the dotted line (· · ·) indicates the 95% limit. Notice that the onset of fouling had an effect on the product, and that this effect could be detected in all cases, by following the T^2 calculated from the product quality properties. Had we used univariate charts, points 52 and 53 would have been missed. This is because, for these points, the individual values for the five Y variables are within the expected limits of the corresponding univariate charts; however, the values of these five variables relative to each other are not justified by the correlation structure of the $\mathbf{Y}$ matrix (determined under normal operating conditions), and this was detected by the multivariate charts.

2.2. Quality control charts based on principal components

When the number of measured quality variables (k) is large, one often finds that they are highly correlated with one another and their covariance matrix $\boldsymbol{\Sigma}$ is nearly singular. A common procedure for reducing the dimensionality of the quality variable space is principal component analysis (PCA) [11,15,16]. The first principal component (PC) of $\boldsymbol{y}$ is defined as that linear combination $t_1 = \boldsymbol{p}_1^{\mathrm{T}} \boldsymbol{y}$ that has maximum variance subject to $|\boldsymbol{p}_1| = 1$. The second PC is that linear combination defined by $t_2 = \boldsymbol{p}_2^{\mathrm{T}} \boldsymbol{y}$ which has next greatest variance subject to $|\boldsymbol{p}_2| = 1$, and subject to the condition that it be uncorrelated with (orthogonal to) the first PC (t_1). Additional PCs up to k are similarly defined. In effect PCA decomposes the observation matrix $\mathbf{Y}$ as:

$$\mathbf{Y} = \mathbf{T}\mathbf{P}^{\mathrm{T}} = \sum_{i=1}^{k} \boldsymbol{t}_i \boldsymbol{p}_i^{\mathrm{T}} \tag{5}$$

PCA is scale dependent, and so the $\mathbf{Y}$ matrix must be scaled in some meaningful way. The most usual form of scaling is to scale all variables to unit variance and then perform PCA on the correlation matrix. Alternatively, in quality control situations, scaling the Ys inversely proportional to their specification limits or some other measure of relative importance is usually more meaningful.

In practice, one rarely needs to compute all the k principal components, since most of the variability in the data is captured in the first few principle components. The NIPALS algorithm [16] is ideal for computing the principal components in a sequential manner when the number of variables is large. The number of PCs that provide an adequate description of the data can be assessed using a number of methods [11] with cross-validation [17] being perhaps the most reliable. By retaining only the first A PCs the $\mathbf{Y}$ matrix is approximated by:

$$\hat{\mathbf{Y}} = \sum_{i=1}^{A} \boldsymbol{t}_i \boldsymbol{p}_i^{\mathrm{T}} \tag{6}$$

In practice 2 or 3 PCs are often sufficient to explain most of the predictable variations in the process.

Having established a PCA model based on historical data collected when only common cause variation was present, future behaviour can be referenced against this 'in-control' model. New multivariate observations can be projected onto the plane defined by the PCA loading vectors to obtain their scores ($t_{i,\mathrm{new}} = \boldsymbol{p}_i^{\mathrm{T}} \boldsymbol{y}_{\mathrm{new}}$), and the residuals $\boldsymbol{e}_{\mathrm{new}} = \boldsymbol{y}_{\mathrm{new}} - \hat{\boldsymbol{y}}_{\mathrm{new}}$, where $\hat{\boldsymbol{y}}_{\mathrm{new}} = \mathbf{P}_A \boldsymbol{t}_{A,new}$, and $\boldsymbol{t}_{A,\mathrm{new}}$ is the $(A \times 1)$ vector of scores from the model and $\mathbf{P}_A$ is the $(k \times A)$ matrix of loadings. Multivariate control charts based on Hotelling's T^2 can be plotted based on the first A PCs, where

$$T_A^2 = \sum_{i=1}^{A} \frac{t_i^2}{s_{t_i}^2} \tag{7}$$

and $s_{t_i}^2$ is the estimated variance of t_i. If $A = 2$, a joint t_1 vs. t_2 plot can be used.

Note that the traditional Hotelling T^2 in Eq. (3) is equivalent [15,18] to

$$T^2 = \sum_{i=1}^{A} \frac{t_i^2}{s_{t_i}^2} + \sum_{i=A+1}^{k} \frac{t_i^2}{s_{t_i}^2} \tag{8}$$

By scaling each t_i^2 by the reciprocal of its variance, each PC term plays an equal role in the computation of T^2 irrespective of the amount of variance it explains in the **Y** matrix. This illustrates some of the problems with using T^2 when the variables are highly correlated and $\boldsymbol{\Sigma}$ is very ill-conditioned. When the number of variables (k) is large, $\boldsymbol{\Sigma}$ is often singular and cannot be inverted. Even if it can, the last PCs ($i = A + 1, \ldots, k$) in Eq. (8) explain very little of the variance of **Y** and generally represent random noise. By dividing these t_is by their very small variances, slight deviations in these t_is which have almost no effect on **Y** will lead to an out-of-control signal in T^2. Therefore, T_A^2 based on the first A (cross-validated) PCs provides a test for deviations in the product quality variables that are of greatest importance to the variance of **Y**.

However, monitoring product quality via T_A^2 based on the first A PCs is not sufficient. This will only detect whether or not the variation in the quality variables in the plane of the first A PCs is greater than can be explained by common cause. If a totally new type of special event occurs which was not present in the reference data used to develop the in-control PCA model, then new PCs will appear and the new observation y_{new} will move off the plane. Such new events can be detected by computing the squared prediction error (SPE_y) of the residuals of a new observation [19].

$$SPE_y = \sum_{i=1}^{k} (y_{new,i} - \hat{y}_{new,i})^2 \qquad (9)$$

This is also often referred to as the Q statistic [11] or distance to the model. It represents the squared perpendicular distance of a new multivariate observation from the projection space. When the process is 'in-control', this value of SPE_y or Q should be small. Upper control limits for this statistic can be computed, from historical data, using approximate results for the distribution of quadratic forms [11,20]. A very effective set of multivariate control charts is therefore a T^2 chart on the A dominant orthogonal PCs ($t_1, \ldots, t_A$) plus a SPE_y chart.

3. Multivariate methods for process monitoring

So far, statistical quality control (SQC) methods based only on product quality data (**Y**) have been discussed. This use of only product quality data has been the common approach to quality control methods developed throughout the statistical literature. However, in these approaches, all of the data on the process variables (**X**) are being ignored. If one truly wants to do statistical process control (SPC), one must look at all of these process data as well. There are often hundreds of process variables, and they are measured much more frequently and usually more accurately than the product quality data (**Y**). Furthermore, any special events which occur will also have their fingerprints in the process data (**X**). Sometimes product quality is only determined by the performance of the product later, in another process (i.e., catalyst conditioning); it would be useful to know if the product is good before using it; monitoring the process would help detect problems during production that may lead to a questionable product.

There are several other reasons why monitoring the process is advantageous. Sometimes, only a few properties of the product are measured, but these are not sufficient to define entirely the product quality. For example, only the relative viscosity (RV) is charted in nylon production, although there are other properties (amine end groups) that affect the dye properties of the product. If process problems that affect amine groups occur, they will not be detected by following the RV only. In these cases the process data may contain more information about events with special causes that may affect the product quality (product performance).

Finally, even if product quality measurements are frequently available, monitoring the process may help in diagnosing assignable causes for an event. When monitoring product quality, even if we determine which quality variable caused the multivariate chart to go out of limits, it may still be difficult to determine what went wrong in the process. For example, in the LDPE process by following the product variables it was determined that for point 56, SCB is the major contributor to the out-of-control signal. However, there may be several reasons (combinations of process conditions) that might have caused this property to change. Monitoring the process would bring us closer to the answer as will be demonstrated later.

Certainly one could apply the previously discussed SQC charting methods directly to the process variables (**X**) as well [18]. However, as discussed previously, with large numbers of highly correlated variables, these methods are impractical. Furthermore, they offer no way of relating the **X** and **Y** data, and least squares regression analysis is also impractical in this situation. Another problem is that

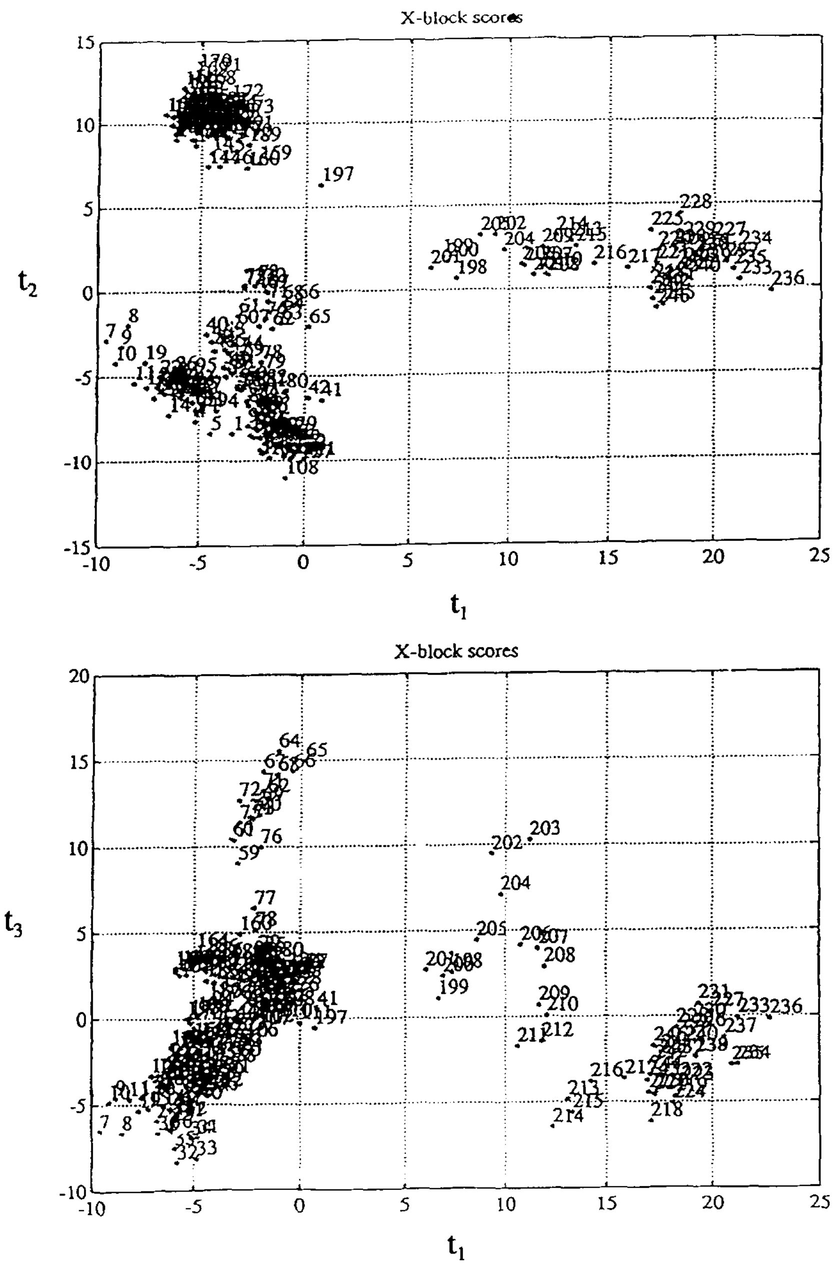

Fig. 3. PLS score plots for 15 days of operation of the catalytic cracking and fractionation section of a refinery (top: t_1 vs. t_2; bottom: t_1 vs. t_3).

these methods cannot handle missing data; sensor failure is a common problem in the process industries. The only practical approaches to multivariate SPC appear to be those based on multivariate statistical projection methods such as PCA and PLS (projection to latent structures or partial least squares). The methods are ideal for handling the large number of highly correlated and noisy process variable measurements that are being collected by process computers on a routine basis; these methods can also handle missing data. PCA has already been described, and a brief overview of PLS follows.

3.1. PLS — partial least squares

Given two matrices, an $(n \times m)$ process variable data matrix $\mathbf{X}$, and an $(n \times k)$ matrix of corresponding product quality data $\mathbf{Y}$, one would like to extract latent variables that not only explain the variation in the process data ($\mathbf{X}$), but that variation in $\mathbf{X}$ which is most predictive of the product quality data ($\mathbf{Y}$). PLS is a method (or really a class of methods) which accomplishes this by working on the sample covariance matrix $(\mathbf{X}^T\mathbf{Y})(\mathbf{Y}^T\mathbf{X})$. In the most common version of PLS [21,22], the first PLS latent variable $t_1 = \boldsymbol{w}_1^T\boldsymbol{x}$ is that linear combination of the x variables that maximizes the covariance between it and the $\mathbf{Y}$ space. The first PLS loading vector $\boldsymbol{w}_1$ is the first eigenvector of the sample covariance matrix $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$. Once the scores $\boldsymbol{t}_1 = \mathbf{X}\boldsymbol{w}_1$ for the first component have been computed the columns of $\mathbf{X}$ are regressed on $\boldsymbol{t}_1$ to give a regression vector $\boldsymbol{p}_1 = \mathbf{X}\boldsymbol{t}_1/\boldsymbol{t}_1^T\boldsymbol{t}_1$ and the $\mathbf{X}$ matrix is deflated to give residuals $\mathbf{X}_2 = \mathbf{X} - \boldsymbol{t}_1\boldsymbol{p}_1^T$. The second latent variable is then computed as $t_2 = \boldsymbol{w}_2^T\boldsymbol{x}$ where $\boldsymbol{w}_2$ is the first eigenvector of $\mathbf{X}_2^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}_2$ and so on. As in PCA the new latent vectors or scores $(\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots)$ and the loading vectors $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots)$ are orthogonal. For large ill-conditioned data sets, it is usually convenient to calculate the PLS latent variables sequentially via the NIPALS algorithm [21] and to stop based on cross-validation criteria.

3.2. Analysis of historical process data sets

With process computers hooked up to most industrial processes, massive amounts of process data are being collected and stored in data bases. Very little analysis and interpretation of these data are being performed because of the overwhelming size of the data bases and because of the very ill-conditioned nature of the routine operating data being collected. Furthermore, the signal to noise ratio is often poor in these data, and there are often significant amounts of missing data. However, all these problems are well addressed by the multivariate statistical projection methods of PCA and PLS. By examining the behaviour of the process data in the projection spaces defined by the small number of latent variables $(t_1, t_2, \ldots, t_A)$, and interpreting process movements in this reduced space by examining the corresponding space defined by the loading vectors $(\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_A)$, or $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_A)$ in the case of PLS, it is often possible to extract very useful information from these data bases, and to use this information to improve the process. Some early notable attempts at using these approaches for the analysis and interpretation of data bases are the works of Denney et al. [23] on a sulphur recovery unit, and Moteki and Arai [24] on a low-density polyethylene process. The latter work is particularly notable in that the analysis was able to lead them quickly to process conditions that yielded desired new lamination grades and injection grades of polyethylene.

Slama [25] used PCA and PLS to analyze data on more than 300 process variables and 11 product grades from the fluidized bed catalytic cracking and fractionation section of a refinery. The difficulty with such massive data sets is first to find out where in the data there is useful information. The projections of hourly average data from fifteen days of continuous operation into the planes t_1–t_2 and t_1–t_3 defined by the first three latent variables are shown in Fig. 3. The data appears to cluster into about five distinct regions, operating in a stable manner for several days at each condition before shifting to another region. There is very little information about the process within each stable data cluster. However, by focusing attention on the transitions between the regions at time periods 58–59, 76–77, 110, 197 and 212–213 we can probably learn most of what there is to know about the 15-day period of operation. To help diagnose the reasons for these shifts in process operation, one can interrogate the underlying multivariate model (as discussed below in Section 3.4) and display the process variable contributions to these shifts.

There are several interesting examples of using these methods to analyze process data. Wise et al. [26] applied PCA to analyze and diagnose systematic variations in the behaviour of a slurry-fed ceramic

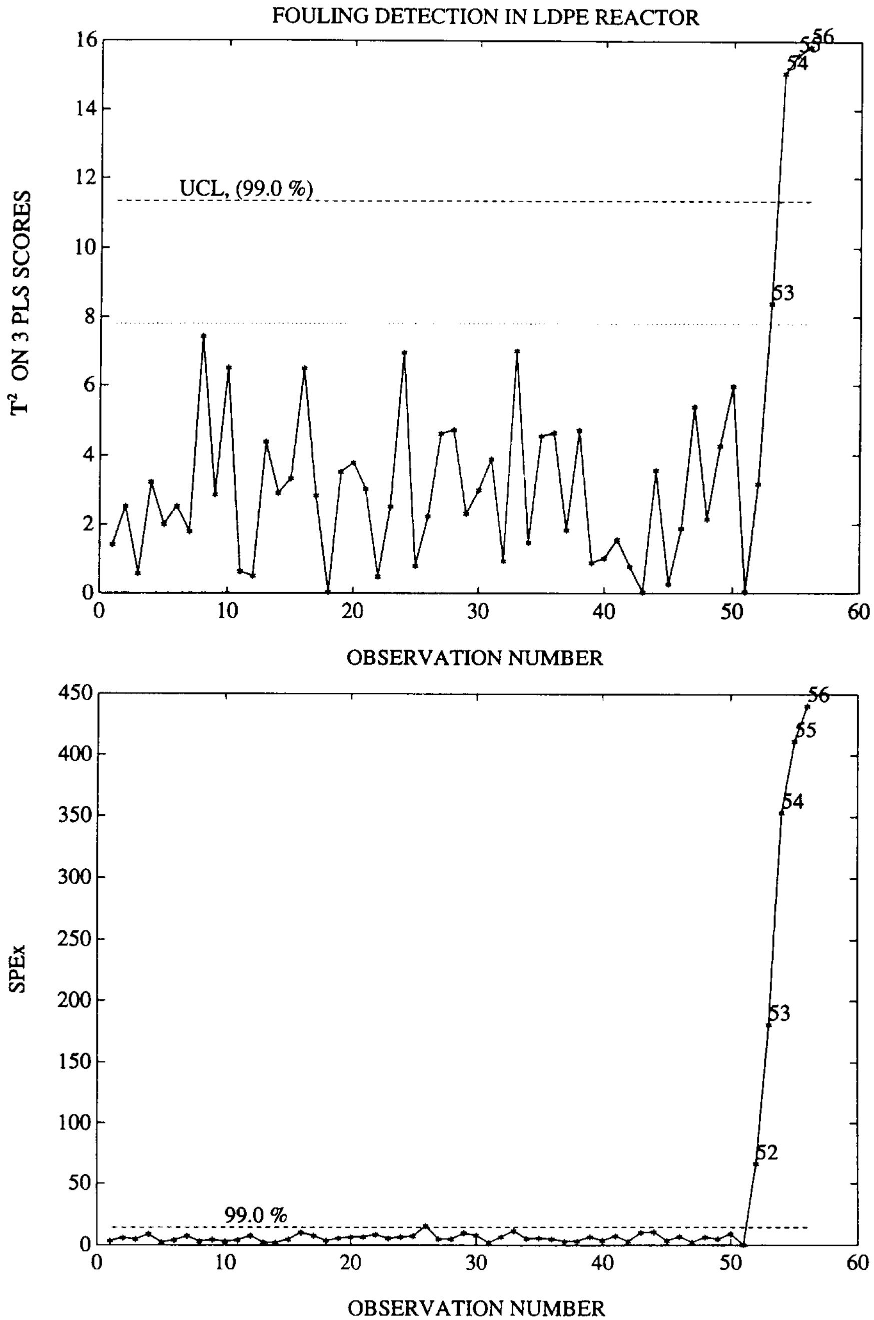

Fig. 4. T^2 chart on three PLS scores and squared prediction error (SPE_x) chart for monitoring the LDPE process. Points 52–56 denote a period where fouling occurred in the first zone of the reactor.

melter process. Skagerberg et al. [27] applied PLS to predict polymer properties from measured temperature profiles in a tubular low-density polyethylene reactor and to interpret the behaviour of this process. Hodouin et al. [28] used PCA and PLS to analyze and interpret the behaviour of mineral flotation and grinding circuits in a large mineral processing plant. Dayal et al. [29] used PLS to model the dynamic behaviour of a continuous Kamyr digester in a pulp mill, and diagnosed the reasons for poor control of Kappa number by examining the loading plots (w_1, w_2).

3.3. Monitoring continuous processes

Although the analysis of historical data bases is an important first step towards process improvement, establishing multivariate control charts to detect special events as they occur, and to diagnose possible causes for them while the information is fresh, is an essential part of SPC. The philosophy applied in developing multivariate SPC procedures based on projection methods is the same as that used for the univariate or multivariate Shewhart charts. An appropriate reference set is chosen which defines the normal operating conditions for a particular process. In other words, a PCA or PLS model must be built based on data collected from various periods of plant operation when performance was good. Any periods containing variations arising from special events that one would like to detect in the future are omitted at this stage. The choice of the reference set is critical to the successful application of the procedure as discussed in Kresta et al. [19].

When the data are serially autocorrelated the **X** and **Y** matrices can be augmented with time-lagged values, in order to account for the dynamics of the process and the disturbances. Multivariate time series analysis is discussed for PCA by Jollife [30], and Jackson [11] and for PLS by Wold et al. [31] and MacGregor et al. [32]. Dead times between variables are accounted for by time shifting. An industrial example where plant data had been both time shifted to account for dead times between X and Y, and lagged to account for autocorrelations in Y, is described in Dayal et al. [29].

The multivariate control chart is now a T^2 chart

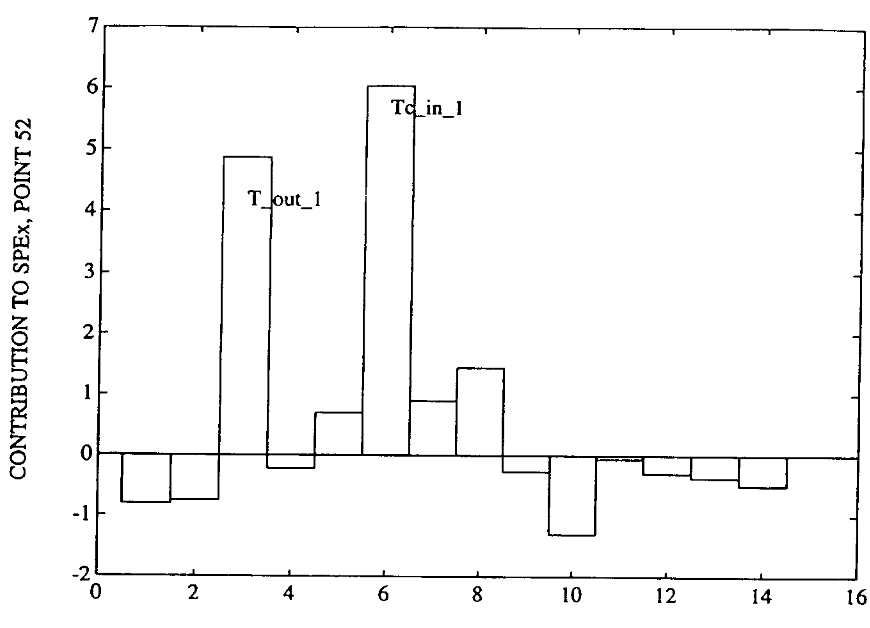

Fig. 5. Contribution plot showing the process variable contributions to the SPE_x, for point 52.

on the first A latent variables (Eq. (7)). Added to this, is a chart on SPE_x where

$$SPE_x = \sum_{i=1}^{m} (x_{new,i} - \hat{x}_{new,i})^2 \quad (10)$$

where $\hat{\boldsymbol{x}}_{new}$ is computed from the reference PLS or PCA model. This latter plot will detect the occurrence of any new events which cause the process to move away from the hyperplane defined by the reference model. Control limits for the T^2 charts are chosen in the same manner as previously discussed, and the UCL on SPE_x is based on the χ^2 approximation (Q statistic [11,20]).

The main concepts behind the development and use of these multivariate SPC charts for monitoring continuous processes were laid out by Kresta et al. [19], Wise et al. [26], Wise and Ricker [33], and MacGregor et al. [34,35]. Several illustrations of the methods were also presented in those papers along with the algorithms and details on estimating control limits.

To illustrate the basic approach, consider the monitoring of the simulated multi-section high-pressure tubular reactor process for the manufacture of low-density polyethylene (LDPE) [14]. The reaction kinetics and the fundamental modelling of this LDPE process can be found in a review by Kiparissides et al. [36]. Measurements are available on a frequent basis on all process variables (X) — reactor temperature profiles in each section, feed rates on all component streams, cooling system flows and temperatures, and pressures in each reactor section. Every hour or so, measurements are available on product quality and productivity (Y) — polymer molecular weights and branching properties, and conversion of monomer to polymer. Using data collected (**X**, **Y**) when the process was operating well, and no special events were present, a PLS model using only three latent variables ($A = 3$) was able to explain 90.0% of the variation in the **Y** data. Fig. 4 illustrates the use of a T^2 chart (on three PLS scores) and an SPE_x chart to monitor the behaviour of the reactor when there is an increasing level of fouling in the first section of the reactor. Unnumbered points indicate past conditions of normal operation. Fouling starts at point 52. Notice that the squared prediction error plot quickly detected the onset of this special event and alarmed an out-of-control situation, on-line, before laboratory data on product quality became available. The T^2 plot signalled later. As already discussed, the two plots are complementary in detecting special events; both of them are required for proper monitoring.

3.4. Diagnosing assignable causes

Both univariate and multivariate SPC charts are based on statistical tests to detect any deviations from the in-control reference distribution upon which the models and charts have been built. In classical quality control approaches which chart only quality variables (Y), once an out-of-control signal has been given, it is then left up to the process operators and engineers to try to diagnose an assignable cause using their process knowledge and a one-at-a-time inspection of process variables. However, multivariate charts based on PLS or PCA provide a much greater capability for diagnosing assignable causes. By interrogating the underlying PLS or PCA model at the point where an event has been detected, one can extract diagnostic or contribution plots which reveal the group of process variables making the greatest contributions to the deviations in the SPE_x and the scores [33,34,37]. Although these plots will not unequivocally diagnose the cause, they will provide much greater insight into possible causes and thereby greatly narrow the search.

Consider the out-of-control alarms shown in Fig. 4 for the LDPE process. Diagnostic plots showing the contribution of the process variables to the SPE_x at point 52 are shown in Fig. 5. These contribution plots point to the temperature of the reaction mixture at the exit from zone 1 and the temperature of the cooling agent into the jacket of the first zone as being the main process variables that are showing inconsistency (by contributing significantly to the large values of SPE_x). This combination of variables would imply heat transfer problems and could lead the operator to suspect fouling.

3.5. Multi-block PLS

When a large number of variables is included in the **X** space, the monitoring and diagnosing charts

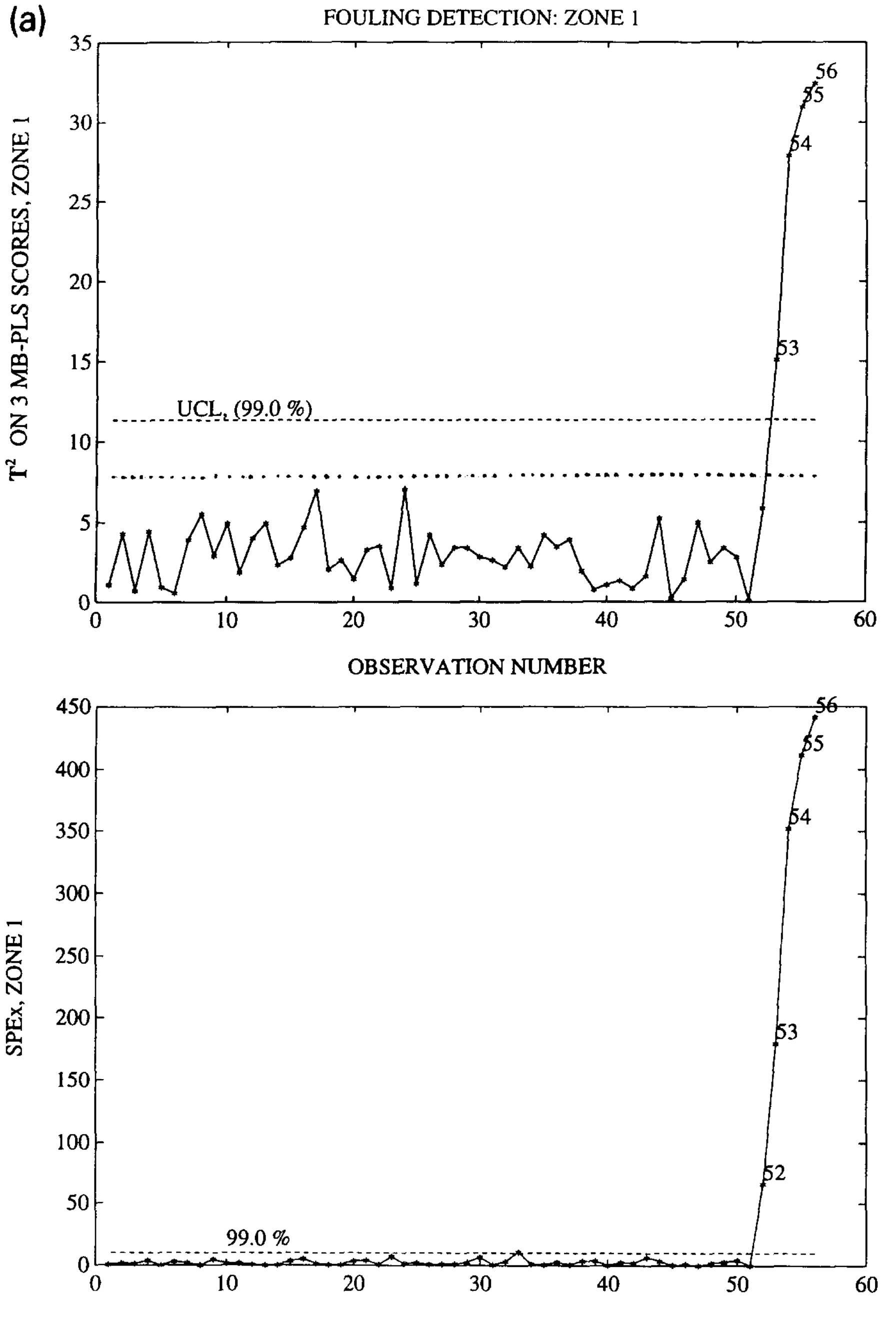

Fig. 6. Monitoring the LDPE process using multi-block PLS. The disturbance is fouling in zone 1. Monitoring of individual zones detects the problematic zone. (a) T^2 chart on three MB-PLS scores and SPE_x for the process variables of zone 1. (b) T^2 chart on three MB-PLS scores and SPE_x for the process variables of zone 2.

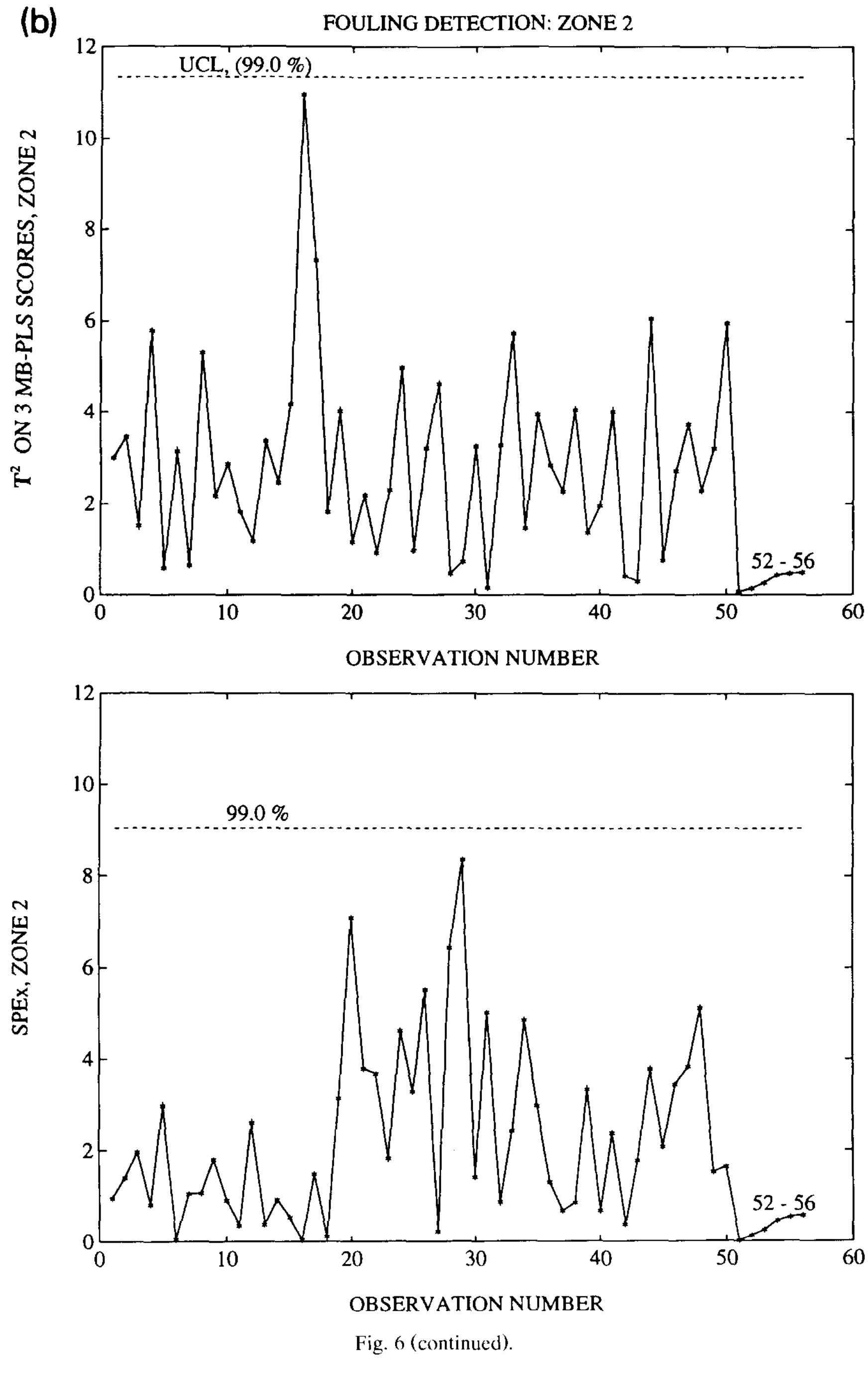

Fig. 6 (continued).

discussed in the previous sections may be difficult to interpret. The combined use of multi-block PLS (MB-PLS) and contribution plots may facilitate this task. In the MB-PLS approach, large sets of process variables (**X**) are broken into meaningful blocks; usually each block corresponds to a process unit or a section of a unit. Multivariate monitoring charts for important subsections of the plant as well as for the entire process can be constructed. The principles behind multi-block data analysis methods and their algorithms can be found in Wold [38] and Wangen and Kowalski [39]. MacGregor et al. [14] discuss an application of MB-PLS to process monitoring and diagnosing for the LDPE reactor. Each block corresponds to one zone. Plots of t_1 vs. t_2 and SPE_x obtained for each block of the process were utilized to detect an abnormal event in the zone it occurred; then contribution plots were successfully used to assign causes for it. When the number of latent variables used for modelling is more than two, one should combine the information of the scores in a statistic, rather than plotting scores pair-wise. The following example demonstrates a monitoring procedure utilizing a chart of T^2 calculated from the scores used for the MB-PLS model and a SPE_x chart.

Fig. 6a gives the T^2 chart on three scores (calculated from scores t_1, t_2 and t_3 of block 1) and the SPE_x chart for block 1 (corresponding to zone 1) for the same simulated process conditions of Fig. 4 (fouling in zone 1). Notice that by monitoring block 1 (zone 1, only) problems are detected for observations 52–56. Fig. 6b gives the T^2 chart on three scores and the SPE_x chart for block 2 (zone 2), for the same simulated fouling case. Notice that no problems were detected in zone 2 for observations 52–56. MB-PLS successfully detected that the problem is in zone 1 and that zone 2 operates normally. Utilizing the contribution plots for fault diagnosis (on the scores and SPE_x of zone 1) has revealed that process variables with unusual values were the temperature of the reacting mixture at the exit of zone 1 and the temperature of the cooling agent in zone 1. Although the monitoring and diagnosis procedures based on MB-PLS and PLS gave comparable results for this system with only 14 process variables, MB-PLS offers an advantage when larger systems with tens or hundreds of variables are involved.

3.6. Monitoring batch processes

Recent trends in most industrialized countries have been towards the manufacture of higher value added specialty chemicals (specialty polymers, pharmaceuticals and biochemicals) that are produced mainly in batch reactors. There are also many other batch type operations, such as crystallization and injection molding, which are very important to the chemical and manufacturing industries. Monitoring these batch processes is very important to ensure their safe operation and to assure that they produce consistent high-quality products. The use of the multivariate statistical projection methods has been extended to the analysis and the on-line monitoring and diagnosis

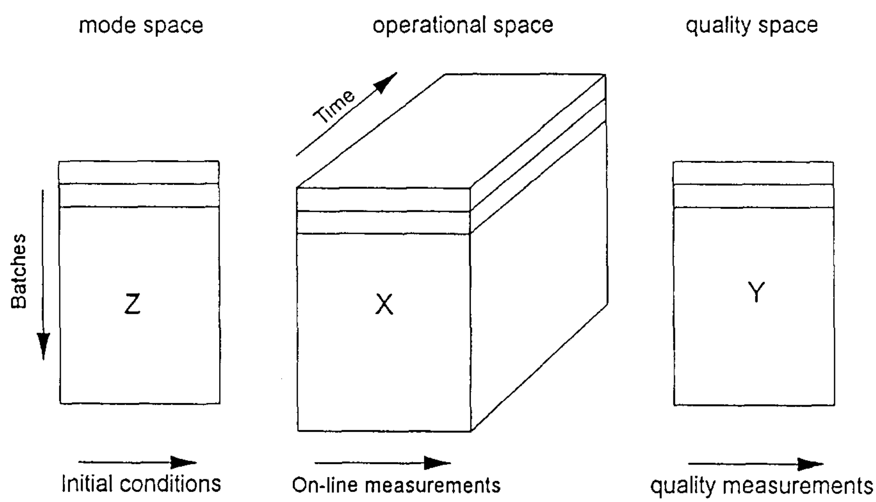

Fig. 7. Nature of batch data. The batch process is described by **X**, quality variables by **Y** and feed properties by **Z**.

of batch processes by MacGregor and Nomikos [40] and Nomikos and MacGregor [20,41,42]. Typical data from batch processes include time-varying trajectories of all the measured process variables throughout the duration of each batch (X), product quality measurements (Y) at the end of each batch, and batch recipe and charge conditions (Z) at the start of each batch. If such data are available in a historical data base on many past batches, multivariate PCA and PLS models can be developed for analyzing these historical batches and for establishing on-line SPC charts for monitoring the progress of each new batch.

The nature of the data available in a batch monitoring problem is illustrated in Fig. 7. The **X** matrix is a $(I \times J \times K)$ array, where I is the number of batch runs, J is the number of variables and K is the time intervals throughout the batch. Each horizontal slice through this array is a $(J \times K)$ matrix containing the trajectories of all the variables from a single batch. Each of its vertical slices is a $(I \times J)$ matrix representing the values of all the variables for all the batches at a common time interval (k). The final product quality measurements are taken at the end of each batch, for a few variables, L. These are summarized in the $(I \times L)$ matrix **Y**. For each batch, measured feed-stock properties and other variable initial conditions may be available; these are summarized in a matrix **Z**.

Since the process data (**X**) are now a three-dimensional array (batch run × variable × time), Nomikos and MacGregor used three-dimensional or multi-way PCA (MPCA) and PLS (MPLS) methods. Multi-way PCA and PLS methods have been discussed in a series of articles [43–46]. Nomikos and MacGregor proposed approaches for handling the fact that one dimension (time) is evolving during the progress of a new batch, and for establishing control limits on the

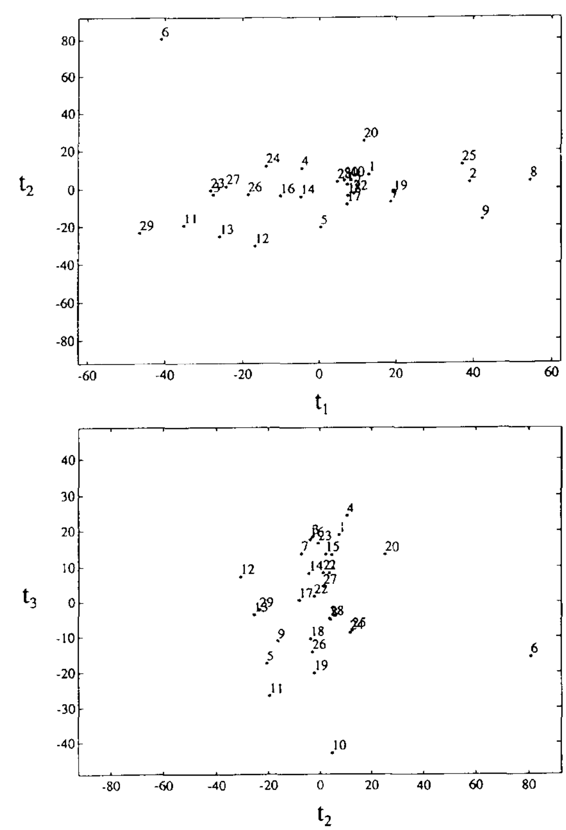

Fig. 8. Plots of t_1–t_2 and t_2–t_3 for 30 batches. Batch No. 6 is abnormal.

multivariate SPE and score plots. They transformed the three-dimensional array **X** to a two-dimensional array by unfolding **X** in such a way as to put each of its vertical slices $(I \times J)$ side by side to the right, starting with the one corresponding to the first time interval. The resulting two-dimensional matrix has size $(I \times JK)$. This unfolding allows for analyzing the variability among the batches in **X** by summarizing the information in the data with respect both to variables and their time variation. With this particular unfolding, by subtracting the mean of each column prior to performing the MPCA, one is decomposing the variation about the mean trajectories of all the variables.

The MPCA approach classifies batches as good or bad based on their similarity to a group of previous batches that produced an acceptable product. Information from quality measurements is not utilized directly. MPLS may be used to utilize information from the product quality. Once the **X** matrix has been unfolded into a two-dimensional matrix, PLS can be performed between **Y** and this new matrix, to relate the quality characteristics to the process conditions. By utilizing the quality measurements the batches may be classified in a way that they are more predictive of **Y** — in this case variables that exhibit high variability but do not affect the quality of the product are weighted less heavily; as a result, disturbances in these variables will be flagged but not cause unnecessary alarms [42]. When extra information relevant to the batch process is available (in the form of matrix **Z** in Fig. 7) this information may also be utilized, by performing multi-way multi-block PLS. Matrix **Z** and the unfolded **X** matrix may be treated as two blocks, weighted appropriately.

The use of MPCA to monitoring batch processes is illustrated here with an example. Data from 30 batches from an industrial process were provided. There were no product quality measurements; the quality of the batch ('good' or 'bad') was assessed from the performance of the batch product in another process, later. For each batch, the trajectories of 4 variables for 375 time intervals were provided. One of the batches was characterized as 'bad' by the company. In a preliminary analysis, MPCA was performed on all the batches (i.e., on the three-way array **X** with dimensions $30 \times 4 \times 375$), to test if the method would be able to discriminate between 'good' and 'bad' batches with the available process data; in other words to assess if the system was *observable*. Fig. 8 shows the projections of these 30 batches on the score planes (t_1–t_2 and t_2–t_3) defined by the three first principal components. It can be seen that batch No. 6 (the one characterized by the company as 'bad') is out of the main cluster (normal operating region) formed by the rest of the batches.

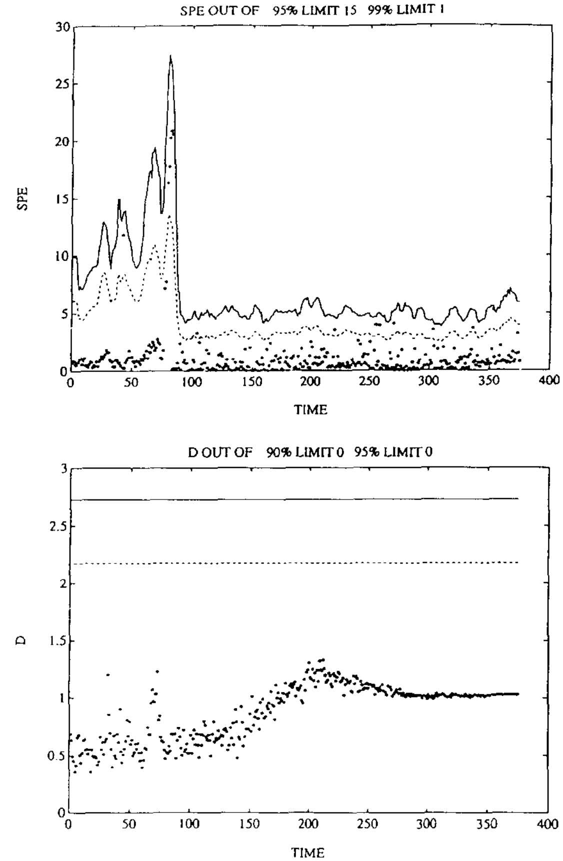

Fig. 9. Monitoring a good batch. SPE_x and D statistic.

Having established the observability of faults with the analysis of past data, a model was built to summarize the information contained in the 29 good batches about the normal operating region of the process. This model was then used as statistical reference to classify new batches as normal ('good') or abnormal ('bad'). The model was used for the classification of new batches in the way described in Nomikos and MacGregor [41]. New batches are classified by monitoring a statistic, D (essentially a

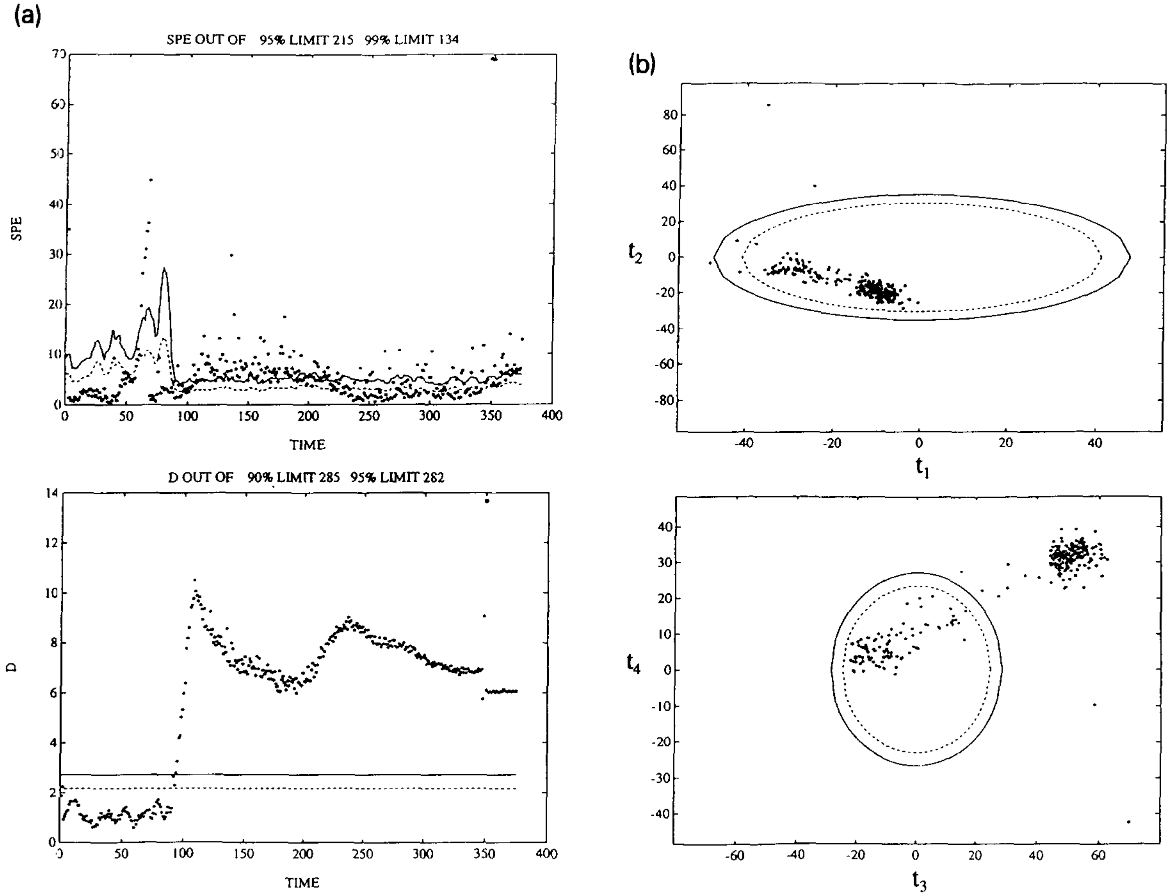

Fig. 10. Monitoring batch No. 6 (a bad batch). (a) SPE_x and D statistic. (b) Plots of t_1–t_2 and t_3–t_4 for the duration of the batch.

Hotteling T^2 calculated from the A latent variables at each time interval k), and by monitoring SPE at each time interval k.

Fig. 9 shows the SPE response as a function of time, and the D statistic of a batch that was eventually classified as 'good'. Notice that both of these quantities remain well within the control limits throughout the batch. (The solid line corresponds to a 99% limit and the dashed line to a 95% limit for SPE; at the top of the figure it is indicated that for the current batch, 15 points, out of 375, were out the 95% limit and one point out the 99% limit.) Fig. 10 shows how batch No. 6 would have behaved, had the model been in use on-line, when the data for this batch were becoming available. Fig. 10a shows the SPE behaviour and the D statistic. Notice that 215 points out of 375 are out of the 95% limit in the SPE, while the D statistic goes out of limits around 100 min into the batch run. Plots of t_1 vs. t_2 and t_3 vs. t_4 (Fig. 10b) reveal that mainly t_3 and t_4 scores show abnormalities. Indeed, individual plots of t_3 and t_4 (not shown here) revealed that these latent variables were out of limits after 100 min into the run.

The proposed monitoring charts are in accordance with the SPC requirements in that they can be easily displayed and interpreted, and they can quickly detect a fault. Furthermore, it is also possible to provide the operators with diagnostic information by interrogating the underlying MPCA, MPLS or multi-way multi-block PLS model. Other industrial applications of these methods have been reported for the analysis of historical batch data bases by Kosanovich et al. [47], and for the monitoring of a batch polymerization by Nomikos and MacGregor [20].

4. Summary

This paper has provided an overview of the concepts behind multivariate statistical process control. Justifications for treating the data in a truly multivariate manner are given. To genuinely do multivariate statistical process control (SPC) one must utilize not just the final product quality data (**Y**), but all the data on process variables (**X**) being collected routinely by process computers. SPC approaches based on multivariate statistical projection methods (PCA and PLS) have been developed for this purpose. The ideas behind these new approaches and the literature on them is reviewed. Multivariate control charts in the projection spaces provide powerful methods for both detecting out-of-control situations, and for diagnosing assignable causes, and they are applicable both to continuous and batch processes. The only requirement for applying these methods is the existence of a good data base on past operations. For this reason, they have attracted wide interest, and are rapidly being applied in many industries. Recent advances in the traditional multivariate SPC methods for monitoring and diagnosing process operating performance are reported and compared to projection method approaches in Kourti and MacGregor [48].

Acknowledgements

The authors wish to thank the 15 companies of the McMaster Advanced Control Consortium for their financial support. Also, for their active involvement in joint research projects which provide the opportunity to test new research ideas in industry.

References

[1] W.A. Shewhart, *Economic Control of Quality of Manufactured Product*, Van Nostrand, Princeton, NJ, 1931.

[2] R.H. Woodward and P.L. Goldsmith, *Cumulative Sum Techniques*, Oliver and Boyd, London, 1964.

[3] J.S. Hunter, Exponentially weighted moving average, *Journal Quality Technology*, 18 (1986) 203–210.

[4] S.J. Wierda, Multivariate statistical process control — recent results and directions for future research, *Statistica Neerlandica*, 48 (1994).

[5] R.S. Sparks, Quality control with multivariate data, *Australian Journal of Statistics*, 34 (1992) 375–390.

[6] H. Hotelling, Multivariate quality control, illustrated by the air testing of sample bombsights, in C. Eisenhart, M.W. Hastay and W.A. Wallis (Editors), *Techniques of Statistical Analysis*, McGraw-Hill, New York, 1947, pp. 113–184.

[7] F.B. Alt, *Economic Design of Control Charts for Correlated, Multivariate Observations*, Ph.D. Dissertation, Georgia Institute of Technology, 1977.

[8] F.B. Alt, Multivariate quality control, in S. Kotz and N.L. Johnson (Editors), *Encyclopedia of Statistical Sciences*, Wiley, New York, 1985, Vol. 6, pp. 110–122.

[9] F.B. Alt and N.D. Smith, Multivariate process control, in P.R. Krishnaiah and C.R. Rao (Editors), *Handbook of Statistics*, North-Holland, Amsterdam, 1988, Vol. 7, pp. 333–351.

[10] T.P. Ryan, *Statistical Methods for Quality Improvement*, Wiley, New York, 1989.

[11] J.E. Jackson, *A User's Guide to Principal Components*, Wiley, New York, 1991.

[12] C. Fuchs and Y. Benjamini, Multivariate profile charts for statistical process control, *Technometrics*, 36 (1994) 182–195.

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

[14] J.F. MacGregor, C. Jaeckle, C. Kiparissides and M. Koutoudi, Monitoring and diagnosis of process operating performance by multi-block PLS methods with an application to low density polyethylene production, *AIChE Journal*, 40 (1994) 826–838.

[15] K.V. Mardia, J.T. Kent and J.M. Bibby, *Multivariate Analysis*, Academic Press, London, 1989.

[16] S. Wold, K. Esbensen and P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 37–52.

[17] S. Wold, Cross-validatory estimation of the number of components in factor and principal components model, *Technometrics*, 20 (1978) 397–405.

[18] T. Kourti and J.F. MacGregor, Multivariate SPC methods for monitoring and diagnosing process performance, in En Sup Yoon (Editor), *PSE '94, Fifth International Symposium on Process Systems Engineering, Preprints*, 1994, Vol. II, pp. 739–746.

[19] J. Kresta, J.F. MacGregor and T.E. Marlin, Multivariate statistical monitoring of process operating performance, *Canadian Journal of Chemical Engineering*, 69 (1991) 35–47.

[20] P. Nomikos and J.F. MacGregor, Multivariate SPC charts for monitoring batch processes, *Technometrics*, 37(1) (1995).

[21] P. Geladi and B.R. Kowalski, Partial least-squares regression: a tutorial, *Analytica Chimica Acta*, 185 (1986) 1–17.

[22] A. Höskuldsson, PLS regression methods, *Journal of Chemometrics*, 2 (1988) 211–228.

[23] D.W. Denney, J. MacKay, T. MacHattie, C. Flora and E. Mastracci, Application of pattern recognition techniques to process unit data, *Canadian Society of Chemical Engineering Conference*, Sarnia, Ontario, Canada, 1985.

[24] Y. Moteki and Y. Arai, Operation planning and quality design of a polymer process, *Proceedings, IFAC Symposium, DYCORD-86, Bournemouth, UK*, 1986, pp. 159–166.

[25] C.F. Slama, *Multivariate statistical analysis of data obtained from an industrial fluidized catalytic process using PCA and PLS*, M.Eng. Thesis, Department of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada, 1991.

[26] B.M. Wise, D.J. Veltkamp, N.L. Ricker, B.R. Kowalski, S. Barnes and V. Arakali, Application of multivariate statistical process control (MSPC) to the West Valley slurry-red ceramic melter process, in Post and Wacks (Editors), *Waste Management '91 Proceedings*, University of Arizona Press, Tucson, AZ, 1991.

[27] B. Skagerberg, J.F. MacGregor and C. Kiparissides, Multivariate data analysis applied to low-density polyethylene reactors, *Chemometrics and Intelligent Laboratory Systems*, 14 (1992) 341–356.

[28] D. Hodouin, J.F. MacGregor, M. Hou and M. Franklin, Multivariate statistical analysis of mineral processing plant data, *CIM Bulletin of Mineral Processing*, 86 (1993) 23–34.

[29] B. Dayal, J.F. MacGregor, P.A. Taylor, R. Kildaw and S. Marcikic, Application of feedforward and neural networks and partial least squares regression for modelling kappa number in a continuous Kamyr digester, *Pulp and Paper Canada*, 95 (1994) 26–32.

[30] I.T. Jollife, *Principal Component Analysis*, Springer, New York, 1986.

[31] S. Wold, C. Albano, W.J. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg and M. Sjöström, Multivariate data analysis in chemistry, in B. Kowalski (Editor), *Chemometrics. Mathematics and Statistics in Chemistry*, Reidel, Dordrecht, 1984, pp. 17–95.

[32] J.F. MacGregor, T. Kourti and J. Kresta, Multivariate identification: a study of several methods, in K. Najim and J.P. Babary (Editors), *IFAC International Symposium, ADCHEM'91, Proceedings*, Paul Sabatier University, Toulouse, 1991, pp. 369–375.

[33] B.M. Wise and N.L. Ricker, Recent advances in multivariate statistical process control: improving robustness and sensitivity, in K. Najim and J.P. Barbary (Editors), *IFAC International Symposium, ADCHEM'91, Proceedings*, Paul Sabatier University, Toulouse, 1991, pp. 125–130.

[34] J.F. MacGregor, T.E. Marlin, J. Kresta and B. Skagerberg, Multivariate statistical methods in process analysis and control, *AIChE Symposium Proceedings of the Fourth International Conference on Chemical Process Control*, AIChE Publ. No. P-67, New York, 1991, pp. 79–99.

[35] J.F. MacGregor, B. Skagerberg and C. Kiparissides, Multivariate statistical process control and property inference applied to low density polyethylene reactors, in K. Najim and J.P. Barbary (Editors), *IFAC International Symposium, ADCHEM'91 Proceedings*, Paul Sabatier University, Toulouse, 1991, pp. 131–135.

[36] C. Kiparissides, G. Verros and J.F. MacGregor, Mathematical modelling, optimization and control of high pressure ethylene polymerization reactors, *Journal of Macromolecular Science, Reviews in Macromolecular Chemistry*, C33 (1993) 437.

[37] P. Miller, R.E. Swanson and C.F. Heckler, Contribution plots: the missing link in multivariate quality control, *37th Annual Fall Conference ASQC, Rochester, NY*, 1993.

[38] H. Wold, Soft modelling. The basic design and some extensions, in K.G. Joreskog and H. Wold (Editors), *Systems under Indirect Observation*, Vol. 2, North-Holland, Amsterdam, 1982, Chap. 1.

[39] L.E. Wangen and B. Kowalski, A multiblock partial least squares algorithm for investigating complex chemical systems, *Journal of Chemometrics*, 3 (1988) 3–20.

[40] J.F. MacGregor and P. Nomikos, Monitoring batch processes, in Reklaitis, Rippin, Hortacsu and Sunol (Editors), *NATO Advanced Study Institute (ASI) for Batch Processing Systems Engineering, May 29–June 7, 1992, Antalya, Turkey*, Springer, New York.

[41] P. Nomikos and J.F. MacGregor, Monitoring and batch processes using multi-way principal component analysis, *AIChE Journal*, 40 (1994) 1361–1375.

[42] P. Nomikos and J.F. MacGregor, Multi-way partial least squares in monitoring batch processes, *First International Chemometrics Internet Conference*, September 1994, to be published in *Chemometrics and Intelligent Laboratory Systems*.

[43] J. Löhmoller and H. Wold, presented at the *European Meeting of Psychometrics Society*, Groningen, The Netherlands, 1980.

[44] S. Wold, P. Geladi, K. Esbensen and J. Ohman, Multi-way principal components and PLS analysis, *Journal of Chemometrics*, 1 (1987) 41–56.

[45] P. Geladi, Analysis of multi-way (multi-mode) data, *Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 11–30.

[46] A.K. Smilde and D.A. Doornbos, Three-way methods for the calibration of chromatographic systems: comparing PARAFAC and three-way PLS, *Journal of Chemometrics*, 5 (1991) 345–360.

[47] K.A. Kosanovich, M.J. Piovoso, K.S. Dahl, J.F. MacGregor and P. Nomikos, Multi-way PCA applied to an industrial batch process, *Proceedings of the American Control Conference, Baltimore, MD, 1994*.

[48] T. Kourti and J.F. MacGregor, Recent developments in multivariate SPC methods for monitoring and diagnosing process and product performance, submitted to the *Journal of Quality Technology*, 1994.