4.02 Multivariate Statistical Process Control and Process Control, Using Latent Variables

T. Kourti, McMaster University, Hamilton, ON, Canada

4.02.1 Introduction

Databases containing process data collected in industry are becoming increasingly large. Practitioners frequently turn to these data to gain process understanding with the aim to improve the process, to establish procedures for process monitoring, and if possible to establish procedures for some form of process control in real time. The types of data required to address each one of these objectives as well as the types of models and procedures that are suitable for such purposes depend on the objective. Utilizing databases for such activities should be planned carefully to obtain scientifically viable results. In the industrial database, some of the data may have been collected by applying some form of design of experiments (DOEs) but the majority of them have been collected during routine production. As a result, most of the data are noncausal in nature. They consist of highly correlated variables with many missing measurements and low content of information in any one variable, due to the low signal-to-noise ratios. Latent variable methods can deal with such data and as a result they have received a lot of attention by industrial practitioners for troubleshooting, process understanding and improvement multivariate monitoring and with the appropriate amount of Designed Experiments for control schemes and product transfer.

Traditionally, monitoring in the form of statistical process control (SPC) has been synonymous with monitoring product quality variables or some key process variables in a univariate way. The direct result of this is the large number of control charts that are usually present in a control room and that the operators have to attend to. When there is an abnormality in the plant operation, several of these charts alarm in a short period from each other or simultaneously. This happens simply because process variables are correlated, and an abnormal event may affect more than one variable at the same time. When such a situation occurs, it is difficult for the operator to isolate and determine the source of the problem, which may lie with only one of the many correlated alarming variables, or, as it is most frequently the case, it may simply be a nonmeasurable variable (e.g., impurities, leakage, plugged pipe, or blockage of a sensor) that causes several other measured variables to go out of control. The use of latent variable methods has revolutionalized the idea of SPC for multivariate processes. The performance of an entire unit, or even a plant, can be monitored by the operator looking at only a few multivariate control charts that can be thought of as process performance indices or indices of wellness.[1–5]

Lately, latent variable methods have found their way to process control. Empirical latent variable models for the control of batch product quality have been applied in industrial problems. Work has also been reported for complicated control problems where adjustments are required for the full manipulated variable trajectories (MVTs). The process is controlled in the reduced space of a latent variable model rather in the real space of the manipulated variables.[6] The reader may want to consult Section 4.02.10, which covers multivariate statistical process control, for additional background.

Finally, work on multivariate image analysis (MIA) methods based on latent variables and their extension to online monitoring provides a breakthrough in this area and has led to the use of imaging sensors for monitoring and controlling industrial processes.[7] Color video cameras can easily be installed to monitor the state of combustion in furnaces and to detect defects and quality problems in sheet and film forming processes. Digital imagery makes it possible to monitor solids and other heterogeneous materials (e.g., pulp and paper products, polymer films, and multiphase streams).

Several companies have enthusiastically adopted the methods and have reported many success stories. Applications where multivariate statistical process control (MSPC), fault detection and diagnosis is achieved by utilizing the latent variable space, have been reported for continuous and batch processes, as well as, for process transitions as for example start-ups and restarts.[8]

The objective of this work is to provide an overview of (1) the latest developments of MSPC and its applications for fault detection and isolation (FDI) in industrial processes and (2) recent applications of latent variable methods to process control as well as image analysis for monitoring and feedback control. It provides a critical review of the methodology and describes how it is transferred to the industrial environment, and gives results applied to real industrial applications. The application studies reported in this chapter include linear, nonlinear, and time-varying continuous and batch processes. Practical issues related to transferring the methodology to industrial settings are discussed.

4.02.2 Traditional SPC Charts

SPC concepts and methods have become very important in the manufacturing and process industries. SPC can also be useful in a laboratory environment where consistency of sets of experiments is required over a period of time.

The aim of SPC is to monitor the performance of repetitive operations over time to verify that they are remaining in a state of statistical control. Such a state is said to exist if certain variables remain close to their desired values and the only source of variation is common-cause variation, that is, variation that affects the process all the time and it is essentially unavoidable within the current operation. Various control charts are used to monitor key variables in the process or the product to detect the occurrence of any event having a special or assignable cause. Long-term improvements in the process and in the product quality can be achieved by finding assignable causes and eliminating the causes.

Until recently, the usual practice in industry was to examine only a small number of variables, independently of one another, using univariate control charts (e.g., Shewhart, cumulative sum (CUSUM), and exponentially weighted moving average (EWMA)). When several variables (properties) describe the quality of a product, it is not correct to monitor them independently, as it may generate serious decision errors. In that case, multivariate control charts should be used. They can be based on the traditional approach of using the measured data directly or the data can be transformed into latent variables first. The latent variable approach allows a large number of process variables as well as product variables to be included in the process monitoring scheme.

In this section, we look at all the three approaches: univariate, traditional multivariate, and latent variable multivariate.

4.02.2.1 Univariate Charts

4.02.2.1.1 Shewhart charts

A typical control chart is shown in **Figure 1**. Data from a single variable Q are plotted in time order. The chart contains a center line (CL), which represents the average value of Q expected when the process is in control. Two other lines are the upper control limit (UCL) and the lower control limit (LCL). These are chosen such that when the process is in control, nearly all the data points fall between them. If a point falls outside the limits (observation 40 in **Figure 1**), then the process is investigated to identify the cause of the deviation. The charts are named after Dr. W. A. Shewhart of the Bell Telephone Laboratories, USA, who first proposed them.[9]

The measured variable can be a product quality variable (the molecular weight or melt index from a polymerization process) or a process variable (temperature of the reactor). The value at each point on these charts, Q, can be either the mean value of measurements on a variable for n samples $\bar{X}$-chart, or the range (R-chart) of the variable or its standard deviation (S-chart) over the n samples.

When $n = 1$, the $\bar{X}$-chart is simply a plot of the individual values of the variable. However, range and standard deviation are not defined for $n = 1$. To overcome this, the moving range or moving standard deviation can be used.[10,11]

Samples may be collected at periodic intervals (e.g., every 15 min, 1 h, or 1 day) for continuous processes, or after each batch is completed for products manufactured by batch and semibatch processes, or after several completed batches or number of parts produced when many such batches or parts are produced per day. The number of samples collected per interval depends on individual circumstances. For example,

- the molecular weight of a polymer product reported for four batches selected at random from those produced in a day ($n = 4$),
- the assay of six tablets randomly chosen from one batch ($n = 6$), and
- the chromatographic peak area of a reference sample inserted every 15 test samples ($n = 1$).

Similarly, the sampling frequency needs to be considered carefully, to allow appropriate action.

When calculating the limits for control charts from past historical data, one should be careful about the data used. Points corresponding to assignable causes should not be included as this will result in wide control limits and reduced ability to detect small shifts.

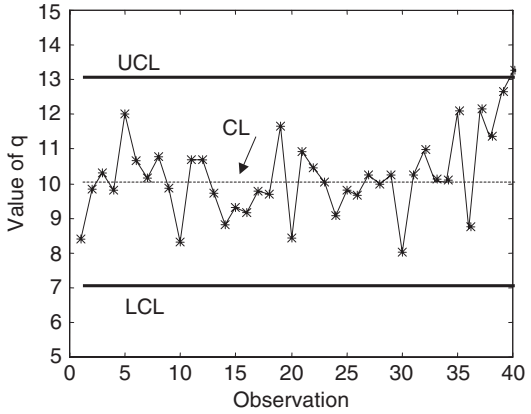

Figure 1 A typical control chart.

4.02.2.1.1(i) A general model for Shewhart charts If Q is the sample statistic ($\bar{X}$ or R) of the variable we wish to monitor, we can define μ_Q as the population mean of Q and σ_Q as the population standard deviation of Q. When σ and μ of the measurement populations are not known, they must be estimated from samples.[10]

The CL and the control limits of the chart are given as

$$\text{UCL} = \mu_Q + \lambda\sigma_Q, \quad \text{CL} = \mu_Q, \quad \text{UCL} = \mu_Q - \lambda\sigma_Q \tag{1}$$

where λ gives the distance of the control limits from the CL. A common choice is $\lambda = 3$, that is, the control limits are located at three standard deviations from the target (3σ control limits). For Q normally distributed, we expect $100(1-\alpha)\%$ of the plotted Q values to fall within the limits when the process is in control.

Very frequently, the 1σ and 2σ limits are drawn on the Shewhart chart in addition to the 3σ limits. These limits should contain 68.3 and 95.5% of Q values, respectively, if Q has a normal distribution. The additional limits allow us to do a more detailed checking of the Q values to determine unusual conditions in the process by looking for patterns in the sequences of values. Many patterns have been described, the most common being those given in the *Western Electric Handbook* rules (1954); check for an unusual event, if any of the following occurs:

1. One point plots outside 3σ control limits.
2. Two out of three consecutive points plot beyond the same 2σ limit.
3. Four out of five consecutive points plot beyond the same 1σ limit.
4. Eight consecutive points plot on one side of the CL.

Other charts are available (p-charts for percent defective items, c-charts for count data, m-charts for medians of the subgroups, etc.). Almost any statistic computed from repeated samples can be displayed as a Shewhart chart. Variable sample size is also possible, but a weighted average approach is used to calculate the mean and standard deviation for the $\bar{X}$- and S-charts.[12]

4.02.2.1.2 Univariate CUSUM charts

The Shewhart chart is relatively insensitive to shifts in the measured value that are smaller than $\sim 1.5\sigma$ when the sample size is small. In **Figure 1** for example, the process mean was shifted by 1σ after observation 30. This shift was not detected with the Shewhart chart ($n = 1$) until observation 40. The Western Electric rules may solve this problem to some extent; note that in **Figure 1**, applying rule 2 would have detected the problem at point 36, as two (34, 36) out of three consecutive points are beyond 2σ. However, applying these rules tends to increase the false alarm rate. An alternative to Shewhart chart is the CUSUM chart, which plots the cumulative sum of the deviations of the Q values from a target value. The chart is very effective for plotting individual measurement data ($n = 1$). The assumption in this chart is that observations of the variable are independent with fixed mean and constant variance.[12]

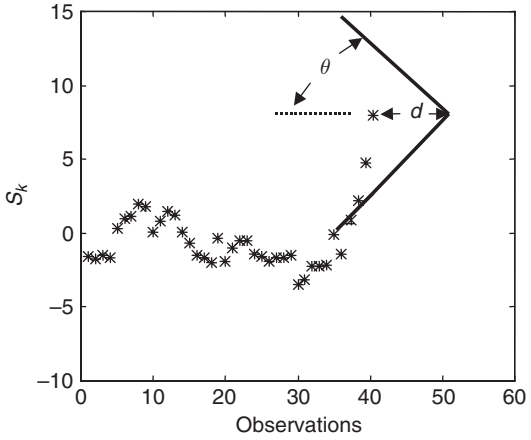

Figure 2 A CUSUM chart with V-mask.

Suppose that $\bar{x}_j$ is the average value of the jth sample (size $n \geq 1$). If the objective is to run the process such that the mean value of the property is at a target value μ, then the following quantity is plotted for the CUSUM chart, for the current sample k:

$$S_k = \sum_{j=1}^{k} (\bar{x}_j - \mu) = (\bar{x}_k - \mu) + S_{k-1} \tag{2}$$

S_k is the cumulative sum up to sample k. The CUSUM chart for the same data as in **Figure 1** is shown in **Figure 2**.

When the process is on target, the value $(\bar{x}_j - \mu)$ is a random error with mean 0 and the cumulative sum should fluctuate around 0. If the process mean shifts upward, then a positive drift will develop for S_k, as happens after observation 30 in **Figure 2**. Sometimes, however, even when the process is at target, S_k can wander remarkably far from 0 and give the appearance that there has been a process shift, for example, between observations 5 and 10. This is a symptom of the random walk. A V-mask is used to differentiate between random walk and a shift in process mean. The V-mask is applied to each new point on the CUSUM chart with its vertex at distance d from the current point with its axis parallel to the time axis. If all previous points lie within the arms of the V-mask, then the process is in control. In **Figure 2**, the mask is placed on observation 40 to illustrate an out-of-control case.

The construction of the V-mask is essentially the calculation of the angle θ, shown in **Figure 2**, which is given by $\theta = \tan^{-1}(\Delta/2A)$. Here Δ is the shift we wish to detect and A is the scale factor of the S_k axis. The leading distance d is calculated as

$$d = \left(\frac{2}{\delta^2}\right) \ln\left(\frac{1-\beta}{\alpha}\right) \quad \text{with} \quad \delta = \frac{\Delta}{\sigma_Q}$$

where σ_Q is the standard deviation of Q, α is the probability of false alarm (type I error), and β is the probability of failing to detect the shift in mean (type II error). Even without the V-mask, the CUSUM is a powerful visual device for informing the user about the status of the process and can be used as a companion to Shewhart chart. CUSUM charts can be constructed for historical data and these are particularly useful for checking the validity of data to be used to estimate control limits for Shewhart charts.

An alternative technique (that sets the limits without a V-mask) is the tabular CUSUM, which is attractive for computer implementations and it is described in Montgomery.[12] One-sided CUSUM charts are also possible,[13] which are useful when only deviations of one side of the target value are important (i.e., purity should never fall below 95% or concentration of an impurity should be kept below 0.02%). Finally, it should be noted that the Western Electric procedures do not apply for the CUSUM chart. Charts to detect unknown abrupt changes have also been proposed.[14]

4.02.2.1.3 Univariate EWMA charts

For each point on a Shewhart chart, only the current observation is important; all past observations are ignored. In CUSUM charts, all observations, old and new, are weighted equally at each point. In the EWMA charts, past observations are accounted for, but they are given a smaller weight as they become older. The EWMA is defined as

$$Z_k = r\bar{x}_k + (1-r)Z_{k-1} \tag{3a}$$

where $0 < r \leq 1$. The starting value for EWMA at $k = 0$ is $Z_0 = \mu$ (the target value for the property).

Note that an alternative way for constructing a univariate EWMA is using the following statistic instead that of Equation (3a) with $Z_0 = 0$:

$$Z_k = r(\bar{x}_k - \mu) + (1-r)Z_{k-1} \tag{3b}$$

The values of Z_k (from either Equation (3a) or (3b)) are plotted on a chart with a CL at Z_0 and control limits calculated from

$$\sigma_{z_k}^2 = \sigma_Q^2 \left(\frac{r}{2-r}\right)\left[1-(1-r)^{2k}\right] \tag{4}$$

The variance of the EWMA chart eventually converges to $\sigma_{\text{EWMA}}^2 = \sigma_Q^2 [r/(2-r)]$. The limits for the chart are at $Z_0 \pm \lambda\sigma_{z_k}$, with $\lambda = 3$ for values of r between 0.1 and 0.25, which are typically used in practice. An EWMA chart is plotted in **Figure 3** for the data set of **Figure 1**, using Equation (3a), $Z_0 = \mu$, $\lambda = 3$, and $r = 0.2$. For this set of data, the 1σ shift was detected at observation 39, only one observation earlier than that detected by the Shewhart chart. Of course the users can fine-tune the detection mechanism (values of λ, r) to achieve a desired average run length (ARL) for small shifts for the particular system they are working with. The choices for the value of r and λ of EWMA charts to detect small shifts are discussed in Montgomery.[12]

As $r \to 1$, $Z_k \to \bar{x}_k$ and the chart of Equation (3a) becomes a Shewhart chart. When r has very small values ($r \to 0$), the most recent observation has a very small weight and previous observations have near equal (though very low) weights. In these cases, the EWMA takes the appearance of a CUSUM. The EWMA chart for $0 < r < 1$ stands between the Shewhart and CUSUM chart in its use of historical data.

Note that Z_k contains all the information on past data needed for the next interval $k+1$, and there is no need to keep information on previous observations and their weights. A detailed discussion can be found in Hunter,[13] where it is also shown how the EWMA can be used to provide a forecast of where the process will be in the next instance in time and, thus, provide a mechanism for dynamic process control.

Harris and Ross[15] discuss the impact of serially correlated data on the performance of these charts. It is shown that serious errors concerning the 'state of SPC' may result if the correlation structure of the observations is not taken into account.

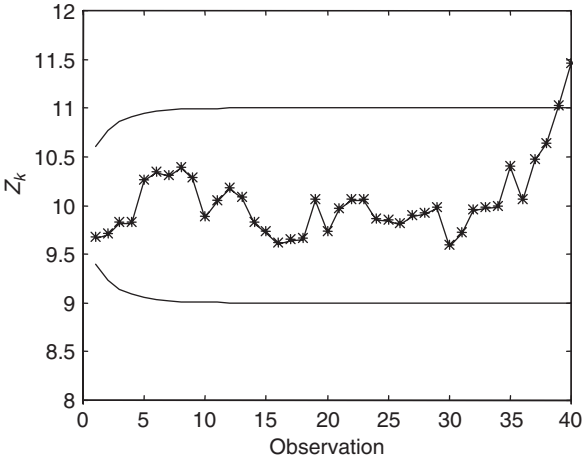

Figure 3 An EWMA chart.

4.02.2.2 Multivariate Charts for Statistical Quality Control

Most industries use univariate charts to monitor product characteristics or key process variables that in some way affect the quality of the final product. The problem with using univariate control charts for separately monitoring key variables on the final product is that most of the time the variables are not independent of one another, and none of them adequately define product quality by itself. Product quality is defined by the simultaneously correct values of all the measured properties, that is, it is a multivariate property.

In spite of the misleading nature of univariate quality control charts, they continue to be almost the only form of monitoring used by most industries. However, several multivariate extensions of the Shewhart, CUSUM, and EWMA based on Hotelling's T^2 statistic have been proposed in the literature.

4.02.2.2.1 Hotelling's T^2 and χ^2 multivariate charts

Hotelling's T^2 and χ^2 charts are the multivariate equivalents of Shewhart charts used when a vector of q variables $\mathbf{y}_k$ $(q \times 1)$ is observed at each time period k $(q \geq 2)$. Details can be found in Montgomery.[12]

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

$$\chi_k^2 = (\mathbf{y}_k - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_k - \boldsymbol{\mu}) \tag{5}$$

The χ^2 statistic in Equation (5) represents the directed or weighted distance (Mahalanobis distance) of any point from the target $\boldsymbol{\mu}$.

This statistic is distributed as a central χ^2 distribution with q degrees of freedom if the mean is on target $\boldsymbol{\mu}$. A multivariate χ^2 control chart can be constructed by plotting χ^2 versus time with a UCL given by $\chi_\alpha^2(q)$, where α is an appropriate level of significance for performing the test (e.g., $\alpha = 0.01$).

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

$$\mathbf{S} = (n-1)^{-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^{\mathrm{T}} \tag{6}$$

When a new multivariate observation $(\mathbf{y})$ is obtained, Hotelling's T^2 statistic is given by

$$T^2 = (\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{S}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \tag{7}$$

This can be plotted against time in a way similar to χ^2. A UCL on this chart is given by

$$T_{\mathrm{UCL}}^2 = \frac{(n-1)(n+1)q}{n(n-q)} F_\alpha(q, n-q) \tag{8}$$

where $F_\alpha(q, n-q)$ is the upper $100\alpha\%$ critical point of the F-distribution with q and $n-q$ degrees of freedom.[16]

The above charts are for a single new multivariate observation vector at each time. If an average of m new multivariate observations is used at each time, or if the estimate of the variance $\mathbf{S}$ is based on pooling estimates from rational subgroups, then the χ^2- and T^2-charts and their UCLs must be correspondingly redefined.[17] 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.[16,17]

Once an out-of-control signal is detected, the challenge is to determine the variables that are responsible for it. Several approaches have been suggested for this. Kourti and MacGregor[4] reviewed these approaches and suggest a computationally fast and easy algorithm based on principal components (PCs) to determine the variables that numerically contribute to the out-of-control signal. This approach is detailed later in this work.

4.02.2.2.2 Multivariate CUSUM charts

There have been several suggestions for multivariate CUSUM (MCUSUM) charts by various researchers. These are reviewed by Wierda.[17] The author investigated whether an MCUSUM chart could detect small shifts earlier than a Hotteling's T^2-chart, and ranked various MCUSUM charts based on their performance in

simulated cases. The best performer was that of Pignatello and Runger[18] described below. The charts were evaluated for their ability to detect a shift and the interpretability of the out-of-control signal. We describe briefly some of these charts and refer the reader to Wierda[17] and Lowry *et al.*[19] for more discussions and comparisons of their performance.

MCUSUM charts can be thought of as sequential probability ratio tests. Suppose we have a vector $\mathbf{y}_k$ of measurements on q variables, taken at each interval k, so we have a sequence of $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots$ The $\mathbf{y}$ variables are distributed independently with an in-control covariance $\boldsymbol{\Sigma}$, as N_q ($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$), and it is desired to test the null hypothesis H_0: the mean is at $\boldsymbol{\mu}$ versus H_1: the mean is at $\boldsymbol{\mu}_1$. Let

$$d^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu})$$

define the squared length of the shift in mean we wish to detect. Then the sequential probability ratio test rejects the null hypothesis whenever

$$\sum_{i=1}^{k} \left\{ d^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_k - \boldsymbol{\mu}) - \frac{d}{2} \right\} > -\frac{\log \alpha}{d} \tag{9}$$

where α is the level of significance chosen (the probability for type I error). Healy[20] used this result to propose plotting the CUSUM.

$$C_k = \max \left\{ 0, C_{k-1} + d^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_k - \boldsymbol{\mu}) - \frac{d}{2} \right\} \tag{10}$$

Recall that with this CUSUM, an out-of-control signal is interpreted as a shift in mean from $\boldsymbol{\mu}$ to $\boldsymbol{\mu}_1$.

Crosier[21] proposed the COT scheme (CUSUM of T). This consists of computing T_k^2 (Equation (7)) at each point in time k and then forming a CUSUM of the scalars T_k as

$$C_k = \max\{0,\ C_{k-1} + T_k - \varphi\} \tag{11}$$

with initial condition $C_0 \geq 0$ and $\varphi > 0$, where φ is a shrinkage factor (the updated C_k is shrunk toward 0 by φ). The value of φ is chosen based on ARL considerations. (The average number of points that must be plotted before a point indicates an out-of-control condition is called the ARL.) This CUSUM scheme signals an out-of-control situation when $C_k > h$, where h is the control limit for the CUSUM.

Crosier[21] also proposed replacing the scalar quantities of the univariate CUSUM by their vector counterparts and computing the vector CUSUM.

$$\mathbf{s}_k = \begin{cases} 0 & \text{if } C_k \leq \varphi \\ (\mathbf{s}_{k-1} + \mathbf{y}_k - \boldsymbol{\mu}) \left(1 - \dfrac{\varphi}{C_k} \right) & \text{if } C_k > \varphi \end{cases} \tag{12}$$

where C_k is the weighted length $\{(\mathbf{s}_{k-1} + \mathbf{y}_k - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{s}_{k-1} + \mathbf{y}_k - \boldsymbol{\mu})\}^{1/2}$, $\mathbf{s}_k$ is the vector of the individual S_k's of each variable, and $\mathbf{s}_0$ is a vector of 0's.

The scheme signals an out-of-control situation whenever

$$\max\{0, C_k - \varphi\} > h \tag{13}$$

A reference value $\varphi = d^2/2$ is usually chosen to detect any shift in the mean vector yielding square length d. This choice minimizes the ARL at deviation d for a given on-target ARL. The on-target ARL is determined by the choice of the control limit h.

Pignatello and Runger[18] suggest a chart based on the following vectors of cumulative sums:

$$\mathbf{d}_k = \sum_{j=k-n_k+1}^{k} (\mathbf{y}_j - \boldsymbol{\mu})$$

n_k is the number of observations since the most recent renewal (the last zero value) of the CUSUM chart. The chart is constructed for the statistic

$$MC_k = \max\left\{0,\ \left(\mathbf{d}_k^T \boldsymbol{\Sigma}^{-1} \mathbf{d}_k\right)^{1/2} - \xi n_k\right\}$$

where $\xi > 0$. $n_k = n_{k-1} + 1$ if $0 < MC_{k-1} < h$ and $n_k = 1$, otherwise. An out-of-control signal is given when $MC_k > h$, where $h > 0$ is the control limit.

4.02.2.2.3 Multivariate EWMA charts

Lowry *et al.*[19] extended Equation (3b) to the multivariate exponentially weighted moving average (MEWMA). Suppose that random vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots$ on q process variables are observed over time with covariance matrix $\boldsymbol{\Sigma}$. At time k, for the observed vector $\mathbf{y}_k$,

$$\mathbf{z}_k = \mathbf{R}(\mathbf{y}_k - \boldsymbol{\mu}) + (\mathbf{I} - \mathbf{R})\mathbf{z}_{k-1} \tag{14}$$

where $\mathbf{z}_k$ is essentially the vector of univariate Z_k's calculated at interval k for each variable; the vector $\boldsymbol{\mu}$ gives the targets of the individual variables and $\mathbf{R} = \text{diagonal}\{r_1, r_2, \ldots, r_q\}$ and $0 < r_j \leq 1; j = 1, \ldots, q$. The MEWMA gives an out-of-control signal when

$$Q_k^2 = \mathbf{z}_k^T \boldsymbol{\Omega}_k^{-1} \mathbf{z}_k > h \tag{15a}$$

where the control limit h is chosen to achieve a specified in-control ARL and $\boldsymbol{\Omega}_k$ is the covariance matrix of $\mathbf{z}_k$. When all the r_j are equal to each other ($r_j = r$, for $j = 1, 2, \ldots, q$), the covariance matrix is given by

$$\boldsymbol{\Omega}_k = \frac{r[1 - (1-r)^{2k}]}{2 - r}\boldsymbol{\Sigma} \tag{15b}$$

Lowry *et al.*[19] showed that depending on the type of process and the shifts to be detected, this multivariate chart, although simple to construct, performs equally well as, and sometimes better than, the CUSUM of Crosier[21] – not the COT – and Pignatello and Runger[18] based on ARL comparisons.

4.02.3 Latent Variable Based Process Monitoring

Latent variable methods allow monitoring a large number of process as well as product variables. Charts based on latent variables are simple, easy to understand by the operators, and have found quick acceptance in the control rooms. They improve early fault detection capabilities, because they are able to detect the onset of a fault at the same time as, but in most situations earlier than, the many univariate charts. More importantly, however, they detect problems that manifest themselves as changes in the covariance structure of the process variables, which univariate charts will miss if the variables remain within their expected univariate operation limits. The methodology based on latent variables also provides diagnostic tools that help the operators to determine quickly and efficiently the source of the problem.

Venkatasubramanian *et al.*[22–24] give a three-part review of process fault detection and diagnosis techniques, where quantitative model-based methods, qualitative models and search strategies, and process history-based methods are discussed, respectively. Latent variable-based fault detection methods fall into the category of process history-based methods, because they use historical databases to derive the empirical models and to set the limits for acceptable operation. Historical databases are noncausal in nature. The differences between latent variable methods and causal model-based approaches to FDI are discussed by Yoon and MacGregor.[25]

Latent variable methodology exploits the main characteristic of process databases, namely that although they consist of measurements on a large number of variables (hundreds), these variables are highly correlated and the effective dimension of the space in which they move is very small. Typically only a few process disturbances or independent process changes routinely occur, and the hundreds of measurements on the process variables are only different reflections of these few underlying events. Latent variable methods that exploit the correlation of process variables, and therefore model the structure of the process space, are extremely powerful when dealing with two other problematic characteristics of industrial historical databases, namely missing data and low content of information in any one variable (due to the low signal-to-noise ratios). The last 15 years have seen an

upsurge in interest in the application of multivariate statistical methods based on latent variables to FDI problems from industry, as can be seen from the number of industrial publications and patents listed in several publications.[2,5,8,26]

4.02.3.1 Hotelling's T^2 and PC Transformation

When the number of measured variables is large, it is frequently the case that they are highly correlated with one another and their covariance matrix $\mathbf{\Sigma}$ (required for the χ^2 statistic calculation, Equation (5)) is nearly singular. Principal component analysis (PCA) is a procedure for reducing the dimensionality of the variable space by representing it with a few orthogonal (uncorrelated) variables that capture most of its variability. New variables are calculated as linear combinations of the original variables. The new variables are independent of one another and are calculated such that the first one explains the highest amount of variability in the system, the second the next highest amount, and so on. A few PCs may explain a very high percentage of variation of the system. Control charts using these first few variables can then be developed. PCA is covered in Chapter 2.13.

Suppose we have a sample of mean-centered and scaled measurements with n observations on q variables, $\mathbf{Y}$. The first PC is defined as the linear combination $\mathbf{t}_1 = \mathbf{Y}\mathbf{p}_1$ that has maximum variance subject to $|\mathbf{p}_1| = 1$. The second PC is the linear combination defined by $\mathbf{t}_2 = \mathbf{Y}\mathbf{p}_2$ that has the next greatest variance subject to $|\mathbf{p}_2| = 1$ and subject to the condition that it is uncorrelated with (orthogonal to) the first PC ($\mathbf{t}_1$). Up to q, PCs are similarly defined. The sample PC loading vectors $\mathbf{p}_i$ are the eigenvectors of the covariance matrix of $\mathbf{Y}$ (in practice, the covariance matrix is estimated by $(n-1)^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{Y}$). The corresponding eigenvalues give the variance of the PCs (i.e., $\mathrm{var}(\mathbf{t}_i) = \lambda_i$). In effect PCA decomposes the observation matrix $\mathbf{Y}$ as

$$\mathbf{Y} = \mathbf{T}\mathbf{P}^{\mathrm{T}} = \sum_{i=1}^{q} \mathbf{t}_i \mathbf{p}_i^{\mathrm{T}} \tag{16}$$

PCA is scale-dependent in that if the $\mathbf{Y}$ matrix has variables that numerically differ in orders of magnitude (e.g., temperature in the range 200–300 °C and viscosity in the range 0.1–0.3 P), then the first few PCs are dominated by the numerically large variables. So, the $\mathbf{Y}$ matrix must be scaled in some meaningful way to remove the effect of numerically large values. The commonest approach is to scale all variables to unit variance. This means that PCA is carried out on the correlation matrix rather than the covariance matrix.

In practice, one rarely needs to compute all q eigenvectors, as most of the predictable variability in the data is captured in the first few PCs. The NIPALS algorithm[27] is ideal for computing the PCs 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, with cross-validation[28] being perhaps the most reliable. By retaining only the first A PCs, the $\mathbf{Y}$ matrix is approximated by

$$\hat{\mathbf{Y}} = \sum_{i=1}^{A} \mathbf{t}_i \mathbf{p}_i{}^{\mathrm{T}} \tag{17}$$

The traditional Hotelling's T^2 on the original variables (Equation (7)) can be written in terms of the PCs of the variables,[4]

$$T^2 = \sum_{i=1}^{q} \frac{t_i^2}{\lambda_i} = \sum_{i=1}^{q} \frac{t_i^2}{s_{t_i}^2} \tag{18}$$

4.02.3.2 Partial Least Squares

Suppose that two matrices are available, an $(n \times m)$ process variable data matrix, $\mathbf{X}$, and an $(n \times q)$ matrix of corresponding product quality data, $\mathbf{Y}$. It would be very useful to extract latent variables that explain the high variation in the process data, $\mathbf{X}$, which is most predictive of the product quality data, $\mathbf{Y}$. Then we can create charts to monitor the process variables but with such control limits that an alarm signals when a change in the

process variables will affect the product. Partial least squares (PLS) is a method (or really a class of methods) that 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,[29,30] the first PLS latent variable $\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1$ is the linear combination of the x-variables that maximizes the covariance between $\mathbf{t}_1$ and the $\mathbf{Y}$ space.

The first PLS weight vector $\mathbf{w}_1$ is the first eigenvector of the sample covariance matrix $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$. Once the scores for the first component have been computed, the columns of $\mathbf{X}$ are regressed on $\mathbf{t}_1$ to give a regression vector, $\mathbf{p}_1 = \mathbf{X}\mathbf{t}_1/\mathbf{t}_1^T\mathbf{t}_1$, and the $\mathbf{X}$ matrix is deflated (the $\hat{\mathbf{X}}$ values predicted by the model formed by $\mathbf{p}_1$, $\mathbf{t}_1$, and $\mathbf{w}_1$ are subtracted from the original $\mathbf{X}$ values) to give residuals $\mathbf{X}_2 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1^T$. $\mathbf{Q}$ are the loadings in the $\mathbf{Y}$ space. In NIPALS, $\mathbf{q}_1$ is obtained by regressing $\mathbf{t}_1$ on $\mathbf{Y}$, then $\mathbf{Y}$ is deflated $\mathbf{Y}_2 = \mathbf{Y} - \mathbf{t}_1\mathbf{q}_1^T$. The second latent variable is then computed from the residuals as $\mathbf{t}_2 = \mathbf{X}\mathbf{w}_2$, where $\mathbf{w}_2$ 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 ($\mathbf{t}_1$, $\mathbf{t}_2$, ...) and the weight vectors ($\mathbf{w}_1$, $\mathbf{w}_2$, ...) are orthogonal. PCA and PLS are frequently referred to as projection methods because the initial information is projected on to a lower-dimensional space.

4.02.3.3 General Form of Latent Variable Models

There are several latent variable methods and they can be described by a common framework.[31] For a set of historical process data consisting of an $(n \times k)$ matrix of process variable measurements $\mathbf{X}$ and a corresponding $(n \times m)$ matrix of product quality data $\mathbf{Y}$, the latent variable models for linear spaces are given by the following equations:[1,4,5]

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{19}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^T + \mathbf{F} \tag{20}$$

$\mathbf{E}$ and $\mathbf{F}$ are error terms, $\mathbf{T}$ is an $(n \times A)$ matrix of latent variable scores, and $\mathbf{P}$ $(k \times A)$ and $\mathbf{Q}$ $(m \times A)$ are loading matrices that show how the latent variables are related to the original $\mathbf{X}$ and $\mathbf{Y}$ variables. Latent variable models assume that both the process and product data spaces are observed with error and that both are effectively of very low dimension (i.e., nonfull rank). The dimension A of the latent variable space is often quite small compared to the dimension of the process data space, and it is determined by cross-validation or some other procedure. Effectively, these models reduce the dimension of the problem through a projection of the high-dimensional $\mathbf{X}$ and $\mathbf{Y}$ spaces onto the low-dimensional latent variable space $\mathbf{T}$, which contains most of the important information. By working in this low-dimensional space of the latent variables (t_1, t_2, ..., t_A), the problems of process analysis, monitoring, control, and optimization are greatly simplified.

PCA models only a single space ($\mathbf{X}$ or $\mathbf{Y}$) by finding the latent variables that explain the maximum variance. Projection to latent structures or partial least squares (PLS) maximizes the covariance of $\mathbf{X}$ and $\mathbf{Y}$ (i.e., the variance of $\mathbf{X}$ and $\mathbf{Y}$ explained, plus correlation between $\mathbf{X}$ and $\mathbf{Y}$). Reduced rank regression (RRR) maximizes the variance of $\mathbf{Y}$ and the correlation between $\mathbf{X}$ and $\mathbf{Y}$. Canonical variate analysis (CVA), or canonical correlation regression (CCR), maximizes only the correlation between $\mathbf{X}$ and $\mathbf{Y}$. The choice of the method depends on the objectives of the problem; however, all of them lead to a great reduction in the dimension of the problem. Methods that model the variation in the $\mathbf{X}$ space as well as in the $\mathbf{Y}$ space (principal component regression (PCR) and PLS) are excellent candidates for applications such as process monitoring and FDI,[1] control in the latent variable space,[6] and for dealing with missing data.[32–34]

4.02.3.4 Multivariate Control Charts Based on Latent Variables

Ever since control charts based on latent variables were introduced, their use in industry is increasing. The charts answer the need of process industries for a tool that allows them to utilize the massive amounts of data being collected on hundreds of process variables, as well as the spectral data collected from modern analyzers.

Latent variable control charts can be constructed to monitor either a group of response variables $\mathbf{Y}$ (e.g., product quality variables) or a group of predictor variables $\mathbf{X}$ (process variables). However, a very important advantage of latent variables is that they can be used to monitor predictor variables taking into account their effect on the response variables. A model is built to relate $\mathbf{X}$ and $\mathbf{Y}$ using available historical or specially

collected data. Monitoring charts are then constructed for future values of **X**. This approach means that the process performance can be monitored even at times when the product quality measurements, **Y**, are not available.

The main approach of statistical quality control (SQC) methods developed throughout the statistical literature has been to monitor only product quality data (**Y**) and, in some cases, a few key process variables (**X**). However, often hundreds of process variables are measured much more frequently (and usually more accurately) than the product quality data. So monitoring the process data is expected to supply much more information on the state of the process and supply this information more frequently. Furthermore, any special event that occurs will also have its fingerprints in the process data. So, once a special event is detected, it is easier to diagnose the source of the problem as we are dealing directly with the process variables. On the contrary, control charts on the product variables only indicate that the product properties are no longer consistent with specification, and they do not point to the process variables responsible for this.

Control charts on process variables are useful in multistep operations when quality data are not available between successive steps. For example, if a catalyst is conditioned in a batch process before being used for polymer production, the quality of the catalyst (success of conditioning) is assessed by its performance in the subsequent polymer production. It would be useful to know if the catalyst will produce good product before using it; monitoring the batch process variables with a latent variable chart would give early detection of poor quality product. Similarly, the few properties measured on a product are sometimes not sufficient to define product performance for several different customers. For example, if only viscosity of a polymer is measured, end-use applications that depend on chemical structure (e.g., branching, composition, and end-group concentration) are unlikely to receive good material. In these cases, the process data may contain much more information about events with special causes that affect the hidden product quality variables.

Multivariate control charts (Hotelling's T^2, MCUSUM, etc.) have been used to monitor product quality where a few quality variables are considered. These charts could be applied also to a few process variables. However, when the number of process variables becomes very large, one faces singularity problems in the calculation of the Hotelling's T^2.

Now, having defined the PCs, we can illustrate some of the problems with using T^2 when the variables are highly correlated and $\mathbf{\Sigma}$ is very ill-conditioned. From Equation (18) we have:

$$T^2 = \sum_{i=1}^{q} \frac{t_i^2}{\lambda_i} = \sum_{i=1}^{q} \frac{t_i^2}{s_{t_i}^2} = \sum_{i=1}^{A} \frac{t_i^2}{s_{t_i}^2} + \sum_{i=A+1}^{q} \frac{t_i^2}{s_{t_i}^2}$$

In effect, each t_i^2 is scaled by the reciprocal of its variance bringing them to the same numerical scale. Thus, 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. The last few PCs ($i > A+1$) are divided by very small eigenvalues and as a result very small deviations in their value (sometimes owing to round off error) are magnified and can cause large changes in T^2. As a consequence, if these PCs, which explain very little variance in **Y** and generally represent random noise, are used to calculate T^2, they may cause the generation of out-of-control signals.

By using only the first A important PCs and calculating T^2 only from these,

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

the distorting effect of the minor PCs is removed.

The philosophy applied in developing multivariate SPC procedures based on projection methods is the same as that used for the univariate Shewhart charts. An appropriate reference set that defines the normal operating conditions for a particular process is chosen. Future values are compared against this set. A PCA or PLS model is built based on data collected from periods of plant operation when performance was good. Periods containing variations due to special events are omitted at this stage. The choice and quality of this reference set are critical to the successful application of the procedure.

When we are interested in developing a chart on a set of variables **Z** (**Z** could be either the **X** predictor or process space, or the **Y** response or quality space), a PCA model can be used. From the historical data, we develop the model

$$\hat{\mathbf{Z}} = \sum_{i=1}^{A} \mathbf{t}_i \mathbf{p}_i{}^{\mathrm{T}} \tag{21}$$

We calculate the variance of the scores from the model scores per component $s_{t_i}^2 (= \lambda_i)$ and use them together with $\mathbf{P}_A = [\mathbf{p}_1\ \mathbf{p}_2 \ldots \mathbf{p}_A]$ for future monitoring calculations. Future behavior can now be referenced against this 'in-control' model. Using Equation (22), we construct a chart for T_A^2, for each new multivariate observation $\mathbf{z}_{\mathrm{new}}$ $(q \times 1)$.

$$T_A^2 = \sum_{i=1}^{A} \frac{t_{i,\mathrm{new}}^2}{s_{t_i}^2} \tag{22}$$

by calculating scores and residuals as

$$\mathbf{t}_{i,\mathrm{new}} = \mathbf{p}_i^{\mathrm{T}} \mathbf{z}_{\mathrm{new}} \quad \text{and} \quad \mathbf{e}_{\mathrm{new}} = \mathbf{z}_{\mathrm{new}} - \hat{\mathbf{z}}_{\mathrm{new}}$$

where $\hat{\mathbf{z}}_{\mathrm{new}} = \mathbf{P}_A \mathbf{t}_{A,\mathrm{new}}$, $\mathbf{t}_{A,\mathrm{new}}$ is the $(A \times 1)$ vector of scores, and $\mathbf{P}_A$ is the $(q \times A)$ matrix of loadings.

Monitoring $\mathbf{z}_{\mathrm{new}}$ via T_A^2 will only detect whether or not the variation in the variables z in the plane of the first A PCs is greater than can be explained by common cause. However, this is not sufficient. Another chart to monitor the residuals is also necessary. By monitoring the residuals, we can detect whether or not the noise in the system is similar to the one that existed during model development. A special event that was not included in the reference data used to develop the in-control PCA model may result in a change in the covariance structure of $\mathbf{Z}$. The squared prediction error (SPE_z) of the residuals of new observations is sensitive to this change.

$$\mathrm{SPE}_z = \sum_{i=1}^{q} \left(z_{\mathrm{new},i} - \hat{z}_{\mathrm{new},i} \right)^2 \tag{23}$$

This statistic is also referred to as the Q-statistic[35] or distance to the model. It represents the squared perpendicular distance of a new multivariate observation from the plane defined by the model. When the process is 'in-control', SPE_z should be within control limits developed from historical data, computed as discussed later. A very effective set of multivariate control charts is therefore a T_A^2-chart on the A dominant orthogonal PCs ($\mathbf{t}_1, \ldots, \mathbf{t}_A$) and an SPE_z chart.

When both $\mathbf{X}$ and $\mathbf{Y}$ historical data are available, the scores $\mathbf{t}$ in the X-space, the weights $\mathbf{w}$, and the loadings $\mathbf{p}$ are calculated from a PLS model between $\mathbf{X}$ and $\mathbf{Y}$. The standard deviation of the $\mathbf{t}$ scores is calculated to be used on the T_A^2-chart. For process monitoring, the new observations are available only for $\mathbf{X}$. New scores are calculated for the new observation $\mathbf{x}_{\mathrm{new}}$ $(m \times 1)$ as

$$\mathbf{t}_{j,\mathrm{new}} = \mathbf{w}_j^{\mathrm{T}} \mathbf{x}_{j,\mathrm{new}}$$

where $\mathbf{x}_{1,\mathrm{new}} = \mathbf{x}_{\mathrm{new}}$ and for $j > 1$, $\mathbf{x}_{j,\mathrm{new}} = \mathbf{x}_{j-1,\mathrm{new}} - \mathbf{t}_{j-1,\mathrm{new}} \mathbf{p}_{j-1}$.

Multivariate control is now achieved with a T_A^2-chart on the first A latent variables (Equation (22)) and an SPE_x chart where

$$\mathrm{SPE}_x = \sum_{i=1}^{m} \left(x_{\mathrm{new},i} - \hat{x}_{\mathrm{new},i} \right)^2 \tag{24}$$

where $\hat{\mathbf{x}}_{\mathrm{new}} = \mathbf{P}_A \mathbf{t}_{A,\mathrm{new}}$. As with the PCA method, the SPE_x plot will detect the occurrence of events that cause the process to move away from the hyperplane defined by the reference model.

4.02.3.4.1 Calculation of chart control limits

For a Hotelling's T^2-chart (either for PCA or PLS) a UCL based on the first A PCs and derived from n observations is obtained using the F-distribution and given by

$$T_{A,\mathrm{UCL}}^2 = \frac{(n^2 - 1)A}{n(n - A)} F_{\alpha(A,n-A)} \tag{25}$$

where $F_{\alpha(A,n-A)}$ is the upper $100\alpha\%$ critical point of the F-distribution with $(A, n-A)$ degrees of freedom.[4]

For the SPE_x chart, limits can be computed using approximate results from the distribution of quadratic forms. Jackson and Mudholkar[36] showed that the critical upper $100(1-\alpha)\%$ confidence interval on SPE is given as

$$\theta_1\left[\frac{z_a\sqrt{2\theta_2 h_0^2}}{\theta_1}+\frac{\theta_2 h_0(h_0-1)}{\theta_1^2}+1\right]^{1/h_0} \tag{26}$$

where z_a is the unit normal deviate corresponding to the upper $100(1-\alpha)\%$, α is the chance taken to incorrectly declare a fault because of the type I error

$$\theta_i=\sum_{j=A+1}^{m}\lambda_j^i=\mathrm{Tr}(\mathbf{E}^i)\quad \text{for } i=1,2,3,\ldots \tag{27}$$

where λ_i is the ith eigenvalue referring to the covariance matrix, and $h_0=1-(2\theta_1\theta_3/3\theta_2^2)$.

Nomikos and MacGregor[37] used an approximation based on the weighted χ^2 distribution $[g\chi^2(h)]$ proposed by Box.[38] They suggested a simple and fast way to estimate the g and h, which is based on matching moments between a $g\chi^2(h)$ distribution and the reference distribution of SPE at any time interval. The mean $[\mu=gh]$ and the variance $[\sigma^2=g^2(2h)]$ of the distribution are equated to the sample mean (b) and variance (v) at each time interval. Therefore, the g and h are estimated from the equations $\hat{g}=v/2b$ and $\hat{h}=2b^2/v$.

Hence, the UCL on the SPE at significance level α is given by

$$\frac{v}{2b}\chi_\alpha^2\left(\frac{2b^2}{v}\right) \tag{28}$$

It should be emphasized that the models built for process monitoring model only common-cause variation and not causal variation. The main concepts behind the development and use of these multivariate SPC charts based on latent variables for monitoring continuous processes were laid out in the early 1990s.[3,4,39–41] Several illustrations of the projection methods are also presented in these papers, along with the algorithms and details on estimating control limits.

4.02.3.5 Fault Diagnosis and Contribution Plots

In classical quality control where only quality variables are monitored, it is up to process operators and engineers to try to diagnose an assignable cause for an out-of-control signal, using their process knowledge and a one-at-a-time inspection of process variables. When PLS or PCA models are used to construct the multivariate control charts, they provide the user with the tools to diagnose assignable causes. Diagnostic or contribution plots can be extracted from the underlying PLS or PCA model at the point where an event has been detected. The plot reveals the group of process variables making the greatest contributions to the deviations in SPE_x and the scores. Although these plots will not unequivocally diagnose the cause, they provide an insight into possible causes and thereby greatly narrow the search.

4.02.3.5.1 Contributions to SPE

When an out-of-control situation is detected on the SPE plot, the contribution of each variable of the original data set is simply given by $\left(x_{\mathrm{new},j}-\hat{x}_{\mathrm{new},j}\right)^2$. Variables with high contributions are investigated.

4.02.3.5.2 Contributions to Hotelling's T^2

Contributions to an out-of-limits value in the Hotelling's T^2-chart are obtained as follows: A bar plot of the normalized scores $\left(t_i/s_{t_i}\right)^2$ is plotted and scores with high normalized values are further investigated by calculating variable contributions. A variable contribution plot indicates how each variable involved in the calculation of that score contributes to it. The contribution of each variable of the original data set to the score of component q is given by

$$c_j=p_{q,j}(x_j-\bar{x}_j)\ \text{for PCA}\quad \text{and}\quad c_j=w_{q,j}(x_j-\bar{x}_j)\ \text{for PLS} \tag{29}$$

where c_j is the contribution of the jth variable at the given observation, $p_{q,j}$ is the loading and $w_{q,j}$ is the weight of this variable to the score of the PC q, and $\bar{x}_j$ is its mean value (which is 0 for mean-centered data).[4] Variables on this plot that appear to have the largest contributions to it, but also the same sign as the score should be investigated (contributions of the opposite sign will only make the score smaller).

When there are K scores with high values, an 'overall average contribution' per variable is calculated, over all the K scores, as shown below:

Step 1: Repeat for all the K high scores ($K \leq A$)

- Calculate the contribution of a variable x_j to the normalized score $(t_i/s_{t_i})^2$:

$$\text{cont}_{i,j} = \frac{t_i}{s_{t_i}^2} p_{i,j}\left(x_j - \bar{x}_j\right)$$

- Set contribution equal to 0 if it is negative (i.e., sign opposite of sign of score)

Step 2: Calculate the total contribution of variable x_j: $\text{CONT}_j = \sum_{i=1}^{K}(\text{cont}_{i,j})$
Step 3: Investigate variables with high contributions

The variable contribution plots provide a powerful tool for fault identification. However, the user should be careful with its interpretation. In general, this approach will point to a variable or a group of variables that contribute *numerically* to the out-of-control signal; these variables and any variables highly correlated with them should be investigated to assign causes. The role of the contribution plots to fault isolation is to indicate which of the variables are related to the fault rather than to reveal the actual cause of it. Sometimes, the cause of a fault is not a measured variable. The variables with high contributions to the contribution plots are simply the signature of such faults. Reactor fouling and reaction impurities are two characteristic examples of complex process faults where they manifest themselves on other measured variables.[1,4]

4.02.3.6 Other Latent Variable Charts

Further to the charts discussed above, plots of time-series normalized scores for one component (with limits ± 3) or scatter score plots are also used for detecting and diagnosing out-of-control situations.

In time-series score plots, one may wish to calculate contributions to a move in the value of the scores between two observations. From Equation (29), the contribution of a variable j to the move of the score values between two observations (say, 464 and 465) for component q is calculated as $p_{qj}\,(x_{j,465} - x_{j,464})$ for PCA and w_{qj} $(x_{j,465} - x_{j,464})$ for PLS.

The user should be careful with scatter plots and their interpretation. One could plot t_i versus t_j or (t_i/s_{t_i}) versus $\left(t_j/s_{t_j}\right)$ with $i \neq j$ and $1 \leq i, j \leq A$. The first plot preserves the Euclidean distance and the rotation of the initial space and it is good to interpret PCA modeling. The second plot helps distinguish easier out-of-control points. An in-depth discussion of the scatter plots that have been used in the literature and their interpretation can be found in Geladi *et al.*[42]

4.02.3.7 Online Process Monitoring

The procedure that leads to online monitoring is schematically shown in **Figure 4**. A set of data consisting of process (X) and product variables (Y) and representing routine operation where a good quality product is produced is used as the training set to derive the model.

A PCA model is derived from the process data ($\mathbf{X}$) and will contain vectors with information about centering, scaling and transformation of the original variables, the standard deviation of the PCs required for the T^2 calculation $(\mathbf{c}^x, \mathbf{s}^x, \mathbf{f}^x, \mathbf{s}_t^x)$, the loadings matrix ($\mathbf{P}$), and the scalar limits of the SPE and Hotelling's T^2 plots.

A PLS model is derived from the process data ($\mathbf{X}$) and the corresponding quality data ($\mathbf{Y}$) and will contain vectors with information about centering, scaling and transformation of the original variables for both spaces, the standard deviation of the PCs in $\mathbf{X}$ space $\left(\mathbf{c}^x, \mathbf{s}^x, \mathbf{f}^x, \mathbf{s}_t^x\right)$ and $(\mathbf{c}^y, \mathbf{s}^y, \mathbf{f}^y)$, the weights matrix ($\mathbf{W}$), the loadings matrices for both spaces ($\mathbf{P}$, $\mathbf{Q}$), and the scalar limits for the SPE and Hotellings T^2 plots.

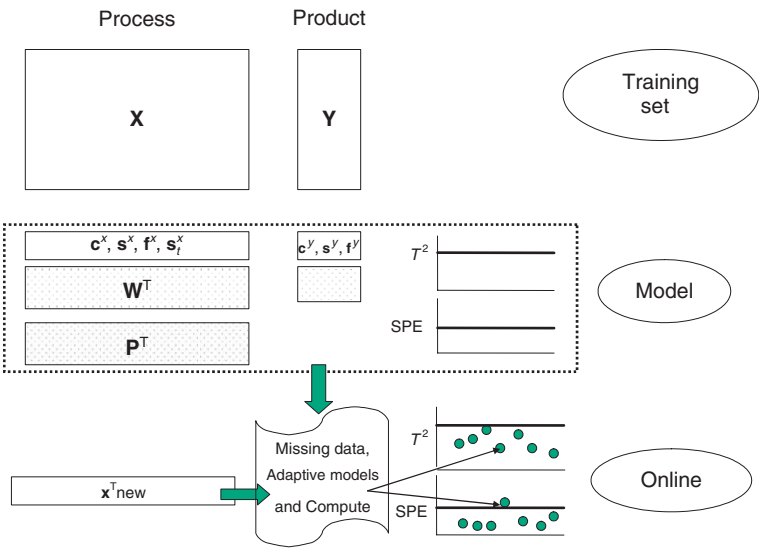

Figure 4 MSPC procedure. A set of data consisting of process (*X*) and product variables (*Y*) and representing routine operation where a good quality product is produced is used as the training set. A model is developed. For online operation, a vector of new observations $\mathbf{x}_{new}^T$ is collected at a predetermined time interval. Utilizing information from the model, the scores corresponding to $\mathbf{x}_{new}^T$ are calculated and subsequently the points on the T^2-charts and SPE charts. When there are missing data and/or the process is time variant, missing data methodology and adaptive modeling supplement the model to calculate the scores.

The information from the model will be used for the online calculation of the scores, as a new vector of observations $\mathbf{x}_{new}$ becomes available. The new vector may have missing data and/or the process may be time variant. Missing data procedures (described in Section 4.02.4) and/or adaptive modeling procedures (described in Section 4.02.5) are included in the algorithm and combined with the model to calculate the scores. From the scores, the value of the Hotellings T^2 and the value of the SPE are calculated (each new vector of observations $\mathbf{x}_{new}$ represents a point in the SPE and Hotellings T^2 graphs in **Figure 4**). If the points remain within the limits, the process is in control.[4,5]

In industry, it is common to divide the values of T_A^2 and SPE_x with their corresponding limits for normal operation to create two easy to follow indices that alarm every time their value is larger than 1. Other charts can be added to these basic charts if necessary. For example, if the detection of a specific fault that manifests itself at a certain combination of scores is important, a Hotelling's T^2 on these scores can also be plotted.[1,2]

To summarize, multivariate process/product monitoring requires three charts. The process is being monitored (online or offline) with SPE_x charts and T_A^2-charts, constructed from measurements as they become available. Once a deviation from limits is detected, contribution charts point to the variable(s) that contribute numerically to the signal. T_A^2 checks that variables causing the main variation in the system are consistent with past operation. SPE_x checks that the disturbances (noise in the system) remain within acceptable limits, consistent with past operation.

4.02.3.8 Practical Considerations

There are several issues to consider when implementing MSPC in real-time industrial applications. They include the choice of the reference set that will be used for modeling, the choice of variables to include in the monitoring scheme, the choice of weights for the variables, frequency of sampling, to name a few. Process-related knowledge and other information can be used in the model and can also help the choice of variable transformations and variable grouping with different weights. Sometimes, process knowledge is required to choose a subset of variables (and/or use nonlinear transformations) such that the detection scheme becomes more sensitive to the specific event.

4.02.3.8.1 Reference data set for SPC modeling

When dealing with empirical modeling (PCA or PLS), the data set upon which the model will be based must be chosen carefully to satisfy the needs of the intended application.

For inferential modeling and response surface modeling, one needs data from designed experiments over a prescribed range of **X** and **Y** variables. Usually, in this case, a wide range of process conditions is considered to choose the optimal operating region. When the model is to be used for SPC, only a specific operating region is tackled. Historical process and/or product data corresponding to this operating region should be used. The objective is to model the good process behavior in this operating region and to test for any future deviations from this model. All the data should correspond to in-control operation and faults or disturbances are excluded from this model. If the preliminary analysis of the historical data in this region indicates clusters containing only a few points, or shows individual outliers, these data should not be discarded, but investigated. True outliers (measurement errors) should be discarded if they can be identified. If the small clusters reflect some real, unusual event that still produces acceptable product, then the data could be included in the model (with appropriate centering/scaling/transformation) only if more data points in this region can be collected to establish a robust model. Otherwise, these data should be left out during modeling and then tested with the model as if they were new data, in order to identify what alarm signal they produce (direction of PCs), and store the information as a warning that the particular signal corresponds to this known 'rare situation' and not a bad product.

Finally, when we are exploring the process through historical data analysis, all the data should be used initially so that outliers are identified and discarded. Then the rest of the data should be used in the projection for the analysis of past behavior.

4.02.3.8.2 Data validity checking

This stage should precede the step of creating multivariate control charts. It helps us understand the behavior of the process, detect outliers, and find clusters of data. This stage should also precede any other empirical model building, no matter what type of regression model is used (multiple linear regression (MLR), PLS, or neural nets).

4.02.3.8.3 Exploration and analysis of process databases/troubleshooting

Multivariate latent variable methods can be used for exploring databases to identify periods of unusual/abnormal process behavior and to diagnose possible causes for such behavior. By examining the behavior of the process data in the projection spaces defined by the small number of latent variables (t_1, t_2, ..., t_A), regions of stable operation, sudden changes, or slow process drifts may be readily observed. An interpretation of the process movements in this reduced space can be found by examining the loading vectors ($\mathbf{p}_1$, $\mathbf{p}_2$, ..., $\mathbf{p}_A$), or ($\mathbf{w}_1$, $\mathbf{w}_2$, ..., $\mathbf{w}_A$) in the case of PLS, and the contribution plots. The use of latent variable methods to explore and analyze historical data is illustrated by a troubleshooting problem on a continuous recovery process.

The feed stream with a concentration of ~20% in component A is passed through a series of separators until a stream with high purity in component A is produced. The principal operating objectives for the plant are to maintain the concentration of A at a specified level ($\geq$99.5%) while achieving a certain minimum recovery (at least 92%) of A in this stream. However, for the last 3 months of operation, the recovery dropped significantly below 92% (**Figure 5**). The company supplied daily averages on 447 process and product variables for a period of 498 days. The data set had several missing data.

Multivariate projection methods were used for the analysis. PLS models were built between the process variables and purity and recovery. All the historical data were used for the initial model. Projection of all the process variables on the first two PCs (**Figure 6**) indicated that for the last 3 months (points 400–490), where the recovery was low, the process behavior had changed; the location of the projection of points 400–490 was different from the location of the first 400 points. It also showed that the abnormal event manifested itself along the first PC. In other words, a combination of variables from the first PC seems to be related with the event. Contribution plots can be constructed using information from the loadings and the weights to identify the process variables that numerically contributed to this move along t_1. Take for example a point/day when the process behaved normally (say, point 389) and a point after the recovery started dropping (say, 464). From point 389 to point 464, the value of t_1 decreases. A contribution plot is a bar plot of the numerical contribution of each

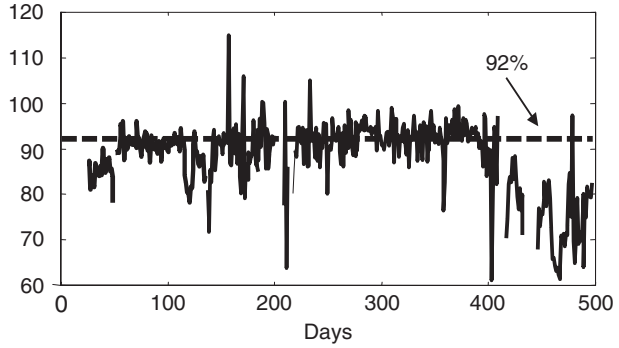

Figure 5 Recovery history of product A. Reproduced with permission from MacGregor, J.F.; Kourti, T. Multivariate Statistical Treatment of Historical Data for Productivity and Quality Improvements, AIChE Symposium Series. In *Proceedings of the 3rd International Conference on Foundations of Computer Aided Process Operations*; Pekny, J.F., Blau, G.E., Carnahan, B., Eds; CACHE and AIChE, 1998; Vol. 4, no. 320, pp. 31–41. Copyright © 1998 American Institute of Chemical Engineers, and CACHE Corporation.

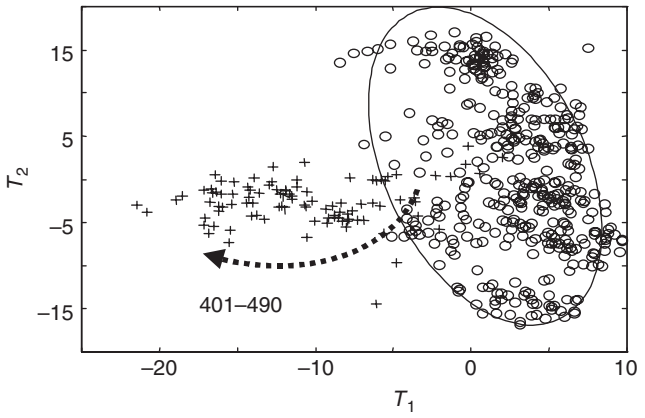

Figure 6 Projection of process history on latent variable space. Reproduced with permission from MacGregor, J.F.; Kourti, T. Multivariate Statistical Treatment of Historical Data for Productivity and Quality Improvements, AIChE Symposium Series. In *Proceedings of the 3rd International Conference on Foundations of Computer Aided Process Operations*; Pekny, J.F., Blau, G.E., Carnahan, B., Eds; CACHE and AIChE, 1998; Vol. 4, no. 320, pp. 31–41. Copyright © 1998 American Institute of Chemical Engineers, and CACHE Corporation.

variable to this decrease. **Figure 7** shows the contributions in the move along the direction of t_1, between days 389 and 464. The contribution of variable j to the move of the score values between two observations (say, 389 and 464) for component q is calculated as

$$P_{jq}\left(x_{j,389} - x_{j,464}\right) \text{for PCA}$$

and

$$w_{jq}\left(x_{j,389} - x_{j,464}\right) \text{for PLS}$$

where p_{jq} is the loading of variable j on component q and w_{jq} is the weight of variable j on component q.

Note that in **Figure 7**, four variables have the highest contribution to this move. Three of these variables (33, 207, and 277) are process conditions that can be controlled to improve the process performance; the fourth variable (158) is correlated to them, mainly as a result of these process conditions. From the 442 variables, the projection methods isolated three process conditions that were not properly controlled as responsible for the change in recovery. The time-series plots of these variables together with the percent recovery of A were investigated. All these variables show a change in behavior, a slope that is positively or negatively correlated to

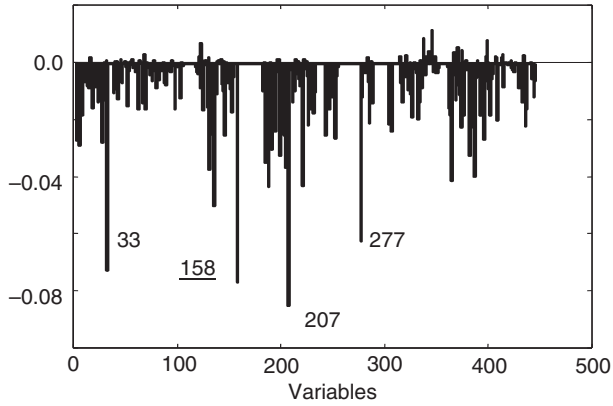

Figure 7 Contribution plots, days 389–464. Reproduced with permission from MacGregor, J.F.; Kourti, T. Multivariate Statistical Treatment of Historical Data for Productivity and Quality Improvements, AIChE Symposium Series. In *Proceedings of the 3rd International Conference on Foundations of Computer Aided Process Operations*; Pekny, J.F., Blau, G.E., Carnahan, B., Eds; CACHE and AIChE, 1998; Vol. 4, no. 320, pp. 31–41. Copyright © 1998 American Institute of Chemical Engineers.

the slope in recovery trajectory for the time period after 400. By exploiting these correlations, we are able to isolate the problem to a few variables out of 442. It is worth mentioning that, before applying multivariate analysis, company engineers spent considerable time determining what process conditions were leading to low recovery and had independently determined a set of variables. The multivariate analysis pointed to those same variables in a small fraction of the time spent by the engineers.

One should keep in mind that these methods will not be able to unequivocally identify the cause of any problem (because of the highly correlated and noncausal nature of the data). However, they can almost always identify unusual operating periods and can usually isolate the region of the plant and the group of process variables that are related to the problem. Thus, they are a powerful tool for focusing the attention of the operations engineers on a much smaller area, allowing them to better use their engineering knowledge to diagnose the cause of any abnormal situation and thereby improve the process.

The enormous potential of these methods for troubleshooting was demonstrated almost immediately after they started being used, as it is evident by the industrial applications referenced in early publications.[1,2,4,5,8,43–46]

4.02.3.8.4 Practical Considerations for Online Applications

Before any online implementation, the model should be tested offline, utilizing past data with faults, to make sure that these faults can be detected. The new monitoring scheme (based on the latent variables) runs for some time period in parallel with the scheme/tool that the operator has been using for the past years. The operators switch to the new scheme after they feel comfortable that they can rely on it.

Sometimes, it may be necessary to utilize more than one model for the same online monitoring scheme (depending on different grades, etc.). Zhang *et al.*[47] mention that they maintain six different PCA models and switch to proper model according to operation. Other times, with proper preprocessing the same model can be used for several grades.[48] Finally, although nonlinear methods are available,[49] linear models are sufficient to describe operation around a nominal point or a nominal trajectory.

Most of the industrial MSPC applications reported in the literature and others known to the author from personal experience use the methodology to monitor the process performance of specific units, rather than that of the entire plant. The reader is referred to a paper written by industrial practitioners,[2] where several issues, related to implementing MSPC in industry, are discussed in detail based on their experience. They include discussions on data selection and preparation, model development, data preprocessing, filtering, and model evaluation. They also discuss how systems are evaluated as candidates for online applications and they describe the four stages of online implementation (system design, integration, evaluation with online systems, and maintenance).

Finally, online monitoring schemes are used as a quick and easy way to test the reliability of analyzers.

4.02.3.9 Other applications of multivariate charts

The principles presented so far are applicable to any scale of operation: laboratory, pilot-scale experiments/production or full-scale process units. In this section, we shall look at some applications of multivariate charting methods not related directly to controlling full-scale processes.

Standard or check samples (samples with known properties) are used to test analytical methods periodically to ensure that they are in control. For example, latex suspensions with known particle size are used to test the calibration in particle sizing methods and samples with known concentration of an analyte are used to test gas chromatography. The results of these samples can be plotted on control charts to test that the value measured exhibits only common-cause variation. Box *et al.*[50] (p 559) give an example of how a Shewhart chart can be used for this purpose. When the response obtained from the standard is not a single number (i.e., instead of composition of one analyte, we need composition of several analytes), one should consider multivariate charts.

In chromatographic systems, variables such as flow rate and pressure affect the performance of the system, whereas, for example, column aging or a contaminated detector can take the system out of control. Multivariate monitoring of a chromatographic system has been carried out using a check sample containing five analytes to test column performance.[51] A T_A^2-chart and an SPE_x chart were used to monitor analyte peak area% of the five analytes. The results indicated that false alarms, which would have occurred with univariate charts, were avoided and points out of control due to change in correlation could be detected (impossible with univariate charts).

4.02.4 Handling Future Observations with Missing Data

Missing measurements are a frequent occurrence in process industries. Therefore, the new observation vector $\mathbf{x}_{new}$ (**Figure 4**) may have a few elements missing. Latent variable methods that model the process space (PCA, PCR, and PLS) make it possible to infer the corresponding score values of $\mathbf{x}_{new}$ by using the available elements in the vector together with the model built from the training data set.

The fact that process variables are highly correlated and that there is redundancy in process data (i.e., many variables are affected by the same event) makes this possible. Redundancy is beneficial for handling missing data. More details on methods for treatment of missing data in regression can be found in Chapter 3.06.

A variety of algorithms have been suggested[32,33] to handle missing data, with different degree of complexity: trimmed score method (TRI), single-component projection (SCP), projection to the model plane (PMP) – using PLS or ordinary least squares (PMP_{PLS}, PMP_{OLS}), iterative imputation of missing data (II), a method based on the minimization of the squared prediction error (SPE), conditional mean replacement (CMR), trimmed score regression (TSR), and regression on known data (KDR).

Suppose that $\mathbf{x}_{new}^T = [\mathbf{x}^* \;\; \mathbf{x}\#]^T$, where without loss of generality we assume that $\mathbf{x}\#$ is the vector of missing observations. (Following this convention, $\mathbf{p}^*$ and $\mathbf{P}^*$ are loadings corresponding to the known $\mathbf{x}^*$.) The methods can be seen as different ways to impute values for the missing variables vector, $\mathbf{x}\#$. By setting the missing values equal to their expected mean value (i.e., for mean-centered data $\mathbf{x}\# = 0$), we have the TRI method.[33]

SCP is the simplest but also the poorest performing approach: It calculates each of the scores independently and sequentially as $\hat{t}_i = \mathbf{z}^*\mathbf{p}_i^*/\mathbf{p}_i^{*T}\mathbf{p}_i^*$, where $\mathbf{z}^*$ is $\mathbf{x}^*$ deflated by the first $i-1$ components.

Nelson *et al.*[32] showed that superior results can be obtained by calculating all of the scores at once by projecting onto the hyperplane formed by the $\mathbf{P}^*$ vectors. In the PMP method, the known $\mathbf{x}^*$ vector is regressed onto the matrix $\mathbf{P}^*$. Sometimes, depending on the measurements missing, some of the columns of $\mathbf{P}^*$ may become highly correlated and $\mathbf{P}^{*T}\mathbf{P}^*$ becomes ill-conditioned. It was suggested[32] to use PLS, PCR, or regularized least squares regression for the projection.

CMR[32] and TSR[33] use the known score $\mathbf{T}$ matrix from the training data together with the loadings ($\mathbf{P}^*$) and the available measurements ($\mathbf{x}^*$) to estimate the score vector. A singularity problem that may arise in CMR may be solved by a procedure suggested by Nelson *et al.*,[32] where the estimated score vector is calculated in two steps: First a parameter $\boldsymbol{\beta}$ is computed using PLS from $\mathbf{T}=\mathbf{X}^*\boldsymbol{\beta}$, where $\mathbf{T}$ and $\mathbf{X}_,^*$ respectively, represent the score matrix and those columns from the training data set corresponding to known values; then $\boldsymbol{\beta}$ along with the current available data vector $\mathbf{x}^*$ is used to compute an estimate of the score vector.

In iterative imputation, one may use an initial estimate of the final scores (say, those given by SCP method) to forecast the missing values $\hat{\mathbf{x}}\#$, (using their corresponding loadings), then create the new vector and recalculate a score estimate, and iterate until convergence.

Arteaga and Ferrer[33] presented an extensive study on the various methods. Iterative imputation and SPE methods are equivalent to PMP; KDR is equivalent to CMR. They concluded that based on the best prediction of the missing values, KDR is statistically superior to the other methods. The TSR is practically equivalent to the KDR and has the advantage that a much smaller matrix needs inversion. Additionally, TSR is statistically superior to PMP method.

Before the system is implemented online, there should be a plan for the operators as to how to respond if the values of several variables stop being recorded. For example, if there are three thermocouples in a reactor and one fails, common sense dictates that we can afford to continue the monitoring scheme. On the contrary, if there is only one sensor for a variable uncorrelated with any other, the value for this variable cannot be assessed from the rest of the variables in the system; therefore depending of the importance of this variable, one may not be able to rely on the monitoring scheme until the failed sensor is replaced. This idea was treated quantitatively by Nelson[52] and Nelson *et al.*,[53] where they analyzed the uncertainty resulting from missing measurements for the predictions of the values of the Hotelling's T^2 and the SPE. Rather than representing an object with missing measurements by a single point, an estimate of the uncertainty regions in the score, Hotteling's T^2, and SPE spaces arising from the missing measurements is provided. They suggested measures to distinguish between situations where model performance will continue to be acceptable and situations where it will be unacceptable, and therefore if the missing measurements cannot be recovered the application must be shut down.

Missing data methods did find their way in the industrial applications. In their industrial perspective on implementing online applications of multivariate statistics, Miletic *et al.*[2] emphasize that missing data handling is a necessary feature for both the offline modeling and the online systems and report that are using the methods proposed by Nelson *et al.*[32]

4.02.5 Adaptive Latent Variable Models

Most of the industrial processes are time varying and may require an adaptive, rather than a fixed, model to describe their behavior; a fixed model would lead to frequent false alarms. Sometimes, variability is simply related to variable throughput and simple data laundering methods can be used.[54] For more complicated cases, several approaches for recursively updating PLS and PCA models have been proposed.[55–59] Care should be exercised when using these algorithms, because the recursive algorithm may adapt for process faults if the right window length is not used. Capron *et al.*[60] studied the effect of the number of new samples and the weight given to these samples on the performance of updated models. They investigated the possibility to use only very few samples and to give them more weight, by including several copies of them. Of course, this has an impact on the covariance structure of the new model. The authors arrived at the intuitive conclusion that the weight applied to samples used for updating the model has less importance than the number of samples and their 'representativity' as they call it (i.e., representative methods of selection of samples lead to better results than the other ones).

Dayal and MacGregor[55] developed a recursive exponentially weighted PLS algorithm. The covariance matrices are updated as

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})_t = \lambda(\mathbf{X}^{\mathrm{T}}\mathbf{X})_{t-1} + \mathbf{x}_t^{\mathrm{T}}\mathbf{x}_t \tag{30}$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{Y})_t = \lambda(\mathbf{X}^{\mathrm{T}}\mathbf{Y})_{t-1} + \mathbf{x}_t^{\mathrm{T}}\mathbf{y}_t \tag{31}$$

where $\mathbf{x}_t$ and $\mathbf{y}_t$ are the new predictor and response vectors observed at time t, and $(\mathbf{X}^{\mathrm{T}}\mathbf{X})_t$ and $(\mathbf{X}^{\mathrm{T}}\mathbf{Y})_t$ are the updated covariance matrices at time t. At each new sampling period, the previous data in the covariance matrices are being exponentially discounted, with a forgetting factor λ_t ($0 < \lambda_t \leq 1$), and the new data are being added. Instead of having a scheme for updating the mean and the variance (to account for possible change in the

mean level of the variables in the time-varying process), they augment $\mathbf{x}_t$ with a unity element to account for the constant term or the intercept: $\mathbf{x}_t = [x_{1,t}\ x_{2,t} \ldots\ x_{m,t}\ 1]$.

The mean of each variable can be computed using the last column or the row, $(\boldsymbol{\Sigma}\mathrm{x_i})_t = \lambda_t(\boldsymbol{\Sigma}\mathrm{x_i})_{t-1} + x_{i,t}$, and the current effective memory of the data, $N_t = \lambda_t(\boldsymbol{\Sigma}1)_{\mathrm{t-1}} + 1 = \lambda_t \mathrm{N_{t-1}} + 1$. Therefore

$$\bar{x}_{i,t} = \frac{(\sum x_i)_t}{N_t} \tag{32}$$

The updated variance can accordingly be computed as

$$\mathrm{var}(x_i)_t = \frac{(\sum (x_i - \bar{x}_i)^2)_t}{N_t - 1} = \frac{\left(\sum x_i^2\right)_t - N_t \bar{x}_{i,t}{}^2}{N_t - 1} \tag{33}$$

Wang *et al.*[59] introduced a monitoring scheme that involves recursive PLS algorithms together with an adaptation of the confidence limits for each of the utilized SPE charts and Hotelling's T^2-charts; this combination reduced considerably the number of false alarms as concluded from their study. They applied it to a simulation of a fluid catalytic cracking unit and on data from an industrial distillation process.

An industrial application where an adaptive PLS algorithm is used in a feedforward control scheme has been reported by Dofasco.[58] A modified version of the recursive Kernel algorithm[55] was used. At regular time intervals, a set of new observations $\mathbf{Y}_{\mathrm{new}}$ and $\mathbf{X}_{\mathrm{new}}$ are downloaded from the database. Their covariance structure is calculated by

$$\begin{aligned} \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)_{\mathrm{new}} &= \frac{1}{n_{\mathrm{new}} - 1}\mathbf{X}_{\mathrm{new}}^{\mathrm{T}}\mathbf{X}_{\mathrm{new}} \\ \left(\mathbf{X}^{\mathrm{T}}\mathbf{Y}\right)_{\mathrm{new}} &= \frac{1}{n_{\mathrm{new}} - 1}\mathbf{X}_{\mathrm{new}}^{\mathrm{T}}\mathbf{Y}_{\mathrm{new}} \end{aligned} \tag{34}$$

where n_{new} is the number of observations in $\mathbf{Y}_{\mathrm{new}}$ and $\mathbf{X}_{\mathrm{new}}$. The covariance structure is updated by

$$\begin{aligned} \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)_{\mathrm{updated}} &= \alpha\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)_{\mathrm{current}} + (1-\alpha)\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)_{\mathrm{new}} \\ \left(\mathbf{X}^{\mathrm{T}}\mathbf{Y}\right)_{\mathrm{updated}} &= \alpha\left(\mathbf{X}^{\mathrm{T}}\mathbf{Y}\right)_{\mathrm{current}} + (1-\alpha)\left(\mathbf{X}^{\mathrm{T}}\mathbf{Y}\right)_{\mathrm{new}} \end{aligned} \tag{35}$$

From the updated correlation and cross-correlation matrices, PLS is used to compute values for the coefficients in a regression-type model.

4.02.6 Batch Process Monitoring

Modeling batch operations requires taking into account their nonlinear dynamic nature. The methodology for developing multivariate control charts based on latent variables for batch process monitoring was initially presented by Nomikos and MacGregor[37,61,62] in a series of landmark papers. Any operation of finite duration, such as batch distillation, batch annealing, mixing additives for a finite time, drying, can be modeled by the same methodology.

4.02.6.1 Modeling of Batch Process Data

Historical data collected from a batch process had traditionally been represented by a three-dimensional data array $\underline{\mathbf{X}}$ where a matrix $\underline{\mathbf{X}}$ $(I \times \mathcal{J} \times K)$ indicates that $\mathcal{J}$ process variables are measured at K time intervals, or K aligned observation numbers (AON), for each one of I batches. Kourti[63,64] discussed that in practice it is not necessary that the same number of measurements are available for all the variables for the duration of the batch process. Some variables may not be present or measured for the full

duration of the batch. Furthermore, the frequency of measurements may be different because of several reasons: (1) some variables may be measured more frequently than others (i.e., some every minute and others every 15 min); (2) certain phases in the process may be sampled more frequently to capture important phenomena (in emulsion polymerization, particle nucleation occurring at the very first few minutes in the reaction determines the number of particles and the particle size distribution; one may need to capture this with more frequent sampling at those stages). Therefore, Kourti argued that the data set in such situations does not form a complete cube, but rather a cube where some columns are missing (**Figure 8**). Consequently, the methods that are used to model batch processes should be capable of modeling the structure of this incomplete cube. There are several methods for modeling three-way data. The choice of the method depends on the use of the model (i.e., prediction of final quality, monitoring, or process control) and the types of the data sets available. Critical discussions on batch processes modeling procedures for robust process monitoring, fault detection, and control can be found in Kourti[63,64] and Camacho *et al.*[65,66]

The method presented by Nomikos and MacGregor[37,61,62] is termed in the literature 'batchwise unfolding' and is capable of modeling the incomplete cube structure. Furthermore, it is capable of modeling three-way structures generated when formulating the control problem of batch processes using latent variables, which is discussed later. The method unfolds the three-dimensional structure into a two-dimensional array. In this new array, different time slices are arranged next to each other; variables observed at a given time interval are grouped in one time slice; the number of variables in each time slice may vary. **Figure 8** shows an example of an unfolded matrix where variable x_4 is not measured at time $t = 1$ and variable x_2 is not measured at time $t = 3-4$. Once the three-way structure is unfolded to a two-way matrix **X**, Equations (19) and (20) can be used to model **X** using PCA, or **X** and **Y** using PLS.

Multivariate control charts (Hotelling's T^2 and SPE) can be constructed for batch processes in a straightforward manner.[37,61,62] Multivariate charts have superior detection capabilities compared to univariate charts for batch processes. In the words of a colleague from industry: "In most cases in practice, changes in the covariance structure precede detectable deviations from nominal trajectories. This was the problem that univariate monitoring approaches for batch processes could not address. In most process upsets it is the correlation among the monitored variables that changes first, and later, when the problem becomes more pronounced, the monitored variables deviate significantly from their nominal trajectories. There are cases where a process upset will change dramatically only the correlation among the variables without causing any of the variables involved to deviate significantly from its nominal trajectory. These particular cases, although rare, can result to

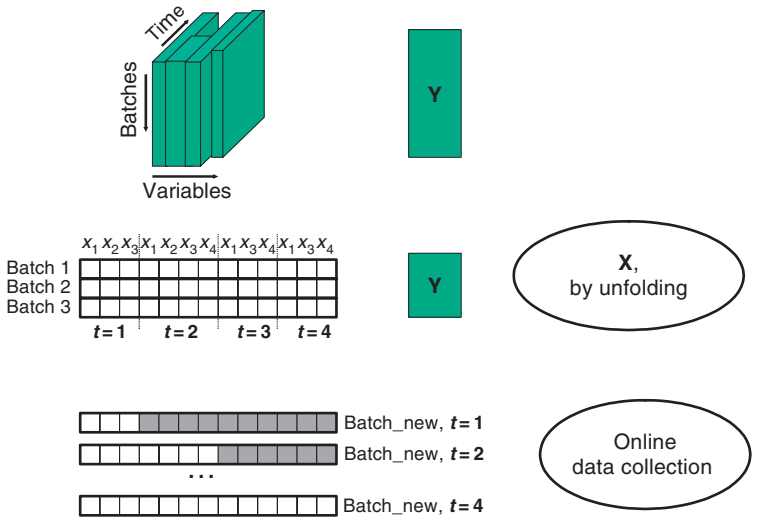

Figure 8 The structure of data in batch processes.

significant cost to a company since they can go unnoticed for long periods of time (usually they are detected from a customer complaint)." (P. Nomikos, personal communication, 2002).

4.02.6.2 Alignment of Batch Processes of Different Time Duration

Sometimes, batches have different time duration. In other words, using the same recipe it may take different time duration to achieve the same conversion. This is due to the fact that time is not the deciding factor for the completion of the batch. Sometimes, the deciding factor may be the rate by which a certain reactant is added, and so forth. One important stage before modeling batch process data is the alignment or synchronization of the data, such that they are expressed against the correct aligning factor (which may not be time). With alignment or synchronization, we must achieve the following: (1) Establish common start points at different phases of the run. For example, we could define that the first observation in all the runs for the first phase of the reaction will correspond to the start of the monomer feed, for the second phase to the initiator injection, and so on. (2) Match the shape of the trajectories of key variables. Once the shapes match, it is not necessary that the length of the batches match.[64]

A first attempt at batch data alignment involved the use of the cumulative monomer added to the reactor as an indicator variable;[67] the variable trajectories were expressed as a function of the indicator variable, rather than time. The extent of the reaction was also used as an indicator variable later.[68] Dynamic time warping, based on speech recognition methods, was suggested by Kassidas *et al.*;[69] they also suggested the use of total time as a variable in the **Z** matrix, as extra information to describe the batch. Taylor[70] suggested to include the cumulative warping, up to a given warped observation, as a new variable trajectory; his argument was that this would provide much richer information on the state of the batch by comparing it to the 'typical batch' and would provide it in real time, rather than waiting for the batch to finish so that we can calculate the total time; the cumulative time spent could be used as an extra trajectory in the case of alignment with an indicator variable. This suggestion was used later with excellent results[71,72] and also provided the basis for designing batches with desired duration;[73] Westerhuis *et al.*[71] used the cumulative warp as an extra variable to take into account time effects on batch quality, when batches were synchronized by dynamic time warping, whereas García-Muñoz *et al.*[72] used the cumulative time when batches were synchronized by the indicator variable approach. A critical discussion of the various synchronization approaches can be found in Kourti.[64] Provided that an indicator variable exists (or can be constructed by nonlinear transformations from other variables and/or process knowledge), the indicator variable approach is usually chosen as the simplest and most convenient for industrial applications. Other synchronization approaches have also been presented to deal with alignment of chromatographic data.[74]

4.02.6.3 Mean Centering and Scaling the Incomplete Cube

Mean centering the two-way matrix, formed by batchwise unfolding of the three-way data, is equivalent to subtracting from each variable trajectory its average trajectory over the I batches, thus converting a nonlinear problem to one that can be tackled with linear methods such as PCA and PLS.[37,61,62]

When the three-way data form a full cube, that is an $\underline{\mathbf{X}}$ $(I \times J \times K)$ matrix, it is common practice to autoscale the two-way matrix formed by unfolding $\underline{\mathbf{X}}$ (i.e., divide each column by its standard deviation). This accomplishes two things: (1) gives an equal weight to all time periods and consequently does not give high weights to noisy phases, or underweight a variable in tight control; after all, a variable is in tight control either because it is important to the product quality or because of safety and/or environmental concerns; and (2) for the case of complete cube gives an equal weight to all the variables considered. However, in the case where variables are sampled less frequently, or are not present for the full run, the weights have to be adjusted accordingly, depending on the objective. To give equal weight to all the variables for the entire run, for example, after autoscaling, each column corresponding to variable j must be divided by $\sqrt{K_j}$, where K_j is the number of times that variable j was sampled in the run. In the example of **Figure 8**, after autoscaling the two-way matrix, all columns corresponding to x_1 and x_3 must be divided by $\sqrt{4}$, those corresponding to x_2 by $\sqrt{2}$, and those corresponding to x_4 by $\sqrt{3}$.

4.02.6.4 Online Monitoring of Batch Processes

Each batch run has a finite duration and the process variables exhibit a dynamic behavior during the run. This means that not only the autocorrelation structure of each variable changes during the run but also that the cross-correlation of the variables changes as well. Models utilizing batchwise unfolding take into account this changing covariance structure across variables and time for the duration of the batch. For online monitoring, it is the covariance structure of each evolving batch that is compared against the typical behavior, as modeled by the training set of batches. The procedure for online monitoring of batch processes is slightly more complicated than that for continuous processes because of the following reasons:

For the online monitoring of continuous processes, at every time instant, we have a vector of new observations $\mathbf{x}_{\text{new}}^{\text{T}}$ that has a length equal to the number of columns in the model matrix $\mathbf{X}$ (and occasionally some measurements may be missing because of sensor failure or some other factor).

- In batch process monitoring, we have a vector with a length equal to the number of columns in the unfolded $\mathbf{X}$ only when the batch run has finished. At any other time, data are missing from this vector, simply because they have not been collected yet. In **Figure 8**, the new vector batch_new is shown for different time intervals; gray areas have not been collected yet. Of course, the part with the collected data may also have missing data because of sensor failure or some other factor.

Therefore, the score calculations and the limits for the multivariate control charts have to be developed in such a way that they take these 'incomplete' measurement vectors into account. The procedure for the development of the multivariate control charts for the duration of the batch was outlined by Nomikos and MacGregor.[37] To deal with the incomplete measurement vector, several approaches have been suggested. García-Muñoz *et al.*[34] recently investigated these approaches and demonstrated that using the *missing data* option and solving the score estimation problem with an appropriate method is equivalent to the use of an accurate forecast for the future samples over the shrinking horizon of the remainder of the batch. As PCA can model the covariance structure of the process variables, it facilitates handling of missing data. In batch processes, a PCA model describes the variance–covariance structure between variables over the entire batch; in other words, there exists information over all combinations of x_{jk}, j being the variable number and k the time interval. (For example, one can find how variable 4 at time 6 is related to itself at time 15, but also to variable 8 at time 35.) Because of the tremendous structural information built into these multivariate PCA models for batch processes, the missing data option for predicting the future trajectory is shown to yield the best performance by all measures, even from the beginning of the batch.

Provided that there are no faults, then for the prediction of the future process variable trajectories, the final scores, and the product quality, these missing data estimation methods are very powerful. They have also been proven critical to the success of the control methods using latent variables.[6] However, for process monitoring and online detection of process faults, all the alternative 'filling in' methods give similar results. When a fault occurs, the model structure is not valid anymore. In that case, the differences among the trajectory estimation methods appear to be much less critical because the control charts used in each case are tailored to the filling in mechanism employed. All the approaches appear to provide powerful charting methods for monitoring the progress of batch processes.

The calculation of monitoring charts and their limits for batch processes is discussed in Nomikos and MacGregor,[37] where the Hotelling's T^2 statistic for the analysis of batch process data (called D statistic) is calculated as

$$D = \frac{\mathbf{t}_R^{\text{T}}\mathbf{S}^{-1}\mathbf{t}_R\mathbf{I}}{(\mathbf{I}-1)^2} \tag{36}$$

where $\mathbf{t}_R$ is the vector containing the R retained components of the model and $\mathbf{S}$ represents the covariance matrix of the R retained score vectors. It is mentioned that $\mathbf{S}$ is a diagonal matrix because of the orthogonality of the scores, which is true for the final score estimate (i.e., when the batch run is complete). García-Muñoz *et al.*[34] discuss that when computing this statistic for the online monitoring of batches, one should consider that the covariance of the scores changes with time and the scores might become nonorthogonal; therefore, they

compute the Hotelling's statistic using the correct and complete variance–covariance matrix that corresponds to each time sample. This time-varying variance–covariance is computed using the reference set of batches. Therefore, the estimate of the Hotelling statistic at time k for batch i, D_{ki}, is a function of the estimate of the score vector for the R retained components at time k ($\hat{\boldsymbol{\tau}}_{Rki}$) for batch i and the covariance matrix of the scores at time k ($\mathbf{S}_k$). Note that D_{ki} will change depending on the method used to solve the missing data problem (or the option selected to 'fill in'), as it is a function of $\hat{\boldsymbol{\tau}}_{Rki}$, which has been shown[34] to differ from method to method and from option to option. Using this corrected version of the Hotelling's statistic dramatically improves abnormality detection.

4.02.6.5 Industrial Practice

Industrial applications for batch analysis, monitoring, and fault diagnosis have been reported.[67,68,72,73,75–78] It should be noted here that several companies choose to use the methodology not necessarily for real time monitoring but as a tool for *real time release* of the batch product. This means that the batch run is not monitored in real time as it evolves; rather, immediately after the batch run finishes, the collected process data are passed through the model and the scores for the complete batch run are investigated. If they are within control limits, the product is released. If there is a problem, the product is sent for analysis in the laboratory. This procedure saves the company time and money. The batch run may last 2–3 h but the product analysis may take many more hours. This means that they do not have to waste batches while they are waiting for the results from the laboratory. By checking the process data as soon as the batch run is complete, they can detect problems before starting a new batch.

Other applications of multiway methods to batch analysis, optimization, and control have been reported and will be discussed later in the corresponding sections. Multiway methods and design of experiments can be used[79] to determine optimal process variable trajectories in a batch process to obtain a desired quality property. Multiway and multi-set methods for regression are covered in more detail in Chapter 2.21.

4.02.7 Monitoring Transitions in Continuous Processes

Data collected during transitions can be modeled with the same methodology as the data from batch processes. These can be transitions from grade to grade, start-up of a continuous process, and restart of a continuous process that went on hold because of a technical problem. These transitions, which are frequent in plant operations, lead to important loss of production time, large amounts of off-grade materials, and inconsistent reproducibility of product grades. Ideally, optimal transition policies can be obtained using fundamental models and constrained optimization. However, if theoretical models are not available, but a good database on prior transitions and steady states exists, then multivariate PCA/PLS methods can be used to analyze and improve these transition problems.[63,80] The ideas of centering, scaling, and aligning apply to transition data as well. Industrial applications of projection methods for troubleshooting and for MSPC of transitions have been reported.[47,80] Both the authors used the indicator variable approach for the alignment of the transition trajectories.

The application by Zhang *et al.*[47] is operating in real time since 2002 for a continuous caster. The strand length acts as the indicator variable because it progresses monotonically in time and has the same starting and ending value for each start-cast. All process trajectories of start-casts are synchronized using interpolation based on a set of predefined scales in the strand length. The synchronization scales are determined by a quadratic function of time such that it possesses small intervals at the beginning of the start-cast duration and large intervals at the end. Such nonuniform scales provide a better opportunity for early detection of abnormal situations. Nonuniform scales were also suggested by Kourti[63,64] to better capture important phenomena in the process.

The application above reported by Dofasco, Canada,[47] combines both continuous and batch latent variable technologies into an integrated monitoring solution. Continuous monitoring is used for continuous, run-time casting operation. Batch monitoring is applied during the start-up operation while the process is in the transition to the run-time operation. This integrated application provides a real-time indication of the stability

of the casting operation, which has resulted in improved process safety and economic performance. The transition from the start-up to continuous application happens seamlessly and without the involvement of the operator. In fact, the monitoring graphs (properly adjusted) remain the same. The methodology employs several methodologies from the multivariate statistics theory, including missing data approaches, dynamic trajectory alignment approaches, filtering, and lagging data for preprocessing. An excellent example of FDI is presented, where a complicated case occurs because of a faulty sensor and an abnormal event happening simultaneously. The problem due to faulty sensor is correctly isolated, the faulty sensor is deselected and the system switches to missing data mode, and then the cause of the abnormal situation is also determined. It is such excellent examples that prove the robustness of these methods.

4.02.8 Multistage Operations – Mutiblock Analysis

Very frequently in industry one encounters processes that are multiphase and/or multistage operations. There are often multiple units in a process (e.g., each stage of a multistage synthesis can be regarded as a process unit), and having a control chart relating to each unit rather than one for the whole process could be helpful to operators. Rather than building a model for each unit, one can build a model for the full process that will take into account the interactions between units and their relative importance to the final product quality by weighting them differently. This is the approach of multiblock PLS (MB-PLS).

In the MB-PLS approach, large sets of process variables (**X**) are broken into meaningful blocks, with each block usually corresponding to a process unit or a section of a unit. MB-PLS is not simply a PLS between each **X** block and **Y**. The blocks are weighted in such a way that their combination is most predictive of **Y**. Several algorithms have been reported for multiblock modeling. It is suggested that the reader consult Westerhuis *et al.*[81] and Qin *et al.*[82] for a good introduction and discussion on the subject, where all of the algorithms are presented and compared, and Höskuldsson,[83] where a unified theory is presented for path and multiblock modeling. In the last publication, the data blocks are arranged in a directional path, such that each data block can lead to one or more data blocks. If a block leads to no other blocks, it is designated as an output block.

Multivariate monitoring charts for important subsections of the process, as well as for the entire process, can then be constructed, and contribution plots could be used for fault diagnosis as before. In a multiblock analysis of a batch process for example, one could have the combination of three blocks (**Z**, $\underline{\mathbf{X}}$, and **Y**); block **Z** could include information available on recipes, preprocessing times, hold times, as well as information on the shifts (which operator was in charge) or the vessels used (i.e., which reactor was utilized), $\underline{\mathbf{X}}$ could include process variable trajectories, and **Y** could be quality. Analysis of this type of data could even point to different ways the operators operate the units and relate product quality to operator, or identify different process behavior of vessels and identify faulty vessels, etc. The reader is referred to the work of Munoz *et al.*[72,73] for detailed examples where the multiblock analysis is utilized in batch processes for troubleshooting and for determining the batch operating policies to achieve specific product quality while minimizing the duration of the batch run.

Several alternative ways to perform multiblock appear in commercial software. One approach that is being frequently used to deal with a data structure of several blocks involves two stages: PCA is performed for each one of the **Z** and **X** blocks and then the scores and/or residuals derived from these initial models are related to **Y** with a PLS. In an alternative version, PLS is performed between **Z** and **Y**, **X** and **Y** and the resulting scores are related to **Y**. The users should exercise caution, because these approaches may fail to take into account combinations of variables from different blocks that are most predictive of **Y**. For example, in situations where process parameters **X** are modified to account for variability of raw material properties **Z** (i.e., when **X** settings are calculated as a feedforward control to deviations of **Z**), a PLS between **Z** and **Y** will show that **Z** is not predictive of **Y** variability; similarly, a PLS between **X** and **Y** will show that **X** is not predictive of **Y**; an MB-PLS of [**Z**, **X**] and **Y** will identify the correct model. Finally, MB-PLS handles missing data in a very effective way.

As might be expected in multistage continuous processes, there can be significant time delays between the moment an event occurs in one unit (and therefore affects the variables of that unit) and the moment its effect will become obvious on a product variable at the end of the process. These delays significantly affect the interaction and correlation structures of the process variables and need to be handled by lagged variables created from the original process variables. Data can be time shifted to accommodate time delays between process units.

In some multistage operations, the path of the product through the various process units can be traced easily, and eventually one can relate a specific lot number to several process stages (via an MB-PLS). In such cases, the process conditions of these units can be used to predict the quality of the product. There are situations, however, where a product (or the composition of the effluent stream of a process) is a result of a multistage operation but its path cannot be traced clearly because of mixing of streams from several parallel units in one vessel and then splitting to a number of other vessels. A discussion on monitoring difficult multistage operations can be found in Kourti.[64] In those cases, the best alternative to achieve consistent operation is to monitor each unit, separately, by a PCA model. By assuring a consistent operation per unit, one hopes for a consistent product. Once an unusual event is detected in one unit, one may decide not to mix the product further, or investigate lab quality before proceeding to the next stage.

4.02.9 Process Control Using Latent Variable Methods

Process control refers to a system of measurements and actions within a process intended to ensure that the output of the process conforms to pertinent specifications. Feedback control is a term used to indicate that we are reactive, that is, the corrective action is taken on the process based on information from the process output (e.g., measurements on the deviation of product quality from target are used to adjust process operating conditions). Feedforward control indicates that we are proactive, that is, the process conditions are adjusted based on measured deviations of the input to the process (e.g., information on raw material quality deviations is used to adjust the process operating conditions). Feedback control can also be applied to a unit on top of the feedforward control, as necessary. Latent variable methods have been utilized in process control applications.

4.02.9.1 Feedforward Estimation of Process Conditions

The concept of adjusting the process conditions of one or more units of a process based on measured disturbances (variability from target in the input(s)) is a concept well known to the process systems engineering community for several decades. Input can be the quality of raw material entering the first unit of the process, or some measured intermediate quality entering a unit, or a soft sensor response of a previous unit indicating deviation of expecting performance that needs to be corrected in the next process unit(s).

To the author's knowledge, there are several unpublished examples in the chemical and other industries where models based on multivariate methods are used for feedforward control. In these models, information on the raw data $\mathbf{Z}$ is used to determine the process conditions $\mathbf{X}$ (settings of manipulated variables for several units) or $\underline{\mathbf{X}}$ (trajectories of manipulated variables in a batch process) to achieve the desired quality $\mathbf{Y}$. Sometimes, such information from $\mathbf{Z}$ may simply be used to determine the length of a batch process run, whereas in other cases it may be a multivariate sophisticated scheme that determines a multivariate combination of trajectories for the manipulated variables. To achieve this, appropriate data from historical databases can be used combined with additional experimental design to develop multiblock models relating $\mathbf{Z}$, $\mathbf{X}$ (or $\underline{\mathbf{X}}$), and $\mathbf{Y}$.

4.02.9.2 Feedback Control

Chen *et al.*[84] proposed reduced dimension controllers for continuous processes based on PCA. A PCA model that incorporated time-lagged variables was used and the control objective was expressed in the score space of this PCA model. A model predictive controller was designed and used to control the equivalent score space representation of the process. The score predictive model for the MPC algorithm was built using PLS.

A much more complicated problem is the use of latent variable approaches for complete trajectory manipulation in batch processes for control of end product quality. Control of batch product quality requires the online adjustment of several MVTs such as temperature and material feed rates. Traditional approaches based on detailed theoretical models are based on either nonlinear differential geometric control or online optimization. Many of the schemes suggested in the literature require substantial model knowledge or are computationally intensive and therefore difficult to implement in practice. Empirical modeling offers the advantage of easy model building. Control through complete trajectory manipulation using empirical models is

possible by controlling the process in the reduced space (scores) of a latent variable model rather than in the real space of the manipulated variables. Model inversion and trajectory reconstruction is achieved by exploiting the correlation structure in the MVTs. The batchwise unfolding method presented by Nomikos and MacGregor[37,61,62] is capable of modeling three-way structures generated when formulating the control problem of batch processes using latent variables.

Empirical models for the control of batch product quality where the control action was restricted to only a few movements in the manipulated variables (additional reactant injections) have appeared in the literature and are applied in industrial problems.[85–87] Obviously, these approaches can be used in situations where few adjustments are enough to compensate for disturbances and bring the desired end product quality within specifications.

There have also been several attempts to solve more complicated control problems, where adjustments are required for the full MVTs. One approach solves the problem by segmenting the trajectories into a small number of intervals and forces the behavior of the MVTs over the duration of each interval to follow a zero- or first-order hold. The start of the interval corresponds to a decision point. Control is accomplished by manipulating the slope or the level at the decision point.[88,89]

Flores-Cerrillo and MacGregor[6] recently argued that such a staircase approximation may not be acceptable by plant personnel who traditionally used smooth MVTs. They also point out that model inversion in the control algorithm would be an additional problem in such approaches because a large number of highly correlated control actions need to be determined at every decision point. They suggested a solution to the problem by recognizing that in the range of normal process operation, all the process variable trajectories (both the MVTs and measured variables) are very highly cross-correlated both at a given instant and also over the time duration of the batch. This implies that their behavior can be represented in a much lower-dimensional space using latent variable models based on PCA or PLS. They presented a methodology by which control through complete trajectory manipulation using empirical models only is achieved by controlling the process in the reduced space of a latent variable model, rather in the real space of the manipulated variables. Model inversion and trajectory reconstruction is realized by exploiting the correlation structure in the MVTs captured by a PLS model. The approach was illustrated with a condensation polymerization example for the production of nylon.

In another application,[90] latent variable methodology was used for soft sensor development that could be used to provide FDI capabilities and that could be integrated within a standard model predictive control framework to regulate the growth of biomass within a fermenter. This model predictive controller is shown to provide its own monitoring capabilities that can be used to identify faults within the process and also within the controller itself. Finally, it is demonstrated that the performance of the controller can be maintained in the presence of fault conditions within the process.

4.02.10 MIA for MSPC and Control

MIA describes a set of techniques that employ multivariate statistical methods such as PCA and PLS to analyze images. With MIA, most of the analysis is done in the latent variable space rather than the image space. The objective of the approach is to extract subtle information from the image that is related to product quality, and use such information for prediction monitoring and control. This approach is different from the traditional digital image processing, which involves methods for altering the visual image in some way to make it more visually appealing or to extract information on the shapes, boundaries, or location of various observable features.[91]

Most of the MIA methods had been applied to the analysis of single still images. Bharati and MacGregor[92] evaluated their potential for monitoring time-varying images, and the methods have been applied for online monitoring of lumber defects and pulp and paper quality, for monitoring and control of the amount of coating applied to the base food product and the distribution of the coating among the individual product pieces in snack food industry, and fast estimation of concentrate grade from RBG color images of the froth collected on the surface of flotation cells.[7,8,93,94]

MIA methods and their extension to online monitoring provide a breakthrough in this area and have led to the use of these imaging sensors for monitoring and controlling industrial processes. Digital imagery makes it possible to monitor solids and other heterogeneous materials (e.g., pulp and paper products, polymer films, and multiphase streams); it provides informative, inexpensive, and robust online sensors for the solids industry. Therefore, it opens new ways for the successful monitoring and control of processes, which was traditionally difficult because of lack of sensors. There is potential of utilizing these methods in pharmaceutical industry.

A new machine vision approach for quantitatively estimating and monitoring the appearance and aesthetics of manufactured products is presented by Liu and MacGregor.[95] The methodology is specifically designed to treat the stochastic nature of the visual appearance of many manufactured products. This nondeterministic aspect of product appearance has been an obstacle for the success of machine vision in many industries. The emphasis of this approach is on the consistent and quantitative estimation of continuous variations in visual appearance rather than on classification into discrete classes. This allows for the online monitoring and the eventual feedback control of product appearance. This approach is successfully applied to the estimation and monitoring of the aesthetic quality of manufactured stone countertops. This new machine vision approach combines methods such as wavelet texture analysis, latent variables methods for the quantitative estimation of visual quality from textural information, and MSPC of visual quality.

4.02.11 Concluding Remarks

Multivariate statistical methods have been used successfully by industry in recent years for process monitoring and fault diagnosis, and for process control, in both continuous and batch operations. Online state-of-the-art multivariate industrial applications are a reality. MIA has been adopted for the development of inexpensive sensors for process monitoring and control. The methods are ideal for industrial applications because of their ability to handle noisy measurements for a large number of highly correlated variables. Latent variable methods that model the process space are superb in handling missing data and ideal for outlier detection. Furthermore, modeling the process space makes fault diagnosis possible. Multivariate control charts are easy to understand, whereas contribution plots offer easy diagnostics. Other extensions to the latent variable modeling methodology for process and product transfer and scale-up and for improving operating policies are well under way.[73,96,97]

It should be emphasized that process knowledge is a must. Based on process knowledge one will choose weights and transformations for the variables, will decide on the data required to model routine operation, and will also be able to assess the quality of data supplied by the operators and analysts, or downloaded by data historians[98].

The data that are involved in the models come from several sources, including laboratory data, real-time analyzers, and other data historians. The quality of the data that are archived by the historians is extremely important for reliable results in multivariate analysis. Data acquisition, compression, and reconstruction, if not done properly, may distort the quality of the historical data and make them useless for multivariate analysis (and in some cases any other analysis). To save storage space, data are compressed; not all data acquired at the designated time intervals are stored. Then the data are reconstructed at the user's request in minute averages, hourly averages, etc. The problem is that univariate data compression methods are used to minimize the amount of data that needs to be stored, and they corrupt the multivariate nature of the data, by introducing spurious correlations or destroying real ones. The type and degree of compression should be correctly decided with the vendors. These issues are discussed in detail in Kourti.[64]

The quality of the laboratory data is also important. As a routine procedure, multivariate analysis (PCA) of the laboratory data can be used to detect outliers and erroneous entries. Also, the number of decimal points that are included when data are recorded is an important issue to the quality of the model. Although it is obvious that quantities should be recorded with the precision they are measured, this is not always the practice; for example, in some cases, although a quantity is available with 1–2 decimal points, the value is rounded to the closest integer. If the rounded integer value is included as $\mathbf{Y}$ in a PLS model, the true versus predicted plots will look segmented and the percent model prediction of $\mathbf{Y}$ may not be good; keep in mind that the process data tend to tell a more detailed story than the quality data. In other words, the process data may be able to tell the difference between $y = 9$ and $y = 9.4$, but if both are rounded to 9.0, the model is forced to be noisy.

Finally, some analyzer manufacturers do not supply the raw spectral data so that they can be archived together with other data of the process. Rather they supply the calculated quality value that the analyzer is used for. It is important that the plant has access to the raw spectral data, so that they can be used in the future, together with other process data, for multivariate analysis to investigate production issues.

By their nature, models developed from data based statistical methods are highly dependent on the data quality; the data IS the model. So emphasis on good dataset quality is a must.

Multivariate Analysis has a very important role to play in the pharmaceutical industry under the current regulatory framework, as it is evident by the introduction of concepts like Design Space and Control Strategy. The International Conference on Harmonization (ICH Draft Step 4, Q8 (R1) Pharmaceutical development revision 1, 11-November-2008) defines the Design Space as ": the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality" [99]

References

1. Kourti, T. Process Analysis and Abnormal Situation Detection: From Theory to Practice. *IEEE Control Syst.* **2002**, *22* (5), 10–25.
2. Miletic, I.; Quinn, S.; Dudzic, M.; Vaculik, V.; Champagne, M. An Industrial Perspective on Implementing On-Line Applications of Multivariate Statistics. *J. Process Control* **2004**, *14* (8), 821–836.
3. Kourti, T. Process Analytical Technology and Multivariate Statistical Process Control. Index of Wellness of Product and Process – Part 1. *J. Process Anal. Technol.* **2004**, *1* (1), 13–19.
4. Kourti, T.; MacGregor, J. F. Recent Developments in Multivariate SPC Methods for Monitoring and Diagnosing Process and Product Performance. *J. Qual. Technol.* **1996**, *28* (4), 409–428.
5. Kourti, T. Process Analytical Technology Beyond Real-Time Analyzers: The Role of Multivariate Analysis. *Crit. Rev. Anal. Chem.* **2006**, *36*, 257–278.
6. Flores-Cerrillo, J.; MacGregor, J. F. Control of Batch Product Quality by Trajectory Manipulation Using Latent Variable Models. *J. Process Control* **2004**, *14*, 539–553.
7. Prats-Montalbán, J. M.; Ferrer, A. Integration of Colour and Textural Information in Multivariate Image Analysis: Defect Detection and Classification Issues. *J. Chemom.* **2007**, *21* (1–2), 10–23.
8. Kourti, T. Symposium Report on Abnormal Situation Detection and Projection Methods. *Chemom. Intell. Lab. Syst.* **2005**, *76*, 215–220.
9. Shewhart, W. A. *Economic Control of Quality of Manufactured Product*; Van Nostrand: Princeton, NJ, 1931.
10. Montgomery, D. C.; Runger, G. C. *Applied Statistics and Probability for Engineers*; Wiley & Sons: New York, 1994.
11. Montgomery, D. C.; Runger, G. C.; Hubele, N. F. *Engineering Statistics*; Wiley & Sons: New York, 1998.
12. Montgomery, D. C. *Introduction to Statistical Quality Control*, 5th ed.; Wiley & Sons: New York, 2005.
13. Hunter, J. S. Exponentially Weighted Moving Average. *J. Qual. Technol.* **1986**, *18*, 203–210.
14. Han, D.; Tsung, F. Detection and Diagnosis of Unknown Abrupt Changes Using CUSUM Multi-Chart Schemes. *Sequential Anal.* **2007**, *26* (3), 225–249.
15. Harris, T. J.; Ross, W. H. Statistical Process Control Procedures for Correlated Observations. *Can. J. Chem. Eng.* **1991**, *69*, 139–148.
16. Tracy, N. D.; Young, J. C.; Mason, R. L. Multivariate Control Charts for Individual Observations. *J. Qual. Technol.* **1992**, *24*, 88–95.
17. Wierda, S. J. Multivariate Statistical Process Control – Recent Results and Directions for Future Research. *Statistica Neerlandica* **1994**, *48* (2), 147–168.
18. Pignatello, J. J. Jr.; Runger, G. C. Comparisons of Multivariate CUSUM Charts. *J. Qual. Technol.* **1990**, *22*, 173–186.
19. Lowry, C. A.; Woodall, W. H.; Champ, C. W.; Rigdon, S. E. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics* **1992**, *34*, 46–53.
20. Healy, J. D. A Note on Multivariate CUSUM Procedures. *Technometrics* **1987**, *29*, 409–412.
21. Crosier, R. B. Multivariate Generalisations of Cumulative Sum Quality-Control Schemes. *Technometrics* **1988**, *30*, 291–303.
22. Venkatasubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S. N. A Review of Process Fault Detection and Diagnosis. Part I: Quantitative Model-Based Methods. *Comput. Chem. Eng.* **2003**, *27*, 293–311.
23. Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S. N. A Review of Process Fault Detection and Diagnosis. Part II. Qualitative Models and Search Strategies. *Comput. Chem. Eng.* **2003**, *27*, 313–326.
24. Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S. N.; Yin, K. A. Review of Process Fault Detection and Diagnosis. Part III: Process History Based Methods. *Comput. Chem. Eng.* **2003**, *27*, 327–346.
25. Yoon, S.; MacGregor, J. F. Relationships between Statistical and Causal Model-Based Approaches to Fault Detection and Isolation. *Am. Inst. Chem. Eng. J.* **2000**, *46*, 1813–1824.
26. Kourti, T. Application of Latent Variable Methods to Process Control and Multivariate Statistical Process Control in Industry. *Int. J. Adapt. Control Signal Process.* **2005**, *19*, 213–246.
27. Wold, H. Non-linear Estimation by Iterative Least Squares Procedures. In *Research Papers in Statistics*; David, F., Ed.; Wiley: New York, 1966; pp 411–444.
28. Bro, R.; Kjeldahl, K.; Smilde, A. K.; Kiers, H. A. L. Cross-Validation of Component Models: A Critical Look at Current Methods. *Anal. Bioanal. Chem.* **2008**, *390*, 1241–1251.

29. Wold, S.; Høy, M.; Martens, H.; Trygg, J.; Westad, F.; MacGregor, J.; Wise, B. M. The PLS Model Revisited. *J. Chemom.* **2008**, Published Online: 9 July 2008, DOI: 10.1002/cem.1171.
30. Bro, R.; Eldén, L. PLS Works. *J. Chemom.* **2008**, Published Online: 18 July 2008, DOI: 10.1002/cem.1177.
31. Burnham, A. J.; Viveros, R.; MacGregor, J. F. Frameworks for Latent Variable Multivariate Regression. *J. Chemom.* **1996**, *10*, 31–45.
32. Nelson, P. R. C.; Taylor, P.; MacGregor, J. F. Missing Data Methods in PCA and PLS: Score Calculations with Incomplete Observations. *Chemom. Intell. Lab. Syst.* **1996**, *35* (1), 45–65.
33. Arteaga, F.; Ferrer, A. Dealing with Missing Data in MSPC: Several Methods, Different Interpretations, Some Examples. *J. Chemom.* **2002**, *16*, 408–418.
34. García-Muñoz, S.; MacGregor, J. F.; Kourti, T. Model Predictive Monitoring for Batch Processes with Multivariate Methods. *Ind. Eng. Chem. Res.* **2004**, *43*, 5929–5941.
35. Jackson, J. E. *A User's Guide to Principal Components*; Wiley & Sons: New York, 1991.
36. Jackson, J. E.; Mudholkar, G. S. Control Procedures for Residual Associated with Principal Components Analysis. *Technometrics* **1979**, *21* (3), 341–349.
37. Nomikos, P.; MacGregor, J. F. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics* **1995**, *37* (1), 41–59.
38. Box, G. E. P. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems: Effects of Inequality of Variance in One-Way Classification. *Ann. Math. Stat.* **1954**, *25*, 290–302.
39. Kresta, J.; MacGregor, J. F.; Marlin, T. E. Multivariate Statistical Monitoring of Process Operating Performance. *Can. J. Chem. Eng.* **1991**, *69*, 35–47.
40. Wise, B. M.; Veltkamp, D. J.; Ricker, N. L.; Kowalski, B. R.; Barnes, S.; Arakali, V. Application of Multivariate Statistical Process Control (MSPC) to the West Valley Slurry-Red Ceramic Melter Process. *Proceedings of the Waste Management 1991*, Tucson, Arizona, 1991.
41. Skagerberg, B.; MacGregor, J. F.; Kiparissides, C. Multivariate Data Analysis Applied to Low-Density Polyethylene Reactors. *Chemom. Intell. Lab. Syst.* **1992**, *14*, 341–356.
42. Geladi, P.; Manley, M.; Lestander, T. Scatter Plotting in Multivariate Data Analysis. *J. Chemom.* **2003**, *17*, 503–511.
43. Schlags, C. E.; Anani, A. S. Application of Dynamic PLS to Waste Treatment Process Analysis. *AIChE Meeting*, Section 224e, Miami Beach, Florida, November 1998.
44. Piovoso, M. J.; Kosanovich, K. A.; Yuk, J. P. Process Data Chemometrics. *IEEE Trans. Instrum. Meas.* **1992**, *41* (2), 262–268.
45. Bendwell, N. Les techniques multivariables au service du controle des procedes. *Vecteur Environ.* **2000**, *33* (2).
46. Vaculik, V.; MacCuish, B.; Dudzic, M.; Miletic, I.; Hough, M.; Smith, K. Innovative Monitoring Technology Improves Caster Operation at Dofasco. *Iron and Steelmaker* **2001**, *28* (9), 23–28.
47. Zhang, Y.; Dudzic, M.; Vaculik, V. Integrated Monitoring Solution to Start-Up and Run-Time Operations for Continuous Casting. *Annu. Rev. Control* **2003**, *27*, 141–149.
48. Ivanov, I.; Wold, S.; Kettaneh-Wold, N. *Multivariate Modeling for a Multigrade Application,Control Systems 2002*, Stockholm, Sweden, 3–5 June 2002.
49. Berglund, A.; Kettaneh-Wold, N.; Uppgard, L. L.; Wold, S.; Bendwell, N.; Cameron, D. R. The GIFI Approach to Nonlinear PLS Modelling. *J. Chemom.* **2001**, *15*, 321–336.
50. Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building*; Wiley Series in Probability and Mathematical Statistics; Wiley & Sons: New York, 1978.
51. Nijhuis, S.; de Jong, B. G. M. Vandeginste. Multivariate Statistical Process Control in Chromatography. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 51–62.
52. Nelson, P. R. The Treatment of Missing Measurements in PCA and PLS Models. Ph.D. thesis, Department of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada, 2002.
53. Nelson, P. R. C.; MacGregor, J. F.; Taylor, P. The Impact of Missing Measurements on PCA and PLS Prediction and Monitoring Applications. *Chemom. Intell. Lab. Syst.* **2006**, *80*, 1–12.
54. Christie, O. H. J. Data Laundering by Target Rotation in Chemistry-Based Oil Exploration. *J. Chemom.* **1996**, *10*, 453–461.
55. Dayal, B. S.; MacGregor, J. F. Recursive Exponentially Weighted PLS and Its Applications to Adaptive Control and Prediction. *J. Process Control* **1997**, *7* (3), 169–179.
56. Qin, S. J. Recursive PLS Algorithms for Adaptive Data Modelling. *Comput. Chem. Eng.* **1998**, *22* (4–5), 503–514.
57. Li, W.; Yue, H. H.; Valle-Cervantes, S.; Qin, S. J. Recursive PCA for Adaptive Process Monitoring. *J. Process Control* **2000**, *10*, 471–486.
58. Quinn, S. L.; Vaculik, V. Improving the Desulfurization Process Using Adaptive Multivariate Statistical Modeling. *AISE Steel Technol.* **2002**, *79* (10), 37–41.
59. Wang, X.; Kruger, U.; Lennox, B. Recursive Partial Least Squares Algorithms for Monitoring Complex Industrial Processes. *Control Eng. Pract.* **2003**, *11*, 613–632.
60. Capron, X.; Walczak, B.; de Noord, O. E.; Massart, D. L. Selection and Weighting of Samples in Multivariate Regression Model Updating. *Chemom. Intell. Lab. Syst.* **2005**, *76* (2), 205–214.
61. Nomikos, P.; MacGregor, J. F. Monitoring of Batch Processes Using Multi-Way Principal Component Analysis. *AIChE J.* **1994**, *40* (8), 1361–1375.
62. Nomikos, P.; MacGregor, J. F. Multiway Partial Least Squares in Monitoring Batch Processes. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 97–108.
63. Kourti, T. Multivariate Dynamic Data Modelling for Analysis and Statistical Process Control of Batch Processes, Start-Ups and Grade Transitions. *J. Chemom.* **2003**, *17* (1), 93–109.
64. Kourti, T. Abnormal Situation Detection, Three Way Data and Projection Methods – Robust Data Archiving and Modeling for Industrial Applications. *Annu. Rev. Control* **2003**, *27* (2), 131–138.
65. Camacho, J.; Picó, J.; Ferrer, A. Bilinear Modelling of Batch Processes. Part I: Theoretical Discussion. *J. Chemom.* **2008**, *22* (5), 299–308.
66. Camacho, J.; Picó, J.; Ferrer, A. Bilinear Modelling of Batch Processes. Part II: A Comparison of PLS Soft-Sensors. *J. Chemom.*, **2008**, *22* (10), 533–547.

67. Kourti, T.; Lee, J.; MacGregor, J. F. Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control. *Comput. Chem. Eng.* **1996**, *20* (Suppl.), S745–S750.
68. Neogi, D.; Schlags, C. E. Multivariate Statistical Analysis of an Emulsion Batch Process. *Ind. Eng. Chem. Res.* **1998**, *37*, 3971–3979.
69. Kassidas, A.; MacGregor, J. F.; Taylor, P. A. Synchronization of Batch Trajectories Using Dynamic Time Warping. *AIChE J.* **1998**, *44*, 864–875.
70. Taylor, P. A. *Personal communication, Computing and Software Department*, McMaster University, Hamilton, Ontario, Canada, May 1998.
71. Westerhuis, J. A.; Kassidas, A.; Kourti, T.; Taylor, P. A.; MacGregor, J. F. On-Line Synchronization of the Trajectories of Process Variables for Monitoring Batch Processes with Varying Duration. *Presented at the 6th Scandinavian Symposium on Chemometrics*, Porsgrunn, Norway,15–19 August 1999.
72. García-Muñoz, S.; Kourti, T.; MacGregor, J. F.; Mateos, A. G.; Murphy, G. Troubleshooting of an Industrial Batch Process Using Multivariate Methods. *Ind. Eng. Chem. Res.* **2003**, *42*, 3592–3601.
73. García-Muñoz, S.; MacGregor, J. F.; Neogi, D.; Latshaw, B. E.; Mehta, S. Optimization of Batch Operating Policies. Part II. Incorporating Process Constraints and Industrial Applications. *Ind. Eng. Chem. Res.* **2008**, *47* (12), 4202–4208.
74. Tomasi, G.; van den Berg, F.; Andersson, C. Correlation Optimized Warping and Dynamic Time Warping as Preprocessing Methods for Chromatographic Data. *J. Chemom.* **2004**, *18*, 231–241.
75. Kosanovich, K. A.; Piovoso, M. J.; Dahl, K. S. Multi-Way PCA Applied to an Industrial Batch Process. *Ind. Chem. Res.* **1995**, *35*, 138–146.
76. Lennox, B.; Montague, G. A.; Hiden, H. G.; Kornfeld, G.; Goulding, P. R. Process Monitoring of an Industrial Fed-Batch Fermentation. *Biotechnol. Bioeng.* **2001**, *74* (2), 125–135.
77. Kourti, T.; Nomikos, P.; MacGregor, J. F. Analysis, Monitoring and Fault Diagnosis of Batch Processes Using Multiblock and Multiway PLS. *J. Process Control,* **1995**, *5*, 277–284.
78. Nomikos, P. Detection and Diagnosis of Abnormal Batch Operations Based on Multiway Principal Component Analysis. *ISA Trans.* **1996**, *35*, 259–267.
79. Duchesne, C.; MacGregor, J. F. Multivariate Analysis and Optimization of Process Variable Trajectories for Batch Processes. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 125–137.
80. Dushesne, C.; Kourti, T.; MacGregor, F. Multivariate SPC for Start-Ups, and Grade Transitions. *AIChE J.* **2002**, *48* (12), 2890–2901.
81. Westerhuis, J.; Kourti, T.; MacGregor, J. F. Analysis of Multiblock and Hierarchical PCA and PLS Models. *J. Chemom.* **1998**, *12*, 301–321.
82. Qin, J. S.; Valle, S.; Piovoso, M. J. On Unifying Multiblock Analysis with Application to Decentralized Process Monitoring. *J. Chemom.* **2001**, *15*, 715–742.
83. Höskuldsson, A. Multi-Block and Path Modelling Procedures. *J. Chemom.* **2008**, *22* (11–12), 571–579.
84. Chen, G.; McAvoy, T.; Piovoso, M. J. A Multivariate Statistical Controller for On-Line Quality Improvement. *J. Process Control* **1998**, *8* (2), 139–149.
85. Yabuki, Y.; Nagasawa, T.; MacGregor, J. F. An Industrial Experience with Product Quality Control in Semi-Batch Processes. *Comput. Chem. Eng.* **2000**, *24*, 585–590.
86. Flores-Cerrillo, J.; MacGregor, J. F. Control of Particle Size Distributions in Emulsion Semi-Batch Polymerization Using Mid-Course Correction Policies. *Ind. Eng. Chem. Res.* **2002**, *41*, 1805–1814.
87. Flores-Cerrillo, J.; MacGregor, J. F. Within Batch and Batch to Batch Inferential Adaptive Control of Batch Reactors: A PLS Approach. *Ind. Eng. Chem. Res.* **2003**, *42*, 3334–3345.
88. Chin, S. I.; Lee, S. K.; Lee, J. H. A Technique for Integrated Quality Control, Profile Control and Constraint Handling of Batch Processes. *Ind. Eng. Chem. Res.* **2000**, *39*, 693–705.
89. Russell, S. A.; Kesavan, P.; Lee, J. H.; Ogunnaike, B. A. Recursive Data Based Prediction and Control of Batch Product Quality. *AIChE J.* **1998**, *44*, 2442–2458.
90. Zhang, H.; Lennox, B. Integrated Condition Monitoring and Control of Fed-Batch Fermentation Processes. *J. Process Control* **2004**, *14*, 41–50.
91. Geladi, P.; Grahn, H. *Multivariate Image Analysis*; Wiley & Sons: Chichester, UK, 1996.
92. Bharati, M. H.; MacGregor, J. F. Multivariate Image Analysis for Real Time Process Monitoring and Control. *Ind. Eng. Chem. Res.* **1998**, *37*, 4715–4724.
93. Yu, H.; MacGregor, J. F. Multivariate Image Analysis and Regression for Prediction of Coating Content and Distribution in the Production of Snack Foods. *Chemom. Intell. Lab. Syst.* **2003**, *67*, 125–144.
94. Liu, J. J.; MacGregor, J. F.; Duchesne, C.; Bartolacci, G. Flotation Froth Monitoring Using Multiresolutional Multivariate Image Analysis. *Miner. Eng.* **2005**, *18* (1), 65–76.
95. Liu, J.; MacGregor, J. F. On the Extraction of Spectral and Spatial Information from Images. *Chemom. Intell. Lab. Syst.* **2007**, *85* (1), 119–130.
96. Garcia Munoz, S.; Kourti, T.; MacGregor, J. F.; Apruzzese, F.; Champagne, M. Optimization of Batch Operating Policies. Part I. Handling Multiple Solutions. *Ind. Eng. Chem. Res.* **2006**, *45*, 7856–7866.
97. Garcia Munoz, S.; MacGregor, J. F.; Kourti, T. Product Transfer between Sites Using Joint Y_PLS. *Chemom. Intell. Lab. Syst.* **2005**, *79*, 101–114.
98. Miletic, I.; Boudreau, F.; Dudzic, M.; Kotuza, G.; Ronholm, L.; Vaculik, V.; Zhang, Y. Experiences in Applying Data-Driven Modelling Technology to Steelmaking Processes. *Can. J. Chem. Eng. 86*, 937–946.
99. Lepore, J.; Spavins, J. PQLI Design Space. *J Pharm Innov* **2008**, (3), 79–87.

Biographical Sketch

Dr. Theodora Kourti is a Senior Technical Director in GlaxoSmithKline and holds an adjunct professor appointment with McMaster University, Canada. She has an extensive experience with methodology for process and product improvement and abnormal situation detection in process industries, and has been involved in more than 90 major industrial applications in North America, Europe, and Asia, in more than 40 diverse industries such as chemicals, pharmaceuticals, semiconductor, mining, pulp and paper, petrochemicals, photographic, and steel industry. This contribution has been recognized, among others, the 2003 University – Industry Synergy Award for Innovation, given by the Natural Science & Engineering Research Council of Canada. She has published extensively in this area and has provided training for numerous industrial practitioners. Previous employment includes Polysar (Sarnia, ON), Exxon Research and Engineering (USA), and Esso Rotterdam Refinery (The Netherlands). She received her diploma in chemical engineering from Aristotle University of Thessaloniki, Greece, and her Ph.D. in polymer reaction engineering and control from McMaster University, Canada.