Multivariate SPC Charts for Monitoring Batch Processes

Paul NOMIKOS and John F. MACGREGOR

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

The problem of using time-varying trajectory data measured on many process variables over the finite duration of a batch process is considered. Multiway principal-component analysis is used to compress the information contained in the data trajectories into low-dimensional spaces that describe the operation of past batches. This approach facilitates the analysis of operational and quality-control problems in past batches and allows for the development of multivariate statistical process control charts for on-line monitoring of the progress of new batches. Control limits for the proposed charts are developed using information from the historical reference distribution of past successful batches. The method is applied to data collected from an industrial batch polymerization reactor.

KEY WORDS: Control charts; On-line monitoring; Polymerization; Principal-component analysis; Reference distribution; Statistical process control.

Batch and semi-batch processes play an important role in the production and processing of high-quality speciality materials and products. Examples include the production of polymers, pharmaceuticals, and biochemicals; the separation and transformation of materials by batch distillation and crystallization; and the processing of materials by injection molding. In general, a batch process is a finite-duration process consisting of the following steps: Charging of the batch vessel with a specified recipe of materials; processing under controlled conditions during which process variables such as temperatures, pressures, agitation, and feedrates are varied according to specified time trajectories; and finally discharging of the product. On completion of the batch, a range of quality measurements is usually made at the quality-control laboratory on a sample of the product.

Batch processes generally exhibit some batch-to-batch variation arising from such things as deviations of the process variables from their specified trajectories, errors in the charging of the recipe of materials, and disturbances arising from variations in impurities. Abnormal conditions that develop during these batch operations can lead to the production of at least one batch or a whole sequence of batches with poor-quality product if the problem is not detected and remedied. In spite of this, most industrial batch processes are run without any effective form of real-time, on-line monitoring to ensure that the batch is progressing in a manner that will lead to a high-quality product or to detect and indicate faults that can be corrected prior to completion of the batch or can be corrected in subsequent batches. For the most part, they rely simply on the precise sequencing and automation of the major steps

in each batch run. Some effort has been made in industry to use relational data base software to try to uncover particular attributes of the measurement trajectories, such as the timing of valve openings or the maximum temperature or pressure attained during an interval, that appear to affect product quality and then to monitor these attributes.

The application of statistical process control (SPC) charts to batch processes has been very limited. Most SPC methods use only the product quality measurements obtained at the end of each batch (e.g., Vander Wiel, Tucker, Faltin, and Doganaksoy 1992) and therefore monitor only the batch-to-batch variation. Hahn and Cockrum (1987) investigated the case in which one also has a few quality measurements taken during the batch run. Marsh and Tucker (1991) recognized that the process variable measurements taken during a batch run, although transient in nature, do follow a certain dynamic pattern, and they proposed a simple SPC technique for monitoring a single measurement variable. Konstantinov and Yoshida (1992) (temporal shapes of time profiles) and Holloway and Krogh (1992) (trajectory encoding) applied qualitative reasoning to tackle the monitoring problem of dynamic processes. Both tried to determine if the on-line observations received from the process up to the present time are consistent with some acceptable dynamic behavior of the system. The lack of statistical reasoning in their work and their univariate orientation are their main drawbacks. Bonvin and Rippin (1990) used target-factor analysis to identify on-line possible reaction stoichiometries from measured composition or thermal data and to detect any batch runaways (Prinz and Bonvin 1992).

41

With on-line computers connected to most batch processes, massive amounts of data are being collected routinely during the batch on many easily measured process variables such as temperatures, pressures, and flowrates. One may have measurements on up to 50 or more variables every few seconds throughout the entire history of a batch. Furthermore, there is usually a history of many past successful and some unsuccessful batches. Not only is the relationship among all of the variables at any one time important, but so is the entire past history of the trajectories of all these variables. The history of the process-variable trajectories during a batch provide a "fingerprint" for each batch, and it should be possible from these data to build an empirical model to characterize the operation of successful batch runs. The major difficulties are how to handle the many measured process variables, their time-varying and highly correlated structure, and the nonlinear finite-time nature of batch operations.

To handle such large multivariable problems in continuous processes operating around a fixed target, multivariate SPC charts based on principal-component analysis (PCA) and partial least squares or projection to latent structures (PLS) have been developed (Kresta, MacGregor, and Marlin 1991; Skagerberg, MacGregor, and Kiparissides 1992; Miller, Swanson, and Heckler 1993; MacGregor, Jaeckle, Kiparissides, and Koutoudi 1994). These methods can use the many highly correlated process measurements that are being continuously collected. The information contained in these data is projected into low-dimensional spaces defined by latent vectors, and control charts that are simple to present and easy to interpret have been proposed and are now used in industry. Furthermore, the diagnostic capabilities of these multivariate methods have been shown to greatly enhance one's ability to isolate assignable causes for violations of these charts (Miller et al. 1993; MacGregor et al. 1994).

Recently MacGregor and Nomikos (1992) and Nomikos and MacGregor (1994) employed multiway PCA (MPCA) to extend multivariate SPC methods to batch processes. By again projecting the information contained in the process-variable trajectories down into low-dimensional latent-variable spaces, both the variable correlations and their time histories could be summarized in a few plots. These multivariate approaches are based on analyzing a historical reference distribution of the measurement trajectories from past successful batches. The variation in the trajectories among those batches (common-cause variation) is characterized in a reduced latent-vector space using MPCA. The behavior of new batches is then compared to this reference distribution to test the following hypothesis: H_o: The on-line measurements of the process-variable trajectories up to the current time in a new batch are consistent with normal batch operation as defined by the historical reference distribution.

The objectives of this article are (a) to present an overview and some new variations of the MPCA method

that have been proposed for the analysis and on-line monitoring of batch processes, (b) to establish statistical control limits for the multivariate SPC charts that arise from these methods, and (c) to illustrate the approach with an application of the analysis and monitoring of an industrial batch polymerization reactor.

The article is organized as follows. MPCA applied to batch processes is outlined in Section 1, and its use in the post-analysis of past batch polymerization runs is illustrated in Section 2. The selection of a suitable historical reference distribution of past normal batches is illustrated and modeled via MPCA in Section 3. Several variations of multivariate SPC charts for on-line monitoring of the progress of new batches are presented in Section 4, and statistical control limits for each of the multivariate SPC charts are developed in Section 5. Two examples of monitoring new batches are given in Section 6, and some engineering issues and future directions are discussed in Section 7.

1. MPCA FOR MODELING BATCH PROCESSES

Consider the problem at hand—namely, analyzing a historical set of batch trajectory data. In a typical batch run, $j = 1, 2, \ldots, J$ variables are measured at $k = 1, 2, \ldots, K$ time intervals throughout the batch. Similar data will exist on several ($i = 1, 2, \ldots, I$) similar process batch runs. This vast amount of data can be organized into a three-way array $\underline{X}(I \times J \times K)$ as illustrated in Figure 1. The dif-
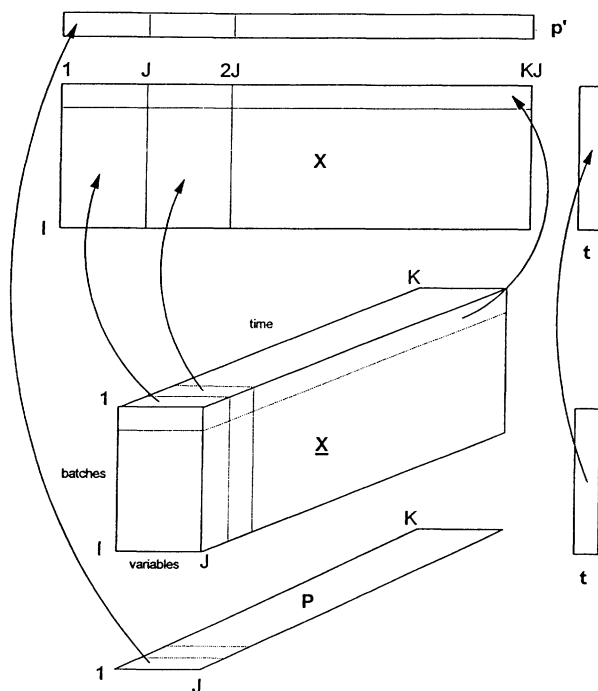

Figure 1. Arrangement of Batch Data in MPCA (lower part) and its Equivalent PCA Form (upper part). The three-way array $\underline{X}$ unfolds into a matrix X where a normal PCA analysis can be performed to extract the t scores and p loadings.

ferent batch runs have been arranged along the vertical axis, the measurement variables are across the horizontal axis, and their time evolution occupies the third dimension. Each horizontal slice through this array is a $(J \times K)$ matrix containing the trajectories of all the variables from a single batch (i). Each vertical slice is an $(I \times J)$ matrix representing the values of all the variables for all the batches at a common time interval (k).

Several multidimensional statistical methods have been proposed for decomposing such data arrays into the sum of a few products of vectors and matrices and to summarize the variation of the data in the reduced dimensions of these spaces. MPCA was introduced by Geladi, Esbenson, and Ohman (1987) and was successfully applied in image analysis (Geladi et al. 1989) and to some cases in chemometrics (Smilde and Doornbos 1991). Other multiway methods (Geladi 1989) such as the Tucker model, the PARAFAC model (Smilde and Doornbos 1991), the canonical decomposition, the three-mode factor analysis (Zeng and Hopke 1990), and the tensor rank (Sanchez and Kowalski 1990) have been proposed for special situations. MacGregor and Nomikos (1992) and Nomikos and MacGregor (1994), however, were able to show that MPCA was well suited to handle multiway batch data.

MPCA is equivalent to unfolding the three-dimensional array $\underline{\mathbf{X}}$ slice by slice (three possible ways), rearranging the slices into a large two-dimensional matrix $\mathbf{X}$ (two possible ways), and then performing a regular PCA. Each of these six possible rearrangements of the data array $\underline{\mathbf{X}}$ into a large data matrix $\mathbf{X}$, followed by a PCA on the matrix $\mathbf{X}$, corresponds to looking at a different type of variability. For analyzing and monitoring batch processes, the most meaningful way of unfolding the array $\underline{\mathbf{X}}$ is to arrange its vertical slices, corresponding to each point of time, side by side into a two-dimensional matrix $\mathbf{X}(I \times JK)$ with the vertical slice corresponding to the first time interval at the left side (Fig. 1). The data are then mean centered and scaled prior to performing a PCA. This unfolding is particularly meaningful because, by subtracting the mean of each column of this matrix $\mathbf{X}$, we are in effect subtracting the mean trajectory of each variable, thereby removing the main nonlinear and dynamic components in the data. A PCA performed on these mean-corrected data is therefore a study of the variation in the time trajectories of all the variables in all batches about their mean trajectories. Figure 2 shows the measurements for a single variable over the whole-batch duration from 36 normal batches, where one can see clearly the kind of variation that MPCA will explain at each time interval. The variables in each column of $\mathbf{X}$ are also scaled to unit variance by dividing by their standard deviation so as to handle differences in the measurement units between variables and to give equal weight to each variable at each time interval. If one wishes to give greater or less weight to any particular variable, however, or to any particular period of time in the batch, these weights are easily changed. Another
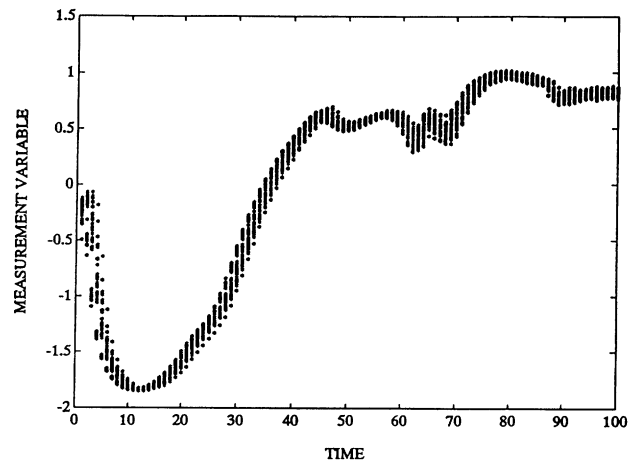

Figure 2. Measurement Trajectories of a Single Variable (temperature) for 36 Normal Batches. At each time interval there are 36 observations, one from each batch. The black band that all these observations create is the variability that MPCA tries to explain.

way of scaling that gives similar results to what we use in this article is scaling each variable at each time interval by its overall (throughout the batch duration) standard deviation.

This form of MPCA decomposes the data ($\underline{\mathbf{X}}$ or $\mathbf{X}$) into a summation of R products of score vectors ($\mathbf{t}$) and loadings matrices ($\mathbf{P}$ or $\mathbf{p}$), plus residuals ($\underline{\mathbf{E}}$ or $\mathbf{E}$), which are as small as possible in a least squares sense.

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{t}_r \otimes \mathbf{P}_r + \underline{\mathbf{E}} \qquad \text{or} \qquad \mathbf{X} = \sum_{r=1}^{R} \mathbf{t}_r \mathbf{p}_r' + \mathbf{E}.$$

Figure 1 illustrates the correspondence between the scores and loadings of MPCA performed on the array $\underline{\mathbf{X}}$ and those of a PCA performed on the equivalent unfolded matrix $\mathbf{X}$. The NIPALS (nonlinear iterative partial least squares) algorithm for sequential computing the dominant principal components, is given in Appendix A.

The t vectors ($\mathbf{t}$) are orthogonal and the loading vectors $\mathbf{p}$ (unfolded $\mathbf{P}$) are orthonormal. Usually, a few principal components can express most of the variability in the data when the variables are highly correlated, as in this case, and can point out any similarities and dissimilarities among batches. Each element of the t vector ($\mathbf{t}$) corresponds to a single batch and depicts the overall variability of this batch with respect to the other batches in the data base throughout the whole batch duration. The loading vectors ($\mathbf{p}$) provide the directions of maximum variability and give a simpler and more parsimonious description of the covariance structure of the data. Each loading vector ($\mathbf{p}$), as one can see from the unfolded matrix $\mathbf{X}$ (Fig. 1), summarizes the time variation of the measurement variables around their average trajectories, and its elements are the weights applied to each variable at each time interval within a batch to give the t score for that batch. The power of MPCA results from

using the joint covariance matrix of the variable deviations from their main trajectories. Thus it uses not just the magnitude of the deviation of each variable from its mean trajectory but also the contemporaneous correlation among all of the variables over the time history of the batches.

1.1 Selecting the Number of Principal Components

The number of principal components needed to build an MPCA model that describes adequately the normal behavior of a batch operation can be found with several criteria. These criteria range all the way from significance tests to graphical procedures (Jackson 1991). One quick and dirty criterion is the broken-stick rule (Jolliffe 1986). This is based on the fact that if a line segment of unit length is randomly divided into z segments the expected length of the rth longest segment is

$$G = 100 \frac{1}{z} \sum_{i=r}^{z} 1/i.$$

As long as the percentage of variance explained by each principal component is larger than the corresponding G, then one can retain the corresponding principal component. The number of segments is the maximum possible rank of $\mathbf{X}$, $z = \min(I, KJ)$, and the rule should be applied only to unit variance-scaled matrices. This criterion is rather crude but still is a quick method to judge if a principal component adds any structural information about the variance in the data or explains only noise.

When the purpose of a PCA analysis is to construct a model that will be used on future observations, then the suggested criterion to obtain the optimum number of components is cross-validation (Efron 1983, 1986; Stone 1974). Cross-validation shows how the prediction power of a PCA model increases as one adds more principal components. It is a simple, but computationally lengthy, procedure similar to the jackknifing method. Given a data base of I normal batches with J variables and K time intervals, the unfolded $\mathbf{X}$ matrix has dimensions $I \times JK$. After scaling the matrix $\mathbf{X}$, one batch is excluded from the data base and a PCA model is built with the remaining $(I - 1)$ batches. This is done for all the batches in the data base, and each time the sum of the squared prediction errors after each principal component is recorded for the batch not included in the model building. At the end, these sum-of-the-squared-prediction errors corresponding to each principal component (r) are added for all the batches to give the Press$_r$.

One way to choose the model dimension is to select the one with minimum Press, but this has been shown to have poor statistical properties (Osten 1988). Wold (1978) and Krzanowski (1983, 1987) proposed two criteria for choosing the optimal number of principal components. Wold checked the ratio $R = \text{Press}_r/\text{RSS}_{r-1}$, where RSS$_r$ is the residual sum of squares after the rth principal com-

ponent based on the PCA model, which is built using the whole data base. This criterion compares the prediction power of a model based on r principal components with the squared differences between observed and estimated data using $r - 1$ principal components. A value of R larger than unity suggests that the rth component did not improve the prediction power of the model and it is better to use only $r - 1$ components. Krzanowski suggested the ratio

$$W = ((\text{Press}_{r-1} - \text{Press}_r)/D_m)/(\text{Press}_r/D_r)$$
$$D_m = I + JK - 2r$$
$$D_r = JK(I - 1) - \sum_{i=1}^{r} I + JK - 2i,$$

where D_m and D_r are numbers indicating the degrees of freedom required to fit the rth component and the degrees of freedom remaining after fitting the rth component, respectively. This statistic is similar to the F test for the inclusion of an additional variable in a linear regression model. It gives the ratio between the improvement in predictive power by adding the rth component and the predictive value of the same component. If W is larger than unity, then this criterion suggests that the rth component is worthwhile to be included in the model.

There is no sound statistical test for the cross-validation procedure. The main problem is not knowing how many degrees of freedom one starts with nor how many degrees of freedom are extracted with each principal component (e.g., Box, Hunter, MacGregor, and Erjavec 1973). Thus the number of principal components needed in a PCA model should be based on the overall picture that these criteria give.

2. POST-ANALYSIS OF INDUSTRIAL BATCH DATA

Data supplied by DuPont from an industrial batch polymerization reactor are used to illustrate the application of the proposed method. The cycle in the reactor consists of two stages, and the time spent processing in both stages is approximately two hours. Ingredients are loaded into the reactor to begin the first stage. Reactor-heating-medium flows are adjusted to establish proper control of pressure and the rate of temperature change. The solvent used to convey ingredients to the reactor is vaporized and removed from the reactor vessel. The vaporization process is vigorous enough that the contents of the vessel do not require stirring. After nearly one hour spent removing solvent, the second stage begins. During this stage the ingredients complete their reaction to yield the final product, a polymer. Once again, vessel pressures and the rate of temperature change are controlled during this processing stage. The batch finishes by pumping the polymer product from the vessel at the end of the second cycle.

The results of a critical property measurement are usually received 12 hours or more after each batch has finished, and therefore there is no time for recipe adjustments

in the next batch. Furthermore, it is often difficult to es-
tablish when a batch is going wrong and to diagnose what
caused the property to deviate from target. This makes the
application of an on-line SPC monitoring method attrac-
tive for this process. Failure to attain on-aim control of
the critical property leads to increased manufacturing costs
either through necessitating blending with other batches
or through downgrading the product to end uses that have
a lower selling price.

A data set of 55 successful and some unsuccessful
batches was provided from the preceding process. Each
batch had a duration of 100 time intervals (K), and 10
measurement variables (J) were monitored throughout
the batch. Variables 1, 2, and 3 are temperature measure-
ments inside the reactor, whereas variables 6 and 7 are tem-
perature measurements in the heating-cooling medium.
Variables 4, 8, and 9 are pressure measurements, and the
rest of the variables are flowrates of materials added to the
reactor during its operation. A plot of the 10 variable tra-
jectories for a typical batch is shown in Figure 3. Batches
40, 41, 42, 50, 51, 53, 54, and 55 had the final quality mea-
surement well outside the acceptable limit, and batches 38,
45, 46, 49, and 52 were above or very close to that limit.
The batches can be compared with an MPCA analysis by
plotting their t scores and their sum of squared errors:

$$Q_i = \sum_{c=1}^{KJ} \mathbf{E}(i, c)^2.$$

The t plot represents the projection of each batch history
onto the reduced plane defined by the principal compo-
nents, and the Q plot represents the squared distance of
each batch perpendicular to this plane.

First, a preliminary MPCA analysis was conducted to
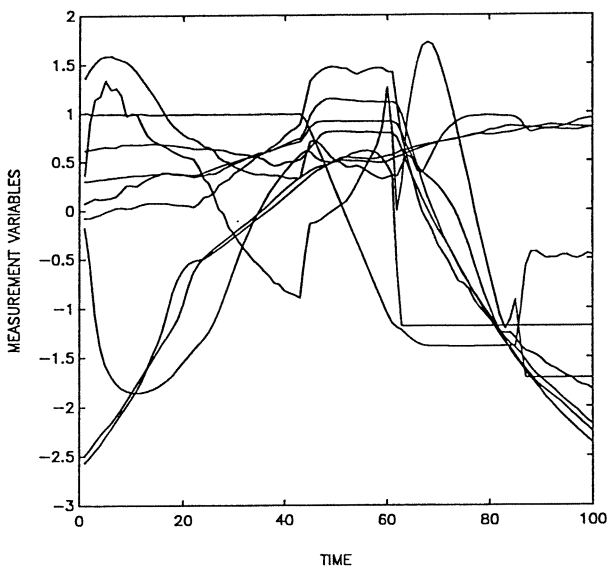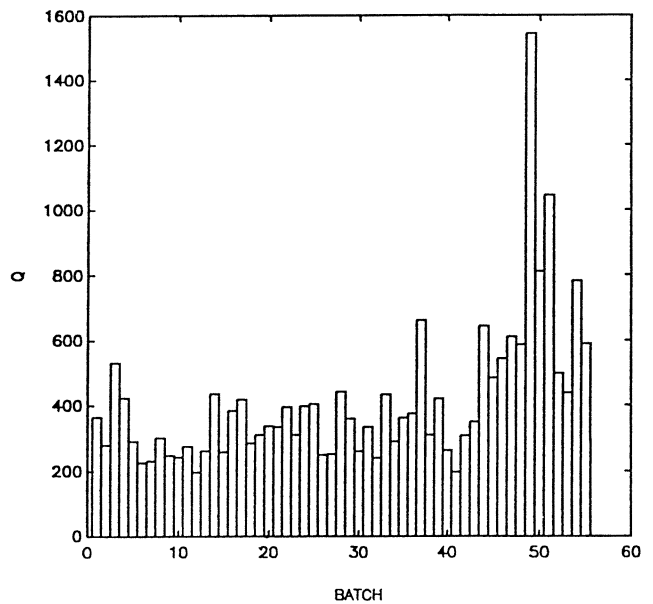identify if there was enough information in the process-

Figure 4. Results of MPCA Analysis (two principal compo-
nents) of the Original 55 Batches. Batches 49 through 55 are
identified as abnormal batches either because of their posi-
tion in the t plot (batch 50 through 55) or their residuals Q
(batch 49).

variable trajectory data to discriminate between normal
and abnormal batches. Two principal components were
extracted, and Figure 4 shows that batches 50, 51, 52, 53,
54, and 55 are clearly identified from their position in the
reduced space (t plot) as being very different from the
rest of the batches. In hindsight their different behavior
can be seen by visual inspection of the measurement tra-
jectories if one superimposes these over the trajectories
from normal batches on the same plot. Due to their struc-
tural similarity (maxima, minima, points of inflection, or

Figure 3. Trajectories From all of the Measurement Variables
From a Typical Batch Run.

discontinuity) with trajectories from normal batches, however, operators can easily see nothing different when they are displayed alone. Batch 49 also is identified as different due to its large residual (Fig. 4). The quality of this batch was barely acceptable, and its main difference compared to normal batches was during a time period (50–65 time intervals) when most of the measurement variables do not usually exhibit much variation (flat trajectories). The p loadings of the first two principal components during this time period are small, so any difference shows up in the residuals. This example with batch 49 points out
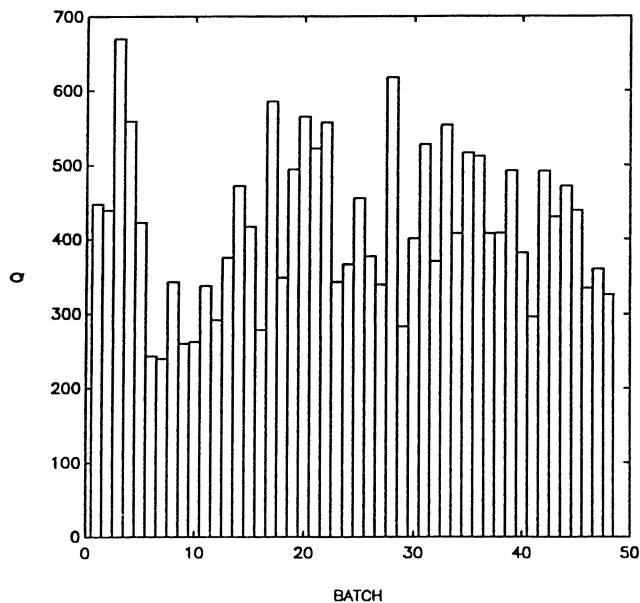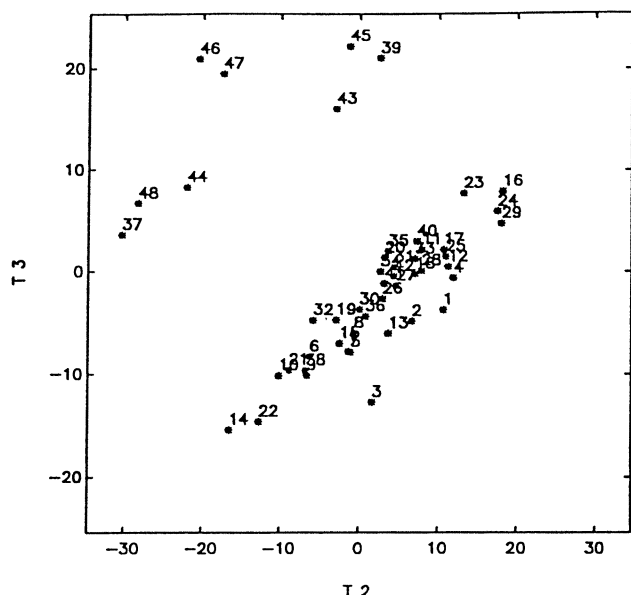

Figure 5. Results of MPCA Analysis (three principal components) of the First 48 Batches. Batches 37, 39, and 43 through 48 can be identified (t plot) as batches with unusual operational behavior.

the complementary nature of the t scores and residuals. Any variation that is not explained by the current principal components is contained in the residuals, and it will be explained in one of the components.

A second MPCA analysis was run, this time excluding these last six batches. The results from this analysis after three principal components are given in Figure 5. Batches 38, 40, 41, and 42 cannot be identified as abnormal batches. There is nothing unusual in their trajectories to be captured by MPCA because the cause of their unacceptable product does not have any effect on the measured trajectories. This shows that there are cases in which the measurements are typical of a successful batch and still the product may not meet the performance standards. The problems in batches like these may have come from poor quality materials, inadequate preprocessing, or something that can be observed only if one measures other on-line variables as well. Again another group of batches clustered away from the main body of batches in the t_2-t_3 plot. Further investigation of what went wrong with these batches must take place because most of them are in sequence (43 through 49) and two of these (45, 46) gave unacceptable product. It is important to note that the differences in the preceding batches could not be detected by a simple visual inspection in the measurement variables. The residuals in Figure 5 suggest that there are no other major differences in the operational behavior of the batches.

3. REFERENCE DISTRIBUTION OF NORMAL BATCH OPERATION

To develop multivariate SPC charts, one must have a history of past successful normal batches that can provide a reference distribution against which future batches can be compared. This reference distribution should contain all of those batches deemed to be subject only to common-cause variation. All batches exhibiting characteristics that one might wish to alarm as special causes in the future should be omitted. Based on the preceding MPCA analysis, 36 batches (I) were selected as the reference data base. All of the other batches were excluded because either they had problematic operation or they gave unacceptable product.

Table 1 shows the results of extracting successive principal components and gives the percent sum of squares explained and the three previously discussed criteria for selecting the required number of principal components. The broken-stick rule indicates that the first two principal components are significant, and the W statistic suggests three components. The R statistic implies that one could use either two or three components. Based on these results we chose to include three (R) principal components in our MPCA model. The fact that three principal components explain only about 55% of the total variability in the data is not disappointing if one considers the many variables (JK) in the unfolded matrix $\mathbf{X}$. The rest of the variabil-
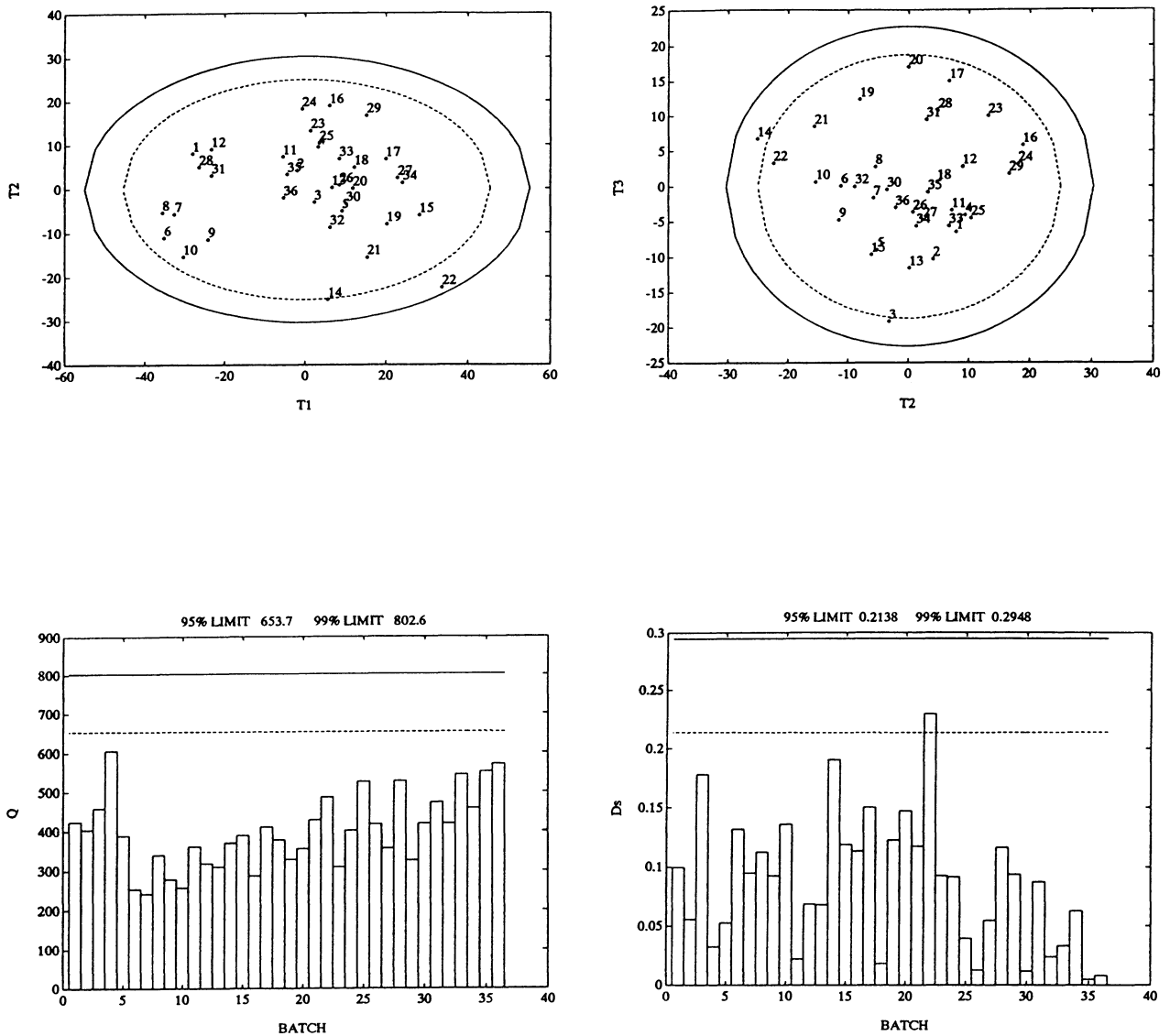
Figure 6. Results of MPCA Analysis (three principal components) of the "In-control" Reference Data Base With the 95% and 99% Confidence Limits for the Residuals Q, the D_S Statistic, and the t Plots. None of the batches exhibit unusual operational behavior.

ity is mainly due to measurement noise and to random variation in normal batch operation.

The t plots with their 95% and 99% confidence ellipsoids in Figure 6 show that none of the 36 batches exhibit any unusual behavior. The slight gap between two clusters in the $t_1 - t_2$ plot was not attributed to any significant difference but simply to not having enough batches in the reference data base to fill this gap and give a smooth variation across the first principal component. The bottom left plot in Figure 6 has the sum of the squared residuals (Q_i) for each batch with their 95% and 99% confidence limits (Appendix B), and shows that no batch exhibits any significantly large residual. The assumption behind these approximate confidence limits for Q is that the variables (JK) in the unfolded matrix X have a multinormal distribution with population mean

0 (Jackson and Mudholkar 1979; Jensen and Solomon 1972). This assumption is reasonable in our case because we have chosen a reference data base of normal operating batches. The t scores of all principal components were found to follow very well a multinormal distribution (Horswell and Looney 1992), a result arising in part because the t scores are linear combinations of random variables. Thus with the assumption that the t scores follow a multinormal distribution with population mean $\mathbf{0}$ and estimated covariance matrix $\mathbf{S}(R \times R)$, which is diagonal due to the orthogonality of the t scores, one has the following Hotelling statistic (Tracy, Young, and Mason 1992):

$$D_s = \mathbf{t}'_R \mathbf{S}^{-1} \mathbf{t}_R I /(I-1)^2 \quad \sim \quad B_{R/2,(I-R-1)/2},$$

where $\mathbf{t}_R$ is the vector containing the scores of a given

Table 1. Percentage of Explained Sum of Squares (cumulative and for each principal component) From the MPCA Analysis of the Reference Data Base and the Results From the Three Criteria to Determine the Optimal Number of Principal Components

PC	% SSX	% PC SSX	G	R	W
1	38.55	38.55	11.59	.65	17.58
2	50.22	11.68	8.82	.91	5.35
3	56.75	6.53	7.43	1.04	2.24
4	61.33	4.57	6.50	1.17	.76

batch from the R retained principal components and the critical values of the beta variable at significance level α can be found from critical values of the F distribution by using the relationship

$$B_{R/2,(I-R-1)/2,\alpha} = (R/(I - R - 1))F_{R,I-R-1,\alpha}$$
$$\div (1 + (R/(I - R - 1))F_{R,I-R-1,\alpha}).$$

Accordingly, the 95% and 99% confidence ellipsoids in the t plots (Fig. 6) are centered on $\mathbf{0}$, and their axis lengths in the direction of the rth principal component are given by (Johnson and Wichern 1988)

$$\pm(\mathbf{S}(r, r)B_{1,(I-2-1)/2,\alpha}(I - 1)^2/I)^{1/2}.$$

The D_s statistic gives a measure of the Mahalanobis distance in the reduced space between the position of a batch (t scores) and the origin that designates the point with the minimum variation in the batch process behavior. Again, no batch in the bottom right plot of Figure 6 shows any unusual variability.

The Q and D_s plots shown in Figure 6 appear to justify the normality assumptions that we have made and corroborate our choice of this data base to describe the normal batch operation. These plots provide the diagnostics to test if we have included any unusual batch in our data base and if we have built a representative model of the normal batch operation.

Figure 7 shows the percent of the sum of squares explained by the model with respect to time and variables. The first principal component concentrates more on the first stage of the process, at which the vaporization takes place, and the variables associated with this period (1, 2, 3, and 10). The second principal component captures most of the variability in the second stage of the process, at which most of the polymerization takes place, and involves mainly variables 6, 8, and 9. When the process has two stages, as in this industrial example, it is common to see one principal component to explain the first stage and another to explain the second stage. MPCA does this because the correlation of the measurement variables in each stage changes, and thus a single principal component may not be able to explain both of them. None of the three principal components explains much of the variation during the transition period (50 through 65) because most of
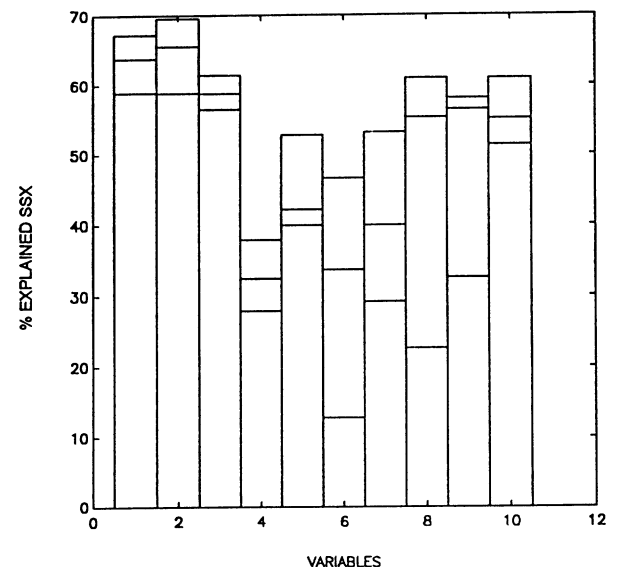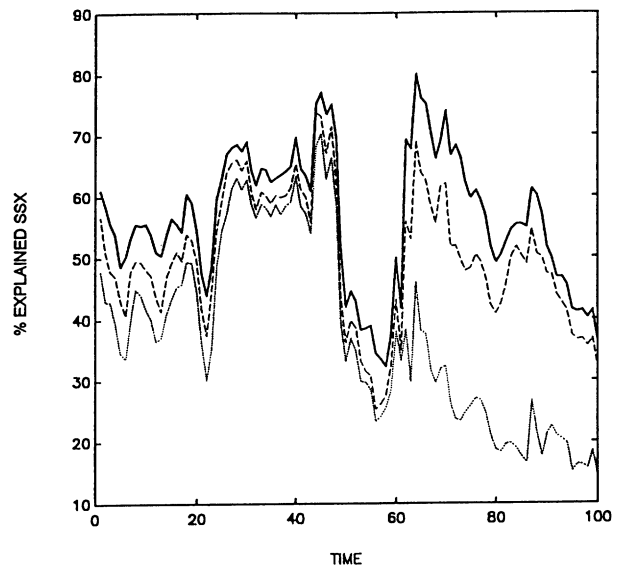
Figure 7. Cumulative Percent of the Sum of Squares Explained, With Respect to Time and Variables, for the Three Principal Components. The lowest line represents the percent explained by the first principal component, the line above it gives the percent explained by the first two principal components, and the top line shows the percent explained by all three principal components.

the variables during this time period have flat trajectories with few deviations from them.

4. MULTIVARIATE SPC CHARTS FOR ON-LINE MONITORING

The p-loading matrices from the MPCA analysis of the reference data base contain most of the structural information about how the variable measurements deviate from their average trajectories under normal operation. If a new batch is to be tested for any unusual process behavior, one

can use the these p-loading matrices to check this hypothesis by obtaining the predicted t scores and residuals for the new batch $\mathbf{X}_{new}(K \times J)$.

Scale and unfold $\mathbf{X}_{new}$ to $\mathbf{x}'_{new}(1 \times JK)$:

$$t_r = \mathbf{x}'_{new}\mathbf{p}_r, \quad \mathbf{e} = \mathbf{x}_{new} - \sum_{r=1}^{R} t_r \mathbf{p}_r.$$

If the t scores of the new batch are close to the origin and its residuals are small, then this indicates that its operation is similar to that in the reference data base of normal batches.

A problem arises when one wants to perform the test sequentially in time as the new batch evolves. In this situation the matrix $\mathbf{X}_{new}$ is not complete until the end of the batch operation. At each time interval during the batch operation, the matrix $\mathbf{X}_{new}$ has all the measurements only up to that time interval. The rest of the $\mathbf{X}_{new}$ matrix from the current time to the end of the batch is still undefined. The most valid way to overcome this problem is to build K different MPCA models, one up to each time interval k using only the information available up to that time. This results in the need to store K loading vectors ($\mathbf{p}$) for each principal component of dimension ($Jk \times 1, k = 1, 2, \ldots, K$), and to apply the one appropriate for the current time k to calculate the scores and residual for that time. Although this is the most correct approach, the computational and storage requirements would be very large except for short duration (small K) batch processes having a relatively few on-line measurement variables (J). An alternative monitoring scheme, without using MPCA, is to make the assumption that the measuring variables are multinormally distributed and at each time interval k to perform an F test based on a Hotelling statistic. This will test if the Jk variables measured up to the current time (k) are too far away from the origin of the multivariate distribution estimated from the reference-good data base. This scheme is unattractive because one has to store at each time interval a large covariance matrix ($Jk \times Jk$) that is ill-conditioned because of the highly correlated variables.

Therefore, we propose several approximate MPCA methods for constructing sequential tests. All of these involve using the full ($JK \times 1$) loading vectors $\mathbf{p}_r$ obtained from the MPCA on the entire histories of the batches in the reference data base and then filling in the future observations in $\mathbf{X}_{new}$ in different ways. All of these approaches will give the same predicted t scores and residuals at the end of the batch when the full $\mathbf{X}_{new}$ is known. To monitor the progress of a new batch as new observations become available, one of the methods discussed in Section 4 is used to fill out the $\mathbf{X}_{new}$ matrix, and then the t scores and residuals are calculated for each time interval. The t scores assert the overall performance of a batch, and the best way to track the particular instant that something behaves differently is to use the squared prediction

error

$$\text{SPE}_k = \sum_{c=(k-1)J+1}^{kJ} \mathbf{e}(c)^2$$

associated with the latest on-line measurements at time interval k from the process. The sum of the squared residuals over all time periods, $Q = \sum_{c=1}^{KJ} \mathbf{e}(c)^2$, is not a good indicator because it does not represent the instantaneous perpendicular distance of a batch from the reduced space as does the SPE and it is affected by the errors associated with the filling in of future unknown observations in the matrix $\mathbf{X}_{new}$.

4.1 Anticipating the Future Observations in $\mathbf{X}_{new}$

Three methods are considered in this article for filling in the unknown data in $\mathbf{X}_{new}$ between the current time interval k and the end of the batch. Recall that the $\mathbf{x}_{new}$ (unfolded $\mathbf{X}_{new}$) after scaling contains the deviations of the measurements from their mean trajectories. Monitoring charts for SPE and the first latent variable t_1 (similar charts are obtained for t_2 and t_3) are shown in Figure 8 for each of these methods. The approximate 95% and 99% control limits and the outermost values (five for the SPE and six for the t scores) at each time interval from the reference-normal data base are also shown. The control limits for the t scores and the SPE are developed in Section 5.

1. The first approach to filling in the unknown observations in $\mathbf{x}_{new}$ is to assume that the future observations are in perfect accordance with their mean trajectories as calculated from the reference data base. Thus the assumption behind this approach is that the batch will operate normally for the rest of its duration with no deviations in its mean trajectories, and one has to fill the unknown part of $\mathbf{x}_{new}$ with zeros. The advantage of this approach is a nice graphical representation of the batch operation in the t plots and the quick detection of an abnormality in the SPE plot (MacGregor and Nomikos 1992). In the upper part of Figure 8 one can see the cone shape of the control limits for the t scores due to the assumption of future normal operation. A new batch always starts from the origin of the t scores in the reduced space and progressively moves out. The drawback of this approach is that the t scores are reluctant, especially at the start of the batch run, to detect an abnormal operation.

2. A second approach is to assume that the future deviations from the mean trajectories will remain for the rest of the batch duration at their current values at time interval k (Nomikos and MacGregor 1994). In this case, the assumption is that the same errors will persist for the rest of the batch run. Under this assumption, the SPE chart is not as sensitive as in the first approach, but the t scores pick up an abnormality more quickly. A compromise between the first two approaches that shares their advantages and disadvantages is to assume that the future deviations will decay linearly or exponentially from their current values to 0 at the end of the batch run.

PAUL NOMIKOS AND JOHN F. MACGREGOR

OUT OF 95% LIMIT = 187 OUT OF 99% LIMIT = 50 OVER 3600 OBSERVATIONS

OUT OF 95% LIMIT = 8 OUT OF 99% LIMIT = 0 OVER 3600 OBSERVATIONS

OUT OF 95% LIMIT = 177 OUT OF 99% LIMIT = 52 OVER 3600 OBSERVATIONS

OUT OF 95% LIMIT = 24 OUT OF 99% LIMIT = 4 OVER 3600 OBSERVATIONS

OUT OF 95% LIMIT = 177 OUT OF 99% LIMIT = 45 OVER 3600 OBSERVATIONS

OUT OF 95% LIMIT = 24 OUT OF 99% LIMIT = 0 OVER 3600 OBSERVATIONS
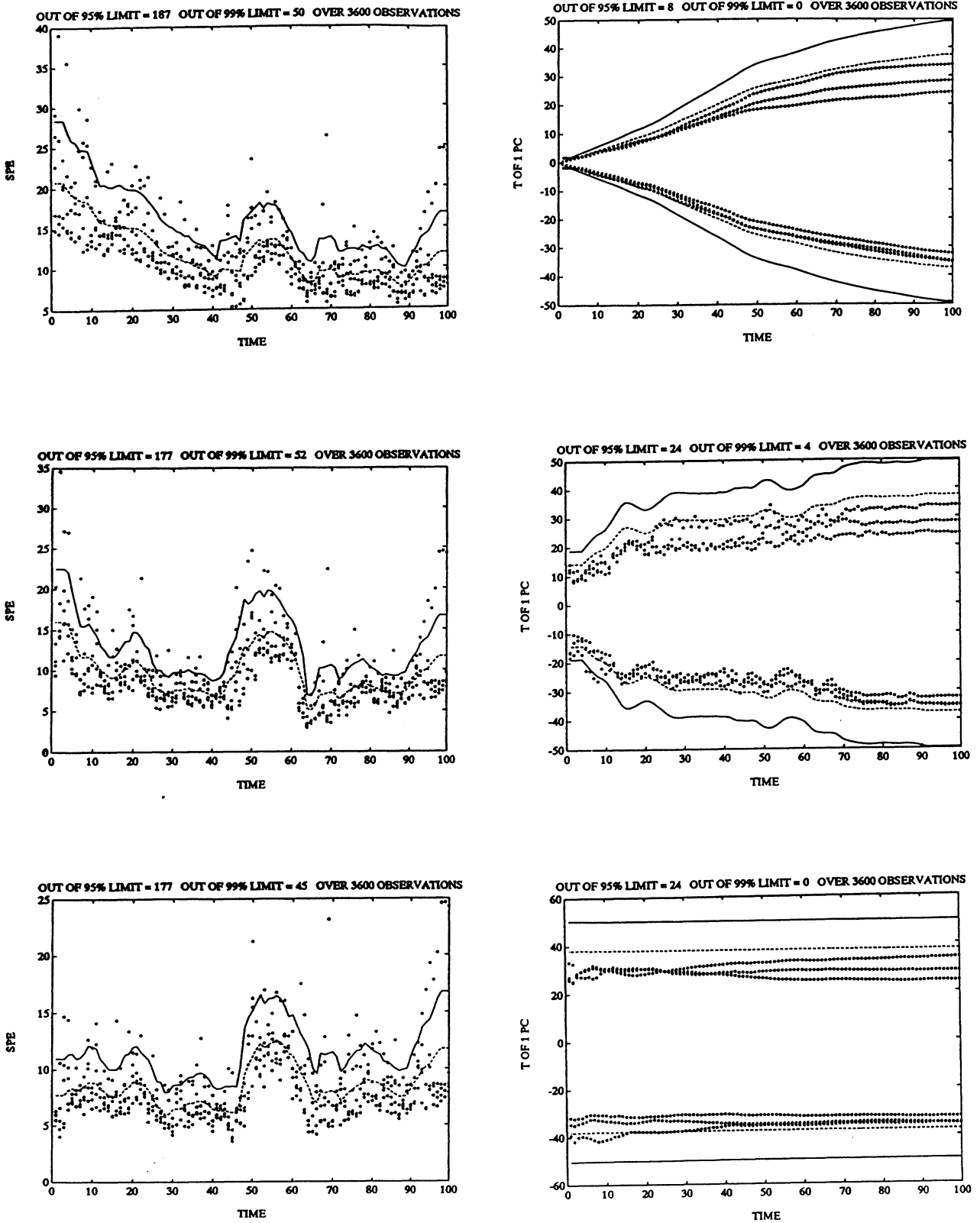
Figure 8.   Control Limits (95% and 99%) for the SPE and t Scores, With the Outermost Values at Each Time Interval for the Three Approaches of Handling Future observations in X_{new}. The upper plots are for the approach with zeros, the middle plots for the approach with current deviations, and the last plots for the approach by projection.

TECHNOMETRICS, FEBRUARY 1995, VOL. 37, NO. 1

3. The last approach uses the ability of PCA to handle missing data. The unknown future observations can be regarded as missing values from an object (batch) in MPCA. Hence one can use the principal components of the reference data base to predict these missing values by restricting them to be consistent with the already observed values up to time interval k and with the correlation structure of the measurement variables in the data base as defined by the p-loading matrices of the MPCA model. MPCA can do this by projecting the already known observations $(\mathbf{x}_{\text{new},k}(kJ \times 1))$ into the reduced space and calculating the t scores at each time interval as (Appendix C)

$$\mathbf{t}_{R,k} = (\mathbf{P}_k'\mathbf{P}_k)^{-1}\mathbf{P}_k'\mathbf{x}_{\text{new},k},$$

where $\mathbf{P}_k(kJ \times R)$ is a matrix having as columns all the elements of the p-loading vectors $(\mathbf{p}_r)$ up to time interval k from all the principal components. The matrix $(\mathbf{P}_k'\mathbf{P}_k)$ is well conditioned because of the orthogonality property of the loading vectors $(\mathbf{p}_r)$. This method appears to be superior to the other methods if at least 10% of the history of a new batch is known. It has the great advantage of giving t scores very close to their actual final values, and thus their control limits have quite constant trajectories (Fig. 8). Caution must be used at the beginning of a new batch in which this method may give quite large and unexplainable t scores because there is so little information to work with.

Which approach to use depends on the specific characteristics of the process under consideration. If the trajectories of the process measurements do not exhibit frequent discontinuities or early deviations, one may use the third approach. If there is knowledge that the disturbances in a given process are quite persistent, then it is better to use the second approach which generally has worked well in most cases we have investigated. If a batch process does not exhibit persistent disturbances or has variables with discontinuities in their trajectories, then it may be better to use the first approach. In general, one can use a combination of the preceding approaches, like starting with one approach and switching after some time to another one, and to build in this way some engineering knowledge in his monitoring scheme.

5. CONTROL LIMITS FOR THE MULTIVARIATE SPC CHARTS

Independent of which way one may choose to handle the future observations in a new batch run, the on-line monitoring will be based on control charts in which the control limits will be established in the same way. These charts will monitor the t scores and the SPE of a new batch as it progresses. If a new batch is still operating in the same way as the batches in the reference data base, but still has a larger than normal variation in its measurements, this will show up clearly as large deviations of the

t scores from the origin of the reduced space. In the case in which a totally new fault that is not represented in the data base occurs, the principal components will then not be able to describe correctly the variation. Thus the new observations will move off the MPCA plane, resulting in large values of the SPE. The residuals account for any disturbance that is not described sufficiently in the data base of good batches, and this makes them very sensitive in detecting such new faults.

The control limits for these charts are calculated by passing each of the batches in the reference data base through the monitoring procedure and collecting their t scores and SPE at each time interval k. These 36 observations for the t scores and SPE at each time interval provide the external reference distribution (Box, Hunter, and Hunter 1978) upon which the control limits can be directly calculated. The assumption is that this external reference distribution is sufficient to capture the common-cause variation in normal batch operations and that this variation will still be functioning in the same manner in future batch runs.

5.1 Control Limits on t Scores and SPE

The t scores are linear combinations of the measurement variables and by the central limit theorem should be approximately Normally distributed. Analysis of the batch data revealed that they were indeed well approximated by a Normal distribution (D'Agostino and Stephens 1986; Lilliefors 1967) except at the first few time intervals. These early deviations from Normality result from the approximations used to handle the future observations in $\mathbf{X}_{\text{new}}$. Under the assumption of Normality the control limits at significance level α, for a new independent t score, at any given time interval are given by (Chew 1968; Hahn and Meeker 1991)

$$\pm t_{n-1,\alpha/2}s_{\text{ref}}(1 + 1/n)^{1/2},$$

where n, s_{ref} are the number of observations and the estimated standard deviation of the t-score sample at a given time interval k (the mean is always 0) and $t_{n-1,\alpha/2}$ is the critical value of the Studentized variable with $n - 1$ df at significance level $\alpha/2$. The Hotelling statistic (Tracy et al. 1992) for a new independent t vector becomes

$$D = \mathbf{t}_{R,k}'\mathbf{S}^{-1}\mathbf{t}_{R,k}I(I - R)/R(I^2 - 1) \sim F_{R,I-R}$$

and the axis lengths of the confidence ellipsoids in the direction of the rth principal component are given by (Johnson and Wichern 1988)

$$\pm(\mathbf{S}(r,r)F_{2,I-2,\alpha}2(I^2 - 1)/I(I - 2))^{1/2}.$$

The SPE is a quadratic form of the errors associated with the least observations at time interval k ($\text{SPE}_k = \sum_{c=(k-1)J+1}^{kJ}\mathbf{e}(c)^2$). These errors were found to be well approximated by a multinormal distribution $N(\mathbf{0}, \mathbf{\Sigma})$.

Box (1954) and Jackson and Mudholkar (1979) derived approximate distributions for such quadratic forms. Box showed that it is well approximated by a weighted chi-squared distribution $(g\chi_h^2)$, where the weight (g) and the degrees of freedom (h) are both functions of the eigenvalues of Σ. Jackson and Mudholkar's approximate distribution, as shown in Appendix C, is very close to that given by Box. In this article we use the $g\chi_h^2$ approximation of Box for the distribution of the SPE to estimate the control limits at any point in time. Although the g and h can be estimated from the eigenvalues of the estimated Σ, a simpler approach is used here based on matching moments between a $g\chi_h^2$ distribution and the reference distribution of SPE at any time interval k. The mean and variance of the $g\chi_h^2$ distribution $(\mu = gh, \sigma^2 = 2g^2h)$ are equated to the sample mean (m) and variance (v) of the SPE sample at each time k (Appendix B). It is a quick way to estimate g and h reasonably well provided that the number of SPE observations is sufficiently large. Thus the control limit on the SPE at significance level α for time interval k are given by

$$\text{SPE}_\alpha = (v/2m)\chi^2_{2m^2/v,\alpha},$$

where $\chi^2_{2m^2/v,\alpha}$ is the critical value of the chi-squared variable with $2m^2/v$ df at significance level α.

Estimates of g and h are shown in Figure 9 for the monitoring scheme used in the last plots in Figure 8 (i.e., filling in the missing data by projection). These results are similar to those obtained using the other two approaches to handling the future observations in X_{new}, with differences occurring mainly in the first 10 to 15 time intervals. These plots of g and h provide information about the changing nature of the distribution of the residuals throughout the duration of the batch. Low values of the degrees of freedom (h) indicate that the distribution is dominated by large variability of only a few of the measurement variables around their mean trajectories. High values of h occur during more stable periods in which deviations from most of the variables are contributing evenly to the SPE. For example, the batch period (10–40) represents a fairly smooth behavior during the well-controlled vaporization stage. Period (45–50) represents the transition from the vaporization to polymerization stage and (65–70) a change in nature of operation during polymerization. In both of these periods a few variables are changed rapidly. Moreover, sudden changes are made to the process at the start and at the end of the batch run. The g is simply a scaling factor to enable one to match the moments.

Because the number of observations in the reference distribution at each time interval may not be very large $(I = 36$ in this example), the control limits on the t scores and SPE can be quite variable. Because most batch processes progress in a reasonably smooth manner and each time interval is closely related to its previous and next ones, one might expect the control limits to change smoothly. Therefore, we have used the idea of window-
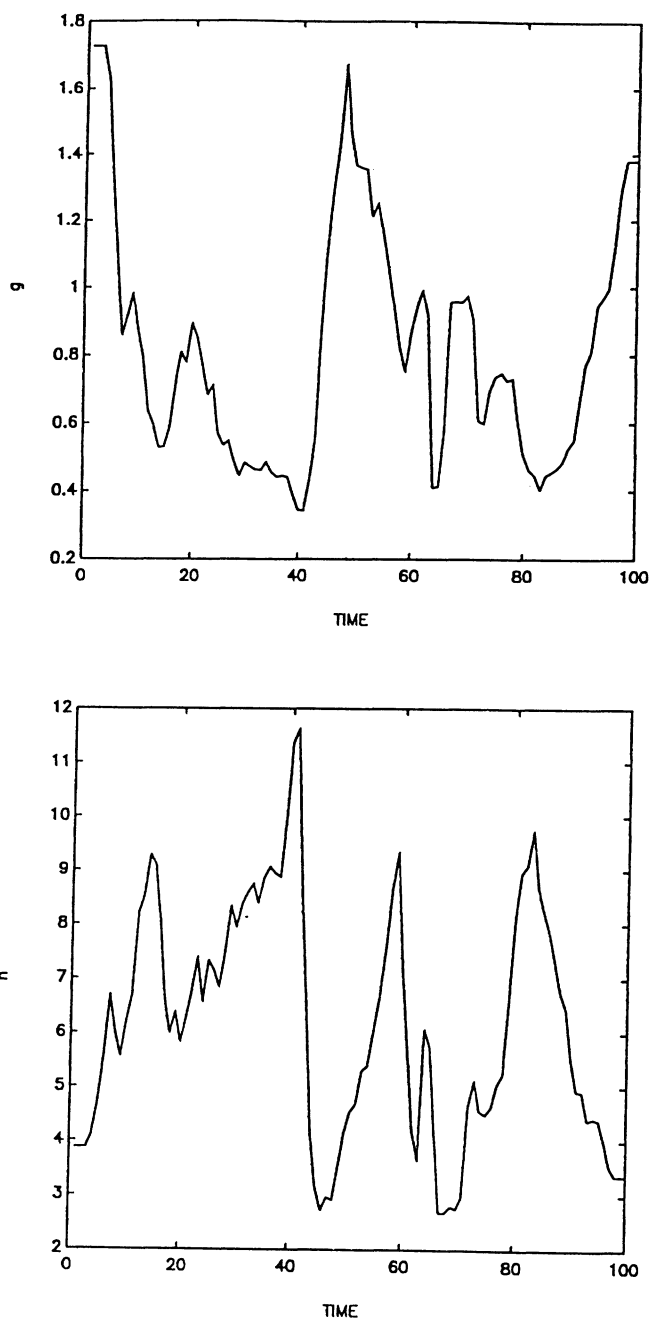

Figure 9.   Plots of Estimated g and h Values From the $g\chi_h^2$ Distribution.

ing taken from spectral analysis (Jenkins and Watts 1968) for the results in Figures 8 and 9. The reference distribution at time interval k was composed of the observations at time intervals $k - 2, k - 1, k, k + 1$, and $k + 2$. In effect, we have used a moving window that is five time intervals wide to combine data for the estimation of the control limits on the t scores and SPE at the center of the window. In our example this provides $n = 5*36$ observations for the calculation of the limits at each time interval. In general the width of the smoothing window will depend on the number of batches in the reference distribution and on the nature of the process itself. In the

Table 2. Overall Significance Levels for the Control Limits of the t Scores and SPE for the Three Approaches to Handling the Future Observations in X_{new}

Approach	Zeros		Current deviations		Projection	
t_1	.002	.000	.006	.001	.006	.000
t_2	.052	.011	.041	.003	.086	.009
t_3	.030	.000	.051	.002	.056	.000
SPE	.052	.014	.049	.015	.049	.012
Instantaneous α	.050	.010	.050	.010	.050	.010

case in which there are many batches in the reference data set, very little smoothing will be necessary. The use of a window helps small sample sizes but relies on the additional assumption that the variance of the statistics varies in a smooth manner with respect to time. In practice (as in spectral estimation), several window widths can be tried and one chosen that gives reasonably smooth control limits (low variance) but does not affect their main shapes (low bias).

5.2 Overall Type I Error

The preceding control limits for the t scores and SPE were based on the approximate distributions of these statistics at any one point in time. Considering only that period of time, the α value in these tests would be the Type I error. Although this is a common procedure for setting control limits (Bauer and Hackl 1980; John 1990; MacGregor and Harris 1993; Montgomery 1994), it is not correct when one considers the sequential application of the procedure over the entire batch run. In general, the Type I error associated with monitoring the entire 100 time intervals will be different from the α value for the instantaneous test. If the statistics were independent over time then the overall Type I error would be given by $1 - (1 - \alpha)^{100}$, which is .634 for $\alpha = .01$. The t scores and SPE values at successive times are not independent, however, and the overall Type I error could only be determined knowing the joint distribution of these statistics over all periods. Fortunately, to establish the control limits each set of batch data in the reference data base is passed through the monitoring procedure, and the number of values of each test statistic falling outside the control limits can be enumerated. Thus an overall Type I error can be estimated for the control chart as the number of values of the test statistic outside the control limits in the reference data base divided by the total number of observations (IK). These overall Type I errors are presented in Table 2 for the two values of the instantaneous test Type I errors. The estimated overall Type I errors for the SPE test are quite close to the instantaneous α values. We have found this to be generally true in every data set we have investigated so far. The overall Type I errors for the t scores are generally close to the nominal α value when the instantaneous $\alpha = .05$, but the approximation is poorer for instantaneous $\alpha = .01$.

6. MONITORING AN INDUSTRIAL BATCH PROCESS

Two examples are given in Figures 10 and 11 for online monitoring of two new batches from the industrial polymerization process using the projection approach for filling in X_{new}. Neither batch was included in the reference data base of normal batches used to develop the MPCA model. The batch in Figure 10 is a new batch that was felt to have exhibited both acceptable operation and acceptable final polymer quality. The batch in Figure 11 is batch 49 discussed earlier in our post-analysis study. This batch yielded a product of marginal quality, in that the quality measurement was right at the acceptable limit.

Figures 10 and 11 show plots for the SPE and one of the t scores along with their control limits. Plots of the joint t_1-t_2 space and the D statistic are also shown. The control limits on these later two charts are only approximate ones because they are based on the covariance matrix S of the t scores from the post-analysis of the reference data base, which has available the measurement variables for the whole batch duration. These two charts evaluate at each time the expected performance of the whole batch duration assuming that the future behavior of the batch is well described by the approach that is used to fill in the unknown observations in X_{new}. To provide more precise control limits for the joint t space and the D-statistic charts would require evaluating and storing the estimated joint covariance matrix $(R \times R)$ of the t scores for each point in time. Note that the t scores are truly orthogonal only at the final time corresponding at the end of the batch.

The batch with acceptable product (Fig. 10) shows no abnormality in any of the SPC charts. The SPE chart for batch 49 clearly signals that something is unusual between time intervals 57 and 65. During this period the p loadings are small, which makes the t scores slow to respond to the change. After the 65th time interval the measurements return to their normal trajectories, as do the t scores, because now the unusual previous behavior plays a less significant role in them. Unlike the SPC charts on the other abnormal batches investigated, in which once a problem was detected the t and SPE values remained outside of the control limits for the remainder of the batch, in this batch (49) the problem disappears shortly after time 65. In spite of its return inside the acceptable control region, this batch is characterized as abnormal because of the violation of
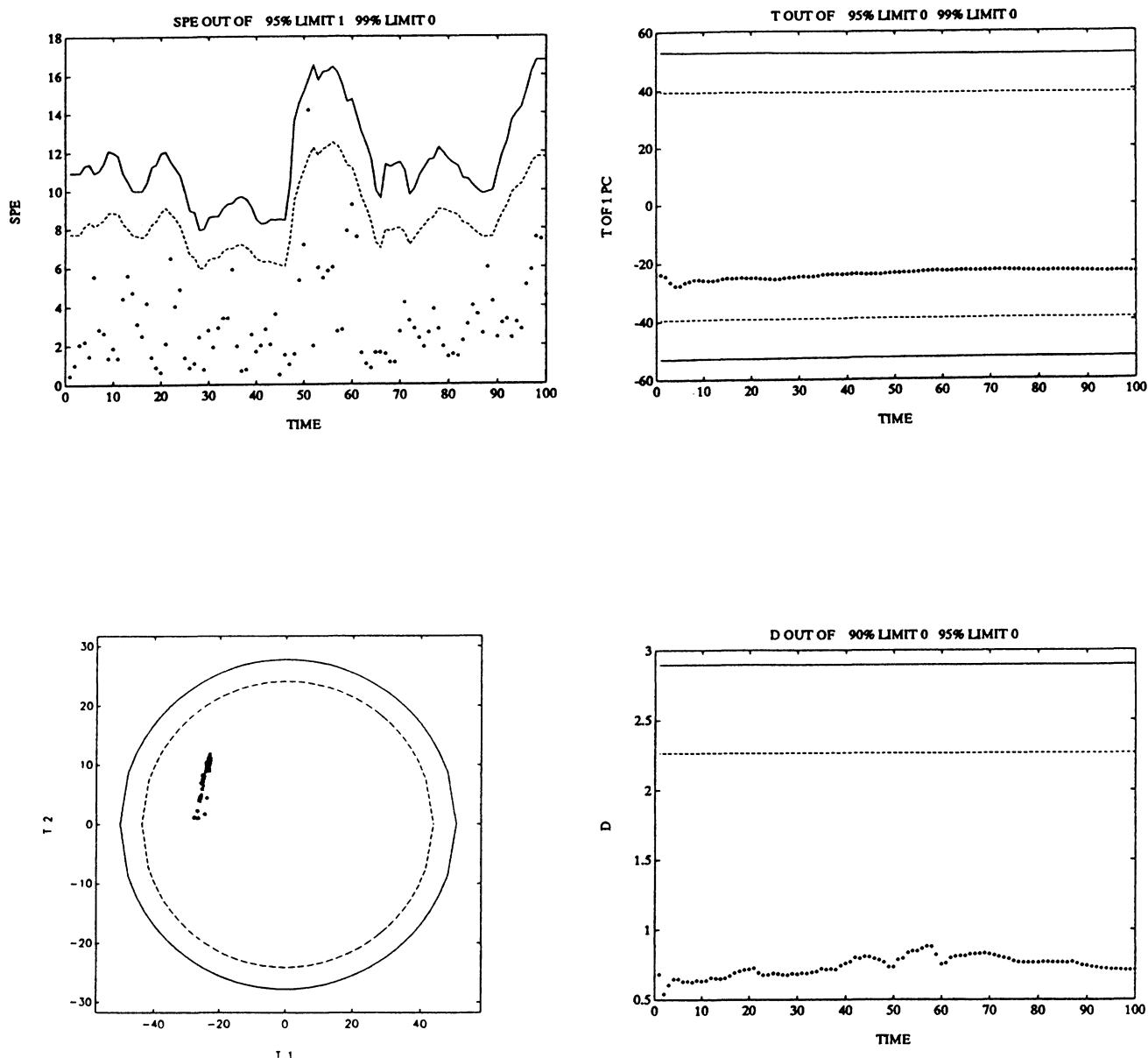
Figure 10. Monitoring Charts for a New Batch With Good Operation. SPE and t_1 plots (95% and 99% control limits), joint t_1-t_2 and D statistic plots (90% and 95% control limits).

the control charts during the period 57–65. Indeed, this deviation did result in a product with a borderline quality.

Once a fault or special event has been detected, it is important to diagnose the event to find an assignable cause. For this aspect of SPC, the multivariate methods are much more useful than univariate methods. By interrogating the underlying MPCA model, the contribution of each measurement variable to the deviations observed in the SPE and in the t scores can be displayed (Miller et al. 1993; MacGregor et al. 1994). These "contribution" or "diagnostic" charts can be immediately displayed on-line by the operator as soon as the special event is detected. Although they may not provide an unequivocal diagnosis, they at least will clearly show the group of variables that are primarily responsible for the detected deviations. In

the case of the industrial batch polymerization run 49, they clearly pointed to simultaneous deviations in four variables (6, 7, 8, and 9) from their average trajectories. On closer examination of these trajectories it was seen that at time period 57 these four variables did deviate in a systematic manner and returned to their mean trajectories around time 67. Therefore, the special event in this batch could be attributed to an operational problem with these variables.

From the preceding examples it is clear that one has to monitor closely the SPE and D charts and, if something goes wrong, to check the individual t-score plots to get a better understanding of what is going on. In some cases it may be possible to identify areas in the reduced space (t plots) corresponding to a particular fault and thus to
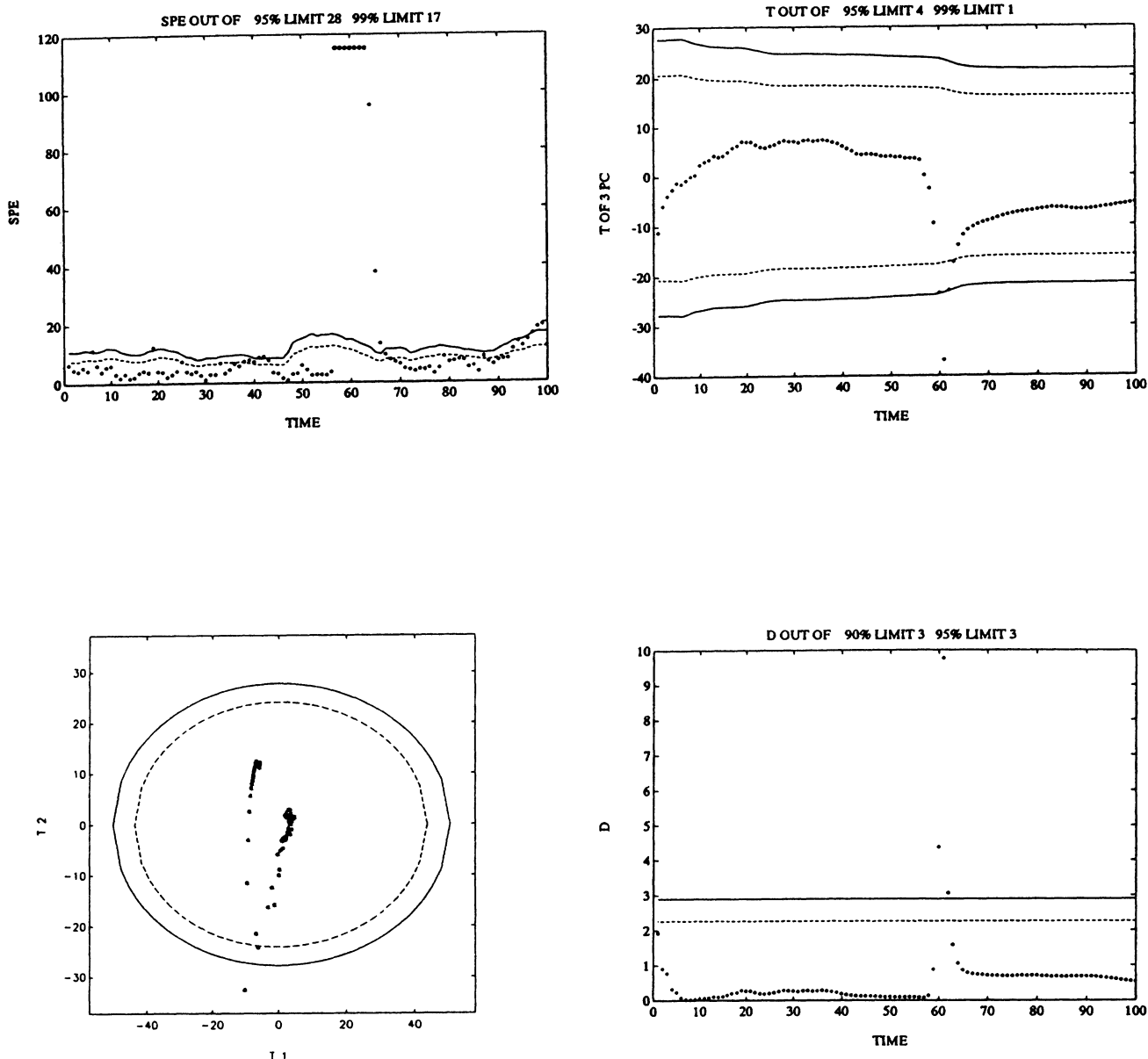
Figure 11. Monitoring Charts for "Bad" Batch Number 49. SPE and t_3 plots (95% and 99% control limits), joint t_1-t_2 and D statistic plots (90% and 95% control limits).

construct an expert system for diagnosis. It should also be noted that a violation of the control charts does not mean that the product will be unacceptable. It only means that the operational behavior of the batch is unusual and this unusual behavior may lead to low-quality product.

7. DISCUSSION AND ENGINEERING ISSUES

MPCA provides an answer to the question if the on-line measurements have enough information to detect unsuccessful batches. Post-analysis by MPCA has proven useful for discriminating between batches with normal and abnormal operation and to identify the time periods and in which measurement variables the differences occur. Large p loadings are good indicators of where to look to explain

differences in the t scores. MPCA has also been shown to provide a useful means of augmenting knowledge gained from the final quality measurements for assessing whether or not a past batch is a good one. This is a rather attractive means of characterization, if one considers how difficult it is to obtain quality measurements and the uncertainty that is involved in them. Certainly, one quality measurement, as in the industrial example of this article, cannot capture all the quality aspects of the final product.

In common with all on-line monitoring methods, these multivariate SPC methods can only detect "observable" events—that is, events that influence at least one or more of the measured variables. This is analogous to the requirement of "observability" in state estimation of mechanistic models (Kuo 1987). Some events that lead to quality prob-

lems may pass undetected if they have no impact on the measurement variables. The only way of improving this situation is to add new measurements that are responsive to these events. Indeed, in another industrial batch process that we investigated, an important quality problem related to surface properties of the product was not able to be detected by these MPCA charts. This had not been unexpected, however, because none of the on-line measurements available were related to surface chemistry.

The batch runs must be comparable for MPCA to be effective. By comparable runs, we mean batches that operate in reactors of similar design, with the same catalysts, the same operational program, and so forth. If something changes in the process, a new MPCA model must be built to accommodate this change. This is a general requirement of any method based on an empirical reference model. In our industrial example, the reactor unit is removed from service every several hundred batches for routine maintenance and cleaning. This cleaning changes the heat-transfer characteristics of the reactor for the first few batches after the reactor is placed in service again, and special precautions and control in the operation of these batches is required. Our batch data base is from a seasoned unit that makes the resulting MPCA model unsuitable for monitoring these first batches after the cleaning.

The only requirement of the proposed method is the availability of a good data base on past batches and the ability to access the same data in real time for new batches. Data-collection systems and data historians on many industrial batch processes are inadequate. Historical data on all the variable trajectories are often not saved but rather summarized by only a few raw statistics. The sensors for measuring the on-line variables must also be well maintained on a regular basis.

The sampling rate must be adequate for capturing the important trajectory information in the process. If there exists prior knowledge that a particular period during the batch is very important for product quality, then the sampling rate should be increased over that period. By increasing the sample rate it will be possible to track faster any deviations from the average trajectories over that period, and it will also weight that period more heavily in the MPCA model because there will be more p loadings corresponding to it. Another way to weight a particular period or variable more heavily is with proper scaling when the MPCA model is being built. The method can also handle different numbers of measurement variables during different stages of the batch operation. One can either put zeros in the time intervals that these variables are not measured or augment columnwise the unfolded matrix $\mathbf{X}$ at the appropriate time intervals at which these variables are measured.

A difficulty with the proposed method, as outlined in this article, occurs when one encounters batch processes in which the duration of each batch or the timing for key events during each batch is different. An example of this

occurs when various decisions during the batch are not automated but left to the discretion of an operator. The batches are usually synchronized at time 0 using a trigger variable whose change indicates the start of the batch. If the batch trajectory shapes are similar from batch to batch and only the elapsed time required to achieve a given endpoint changes, then for post-analysis one can often renormalize the time scale so that all batches have the same duration. This is not feasible for on-line monitoring, however, because the duration is not known a priori except in cases in which one can easily develop rules to anticipate delays between different batch stages. One way to handle varying batch times in on-line monitoring is to replace time by another measured variable that progresses monotonically in time and has the same starting and ending value for each batch. Examples would be an on-line measure of conversion in a chemical reactor or a measure of lance position in injection molding. Numerous other possibilities exist, but each is specific to the nature of the batch process. The data used in this article were from a well-automated process, and the only thing we had to do was to discard a few observations in the original raw data base prior to the start and after the end of the batch operation.

The MPCA used in this article only makes use of the process-variable trajectory measurements ($\underline{\mathbf{X}}$) taken throughout the duration of the batch. Measurements on product quality variables ($\mathbf{Y}$) taken at the end of each batch were used only to help classify a batch as "good" or "bad". Nevertheless, such product-quality data can be used in a much more direct fashion. Multiway PLS (Wold et al. 1987) can be performed using both the process-variable data array $\underline{\mathbf{X}}$ and the product-quality data matrix $\mathbf{Y}$. Rather than focusing only on the variance in $\underline{\mathbf{X}}$, MPLS focuses more on the variance of $\underline{\mathbf{X}}$ that is most predictive for the product quality $\mathbf{Y}$. In addition, one usually has also matrix $\mathbf{Z}$ containing measurements on the initial setup conditions, such as quality measurements on the reactants, and measurements on the initial amount of each ingredient charged to the batch. Changes in these variables may often be as important to the product quality as variations in the process trajectories. MPLS or multiblock versions of MPLS (MacGregor et al. 1994) can easily incorporate all of these data ($\mathbf{Z}$, $\underline{\mathbf{X}}$, $\mathbf{Y}$) into the proposed multivariate SPC monitoring schemes. The charts and methods for setting them up are identical to those presented here using MPCA.

Many articles have been written recently on the topic of combining SPC and automatic process control (APC) (Box and Kramer 1992; Tucker, Faltin, and Vander Weil 1993; Vander Wiel et al. 1992). It is important to note that the multivariate SPC charts proposed in this article are meant to be applied to data collected while the underlying batch process is being controlled by the feedforward and feedback APC schemes. If something goes wrong during a batch run, the controllers may be able to compensate for it. Although its final product may be acceptable, MPCA

will detect this batch as abnormal because of the unusual behavior caused by the excessive control actions.

When discussing these multivariate SPC monitoring schemes, a very appealing feedback control idea has sometimes been proposed—namely, that whenever a fault is detected the MPCA model be inverted to solve for values of the adjustable variables that could be reset to bring the batch back into the in-control region of the t space. This suggestion is not reasonable, however, because the MPCA model is not a cause-and-effect model but rather only a model for the correlation structure of the process variables under routine operating conditions. It cannot be used to predict the effects that independent changes in some of the measurement variables will have on the quality of the final product. Furthermore, once the t scores or the SPE have indicated the occurrence of a special event, simply forcing them back into the control region will not imply acceptable end product, as illustrated in the example with batch 49 (Fig. 11). Of course, this does not mean that nothing should be done when a fault is detected. The nature of any corrective action will depend on the underlying cause and on the time during the batch at which the fault occurred. Therefore, the best form of control action is probably to use on-line diagnostic tools to interrogate the underlying MPCA model for possible reasons for the fault and respond accordingly using one's process knowledge or an expert system. Even if the current batch cannot be saved, SPC philosophy dictates that an assignable cause be found and corrected so as not to affect any future batches.

8. CONCLUSION

Multivariate SPC methods have been presented for the analysis and on-line monitoring of batch and semi-batch processes. The only information needed to develop these methods is a historical data base on measured process-variable trajectories from past successful batches. MPCA is used to summarize and compress the data with respect both to variables and time into low-dimensional spaces that describe the normal batch operation. In spite of the complexity of the original problem, this approach leads to multivariate SPC charts that are as simple as univariate Shewhart charts but are more powerful in their ability to quickly detect and diagnose special events that may occur during the progress of a batch. The methodology is generic so as to be easily applied in most batch and semi-batch processes and provides a continuous basis for improvement. Control limits for the principal components t-score and residual charts are developed using information from the historical reference distribution of past batches. Finally, the methods are illustrated through their application to the analysis and on-line monitoring of an industrial polymerization reactor.

ACKNOWLEDGMENTS

APPENDIX A: NIPALS ALGORITHM FOR MPCA

1. Unfold $\underline{\mathbf{X}}(I \times J \times K)$ into $\mathbf{X}(I \times JK)$.
2. Scale $\mathbf{X}$.
3. Choose a column of $\mathbf{X}$ as $\mathbf{t}$.
4. $\mathbf{p} = \mathbf{X}'\mathbf{t}$.
5. $\mathbf{p} = \mathbf{p}/|\mathbf{p}|$.
6. $\mathbf{t} = \mathbf{Xp}$.
7. If $\mathbf{t}$ has converged, then go to Step 8; otherwise go to Step 4.
8. $\mathbf{E} = \mathbf{X} - \mathbf{tp}'$.
9. Go to Step 4 with $\mathbf{X} = \mathbf{E}$ to extract the next principal component.

APPENDIX B: CONTROL LIMITS FOR QUADRATIC FORMS OF RESIDUALS

Let $\mathbf{x}$ be an observation vector from a multinormal population $N(\mathbf{0}, \mathbf{\Sigma})$ and λ_i the eigenvalues of $\mathbf{\Sigma}$, then approximate control limits for the quadratic form $Q = \mathbf{x}'\mathbf{x}$ with significance level α are given by (Box 1954) $Q_\alpha = g\chi_{h,\alpha}^2$ and (Jackson and Mudholkar 1979) $Q_\alpha = \theta_1[1 - \theta_2 h_0(1 - h_0)/\theta_1^2 + z_\alpha(2\theta_2 h_0^2)^{1/2}/\theta_1]^{1/h_0}$, where χ_h^2 is the chi-squared variable with h df and z is the Normal variable that has the same sign as h_0. The rest of the parameters are $\theta_1 = \sum \lambda_i$, $\theta_2 = \sum \lambda_i^2$, $\theta_3 = \sum \lambda_i^3$, $g = \theta_2/\theta_1$, $h = \theta_1^2/\theta_2$, and $h_0 = 1 - 2\theta_1\theta_3/3\theta_2^2$. The relationship between them becomes clear when one uses the Wilson-Hilferty approximation for the chi-squared variable (Evans, Hastings, and Peacock 1993) and rewrites Box's equation as follows: $Q_\alpha \cong gh[1 - 2/9h + z_\alpha(2/9h)^{1/2}]^3$. Every term (except the second one) in this equation approximates well the corresponding term in Jackson and Mudholkar's equation when one has extracted most of the significant principal components (λ_i with large values), and thus $\theta_2^2 \approx \theta_1\theta_3$.

In our case, we use Jackson and Mudholkar's equation for the control limit on Q, and we estimate the θ_i from the estimated residual covariance matrix. The matrices $\mathbf{E}'\mathbf{E}(JK \times JK)$ and $\mathbf{EE}'(I \times I)$ have the same eigenvalues: $\mathbf{V} = \mathbf{EE}'/(I - 1)$, $\theta_1 = \text{trace}(\mathbf{V})$, $\theta_2 = \text{trace}(\mathbf{V}^2)$, and $\theta_3 = \text{trace}(\mathbf{V}^3)$. The control limits for the SPE are based on Box's equation, and we approximate g and h by matching moments of the $g\chi^2$ distribution. We chose to do this because we have to estimate a control limit at each time interval, and this way is faster than using the traces of powers of the residual covariance matrix ($J \times J$) at each time interval. Thus let m and v be the estimated mean and

variance of the SPE at a particular interval k, then the g and h are approximated by $g = v/2m$ and $h = 2m^2/v$. The method of matching moments is susceptible to error when there are outliers in the data or when the number of observations is small. Outliers have been avoided, however, by the careful selection of the reference normal data base, and the number of SPE values used to estimate the control limit at each time interval is fairly large due the smoothing window we apply in our data.

APPENDIX C: PCA PROJECTION

At each time interval (k) we are searching for a t vector ($\mathbf{t}_{R,k}(R \times 1)$), which, along with the p loadings from the PCA model developed from the training set of good batches, will approximate the observations available up to time period k ($\mathbf{x}_{\text{new},k}(kJ \times 1)$) with the least squared error $\mathbf{x}_{\text{new},k} = \mathbf{P}_k \mathbf{t}_{R,k}$. The $\mathbf{P}_k(kJ \times R)$ is the matrix that has as columns all the elements of the p-loading vectors ($\mathbf{p}_r$) up to time interval k from all R principal components. The least squares solution to the preceding equation, which is the projection of vector $\mathbf{x}_{\text{new},k}$ onto the space defined by $\mathbf{P}_k$, is given by $\mathbf{t}_{R,k} = (\mathbf{P}_k' \mathbf{P}_k)^{-1} \mathbf{P}_k' \mathbf{x}_{\text{new},k}$. The matrix $(\mathbf{P}_k' \mathbf{P}_k)^{-1}$ is well conditioned even for the early time intervals, and it approaches the identity matrix as k approaches K (the final time interval) due to the orthogonality of the p-loading vectors.

[*Received January 1994. Revised June 1994.*]

REFERENCES

Bauer, P., and Hackl, P. (1980), "An Extension of the MOSUM Technique for Quality Control," *Technometrics*, 22, 1–8.

Bonvin, D., and Rippin, D. W. T. (1990), "Target Factor Analysis for the Identification of Stoichiometric Models," *Chemical Engineering Science*, 45, 3417–3426.

Box, G. E. P. (1954), "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems: Effect of Inequality of Variance in One-Way Classification," *The Annals of Mathematical Statistics*, 25, 290–302.

Box G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experiments*, New York: John Wiley.

Box, G. E. P., Hunter, W. G., MacGregor, J. F., and Erjavec, J. (1973), "Some Problems Associated With the Analysis of Multiresponse Data," *Technometrics*, 15, 33–51.

Box, G. E. P., and Kramer, T. (1992), "Statistical Process Monitoring and Feedback Adjustment—A Discussion," *Technometrics*, 34, 251–267.

Chew, V. (1968), "Simultaneous Prediction Intervals," *Technometrics*, 10, 323–330.

D'Agostino, R. B., and Stephens, M. A. (1986), *Goodness-of-fit Techniques*, New York: Marcel Dekker.

Efron, B. (1983), "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation," *Journal of the American Statistical Association*, 78, 316–331.

———(1986), "How Biased Is the Apparent Error Rate of a Prediction Rule?" *Journal of the American Statistical Association*, 81, 461–470.

Evans, M., Hastings, N., and Peacock, B. (1993), *Statistical Distributions*, New York: John Wiley.

Geladi, P., Isaksson, H., Lindqvist, L., Wold, S., and Esbensen, K. (1989), "Principal Component Analysis of Multivariate Images," *Chemometrics and Intelligent Laboratory Systems*, 5, 209–220.

Geladi, P. (1989), "Analysis of Multi-way (Multi-mode) Data," *Chemometrics and Intelligent Laboratory Systems*, 7, 11–30.

Hahn, G. J., and Cockrum, M. B. (1987), "Adapting Control Charts to Meet Practical Needs: A Chemical Processing Application," *Journal of Applied Statistics*, 14, 35–52.

Hahn, G. J., and Meeker, W. Q. (1991), *Statistical Interval. A Guide for Practitioners*, New York: John Wiley.

Holloway, L. E., and Krogh, B. H. (1992), "On-Line Trajectory Encoding for Discrete-Observation Process Monitoring," unpublished paper presented at the On-line Fault Detection and Supervision in the Chemical Process Industries, IFAC Symposium, Newark, Delaware, April 22–24.

Horswell, R. L., and Looney, S. W. (1992), "A Comparison of Tests for Multivariate Normality That Are Based on Measures of Multivariate Skewness and Kurtosis," *Journal of Statistical Computation and Simulation*, 42, 21–38.

Jackson, J. E. (1991), *A User's Guide to Principal Components*, New York: John Wiley.

Jackson, J. E., and Mudholkar, G. S. (1979), "Control Procedures for Residuals Associated With Principal Component Analysis," *Technometrics*, 21, 341–349.

Jenkins, G. M., and Watts, D. G. (1968), *Spectral Analysis and its Applications*, San Francisco: Holden-Day.

Jensen, D. R., and Solomon, H. (1972), "A Gaussian Approximation to the Distribution of a Definite Quadratic Form," *Journal of the American Statistical Association*, 67, 898–902.

John, P. W. (1990), *Statistical Methods in Engineering Quality Assurance*, New York: John Wiley.

Johnson, R. A., and Wichern, D. W. (1988), *Applied Multivariate Statistical Analysis*, Englewood Cliffs, NJ: Prentice-Hall.

Jolliffe, I. T. (1986), *Principal Component Analysis*, New York: Springer-Verlag.

Konstantinov, K. B., and Yoshida, T. (1992), "Real-Time Qualitative Analysis of the Temporal Shapes of (Bio)process Variables," *AIChE Journal*, 38, 1703–1715.

Kresta, J., MacGregor, J. F., and Marlin, T. E. (1991), "Multivariate Statistical Monitoring of Process Operating Performance," *Canadian Journal of Chemical Engineering*, 69, 35–47.

Krzanowski, W. J. (1983), "Cross-Validatory Choice in Principal Component Analysis, Some Sampling Results," *Journal of Statistical Computation and Simulation*, 18, 299–314.

———(1987), "Cross-Validation in Principal Component Analysis," *Biometrics*, 43, 575–584.

Kuo, B. C. (1987), *Automatic Control Systems*, Englewood Cliffs, NJ: Prentice-Hall.

Lilliefors, H. W. (1967), "On the Kolmogorov–Smirnov Test for Normality With Mean and Variance Unknown," *Journal of the American Statistical Association*, 62, 399–402.

MacGregor, J. F., and Harris, T. J. (1993), "The Exponentially Weighted Moving Average," *Journal of Quality Technology*, 25, 106–118.

MacGregor, J. F., Jaeckle, C., Kiparissides, C., and Koutoudi, M. (1994), "Monitoring and Diagnosis of Process Operating Performance by Multi-Block PLS Methods With an Application to Low Density Polyethylene Production," *AIChE Journal*, 40, v838.

MacGregor, J. F., and Nomikos, P. (1992), "Monitoring Batch Processes," in *Batch Processing Systems Engineering: Current Status and Future Directions* (NATO ASI Series F), eds. Reklaitis, Rippin, Hortacso, and Sunol, Heidelberg: Springer-Verlag.

Marsh, C. E., and Tucker, T. W. (1991), "Application of SPC Techniques to Batch Units," *ISA Transactions*, 30, 39–47.

Miller, P., Swanson, R. E., and Heckler, C. E. (1993), "Contribution Plots: The Missing Link in Multivariate Quality Control," unpublished manuscript to be submitted to *Journal of Quality Technology*.

Montgomery, D. C. (1994), *Statistical Quality Control*, New York: John Wiley.

Nomikos, P., and MacGregor, J. F. (1994), "Monitoring of Batch Process Using Multiway Principal Component Analysis," *AIChE Journal*, 40, 1361–1375.

Osten, D. W. (1988), "Selection of Optimal Regression Models via Cross-Validation," *Journal of Chemometrics*, 2, 39–48.

Prinz, O., and Bonvin, D. (1992), "Monitoring Discontinuous Reactors Using Factor-Analytical Techniques," paper presented at IFAC Symposium, DYCORD +92, College Park, Maryland, April 26–29.

Sanchez, E., and Kowalski, B. R. (1990), "Tensorial Resolution: A Direct Trilinear Decomposition," *Journal of Chemometrics*, 4, 29–45.

Skagerberg, B., MacGregor, J. F., and Kiparissides, C. (1992), "Multivariate Data Analysis Applied to Low Density Polyethylene Reactors," *Chemometrics and Intelligent Laboratory Systems*, 14, 341–356.

Smilde, A. K., and Doornbos, D. A. (1991), "Three-Way Methods for the Calibration of Chromatographic Systems: Comparing PARAFAC and Three-Way PLS," *Journal of Chemometrics*, 5, 345–360.

Stone, M. (1974), "Cross-Validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society*, Ser. B, 2, 111–133.

Tracy, N. D., Young, J. C., and Mason, R. L. (1992), "Multivariate Control Charts for Individual Observations," *Journal of Quality Technology*, 24, 88–95.

Tucker, W. T., Faltin, F. W., and Vander Wiel, S. A. (1993), "Algorithmic Statistical Process Control: An Elaboration," *Technometrics*, 35, 363–375.

Vander Wiel, S. A., Tucker, W. T., Faltin, F. W., and Doganaksoy, N. (1992), "Algorithmic Statistical Process Control: Concepts and an Application," *Technometrics*, 34, 286–297.

Wold, S. (1978), "Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models," *Technometrics*, 20, 397–405.

Wold, S., Geladi, P., Esbensen, K., and Ohman, J. (1987), "Multi-Way Principal Components and PLS Analysis," *Journal of Chemometrics*, 1, 41–56.

Zeng, Y., and Hopke, P. K. (1990), "Methodological Study Applying Three-Way Factor Analysis to Three-Way Factor Analysis to Three-Way Chemical Datasets," *Chemometrics and Intelligent Laboratory Systems*, 7, 237–250.