Monitoring of Batch Processes through State-Space Models

Jay H. Lee
Georgia Institute of Technology, School of Chemical Engineering, Atlanta, GA 30332

Andrew W. Dorsey
Purdue University, School of Chemical Engineering, West Lafayette, IN 47907

DOI 10.1002/aic.10114
Published online in Wiley InterScience (www.interscience.wiley.com).

The development of a state-space framework for monitoring batch processes that can complement the existing multivariate monitoring methods is presented. A subspace identification method will be used to extract the dynamic and batch-to-batch trends of the process and quality variables from historical operation data in the form of a "lifted" state-space stochastic model. A simple monitoring procedure can be formed around the state and residuals of the model using appropriate scalar statistical metrics. The proposed state-space monitoring framework complements the existing multivariate methods like the multi-way PCA method, in that it allows us to build a more complete statistical representation of batch operations and use it with incoming measurements for early detection of not only large, abrupt changes but also subtle changes. In particular, it is shown to be effective for detecting changes in the batch-to-batch correlation structure, slow drifts, and mean shifts. Such information can be useful in adapting the prediction model for batch-to-batch control. The framework allows for the use of on-line process measurements and/or off-line quality measurements. When both types of measurements are used in model building, one can also use the model to predict the quality variables based on incoming on-line measurements and quality measurements of previous batches. © 2004 American Institute of Chemical Engineers *AIChE J,* 50: 1198–1210, 2004
Keywords: process control, batch control, monitoring, PCA, subspace identification, multivariate statistics, inferential prediction

Introduction

The primary objective of statistical monitoring of a batch process is to ensure that significant and sustained changes in the product quality (brought on by disturbances and faults) are detected quickly. The traditional way of doing this is to plot the sampled quality variables to see whether they fall within specified limits on a Shewart chart. However, the effectiveness of the conventional Shewart control charts for batch process monitoring has proven to be limited, particularly in terms of detecting small mean shifts and slow drifts that are common in chemical processes. This has led to the more common implementations of cumulative sum (CUSUM)– or exponentially weighted moving average (EWMA)–based methods for their improved ability to detect mean shifts or the presence of strong autocorrelation in the process. However, the lack of on-line quality measurements limits the efficiency of these classical statistical process control (SPC) methods in making quality control or product release decisions.

To alleviate the limitation, many researchers have considered using multivariate statistical models built around on-line process measurements to provide monitoring at the batch level. The use of techniques such as principal-component analysis

Correspondence should be addressed to J. H. Lee at jay.lee@che.gatech.edu.

(PCA) and partial least squares (PLS) for batch process monitoring has been discussed extensively in the literature (Nomikos and MacGregor, 1994, 1995a,b). These multivariate SPC (MSPC) methods rely on the extracted statistical information about how the process measurements vary together around some nominal conditions. MSPC methods have been shown to be effective in detecting and diagnosing events that have either caused a significant change in the dynamic correlation structure among the process variables and/or caused the variables to go outside the normal range as defined by the historical data. In the absence of an immediate quality measurement, these deviations can be used to distinguish between "good" and "bad" product quality.

In practice, it is not uncommon for disturbances to evolve over several batches, leading to a gradual drift of the product quality out of acceptable control limits. In addition, small shifts in the product quality can occur, which may not immediately translate into products outside the normal range but just an increased risk for off-spec products. The occurrence of such changes can be detected by current MSPC methods but tight bounds may be needed that can lead to many false alarms. In other words, by monitoring each batch independently, the existing MSPC methods may not be very efficient in terms of detecting these types of changes. Moreover, changes in batch-to-batch correlation structure can go unnoticed by these methods. Such changes can be relevant for batch process control, given that many industrial batch processes are controlled in a batch-to-batch manner and recipe adjustment rules used by an operator or a controller may be based on a particular correlation behavior. Adjusting recipes based on a wrong correlation behavior can worsen the result and hence the importance of understanding the predominant correlation structure and detecting any significant deviation from it.

As mentioned earlier, the same deficiency was realized with the classical control charts, which in turn led to the widespread implementations of CUSUM- or EWMA-based methods. However, all these methods implicitly assume that deviations in the monitored variables are serially independent. In the case of strongly autocorrelated processes, the detection efficiency of these methods can be enhanced by so-called whitening of the sequence before applying them. For batch processes, it is of interest to develop monitoring methods within the multivariate SPC setting that are capable of detecting onset of such changes. At the simplest level, one could apply a CUSUM or EWMA test to the relevant variables monitored by the MSPC method (such as the scores of PCA). This may help detect drifts and mean shifts faster but it may not help detect changes in the correlation structure in general. Also, in view of the fact that many industrial batches show fairly strong batch-to-batch correlations, even under normal operating conditions, applying these methods directly to the autocorrelated data may not result in the most efficient detection schemes. In addition, it is desirable but not clear how to integrate different types of on- and off-line measurements for monitoring, quality prediction, and control.

Our aim is to develop a batch monitoring framework, which can nicely complement the current MSPC methods by improving the ability to detect the aforementioned types of changes, which may not immediately lead to off-spec batches but nevertheless have strong implications for batch operations. We propose to capture the multivariate dynamic behavior of process and quality measurements, both in time and batch-to-batch senses, in the form of a stochastic state-space model. The state-space framework provides us with the flexibility and generality, such as the ability to integrate different types of on- and off-line measurements and the ability to make quantitative predictions about relevant batch variables and actively control them.

Here, state-space models will be identified directly from process data using subspace identification techniques such as the N4SID method (Van Overschee and De Moor, 1994). Even though these techniques were developed originally for identifying continuous systems, they can be used to model batch processes by using the technique called *lifting*. In our recent conference paper (Dorsey and Lee, 1999), we investigated the use of the resulting model for inferential product quality prediction. In the present article, we will explore the use of the model for process monitoring.

State-space modeling enables us to separate variations seen in normal batch operations into a batchwise correlated part (the state) and a part independent from the previous batches (the innovation). This in turn allows us to use additional tests on the innovation, such as a whiteness test, for detection of changes in the correlation structure. Simple tests like the T^2 monitoring of the innovation as well as its sum over several successive batches or its EWMA will be shown to be very effective for detecting changes in the correlation structure as well as small mean shifts and slow drifts at their onset. A similar method for monitoring continuous systems was proposed by Negiz and Cinar (1997). In their monitoring procedure, however, a single T^2 monitoring statistic was used around the state variables of the model. Going beyond their proposed scheme, we will propose to develop and use scalar statistics for the innovation term for improved detection of various subtle changes.

The primary advantage of using the state-space modeling framework for batch process monitoring is that a more complete representation of the process operation can be extracted from the data and used not only for monitoring but also for inferential prediction and control of product qualities.

The proposed method will nicely complement the existing MSPC methods, which are proven to be very effective at catching events that lead to abnormal batches. In fact, we show that the conventional multi-way PCA-based method can be implemented as a part of the new method by using PCA as a preprocessing step before the identification of the state-space model, to allow for a recursive estimation of the PCA scores (and SPEs) during the batch as well as retaining the ability to monitor each batch independently for abnormality. The rest of the article is organized as follows. We explain how to develop a state-space representation of multidimensional batch variable correlations from normal operating data. We discuss how the model can be used in various ways for on- and off-line monitoring. A case study involving a simulated batch pulp digester is presented to illustrate the useful and other notable features of the new framework. In Section 5, we offer some conclusions.

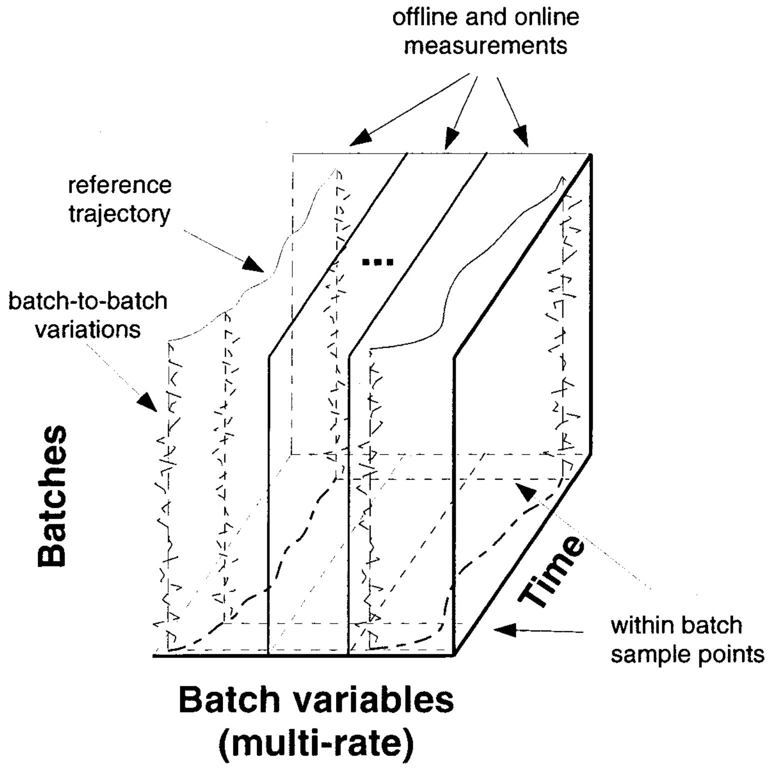

Figure 1. Description of batch data matrix X.

Identification of Batch State-Space Models

Background

Batch process data can be represented by a three-dimensional data array containing the different process variables (indexed through 1, . . . , L) across both the batch (1, . . . , K) and time dimensions (1, . . . , M), as shown in Figure 1. Let **X** ($K \times L \times M$) represent this three-dimensional data array. Traditional multivariate approaches to batch processes begin with the unfolding of this three-dimensional data array into two dimensions before application of the modeling algorithm. The typical choice for unfolding this array is to preserve the batch dimension and rearrange each resulting batch slice into a single vector (by collapsing the "within-batch" dimensions). This unfolding process creates a large data matrix **X** ($K \times LM$), upon which a multivariate statistical modeling technique like PCA or PLS is applied. In that context, the particular way of unfolding results in a model that describes variations of the process variables around their nominal trajectories and treats each batch as an independent observation. Seldom is an attempt made to establish a model along the batch dimension that captures the multivariate statistics among the samples of successive batches.

However, in the presence of drifts, mean shifts, or other changes in the correlation structure, the lack of information on the "batch-to-batch" behavior could limit the effectiveness of the statistical monitoring scheme. For existing MSPC tools to detect such a change at its onset, the detection limits may need to be set very tightly, which can lead to many false alarms. The conventional 3σ bounds or 99% confidence limits may not be very efficient at detecting slow drifts or small mean shifts. The ideal situation is to detect these changes well before the process actually starts producing off-spec products. Then, corrective action could be made before any true quality deviation is experienced, increasing the overall uptime of the plant by reducing the number of batches that are downgraded or disposed.

For a monitoring scheme to detect such changes in the batch-to-batch trends, it helps to model the behavior along the batch dimension of the data matrix **X**. This can be accomplished through developing a dynamic state-space model with respect to the batch dimension using the subspace identification technique. Such a formulation does not treat each batch independently but tries to capture and use multivariate statistical information on the batch-to-batch behavior.

Preliminary steps

The data preparation steps for the identification procedure begin by mean centering and scaling the process measurements so that they are expressed as scaled deviations from the nominal trajectories defined over the batch history. The within-batch dimension of the data is then collapsed into the "lifted" vector. To assign some notation to this procedure, let $\mathbf{y}_k(t)$ be the vector of mean centered and scaled process measurements available at the current sample time t within batch k. Collecting all such vectors for 1, . . . , M sample points, the lifted batch vector for the kth batch can then be expressed as

$$\mathcal{Y}_k^f = [y_k(1)^T, y_k(2)^T, \ldots, y_k(M)^T]^T \quad (1)$$

which contains all of the process measurements within the batch ordered on time. This is done for all the batches (1, . . . ,

K) in the model-building data set, which defines the normal operating condition of the plant. Here, the data defining normal operation should not necessarily include all batches that produced satisfactory end results. One must exclude, for example, a whole sequence of batches along which a problem may have developed later, leading to a higher frequency of abnormal batches or other significant problems.

Model development

The following stochastic model of the process is extracted from the data using a subspace identification technique (Van Overschee and De Moor, 1993)

$$\begin{aligned} \mathcal{X}^f_{k+1} &= A\mathcal{X}^f_k + K\mathcal{E}^f_k \\ \mathcal{Y}^f_k &= C\mathcal{X}^f_k + \mathcal{E}^f_k \end{aligned} \tag{2}$$

The interpretation of the identification procedure in the batch process setting, in which $\mathcal{Y}^f_k$ is the collection of all observations made for the kth batch, is that the state sequence is extracted from the process data based on the relevancy of the previous batch measurements for predicting the future batch measurements. In other words, the state is defined to be a holder of information from previous batches that is relevant for predicting future batches. Therefore, a monitoring algorithm built around the developed state space model delivers a powerful way to detect not only deviations in the level of the process variables but also abnormal deviations in the process behavior from a batch-to-batch dynamics point of view. This can be potentially helpful for detecting unusual changes, both in terms of the magnitudes and batch-to-batch dynamic trends, in the feedstock or other batch operating parameters.

Reducing the dimensionality of the lifted vector

One practical issue that may arise during the application of subspace ID to batch data is the large size of the lifted output vector $\mathcal{Y}^f$. One way to combat the dimensionality problem is to apply PCA before the subspace identification. In most batch processes, variabilities of the measurements are manifested by a relatively small number of parameters. Hence, the data for the lifted vector $\mathcal{Y}^f_k$ are likely to show a high degree of collinearity. One should be able to reduce the lifted sample data trends (other than measurement noises) into a much smaller number of variables that represent the modes of variations in the process variables. If the number of principal components is chosen so that the residuals represent mostly measurement noises, which tend to be timewise and batchwise uncorrelated, and not part of the process trend to be modeled, the scores will retain the original batch-to-batch behavior we want to model. PCA would be applied to project the lifted output vector $\mathcal{Y}^f$ of size (LM) to the score vector $\underline{\mathcal{Y}}$ residing in a lower-dimensional space of size (n). With the compressed output vector, a more compact model can be developed by applying the subspace ID algorithm to obtain the reduced state-space model of the form

$$\begin{aligned} \mathcal{X}_{k+1} &= A\mathcal{X}_k + K\underline{\mathcal{E}}_k \\ \underline{\mathcal{Y}}_k &= C\mathcal{X}_k + \underline{\mathcal{E}}_k \end{aligned} \tag{3}$$

where $\underline{\mathcal{Y}}$ represents the score vector, which is defined by

$$\mathcal{Y}^f = \Theta\underline{\mathcal{Y}} + E \tag{4}$$

where Θ is the matrix whose columns are the principal directions of the PCA model and E is the residual matrix from the PCA. The state-space model output in the reduced dimension can be projected back to the full batch dimension by rewriting the output equation as $\mathcal{Y}^f_k = \Theta(C\mathcal{X}_k + \underline{\mathcal{E}}_k) + E_k$.

State-Space Model–Based Monitoring of Batch Processes

Monitoring procedure

The application of PCA before subspace identification provides some advantages in terms of the monitoring task. The output of the state-space model in this setting becomes the scores of the original PCA model. This is expressed as the sum of $C\mathcal{X}_k$, which is the estimate of the scores from the previous batch measurements, and the prediction error $\underline{\mathcal{E}}$, which is the portion of the scores that are uncorrelated with the previous batch measurements (the "innovation"). This allows the ability to monitor the correlated portion through the state vector and the new portion through the prediction error. Both variables can be monitored using the Hotelling T^2 statistic (as opposed to the Q-statistics that are used to monitor the residuals in the conventional PCA monitoring schemes) because they both lie in the score space and their elements are orthogonal.

The Hotelling T^2 statistic is used based on the assumption that the set of random variables to be monitored are zero mean and follow a multinormal distribution with estimated covariance matrix $\mathbf{\Sigma}$ (Anderson, 1984). The Hotelling T^2 statistic is in the form

$$T^2_k = z_k \sum^{-1} z^T_k \sim \frac{(N-1)n}{(N-n)} F_{n,N-n}(\alpha) \tag{5}$$

where N is the number of samples and n is the dimension of the observation vector $\mathbf{z}$. The F-distribution can be used to establish control limits with significance level α. Hence, if the state variables or the prediction errors in the score space exhibit a loss of orthogonality or a change in level from that defined by the normal operating data, their respective T^2 values will be more likely to exceed the established control limits.

Beyond this "snapshot" analysis, the prediction error ($\underline{\mathcal{E}}$) can also be monitored in terms of its serial correlation. Significant autocorrelation in the residuals is likely an indication of a change in the batch-to-batch dynamics from the normal operating condition, as defined by the model-building data. To detect such a change, we can use a whiteness test of some sort. A simple test one can use for this depends on the situation.

In the case that the normal operation data show little batch-to-batch correlation, states will be negligible and the T^2 monitoring of the prediction error is almost equivalent to the T^2 monitoring of the scores in the PCA approach. When small mean shifts, slow drifts, or changes in the correlation behavior develop, the variance of state and prediction error will increase eventually but it may take a while before the change can be observed through a static test. In such a case, some low-pass–filtered form of the prediction error can be monitored for a quicker detection. For example, assuming the prediction error vector $\underline{\mathscr{E}}$ is multinormal with $N(0, \Xi)$, then summing over a specified number of batches gives a new vector

$$a_k = \sum_{i=k-m+1}^{k} \underline{\mathscr{E}}_i \tag{6}$$

which is also multinormal with $N(0, m\Xi)$, assuming $\underline{\mathscr{E}}$ is an independent, identically distributed (i.i.d.) sequence. Control limits can then be established based on the T^2 calculated around a_k to verify this assumption. Note that the elements of a_k represent mutually orthogonal components, given that $\underline{\mathscr{E}}$ is a vector formed in the score space of PCA, of which the elements are orthogonal. Hence, the T^2 statistic can be used for monitoring this variable as well. The value of m becomes a tuning parameter to achieve the desired sensitivity to drifts and mean shifts of the process. Alternatively, EWMA can be used in place of the moving finite sum.

In the case that the normal operation data show strong batch-to-batch correlation, the states will contain the more significant portion of the scores. A change in the correlation behavior (to a much weaker correlation, for instance) will typically result in an immediate increase in the variance of the prediction error. Hence, the T^2 monitoring of the prediction error can be very effective for the detection of a loss of correlation. The aforementioned T^2 monitoring of the CUSUM or EWMA of the prediction error can still be used to detect sustained shifts and drifts, even though the sensitivity will be lowered simply because the normal data themselves show strong correlation.

One point to emphasize here is that, in practice, in addition to $\mathscr{X}$ and $\underline{\mathscr{E}}$, one should also monitor E_k (through the Q-statistics) because many changes have shown to have strong effects on SPE but not on PCA scores. Hence, the proposed statistics are in addition to the statistics monitored by the conventional MSPC methods and are intended to complement them rather than to replace them.

The additional metrics proposed above offer a better chance to detect relevant changes. Those faults that immediately lead to abnormal batches should show up in the states and in the prediction error within the PCA score space. Subtle changes, such as a slow drift, a mean shift, or a change in the batch-to-batch correlation structure, should be most apparent in the tests on the prediction error because they tend to lead to autocorrelated prediction errors. These issues will be investigated further through the case studies discussed later.

Monitoring formulations

The selection of variables to be included within the lifted vector before identification is greatly dependent on the intent of the model's use. Three possible monitoring formulations can be considered based on the information included in the output vector of the model. These include a formulation that uses off-line quality measurements only, a formulation that uses on-line process measurements, and an integrated approach that uses both sources of information.

Formulation Based on Off-Line Quality Measurements Only. The traditional industrial monitoring schemes rely only on the quality measurements to detect relevant changes in the process behavior. This formulation can easily be incorporated into the state-space formulation by identifying the model directly from the available quality data. Let $\mathbf{q}_k$ be the vector of quality measurements from batch k. Collecting all such vectors over the batch history, the following state-space model can be identified:

$$\begin{aligned} x_{k+1} &= Ax_k + K\varepsilon_k \\ q_k &= Cx_k + \varepsilon_k \end{aligned} \tag{7}$$

As the quality measurements become available, they would be used in the developed Kalman filter to obtain the optimal estimate of the state vector.

$$x_{k+1} = Ax_k + K[q_k - Cx_k] = (A - KC)x_k + Kq_k \tag{8}$$

The above T^2 metrics could then be calculated for the state x_k and prediction errors $(q_k - Cx_k)$ at the end of batch k and monitored. The $\mathbf{\Sigma}$ in Eq. 5 would be calculated separately for both the state and residual based on the data generated around the historical data. In the common situation when quality measurements are delayed considerably, the filter will lag behind the process. Alternatively, the delayed quality variables could be augmented onto the state vector to handle the delays. However, in this setting with no other information available during batch k, there is no significant performance improvement in terms of monitoring.

Formulation Based on On-Line Process Measurements. One attractive feature of the PCA based method is its ability to use on-line measurement information for monitoring. An on-line monitoring procedure can be developed in the present framework by including all of the on-line measurements within the lifted vector $\mathscr{Y}_k^f$ and performing PCA before identification of the state-space model.

$$\begin{aligned} \mathscr{X}_{k+1} &= A\mathscr{X}_k + K\underline{\mathscr{E}}_k \\ \underline{\mathscr{Y}}_k &= C\mathscr{X}_k + \underline{\mathscr{E}}_k \end{aligned} \tag{9}$$

The advantage of this formulation over the purely off-line formulation is that there is no delay from waiting for the off-line quality measurements. The model can be used in a batch-to-batch fashion (analogous to PCA) by applying score vector $\mathscr{Y}$ extracted from the full measurement vector $\mathscr{Y}^f$ to the model Eq. 9 at the end of each batch.

$$\mathscr{X}_{k+1} = A\mathscr{X}_k + K[\underline{\mathscr{Y}}_k - C\mathscr{X}_k] \tag{10}$$

Again, $\mathscr{X}_{k+1}$ and $\underline{\mathscr{Y}}_k - C\mathscr{X}_k$ would be monitored upon completion of batch k.

A disadvantage of the above formulation is that the monitoring result becomes available only after the completion of the batch. Instead, the model can be used in real time as the batch proceeds to provide monitoring throughout the batch. This is an attractive formulation if the process can allow for some corrective action during the batch, or can benefit from the ability to make product release decisions before the end of the batch, either to scrap the batch and begin again or modify the final target of the product.

The "within-batch" monitoring procedure can be implemented by using the same batch-to-batch model of Eq. 9 for recursive estimation through the use of a periodically time-varying (PTV) Kalman filter. To do this it is convenient to create the following augmented system vector

$$\mathfrak{X}_k \equiv \begin{bmatrix} \mathscr{X}_{k+1} \\ \underline{\mathscr{E}}_k \\ \underline{\mathfrak{Y}}_k \end{bmatrix} \tag{11}$$

which contains all of the desired variables that would be monitored for batch k. Notice here that we choose to include $\mathscr{X}_{k+1}$ instead of $\mathscr{X}_k$ in $\mathfrak{X}_k$, which is composed of the variables to be monitored during the kth batch. This is based on the consideration that $\mathscr{X}_{k+1}$ holds the relevant information contained in the measurements up to the kth batch, whereas $\mathscr{X}_k$ holds the information only up to the $(k-1)$th batch. However, defining $\mathfrak{X}_k$ with $\mathscr{X}_k$ would have also been fine as any fresh disturbance information contained in the measurements for the kth batch would be represented in the estimate of $\underline{\mathscr{E}}_k$.

The batch-to-batch transition model of Eq. 9 can be equivalently expressed as the following PTV system

$$\mathfrak{X}_{k+1}(0) = \underbrace{\begin{bmatrix} A & 0 & 0 \\ 0 & 0 & 0 \\ C & 0 & 0 \end{bmatrix}}_{\Phi} \mathfrak{X}_k(M) + \begin{bmatrix} K \\ I \\ I \end{bmatrix} \underline{\mathscr{E}}_{k+1} \tag{12}$$

$$\begin{aligned} \mathfrak{X}_k(t+1) &= \mathfrak{X}_k(t) \\ y_k(t+1) &= H(t)\Theta[0 \quad 0 \quad I]\mathfrak{X}_k(t) + e_k(t+1) \\ & \qquad t \in \{0, \ldots, M-1\} \end{aligned} \tag{13}$$

where $H(t)$ is a matrix that extracts $y_k(t)$ from the lifted vector $\mathfrak{Y}_k^f$ and $e_k(t)$ is the PCA residual at time t, which we will assume to be i.i.d. sequences. Based on the above model representation, a PTV Kalman filter can be constructed (Lee et al., 1992), to estimate $\mathscr{X}_{k+1}$, $\underline{\mathscr{E}}_k$, and $\underline{\mathfrak{Y}}_k$ recursively as each measurement becomes available during the kth batch.

Notice here that the state contains $\underline{\mathfrak{Y}}_k$, which contains the PCA scores. This is certainly redundant given that $\mathscr{X}$ and $\underline{\mathscr{E}}$ define $\underline{\mathfrak{Y}}$. However, by including it, the PTV Kalman filter formulation provides a recursive estimate of the PCA scores, which may be useful. The estimated $\mathscr{X}_{k+1}$ and $\underline{\mathscr{E}}_k$ as well as some filtered form of $\underline{\mathscr{E}}_k$ can be monitored through the T^2 tests as before.

For this formulation the covariance matrix $\mathbf{\Sigma}$ in the T^2 calculations (Eq. 5) will be based on all of the data points for different times obtained by applying the recursive formulation of the state-space model to the historical data base (a total of KLM data points). Alternatively one could calculate the M separate covariance for every sample time within the batch. This would be a tedious task but may be required if the variables to be monitored are dispersed unevenly across the batch trajectory. In that case some sensitivity would be lost in detecting changes at the lower levels of these variables if only a lumped covariance matrix was used.

Integrated Approach. The final and most flexible formulation allows for inferential quality prediction as well as batch monitoring (in a manner analogous to PLS) by including all the measurements from each batch, including the on-line process measurements and the off-line quality variables. Quality variables q_k, which define the terminal condition of the batch, are appended to the lifted on-line measurement score vector $\mathfrak{Y}$ before the identification of the model. We would then identify the model based on this lifted vector

$$\begin{aligned} \mathscr{X}_{k+1} &= A\mathscr{X}_k + K \begin{bmatrix} \mathscr{E}_{\mathfrak{Y}} \\ \mathscr{E}_q \end{bmatrix}_k \\ \begin{bmatrix} \underline{\mathfrak{Y}}_k \\ q_k \end{bmatrix} &= \underbrace{\begin{bmatrix} C_{\mathfrak{Y}} \\ C_q \end{bmatrix}}_{C} \mathscr{X}_k + \underbrace{\begin{bmatrix} \mathscr{E}_{\mathfrak{Y}} \\ \mathscr{E}_q \end{bmatrix}_k}_{\underline{\mathscr{E}}_k} \end{aligned} \tag{14}$$

It is likely that the quality measurement vector q_k will not be available immediately at the end of batch k. For the case when the quality measurements are not available before the next batch begins, one can carry over the quality variable in the state vector to allow for correction based on the laboratory measurements when they become available. To incorporate the lab measurements of the previous batch, the identified model is simply augmented with the quality variables of the previous batch. Any amount of delay in the lab analysis can be handled by augmenting the state with an appropriate number of past batch quality variables. Even when the analysis delay is not fixed, an upper bound on the delay can be established based on typical laboratory analysis times of the quality variables. For example, assuming the quality measurements are delayed by at most two batches, the following augmented state vector can be created

$$Z_k = [x_{k+1}^T, \underline{\mathscr{E}}_k^T, \underline{\mathfrak{Y}}_k, q_k^T, q_{k-1}^T, q_{k-2}^T]^T \tag{15}$$

where $\underline{\mathfrak{Y}}_k$ has been included as well to allow for the recursive formulation of the PCA scores during the batch.

For recursive calculation of Z_k in real time, the following PTV model derived from the above batch-to-batch model can be used

$$Z_k(0) = \underbrace{\begin{bmatrix} A & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ C_{\mathfrak{Y}} & 0 & 0 & 0 & 0 & 0 \\ C_q & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \end{bmatrix}}_{\Phi} Z_{k-1}(M) + \begin{bmatrix} K \\ I \\ \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \\ 0 \\ 0 \end{bmatrix} \underline{\mathscr{E}}_k \tag{16}$$

$$\begin{aligned} Z_k(t+1) &= Z_k(t) \\ y_k(t+1) &= \underbrace{H(t+1)\Theta[0 \quad 0 \quad I \quad 0 \quad 0 \quad 0]}_{\Pi(t)} \\ & \times Z_k(t+1) + e_k(t+1) \quad t \in \{0, \ldots, M-1\} \end{aligned} \tag{17}$$

Based on the above model, a time-varying Kalman filter can be designed to update $Z_k(t)$ as measurements at each time become available. Measurements of the quality variables, whenever they become available, can also be included in the Kalman filter update (by appropriately modifying the measurement matrix $\mathbf{\Pi}$ to include the measured q). This formulation offers an important advantage over previous methods, in that it allows for the use of both on-line and off-line information in the monitoring decision as well as an opportunity to predict the quality variables (possibly for on-line inferential product quality control) all in a single framework.

Case Study

The proposed state-space monitoring formulation will be tested on data generated from a fundamental model of the pulp digester (for model details, see Datta, 1996). For this study the major source of disturbances are assumed to occur from feedstock variations in the wood chips. Variations of the chip moisture, chip size, and chip composition have an impact on the kinetics throughout the batch and lead to deviations in the quality variables, the kappa number, and yield. Measurements of effective alkali, lignin, sulfide, density, and total solids were assumed to be available every 5 min throughout the batch duration of 3 h. This leads to an output vector with dimension of 185 after lifting the process trajectories as described by Eq. 1. The dimension of the full lifted vector $\mathcal{Y}^f$ is prohibitively high for identification of a state space model from limited data. Hence, PCA will be used as before throughout the case study when on-line measurements are involved to reduce the size of the output vector before the identification. Where necessary, laboratory data of the kappa number and yield of the pulp will be combined with these measurements to illustrate the different formulations. In all cases the normal batch history and test data are generated by assuming that the chip properties vary from batch to batch, according to

$$d_k = \phi d_{k-1} + \zeta_k \tag{18}$$

where ζ_k is an i.i.d. sequence; ϕ can be adjusted to model different degrees of correlation in the batch-to-batch behavior. The base case scenario for this article will assume that the batch history is weakly correlated with a value of $\phi = 0.1$. However, in the later part of the study we will see the impact of a stronger correlation in the batch history has on the monitoring performance and set $\phi = 0.9$ to generate the normal history.

To show the complementary nature of the proposed monitoring method, we considered test cases with subtle changes in the feedstock that may not lead to off-spec products on an immediate basis but develop into significant problems in the long term. These types of changes could go unnoticed by the conventional PCA method. Hence it will be shown that the developed framework offers a nice complement to PCA.

Model building

In generating the normal batch history for the base case scenario, it was assumed that the process had very little batch-to-batch correlation and 200 batches were generated, with ϕ set to 0.1 in Eq. 18 to simulate this normal condition. Only on-line measurements from these batches were used to obtain a PCA model with three principal components and in turn a state-space model of order 6 was identified directly from the scores of the PCA model.

Detecting changes in correlation

First we investigate the impact of correlation on the monitoring algorithm, particularly in its ability in detecting changes in the overall degree of batch-to-batch correlation. As we mentioned, this is relevant, given that optimal batch-to-batch adjustment of operating conditions or time varies with the degree of the correlation. Such situations will be simulated by changing the value of ϕ in Eq. 18. Figure 2 shows the behavior of the quality variable and the monitoring performance assuming the process has changed from what is considered to be the normal operating condition of weakly correlated behavior to the stronger correlated case. In this figure and all the subsequent figures, the bounds shown for the quality variable chart are the 3σ bounds and the bounds for all the control charts (shown by the horizontal dotted lines on the plots) are the 99% confidence limits. This change was made after batch 25 by changing the value of ϕ from 0.1 to 0.9 in Eq. 18 and adjusting the covariance of ζ by a factor of $(1 - 0.9)^2/(1 - 0.1)^2$ to keep the total feedstock variance at the same level. The figure shows the scaled kappa number, the T^2 values calculated around the state vector, prediction error, and the sum of the prediction error over five consecutive batches. The performance of the PCA-based monitoring metrics are also shown. In this case the performance of the metrics around the prediction error and the scores of the PCA model are very similar. This is primarily attributed to the presence of little batch-to-batch correlation in the normal operating condition. Therefore, the resulting state-space model appropriately treats the majority of the score variations as being batchwise independent. Although the PCA is unable to detect the change, the T^2 monitoring of the cumulative sum to the state-space prediction error allows for a quick detection of the change in the behavior. Note that the state also indicates a potential problem as the quality extends close to the limits.

Detection of a drift in product quality

Next we are concerned with detecting a drift of the product quality within the normal operating range of the process. To test the monitoring performance, a new data set was generated that had a slow ramp in the feedstock variables midway through 50 batches. In Figure 3, the quality variables and the monitoring performance are shown. The state vector still shows good sensitivity to the change. Most sensitive to the ramp change is once again the sum of the prediction error over five batches, which is indicated in the figure. As expected, the model finds little correlation in the process; hence, when the strong correlated type drift occurs in the process, the test offers a good indication of the change around batch 40. The sensitivity of the CUSUM procedure could be adjusted by increasing the value of m in Eq. 6. The metrics from the PCA model and the state-space prediction error signal faults as the process begins to drift close to the specified limits.

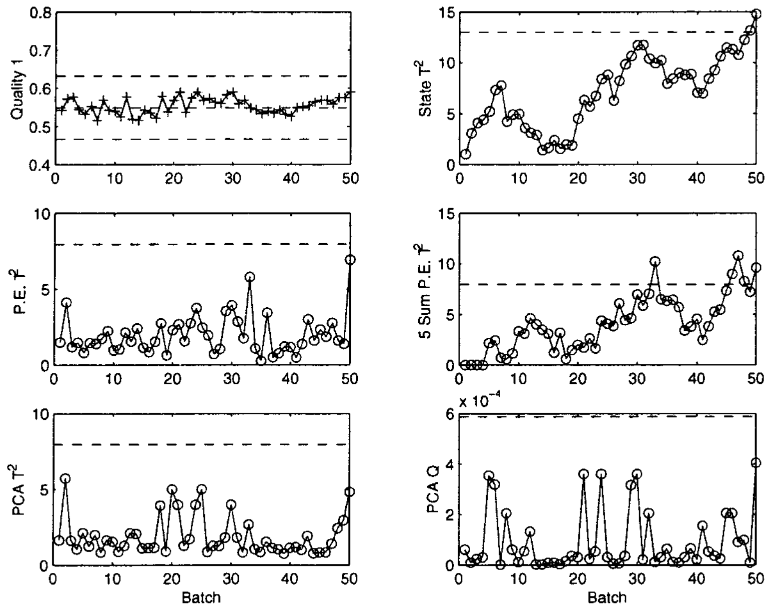

Figure 2. Monitoring performance the state-space model and the conventional PCA model during a change from weak to strongly correlated behavior.

Detection of a small mean shift in product quality

This section considers another example of a subtle change, a small mean shift occurring inside what would for all practical purposes be classified as the normal operating range. Even though such a change may not lead to an off-spec product on an immediate basis, it will surely increase the frequency of off-spec products in the future. For the test data, we simulated 50 batches with a slight mean shift occurring in the chip properties, starting around batch 26. The monitoring results are shown in Figure 4. Again, notice that the T^2 around the state vector is nearly identical in behavior to the T^2 value from PCA (ascribed to the weak correlation in the normal data set). Also notice that the sum of the prediction errors of the previous five batches shows a greater ability in detecting this mean shift in the process.

Impact of a short-lived upset

For this case a single abrupt quality deviation was simulated around batch 26. Using the same base case of the weakly correlated batch history the model was tested again for this new disturbance and is shown in Figure 5. Notice that for this weakly correlated case, the abrupt, short-lived change gives a

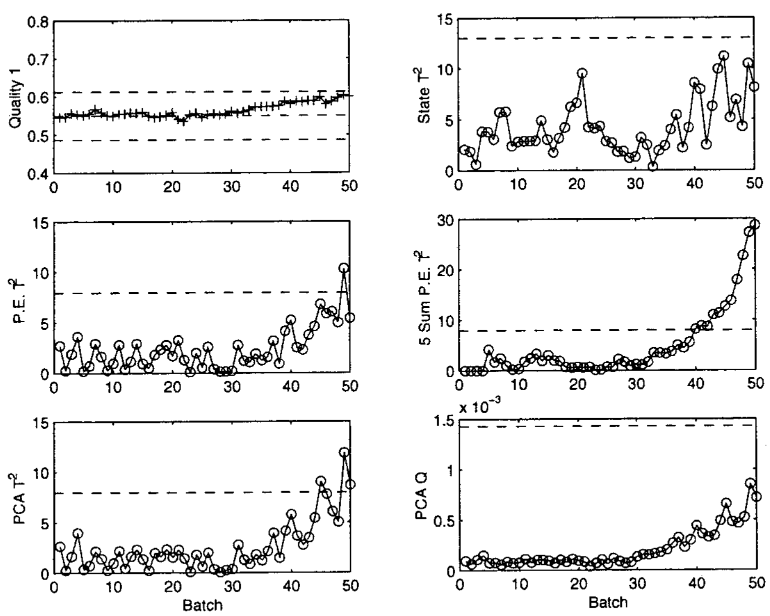

Figure 3. Monitoring performance of the state-space model and the conventional PCA model during a feedstock drift.

This case assumes that only weak batch-to-batch correlation exists during normal historical operations.

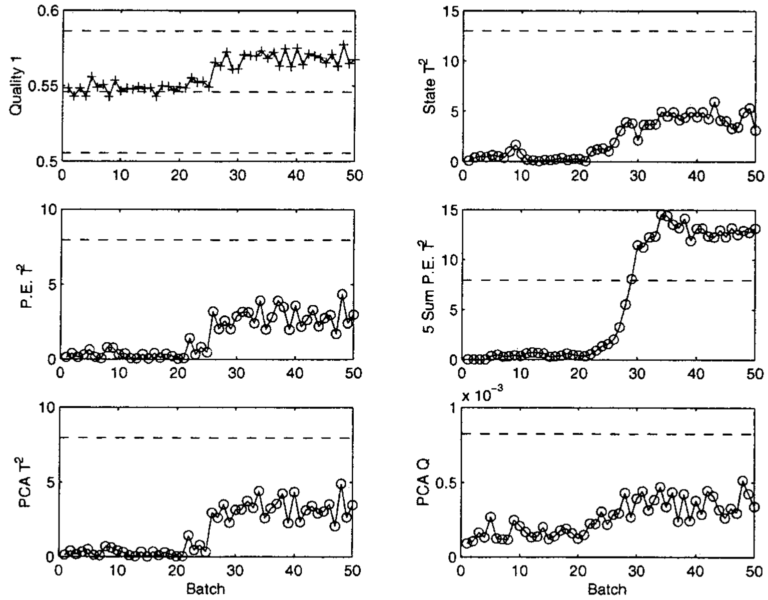

Figure 4. Monitoring performance of the state-space model and the conventional PCA model during a mean shift in the feedstock properties.

This case assumes that only weak batch-to-batch correlation exists during normal historical operations.

large prediction error and PCA score values only for the particular batch. The state is raised appreciably during the event but is not large enough to signal a fault.

Within batch monitoring

As discussed earlier, an on-line formulation can be used to perform monitoring throughout the batch. A model was used that was derived from the normal operating data with weak batch-to-batch correlation ($\phi = 0.1$). The model was then used in the recursive fashion to assess the ability of the framework to detect the change. Figure 6 shows the performance during the same feedstock drift as shown above. Here the dots in the figure represent the points within the batch and the circles are shown to indicate the final point of the batch. Figure 7 shows an example of a within-batch profile taken from batches 45, 49, and 50. For example, in batch 45, the T^2 value around the prediction error goes out of limits 30 min before the end of the batch.

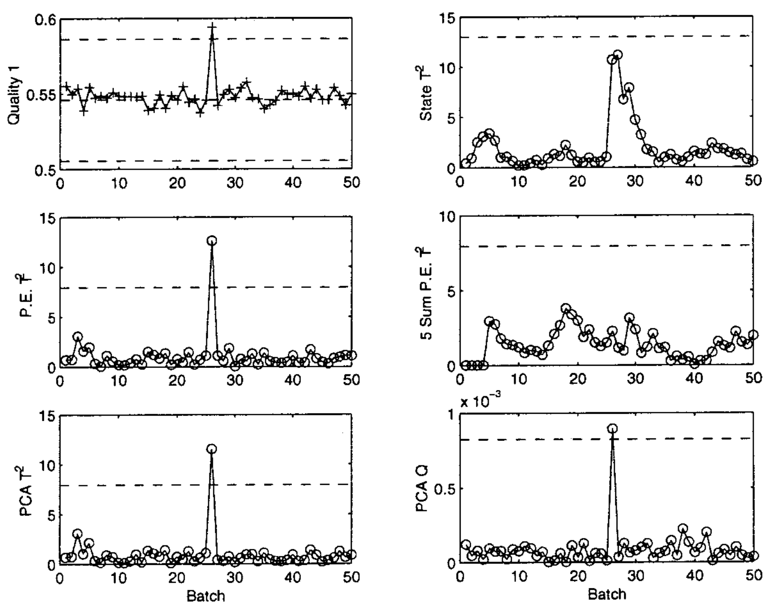

Figure 5. Monitoring performance of the state-space model and the conventional PCA model during an abrupt change in the feed that lasts for a single batch.

This case assumes that only weak batch-to-batch correlation exists during normal historical operations.

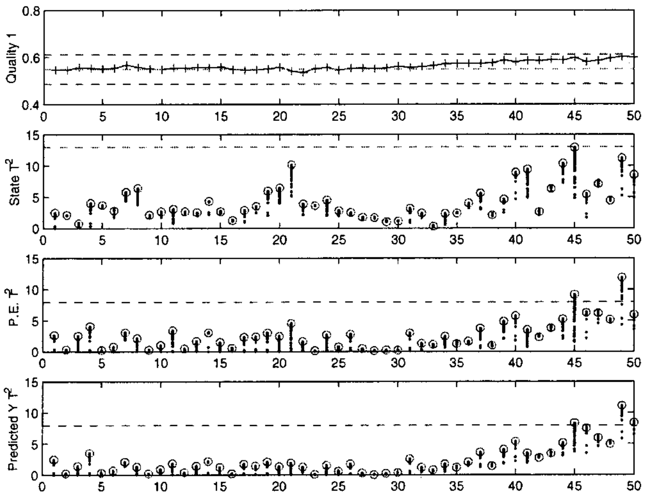

Figure 6. Monitoring performance of the state-space model throughout each batch during a feedstock drift.

This case assumes that only weak batch-to-batch correlation exists during normal historical operations.

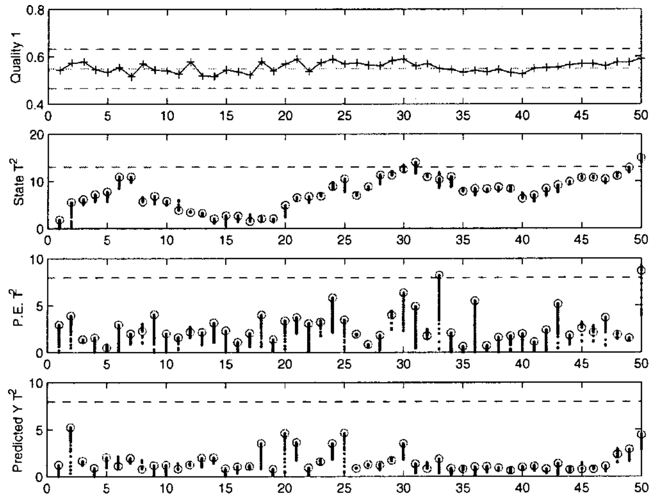

Figure 8. Monitoring performance of the state-space model throughout the batch during a change from weakly to strongly correlated feedstock variations.

Figure 7 also compares the performance of the T^2 value based on the recursive estimate of the score $\mathcal{Y}$ to the actual score vector of the PCA calculation for the three batches that were signaled as faults (45, 49, and 50). Notice that in batches 49 and 50 the method is able to signal quite early (within 15 min) the presence of a potential problem through the recursive formulation of the PCA scores.

Figure 8 repeats the change in feedstock correlation example using the recursive formulation. Here notice that the performance of the metrics around the state and the prediction error are slightly different from that indicated in Figure 2, even though the model and test data were the same in both cases. This is attributed to the differences in the covariance matrices obtained for the T^2 calculation. The actual values of the state and the prediction error at the end of the recursive formulation are equivalent to that obtained by the "snapshot" approach.

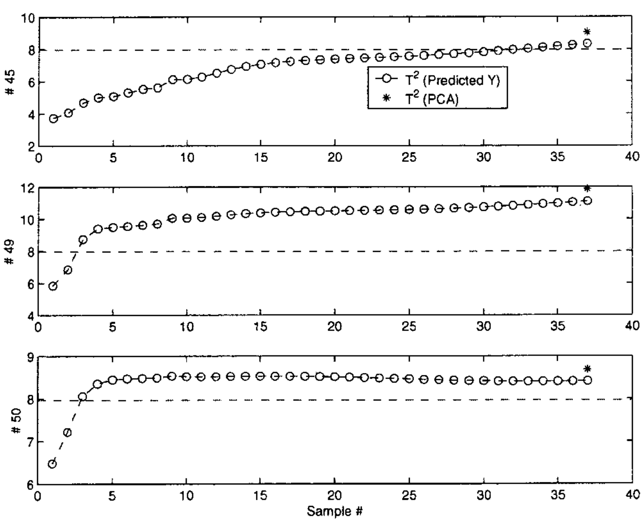

Figure 7. Comparison of T^2 around predicted $\mathcal{Y}$ and conventional PCA for all batches that signaled faults (batches 45, 49, and 50) during a feedstock drift.

Impact of stronger correlations during normal batch operations

In the previous scenarios, the state-space model's prediction errors were very similar in behavior to the PCA scores because of the fact that little correlation was assumed to exist in the batch history. The situation is quite different when the batch history has more substantial correlation. To examine this, we have simulated the case where normal operating data had significant batch-to-batch correlation (ϕ was set to 0.9 in Eq. 18) and built a state-space model with it.

Figure 9 shows the situation where the process changing from what is considered to be the normal operating condition of strongly correlated behavior ($\phi = 0.9$) to the weaker correlated case ($\phi = 0.1$). As we predicted, the T^2 metrics around the prediction error and state vector are both very sensitive to the change, indicating several persistent faults after batch 25.

Figure 10 shows the monitoring performance during a process drift. The identified model is expecting behavior that would be close to a random step changes and when the drift occurs in the process the model prediction continually lag behind the drift leading to the accumulation of the error shown by the cumulative sum of the prediction error. The metrics around the state vector and PCA scores still exhibit good sensitivity, as shown by the ability to signal a fault attributed to the process drifting out of the normal range. However, the statistic on the sum of the prediction error signals the problem before off-spec batches result.

Figure 11 shows the mean shift scenario repeated using a model identified from data with the strong batch-to-batch correlation. Here notice that, even for a relatively small mean shift, the metric around the prediction error shoots up on batch 26, detecting the mean shift in the process. After this batch, however, the metric quickly returns to within its normal range, indicating the ability of the model to follow these correlated

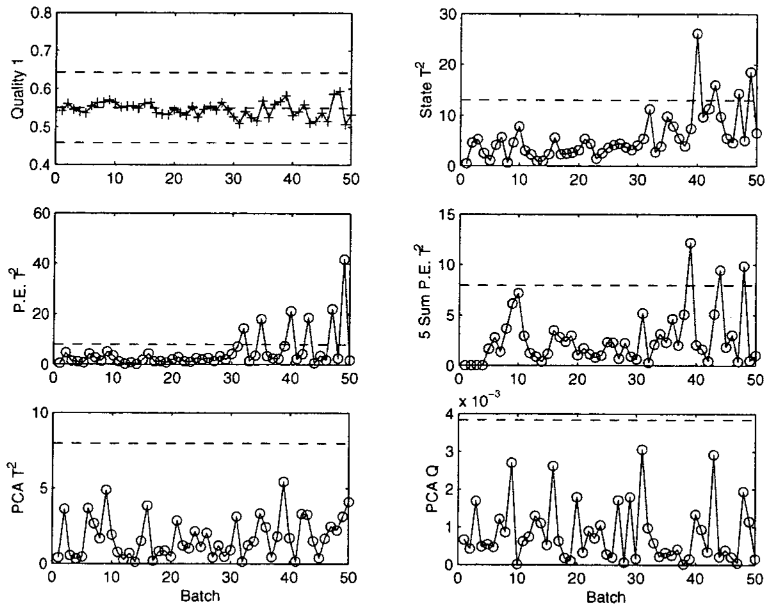

Figure 9. Monitoring performance the state-space model and the conventional PCA model during a change from strong to weakly correlated behavior.

type drifts in the process. This indicates that, by monitoring the innovation term, even small mean shifts (relative to the variance) can be detected even when normal operating data show strong correlations. It has been indicated that detecting a small mean shift in the presence of strong autocorrelation in the data can be very difficult (Harris and Ross, 1991). Even though this is a fundamental problem, we can see that the sensitivity is indeed enhanced by the whitening of the autocorrelated data that the state-space modeling performs implicitly.

Figure 12 shows the case when an isolated upset for a single batch occurs. Notice that in the strongly correlated case the prediction error statistic shows two large signals, the first one when the event occurs and the next one when returning back within the normal range.

On-line with delayed quality measurements

Available quality measurements of pulp kappa and yield can be easily included in the model for purposes of monitoring as well as inferential product quality prediction as discussed earlier. For this case, modeling data were generated by assuming the process was strongly correlated ($\phi = 0.9$). PCA was

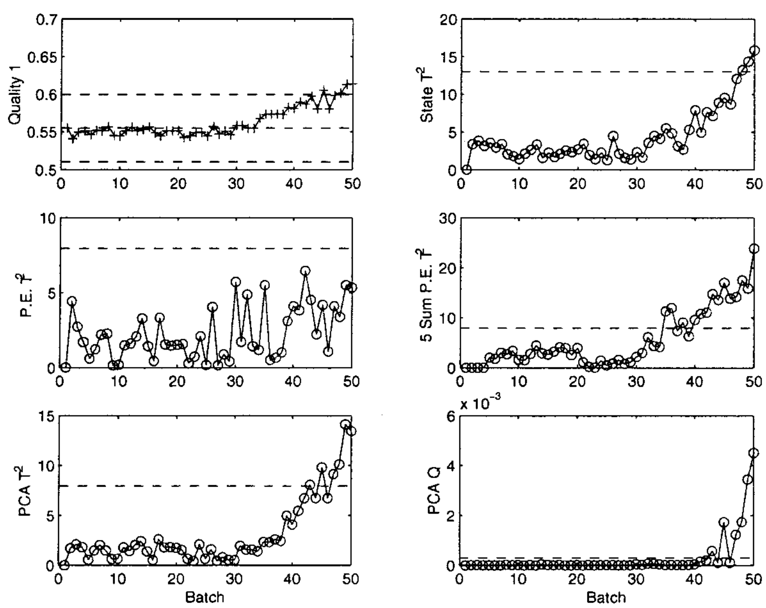

Figure 10. Monitoring performance of the state-space model and the conventional PCA model during a feedstock drift.

This case assumes that strong batch-to-batch correlation exists during normal historical operations.

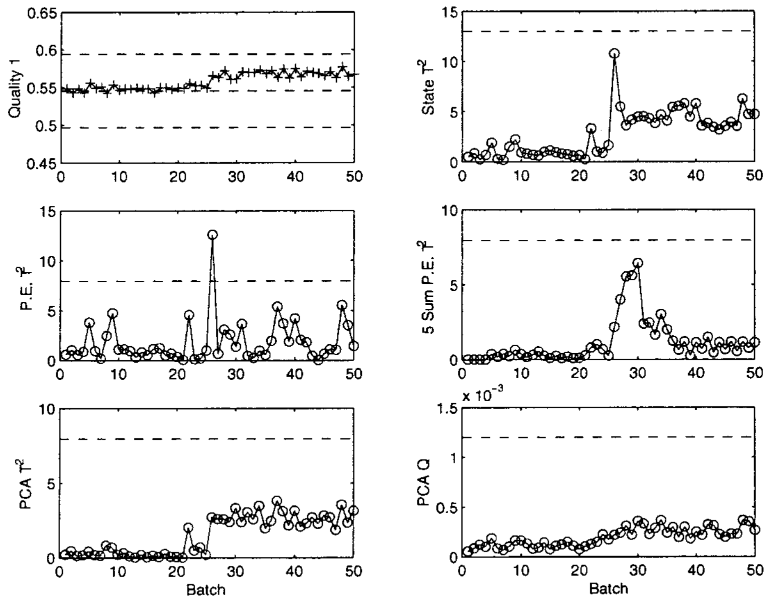

Figure 11. Monitoring performance of the state-space model and the conventional PCA model during a mean shift in the feedstock variables.

This case assumes that strong batch-to-batch correlation exists during normal historical operations.

applied to the lifted on-line measurements $\mathcal{Y}^f$ only to obtain the reduced vector $\underline{\mathcal{Y}}$ upon which the quality variables were appended before identification of the model.

The test data were generated by creating a drift away from the normal condition over 50 batches. Three different inferential prediction scenarios were considered, based on assumptions of the available measurements for updating the model. The first scenario only assumes that on-line measurements are available for model update; the second scenario assumes that only delayed laboratory measurements are available; and the final scenario considers that both pieces of information are available for model update. We note that the same model was used in all three situations and they differed only in the measurements that were entered into the Kalman filter. By considering these different types of model update, some insight can be gained into the importance of using all available sources of information for providing model correction and inferential quality prediction. It will also give some indication of potential model performance in the absence of measurements attributed to a sensor failure or a lab analysis failure.

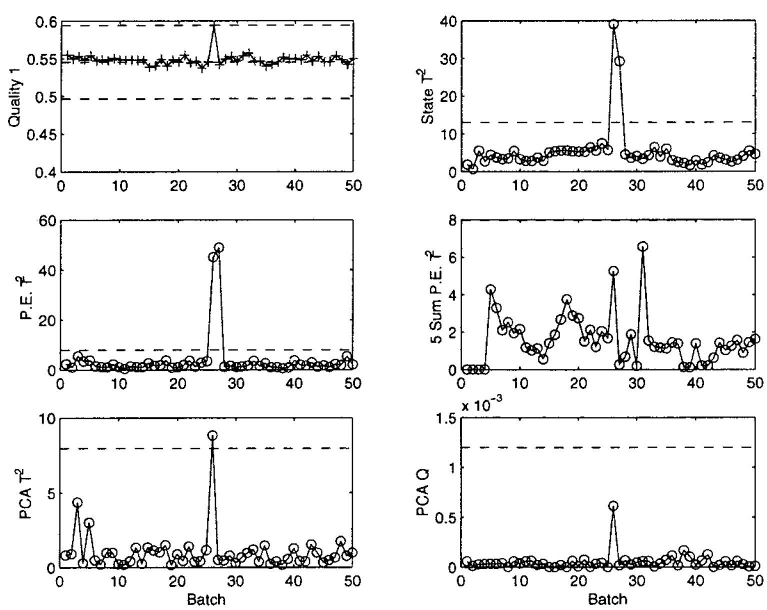

Figure 12. Monitoring performance of the state-space model and the conventional PCA model during an abrupt change in the feed that lasts for a single batch.

This case assumes that strong batch-to-batch correlation exists during normal historical operations.

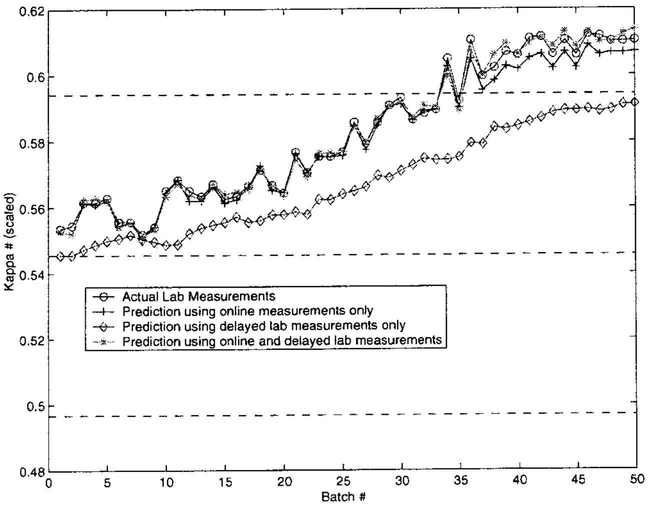

Figure 13. Inferential prediction of the kappa number during a drift in the feedstock quality.

The kappa prediction results for the three updated cases are shown in Figure 13. Using just delayed quality information gives the ability to follow the overall trend of the drift in the process, although the predictions are considerably lagged and biased. Using on-line measurements alone gives good prediction performance in most cases; however, it can be seen from the figure that performance begins to suffer when the process moves out of the normal operating range defined by the model building set. Using the delayed quality information together with on-line measurements offers the ability to correct for this model bias and achieves better prediction performance than just using on-line measurements alone. The inclusion of the delayed quality variables gives the most flexibility in terms of providing a single framework for monitoring along with inferential prediction and control of the product quality.

Summary and Conclusions

In this article, a monitoring framework for batch processes was proposed based on state-space models identified directly from historical operation data through subspace identification. By modeling the batch-to-batch dynamic trends of the variables and coupling it with some additional tests built around the prediction error, the new framework was shown to offer enhanced sensitivity for early detection of slow drifts, mean shifts, and changes in the batch-to-batch correlation structure. The state-space framework rendered us several possible formulations for the monitoring procedure including a formulation based on off-line quality measurements, a formulation based on on-line process measurements, and an integrated formulation that allows for the use of both types of measurements. It nicely complements the existing MSPC methods like the multi-way PCA-based monitoring. Simulation results from a pulp digester using the on-line formulations indicate that the proposed monitoring framework indeed provides increased sensitivities to these types of changes and also provides increased flexibilities to integrate prediction and control with monitoring.

Acknowledgment

J.H.L. gratefully acknowledges the financial support from the National Science Foundation (CTS-0096326).

Literature Cited

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (1984).

Datta, A. K., "Advanced Model Based Control of Pulp Digesters," Ph.D. Thesis, Auburn University, AL (1996).

Dorsey, A., and J. Lee, "Building Inferential Prediction Models of Batch Processes Using Subspace Identification," *J. Process Control*, **13**, 397 (2003).

Harris, T. J., and W. H. Ross, "Statistical Process Control Procedures for Correlated Observations," *Can. J. Chem. Eng.*, **69**, 48 (1991).

Lee, J. H., M. S. Gelormino, and M. Morari, "Model Predictive Control of Multi-rate Sampled Data Systems," *Int. J. Control*, **55**, 865 (1992).

Negiz, A., and A. Cinar, "Statistical Monitoring of Multivariate Dynamic Processes with State-Space Models," *AIChE J.*, **43**(8), 2002 (1997).

Nomikos, P., and J. F. MacGregor, "Monitoring of Batch Processes Using Multi-Way Principal Component Analysis," *AIChE J.*, 40, 1361 (1994).

Nomikos, P., and J. F. MacGregor, "Multi-Way Partial Least Squares in Monitoring Batch Processes," *Chemom. Intell. Lab. Sys.*, **30**, 97 (1995a).

Nomikos, P., and J. F. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, **37**(1), 41 (1995b).

Van Overschee, P., and B. De Moor, "Subspace Algorithms for the Stochastic Identification Problem," *Automatica*, **29**(3), 649 (1993).

Van Overschee, P., and B. De Moor, "N4SID: Subspace Algorithms for the Identification of Combined Deterministic Stochastic Systems," *Automatica*, **30**(1), 75 (1994).

Manuscript received Jan. 7, 2002, revision received July 15, 2003, and final revision received Aug. 11, 2003.