Multivariate Monitoring of Batch Processes using Batch-to-Batch Information

Jesus Flores-Cerrillo and John F. MacGregor
Dept. of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4L7

DOI 10.1002/aic.10147
Published online in Wiley InterScience (www.interscience.wiley.com).

Multiway principal component analysis (MPCA) and multiway partial-least squares (MPLS) are well-established methods for the analysis of historical data from batch processes, and for monitoring the progress of new batches. Direct measurements made on prior batches can also be incorporated into the analysis by monitoring with multiblock methods. An extension of the multiblock MPCA/MPLS approach is introduced to explicitly incorporate batch-to-batch trajectory information summarized by the scores of previous batches, while keeping all the advantages and monitoring statistics of the traditional MPCA/MPLS. However, it is shown that the advantages of using information on prior batches for analysis and monitoring are often small. Its main advantage is that it can be useful for detecting problems when monitoring new batches in the early stages of their operation., the approach and benefits are illustrated with condensation polymerization and emulsion polymerization systems, as examples. © 2004 American Institute of Chemical Engineers *AIChE J,* 50: 1219–1228, 2004
Keywords: Batch monitoring, partial-least squares, principal component analysis, statistical process control, on-line monitoring

Introduction

Batch and semibatch processes are the main production devices in pharmaceuticals, paint-coating, adhesive, and in most of the specialty-chemical industries. However, such processes suffer from variations in initial charge, operating conditions, and impurity concentrations in raw materials. To allow for more consistent operation and final product quality, control and monitoring schemes have been implemented around such processes. The use of multivariate statistical process control methods based on multiway principal component analysis (MPCA), and multiway partial-least squares (MPLS), and their associated monitoring statistics have been shown to be successful with industrial data for both the analysis of completed batches, and for the on-line monitoring of new batches (MacGregor and Nomikos, 1992; Nomikos and MacGregor, 1994, 1995a,b; Kourti and MacGregor, 1995; MacGregor and Kourti, 1995; Kourti et al., 1995, 1996). The success of such methods is largely because of the fact that the type of normal operating data necessary for building the required model is always readily available, and that the statistical control charts used for analysis and monitoring are easily developed from these data. One of the main characteristics of such methods is that the evolving within-batch measurement trajectories are projected into low dimensional latent variable spaces that summarize all the relevant information, and allow monitoring charts to be built in the reduced spaces where visually inspection and interpretation are easier.

Multiway PCA and PLS for the analysis, monitoring, and prediction of final product quality in batch processes were first introduced by Nomikos and MacGregor (1992, 1994, 1995a,b). They illustrated the detection of abnormal batches with several criteria, such as the Q statistic (also known as square prediction error (SPE) or distance to the model in the **X** space (DMODX)), the instantaneous SPE (for on-line monitoring) and the PCA or PLS score plots or equivalently Hotelling's T^2 statistic. Kourti et al. (1995, 1996) used multiblock methods (MBPCA/MBPLS) to incorporate different initial conditions,

Correspondence concerning this article should be addressed to J. MacGregor at macgreg@mcmaster.ca.

modes of operation, and prior processing conditions into the analysis and monitoring of batch processes. These multiblock methods allow one to incorporate any measured variable from prior batches into the analysis and monitoring schemes. In this article, we propose a variation of these multiblock methods to incorporate the PCA or PLS scores from prior batches.

Recently, Dorsey and Lee (2001) proposed a monitoring framework based on state-space models that purported to consider batch-to-batch variability more explicitly than the MPCA and MPLS methods. The approach first uses MPCA to extract the time and covariance structure of the data within batches, and then uses subspace identification to obtain a state-space model for these principal components. However, the proposed methodology is useful mainly for detecting batch-to-batch variations only in the score space obtained from the MPCA model, and has limited potential for detecting changes within batches. By modeling only the score space (obtained by MPCA), the identified state-space model accounts *only* for the batch-to-batch variation in the score space of the normal (in-control) runs. Because the PCA score space is completely orthogonal to the SPE space of the MPCA model, then any charts based on the model states or innovations will *not* be able to detect faults that are detected primarily in the SPE space by the MSPC methods (for example, faults that introduce totally new latent variables or PC's as opposed to those that simply induce larger variations in the existing PC's). Faults that affect both the score and SPE space of the MPCA model might be detected if their effect on the score space is strong enough. The shortcoming of their proposed monitoring methodology results from the fact that all information on the space orthogonal to the PCA score space (that is, the SPE space) is lost. Furthermore, subspace methods for system identification usually require many more training batches to build the state space models that is normally required to establish multivariate PCA/PLS models.

This article introduces a modified MPCA/MPLS procedure that, besides retaining all the advantages of the MPCA/MPLS methods for batch analysis and on-line monitoring, also enables one to incorporate a summary of prior batch trajectories and performance. The approach and potential benefits to be gained from it are illustrated on simulations of two-batch polymerizations processes: the condensation polymerization of nylon and the emulsion polymerization of styrene.

MPCA and MPLS Monitoring Using Batch-to-Batch Information

Preliminaries

An important question concerns "when data from prior batches would be useful for the analysis and monitoring of future batch processes?".Clearly, a minimum requirement is that they contain some information on effects that will have an influence on the performance of the future batches. This implies that there must exist some autocorrelation in important performance variables from batch-to-batch. Such a situation would arise if a common source of raw materials is being used for successive batches, and the materials from this source have some characteristics (for example, impurity concentrations, surface chemistry properties, and so on) which change slowly with time. Then one could expect the performance of future batches to be related to that of recent past batches. If no such batch-to-batch carry-over effect because of common disturbances is present, then the value of incorporating prior batch information into MPCA or MPLS analysis or monitoring schemes would be negligible.

Background

MPCA and MPLS perform PCA/PLS on the three-way matrices arising from batch data (batches (k) $\times$variables (v) $\times$ time (θ)) by unfolding and rearranging them into two-dimensional (2-D) arrays as described in Nomikos and MacGregor (1994, 1995a,b). To remove most of the possible nonlinearities of the batch trajectories, mean centering around the average trajectories and unit variance scaling is performed. Selection of the number of principal components is often performed by cross-validation methods (Wold, 1978; Krzanowski, 1987). Literature describing concepts, fundaments and algorithms on PCA and PLS is vast and, therefore, not addressed here (Geladi and Kowalski, 1986; Jollife 1986; Wold et al., 1987; Höskuldsson, 1988; Geladi, 1988; Geladi, 1989, Jackson, 1991; Kourti and MacGregor, 1995).

Incorporation of batch-to-batch information into MPCA/MPLS

To capture information in prior batches, one could directly use the final quality measurement matrix (**Y**) taken on the product from previous batches. These data can be easily incorporated into a matrix **Z,** and then existing multiblock approaches used for the analysis and monitoring of batch processes (Kourti et al., 1995). However, this may not always be possible. The product quality data for the last batch ($\mathbf{y}^{(k-1)}$) may not be available from the quality control laboratory before the next batch is started. Furthermore, all the product quality data are often not measured for every batch, and even those that are measured may have large measurement error, and may not be sufficient to capture all the relevant information from past batches. Therefore, in this article we propose to use the final PCA or PLS scores values (t_1,t_2. . .t_a) from prior batches to summarize all the variables and their time histories of the prior batches. The idea is then to include these summarizing scores from prior batches into the matrix **Z** for the current batch (k). A multiblock MPCA or MPLS (MBPCA/MBPLS) is then used to combine this batch-to-batch information (**Z** matrix) with the trajectory data (**X**) for the current batches. The structure of the resulting data matrices is illustrated in Figure 1. Each row, $\mathbf{x_k^T}$, of the **X** matrix consists of measurements on all the process variables at all time intervals for the kth batch, and the corresponding row, $\mathbf{z_k^T}$ of the **Z** matrix contains score values for each of the past r batches, that is

$$\mathbf{z_k^T} = [t_1^{(k-1)}, t_1^{(k-2)} \cdots, t_1^{(k-r)}; t_2^{(k-1)}, t_2^{(k-2)} \cdots, t_2^{(k-r)}; \cdots; t_a^{(k-1)}, t_a^{(k-2)} \cdots, t_a^{(k-r)}] \quad (1)$$

Once the multiblock model has been built, this lagging of the prior batch scores in the **Z** matrix poses no problem, and the analysis and monitoring is easily accomplished with existing multiblock MPCA/MPLS approaches (Kourti et al., 1995). However, a problem arises at the model building stage, when to build the MBPCA or MBPLS model, one needs complete data for both the **Z** and **X** matrices for all batches in the training data. However, the score values from prior batches needed to

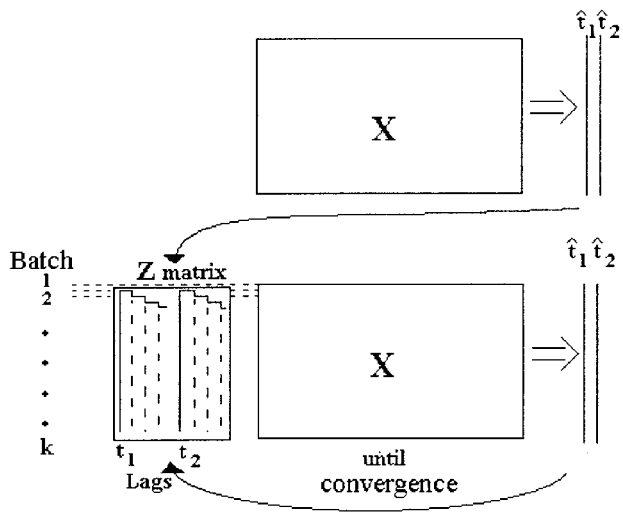

Figure 1. MPCA monitoring scheme with batch-to-batch information.

form the **Z** matrix will not be available until after the model has been built. Faced with this dilemma, the following iterative training approach to building the model is proposed.

(1) An initial MPCA or MPLS is carried out on the **X** matrix only to obtain the scores for each batch belonging to the training data set. The scores from this model are then used to provide initial estimates ($\hat{\mathbf{t}}$) of the scores for prior batches to include in the **Z** matrix that will summarize all relevant prior batch-to-batch information, as illustrated in Figure 1.

(2) The **Z** matrix is then weighted, if desired, relative to the **X** matrix (for example, block scaling), and an augmented [**Z X**] matrix is used to perform another PCA or PLS, to obtain a complete model incorporating both the prior batch information and the current batch data. (If one does not wish to discard the first r batches, then a missing data algorithm (Nelson et al., 1996) can be used to account for the unknown prior batch information arising from the lagging of the first r batches.)

(3) Repeat (2) until convergence of the scores (**T**) is achieved (Figure 1), $(\mathbf{Z}_i - \mathbf{Z}_{(i-1)}) = \mathbf{\Omega}$; $\sum_{j=1}^{a\times r} \sum_{m=1}^{k} \Omega_{m,j}^2(k, a \times r) \leq \varepsilon$, where i is the iteration number. In the examples considered in this work, convergence of the scores was achieved with direct substitution in a few iterations.

(4) At convergence, the final models, that can be used for monitoring, analysis, and prediction are given by:

For PCA

$$[\mathbf{Z}\ \mathbf{X}] = \mathbf{\Gamma}\mathbf{V}^{\mathbf{T}} + \mathbf{E}^* \tag{2}$$

For linear PLS

$$[\mathbf{Z}\ \mathbf{X}] = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^{\mathbf{T}} + \mathbf{F} \tag{3}$$

where **V**, **P** and **Q** are loading matrices determined at the model building stage, and **E***, **E** and **F** error matrices.

Each low dimensional ($1\times a$) row of the ($n \times a$) score matrices $\mathbf{\Gamma}$ or **T** ($\mathbf{\Gamma}$=**[ZX]V** and T=**[ZX]W***) provides all the statistically significant information on the relationships among the prior batch histories (rows of **Z**), and the time histories of all the variables in the current batch (rows of **X**). Separate loadings and scores for the **Z** and **X** matrices in a multiblock scheme then can be obtained directly from the above MPCA or MPLS models (Westerhuis et al., 1998).

The iterative scheme is only necessary for model building. Once the model is identified, the iterative scheme is not needed and monitoring, analysis, and prediction can be performed in the same way as is done in normal MPCA and MPLS methods. For example, for each new batch, the history of prior batches (rows of **Z**), and the process trajectory (a new row, $\mathbf{x}_{new}^{\mathbf{T}}$, of **X**) are both available to calculate the score values $t_1, t_2 \ldots t_a$ for the current batch. Therefore, the use of the proposed approach (Figure 1) will allow one to efficiently incorporate prior batch information into a single model for both the analysis of completed batches and for the on-line monitoring of new ones. Moreover, all the benefits and statistical analysis tools of the conventional MPCA and MPLS methods will be retained. The approach can also be easily extended in cases that a nonlinear PLS method is needed (Wold et al., 1989; Frank, 1990; Wold, 1992; Berglund et al., 1997).

For analysis or monitoring, several different combinations of the matrices shown in Figure 1 can be used. To monitor only batch-to-batch changes, the **Z** matrix contains all the necessary prior batch information, which is summarized by the past scores. Multivariate control charts based on **Z** alone, would be sufficient for detecting changes in the autocorrelation structure of the batch-to-batch behavior, such as might arise from changing feedstocks of raw materials. With the **X** matrix alone would allow one to analyze and monitor only the within batch changes (conventional MPCA/MPLS). With both the **Z** and **X** matrices allows for complete treatment of both batch-to-batch and within batch changes within the same MBPCA/MBPLS model.

Model building: Selection of number of lags

After convergence of the lagged-scores (**Z**) has been achieved, the number of significant lags (number of prior batches, r, used in the **Z** matrix) can be determined by inspecting the values of the loadings for the lagged scores in **Z** ($\mathbf{p_z}$). If the loadings on all the scores in **Z** for all batches beyond a certain lag are small, then these lags can be dropped, and the model reiterated again until convergence. To illustrate this, an example is shown in Figures 2a and 2b (Case study 1 for the detection of change in correlation in the emulsion polymerization of styrene, for large positive batch-to-batch correlation (ϕ=0.9), and for small batch-to-batch correlation ((ϕ;=0.2), respectively.) In Figure 2a, it can be seen that only the two loadings $p_{z,1}^{(k-1)}$, $p_{z,2}^{(k-1)}$ (from the first principal component, $\mathbf{t_1}$) associated with the immediately preceding batch (k-1) are large, indicating that only one lag (and two scores (**T**), Eq. 1) need to be used in model building. Moreover, the method can also detect when the incorporation of previous batch information would be of little value, as shown in Figure 2b, where it can be seen that the values of $\mathbf{p_z}$ are small for the previous scores at all past lags.

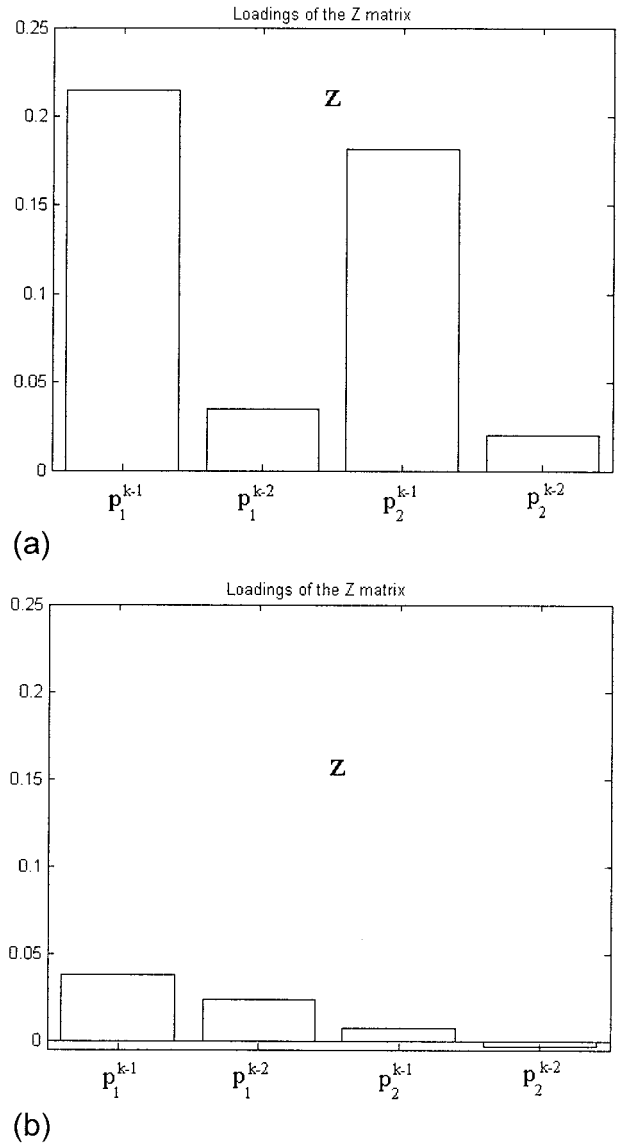

Figure 2. Selection of the number of significant lags for model building.

(a) large (ϕ=0.9), and (b) small (ϕ=0.2) degree of batch-to-batch correlation (Case 1, detection of change in correlation).

Off-line analysis and on-line monitoring studies

Systems. The approach is illustrated with two simulated polymerization systems, one is the condensation polymerization of nylon 6,6 and the other is the emulsion polymerization of styrene. It is important to notice that the usefulness of the previous batch information for on-line monitoring will be greater in early stages of the process, because in later stages most of the information is contained in the on-line measurements taken from the current batch (k). For many industrial processes, as those used here for illustration purposes, the main sources of disturbances occur at the start of the process because of initial charge conditions or impurity variations in the raw materials. Therefore, a monitoring scheme containing batch-to-batch information may be useful in providing a more consistent and faster detection of early disturbances (as long as these have a degree of batch-to-batch correlation). In what follows a brief description of the systems and conditions on which the monitoring studies were performed is presented.

Condensation Polymerization. The first process considered here is the batch condensation polymerization of nylon 6,6. A detailed theoretical model for this process was developed by Russell et al. (1998a), and is used in this work. Details about the model and model parameters are described in the original publication. The focus of this study will be on the initial polymerization stage of the batch, from the initialization of the process to the opening of the vent valve (around 35 min). This is because the decision when to open the vent valve is key to control the achievable product quality (Russell et al., 1998a.) However, the results of this initial polymerization stage are indicative of the complete batch because the main source of disturbance is the fluctuation of the feed water (W) (a single evaporator usually serves several reactors (Russell et al., 1998b)).

For on-line monitoring, measurements include reactor pressure, steam jacket pressure, reactor temperature, and vent flow rate. However, steam jacket and reactor pressure are considered here as manipulated variables that remain at their initial set points during the early phase, and so need not to be included in **X** (If variations exist in these variables, their trajectories should be included in the **X** matrix). Moreover, at the initial heating stage the vent readings are zero because the vent valve is close. Therefore, the available on-line information comes from the reactor temperature readings, T_r (used every 15s. in the PLS model). The end quality variables are amine end groups (NH_2) on the polymer molecules, and the number-average molecular weight (MW) at the end of the reaction (200min). The quality and process measurements are corrupted by normally distributed random error with magnitudes reported in Table 1.

Emulsion Polymerization. A nonlinear model, with simple kinetics, to simulate the styrene emulsion polymerization was developed by Lynch and Kiparissides (1981), and is used in this work. This model, originally developed for tubular reactors with full recycle, has been adapted for use in batch and semi-batch processes. For a complete description of the model and model parameters the reader is referred to the original publication. The focus of this study will be on the initial polymerization stage of the batch, from the initialization of the process up to 40 min. This is because the particle generation is of short duration (<40min), and early detection of abnormal conditions of the batch would allow one to take faster corrective action. Moreover, the results of this initial polymerization stage are indicative of the complete batch because once the particle generation is over there is almost no further change in the number of particles. The main source of variation considered in this study is variation in the surface chemistry properties of the emulsifier. Particularly, these surface chemistry variations affect the surface covering potential a_s of the emulsifier. This disturbance has a great effect on the number of micelles formed and, hence, on the number of polymer particles nucleated (Np),

Table 1. Measurement Noise for Condensation System

Measurements		σ%
X	Tr	0.1
Y	NH_2	0.1
	MW	0.3

Table 2. Measurement Noise for Emulsion System

Measurements		σ%
X	Tr	0.05
	T_i	0.05
Y	C	0.1
	N_p	0.1
	D_p	0.1

as well as on the resulting conversion. On-line reactor T_r and jacket temperature T_j measurements are considered to be available every minute. The end-quality variables are conversion C and number of particles (Np) at the end of the batch (480min). The quality and process measurements are corrupted by normally distributed random error with magnitudes reported in Table 2.

Data History Generation. For the condensation polymerization system, 200 batches were used to generate the normal data history used as a training set, whereas in the case of emulsion polymerization 150 batches were generated. However, adequate MPCA/MPLS models can be built with many fewer batches (Nomikos and MacGregor 1995a; Kourti et al., 1996.) For both systems, the normal batch history was generated by assuming that the disturbances d_k (W for condensation polymerization, and a_s for emulsion polymerization) vary from batch-to-batch in a correlated manner according to the autoregressive model

$$d_k = \phi d_{k-1} + \xi_k \tag{4}$$

where ξ represents normal distributed random error and ϕ represents the degree of batch-to-batch correlation. Two types of correlation are studied for each system: Strong (ϕ=0.8 for condensation and ϕ=0.9 for emulsion) and weak correlation (ϕ=0.2 for both systems). For each type of correlation, a training set was generated.

Case Studies

Several Case studies for the off-line analysis of completed batches and for the on-line monitoring of new batches were performed in both systems for low and high batch-to-batch correlation in the disturbances. Process faults or upsets were introduced including changes in the batch-to-batch correlation structure, slow drifts and short-lived upsets. For the sake of brevity, only some of these Case studies are presented, as shown in Table 3 for the high correlation structure where the benefits of inclusion of prior batch information in the model are more evident. In the Case studies, comparison between MPCA/MPLS with no information on prior batches and the proposed approach (MBPLS/MBPCA with prior batch information, **Z** matrix) is performed.

Table 3. Monitoring Case Studies*

Case Studies	Off-Line	On-Line
1. Change in correlation	Emulsion	—
2. Small drift	Condensation	Condensation
3. Short-lived upset	—	Emulsion

*For strong (ϕ = 0.9 and ϕ = 0.8) and weak (ϕ = 0.2) batch-to-batch correlation in disturbances.

Case Study 1: Detection of changes in correlation from batch-to-batch

The purpose of this section is twofold: (1) to show that a proper SPC monitoring method based on MPCA/MPLS should not alarm for changes in batch-to-batch correlation if these changes are not important to product quality (final product quality only depends on the magnitude of the disturbances not on their time order or equivalently their batch-to-batch correlation structure), and (2) to show that, if needed, changes in correlation can be easily detected with previous batch information.

In this example, a change in the degree of correlation, of the testing data set, from ϕ =0.9 to ϕ =0.2 is performed (Eq. 4) at batch 26 for the emulsion polymerization system (the variance of ξ was adjusted as describe in Dorsey et al. (2001) to keep the total variance of the raw material qualities at the same level).

In Figure 3a, batch-to-batch evolution of the final quality properties (C, Np) is shown for this example with a Hotelling's T_y^2 chart (Tracy et al., 1992). Figure 3b shows the corresponding batch-to-batch evolution of the SPE_x (Q_x statistic) obtained by performing normal MPLS on the batch data only (no inclusion of previous batch information) together with their 95 and 99% confidence limits (Nomikos and MacGregor, 1994, 1995a). Note that the SPE_x does not detect the change in the degree of correlation. However, this is what should happen because, as can be seen in Figure 3a, the end quality properties are clearly still in a state of statistical control. Any method that alarms in this situation (Dorsey and Lee, 2001) is misleading as a monitoring method for the health of the batches. We stress this point because a recent publication (Dorsey and Lee, 2001) has used the lack of detection of this change in batch-to-batch correlation structure by MPCA/MPLS as a negative result for these methods. However, as shown earlier, it is clearly a positive feature if the objective is to monitor the health of the batches.

Changes in the disturbance correlation structure, although usually not important for product quality can be important, and should be monitored, if the disturbance correlation structure is being used somehow in a batch-to-batch control algorithm (Chin et al., 2000). In that case the manipulated variables set points are being changed for new batches based on assuming a previously identified disturbance autocorrelation structure. If this autocorrelation structure were to change suddenly, the batch-to-batch control scheme based on the assumed value should lead to poor results and degrade the quality. In this situation, changes in the batch-to-batch correlation can be detected simply by performing PCA on the **Z** matrix of previous batch scores (Figure 1) and monitoring their SPE_z (Q_z statistic) and Hotelling's T_z^2. In this case, it may be necessary to include enough lags (r) to allow for adequate modeling of the batch-to-batch structure. The detection of a change in the correlation structure of the emulsion process at batch 26 from ϕ=0.9 to ϕ=0.2 when the number of lags r=10, is shown in Figure 4.

Another simple alternative to detect changes in the degree of correlation is by fitting a time series model (for example AR(1)), to the scores **t** obtained from the normal MPLS model on [**X Y**]. The estimated value of $\hat{\varphi}$ would then represent the degree of correlation. Change in the correlation structure of

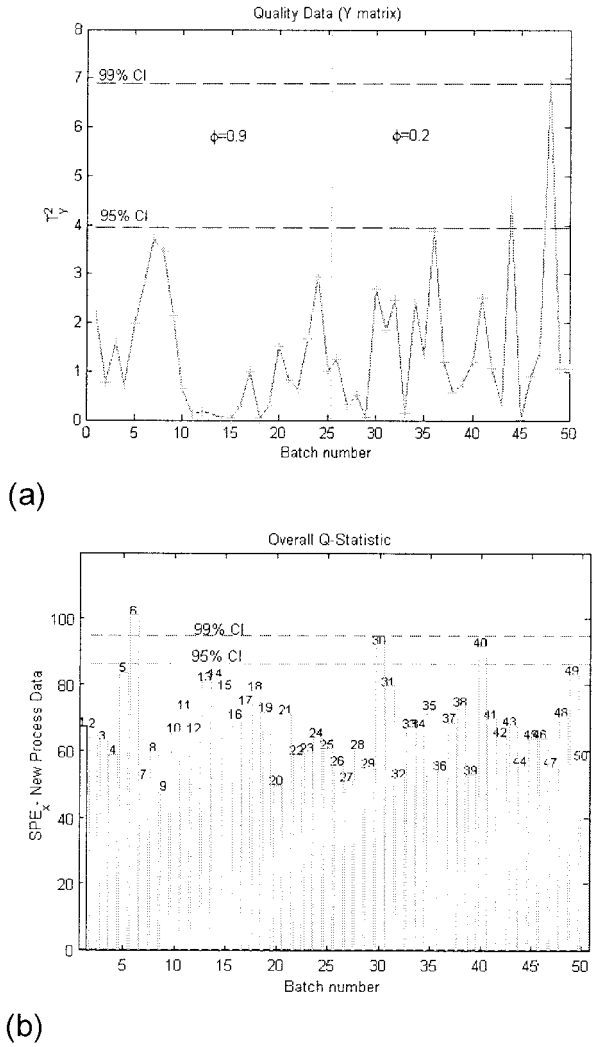

Figure 3. Off-line batch-to-batch monitoring of the emulsion polymerization system for a change in the autocorrelation structure of the disturbances at batch 26.

(a) Hotelling's T^2 for the final quality data (C, Np), (b) SPE_x (Q_x) on **X** data using normal MPLS.

new incoming batches can be monitored with a simple Shewhart chart on the residuals (ξ) of such time series model

$$\xi^k = \mathbf{t}^k - \hat{\varphi}\mathbf{t}^{k-1} \tag{5}$$

and confidence limits established based on the variance of the residuals. This is illustrated in Figure 5 for the case in which the first principal component t_1, of the normal MPLS is modeled.

Case Study 2: Detection of a slow drift in the water content of the salt (W) over many batches

Off-line Analysis. In this example, the objective is to detect a slow drift of the product quality for the condensation polymerization process when there is a slow drift in the initial water content (W) of the incoming batch, as well as a common cause

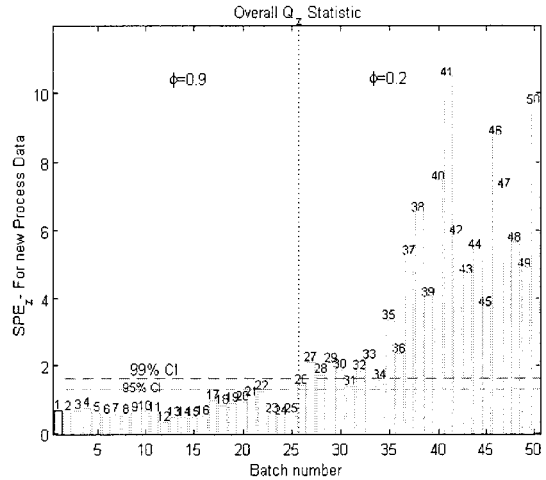

Figure 4. SPE_z for detection a change in correlation from ϕ=0.9 toϕ=0.2 (change at batch 26) using r=10 lags.

batch-to-batch correlation of ϕ=0.8 in W. It is expected that, with batch-to-batch information, faster detection of the drift will be achieved because, as can be seen in Figure 6a, W is progressively getting worse over many batches (drift begins at batch 26). In Figure 6b, a Hotelling's T^2 chart for monitoring the batch-to-batch evolution of the end-quality properties (NH_2, MW) is shown for the 50 batches, whereas in Figure 6c and 6d the t_1 score obtained from normal MPLS and MBPLS (with both **X** and **Z**), is shown, respectively. In these Figures, a statistically significant out of control region (at α=0.01 significance level) is detected only at batch number 50. Note that including information (**Z**) on previous batches did not improve the time to detection (Figure 6d). This result shows that the use of such prior batch information is of limited value once one has complete information on the current batch.

A faster detection of this type of events would be beneficial because would allow one to correct for the disturbance before the quality properties are off-specifications. The simplest way to detect small drifts, trends or mean shifts is by the use of a

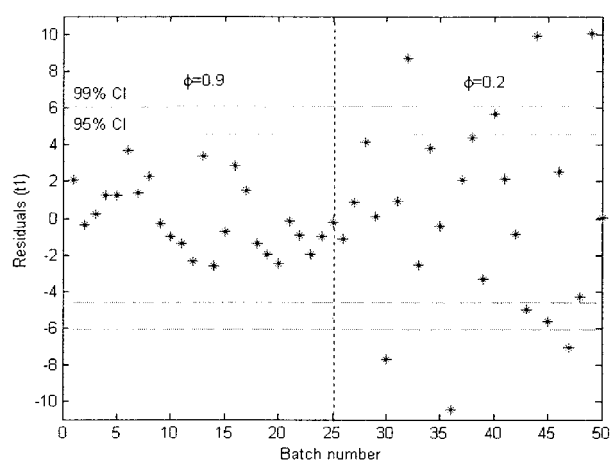

Figure 5. Monitoring of residuals (ξ) for t_1 from normal MPLS.

Change in correlation from ϕ=0.9 to ϕ=0.2.

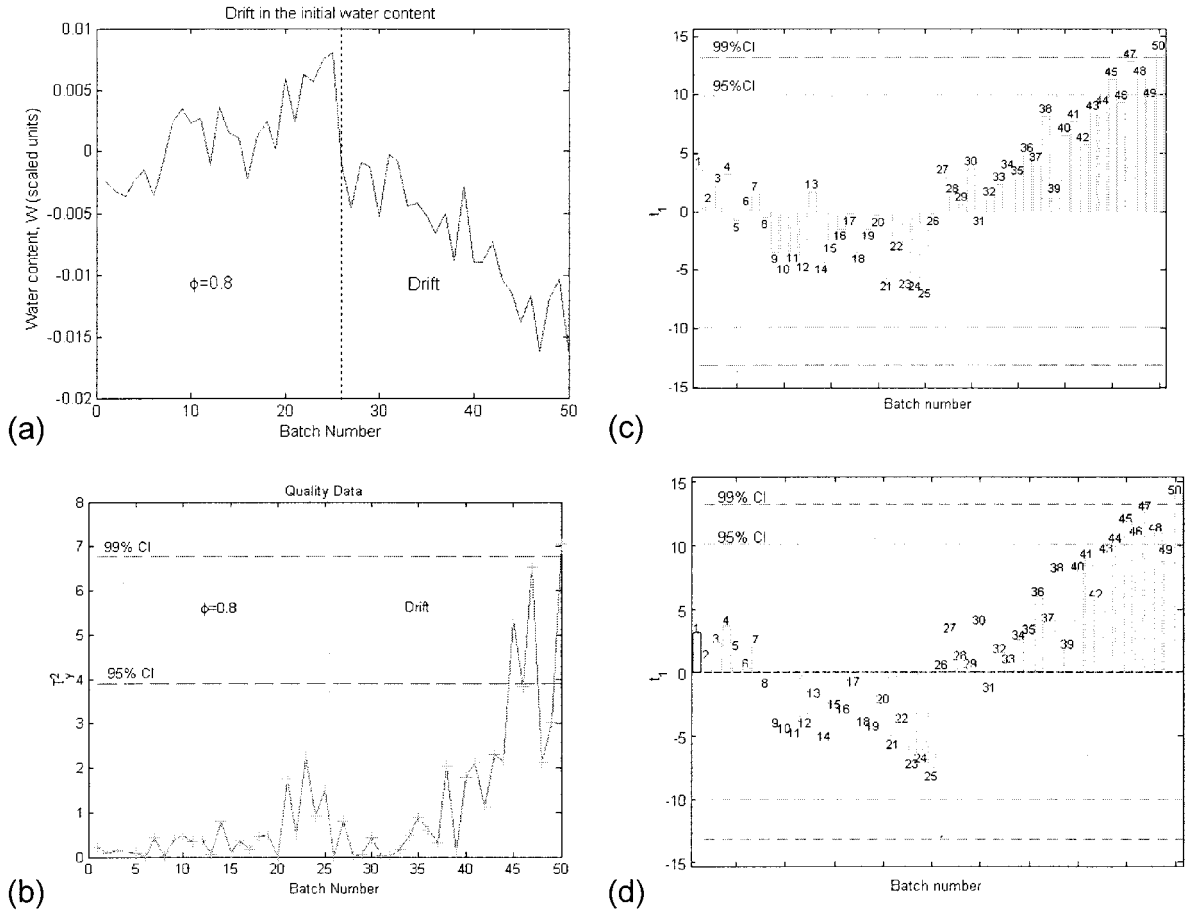

Figure 6. (a) Disturbance trend for water content (W); (b) Hotelling's T^2 on the batch final quality data (Y); (c) t_1 score from MPLS; (d) t_1 score from MBPLS.

cumulative Hotelling's T^2 (ψ) obtained from either multiblock or normal MPLS. Two alternatives are shown: Crosier (1988) proposed computing Hotelling's T^2 at each observation, and then computing the cumulative of the scalar distance T^2 as

$$\psi_i = \max\{0, \psi_{i-1} + T_i - \eta\} \tag{6}$$

This multivariate CUSUM scheme signals an out-of control situation when $\psi_i > h$. The limit h and the parameter η where chosen as suggested by McNeese et al., (1991): $\eta = \sigma/2$ and $h = 4.5\sigma$. Alternatively a finite-horizon cumulative Hotelling's T^2 (ψ_i) may be used (Dorsey and Lee, 2001)

$$\psi_i = \sum_{i=k-r+1}^{k} T_i^2 \tag{7}$$

where r is the number of previous batches over which the summation is taken. Figure 7a shows the CUSUM on Hotelling's T^2 (ψ_i) obtained from Eq. 6, whereas in 7b ψ_i obtained from Eq. 7 with $r=10$ for normal MPLS. It can be seen that, in both approaches, the small drift can be detected around batch 47, slightly earlier than with the direct use of hostelling T^2 (Figure 6c,d).

In this section, we have considered the use of previous batch information in the detection of changes in batch-to-batch correlation, and for the detection of slow drifts for case in which the batches have already been completed (off-line analysis). In the next section, the benefits of the incorporation of prior batch information in the on-line monitoring of new batches are investigated.

On-line Monitoring. In on-line monitoring, at any time during the batch, future unknown measurements need to be estimated. Nomikos and MacGregor (1994, 1995a) proposed three alternatives to estimate such measurements, and found that the approach that uses a missing data algorithm was generally superior and so is used here.

In this example, the slow batch-to-batch drift in the initial water content (W) together with a common-cause batch-to-batch correlation (ϕ=0.8), as shown in Figure 6 is also used to illustrate the effect that the inclusion of previous batch information (**Z** matrix, Figure 1) has on the on-line monitoring of new batches. The on-line t_1 score plots from the MBPLS with prior batch information is shown in Figure 8a for batches 1 (normal operation condition), 47 (out of 95% CI), and 50 (out of 99% CI). The on-line t_1 score plot obtained from normal MPLS without use of prior batch information for the same batches is shown in Figure 8b (note the difference in scales from both figures).

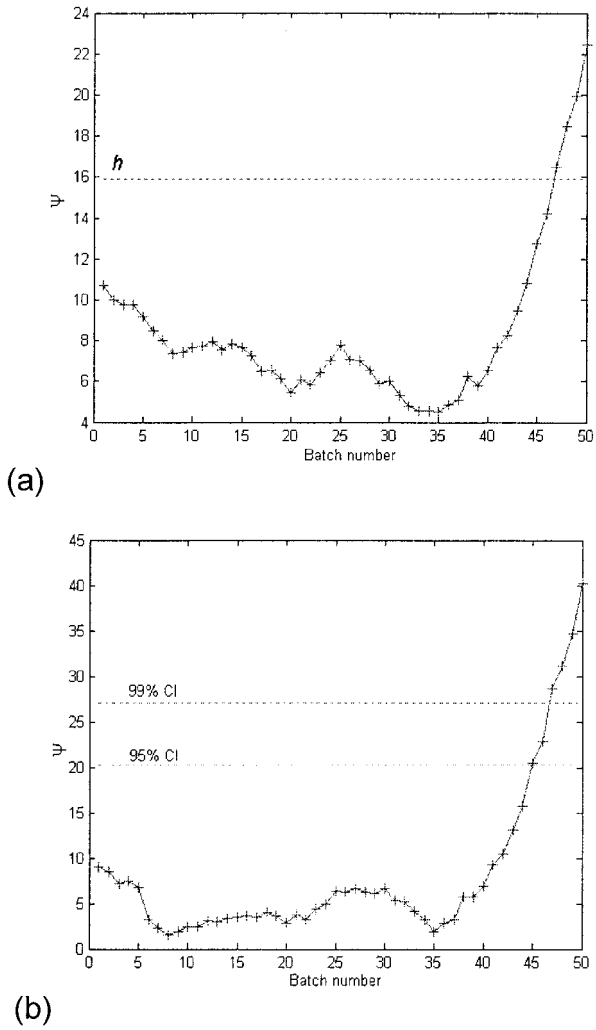

Figure 7. Detection of drift with cumulative Hotelling's T^2 (Ψ).

(a) Eq. 6, and (b) Eq. 7 with $r=10$ for normal MPLS.

From this Figure it is evident that with the incorporation of batch-to-batch information in the on-line monitoring scheme, smaller and more consistent confidence intervals are obtained. Detection of the abnormal situation caused by the slow drift in the inlet water concentration is also achieved faster. As shown in Figure 8a, by incorporating information on previous batches, the starting $\mathbf{t_1}$ score values for each successive batch slowly rises, and by batch 47 to 50 it is evident that a fault is present almost from the first sample point. Without the information on prior batches to confirm this gradual batch-to-batch trend, the detection of the problem takes longer as seen in Figure 8b.

The importance of the previous batch information (**Z** matrix) can be scaled-up or scaled-down, allowing for increased sensitivity on batch-to-batch abnormalities. The effect of scaling-up the **Z** matrix is illustrated in Figure 9 for the same case as discussed in Figures 6, 7 and 8 (slow drift upset in the condensation polymerization system.) To allow for a direct comparison, normalization of the scores and confidence limits (CI) has been performed

$$t_1^* = \frac{t_1}{99\%\ \mathrm{CI}}; \quad \mathrm{CI}^* = \frac{\mathrm{CI}}{99\%\ \mathrm{CI}}$$

Three cases are shown: (i) normal MPLS with no prior batch information (*); (ii) MBPLS including prior batch information, **Z** (+) (unit variance scaling in the **Z** matrix, scaling factor (SF=1)) and (iii) MBPLS when the **Z** matrix is up-weighted by a factor of 2 (SF=2). Figure 9 also more clearly demonstrates the use of prior information (**Z**) in on-line monitoring. Note that, by the end of the batch, all methods end up giving approximately the same normalized score values and indicating an out of control situation. This is consistent with what was shown in the previous section on off-line analysis. However, with the use of prior batch information (**Z**) this slow batch-to-batch drift in the water content of the feed is detected much earlier in the batch. The prior batch information (**Z**) essentially gives the MBPLS a head start by giving initial score values

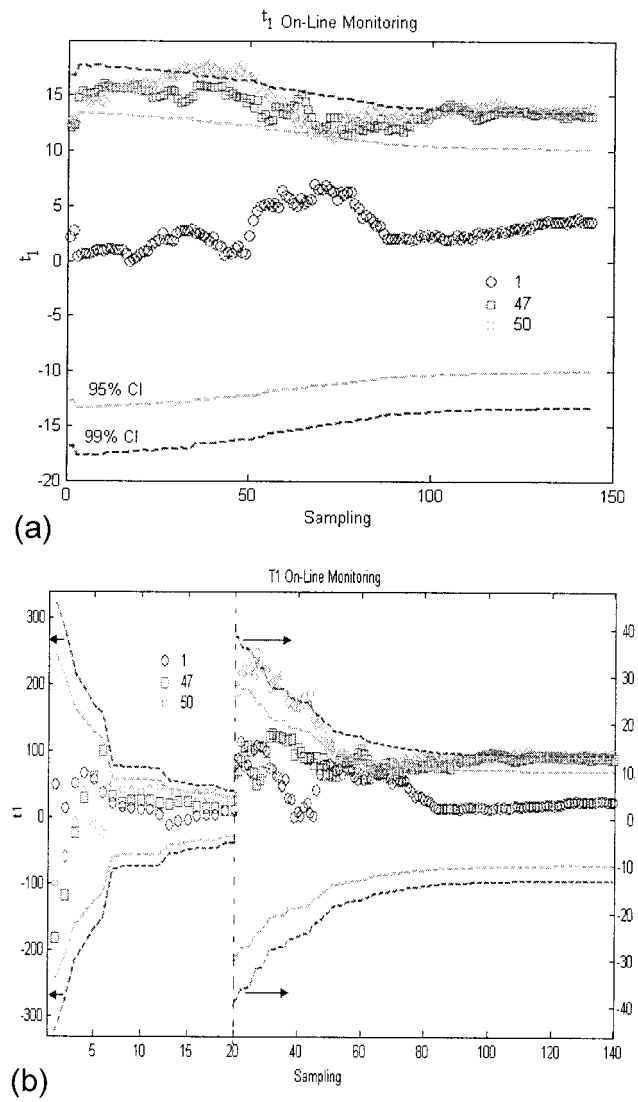

Figure 8. t_1 on-line monitoring for detection of drift with ϕ=0.8.

(a) MBPLS with prior batch information; (b) normal MPLS.

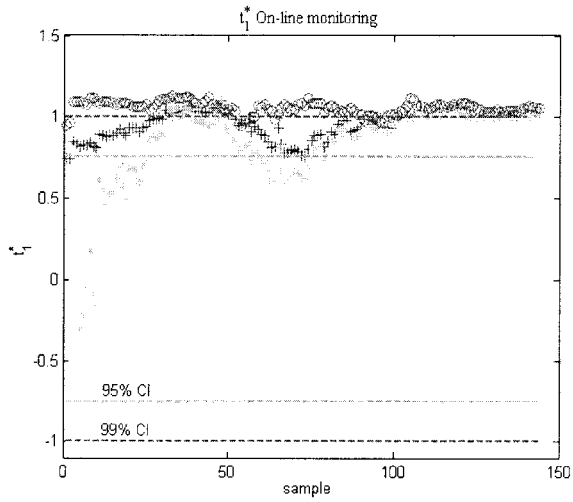

Figure 9. Normalized on-line score plots and the effect of scaling-up the Z matrix.

Normalized t_1^* plot for batch 50 (ϕ=0.8). (*) normal MPLS, (+) MBPLS (unit variance scaling in the **Z** matrix, scaling factor (SF)=1), and (o) MPLS with SF=2 in the **Z** matrix.

close to the final score values of the previous batch, whereas the regular MPLS approach has to start again from scratch and learn from the early data from each new batch.

Use of Contribution Plots in Assessing Past Batch Information. When using previous batch information (**Z** matrix) in MBPLS/MBPCA for on-line monitoring of new batches, once an out of control signal is detected, contribution plots (MacGregor et al., (1994); Kourti and MacGregor (1995)) can be inspected to determine to what extent previous batches ($t_1 \ldots t_r$ scores in the **Z** matrix), and on-line measurements within the current batch (**x** data) are contributing to such a signal. The contribution plot for the $\boldsymbol{t_1}$ score plot in Figure 9 at time interval 4 (99% control limit violated) for batch 50 (SF=2 in **Z** matrix) is shown in Figure 10. The contribution plot clearly shows that it is the variation in prior batch scores that

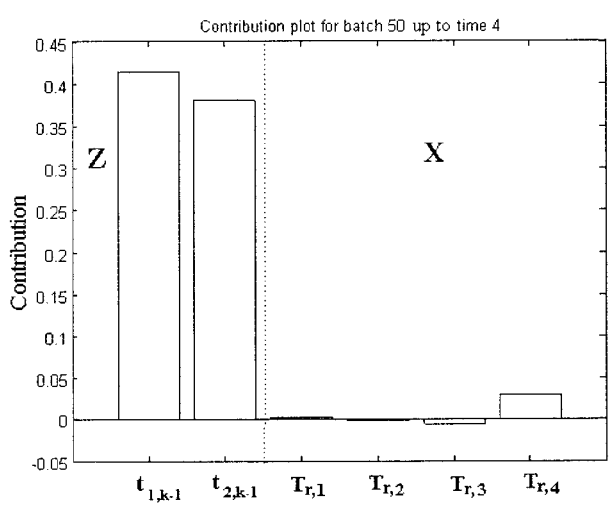

Figure 10. Contribution plot to component 1 up to time 4.

Abnormal batch for slow drift case (batch 50).

Table 4. Monte Carlo Simulation Results for Short-Lived Upset. (The Numbers Indicate the Average Time to Detection of the Fault)

	MBPCA		MPCA	
CI	95%	99%	95%	99%
Average	13.8	23.24	9	18

are contributing to the alarm and not the temperature measurements T_r in the early part of the current batch.

Case Study 3: Detection of a Short-Lived Upset in a Single Batch

In this case, the objective is to show that, for a short-lived upset in the current batch (affecting the end-quality properties of that batch), the use of batch-to-batch information will not add any benefit to an on-line monitoring scheme or off-line analysis. A Monte Carlo simulation study was performed, in which 50 different testing datasets with correlationϕ=0.9 were generated for the emulsion polymerization system. In every new testing dataset a short-lived upset (arising from surfactant variations affecting a_s) was introduced at batch 40. Results from the Monte Carlo simulation are shown in Table 4. The average times to detection of the upset indicated by an out of control signal in the t_1 score (at α=0.05 and 0.01 probability limits) are seen to be even a little slower by including batch-to-batch information than when this is not included. More details are shown in Flores-Cerrillo (2003) including the use of prior batch information to improve the inferential prediction of end quality properties.

Summary and Conclusions

The use of information from previous batches has often been of use in the optimization of batch processes. These batch-to-batch control and optimization methods use the repetitive nature of batch processes to learn about the effects of past optimization moves and, hence, to achieve better operating trajectories in a short time. In these problems prior batch information is essential. However, benefits of using information from previous batches to aid in the monitoring of existing batch processes that are being operated about a fixed set of manipulated variable trajectories is less certain. This latter problem is addressed in this article.

An explicit procedure is presented for the incorporation of information from prior batches into multivariate statistical process control schemes based on multiway PCA/PLS for monitoring batch processes. The approach involves incorporating the scores values, summarizing the operation of immediately preceding batches, into the MSPC scheme for the current batch. At the model building stage, this is shown to require an iterative scheme to develop the necessary PCA/PLS models. These models can then be use in the same straightforward (noniterative) manner, as regular MSPC approaches to monitor new batches.

Simulations on two-batch polymerization systems are used to demonstrate the method and to illustrate its potential. It is shown, that for off-line analysis or monitoring at the end of each batch, incorporating prior batches with this method provides little advantage over the usual MPCA/MPLS methods

based on only the current batch data. In on-line monitoring, by incorporating prior-batch information, smaller and more consistent control limits and scores for the early stages of the process and faster upset detection of abnormal process conditions can be achieved. However, because the past batch information is more important at early stages of the batch and generally vanishes as more on-line measurements are available, only those process that have limited information at the beginning of the batch would benefit much from this methodology. Moreover, for processes that are only weakly batch-to-batch correlated or that suffer from random batch-to-batch disturbances, no benefit should be expected.

Acknowledgments

The authors greatly thank to Dr S. A. Russell for kindly providing us with their model for condensation polymerization. J. Flores-Cerrillo thanks SEP-CONACYT and McMaster University for financial support.

Literature Cited

Berglund, A., and S. Wold, "INLR, implicit Non-linear Latent Variable Regression," *J. of Chemometrics,* **11**, 141 (1997).

Chin, S. I., S. K. Lee, and J. H. Lee, "A Technique for Integrated Quality Control, Profile Control, and Constrain Handling for Batch Processes," *Ind. Eng. Chem. Res.,* **39**, 693 (2000).

Crosier, R. B, "Multivariate Generalizations of Cumulative Sum Quality-Control Schemes," *Technometrics*, **30**, 291 (1988).

Dorsey, W. A., and J. H. Lee, "Monitoring of Batch Processes Through State-Space Models," Pre-prints of the 6th IFAC Symposium on Dynamics and Control of Process Systems, June 4–6, Korea, 245 (2001).

Flores-Cerrillo, J., "Quality Control for Batch Processes using Multivariate Latent Variable Methods," PhD Thesis, McMaster University, Hamilton, Ontario, Canada (2003).

Frank, I, "A Non-Linear PLS Model," *Chemometrics and Intelligent Laboratory Systems,* **8**, 109 (1990).

Geladi, P, "Analysis of Multi-Way (Multi-Mode) Data," *Chemometrics and Intelligent Lab. Sys.*, **7**, 11 (1989).

Geladi, P, "Notes on The History and Nature of Partial Least Squares Modeling," *J. of Chemometrics,* **2**, 231 (1988).

Geladi, P., and B. R. Kowalski, "Partial Least-Squares Regression: A Tutorial," *Anal. Chim. Acta,* **185** (1), 1 (1986).

Höskuldsson, A, "PLS Regression Methods," *J. Chemometrics*, **2**, 211 (1988).

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

Jollife, I. T, *Principal Component Analysis*, Springer, New York, (1986).

Kourti, T., and J. F. MacGregor, "Multivariate SPC Methods for Process and Product Monitoring," *J. of Quality Technol.* **28** (4), 409 (1996).

Kourti, T., J. Lee, and J. F. MacGregor, "Experiences with Industrial Applications on Projection Methods for Multivariate Statistical Process Control," *Computers in Chem. Eng.,* **20**, S745 (1996).

Kourti, T., P. Nomikos, and J. F. MacGregor, "Analysis, Monitoring and Fault Diagnosis of Batch Processes Using Multi-block and Multiway PLS," *J. of Process Control,* **5**, 277 (1995).

Kourti, T., and J. F. MacGregor, "Multivariate Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods – A Tutorial," *Chemometrics and Int. Lab. Syst*., **28**, 3 (1995).

Krzanowski, W. J, "Cross-Validation in Principal Component Analysis," *Biometrics*, **43**, 575 (1987).

Lynch, D., and C. Kiparissides, "Numerical Simulation of a Tubular Polymerization Reactor," *J. of Appl. Poly. Sci.*, **26**, 1283 (1981).

MacGregor, J. F., and T. Kourti, "Statistical Process Control of Multivariate Processes," *Control Eng. Practice*, **3**, 403 (1995).

MacGregor, J. F., C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process Monitoring and Diagnosis by Multi-Block PLS Methods," *AIChE J.,* **40** (5), 826 (1994).

MacGregor, J. F., and P. Nomikos, "Monitoring Batch Processes," NATO Advanced Study Institute (ASI) for Batch Processing Systems Eng. May29–June 7, Antalya, Turkey (1992).

McNeese, W. H., and R. A Klein, *Statistical Methods for the Process Industries*, ASQC Quality Press (1991).

Nelson, P., J. F. MacGregor, and P. A. Taylor, "Missing Data Methods in PCA and PLS: Score Calculations with Incomplete Observations," *Chemometrics and Intelligent Laboratory Systems*, **35**, 45 (1996).

Nomikos P., and J. F. MacGregor, "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J*., **40** (8), 1361 (1994).

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

Nomikos, P., and J. F. MacGregor, "Multi-way Partial Least Squares in Monitoring Batch Processes," *Chemometrics and Intelligent Laboratory Systems*, **30**, 97 (1995).

Russell, S. A., D. G. Robertson, J. H. Lee, and B. A. Ogunnaike, "Control of Product Quality for Batch Nylon 6,6 Autoclaves," *Chem. Eng. Sci.,* **53**, 3685 (1998a).

Russell, S. A., P. Kesavan, J. H. Lee, and B. A. Ogunnaike, "Recursive Data-Based Prediction and Control of Batch Product Quality," *AIChE J.,* **44**, 2442 (1998b).

Tracy, N. D., J. C. Young, and R. L Mason, "Multivariate Control charts for individual observations," *J. of Quality Technol.*, **24**, 88 (1992).

Westerhuis, J. A., T. Kourti, and J. F. MacGregor, "Analysis of Multiblock and Hierarchical PCA and PLS Models," *J. Chemometrics*, **12**, 301 (1998).

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

Wold, S, "Nonlinear Partial Least Squares Modeling II. Spline Inner Relation," *Chemometrics and Intelligent Laboratory Systems*, **14**, 71 (1992).

Manuscript received April 11, 2003, received Aug. 25, 2003, and final revision received Feb. 13, 2004.