Application of Multivariate Statistical Analysis in Batch Processes

L. Lu. Zheng, Thomas J. Mc Avoy,* Yunbing Huang, and Gang Chen

Department of Chemical Engineering, Institute of System Research, University of Maryland, College Park, Maryland 207402

Multivariate statistical analysis methods such as principal component analysis (PCA) and partial least squares (PLS) are powerful tools for chemical process modeling and monitoring. This paper applies PCA/PLS techniques and their variants, multiway PCA/PLS and orthogonal PCA, to an industrial batch process. The study utilizes an existing large historical process data set and combines multivariate statistical methods with batch time optimization calculations to identify possibilities for process improvement. The objective is to increase throughput by shortening the batch reaction time. The batch time optimization calculations provide feasible setpoint operational suggestions while maintaining the underlying data correlation structure. A pseudo-setpoint approach is also proposed to investigate the reaction period during which the setpoint profiles remain constant. Results for an industrial reactor indicate that the batch time can be shortened by approximately 4.3%.

Introduction

Batch processes are widely used in the chemical industry, particularly in the manufacture of specialty materials and products. These processes are usually highly profitable. However, they often exhibit high complexity and nonlinear process behavior. With the help of modern computer technology, researchers and engineers have been exploring new ways of better modeling batch processes using multivariate statistical data analysis techniques. The advantage of applying these techniques is that they require no prior process knowledge. Only the historical data representing day to day normal process operations are needed. The two most commonly used multivariate statistical analysis methods are principal component analysis (PCA) and partial least squares (PLS). Many studies on PCA and PLS have been conducted and recorded in the literature, such as those by Geladi and Kowalski,[1] Wold et al.,[2] and Höskuldson.[3] PCA is a technique that finds a lower dimensional space capturing the maximum amount of variance in an input data matrix, **X**, without losing any useful information. PLS is a similar approach to PCA except it reduces the dimension of both input and output data matrices, **X** and **Y**, by capturing the maximum amount of covariance between **X** and **Y**, to best predict **Y**. PCA/PLS methods are very useful in the areas of process modeling, monitoring, and fault detection.

Because the data structure of a batch process is three-dimensional, I(batch) × J(variable) × K(time), one needs a technique that transforms a multidimensional data structure into a two-dimensional data matrix. Wold et al.[2] and Geladi[4] presented a multiway PCA/PLS approach which is designed to handle multidimensional process data. Nomikos and MacGregor[5,6] successfully applied the multiway PCA/PLS technique to batch processes. Dong et al.[7] developed a nonlinear PCA/PLS algorithm to better describe the nonlinear behavior in a batch process. The PCA/PLS techniques and their variants are also very useful in process improvement and process optimization. Piovoso and Kosanovich[8] discussed applications of PCA/PLS in process monitoring and controller design for a batch polymer process. Dong et al.[7] proposed a batch to batch optimization algorithm that calculates the gradient of the objective function from a nonlinear PLS model. The gradient is then corrected using the plant data as feedback.

In this paper, attention is also focused on batch process improvement using PCA/PLS and their variants. Multiway PCA/PLS are used to deal with the three-dimensional data array involved in a batch process and to provide a predictive model for batch time. The objective of the modeling is to improve manipulated variable operating policies to minimize the batch time. To help achieve this goal, this paper uses a variant of PCA called orthogonal PCA. Orthogonal PCA/PLS is used to reduce the size of the optimization problem. The batch time optimization calculations pinpoint areas for potential process improvement from extremely large and messy process data. Setpoints are used as the optimization variables. The optimization calculations help to identify better setpoint operating policies associated with shorter batch times. A pseudo-setpoint approach is introduced to investigate periods during which the setpoints are constant. Results are presented for the application of the approach to a real industrial batch reactor. Additional results for a second batch reactor are given by Zheng.[9]

Techniques Used

(a) PCA/PLS and Multiway PCA/PLS. Historical process data can be used to develop a comprehensive representation of a chemical process. How to extract useful information from process data is a very important issue. PCA and PLS are designed to handle multivariate data from chemical processes by characterizing the correlation structure in the process measurements using only a few variables known as the principal components or latent variables. These methods project the information contained in a group of highly correlated process variables onto low-dimensional spaces defined by a few

10.1021/ie000630+ CCC: $20.00
Published on Web 03/10/2001

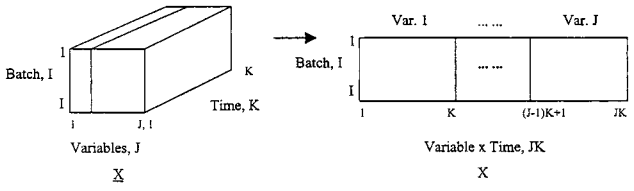

Figure 1. Multiway array unfolding.

principal components. The mathematical representations of PCA and PLS are as follows:

PCA

$$X = \sum_{r=1}^{R} \mathbf{t}_r \mathbf{p}_r^{\mathrm{T}} + \mathbf{E} \tag{1}$$

PLS

$$X = \sum_{r=1}^{R} \mathbf{t}_r \mathbf{p}_r^{\mathrm{T}} + \mathbf{E} \tag{2}$$

$$Y = \sum_{r=1}^{R} \mathbf{u}_r \mathbf{q}_r^{\mathrm{T}} + \mathbf{F} \tag{3}$$

$$Y = X\boldsymbol{\beta} + \Gamma \tag{4}$$

where X is the input variables, such as the process measurements, and $\mathbf{t}_r$, $\mathbf{p}_r$, and $\mathbf{E}$ are known as the scores, loadings, and residuals associated with X; Y is the output variables, such as quality variables, and $\mathbf{u}_r$, $\mathbf{q}_r$, and $\mathbf{F}$ are the scores, loadings, and residuals associated with Y; $\boldsymbol{\beta}$ is the matrix of regression coefficients; and Γ is the regression residual. The principal components are ordered in a way such that the amount of variance described by each principal component decreases as the number of principal components increases. The score vectors define the magnitude of the principal components. The loading vectors are orthonormal and provide the directions of the principal components. The optimum number of principal components retained for the model is determined using cross-validation.[10]

PCA and PLS assume that process data are given by a two-dimensional data matrix $\mathbf{X}$ ($m \times n$). To deal with the three-dimensional batch process data structure, multiway PCA/PLS is applied.[5] Multiway PCA and PLS are modified PCA/PLS techniques that are capable of handling a three-dimensional data array, $\mathbf{X}$, by unfolding it to form a large two-dimensional matrix, $\mathbf{X}$, and then performing PCA and PLS on $\mathbf{X}$. Figure 1 demonstrates how a three-dimensional data array is transformed into a two-dimensional data matrix.

To detect abnormal behavior in a particular batch, the squared prediction error (SPE) and the Hotelling T^2 statistic are used. These two parameters are defined as

$$\mathrm{SPE}_k = \sum_{c=(k-1)J+1}^{kJ} \epsilon_c^{\,2} = x_k(I - \mathbf{P}\mathbf{P}^{\mathrm{T}})x_k^{\mathrm{T}} \tag{5}$$

$$T_k^{\,2} = t_k \boldsymbol{\lambda}^{-1} t_k^{\mathrm{T}} \tag{6}$$

where k denotes a particular time interval and $\boldsymbol{\lambda}^{-1}$ is a diagonal matrix containing the inverse eigenvalues associated with the eigenvectors retained in the model.

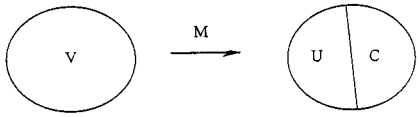

Figure 2. Orthogonal matrix transformation.

The SPE detects any new variations occurring in a process, while T^2 captures larger than normal process variations. The SPE and T^2 confidence limits calculations can be found in Jackson and Mudholkar[11] and Tracy et al.[12]

(b) Orthogonal PCA and Its Use for Preprocessing Data. In addition to the PCA/PLS and multiway PCA/PLS methods, this paper uses another variant of PCA, orthogonal PCA. The mathematical representation of orthogonal PCA as used in this paper is

$$\max_{\mathbf{m}_i} (\mathbf{V}\mathbf{m}_i)^{\mathrm{T}}(\mathbf{V}\mathbf{m}_i) \tag{7}$$

subject to

$$\mathbf{m}_i^{\mathrm{T}}\mathbf{m}_i = 1$$

$$(\mathbf{V}\mathbf{m}_i)^{\mathrm{T}}\mathbf{S} = 0$$

where $\mathbf{V}$ ($m \times n$) are the process measurements, $\mathbf{S}$ ($m \times k$) are the manipulated setpoints, and $\mathbf{m}_i$ are the orthonormal loading vectors. The PCA formulation given in eq 7 was presented by Rao,[13] who gave a solution to the problem. Without the constraint involving $(\mathbf{V}\mathbf{m}_i)^{\mathrm{T}}\mathbf{S}$, eq 7 represents ordinary PCA. By adding the constraint on $(\mathbf{V}\mathbf{m}_i)^{\mathrm{T}}\mathbf{S}$, the optimization procedure is forced to find a normalized linear combination of the measurements that is orthogonal to and not correlated with the measured $\mathbf{S}$ variables and that has a maximum variance.

In this paper an approach different from that taken by Rao is used. Before PCA is applied to the $\mathbf{V}$ data, the data are first separated into a set correlated with $\mathbf{S}$ and a set uncorrelated with $\mathbf{S}$ by multiplication by a square transformation matrix $\mathbf{M}$ as

$$\mathbf{Z} = \mathbf{V}^*\mathbf{M} = [\mathbf{C}, \mathbf{U}] \tag{8}$$

$\mathbf{C}$ is a matrix of transformed measurements that are correlated with $\mathbf{S}$ and $\mathbf{U}$ is a matrix of measurements that are not correlated with $\mathbf{S}$. If a matrix $\mathbf{A}$ is defined as $\mathbf{V}'\mathbf{S}$ and if the singular value decomposition of $\mathbf{A}$ is $\boldsymbol{\mu}\boldsymbol{\Sigma}\boldsymbol{\nu}'$, then as shown in the appendix the transformation $\mathbf{M}$ is equal to $\boldsymbol{\mu}$. Figure 2 provides a graphical representation of the transformation approach used here.

The major advantage of the transformation approach is that it reduces the size of the data set that needs to be examined for analyzing batch processes. For example, in the case of the industrial reactor treated later on, there are 22 process measurements and 4 setpoints. Thus, the dimension of $\mathbf{V}$ is $n \times 22 \times T$, where n is the number of batches and T is the batch duration. By contrast, the dimension of $\mathbf{C}$ is only $n \times 4 \times T$, which is roughly a factor of 5−6 smaller. The transformation $\mathbf{M}$ focuses the data into a set, $\mathbf{C}$, that contains useful information predicting the quality variables associated with a batch process. It should be noted that the transformation $\mathbf{M}$ maps unmeasured process disturbances into both $\mathbf{U}$ and $\mathbf{C}$. The following example illustrates the use of orthogonal PCA.

Table 1. Results after Orthogonal Transformation of Data

case	*X* variables	model	no. of LV	% *y* block explained
1	s_1, s_2	$\mathbf{C} - y$	2	97.2
		$\mathbf{U} - y$	1	0.00
		$[\mathbf{U}, \mathbf{C}, \mathbf{S}] - y$	3	97.2
2	s_1, s_2, d_1	$\mathbf{C} - y$	2	94.2
		$\mathbf{U} - y$	1	11.6
		$[\mathbf{U}, \mathbf{C}, \mathbf{S}] - y$	3	94.9
3	s_1, s_2, d_1, d_2	$\mathbf{C} - y$	2	34.6
		$\mathbf{U} - y$	1	5.4
		$[\mathbf{U}, \mathbf{C}, \mathbf{S}] - y$	3	35.2

To illustrate the transformation approach, consider the following simple example, given by $\mathbf{x} = (\mathbf{Kz})^T$, where **x** consists of four measurements, and the **K** matrix is

$$\mathbf{K} = \begin{bmatrix} -1.600 & 2.000 & -3.000 & 5.000 \\ 0.500 & -1.100 & 1.000 & 0 \\ -1.286 & 1.000 & -0.857 & 0 \\ 1.500 & -1.400 & 1.500 & 0 \end{bmatrix} \quad (9)$$

and **z** is given by

$$\mathbf{z} = \begin{bmatrix} s_1 \\ s_2 \\ d_1 \\ d_2 \end{bmatrix} \quad (10)$$

In this example, **x** can be considered as process measurements, which result from the multiplication of **K** and s_1 through d_2. Here, x_1 is chosen to be the quality variable y. s_1 and s_2 can be considered as the setpoint variables; d_1 can be considered as an unmeasured disturbance that affects all of the x variables; and d_2 can be considered as an unmeasured disturbance that only affects the predicted variable, x_1 or y, e.g. the batch time, and not the process variables x_2 to x_4. The random variables d_1, d_2, s_1, and s_2 are all taken as having zero mean and unit variance. Some random noise (zero mean and variance = 0.1) is added to the all of the x's. The objective is to model y by using PLS with process measurements x_2–x_4. In this example the process measurement matrix **V** is [x_2, x_3, x_4], and the setpoint matrix is $\mathbf{S} = [s_1, s_2]$.

Three different cases are investigated. In the first case, only s_1 and s_2 are used as inputs and d_1 and d_2 are zero. Case 2 adds d_1 as an input, and case 3 includes both d_1 and d_2 as inputs. In each of these cases, the **C** and **U** matrixes are first calculated. Then PLS models are generated that correlate **C** with y, **U** with y, and [**U**, **C**, **S**] with y. Table 1 lists the results for this example. From this table, one can see that **C** is capable of capturing almost all variance in the data when there are no unmeasured disturbances, d_1 and d_2, affecting the process variables and the quality variable. Therefore, in case 1, **C** has almost perfect predictive capability and **U** has no predictive capability because it does not pick up any variability at all. When an unmeasured disturbance, d_1, is present influencing the process measurements (case 2), the predictive capability in **C** is reduced slightly. Because s_1, s_2, and d_1 are mapped into **C**, **C** is able to model essentially all of the variability in y. The unknown disturbance d_1 shows up as well in **U**, and the predictive capability of **U** improves because it captures some of the variance caused by the presence of d_1 in each measurement. In case 3, unmeasured disturbances, d_1 and d_2, influence the batch time. However, d_2 does not affect the process measurements, and it only contributes to the batch time and nothing else. By contrast d_1 does affect the process measurements. The predictive capability in **C** is reduced dramatically because the 1, 4 element **K**, which gives the effect of d_2 on y, is much larger than the rest of the coefficients in the **K** matrix, implying that d_2 has a significant impact on the batch time. **U** has little predictive capability because d_2 is not picked up in the process measurements. Later in the discussion it will be shown that the real batch process examined behaves similarly to cases 1 and 3.

Minimizing the Batch Time

The PLS model for batch time is expressed as follows:

$$t_B = [\mathbf{V}, \mathbf{S}]\boldsymbol{\beta} = \mathbf{V}\beta_1 + \mathbf{S}\beta_2 \quad (11)$$

where β_1 and β_2 come from the final PLS model. Orthogonal PCA calculations can be performed on **V** to give $\mathbf{VM} = [\mathbf{U}, \mathbf{C}]$, which leads to

$$t_B = \mathbf{U}\beta_3 + \mathbf{C}\beta_4 + \mathbf{S}\beta_2 \quad (12)$$

The **U** variables are orthogonal to **S**, and they are not correlated with **S**. The **C** variables are correlated with **S**, and they change when **S** changes. When trying to optimize the batch time t_B, it is only necessary to account for the changes in the **C** variables with **S**. Because the **U** variables do not vary with **S**, they can be eliminated from the optimization by setting them to their mean values, 0. To account for the effect of **S** on **C**, one can generate a PLS model relating **C** to **S** as

$$\mathbf{C} = \mathbf{S}\beta_5 \quad (13)$$

The batch time optimization becomes

$$\min_s \{\mathbf{U}\beta_3 + \mathbf{C}\beta_4 + \mathbf{S}\beta_2\} \rightarrow \min_s \mathbf{S}(\beta_2 + \beta_5\beta_4) \quad (14)$$

Notice here that if one were not using orthogonal PCA, a significantly larger model would be required because the dimension of **V** (dim = 22) is much larger than the dimension of **C** (dim = 4).

To maintain the correlation structure present in the original data, it is necessary to include a SPE constraint at each time step through the batch. These constraints ensure that when **S** is being optimized for a batch, the correlation structure in the data is not violated at any time during the batch. From **U**, **C**, and **S**, one can determine a data matrix to be used in the SPE constraints. Because the **U** variables are all set to zero, for optimization one can estimate **V** as

$$\hat{\mathbf{V}} = [\mathbf{C}, \mathbf{U}]\mathbf{M}^{-1} = [\mathbf{C}, 0]\mathbf{M}^{-1} = [\mathbf{S}\beta_5, 0]\mathbf{M}^{-1} \quad (15)$$

Substituting into the SPE constraint gives

$$|[\hat{\mathbf{V}}, \mathbf{S}]\cdot[I - \mathbf{PP}^T]|^2 = |\mathbf{S}[[\beta_5, 0]\mathbf{M}^{-1}, I]|^2 \leq \epsilon_k \quad (16)$$

where ϵ_k is the SPE for every time interval k. If a batch lasts 100 time units with a time step of 5 units, then there would be 20 SPE constraints, one at each time step. Values of ϵ_k are calculated from the original batch data.[11] The loading matrix **P** is determined from a PCA model describing $[\mathbf{V}, \mathbf{S}] = \mathbf{TP}$ + residual. The manipulated variables **S** also need to be constrained to prevent

large setpoint variable movements. If large setpoint changes are made, the linear relationship between t_B and **S** will not be valid. For this paper, the setpoint movements are bound to ±1 standard deviation, σ_k where $k = 1, ..., \mathbf{K}$. Constraining setpoint changes to this region involves interpolating within the available process data for conditions favoring shorter batch times.

When the above calculation steps are summarized, the final objective function is

$$\min_{S} \mathbf{SW} \tag{17}$$

subject to

$$|S_k\mathbf{Z}|^2 \leq \epsilon_k \qquad k = 1, 2, ..., \mathbf{K}$$

$$S_k \leq \pm 1\sigma_k$$

where k represents the time steps and

$$\mathbf{W} = \beta_2 + \beta_5\beta_6 \tag{18}$$

$$\mathbf{Z} = [[\beta_5, 0]\mathbf{M}^{-1}, I] \tag{19}$$

Because the objective function is linear in **S**, the solution of eq 17 will lie at one of the constraints. The MATLAB Optimization Toolbox routine CONSTR[14] is used to solve eq 17.

Industrial Batch Example

(a) Overview. The batch time minimization calculations are applied to an industrial batch process. Process data are captured from a distributed control system. The data need to be scaled and reviewed for outliers before use in modeling. Multiway PCA/PLS models and variable contribution plots are generated to detect abnormal batch process behavior. The most accurate batch time prediction models are obtained from a repeated process of outlier removal and model regeneration. The final models are used in the batch time optimization calculations. The solution of eq 17 suggests the best setpoint operating policy changes to reduce batch time. Good engineering judgment, based on fundamental process knowledge and process control, is essential in determining which of the mathematically derived improvement suggestions are realizable.

(b) Data Collection and Preprocessing Issues. With the advance of computer technology, process measurements are available continuously. However, not all process measurements should be blindly inputted into the process model. Adequate process knowledge is needed to choose the correct variables for the model. After carefully choosing the proper process measurements, one needs to truncate the process data for all of the batches to the same length in order for the matrix operations associated with multiway PCA/PLS to be carried out. The shortest batch reaction time is chosen to be the standard batch length in a data matrix. The assumption made is that the batch reaction should be either completed or close to being completed at the chosen batch time. The data included up to the standard batch time should be adequate to develop accurate PCA/PLS models. Then the three-dimensional data array is unfolded to a two-dimensional data matrix before PCA/PLS is performed. The three-way array is unfolded in such a fashion that all measurements of a batch over the entire course of that batch are lined up in one row and each row of the matrix represents one batch, batch × (variable × time). In addition, the process data are scaled to zero mean and unit variance to remove the size effects from the measurements.

Preprocessing data correctly is extremely important, because the way the data are processed directly affects the accuracy of the multivariate statistical models. Solid process knowledge and good engineering judgment are involved in most aspects of data preprocessing, such as process measurement selection, batch time determination, and batch length truncation. The person who is responsible for performing data preprocessing needs to be able to identify what conditions determine the beginning of a batch and what conditions indicate the end of a batch. The same individual also has to know where the most valuable information lies in process data and what information is not relevant for the purpose of model development. The size of the computation can be considerably reduced when all of the irrelevant or unimportant process information is removed from the data matrix.

(c) Modeling Aspects. Once the data are ready, one can perform PCA and PLS calculations on it. In this research, the main goal is to obtain an accurate batch time prediction model. A PLS model is used to model the batch time. A PCA model for the input measurements, **V**, is used to detect outliers and thereby to improve the accuracy of the PLS batch time prediction model. A batch with either its PCA scores or its SPE falling outside the confidence limits is considered as an outlier. Outliers can result from different types of abnormal batch behavior, and they are usually associated with poor product quality or long batch time. However, some batches have larger than average variation in the process, resulting in either high score values or above limit SPEs, but the final quality of the product or the batch time remain unaffected. These batches should be retained in the process model because they could very well be important in the optimization. That is why all outliers must be investigated individually and thoroughly to determine the source of the unusual phenomenon. Only batches in which the final product qualities are poor as a result of some abnormal process behavior can be identified as bad batches, and only these batches (bad outliers) should be eliminated from the final batch time prediction model. All other batches (good outliers) should be retained to keep the data set as rich as possible.

Using the results of multiway PCA/PLS, Nomikos and McGregor[6] used corresponding statistical process control (SPC) charts to detect the outliers and to monitor batch operations. There are several different types of SPC charts. The variable contribution plots to the SPE and the variable contribution plots to T^2 are the two most common ones. These plots indicate how much a variable contributes to SPE and T^2. The SPE and T^2 contribution plots can also be generated over the entire duration of the batch. By doing so, one can track when a batch starts to behave out of the ordinary and it should be examined as an outlier. Once the bad outliers are removed, a PLS model can be regenerated and the remaining batches can be reevaluated for abnormality. This procedure is repeated until all of the bad outliers are removed from the model. Each time the model is regenerated the correlation structure of the model changes. The guideline that has been followed during

Table 2. Process Data Preprocessing Result Using Orthogonal PCA

model	no. of LV	% *x* block explained	% *y* block explained
C batch time	1	33.5	28.1
U batch time	1	86.7	0.63
[**U**, **C**, **S**] batch time	1	35.5	27.7
S−**C**	8	91.3	55.0
[**V**, **S**] (PCA)	3	51.7	

the model building phase of the research is that the batch time prediction accuracy should improve progressively with each bad outlier removed. The best PLS model is obtained when the batch time prediction accuracy does not improve when any additional batch is removed.

(d) Process Calculations. The industrial process studied is an almost fully automated batch process. The production facilities consist of a batch reactor, followed by a continuous finishing section. The chemical reaction involved is reversible and can be described as follows:

$$L + M \leftrightarrow N + O$$

The feed and the catalytic materials are charged batchwise to the reactor. An excess amount of material L is present to drive the reaction forward. An important aspect of this batch reaction is the removal of one product, O, produced during the course of the reaction. The reactor overhead system continually removes O from an accumulator drum, with excess L being refluxed back to the reactor. Keeping the content of O in the reaction mass low helps move the reaction toward ultrahigh final conversion.

The final PLS model for this batch process is developed from 88 batches with 22 variables. The shortest batch time is found to be 333 time units, which is chosen to be the batch length for data truncation. By analyzing the batch data, one can develop better setpoint movements to improve process operations and, ultimately, shorten the final batch reaction time. As discussed earlier, orthogonal PCA can be applied to determine variables that are correlated with the setpoint, **C**. Knowing the batch time predictive capability of **C**, one can determine whether there are unmeasured disturbances affecting the batch time but not the process measurements. If there are no unmeasured disturbances that affect the process measurements, the variables that are not correlated with the setpoint, **U**, show little correlation with the batch time. By preprocessing the data using orthogonal PCA and then correlating the results with the batch time, one can find out if any of these disturbances take place during the process.

There are four setpoint variables in the process data set. They are the reactor heat input, the reactor temperature, the reactor pressure, and the reactor overhead accumulator level. The results, as listed in Table 2, show that the process data are similar to those of cases 1 and 3 in Table 1. The PLS model correlating **U** with the batch time captures only 0.63% of the variance, indicating that there are no disturbances in the process affecting the process measurements (case 1 in Table 1). The PLS model correlating **C** to batch time captures 28.1% of variance, suggesting that there are other variables outside of the process data set that are influencing the length of the batch time without affecting the process measurements (case 3 in Table 1). The PLS model between **S** and **C** shows that the variances in both the *X* block (**C**) and *Y* block (**S**) data are reasonably well explained by the model. The number of latent variables used is determined by cross-validation calculations.

In the batch time optimization calculations, the setpoint variables are chosen to be the optimization variables because they are the only variables that can be manipulated. Before the optimization calculations are carried out, it is recommended to have a knowledgeable process engineer go over each phase of the process and determine if all setpoint variables are independent. In some cases, certain setpoints may depend on others during a particular phase. A good example would be a cascade controller where the output of one setpoint is determined by the output of another setpoint variable at some time during a batch. Under these circumstances, the data need to be divided into stages based on the control structure being used. Then, optimization calculations are carried out for individual stages. All setpoint variables should be included in the stagewise optimization calculations as optimization variables if they are independent from each other during that particular phase of the reaction. This type of calculation is known as a multistage calculation.[7] The setpoint variables that are not independent should be removed from the optimization variables because they cannot be manipulated independently during that particular phase of the batch. The dependent setpoints can be included in the data set as regular process variables.

In this process, the control scheme is set up in such a way that the reactor temperature at the beginning of the reaction phase rises until it reaches a target temperature. Then a temperature to pressure cascade is closed, allowing the pressure setpoint to be determined by the amount of positive deviation between the temperature measurement and the temperature setpoint. Once the cascade is turned on, the pressure setpoint is no longer an independent variable. Because of the presence of the temperature−pressure cascade control structure, the data matrix has to be separated into two stages, one before the cascade is closed and one after, and the optimization calculations need to be performed on both stages individually. The shortest length for the time period before the cascade closure is found to be 97 time units, and the shortest length for the time period after the cascade closure is found to be 223 time units. These values are used to separate the data into two stages. Figure 3 gives the optimization results for the first stage of the reaction period during which reactor pressure setpoint is acting as an independent variable. The zero lines in Figure 3 indicate the variable means and the solid black lines are the suggested setpoint movements from the optimizer. The optimization results indicate that the time to cascade closure (first stage of a batch) can be shortened by 14.4 time units on average. However, making operational changes in all of the setpoints indicated in Figure 3 is not possible. These changes need to be evaluated for their feasibility of implementation.

Applying engineering judgment and process knowledge, one would find that suggested setpoint changes for the reactor overhead accumulator level and the reactor pressure and temperature are not feasible. In stage 1, the first drop of feed entering the reactor determines the reaction starting time. The measurements of the overhead accumulator level, pressure, and temperature are allowed to vary freely until the reactor

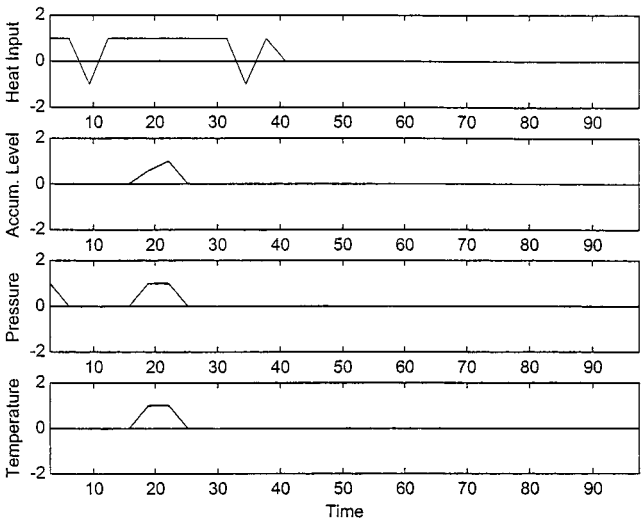

Figure 3. Process optimization result before cascade closure.

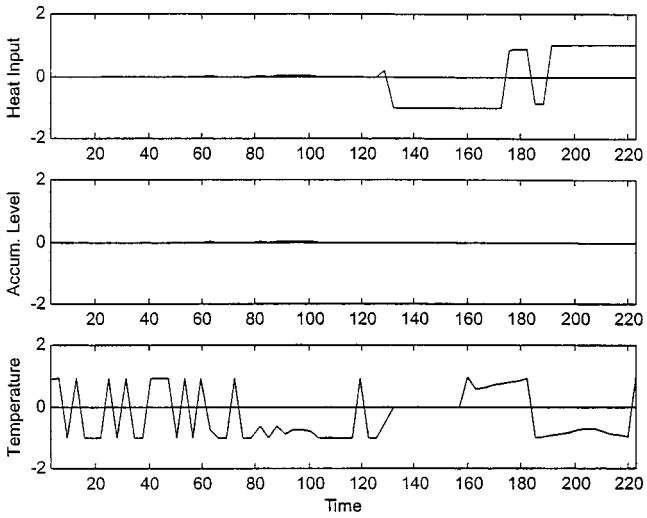

Figure 4. Process optimization results after cascade closure.

finishes charging the feed material L and M. The variation, between 16 and 24 time units, corresponds to the completion of the reactor feed charging process. In the original data set these variables appeared to be independent; however, these setpoint changes involve process-driven events that cannot be manipulated. The optimization results for the overhead accumulator level, pressure, and temperature setpoints indicate that the setpoint changes should be made early. If the feed charging finished quicker, then the setpoint changes can be made earlier and it would take less time to reach the cascade closure point. However, the amount of time it takes to charge the feed varies from batch to batch. In some cases the feed sticks, and this fact leads to a prolonged charging time. Therefore, setpoint operation policy changes associated with the reactor overhead accumulator level, the reactor pressure, and the reactor temperature in the early stage of the reaction cannot be implemented in practice.

Looking at the first graph in Figure 3, the optimizer suggests raising the reactor heat input at the beginning of the first stage. This change agrees with process knowledge because more heat input at the beginning of the reaction raises the reactor temperature faster, which leads to earlier cascade closure. From a process standpoint, this move should be considered. When all other setpoint variables are allowed to operate in the way they normally do and the heat input is allowed to follow the operating profile suggested by the optimizer, the reactor could reach the point of cascade closure in 111.5 time units, compared to an average of 122 time units. Thus, the results predict that the heatup stage can be shortened by 10.5 time units.

The optimization results provide no suggestions on how to improve the batch setpoint operations for the time period after about 40 time units and up to the end of the reactor heatup stage. The reason is that all of the setpoint variables have the same values from 40 time units on, and therefore, there is no variability in the process data. Because the optimizer calculates the best operating policies from the input data, it is not possible to determine any improvement if there is no variability in the data. An approach to determine better operating policies during this period is discussed in the next section.

The optimization results for the second stage after cascade closure are plotted in Figure 4. During this stage, the cascade has been turned on, and the reactor pressure setpoint is no longer independent. For optimizing this part of the operation, the setpoint variables only include reactor heat input, reactor overhead accumulator level, and reactor temperature. It should be noted that the reactor overhead accumulator profile is exactly the same in every batch after the temperature−pressure cascade is closed. No variability exists in the input data set.

The solution calculated for the reactor temperature setpoint is somewhat difficult to interpret. There is a good deal of fluctuation at the beginning of the second stage. This fluctuation is the result of the optimizer trying to pick the best temperature setpoint policy to shorten batch time. The reactor temperature setpoint values in this process are determined from on-line process control programs. Therefore, the temperature profile remains approximately constant in most of the batches up to 157 time units. $\pm 1\sigma$ movements in the temperature setpoint involve almost no movement at all because the variability lies in a very narrow band during this period of time. The temperature setpoint for the last 20 time units of the reaction is determined by the reactor conversion obtained from reactor sampling and by downstream process equipment capacity. There is some variability during this period. The optimization results indicate that, toward the end of the reaction, the reactor temperature should be kept high at first and then low. If the temperature movements can be carried out as suggested, the reaction time after cascade closure can be shortened from 264.2 to 262.4 time units, only a 1.8 time unit reduction.

The optimization results indicate potential setpoint policy changes in reactor heat input. It makes sense to keep the reactor heat input setpoint lower than average until about 126 time units after the cascade is closed and then raise it to a higher than average value at the end of a batch. Lowering the heat input below the average allows the overhead jet system to work better. Therefore, the coproduct O can be removed from the reaction mass more efficiently, and the reaction continues to move forward. When the reaction is almost finished, heat input can be increased again to help boil off the remaining coproduct O in the reaction mass, which pushes the reaction toward its final ultrahigh conversion. If the reactor heat input moves are carried out according to the optimizer, the length of the reaction

time after cascade closure would be shortened from 264.2 to 258.2 time units on average, a 6 time unit reduction.

Looking at the overall batch time optimization results, one can see that the most promising setpoint suggestions lie in the reactor heat input. When the heat flow is increased at the beginning of a batch, the reactor is heated faster, which allows the reactor temperature to reach its cascade closure condition earlier. The recommendation for the heat input at the end of the reaction is also potentially useful. Lowering the heat input toward the end of the reaction enhances the coproduct O removal during that period. Raising the heat input at the very end of the reaction stimulates the final coproduct boiloff, which pushes the reaction toward its final conversion. The suggested heat input policy changes are suitable for implementation because they do not violate any process principles and constraints. When the suggested setpoint changes are applied in the reactor heat input, the total batch time can potentially be decreased by 4.3%. If larger setpoint changes are allowed, i.e., larger than $\pm 1\sigma$, a larger improvement may be possible.

Pseudo-Setpoint Approach

Figures 3 and 4 indicate how setpoint adjustments in reactor heat input can be made to shorten the batch time. However, one can also see that there is almost no variability in some periods of the setpoint profiles, because they follow the same control scheme in every batch. To determine heat-input strategies during these periods, the concept of pseudo-setpoints is introduced. The process measurements associated with the setpoints, labeled here as pseudo-setpoints, are substituted for the real setpoint variables. The rationale for this approach is that the process measurements have variation even though the setpoints are constant. The optimization calculations are carried out using the pseudo-setpoints as optimization variables. Using process measurements as pseudo-setpoints allows one to investigate the "blind periods" that occur in the setpoint profiles and provides information on the process during these periods. Certain types of noise in process measurements, i.e., biased measurements, might cause problems for the pseudo-setpoint approach. Although a general treatment of noise issues is difficult, if the noise is random and zero mean, then meaningful results should be achieved with the pseudo-setpoint approach because the effect of this type of noise should be filtered by PCA and PLS algorithms.

The results of pseudo-setpoint optimization calculations are plotted in Figures 5 and 6 both before and after cascade closure. These figures show results that are consistent with those shown in Figures 3 and 4. The results provide more insight into what kind of process variable movements could lead to shorter batch time. Figures 5 and 6 show that the process measurement of the reactor heat input should be kept higher than the average operating values at the beginning of the batch, lower toward the end of the reaction, and higher again at the very end of the reaction. These suggestions agree with the results from the optimization calculations using setpoints as optimization variables. In addition, the results from the optimization calculations using the pseudo-setpoint variables provide information about the time periods during which the real setpoint variables do not vary. Figures 5 and 6 indicate that the reactor

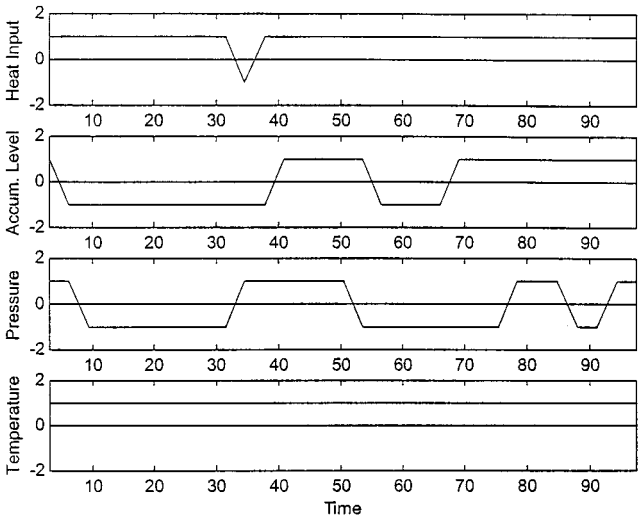

Figure 5. Process pseudo-setpoint optimization results before cascade closure.

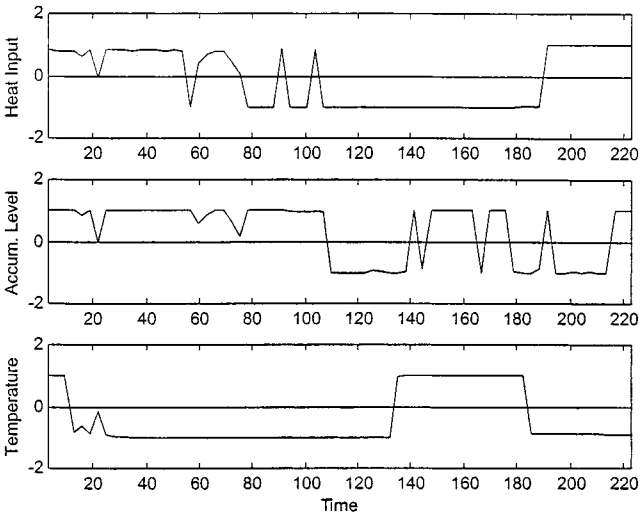

Figure 6. Process pseudo-setpoint optimization results after cascade closure.

heat input should be kept higher than average from the beginning of a batch reaction until about 25 time units after the cascade closure. This information is not available in the setpoint optimization results.

A second batch process has also been studied, and it is primarily manually operated. The chemical reaction involved is similar to that in the first batch process. Instead of having a continuous finishing section, like the first process, every step of the second batch process is carried out in batch fashion. The reaction is driven to completion by having excess reactant and constant removal of one byproduct. Unlike the first process the second process is not operated at maximum capacity because it encounters downstream holdups due to equipment limitations. When the first process encounters problems, equipment and labor from the second process are usually sacrificed to ensure smooth operation of the first process. Because of the operating conditions in the second process, its data set is richer and it has much more variability. The same types of calculations that are carried out for the first process are also carried out for the second one. The optimization results provide similar setpoint suggestions for the reactor heat input, and these suggestions could help shortening the batch time in the second process by 7.5%.

Because of space limitations, a detailed discussion of the results for the second process is not included in this paper but can be found in work by Zheng.[9]

Conclusions

This paper has focused on using existing data to improve batch process operations by applying multivariate statistical methods. The results from an industrial batch process show that these techniques effectively extract useful information from the process data without violating the underlying variable correlation structure. The research demonstrates the utility of preprocessing the data using orthogonal PCA to separate the setpoint correlated variables from the setpoint uncorrelated variables. The recommended setpoint profiles from the optimization calculations point out possible areas for improvement. These suggestions agree with process knowledge. It is found that setpoint improvements can be achieved on reactor heat input, which would help to shorten the batch time. A discussion of how to extract useful information during periods with no setpoint variability using pseudo setpoints has also been given. For the industrial batch reactor considered, it is estimated that batch time can be shortened by approximately 4.3%. Because of production constraints, it was not possible to verify this estimation on the real reactor.

Appendix

Rather than directly solving eq 7, the process measurements are first transformed using eq 8 into a subspace (**C**) that involves linear combinations of measurements that are correlated with **S** and a subspace (**U**) that contains linear combinations of measurements that are not correlated with **S**.

$$\mathbf{Z} = \mathbf{V}^*\mathbf{M} = [\mathbf{C}, \mathbf{U}] \tag{8}$$

By definition **U** satisfies the constraint

$$\mathbf{U}^T\mathbf{S} = 0 \tag{A-1}$$

and **C** satisfies the inequality

$$\mathbf{C}^T\mathbf{S} \neq 0 \tag{A-2}$$

Further **U** is the maximum dimension subspace satisfying eq A-1. Once **U** is determined, PCA can be carried out on **U**, and the resulting solution is also a solution for the orthogonal loading vectors given by eq 7.

To determine **C** and **U**, the matrix **A** is defined as

$$\mathbf{A} \equiv (\mathbf{V})^T\mathbf{S} \tag{A-3}$$

The dimension of **A** is $J \times s$, where J is equal to the number of measurements and s is equal to the number of setpoints. In the development below, it is assumed that the rank of **S** is s and that $J > s$. Let the singular value decomposition of **A** be given as

$$\mathbf{A} = \boldsymbol{\mu}\boldsymbol{\Sigma}v^T \tag{A-4}$$

then $\mathbf{M} = \boldsymbol{\mu}$. Further the subspace **U** results from using columns $s + 1$ to J of $\boldsymbol{\mu}$.

Proof. Consider $\mathbf{Z}^T\mathbf{S}$ after eliminating **A** using eq A-4:

$$\mathbf{Z}^T\mathbf{S} = \mathbf{M}^T\mathbf{A} = \mathbf{M}^T\boldsymbol{\mu}\boldsymbol{\Sigma}\boldsymbol{v}^T = \boldsymbol{\Sigma}\boldsymbol{v}^T \tag{A-5}$$

Equation A-5 makes use of the fact that

$$\mathbf{M}^T\boldsymbol{\mu} = \boldsymbol{\mu}^T\boldsymbol{\mu} = I \tag{A-6}$$

$\boldsymbol{\Sigma}$ is a $J \times s$ matrix with the structure

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{diag}(\sigma_{ii}) \\ 0 \end{bmatrix} \tag{A-7}$$

where the dimension of the diagonal matrix of singular values is $s_x s$ and the dimension of the null matrix is $(J - s)_x s$. Substitution of eqs 8 and A-7 into eq A-5 gives

$$\mathbf{Z}^T\mathbf{S} = \begin{bmatrix} \mathbf{C}^T \\ \mathbf{U}^T \end{bmatrix}\mathbf{S} = \begin{bmatrix} \mathrm{diag}(\sigma_{ii}) \\ 0 \end{bmatrix}\boldsymbol{v}^T \tag{A-8}$$

From eq A-8, eqs A-1 and A-2 follow directly. Because the setpoints are assumed to be independent, all of the elements of the diagonal matrix are nonzero and the dimension of **C** is $s_x s$. The dimension of **U** is $(J - s)_x s$, and **U** is given by multiplying **V** by the last $J - s$ columns of $\boldsymbol{\mu}$. Further the rank of **A** is s, which is equal to the dimension of the null space of **A** defined by

$$\mathbf{A}^T x = 0 \tag{A-9}$$

If the original measurement space has dimension J, then solutions satisfying eq A-9 fall into a subspace whose dimension is reduced by s; i.e., its dimension is $J - s$. Thus, **U** has the maximum dimension of subspaces satisfying eq A-9.

Notation

$\mathbf{A} = \mathbf{V}^T\mathbf{S}$
C = variables that are correlated with the setpoints
d = disturbances in the orthogonal PCA example
E = residual matrix for X-block variables
F = residual matrix for Y-block variables
I = number of batches
J = number of measurements
k = time index
K = batch length, matrix defined in eq 9
m = number of rows in a matrix
$\mathbf{m}_i$ = loading vectors for the orthogonal PCA
M = transformation matrix
n = number of columns in a matrix
p = X-block loading vectors
P = matrix of **p** vectors
q = Y-block loading vectors
Q = matrix of **q** vectors
r = index
R = number of factors retained
s = setpoint variables in the orthogonal PCA example
S = matrix of setpoint variables
SPE = squared prediction error defined by eq 5
t = X-block score vector
t_B = batch time
T = matrix of **t** vectors
T^2 = squared scores, defined by eq 6
u = Y-block score vector
U = matrix of **u** vectors
V = matrix of process measurements
W = matrix of optimization parameters
x = row vector of process measurements
X = process measurements (inputs)

Y = quality variables (outputs)
$\mathbf{z}$ = defined in eq 10
$\mathbf{Z}$ = transformed process measurements
$\boldsymbol{\beta}$ = matrix of regression coefficients
ϵ = error between model and process measurements
Γ = residual in the PLS model
$\boldsymbol{\lambda}$ = diagonal matrix of eigenvalues
$\boldsymbol{\mu}$ = matrix in singular value decomposition
$\boldsymbol{\nu}$ = matrix in singular value decomposition
σ = standard deviation of setpoints
$\boldsymbol{\Sigma}$ = matrix of singular values

Literature Cited

(1) Geladi, P.; Kowalski, B. R. Partial Least Squares Regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.

(2) Wold, S.; Gelaid, P.; Esbensen, K. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37.

(3) Hoskuldsson, A. PLS regression methods. *J. Chemom.* **1988**, *2*, 211−228. Dong, D.; McAvoy, T. J. Batch Tracking via Nonlinear Principal Component Analysis. *AIChE J.* **1996**, *42*, 2199−2208.

(4) Geladi, P. Analysis of Multi-way (Multi-mode) Data. *Chemom. Intell. Lab. Syst.* **1989**, *7*, 11−30.

(5) Nomikos, P.; MacGregor, J. F. Monitoring Batch Processes Using Multi-way Principal Component Analysis. *AIChE J.* **1994**, *40*, 1361−1375.

(6) Nomikos, P.; MacGregor, J. F. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics* **1995**, *37*, 41−59.

(7) Dong, D.; McAvoy, T. J.; Zafiriou, E. Batch to Batch Optimization Using Neural Network Models. *Ind. Eng. Chem. Res.* **1995**, *35*, 2269−2276.

(8) Piovoso, M. J.; Kosanovich, K. A. Applications of Multivariate Statistical Methods to Process Monitoring and Controller Design. *Int. J. Control* **1993**, *59*, 743−765.

(9) Zheng, L. L. Application of Multivariate Statistical Analysis in Batch Processes. M.S. Thesis, University of Maryland, College Park, MD, 1999.

(10) Osten, D. W. Selection of Optimal Regression Models via Cross-validation. *J. Chemom.* **1988**, *2*, 39−48.

(11) Jackson, J. E.; Mudholkar, G. S. Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics* **1979**, *21*, 341−348.

(12) Tracy, N. D.; Young, J. C.; Mason, R. L. Multivariate Control Charts for Individual Observations. *J. Qual. Technol.* **1992**, *24*, 88−95.

(13) Rao, C. The Use and Interpretation of Principal Component Analysis in Applied Research. *Sankhya A* **1964**, *26*, 329−358.

(14) *Optimization Toolbox User's Guide*; Mathworks, Inc.: Natick, MA, 1996.

Received for review June 30, 2000
Accepted January 24, 2001

IE000630+