Extended PLS Approach for Enhanced Condition Monitoring of Industrial Processes

U. Kruger
Intelligent System and Control Group, The Queens University of Belfast, Belfast BT9 5AH, U.K.

Q. Chen and D. J. Sandoz
Control Technology Centre, Division of Mechanical Engineering, University of Manchester, Manchester, U.K.

R. C. McFarlane
Invensys Performance Solutions, Houston, TX 77042

An extended partial-least squares (EPLS) algorithm is introduced to correct a deficiency of conventional partial least squares (PLS) when used as a tool to detect abnormal operating conditions in industrial processes. In the absence of feedback control, an abnormal operating condition that affects only process response variables will not be propagated back to the process predictor (or input) variables. Thus monitoring tools developed under the conventional PLS framework and based only on the predictor matrix will fail to detect the abnormal condition. The EPLS algorithm described removes this deficiency by defining new scores that are based on both predictor and response variables. The EPLS approach provides two monitoring charts to detect abnormal process behavior, as well as contribution charts to diagnose this behavior. To demonstrate the utility of the new approach, the extended algorithm and monitoring tools are applied to a realistic simulation of a fluid catalytic cracking unit and to a real industrial process that involves a complex chemical reaction.

Introduction

The detection and diagnosis of abnormal situations in the operation of industrial processes is a problem of considerable challenge that is attracting wide attention in both academe and industry. Nimmo (1995) showed that the U.S.-based petrochemical industry alone could save up to $10 billion per year if abnormal situations could be detected, diagnosed, and appropriately dealt with. The consequences of not being able to detect such abnormal situations range from increased operational costs to costly plant shutdowns.

Industrial processes often present a large number of process variables, such as temperatures, pressures, flow rates, compositions, which are typically sampled at a frequency of one minute. Identifying and troubleshooting abnormal operating conditions simply by observation are difficult tasks with such a large amount of data, particularly since the process variables are usually highly correlated (MacGregor et al., 1991). However, the sampled data have embedded within them information for revealing the state of the process operation. The difficult issue is to extract this information from the data and to present it in a way that can be easily interpreted.

To address the issue of detection and diagnosis of abnormal operating conditions, multivariate statistical process control (MSPC) approaches have been successfully employed (Kresta et al., 1991; MacGregor and Kourti, 1995; Kourti and MacGregor, 1995). The MSPC techniques aim to reduce the number of variables required to describe significant variation of the process. The recorded data are thereby compressed into an alternate set containing fewer variables that are more manageable and interpretable.

One such MSPC approach is partial least squares (PLS), pioneered by H. Wold (Wold, 1966a,b). The PLS method identifies a parametric regression matrix based upon predictor and response matrices that are constructed from reference data of the process. The predictor matrix comprises the measured input variables to the process (predictor variables), while the response matrix comprises the process output variables (response variables).

The PLS algorithm decomposes the predictor and response matrices into rank-one component matrices. Each

Correspondence concerning this article should be addressed to U. Kruger.

component matrix is composed of a vector product in which one vector describes the variation (score vector) and the other the contribution (loading vector) of the score vector to its corresponding data matrix, that is, the predictor or the response matrix. The decomposition is an iterative procedure in which a pair of component matrices (one for the predictor and one for the response) is calculated at each iteration step. The data reduction is essentially a projection of the original data down to a low-dimensional space defined by a few latent variables (LVs) that describe significant variation in the process. To determine the number of LVs to be retained, a number of procedures have been proposed, including cross validation (Wold, 1978) and analysis of variance (Jackson, 1991).

MacGregor and Kourti (1995) and Wise et al. (1996) established that the PLS decomposition of the predictor matrix can be employed to detect abnormal operating conditions (condition monitoring) of continuous industrial processes by deriving statistics that are analogous to those commonly used in principal-component analysis (Jackson, 1991). PLS decomposition of the predictor matrix allows the calculation of two statistics. The T-squared statistic describes variation of the predictor variables that is significant for predicting the response variables. The Q statistic provides information on the variation of the predictor variable residuals that remain after the PLS decomposition. Both statistics can be plotted in univariate statistical monitoring charts with a time base. This approach is referred to here as Approach I.

Although successful applications of Approach I have been introduced (Kourti and MacGregor, 1995; Wise et al., 1996; Morud, 1996), they do not necessarily detect every kind of abnormal process behavior. This is particularly true if abnormal process conditions manifest themselves only in response variables that are not under closed-loop control. In this case, the abnormal behavior does not propagate through to the predictor variables by controller feedback, and therefore remains undetected because the charted statistics are derived from the predictor variables only.

Another approach for exploiting PLS as a condition monitoring tool, in this article referred to as Approach II, is discussed by MacGregor et al. (1991) and Kresta et al. (1991). In this approach, various statistical plots are used to detect abnormal process behavior:

1. Plots of the squared prediction error of the response variables vs. the score values of each score vector representing the predictor variables (monitoring charts);
2. Plots of each combination of two score vectors representing the predictor variables (scatter plots);
3. Plots of the squared prediction error of either the predictor or response variables vs. time (SPE charts).

This approach can be cumbersome and difficult to use, particularly if the required number of component matrices is greater than two or three. For example, if six component matrices are required for each of the predictor and the response matrices, then Approach II would potentially require 23 charts.

In this article, an extension to the standard PLS aglorithm, referred to as extended PLS or EPLS, is introduced. The extension results in the determination of two new PLS scores, denoted as generalized scores. These generalized scores are calculated by augmenting the matrices that result from the standard PLS decomposition, using both predictor and response information. On the basis of the generalized scores, univariate statistics are calculated, analogous to those calculated in Approach I. These statistics are plotted vs. time as they are in Approach I to allow process monitoring, but in this case they contain information from both the predictor and response variables. Thus, the monitoring method is also sensitive to abnormal conditions that manifest themselves in response variables, even when no feedback on these variables is present. This represents a useful extension to the capability of the standard PLS Approach I for condition monitoring, and a useful alternative to Approach II, which can be cumbersome when the number of charts is large.

To demonstrate the usefulness of the EPLS monitoring charts, and to make comparisons with Approach I, two case studies are considered. The first test process is a detailed mechanistic simulation of a fluid catalytic cracking unit (FCCU) introduced by McFarlane et al. (1993), and the second is a real industrial process that produces two different solvents as a result of a complex chemical reaction carried out in parallel fluidized-bed reactors. In each process, abnormal operating conditions resulting from external disturbances and internal faults are analyzed.

In the next section, the results of the standard PLS decomposition are summarized as are the statistics that are generated from PLS for use in process condition monitoring. The reader is referred to the Appendices and the cited literature for further details. Building on this result, details of EPLS algorithm are then presented, including the derivation of univariate EPLS statistics useful for process condition monitoring. Finally, the two case studies are presented to demonstrate the advantages of EPLS over the conventional PLS approaches for condition monitoring.

Condition Monitoring Using PLS and EPLS

Standard PLS aglorithm

The standard PLS identification technique relies on decomposing the predictor matrix, $\boldsymbol{X} \in \mathbb{R}^{K \times M}$, and the response matrix $\boldsymbol{Y} \in \mathbb{R}^{K \times N}$ to a sum of rank one component matrices (Geladi and Kowalski, 1986). Here K is the number of measurements, N is the number of response variables, and M is the number of predictor variables. Typically, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are mean centered and appropriately scaled prior to the identification procedure. The PLS decomposition of $\boldsymbol{X}$ and $\boldsymbol{Y}$ (see Appendix A for a summary of this procedure) results in the following:

$$\boldsymbol{X} = \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i + \boldsymbol{F_n} = \sum_{i=1}^{n} \boldsymbol{t}_i \boldsymbol{p}_i^T + \boldsymbol{F}_n = \boldsymbol{T}_n \boldsymbol{P}_n^T + \boldsymbol{F}_n = \boldsymbol{X_n} + \boldsymbol{F_n}$$

$$\boldsymbol{Y} = \sum_{i=1}^{n} \tilde{\boldsymbol{Y}}_i + \boldsymbol{E}_n = \sum_{i=1}^{n} \hat{\boldsymbol{u}}_i \boldsymbol{q}_i^T + \boldsymbol{E}_n = \hat{\boldsymbol{U}}_n \boldsymbol{Q}_n^T + \boldsymbol{E}_n = \hat{\boldsymbol{Y}}_n + \boldsymbol{E}_n, \tag{1}$$

where n represents the number of rank one component matrices, $\tilde{\boldsymbol{X}}_i$ and $\tilde{\boldsymbol{Y}}_i$, retained in the decomposition. The vectors $\boldsymbol{t}_i$ and $\hat{\boldsymbol{u}}_i$ are referred to as the t- and u-score vectors, or LVs, while the vectors $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$, are loading vectors, and $\boldsymbol{E}_n$ and $\boldsymbol{F}_n$ are residual matrices. The u-scores, $\hat{\boldsymbol{u}}_i$, can be estimated from the t-score vectors as follows:

$$\hat{\boldsymbol{U}}_n = \left[\hat{\boldsymbol{u}}_1 \vdots \cdots \vdots \hat{\boldsymbol{u}}_n \right] = \left[\boldsymbol{t}_1 b_1 \vdots \cdots \vdots \boldsymbol{t}_n b_n \right] = \boldsymbol{T}_n \boldsymbol{B}_n, \tag{2}$$

where $\boldsymbol{B}_n$ represents a diagonal matrix containing the regression coefficients of the score model, b_i, determined by the PLS algorithm (see Appendix A).

Most industrial processes have strongly correlated process variables, and often only a few LVs are needed to describe most of the process variation. The number of required LVs is typically determined by cross validation, as demonstrated, for example, by MacGregor et al. (1991), MacGregor and Kourti (1995), and Morud (1996).

Different variations on the basic PLS aglorithm have been proposed, for example, by Manne (1987), Geladi and Kowalski (1986), and de Jong (1993), among others.

Review of statistics for PLS condition monitoring

As mentioned previously, in the conventional PLS Approach I, two statistical monitoring charts are derived, based on the Hotellings T^2 and the Q statistics, commonly employed in PCA (Jackson, 1991). They are denoted here as PLS-T^2 and PLS-Q, respectively, and are written as follows for time instance k:

$$PLS - T_k^2 = \sum_{i=1}^{n} \left(\frac{t_{ki}}{{}_T\sigma_i} \right)^2$$

$$PLS - Q_k = \sum_{i=1}^{M} \left(\frac{f_{ki}}{{}_{\hat{x}}\sigma_i} \right)^2, \tag{3}$$

where t_{ki} denotes the value of the ith t-score and ${}_T\sigma_i$ the standard deviation of the ith score vector of the predictor variables; f_{ki} represents the residual of the ith predictor variable; and ${}_{\hat{x}}\sigma_i$ is the standard deviation of the ith residual variable. Each statistic can be plotted in time to form a univariate monitoring chart. The statistics each follow a central chi-squared distribution, with n and M degrees of freedom, respectively, if the predictor variables are normally distributed, and thus confidence limits are easily calculated [see, for instance, Box (1954)].

It should be noted that the normalization of t_{ki} and f_{ki} is essential to provide a sensitive statistic. If this is not done, then the t-scores with a large variance, usually the first few, dominate the resultant value of PLS-T^2, and the residuals of the predictor variables that have a large variance overshadow residuals with relatively small variance. A fault condition that affects primarily the t-scores or residuals that have small variance, may remain undetected in such cases.

Calculating PLS-T^2 and PLS-Q values for newly observed data provides information on the current operating condition. Newly calculated PLS-T^2 values that exceed the confidence limit calculated from the reference data set (while taking into consideration that a small number will exceed this limit by chance alone) imply that the general process behavior has changed significantly or the process has moved to a new operating region. In contrast, PLS-Q values that exceed their confidence limit indicate that the relationship between the predictor variables have changed relative to the relationship prevalent within the reference data.

Derivation of the EPLS algorithm and the generalized scores

The EPLS algorithm produces generalized scores that contain information on the variation of the predictor and response variables. These scores provide the basis for more effective process-condition monitoring than existing approaches based on scores that describe variation in the predictor variables only. The generalized scores are calculated after the weight and loading matrices are determined by the standard PLS algorithm (see Appendix A), and are based on an augmented version of the matrix equation (Eq. 1):

$$\left[\boldsymbol{Y} \vdots \boldsymbol{X}\right] - \left[\hat{\boldsymbol{Y}}_n \vdots \boldsymbol{X}_n\right] = \left[\boldsymbol{E}_n \vdots \boldsymbol{F}_n\right]. \tag{4}$$

Incorporating Eqs. 1 and 2, Eq. 4 can be rewritten as

$$\left[\boldsymbol{Y} \vdots \boldsymbol{X}\right] - \boldsymbol{T}_n\left[\boldsymbol{B}_n\boldsymbol{Q}_n^T \vdots \boldsymbol{P}_n^T\right] = \left[\boldsymbol{E}_n \vdots \boldsymbol{F}_n\right]. \tag{5}$$

As shown in Appendix B, the matrix product $\boldsymbol{B}_n\boldsymbol{Q}_n^T$ is equivalent to $\boldsymbol{P}_n^T\boldsymbol{B}_{PLS}^{(n)}$, where $\boldsymbol{B}_{PLS}^{(n)}$ is the PLS regression matrix with n LVs retained. Using this result, Eq. 5 becomes

$$\left[\boldsymbol{Y} \vdots \boldsymbol{X}\right] - \boldsymbol{T}_n\boldsymbol{P}_n^T\left[\boldsymbol{B}_{PLS}^{(n)} \vdots \boldsymbol{I}_M\right] = \left[\boldsymbol{E}_n \vdots \boldsymbol{F}_n\right], \tag{6}$$

where $\boldsymbol{I}_M$ denotes the $M \times M$ identity matrix. Carrying out a postmultiplication of Eq. 6 by the generalized inverse of $[\boldsymbol{B}_{PLS}^{(n)} \vdots \boldsymbol{I}_M]$ results in

$$\left[\boldsymbol{Y} \vdots \boldsymbol{X}\right]\left[\boldsymbol{B}_{PLS}^{(n)} \vdots \boldsymbol{I}_M\right]^{\perp} - \boldsymbol{T}_n\boldsymbol{P}_n^T = \left[\boldsymbol{E}_n \vdots \boldsymbol{F}_n\right]\left[\boldsymbol{B}_{PLS}^{(n)} \vdots \boldsymbol{I}_M\right]^{\perp}, \tag{7}$$

where $^{\perp}$ denotes the generalized inverse. As shown in Appendix C, the postmultiplication of Eq. 7 by $\boldsymbol{R}_n$ (from Appendix A) leads to a formula for calculating the scores of the predictor matrix, $\boldsymbol{T}_n$:

$$\boldsymbol{T}_n = \left[\boldsymbol{Y} \vdots \boldsymbol{X}\right]\left[\boldsymbol{B}_{PLS}^{(n)} \vdots \boldsymbol{I}_M\right]^{\perp}\boldsymbol{R}_n - \left[\boldsymbol{E}_n \vdots \boldsymbol{F}_n\right]\left[\boldsymbol{B}_{PLS}^{(n)} \vdots \boldsymbol{I}_M\right]^{\perp}\boldsymbol{R}_n. \tag{8}$$

Finally, Eq. 8 can be written simply as

$$\boldsymbol{T}_n = \boldsymbol{T}_n^* - \boldsymbol{E}_n^*, \tag{9}$$

where $\boldsymbol{T}_n^* = [\boldsymbol{Y} \vdots \boldsymbol{X}]\boldsymbol{C}_{PLS}^{(n)}$, $\boldsymbol{E}_n^* = [\boldsymbol{E}_n \vdots \boldsymbol{F}_n]\boldsymbol{C}_{PLS}^{(n)}$ and $\boldsymbol{C}_{PLS}^{(n)} = [\boldsymbol{B}_{PLS}^{(n)} \vdots \boldsymbol{I}_M]^{\perp}\boldsymbol{R}_n$. The matrix $[\boldsymbol{E}_n \vdots \boldsymbol{F}_n]$ is referred to as the augmented residual matrix to $\boldsymbol{F}_n$.

The columns of the matrix $\boldsymbol{T}_n^*$ are further referred to as the generalized t-score vectors while the columns of the matrix $\boldsymbol{E}_n^*$ are denoted as generalized residual score vectors. Equation 9 forms the basis for derivation of statistics and monitoring charts, described in the next section, that are useful for condition monitoring.

Statistics for EPLS

The statistical properties of the generalized scores are summarized below:

1. The generalized t-score vectors, $\boldsymbol{t}_k^*$, as well as the generalized residual score vectors, $\boldsymbol{e}_k^*$, describe mean-centered

signals if the columns in the predictor and the response matrices have been mean centered prior to the PLS identification.

2. The t-score vectors of the standard PLS aglorithm are orthogonal (Hoskuldsson, 1988). In contrast to the standard t-scores, both generalized score types are not orthogonal, irrespective of the number of retained LVs. A proof of this is provided in Appendix D.

To utilize the generalized scores for process condition monitoring, it is desirable to have statistically independent scores, which requires orthogonality. In order to achieve mutually orthogonal $\boldsymbol{t}_k^*$ and $\boldsymbol{e}_k^*$ score vectors, a singular value decomposition (SVD) (Golub and Van Loan, 1996) of the corresponding matrices, $\boldsymbol{T}_n^*$ and $\boldsymbol{E}_n^*$, can be carried out, which results in

$$\begin{aligned}\boldsymbol{T}_n^* &= \boldsymbol{V}_n^{(T^*)}\boldsymbol{\Lambda}_n^{(T^*)}\boldsymbol{W}_n^{(T^*)T}\\ \boldsymbol{E}_n^* &= \boldsymbol{V}_n^{(E^*)}\boldsymbol{\Lambda}_n^{(E^*)}\boldsymbol{W}_n^{(E^*)T},\end{aligned} \tag{10}$$

where $\boldsymbol{V}_n^{(T^*)}$, $\boldsymbol{\Lambda}_n^{(T^*)}$, and $\boldsymbol{W}_n^{(T^*)}$ describe the SVD of the generalized t-score matrix and $\boldsymbol{V}_n^{(E^*)}$, $\boldsymbol{\Lambda}_n^{(E^*)}$, and $\boldsymbol{W}_n^{(E^*)}$ represent the SVD of the generalized residual matrix. The dimensions of these matrices are as follows: $\boldsymbol{V}_n^{(T^*)}$ and $\boldsymbol{V}_n^{(E^*)}$ are $K \times n$ matrices, and $\boldsymbol{\Lambda}_n^{(T^*)}$, $\boldsymbol{\Lambda}_n^{(E^*)}$ $\boldsymbol{W}_n^{(T^*)}$, and $\boldsymbol{W}_n^{(E^*)}$ are $n \times n$ matrices. The columns of the matrices $\boldsymbol{V}_n^{(T^*)}$, $\boldsymbol{V}_n^{(E^*)}$, $\boldsymbol{W}_n^{(T^*)}$, and $\boldsymbol{W}_n^{(E^*)}$ are orthonormal, and $\boldsymbol{\Lambda}_n^{(T^*)}$ and $\boldsymbol{\Lambda}_n^{(E^*)}$ are diagonal. Based on the SVD of the generalized score matrices of the reference data, the following relationship provides orthogonal scores:

$$\begin{aligned}\boldsymbol{\Omega}_{(T^*)} &= \boldsymbol{T}_n^*\boldsymbol{W}_n^{(T^*)}\boldsymbol{\Lambda}_n^{(T^*)^{-1}}\sqrt{K-1}\\ \boldsymbol{\Omega}_{(E^*)} &= \boldsymbol{E}_n^*\boldsymbol{W}_n^{(E^*)}\boldsymbol{\Lambda}_n^{(E^*)^{-1}}\sqrt{K-1},\end{aligned} \tag{11}$$

in which the columns of $\boldsymbol{\Omega}_{(E^*)}$ and $\boldsymbol{\Omega}_{(T^*)}$ are orthogonal. Utilizing Eq. 9, $\boldsymbol{\Omega}_{(E^*)}$ and $\boldsymbol{\Omega}_{(T^*)}$ can be directly calculated from the augmented data and residual matrices, $[\boldsymbol{Y}\dot{:}\boldsymbol{X}]$ and $[\boldsymbol{E}_n\dot{:}\boldsymbol{F}_n]$, as

$$\begin{aligned}\boldsymbol{\Omega}_{(T^*)} &= \left[\boldsymbol{Y}\dot{:}\boldsymbol{X}\right]\boldsymbol{C}_{PLS}^{(n)}\boldsymbol{W}_n^{(T^*)}\boldsymbol{\Lambda}_n^{(T^*)^{-1}}\sqrt{K-1}\\ \boldsymbol{\Omega}_{(E^*)} &= \left[\boldsymbol{E}_n\dot{:}\boldsymbol{F}_n\right]\boldsymbol{C}_{PLS}^{(n)}\boldsymbol{W}_n^{(E^*)}\boldsymbol{\Lambda}_n^{(E^*)^{-1}}\sqrt{K-1}\,.\end{aligned} \tag{12}$$

For the generalized score vectors of the kth data point, ${}_k\boldsymbol{\omega}_{(T^*)}$ and ${}_k\boldsymbol{\omega}_{(E^*)}$, the sum of the squared elements can be used to define a univariate statistic for each vector. These statistics represent the EPLS-T^2 and the EPLS-Q statistic and are defined as follows:

$$\begin{aligned}EPLS - T_k^2 &= \sum_{i=1}^{n} {}_{ki}\omega_{(T^*)}^2\\ EPLS - Q_k &= \sum_{i=1}^{n} {}_{ki}\omega_{(E^*)}^2.\end{aligned} \tag{13}$$

The definition of the confidence limits for the EPLS$-T^2$ and EPLS-Q statistic is analogous to that for the corresponding PLS statistics. If $\boldsymbol{t}_k^*$ and $\boldsymbol{e}_k^*$ describe normally distributed random signals, the EPLS$-T^2$ and Q statistics follow a chi-squared distribution with n degrees of freedom. If a new EPLS$-T^2$ or Q value is below its corresponding confidence limit, which is typically selected as 95% or 99%, the hypothesis that the process behaves normally is accepted, otherwise it is rejected and it is concluded that the process is behaving abnormally. It should be noted, however, that 5% and 1% of the values of EPLS-T^2 and Q will exceed the 95% and 99% confidence limits, respectively, by chance alone.

Abnormally large EPLS-T^2 and/or Q values can occur if the relationship between the predictor and response variables has changed (such as in a time-variant process) or the disturbance or measurement characteristics have changed. Other reasons may be that the process is operating at a different operating point, or abnormal process behavior has occurred. The hypothesis test is therefore a comparison be-

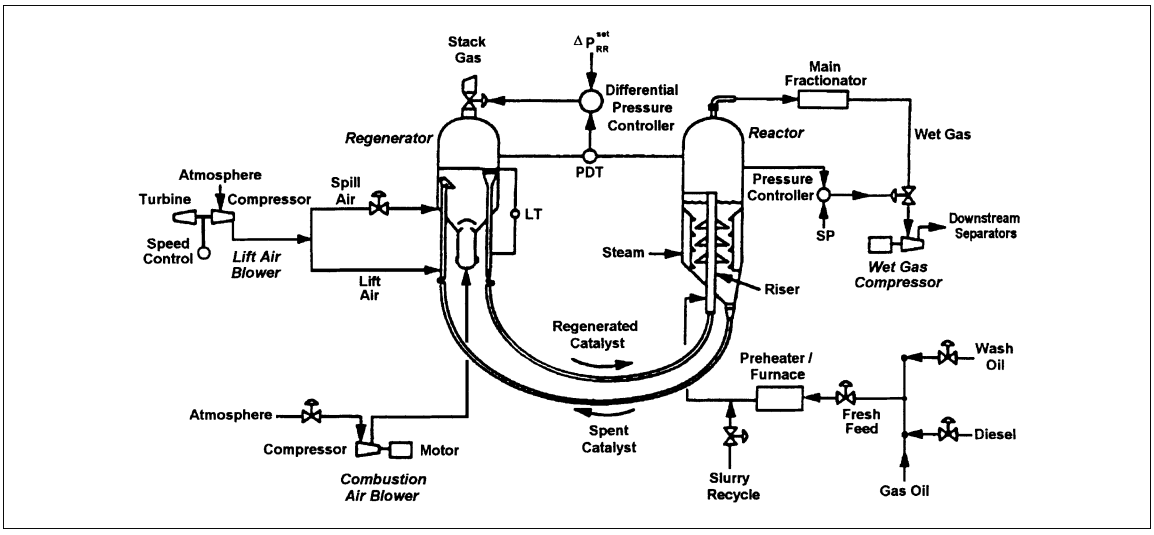


Figure 1. Fluid catalytic cracking unit.

tween the current process operation and the process operation captured in the reference data. Note that the reference data describe the process under normal operation and must capture every variation that can occur under normal operation, otherwise the statistical hypothesis test will be too sensitive.

Wise et al. (1996) emphasised that the scores may not be normally distributed, so that conclusions drawn from charts based on the assumption of normality may be misleading. Dunia et al. (1996) analyzed the influence of an exponential weighted moving average (EWMA) upon the Q statistic incorporating PCA. It was found that the average run length (ARL)—as the average time passed until abnormal process behavior is detected—for detecting faulty conditions on sensors could be reduced by invoking the EWMA Q-statistic. In this article, an EWMA approach is applied to the PLS-Q and the EPLS-Q statistic, and each confidence limit is empirically determined as suggested by Box et al. (1978).

Case Studies

Fluid catalytic cracking unit

A fluid catalytic cracking unit (FCCU) is an important economic unit in refining operations. It typically receives several different heavy feedstocks from other refinery units and cracks these streams to produce lighter, more valuable components that are eventually blended into gasoline and other products. The particular Model IV unit described by McFarlane et al. (1993) is illustrated in Figure 1. The principal feed to the unit is gas oil, but heavier diesel and wash oil streams also contribute to the total feed stream. Fresh feed is preheated in a heat exchanger and furnace and then passed to the riser, where it is mixed with hot, regenerated catalyst from the regenerator. Slurry from the main fractionator bottom is also recycled to the riser. The hot catalyst provides the heat necessary for the endothermic cracking reactions. The gaseous cracked products are passed to the main fractionator for separation. Wet gas off the top of the main fractionator is elevated to the pressure of the lights-end plant by the wet-gas compressor. Further separation of light components occurs in this light-ends separation section.

As a result of the cracking process, a carbonaceous material, coke, is deposited on the surface of the catalyst, which depletes its catalytic properties. For this reason, spent catalyst is recycled to the regenerator where it is mixed with air in a fluidized bed for regeneration of its catalytic properties. Oxygen reacts with the deposited coke to produce carbon monoxide and carbon dioxide. Air is pumped to the regenerator with a high-capacity combustion air blower and a smaller lift air blower. In addition to contributing to the combustion process, air from the lift air blower assists with catalyst circulation. Complete details of the mechanistic simulation model for this particular Model IV FCCU can be found in McFarlane et al. (1993).

The selected predictor variables for the FCCU case study are given in Table 1. All of these variables belong to the feed section of the unit. To simulate realistic disturbance conditions, various different autoregressive integrated moving average (ARIMA) and autoregressive moving average (ARMA) signals were superimposed on these variables, as indicated in Table 1. Wash oil flow rate was held constant.

Table 1. Selected Predictor Variables for FCCU Case Study

Predictor Variables	Signal
Diesel flow rate	ARMA Sequence
Total fresh feed	ARIMA Sequence
Slurry flow rate	ARIMA Sequence
Preheater outlet temperature	ARIMA Sequence

The response variables included excess oxygen in the flue gas, concentration of carbon monoxide in the flue gas, riser temperature, regenerator bed temperature, regenerator standpipe level, reactor pressure, as well as eight other measured variables from the system. See McFarlane et al. (1993) for a complete list of measured variables for the FCCU system.

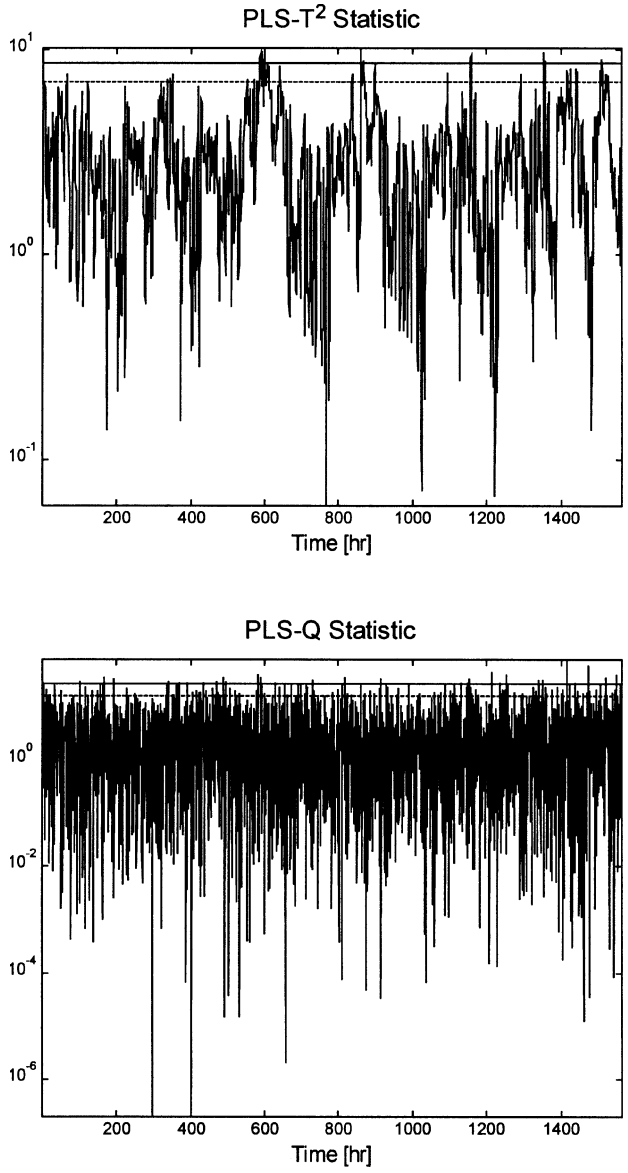


Figure 2. PLS monitoring charts for approximately 62 days of normal operation (Upper Chart: PLS-T^2; Lower Chart: PLS-Q).

To test and compare the PLS and EPLS condition monitoring approaches, the FCCU simulator included several preprogrammed faults that could be applied on command. The first was a step change to the coke formation factor of the feed, which simulated a plug of heavier-than-normal feed entering the unit. The second simulated a disruption in the flow of regenerated catalyst between the regenerator and riser, which is typically caused by partial or complete plugging of steam injectors located in this line.

In the first two runs, no advanced control system was present, only regulatory control for the air compressor flow rates and the reactor pressure. None of these controllers result in feedback between any of the defined response and predictor variables.

Using cross validation, the number of latent variables was determined to be three. Figures 2 and 3 show the resulting T^2 and Q monitoring charts for the PLS and EPLS approaches, respectively, for a period of approximately 62 days of normal operation. In all figures where T^2 and Q plots are presented, the upper solid horizontal line represents the 99% confidence limit for the particular statistic plotted, while the bottom dotted line represents the 95% confidence limit. Futhermore, the ordinate of each T^2 and Q plot is logarithmic to the basic of 10. The sampling period was 30 min.

In Figure 4, the responses of the PLS-T^2 and Q statistics are shown for the first fault, injected at 190.5 h. Since this fault simulates a change in composition of the feed—a plug of heavier feed—its effect is felt immediately in the riser and subsequently in other parts of the unit that are affected by a

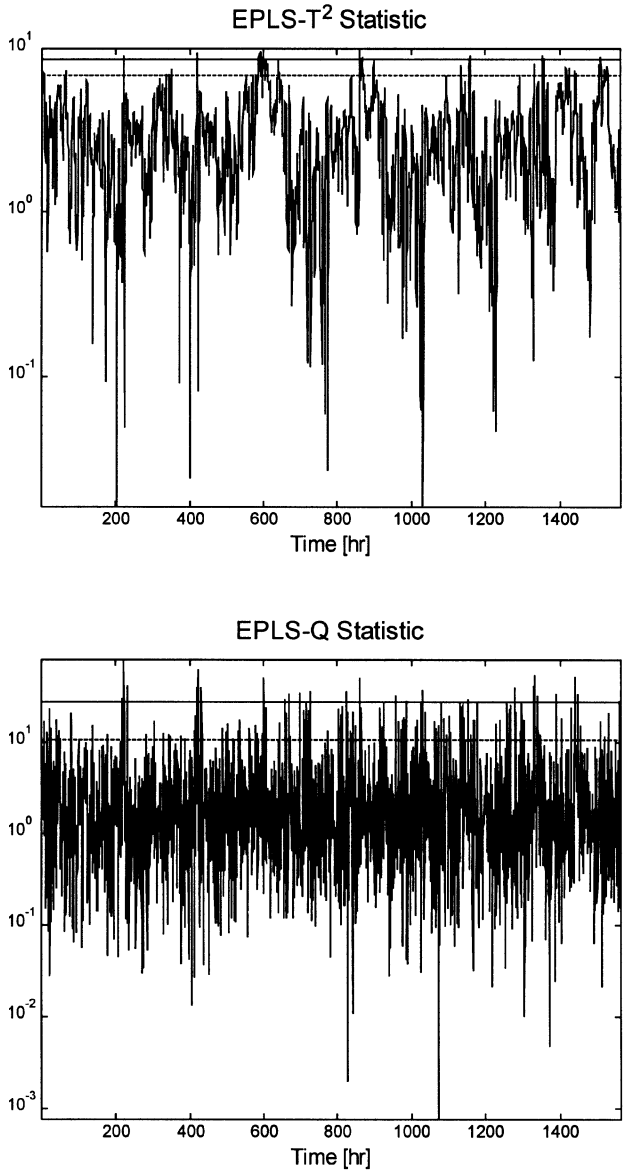


Figure 3. EPLS monitoring charts for approximately 62 days of normal operation (Upper Chart: EPLS-T^2; Lower Chart: EPLS-Q).

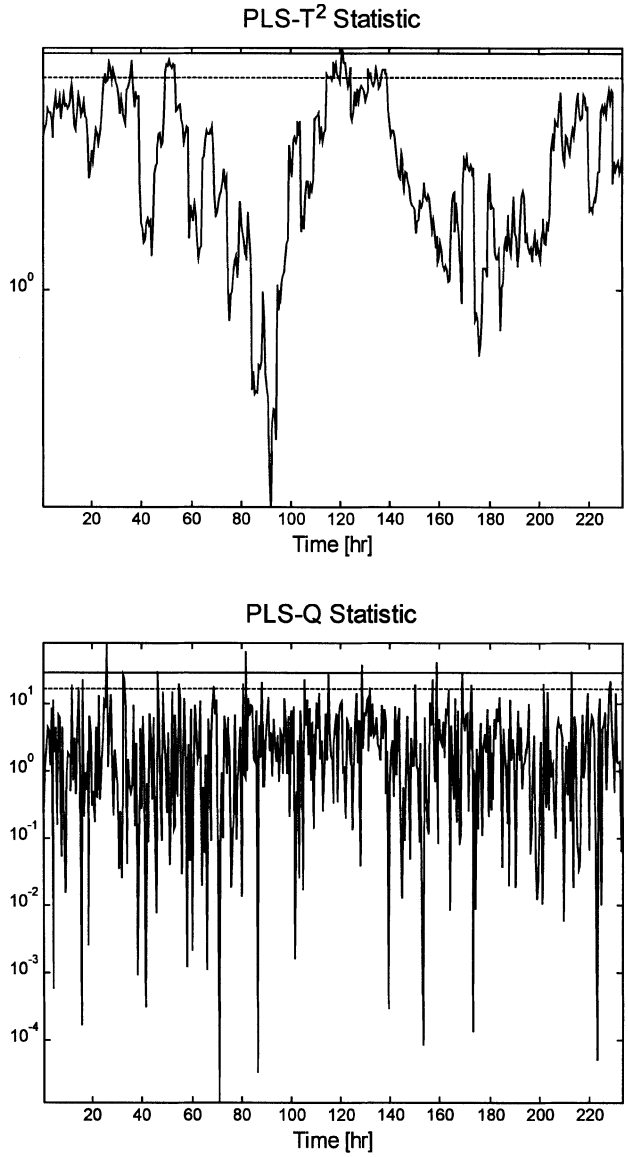
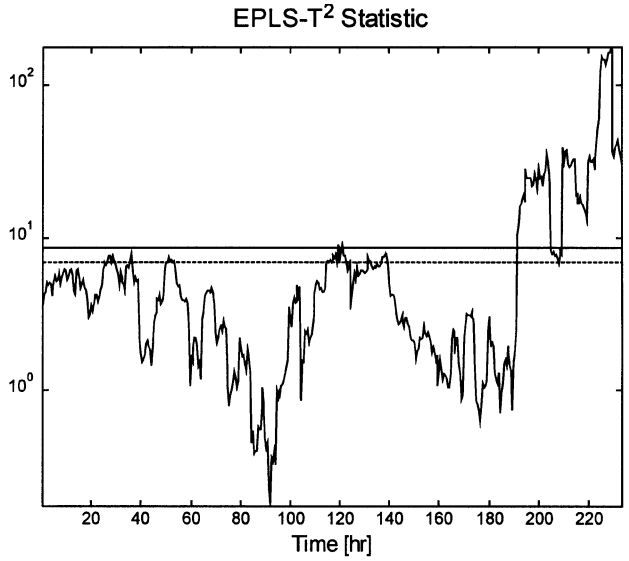


Figure 4. PLS monitoring charts for the unmeasured disturbance at $t = 190.5$ h (Upper Chart: PLS-T^2; Lower Chart: PLS-Q).



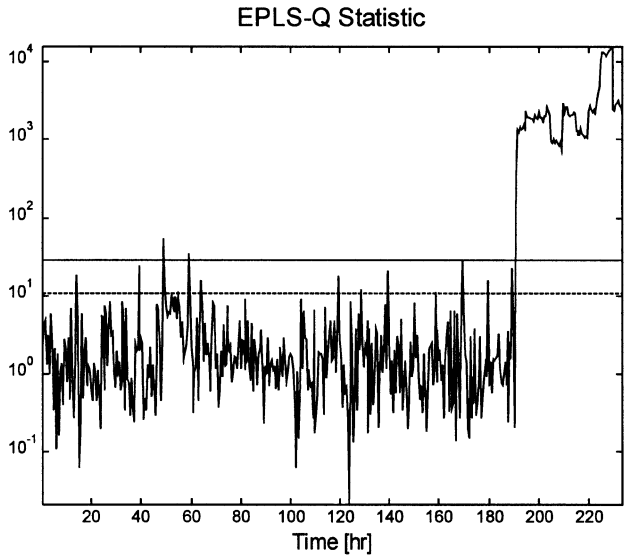


Figure 5. EPLS monitoring charts for the unmeasured disturbance at $t = 190.5$ h (Upper Chart: EPLS-T^2; Lower Chart: EPLS-Q).

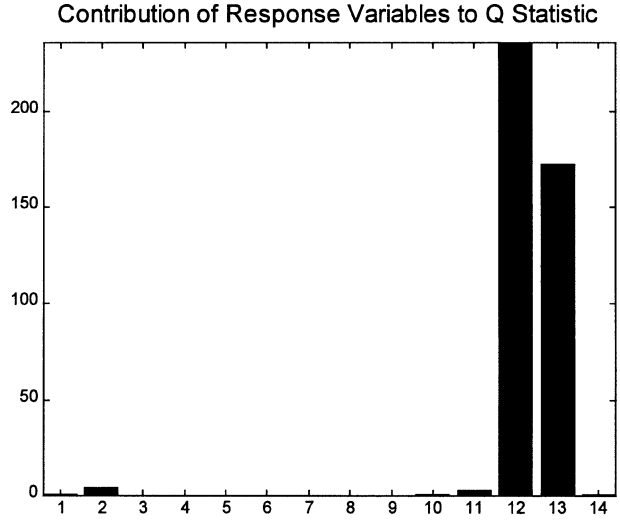


Figure 6. Error contribution chart at $t = 191$ h.

change in riser conditions. This feed composition change does not directly affect any of the defined predictor variables. Neither is there a process path or feedback link between any of the response variables and the predictor variables. Therefore, this disturbance goes completely undetected in the predictor variables, and the T^2 and Q charts based in the basic PLS Approach I provide no indication at all that this disturbance has created an abnormal condition.

In contrast, the EPLS-T^2 and Q statistics plotted in Figure 5 clearly identify the abnormal condition at the 99% confidence level. An error contribution (EC) chart [see Kourti et al. (1995)] corresponding to the time at which this event is apparent (191 h) is shown in Figure 6. An EC chart provides information on what variables most probably contributed to a statistically significant event in the Q chart. Variables 12 and 13, excess oxygen in the flue gas, and concentration of carbon monoxide in the flue gas, respectively, are clearly contributors to the event. This makes physical sense, since a plug of heavier feed will cause a rapid increase in the amount of coke deposited on the catalyst in the riser and transported to the regenerator, having a direct effect on oxygen consumption and production of carbon monoxide. The contribution chart does not point directly to the potential source of the fault, but does provide an experienced plant operator with information that would assist in narrowing down potential causes.

In the second run, the regenerated catalyst fault was applied after 322 h. Again, since the predictor variables all come from the feed section of the unit, a fault or disturbance occurring in either the reactor or regenerator, or the connecting catalyst lines, will have no mechanistic path back to these variables. In this case, the fault only affects response variables, and conventional PLS-T^2 and Q charts will not detect the event. This is demonstrated in Figure 7. However, the EPLS-T^2 and Q charts (Figure 8) clearly detect the abnormal condition after 322.5 h. The corresponding EC chart for the response variables, presented in Figure 9, indicates that excess oxygen in the flue gas and standpipe level in the regenerator have significant contributions. This is easily explained, since any change in flow of regenerated catalyst will affect the material balance in the standpipe, and hence its level. A change in regenerated catalyst flow will also affect catalyst-to-feed ratio in the riser, resulting in a change in the amount of coke deposited on spent catalyst and subsequently the level of oxygen usage in the regenerator.

For runs three and four, the position of the wet-gas compressor suction valve was included as a predictor variable. Thus, any disturbance or fault that affected reactor pressure was transferred to the predictor variable set through the feedback action of the reactor pressure controller. In this case, both PLS and EPLS would be expected to detect an abnormal condition arising from this type of fault. For these runs, application of cross validation resulted in four latent variables.

That both PLS and EPLS are capable of identifying abnormal conditions in this operating configuration is clearly

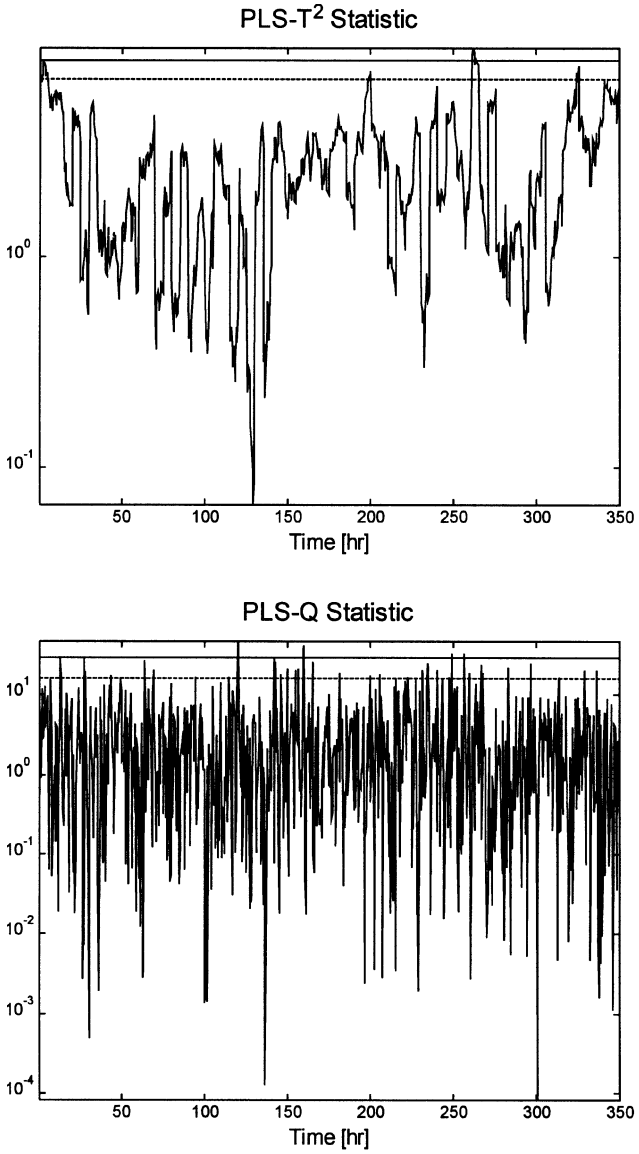


Figure 7. PLS monitoring charts for the regenerated catalyst fault at 322 h (Upper Chart: PLS-T^2; Lower Chart: PLS-Q).

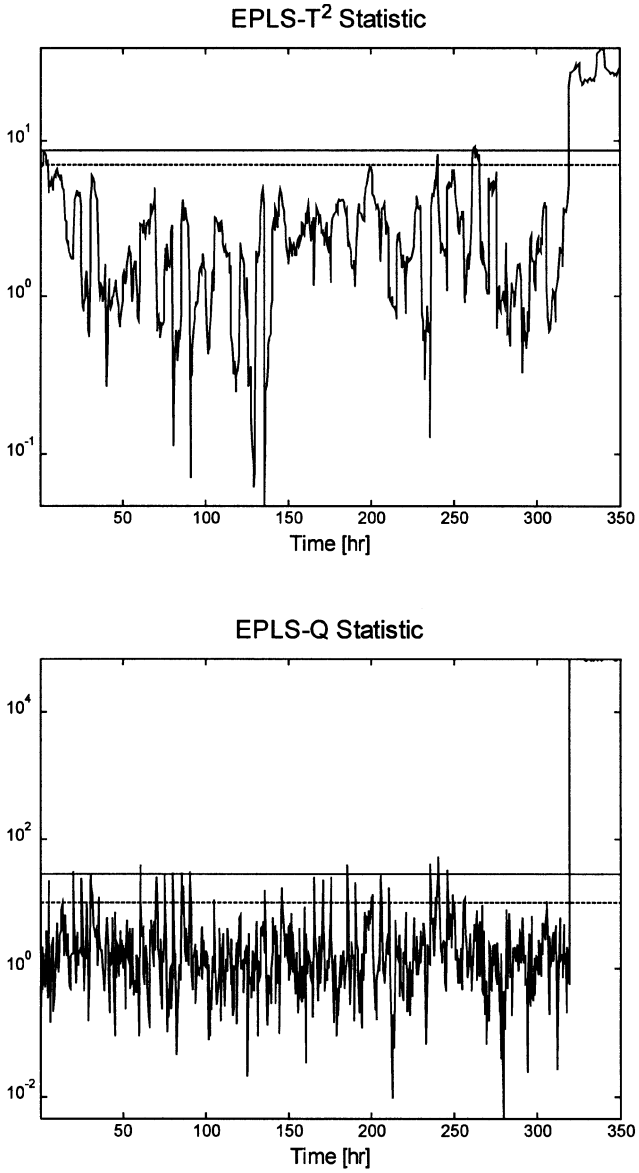


Figure 8. EPLS monitoring charts for the regenerated catalyst fault at 322 h (Upper Chart: EPLS-T^2; Lower Chart: EPLS-Q).

demonstrated in Figures 10 (PLS charts) and 11 (EPLS charts) for the first fault, and Figures 12 (PLS charts) and 13 (EPLS charts) for the second fault.

Note that for these runs, application of conventional PLS Approach II would have required a maximum of 12 charts, rendering it very difficult to monitor.

The application of Approach I in the FCCU case study demonstrated one of its particular weaknesses, namely, that it will not detect abnormal process behavior when only response variables with no feedback are affected. In contrast, the EPLS approach does not suffer from this drawback, and easily detects faults under these conditions.

Fluidized-bed reactor

This industrial process, illustrated in Figure 14, produces two solvent chemicals, denoted as F and G, and consists of several unit operations. The core elements of this plant are five parallel fluidized-bed reactors, each producing compounds F and G by complex exothermic chemical reactions. The reactors are fed with five different reactants. One of the parallel reactor systems is shown in Figure 14.

Streams A, B, and C represent fresh reactant feed supplying pure components A, B, and C, while feedstream D is from an upstream unit. Stream E is plant recycle. Streams D and E are vaporized before entering the reactor. After leaving the reactors, the separation of components F and G is achieved by downstream distillation units.

Each reactor consists of a large shell and a number of vertically oriented tubes in which the chemical reaction is carried out, supported by fluidized catalyst. There is a thermocouple at the bottom of each tube to measure the temperature of the fluidized bed. To remove the heat of the exother-

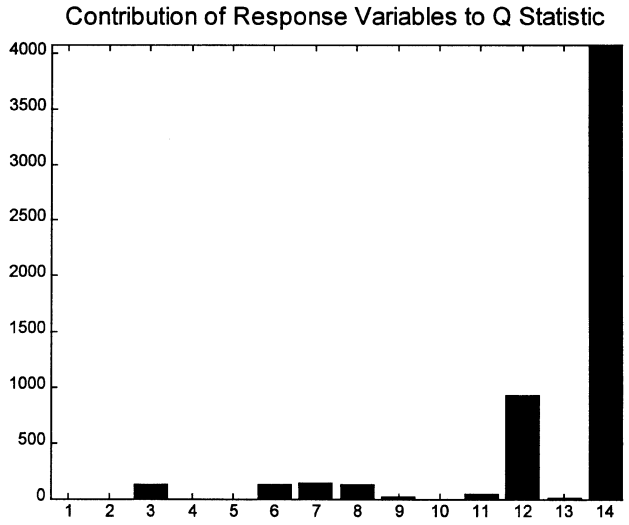


Figure 9. Error contribution chart at $t = 322.5$ hr, for the regenerated catalyst fault

mic reaction, oil circulates around the tubes. The ratio of components F and G is obtained from a lab analysis at 8-h intervals. Based on this analysis, operators adjust the F:G ratio by manipulating reactor feedrates. Furthermore, to keep the catalyst fluidized at all times, the fluidization velocity is maintained constant by adjusting reactor pressure relative to the total flow rate.

The chemical reaction is affected by unmeasured disturbances and changes in the fluidization of the catalyst. A common unmeasured disturbance is caused by pressure upsets in the steam to the vaporizer. Unmeasured disturbances also can be caused by the coolant, which is provided by a separate unit. Fluidization problems appear if the catalyst density is considerably greater at the bottom of the tube, which additionally enhances chemical reaction in the tube, resulting in a significant increase in tube temperature.

During a period of several weeks, normal steady-state operating data as well as data containing process abnormalities were obtained for a single reactor. The data set for capturing normal process operation (reference data set) had to be selected with care, to ensure that the reference data did not capture disturbances as described earlier or fluidization problems of one or more tubes. Conversely, if the size of the reference data set was too small, then normal variation within the system would not be adequately represented. For identifying a steady-state PLS model, the predictor variables were the flow rates of feed streams A, B, D, and E, the steam flow to the varporizer, and the flow of an additional stream required for reducing the pressure in the vaporizer. The response variables were selected as the 35 tube temperatures.

A preanalysis of the data revealed significant correlation between tube temperatures and between the predictor variables. It was determined by cross validation that six LVs were necessary to adequately predict the response variables, the same as the number of predictor variables. This would be considered an unusual situation in most industrial processes where hundreds or even thousands of response and predictor variables would typically be involved in the PLS analysis. Nevertheless, the system represents a real industrial process

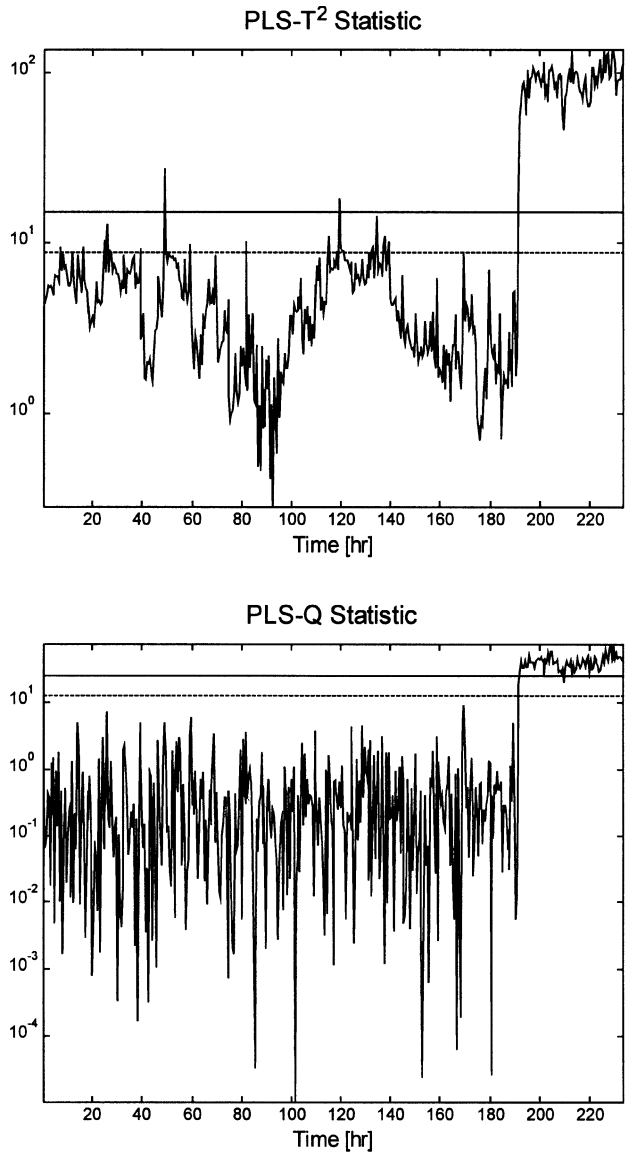


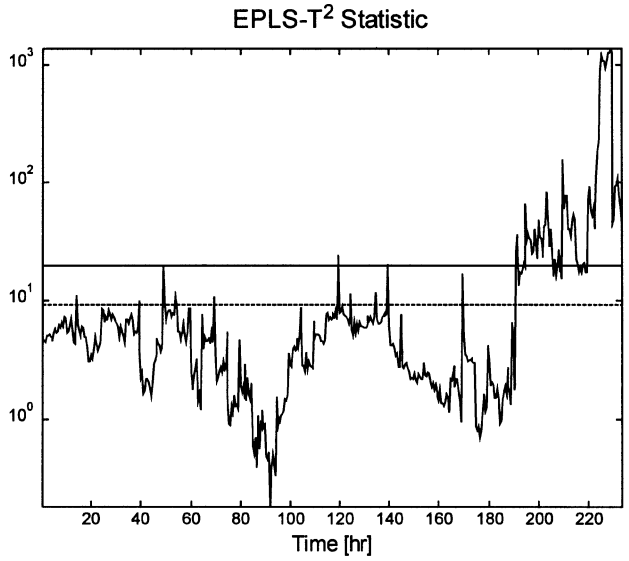
Figure 10. PLS monitoring charts for the unmeasured disturbance at $t = 190.5$ h (Upper Chart: PLS-T^2; Lower Chart: PLS-Q).

that produces data that are difficult to interpret, thereby suggesting it as a candidate for the developed procedures.

Although no reduction in dimensionality of the predictor space could be achieved, PLS did reduce the response space from 30 variables to 6 latent variables. From these 6 LVs, the PLS-T^2, EPLS-T^2, and Q statistics and their corresponding confidence limits for monitoring charts were calculated. Note that the PLS-Q statistic cannot be determined since the number of score vectors was equal to the number of predictor variables, and thus no residuals were calculated.

Normal operating data are represented in Figure 15 (PLS-T^2 chart) and Figure 16 (EPLS-T^2 and Q charts).

The first observed abnormal process event resulted from a large disturbance, a drop in steam pressure supplied to the vaporizer at about 21.5 h into the run. The resultant monitoring charts are shown in Figures 17 (PLS-T^2 chart) and 18



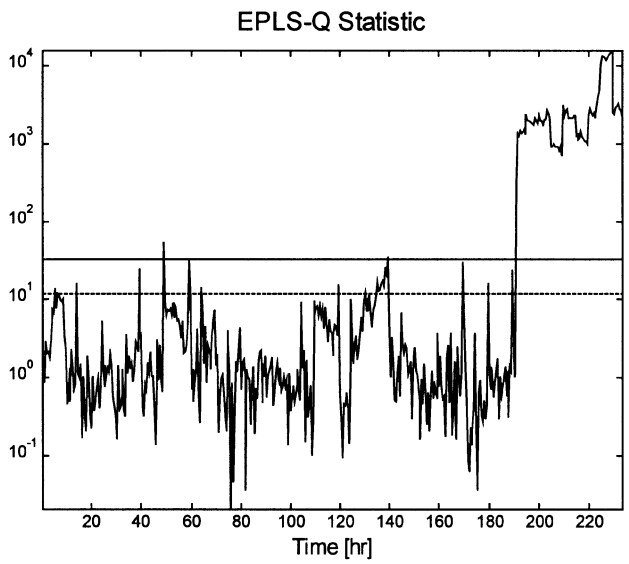
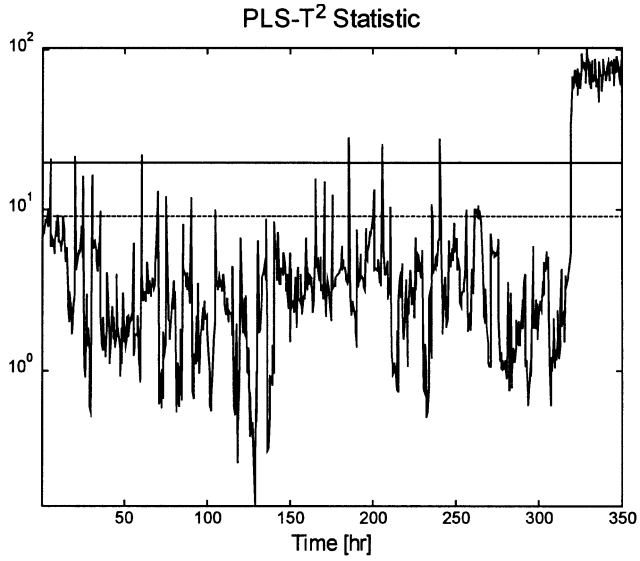


Figure 11. EPLS monitoring charts for the unmeasured disturbance at $t = 190.5$ h (Upper Chart: EPLS-T^2; Lower Chart: EPLS-Q).



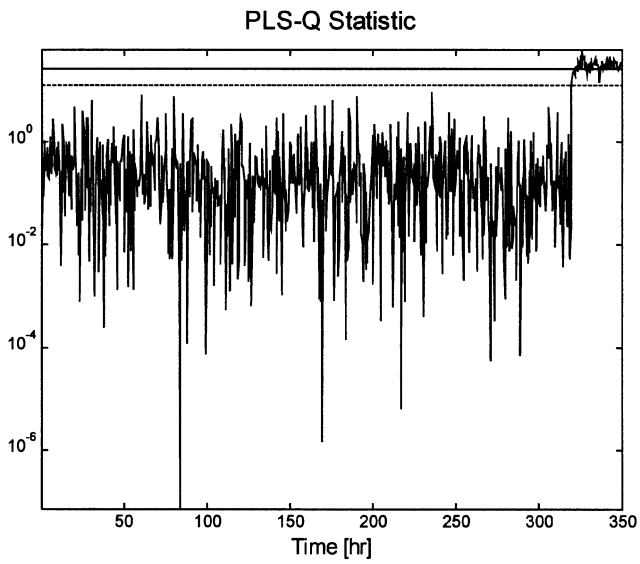
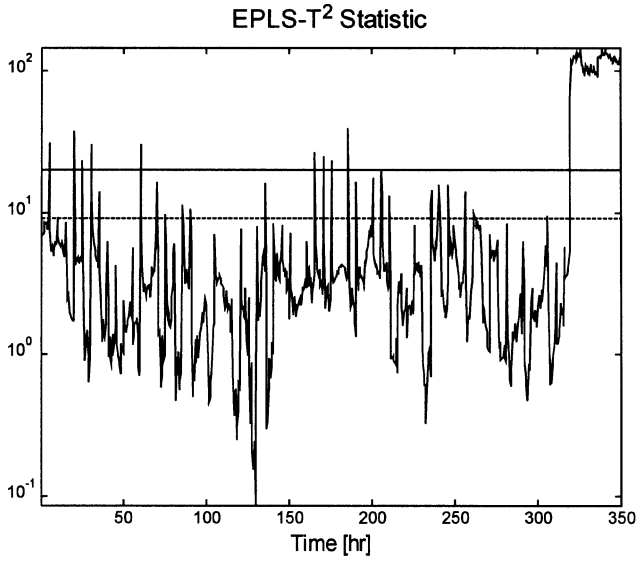


Figure 12. PLS monitoring charts for the regenerated catalyst fault at 322 h (Upper Chart: PLS-T^2; Lower Chart: PLS-Q).

(EPLS-T^2 and Q charts). Although the steam rate remains constant, the enthalpy balance within the vaporizer changes and affects the enthalpy of the reactor feed steam. This change in enthalpy of the feed stream affected none of the predictor variables, and therefore the PLS-T^2 chart (Figure 17) did not detect it. However, it was detected by the EPLS charts (Figure 18).

An error contribution chart at $t = 22$ h (Figure 19) shows that a large majority of the tube temperatures are involved in the event. This is expected because this particular feed stream is fed to all the tubes, and a change in its enthalpy affects the reactions in all of them.

The second abnormal process event to be examined resulted from a fluidization problem in one of the tubes. There are some manipulations that a plant operator can carry out to improve the fluidization and hence bring the tube temperature back to a normal operating level. However, incipient fluidization problems often go undetected by plant operators, as illustrated by this example. Figure 20 shows the PLS-T^2 chart and Figure 21 the EPLS-T^2 and Q charts for the data sequence under study. An abnormal condition developed at approximately 7 h, and was later determined to be caused by an incipient fluidization problem in one of the tubes. The problem was undetected by operators and no corrective action was taken. It should be noted that the PLS and EPLS analyses were performed on historical process data, and that these charts were not available to the operator at the time the events occurred. The event was detected by the PLS and EPLS charts, at the 95% level by the PLS-T^2 chart (Figure 20), and at the 99% level by the EPLS charts (Figure 21). In the case of PLS-T^2 statistic, the event was detected due to the presence of feedback action from a controller that at-



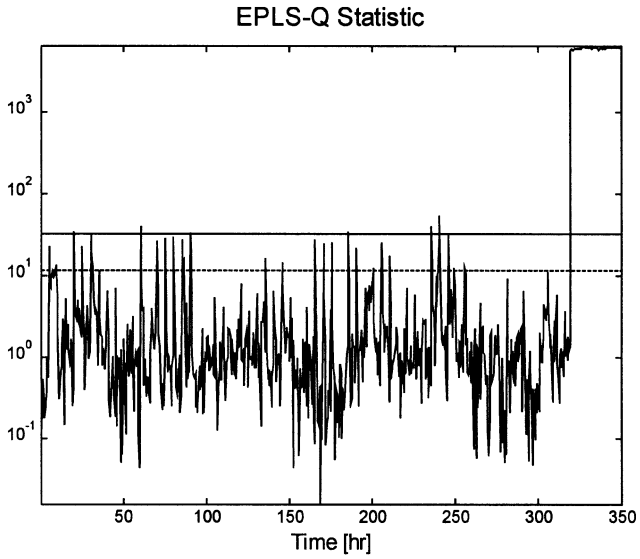


Figure 13. EPLS monitoring charts for the regenerated catalyst fault at 322 h (Upper Chart; EPLS-T^2; Lower Chart: EPLS-Q).

tempts to maintain fluidization conditions in the tubes. The fluidization problem emerged again at around 15 h, this time with greater impact. Operator invention failed to reverse the condition and the tube was eventually shut down. The sharp spike in Figures 20 and 21 at around 15 h correspond to manual operator action, taken to control the runaway reaction conditions. An error contribution chart corresponding to $t =$ 7.3 h (Figure 22) and $t = 15.1$ h (Figure 23) clearly points to tube 11 as the source of the abnormal event. The monitoring and EC charts provided an early indication of the impending fluidization problems in tube 11, and would have allowed an operator to detect an abnormal event that otherwise went undetected.

This industrial example further showed how the conventional PLS Approach I can be insensitive to certain types of process abnormal operating conditions and that EPLS can correct this deficiency.

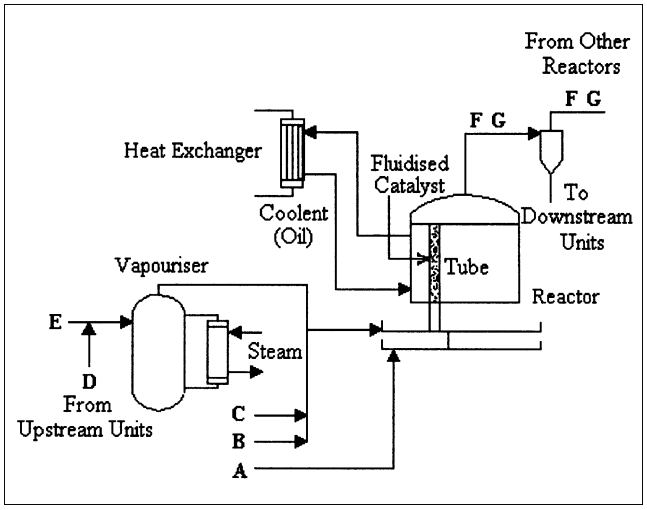


Figure 14. Fluidized-bed reactor and its adjacent units.

Not discussed in these examples was the PLS Approach II. In the industrial example, Approach II would be impractical, since 15 scatter plots, 6 monitoring charts, and 2 SPE charts would be required to comprehensively monitor the process.

Conclusions

In this article, conventional PLS approaches for the condition monitoring of continuous industrial processes, as described in MacGregor et al. (1991), MacGregor and Kourti (1995), Kresta et al. (1991), and Wise et al. (1996) are revisited and problem areas are highlighted. The analysis reveals that the conventional PLS process monitoring charts can be either insensitive in detecting abnormal process behavior, under certain conditions, or too complex to be of practical value. This article presents an extension to the standard PLS algorithm, referred to as EPLS, which leads to the definition of

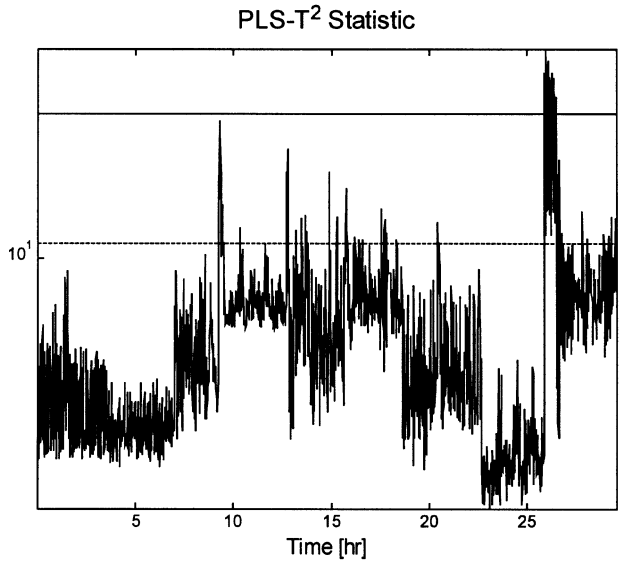


Figure 15. PLS-T^2 chart for normal-reactor operating data.

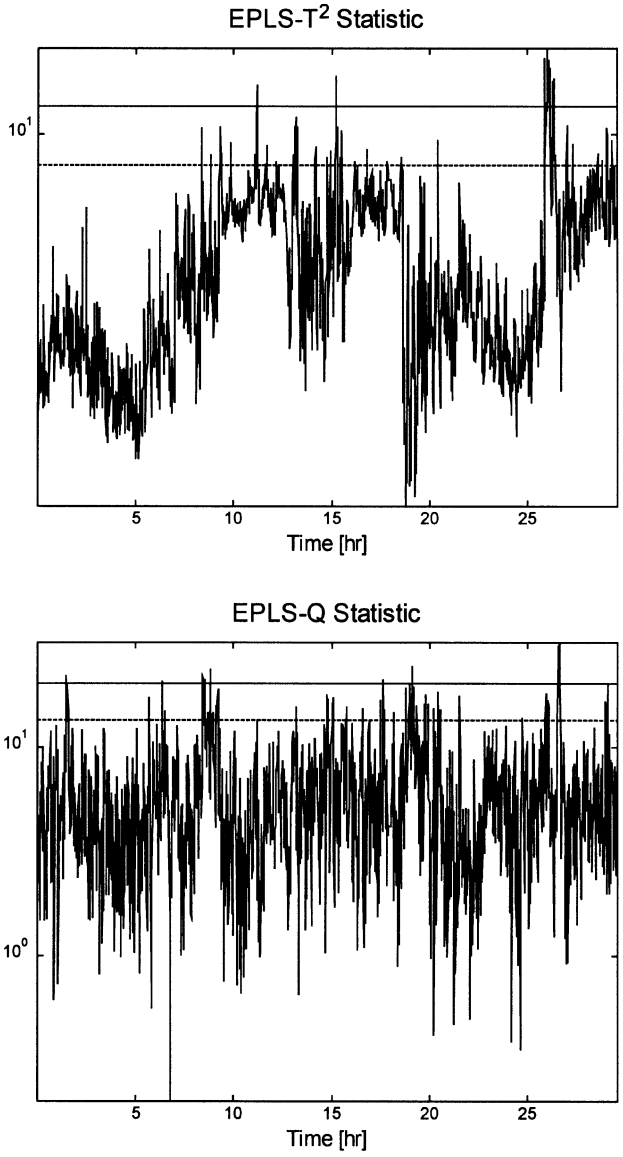


Figure 16. EPLS monitoring charts for normal-reactor operating data (Upper Chart: EPLS-T^2; Lower Chart: EPLS-Q).

new PLS scores, denoted as generalized scores. In similar fashion to conventional PLS approaches, univariate statistics are defined based on the generalized scores that can be plotted vs. time on monitoring charts. These monitoring charts provide information on overall variation in predictor and response variables (EPLS-T^2 chart) and their residuals (EPLS-Q chart). Therefore, the approach is capable of detecting abnormal process conditions that manifest themselves in process response variables as well as predictor variables, even when no process feedback is present. Therefore, it is a more sensitive process monitoring tool than Approach I, based on conventional PLS. Furthermore, since it relies on only two charts based on univariate statistics, it is less complex to interpret than the other common implementation of PLS for process monitoring, namely Approach II as described in this article.

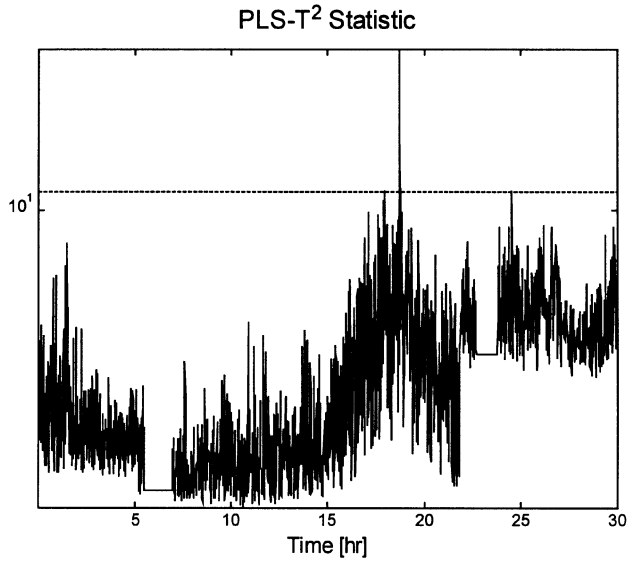


Figure 17. PLS-T^2 chart for steam pressure disturbance at $t = 21.5$ h.

The results of the FCCU and industrial case studies clearly support the conclusions just stated with respect to the advantages of EPLS over the conventional PLS approaches. The examples demonstrated that when feedback was absent, Approach I failed to identify a process fault that affected only process responses, whereas EPLS quickly identified it.

Acknowledgment

The authors thank Invensys Advanced Technology Division for funding aspects of this project. Uwe Kruger is grateful for financial support from the EPSRC. The authors also thank Dr. K. Smith of Simulation Sciences (UK) Ltd. for providing helpful advice in interpreting the process data of the chemical reaction process. They gratefully acknowledge ICI Polymer Chemicals for providing access to the operating data used in this article and for permission to present associated results.

Literature Cited

Box, G. E. P., "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems: Effects of Inequality of Variance in One-Way Classification," *Ann. Math. Stat.*, **25**, 290 (1954).

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

de Jong, S., "SIMPLS, an Alternative Approach to Partial Least Squares Regression," *Chemometrics Intell. Lab. Syst.*, **18**, 251 (1993).

Dunia, R., S. J. Qin, T. F. Edgar, and T. J. McAvoy, "Identification of Faulty Sensors Using Principal Component Analysis," *Process Syst. Eng.*, **42**, 2797 (1996).

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

Golub, G. H., and C. F. Van Loan, *Matrix Computations*, John Hopkins Univ. Press, Baltimore (1996).

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

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

Kaspar, M. H., and W. H. Ray, "Partial Least Squares Modeling as Successive Singular Value Decomposition," *Comput. Chem. Eng.*, **17**, 985 (1993).

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

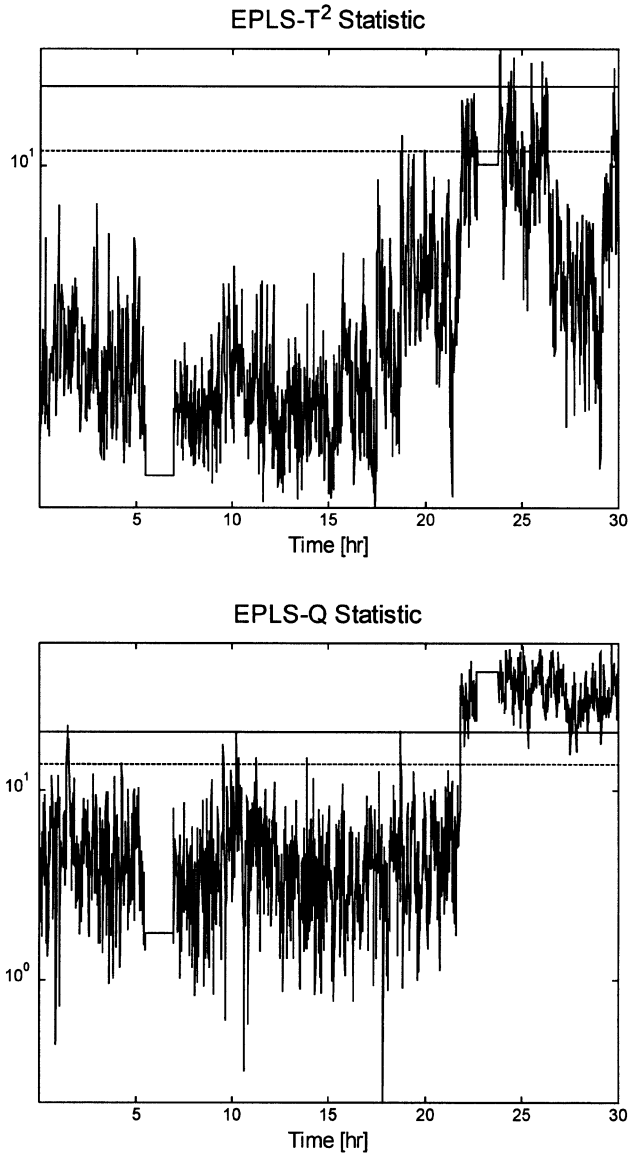


Figure 18. EPLS monitoring charts for the steam-pressure disturbance at $t = 21.5$ h: (Upper Chart: EPLS-T^2; Lower Chart: EPLS-Q).

Kresta, J. V., J. F. MacGregor, and T. E. Marlin, "Multivariate Statistical Monitoring of Process Operating Performance," *Can. J. Chem. Eng.*, **69**, 35 (1991).

Lindgren, F., P. Geladi, and S. Wold, "The Kernel Algorithm for PLS," *J. Chemometrics*, **7**, 45 (1993).

MacGregor, J. F., T. E. Marlin, J. V. Kresta, and B. Skagerberg, "Multivariate Statistical Methods in Process Analysis and Control," *AIChE Symp. Ser.*, **67**, 79 (1991).

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

Manne, R., "Analysis of Two Partial-Least Squares Algorithms for Multivariate Calibration," *Chemometrics Intell. Lab. Syst.*, **2**, 283 (1987).

McFarlane, R. C., R. C. Reineman, J. F. Bartee, and C. Georgakis, "Dynamic Simulator for a Model IV Fluid Catalytic Cracking Unit," *Comput. Chem. Eng.*, **17**, 275 (1993).

Morud, T. E., "Multivariate Statistical Process Control; Example from the Chemical Process Industry," *J. Chemometrics*, **10**, 669 (1996).

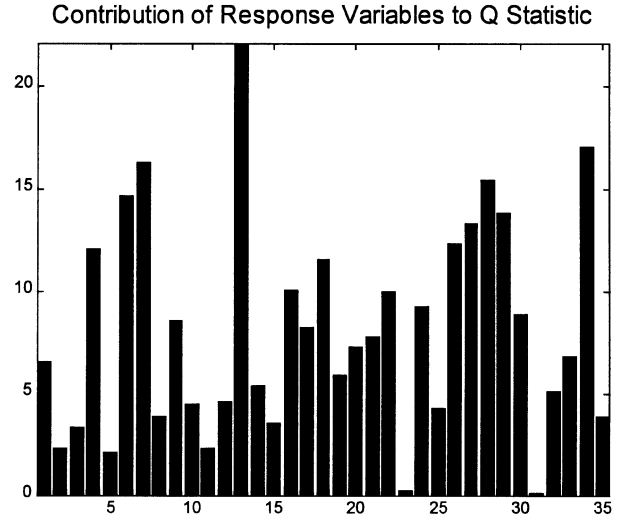


Figure 19. Error contribution chart at 22 h, corresponding to the steam-pressure disturbance.

Nimmo, I., "Adequately Address Abnormal Situation Operations," *Chem. Eng. Prog.*, **91**, 36 (1995).

Wise, B. M., N. B. Gallagher, and J. F. MacGregor, "The Process Chemometrics Approach to Process Monitoring and Fault Detection," *J. Process Control*, **6**, 329 (1996).

Wold, H., *Research Papers in Statistics*, F. David, ed., Wiley, New York, p. 411 (1966a).

Wold, H., *Multivariate Analysis*, P. Krishnaiah, ed., Academic Press, New York (1966b).

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

Appendix A: Summary of the PLS Algorithm

The PLS identification algorithm relies on determining each pair of component matrices, $\tilde{X}_k$ and $\tilde{Y}_k$ (see Eq. 1), by an iterative procedure. After the kth iteration step has been carried out, the calculated component matrices are subtracted from the predictor and the response matrices, respec-

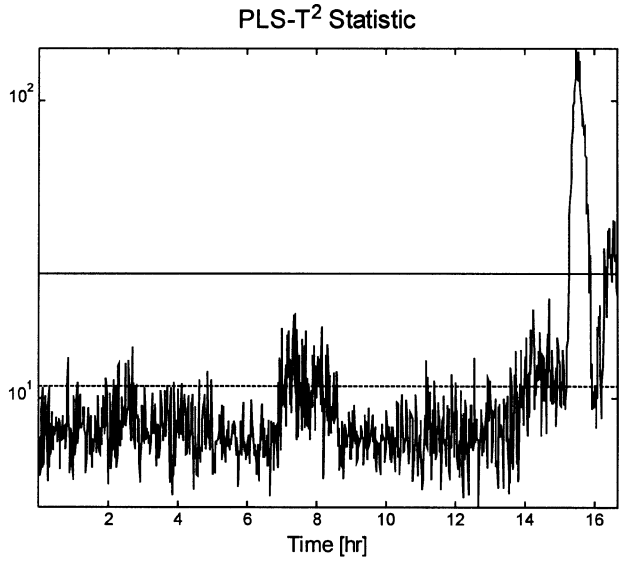
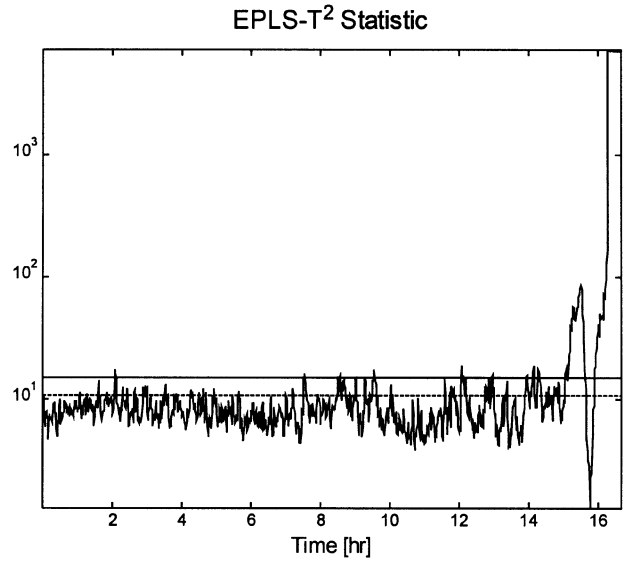


Figure 20. PLS-T^2 chart for the fluidization problem.



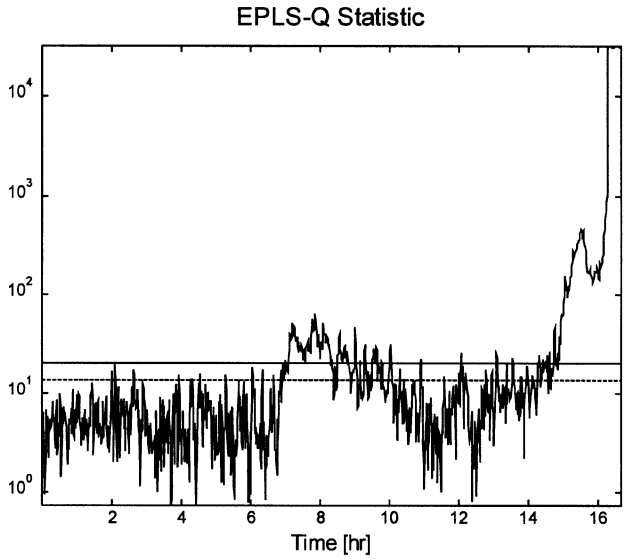


Figure 21. EPLS monitoring charts for the fluidization problem (Upper Chart: EPLS-T^2; Lower Chart: EPLS-Q).

tively, prior to computing the $(k+1)$-th iteration step. The subtraction of the component matrices, or deflation, as it is referred to in the literature, is performed according to

$$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k - \tilde{\boldsymbol{X}}_k = \boldsymbol{X}_k - \boldsymbol{t}_k \boldsymbol{p}_k^T$$

$$\boldsymbol{Y}_{k+1} = \boldsymbol{Y}_k - \tilde{\boldsymbol{Y}}_k = \boldsymbol{Y}_k - \hat{\boldsymbol{u}}_k \boldsymbol{q}_k^T, \tag{A1}$$

where $\boldsymbol{X}_{k+1}$, $\boldsymbol{Y}_{k+1}$ and $\boldsymbol{X}_k$, $\boldsymbol{Y}_k$ represent the predictor and response matrices after $k+1$ and k deflation steps, respectively.

The score vectors, $\boldsymbol{t}_k$, $\hat{\boldsymbol{u}}_k$, and loading vectors $\boldsymbol{p}_k$ and $\boldsymbol{q}_k$ are determined with respect to the following criteria:

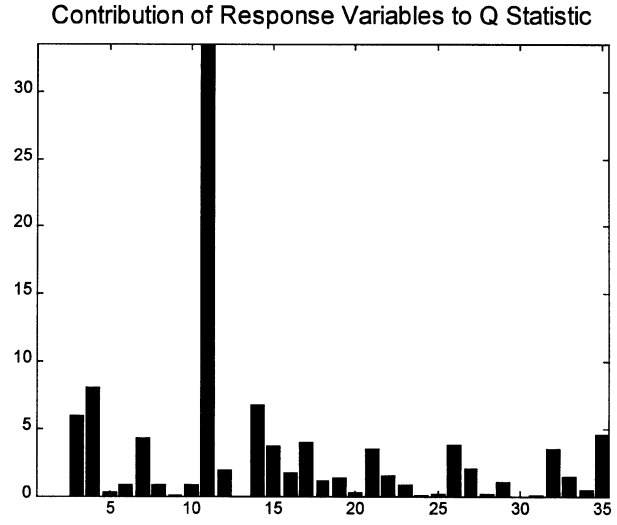


Figure 22. Error contribution chart at $t = 7.3$ h, corresponding to the first occurrence of the fluidization problem.

$$\boldsymbol{t}_k = \boldsymbol{X}_k \boldsymbol{w}_k; \qquad \| \boldsymbol{w}_k \|_2^2 - 1 = 0$$

$$\boldsymbol{u}_k = \boldsymbol{Y}_k \boldsymbol{v}_k; \qquad \| \boldsymbol{v}_k \|_2^2 - 1 = 0$$

$$J_{wv} = \boldsymbol{t}_k^T \boldsymbol{u}_k = \boldsymbol{w}_k^T \boldsymbol{X}_k^T \boldsymbol{Y}_k \boldsymbol{v}_k \tag{A2}$$

$$\boldsymbol{u}_k = b_k \boldsymbol{t}_k + \boldsymbol{e}_k = \hat{\boldsymbol{u}}_k + \boldsymbol{e}_k$$

$$J_e = \boldsymbol{e}_k^T \boldsymbol{e}_k = \boldsymbol{u}_k^T \boldsymbol{u}_k - 2 b_k \boldsymbol{t}_k^T \boldsymbol{u}_k + b_k^2 \boldsymbol{t}_k^T \boldsymbol{t}_k \tag{A3}$$

and

$$J_p = \text{trace}\left\{ \left[\boldsymbol{X}_k - \boldsymbol{t}_k \boldsymbol{p}_k^T \right]^T \left[\boldsymbol{X}_k - \boldsymbol{t}_k \boldsymbol{p}_k^T \right] \right\}$$

$$J_q = \text{trace}\left\{ \left[\boldsymbol{Y}_k - \hat{\boldsymbol{u}}_k \boldsymbol{q}_k^T \right]^T \left[\boldsymbol{Y}_k - \hat{\boldsymbol{u}}_k \boldsymbol{q}_k^T \right] \right\}. \tag{A4}$$

Solution to the three cost functions have to be calculated successively. Beginning with Eq. A2, $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are referred

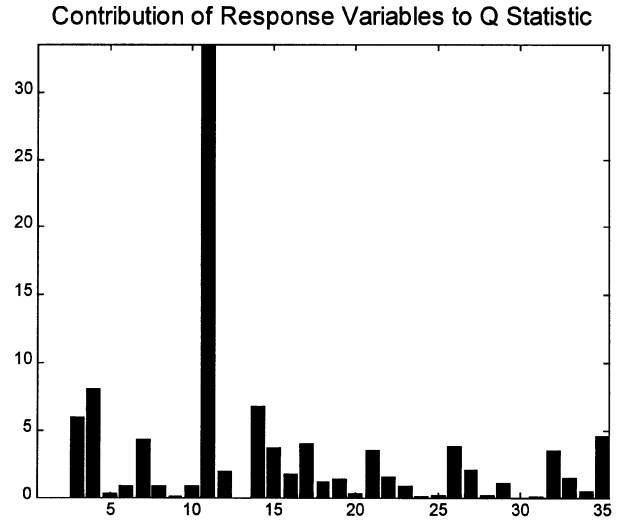


Figure 23. Error contribution chart at $t = 15.1$ h, corresponding to the second occurrence of the fluidization problem.

to as the weight vectors of the predictor and response matries, respectively, and J_{wv} represents the value of the corresponding cost function to be maximized. According to Hoskuldsson (1988), including the constraints on the length of the weight vectors, Equation A2 can also be rewritten as

$$\begin{aligned} \boldsymbol{X}_k^T\boldsymbol{Y}_k\boldsymbol{Y}_k^T\boldsymbol{X}_k\boldsymbol{w}_k &= \lambda_k\boldsymbol{w}_k \\ \boldsymbol{Y}_k^T\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{Y}_k\boldsymbol{v}_k &= \lambda_k\boldsymbol{v}_k, \end{aligned} \tag{A5}$$

where $\boldsymbol{w}_k$ is the eigenvector associated with the largest eigenvalue of the matrix $\boldsymbol{X}_k^T\boldsymbol{Y}_k\boldsymbol{Y}_k^T\boldsymbol{X}_k$ and $\boldsymbol{v}_k$ is the eigenvector corresponding to the largest eigenvalue of the matrix $\boldsymbol{Y}_k^T\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{Y}_k$. Note that $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are also the right and left singular vectors of the matrix $\boldsymbol{X}_k^T\boldsymbol{Y}_k$ (Kaspar and Ray, 1993). In Eq. A3, b_k is the regression coefficient between the kth pair of score vectors, $\boldsymbol{t}_k$ and $\boldsymbol{u}_k$ and J_e is the value of the cost function to be minimized. The solution to Eq. A3 is the ordinary least-squares solution for b_k. For Eq. A4, J_p is the related cost function for determining $\boldsymbol{p}_k$ and J_q is for computing $\boldsymbol{q}_k$. Minimizing J_p and J_q results in the following solution Geladi and Kowalski, 1986):

$$\begin{aligned} \boldsymbol{p}_k &= \frac{\boldsymbol{X}_k^T\boldsymbol{t}_k}{\boldsymbol{t}_k^T\boldsymbol{t}_k} \\ \boldsymbol{q}_k &= \frac{\boldsymbol{Y}_k^T\hat{\boldsymbol{u}}_k}{\hat{\boldsymbol{U}}_k^T\hat{\boldsymbol{u}}_k}. \end{aligned} \tag{A6}$$

Finally, a direct regression relation between $\boldsymbol{Y}$ and $\boldsymbol{X}$ can be derived [see Lindgren et al. (1993) for details].

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X}\cdot\boldsymbol{B}_{PLS}^{(n)} + \boldsymbol{E}_n \\ \boldsymbol{B}_{PLS}^{(n)} &= \boldsymbol{W}_n\left[\boldsymbol{P}_n^T\boldsymbol{W}_n\right]^{-1}\boldsymbol{B}_n\boldsymbol{Q}_n^T, \end{aligned} \tag{A7}$$

where $\boldsymbol{W}_n$, $\boldsymbol{P}_n$, and $\boldsymbol{Q}_n$ are formed by the n column vectors $\boldsymbol{w}_k$, $\boldsymbol{p}_k$, and $\boldsymbol{q}_k$, respectively, and as before n is the number of retained latent variables. According to Eq. A2, the weight vector for the predictor matrix, $\boldsymbol{w}_k$, is multiplied with the deflated predictor matrix, $\boldsymbol{X}_k$, to determine the score vector, $\boldsymbol{t}_k$. Lindgren et al., (1993), however, outlined that the score vector, $\boldsymbol{t}_k$, can also be calculated directly from the original predictor matrix, $\boldsymbol{X}$, as follows:

$$\boldsymbol{t}_k = \boldsymbol{X}\boldsymbol{r}_k; \qquad \boldsymbol{r}_k = \left[\prod_{i=1}^{k=1}\left[\boldsymbol{I}_M - \boldsymbol{w}_i\boldsymbol{p}_i^T\right]\right]\boldsymbol{w}_k$$

and $\qquad \boldsymbol{r}_1 = \boldsymbol{w}_1$

$$\boldsymbol{T}_n = \boldsymbol{X}\left[\boldsymbol{r}_1\cdots\boldsymbol{r}_n\right] = \boldsymbol{X}\boldsymbol{R}_n; \qquad \boldsymbol{R}_n = \boldsymbol{W}_n\left[\boldsymbol{P}_n^T\boldsymbol{W}_n\right]^{-1}, \tag{A8}$$

where $\boldsymbol{I}_M$ denotes the $M\times M$ identity matrix.

Appendix B: Proof Required for Eq. 6

The prediction of the response matrix based on n retained latent variables is as follows:

$$\hat{\boldsymbol{Y}}_n = \boldsymbol{T}_n\boldsymbol{B}_n\boldsymbol{Q}_n^T. \tag{B1}$$

The t-scores can directly be calculated from the original predictor matrices (see Eq. A8):

$$\boldsymbol{T}_n = \boldsymbol{X}_n\boldsymbol{R}_n. \tag{B2}$$

The PLS regression matrix can therefore be determined as

$$\boldsymbol{B}_{PLS}^{(n)} = \boldsymbol{R}_n\boldsymbol{B}_n\boldsymbol{Q}_n^T. \tag{B3}$$

Finally, premultiplication with $\boldsymbol{P}_n^T$ provides the required equality (see Appendix C):

$$\boldsymbol{P}_n^T\boldsymbol{B}_{PLS}^{(n)} = \boldsymbol{B}_n\boldsymbol{Q}_n^T. \tag{B4}$$

Appendix C: Proof Required for Eq. 8

The definition of the matrix $\boldsymbol{R}_n$ (see Eq. A8) can be used to prove that $\boldsymbol{R}_n^T\boldsymbol{P}_n = \boldsymbol{P}_n^T\boldsymbol{R}_n = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ denotes the $n\times n$ identity matrix. The elements of the matrix product are defined as follows:

$$\boldsymbol{p}_i^T\boldsymbol{r}_j = \boldsymbol{p}_i^T\prod_{k=1}^{j-1}\left[\boldsymbol{I}_M - \boldsymbol{w}_k\boldsymbol{p}_k^T\right]\boldsymbol{w}_j; \qquad 1\le i,j\le n. \tag{C1}$$

For i greater than j, $\boldsymbol{p}_i^T\boldsymbol{r}_j = 0$, since $\boldsymbol{p}_i^T\boldsymbol{w}_m = 0$, $1\le m\le j$ (Hoskuldsson, 1988). For i less than j, the factors of the matrix product can be reduced up to the ith factor, resulting in

$$\boldsymbol{p}_i^T\boldsymbol{r}_j = \left(\boldsymbol{p}_i^T - \boldsymbol{p}_i^T\boldsymbol{w}_i\boldsymbol{p}_i^T\right)\prod_{k=i+1}^{j-1}\left[\boldsymbol{I}_M - \boldsymbol{w}_k\boldsymbol{p}_k^T\right]\boldsymbol{w}_j^T. \tag{C2}$$

Since $\boldsymbol{p}_i^T\boldsymbol{w}_i = (\boldsymbol{t}_j^T\boldsymbol{X}_{j-1}\boldsymbol{w}_i)/(\boldsymbol{t}_i^T\boldsymbol{t}_i) = 1$, the first term on the right-hand side is equal to zero, and so $\boldsymbol{p}_i^T\boldsymbol{r}_j = 0$. It follows that $\boldsymbol{p}_i^T\boldsymbol{r}_i = 1$. In summary, Eqs. C1 and C2 demonstrate that the matrix product $\boldsymbol{R}_n^T\boldsymbol{P}_n$ is equal to the $n\times n$ identity matrix.

Appendix D: Orthogonality of both Types of Generalized Score Vectors

In order to prove that both generalized scores are not orthogonal, the covariance matrices for both scores, based on n LVs, are investigated. The covariance matrix of the generalized t-score vectors is given in the Eq. D1.

$$\begin{aligned} \boldsymbol{S}_{T^*T^*}^{(n)} &= \frac{1}{K-1}\left[\boldsymbol{T}_n + \boldsymbol{E}_n^*\right]^T\left[\boldsymbol{T}_n + \boldsymbol{E}_n^*\right] = \boldsymbol{S}_{TT}^{(n)} \\ &\quad + \boldsymbol{S}_{TE^*}^{(n)} + \boldsymbol{S}_{TE^*}^{(n)^T} + \boldsymbol{S}_{E^*E^*}^{(n)} \\ \boldsymbol{S}_{TE^*}^{(n)} &= \frac{1}{K-1}\boldsymbol{R}_n^T\boldsymbol{X}^T\left[\boldsymbol{E}_n\vdots\boldsymbol{F}_n\right]\boldsymbol{C}_{PLS}^{(n)} \\ &= \boldsymbol{R}_n^T\boldsymbol{S}_{XX}^{(n)}\left[\boldsymbol{I}_M - \boldsymbol{R}_n\boldsymbol{P}_n^T\right]\left[\boldsymbol{B}_{PLS}^{(n)}\boldsymbol{B}_{PLS}^{(n)T} + \boldsymbol{I}_M\right]^{-1}\boldsymbol{R}_n \\ \boldsymbol{S}_{E^*E^*}^{(n)} &= \boldsymbol{R}_n^T\left[\boldsymbol{B}_{PLS}^{(n)}\boldsymbol{B}_{PLS}^{(n)^T} + \boldsymbol{I}_M\right]^{-1}\left[\boldsymbol{B}_{PLS}^{(n)}\boldsymbol{S}_{E_nE_n}^{(n)}\boldsymbol{B}_{PLS}^{(n)^T}\right. \\ &\quad \left. + \boldsymbol{S}_{F_nE_n}^{(n)^T}\boldsymbol{B}_{PLS}^{(n)^T} + \boldsymbol{B}_{PLS}^{(n)}\boldsymbol{S}_{F_nE_n}^{(n)} + \boldsymbol{S}_{F_nF_n}^{(n)}\right]\left[\boldsymbol{B}_{PLS}^{(n)}\boldsymbol{B}_{PLS}^{(n)^T} + \boldsymbol{I}_M\right]^{-1}\boldsymbol{R}_n^T, \end{aligned} \tag{D1}$$

where $\boldsymbol{I}_M$ is the $M \times M$ identity matrix. In Eq. D1, $\boldsymbol{S}^{(n)}_{T^*T^*}$, $\boldsymbol{S}^{(n)}_{TT}$, $\boldsymbol{S}^{(n)}_{E^*E^*}$, $\boldsymbol{S}^{(n)}_{XX}$, $\boldsymbol{S}^{(n)}_{E_nE_n}$, and $\boldsymbol{S}^{(n)}_{F_nF_n}$ represent the covariance matrices of the generalized t-scores, the standard t-scores, the generalized residual scores, the predictor variables, the prediction error of the response variables, and the residuals of the predictor variables, respectively. Furthermore, $\boldsymbol{S}^{(n)}_{TE^*}$ denotes the covariance matrix of the standard t-scores and the generalized residual scores, and $\boldsymbol{S}^{(n)}_{F_nE_n}$ is the covariance matrix of the prediction error of the response matrix and the residuals of the predictor matrix. Note that only the covariance matrix $\boldsymbol{S}^{(n)}_{TT}$ is of diagonal type, because the standard t-scores are mutually orthogonal (Hoskuldsson, 1988). According to Eq. D1, it can be concluded that the matrices $\boldsymbol{S}^{(n)}_{E^*E^*}$ and $\boldsymbol{S}^{(n)}_{TE}$ are not generally diagonal, even under the assumption that the columns of the residual matrices, $\boldsymbol{E}_n$ and $\boldsymbol{F}_n$, as well as the predictor matrix $\boldsymbol{X}$ consist of white noise signals. Consequently, the covariance matrix $\boldsymbol{S}^{(n)}_{T^*T^*}$, cannot be diagonal if n LVs are retained. If all M LVs are retained, then Eq. D1 reduces to

$$\boldsymbol{S}^{(n)}_{T^*T^*} = \boldsymbol{S}^{(n)}_{TT} + \boldsymbol{R}_M^T \left[\boldsymbol{B}^{(M)}_{PLS} \boldsymbol{B}^{(M)T}_{PLS} + \boldsymbol{I}_M \right]^{-1} \boldsymbol{B}^{(M)}_{PLS} \boldsymbol{S}^{(n)}_{E_ME_M} \boldsymbol{B}^{(M)^T}_{PLS} \left[\boldsymbol{B}^{(M)}_{PLS} \boldsymbol{B}^{(M)^T}_{PLS} + \boldsymbol{I}_M \right]^{-1} \boldsymbol{R}_M. \quad \text{(D2)}$$

Based on the assumption that $\boldsymbol{E}_n$ consists of white-noise signals, the covariance matrix $\boldsymbol{S}_{E_NE_N}$ is clearly diagonal. Because of the pre- and postmultiplication with generally nondiagonal matrices, however, the result is that $\boldsymbol{S}^{(n)}_{E^*E^*}$, and therefore $\boldsymbol{S}^{(n)}_{T^*T^*}$, are solely of the symmetric type. The interpretation of the Eqs. D1 and D2 has revealed that neither covariance matrix $\boldsymbol{S}^{(n)}_{E^*E^*}$ or $\boldsymbol{S}^{(n)}_{T^*T^*}$ is diagonal. Hence both generalized score types are generally not orthogonal.

Manuscript received Apr. 18, 2000, and revision received Feb. 28, 2001.