Ethylene Compressor Monitoring Using Model-Based PCA

Y. Rotem, A. Wachs, and D. R. Lewin
Wolfson Dept. of Chemical Engineering, Technion I.I.T., Haifa 32000, Israel

Principal component analysis (PCA), which is widely used in process monitoring, performs best when the system variables are linearly correlated. In practice, however the variables are often nonlinearly related and may be subject to periodic forcing, both of which compromise the performance of conventional PCA. In model-based PCA (MBPCA), multivariate statistics are used to analyze the portion of the observed variance that cannot be predicted using a model of the process and thus significantly enhances the attainable diagnostic resolution. Here, MBPCA is used for fault-detection monitoring of an ethylene compressor, which operates under a significant periodic disturbance caused by the ambient temperature. An analytical expression is derived to predict the limits of identifiable faults given bounds on the parametric model uncertainty.

Introduction

Techniques for diagnosis of disturbances and faults during batch processes and continuous processes in which grade changes are prevalent are important in the chemical industry. Commonly, these processes generate large volumes of correlated data, and thus are suitable candidates for analysis using multivariate statistical methods. However, these dynamic nonlinear processes are difficult to monitor using methods such as PCA or PLS, since they generally assume linear correlation among the variables that may compromise abnormality diagnosis. The theory and practice of PCA and PLS are discussed by Jackson (1991) and Kresta et al. (1991).

Applications of nonlinear methods within PCA reported in the literature have been based mainly on neural networks (Kramer, 1991; Dong and McAvoy, 1996a,b) and Genetic Programming (Hiden et al., 1999). The neural network approach has potential overfitting problems, and its application leads to black-box models models that do not lend themselves to interpretation, and when changes are made in the process operating conditions, require the modeling step to be repeated. Monitoring batch process with multiway PCA, as suggested by Wold et al. (1987), Nomikos and MacGregor (1994) and Kosanovich et al. (1996), can be problematic. This method assumes that all batch times are the same, which is very unusual in practice, or employs techniques to deal with different batch times, which invariably leads to a reduction in the fault-detection capability (Rothwell et al., 1998).

Wachs and Lewin (1998) present a model-based PCA (MBPCA) approach that accounts for the nonlinearity of process by using known physical relations among process variables, intended to deal with processes that can be adequately described by first-principles models. A block diagram illustrating the application of MBPCA to a closed-loop process is given in Figure 1. In the method, only the residuals of the raw data that cannot be explained by the model are analyzed using multivariate statistics. In theory, when the model is perfect, the residuals will be relatively insensitive to variations caused from the nonlinearity of the process or changes in operating conditions, thus significantly enhancing process diagnosis capability. Since the model in use may be subject to parametric and structure uncertainty, however, it is important to predict the potential fault detection attainable using a given model. Wachs and Lewin (1998) propose a fault-detection factor calculated by using the measured data. This factor provides a quantitative estimate of the capability of the detection algorithm to diagnose a fault, and its resolution sensitivity to model uncertainty. While an analytical expression of the model uncertainty effect is preferred, the derivation of such an expression is too complicated for most processes. Thus, Monte-Carlo simulations and combinations of simulations and logic investigation of the process (such as Gazi et

Correspondence concerning this article should be addressed to D. R. Lewin.
Present address for Y. Rotem and A. Wachs: Tower Semiconductors Ltd., Migdal HaEmek, Israel.

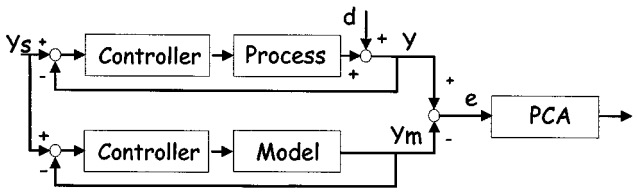

Figure 1. MBPCA principle.

al., 1996) should be used for the estimation of the uncertainty effect. As a consequence of the recommendations afforded by the uncertainty effect analysis, the critical model parameters can be identified and additional effort made to improve their resolution.

In this article, the MBPCA algorithm applied to the diagnosis of an industrial ethylene compressor, and its capabilities displayed using plant data that have been collected from the plant control system over 15 months. This compressor operates under a significant periodic disturbance caused by the ambient temperature, and thus cannot be monitored effectively using conventional PCA. As will be shown, applying MBPCA to the raw data reduces the variance of the residuals and enhances the monitoring resolution. First, the monitoring algorithm is described in detail, and then an analytical expression is derived to quantify the effect of uncertainty on a process subjected to cyclic disturbances. A demonstration of the application of MBPCA for the monitoring of an industrial compressor is then presented, which includes the analysis of three real faults observed over a 15-month period.

The Model-Based PCA Algorithm

The application of MBPCA involves the following steps:

Data acquisition and handling

Two databases should be collected: (1) a set of data describing normal operating conditions (NOC), and (2) a database of previously identified faults. The former is used to detect faults in the plant, and the latter is used to recognize the fault type. Several aspects of NOC data handling should be considered:

- *Selecting the Sampling Rate*. This should be selected to be fast enough to track the expected rate that faults develop, and sufficiently slow to avoid autocorrelation of the data. Selecting excessively high sample rates results in autocorrelated data that prevents analysis by the proposed statistical tools. This is avoided by selecting a sampling interval significantly larger than the characteristic time constants associated with the process and its control system. In the compressor application, although data were available at 15-min intervals, it was resampled at 6-h intervals for use in MBPCA.

- *Determine when data acquisition for NOC should be initiated.* After each maintenance operation, the plant data are abnormal for a period. Experience with compressor indicates that this period may last for several days.

- *Determine the optimal NOC length, n.* While a small NOC data set can cause false alarms, a large set can reduce the fault-detection resolution for slow expected shifts of the system operating point (such as fouling and catalyst decay). Qin (1998) proposes an algorithm to quantitatively determine the database length for PLS modeling, involving a moving window and a forgetting factor to control the age of the data in a database, whose parameters are determined by optimization of a defined objective function derived from the PLS modeling capabilities. This method is designed for modeling purposes, and thus it disregards the effect of NOC age on the fault-detection performance, although it might provide a reasonable solution resulting from an improved model. A more straightforward and direct approach is to determine n by analysis of a known fault, and maximizing the fault-alert time (the time that the fault is detected before the process must be interrupted) while reducing the false-alarm rate to an acceptable level. For the compressor application, a value of $120 < n < 240$ samples (that is, one to two months of data) was empirically found to provide adequate performance. Figure 2 illustrates how this range is selected: the minimum number of samples is selected to be the point where the rate of false alarms is negligible, and alert time becomes unacceptable.

- *Replacement of the NOC data set:* Many maintenance operations, faults, and plant cessation periods require the replacement of the NOC file. The NOC replacement ends the monitoring operation, while keeping an inappropriate NOC data set may cause false alarms.

As illustrated in Figure 3, the data-handling step is repeated each time that a new sample is obtained. As shown, sequence (a) is the number of samples that are discarded immediately after a maintenance operation, since the process is still considered to be abnormal; sequence (c) is a moving win-

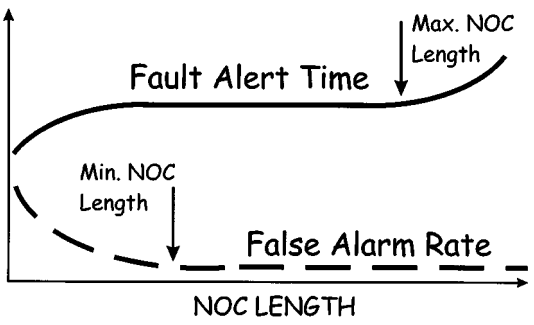

Figure 2. NOC length-selection principle.

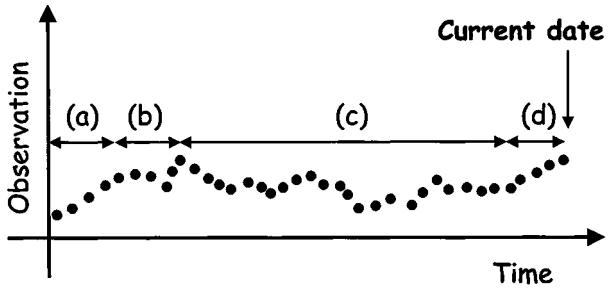

Figure 3. Data acquisition and handling.

(a) Abnormal data immediately after maintenance; (b) old, discarded data; (c) NOC data; (d) monitored data.

dow of samples that are processed as NOC data; sequence (b) is therefore the out-of-date sequence of old data that is discarded; sequence (d) is a sequence of new data that is monitored using MBPCA constructed from the NOC data, which are subsequently inserted into the NOC set if they do not trigger alarms. If a new data point triggers an alarm, its projection is compared to the existing database to attempt its identification. If no such alarm exists in the database, the data point should be identified with the help of plant personnel and then cataloged and stored in the database. Wachs and Lewin (1999) suggest such an approach.

Data manipulation

The dynamic equations describing the process are formulated with as much parametric accuracy as possible, but still simple in a manner that allows on-line calculation of the process variables:

$$\dot{x}_M = f(x,u), \quad y_M = g(x,u). \tag{1}$$

Equation 1 is a generalized unsteady-state nonlinear process model. In practice, this is implemented as a process model of the required precision. In the application treated in this article, a linear pseudo-steady-state model is used to adequately model a compressor.

A matrix Y of output data is collected in normal operation conditions by recording n observations (rows) of m sensors (columns). The matrix of expected values, Y_M is computed at the same time intervals as those of Y by numerical solution of Eq. 1. Next, the residuals matrix, $E=(Y-Y_M)$, is computed. This captures the measurement noise as well as process departures from the model, both due to disturbances and model inaccuracies. The residuals are autoscaled and stored in the $n\times m$ scaled residuals matrix, $\tilde{E}$, with elements:

$$\tilde{e}_{ij} = \left(e_{ij} - \bar{e}_j\right)/\sigma(e_j), \quad j=1,\ldots,m, \quad i=1,\ldots,n. \tag{2}$$

Reduced autocorrelation in the residuals is achieved by using accurate models. The following example demonstrates the effect of the model accuracy over the autocorrelation of the residuals.

Example 1

The measured output, $x_n(k)$ of a discrete first-order exponential filter is described by the following equations:

$$x(k) = \alpha \cdot x(k-1) + (1-\alpha)u(k) \tag{3a}$$

$$x_n(k) = x(k) + p(k), \tag{3b}$$

where $u(k)$ is the input signal, $x(k)$ is the noise-free filter output, $x_n(k)$ is the measured output, $p(k)$ is white noise, and $0<\alpha<1$. For the sake of simplicity, consider the performance of a model of this filter subjected to uncertainty in α. The model prediction is:

$$x_m(k) = \alpha_M \cdot x_M(k-1) + (1-\alpha_M)u(k). \tag{4}$$

Thus, the obtained residuals are

$$e(k) = \left[\alpha \cdot x(k-1) - \alpha_M \cdot x_M(k-1)\right] + \left[\alpha - \alpha_M\right]u(k) + p(k). \tag{5}$$

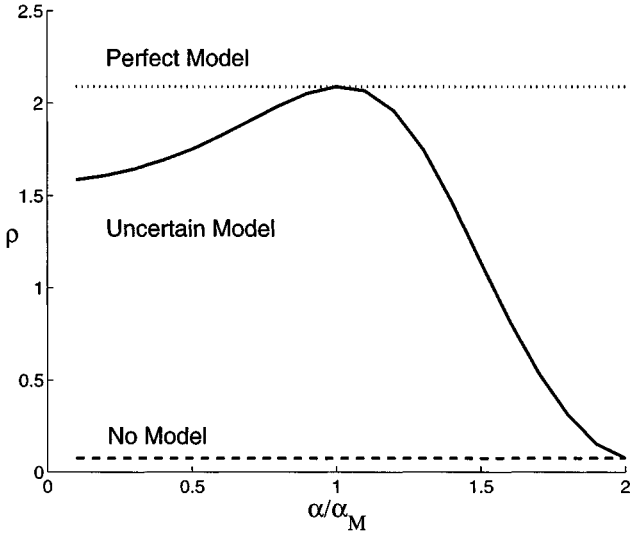

Figure 4. Setup used to show the effect of model inaccuracy on autocorrelation.

The autocorrelation in the scaled residuals is elucidated using the Durbin-Watson test:

$$\rho = \frac{\sum_{k=2}^{n}\left[e(k)-e(k-1)\right]^2}{\sum_{k=1}^{n} e(k)^2}. \tag{6}$$

In Eq. 6, a value of $\rho=2$ implies no autocorrelation. Positive autocorrelation is indicated by values in the range $0\le\rho<2$, whereas negative autocorrelation is implied if $2<\rho\le 4$. Figure 4 illustrates the autocorrelation factor computed for a series of step changes in the filtered signal, as a function of the gain uncertainty. It is noted that the autocorrelation between the input and output signals is almost eliminated using a perfect model, and less so, for cases featuring model inaccuracies.

PCA procedure

The $m\times m$ correlation matrix is formed: $A=\tilde{E}^T\tilde{E}/(n-1)$. The eigenvalues of matrix A are arranged in decreasing order of magnitude, and their corresponding eigenvectors are the *principal component* directions, p_i. The scaled residuals can be expressed in terms of the projected scores, t_{ij}, on the new coordinate system:

$$\tilde{E} = t_1 p_1^T + t_2 p_2^T + \ldots + t_k p_k^T + t_{k+1} p_{k+1}^T + t_m p_m^T. \tag{7}$$

Model-reduction results from truncation of Eq. 7 to k terms, capturing the variability of the data to an acceptable degree. This follows from the fact that the principal component di-

rections are arranged in decreasing contribution to the variance in $\tilde{E}$.

The NOC confidence limits should be derived for each process specifically. Because the distribution of the NOC scores is unknown, a kernel-density estimator (KDE) is recommended. This method is a data-based distribution estimation method (Scott, 1992), and has recently been used in several studies (such as Martin and Morris, 1996; Chen et al., 1996; Robertson et al., 1998). When the scores distribution is normally distributed, Hotelling's T^2 and the Q-statistic, which have computational loads significantly lower than that required by the KDE, are preferred.

On-line process monitoring

New data are monitored against the NOC and its alarm limits. Each new observation vector is subtracted by the predicted values of the model to give a residuals vector, z:

$$z_{rj} = y_{rj} - y_{M,rj}, \quad j = 1, \ldots, m, \tag{8}$$

where r is the current instant. This is autoscaled to make it conform to the principal component space:

$$\tilde{z}_{rj} = \left(z_{rj} - \bar{e}_j \right) / \sigma\left(\bar{e}_j \right), \quad j = 1, \ldots, m \tag{9}$$

$$t = \tilde{z} \cdot [\, p_1 p_2 \cdots p_k \,] \tag{10}$$

An alarm for a shift in the process occurs when the values of scores cross the defined NOC envelope. The suspected fault scores are then compared to the investigated fault database in order to identify the fault type. The maximum and minimum time period from the time stamp of the last measurement in the NOC data set to the occurrence of the new sample are two important degrees of freedom. If this period is too long, false alarms will be triggered because of model uncertainty and slow normal shifts in the system operating points. Conversely, if this period is too short, the failure resolution may be reduced, because data samples exhibit slowly developing faults that can inadvertently be augmented to the NOC data set before the fault is recognized.

Each fault in the database is assumed to be independent, and is logged as a departure from the NOC envelope in a given direction in the scores space. An example of a fault database is given in Wachs and Lewin (1999), who demonstrate the collection of abnormal data for disturbance diagnosis in the Tennessee Eastman process, and its use in disturbance isolation.

Quantifying the Effect of Model Uncertainty

The capability of detecting faults by MBPCA is limited mainly by model uncertainty. This motivates the diagnosis of the potential fault detection as a function of the model uncertainty. A general methodology for quantifying the effect of uncertainty is lacking, with the proposed techniques ranging from analytical methods that are are applicable mostly for linear, low-complexity systems with bounded uncertainty, to statistical methods based on Monte-Carlo simulations that can handle complex multivariate systems. Methods that combine the two types of analyses have been suggested (such as Gazi et al., 1996). Wachs and Lewin (1998) proposed a measure that can be computed analytically or by simulations, according to the system complexity. It relies on the ratio between the mean shift in the scores from NOC to alarm—Δt_j^d—and the range of variability of the scores within NOC—Δt_j. These two measures can be computed for each principal component, j:

$$\Delta t_j = \left(t_{j,\max} - t_{j,\min} \right) / 2, \; j = 1, \ldots, k \tag{11}$$

$$\Delta t_j^d = t_j^d - j_{j,NOC}, j = 1, \ldots, k \tag{12}$$

This definition of Δt_j, as the range between the maximum and minimum score is conservative and may be replaced by other known statistical limits. The fault-detection factor, F_{MB}, is calculated as the Euclidean norm of the scaled perturbations of the scores:

$$F_{MB} = \sqrt{\sum_{j=1}^{k} \left(\Delta t_j^d / \Delta t_j \right)^2}. \tag{13}$$

Values of the F_{MB} indicate which disturbances are potentially easy to identify ($F_{MB} > 1$) and those that are difficult to detect ($F_{MB} < 1$). It is noted that F_{MB} is affected by measurement noise and model uncertainty, both of which tend to increase the magnitude of Δt_j. Furthermore, the measure is limited to normal distributed data. An interpretation of the failure detection factor is given in Figure 5, which illustrates the $F_{MB} = 1$ contour in the scores-place of a two-dimensional system, as well as the F_{MB} coordinates for several known disturbances. Clearly, the disturbances that manifest themselves on the top and lower left of the scores plane are easily identified, while that with $F_{MB} = 0.3$ is hidden within the NOC.

Example 2

Consider the application of MBPCA to monitor a linear system which, in normal operating conditions, is forced by two periodic disturbances. It is seen later that this is a simpli-

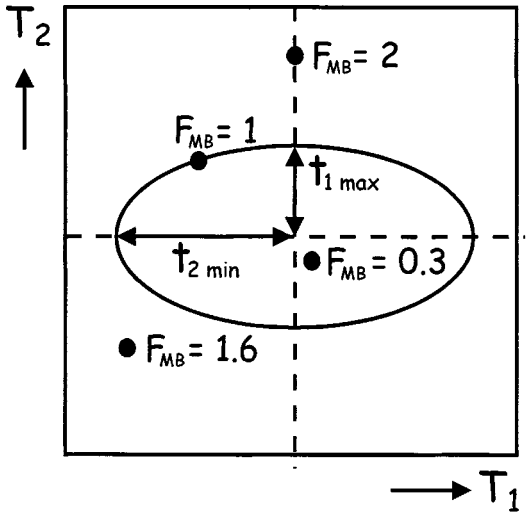

Figure 5. Interpretation of the fault-detection factor, F_{MB}.

fication of the true signals measured from the compressor, which is also characterized by two characteristic frequencies. The system is modeled by the following equations:

$$x_D = c_X + (A_1 + a_1)\sin(w_1 t) + (A_2 + a_2)\sin(w_2 t) \quad (14)$$

$$y_D = f(x_D) + c_y = f\left[c_x + (A_1 + a_1)\sin(w_1 t) + (A_2 + a_2)\sin(w_2 t)\right] + c_y \quad (15)$$

$$\begin{bmatrix} c_x \\ c_y \end{bmatrix} = N_2(\mu, \Sigma), \quad \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix}. \quad (16)$$

In Eqs. 14 and 15, A_1 and A_2 are known coefficients, with a_1 and a_2, respectively, bounded uncertainties; f is a linear function [$f(x) = \alpha \cdot x$], and w_1 and w_2 are two characteristic frequencies. The signals c_x and c_y, representing measurement noise, are normally distributed with zero mean and standard deviations as defined in Eq. 16. MBPCA assumes the following model:

$$x_M = A_1 \sin(w_1 t) + A_2 \sin(w_2 t) \quad (17)$$

$$y_M = f\left[c_x + A_1 \sin(w_1 t) + A_2 \sin(w_2 t)\right]. \quad (18)$$

The resulting residual variables are:

$$\hat{x} = c_x + a_1 \sin(w_1 t) + a_2 \sin(w_2 t) \quad (19)$$

$$\hat{y} = c_y + b_1 \sin(w_1) + b_2 \sin(w_2), \quad b_1 = f(a_1), \quad b_2 = f(a_2). \quad (20)$$

The distribution of the noise is according to Eq. 16; thus, the residuals can be normalized:

$$\tilde{x} = \frac{\hat{x} - \bar{\hat{x}}}{\sigma(\hat{x})} = \frac{c_x + a_1 \sin(w_1 t) + a_2 \sin(w_2 t) - \left[\overline{c_x + a_1 \sin(w_1 t) + a_2 \sin(w_2 t)}\right]}{\sqrt{\frac{1}{n-1}\sum\left\{c_x + a_1 \sin(w_1 t) + a_2 \sin(w_2 t) - \left[\overline{c_x + a_1 \sin(w_1 t) + a_2 \sin(w_2 t)}\right]\right\}^2}}$$

$$= \frac{c_x + a_1 \sin(w_1 t) + a_2 \sin(w_2 t)}{\sqrt{\sigma_x^2 + 0.5a_1^2 + 0.5a_2^2}} \quad (21)$$

$$\tilde{y} = \frac{\hat{y} - \bar{\hat{y}}}{\sigma(\hat{y})} = \frac{c_y + b_1 \sin(w_1 t) + b_2 \sin(w_2 t)}{\sqrt{\sigma_y^2 + 0.5b_1^2 + 0.5b_2^2}}. \quad (22)$$

The principal components of a two-dimension correlation matrix are known to be

$$p = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \Rightarrow t_1 = \frac{\tilde{x} + \tilde{y}}{\sqrt{2}}, \quad t_2 = \frac{\tilde{x} - \tilde{y}}{\sqrt{2}} \quad (23)$$

The maximum and minimum scores are expected in known angles and the noise limits are assumed as $\pm 3\sigma$; thus, NOC limits can be computed:

$$\Delta t_1 = \left\{ \frac{1}{\sqrt{2}} \left(\frac{c_x + a_1 \sin(w_1 t) + a_2 \sin(w_2 t)}{\sqrt{\sigma_x^2 + 0.5a_1^2 + 0.5a_2^2}} + \frac{c_y + b_1 \sin(w_1 t) + b_2 \sin(w_2 t)}{\sqrt{\sigma_y^2 + 0.5b_1^2 + 0.5b_2^2}} \right) \right\}_{\max}$$

$$= \left(\frac{1}{\sqrt{2}} \left(\frac{3\sigma_x + a_1 + a_2}{\sqrt{\sigma_x^2 + 0.5a_1^2 + 0.5a_2^2}} + \frac{3\sigma_y + b_1 + b_2}{\sqrt{\sigma_y^2 + 0.5b_1^2 + 0.5b_2^2}} \right) \right) \quad (24)$$

$$\Delta t_2 = \left\{ \frac{1}{\sqrt{2}} \left(\frac{c_x + a_1 \sin(w_1 t) + a_2 \sin(w_2 t)}{\sqrt{\sigma_x^2 + 0.5a_1^2 + 0.5a_2^2}} - \frac{c_y + b_1 \sin(w_1 t) + b_2 \sin(w_2 t)}{\sqrt{\sigma_y^2 + 0.5b_1^2 + 0.5b_2^2}} \right) \right\}_{\max}$$

$$= \left(\frac{1}{\sqrt{2}} \sum \left| \frac{a_i}{\sqrt{\sigma_x^2 + 0.5a_1^2 + 0.5a_2^2}} - \frac{b_i}{\sqrt{\sigma_y^2 + 0.5b_1^2 + 0.5b_2^2}} \right| + \frac{3\sigma_x}{\sqrt{\sigma_x^2 + 0.5a_1^2 + 0.5a_2^2}} + \frac{3\sigma_y}{\sqrt{\sigma_y^2 + 0.5b_1^2 + 0.5b_2^2}} \right). \quad (25)$$

Finally, the fault-detection factor can be formulated by substituting Eqs. 24 and 25 into Eq. 13:

$$F_{MB}=\left\{\left[\Delta t_1^d\Big/\frac{1}{\sqrt{2}}\left(\frac{3\sigma_x+a_1+a_2}{\sqrt{\sigma_x^2+0.5a_1^2+0.5a_2^2}}+\frac{3\sigma_y+b_1+b_2}{\sqrt{\sigma_y^2+0.5b_1^2+0.5b_2^2}}\right)\right]^2\right.$$
$$\left.+\left(\Delta t_2^d\Big/\frac{1}{\sqrt{2}}\sum\left|\frac{a_i}{\sqrt{\sigma_x^2+0.5a_1^2+0.5a_2^2}}-\frac{b_i}{\sqrt{\sigma_y^2+0.5b_1^2+0.5b_2^2}}\right|+\frac{3\sigma_x}{\sqrt{\sigma_x^2+0.5a_1^2+0.5a_2^2}}+\frac{3\sigma_y}{\sqrt{\sigma_y^2+0.5b_1^2+0.5b_2^2}}\right)^2\right\}^{1/2}. \quad (26)$$

As a numerical example, the fault-detection factor is calculated for two sets of uncertainty values. Table 1 summarizes the parameter values, and Figure 6 displays the principal component scores for the two cases. It is noted that the T^2 limits are somewhat conservative, and indicate that the scores are not multivariate normally distributed. The numerical example emphasizes the general observation that fault-detection potential is strongly dependent on failure direction. The detection of Fault 1, which manifests itself in the first principal axis (marked as ●), is highly dependent on the uncertainty, and thus, reducing the uncertainty enhances its detection. In contrast, the detection of Fault 2, which is aligned in the second principal axis (marked as ■), is indifferent to the uncertainty and affected mostly by the noise.

Applying MBPCA to Ethylene Compressor Monitoring

Problem definition

The primary ethylene compressor is an important unit operation in the polyethylene manufacturing process. It is used to compress gaseous ethylene from 15 to 250 atm and is a part of a sequence of three compression steps, which compresses the ethylene up to 1300 atm. The primary compressor contains three reciprocating compressors (''compression stages'') equipped with two interstage heat exchangers. The ambient temperature periodically effects the compressor data, thus the data is time-dependent and its variance is mostly due to this disturbance. As shown in Figure 7, which presents data for the outlet temperature of the first compression stage, the system response consists primarily of two characteristic frequencies.

Table 1. Fault Detection Indices for Two Sets of Uncertainty Cases

	Variable	Case 1 (High Uncertainty)	Case 2 (Low Uncertainty)
Δt^d	Fault 1(●)	$\Delta t^d=(1, 2)$	
	Fault 2(■)	$\Delta t^d=(0.1, -0.2)$	
	$a_1=a_2$	1	0.2
	$b_1=b_2$	2	0.4
	c_x	0.1	0.1
	c_y	0.15	0.15
F_{MB}	Fault 1(●)	0.44	2.08
	Fault 2(■)	0.38	0.34

An attempt to monitor the compressor using PCA directly will lead to unavoidable difficulties because of the oscillatory nature of the measured data. This is best illustrated by applying PCA on simulated output that is a linear function of the input plus an oscillatory disturbance representing one year of typical plant data. Figure 8 shows the input and output data for a normal period of operation (samples 0–80) and for a fault affecting the gain between the input and output (samples 80–100). Selecting short or long NOC lengths will result in one of the two following scenarios.

1. An alarm system constructed based on an NOC envelope generated using data taken over a short period of operation will trigger many false alarms, since even normal data taken from a different season will be very different from the NOC data. The consequences of this scenario are illustrated in Figure 9a. It is noted that even normal data in the first 75% of the data recorded trigger alarms when only the first 15% of the raw data are used to generate the NOC limits.

2. In contrast, an alarm system based on an NOC generated using data taken over a year's operation will be largely insensitive to true alarms. This is because of the large variance of the process outputs as they respond to the large swings in ambient temperature. The consequences of this scenario are illustrated in Figure 9b. It is noted that faulty data that

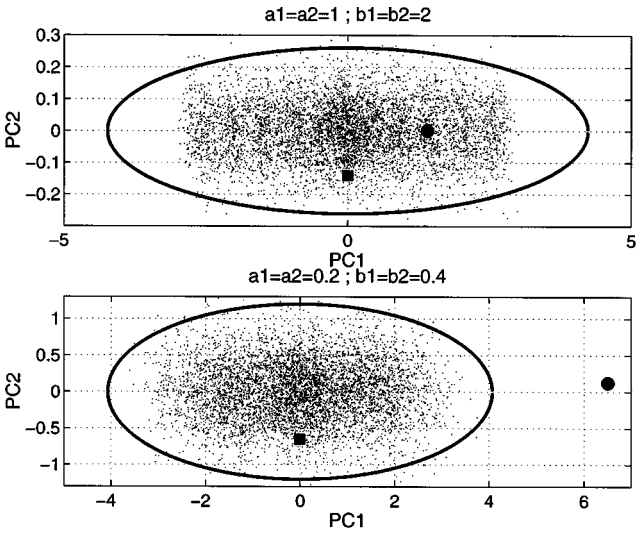

Figure 6. Principal components of the simulated example.

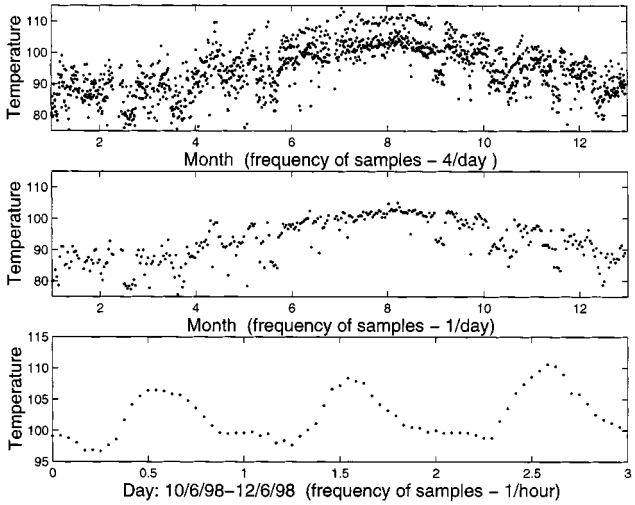

Figure 7. Outlet temperature of the first compression state.

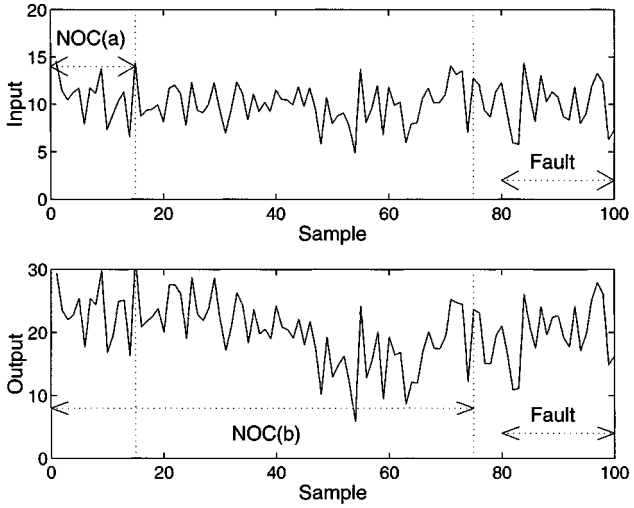

Figure 8. Example data to illustrate problems with PCA.

occur in the last 20% of the data set are difficult to isolate when the PCA NOC data set is taken from the first 75% of the raw data.

A decentralized analysis (Georgakis et al., 1996; Wachs and Lewin, 1999) is employed to focus the fault detection on the problematic compression stage. To filter out the effect of disturbances, a model-based approach is employed. For each compression stage, the discharge temperature at steady state can be calculated according to the compression ratio:

$$T_{\text{out}} = T_{\text{in}}(P_{\text{out}}/P_{\text{in}})^{(r-1)/r}, \qquad r \le C_p/C_v. \tag{27}$$

This relationship is almost linear in the range of operations of each stage, and thus the model is simplified:

$$T_{\text{in},i} = f(d_i, P_{\text{out},i-1}, T_{\text{out},i-1}) + e_{T_{\text{in}}} \tag{28}$$

$$T_{\text{out},i} = T_{\text{in},i} + k(P_{\text{out},i} - P_{\text{in},i}) + e_{T_{\text{out}}} \tag{29}$$

$$P_{\text{in},i} = g(d_i, P_{\text{out},i-1}, T_{\text{out},i-1}) + e_{P_{\text{in}}} \tag{30}$$

$$P_{\text{out},i} = P_{\text{in},i} + \Delta P_i + e_{P_{\text{out}}} \tag{31}$$

$$k = h(P_{\text{in},i}, P_{\text{out},i}, T_{\text{in},i}) \approx \text{Const.} \tag{32}$$

In the preceding, i is indicative of compressor stage i. When the inlet data are subtracted from the model prediction, a linear model is obtained:

$$\hat{T}_{\text{in},i} = T_{\text{in},i} - T_{\text{in},i} = 0 \tag{33}$$

$$\hat{T}_{\text{out},i} = T_{\text{out},i} - T_{\text{in},i} = \Delta T_i = k\Delta P_i \tag{34}$$

$$\hat{P}_{\text{in},i} = P_{\text{in},i} - P_{\text{in},i} = 0 \tag{35}$$

$$\hat{P}_{\text{out},i} = P_{\text{out},i} - P_{\text{in},i} = \Delta P_i. \tag{36}$$

As a consequence of using this model-based approach, each compressor stage is modeled by two variables, and disturbances caused by the ambient temperature and the previous compression stages are almost totally filtered. The use of the residuals $[\Delta T_i, \Delta P_i]$ for each stage automatically removes all periodic components from the data, and since the resulting model is linear, PCA can be used directly. The value of the k parameter is uncertain. However, simulation of this parameter in the region of the operating points using HYSYS, indicates that the uncertainty is only few percent. Figure 10 illustrates the MBPCA procedure for the second compressor stage.

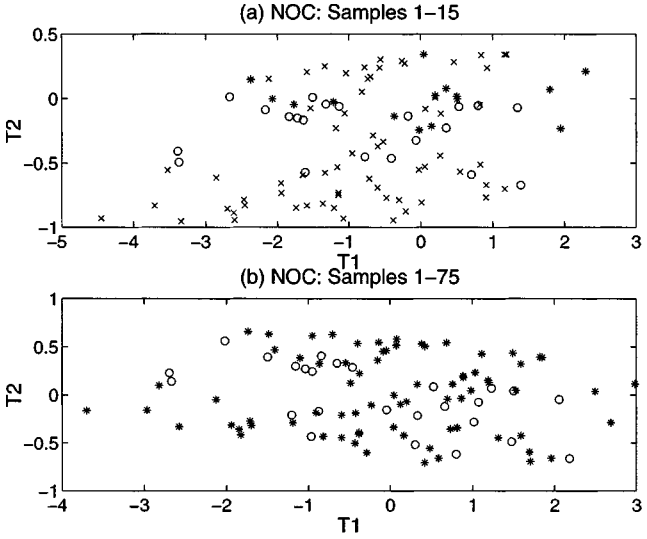

Figure 9. Consequences of applying PCA directly on periodically perturbed data.

(a) Scenario I: The first third of the normal data used to generate NOC; (b) Scenario II: All normal data used to generate NOC envelope. Legend: ∗—normal data used for NOC; ×—normal data not used for NOC; ○—faulty data.

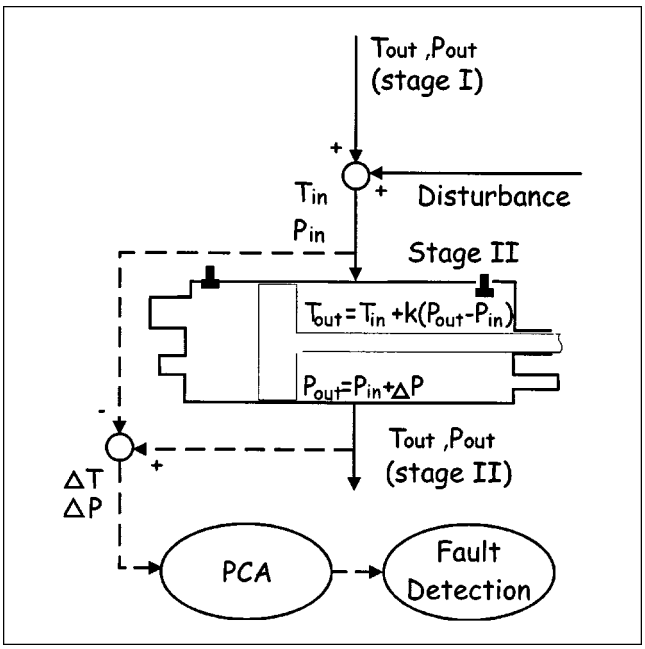

Figure 10. MBPCA procedure for a single compressor stage (Stage II).

Operating data from the compressor was made available, including 15 months of measurements, collected every 15 min, as well as detailed descriptions of all faults and maintenance operations. The data were filtered from recording faults and sampled every 6 h by taking each 24th recorded sample. This frequency was selected to avoid autocorrelated data, and mindful of the expected rate of fault development (over weeks). In retrospect, perhaps twice the sampling rate used would have been desirable to speed up the NOC-generation stage after each maintenance. The MBPCA algorithm as described was implemented off-line to analyze the data.

The NOC minimum length has been empirically estimated to be at least one month (about 120 data points, at a 6-h sample rate). In practice, however, the NOC data set had to be reinitialized every 2 to 3 months due to maintenance operations and faults. Consequently, the fault-detection system monitored the compressor for only half of the total time. This may be typical for statistical fault-detection systems, and one of their major disadvantages. In this application, the residuals are multivariate normal distributed, and thus, the Hotelling T^2 was used to determine the NOC confidence limits.

Performance of MBPCA in failure prediction

The compressor experienced three different valve faults during the monitoring period. The results in off-line detection performance for each fault are described next.

Analysis of Compressor Failure on 8/16/98. The main events in the fault detection, in chronological order, were:

5/10/98 (Day 130): NOC data-set acquisition initiated.

7/2/98 (Day 183): Fault is detected by MBPCA in the **first** compression stage.

7/11/98 (Day 192): The plant operators report a reduction in the compressor performance, probably because of fault in the **second** compression stage.

8/16/98 (Day 228): The compressor is stopped due to the failure of valves in the **first** compression stage. On investigation of the compressor internals, it is discovered that all three of the outlet valves and two (of three) of the inlet valves were faulty.

This fault developed slowly for over a month, and the monitoring system could detect it in advance and locate the problematic compression stage. A graphical display of the monitoring system was designed to display the fault-detection status. The MBPCA analysis performed on each compressor stage is displayed separately. Each plot presents the analysis results for the last 30 samples of scores (about a week of operation), with the latest sample result shown on the right, and the oldest result shown on the left. A bar in the graph displays the T^2 value for each sample, with the value of T^2 above 9 considered to represent a fault. Figures 11–14 display the monitoring system a few days before the fault detection, at the day of the fault detection, and a few days after the fault detection, and when plant operators first noticed a problem. In Figure 11, showing the situation on 6/25/98 (Day 176), the compressor appears to be operating within NOC, with a false alarm having been recorded in one sample approximately 6 days previously. On 7/1/98 (Day 182), as shown in Figure 12, two consecutive high alarms have been registered in Stage 1, echoed in Stage 2. Four days later, the graphical display (Figure 13) indicates a sustained alarm sta-

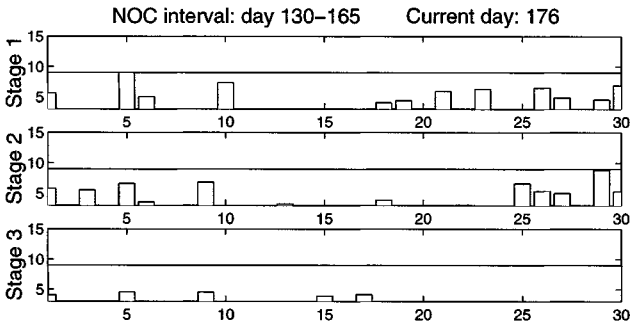

Figure 11. Monitoring system display status six days before the detection of the 8/16/98 fault.

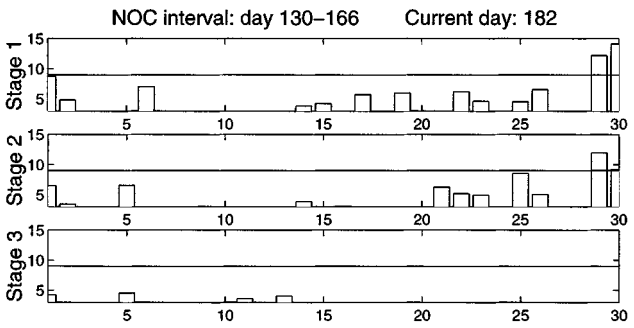

Figure 12. Monitoring system display status at the day of the detection of the 8/16/98 fault.

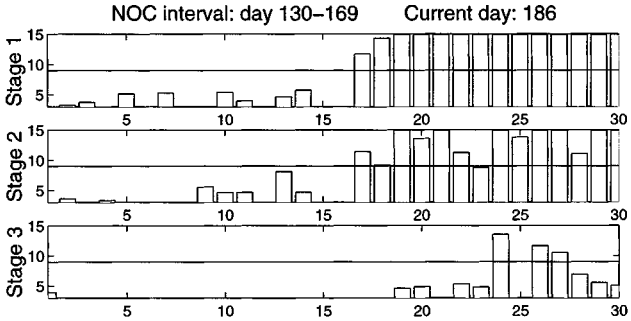

Figure 13. Monitoring system display status four days after the detection of the 8/16/98 fault.

tus on Stage 1 in every sample since 7/1/98 (Day 182). For comparison, Figure 14 illustrates the situation on 7/11/98 (Day 192), when operators first noticed the evidence of abnormalities, and concluded falsely that there was a possible malfunction in Stage 2.

The ability to use only two variables allows the first stage data to be displayed in two dimensions: Figure 15 (top) displays the residuals, while Figure 15 (bottom) shows the scores. The NOC samples are displayed as dots and the fault samples are displayed as open circles. It can be seen that the fault is slowly advancing along the second principal-component axis, and thus it is easily detected.

Analysis of Compressor Failure on the 11/1/98. The main events in the fault detection, in chronological order, were:

8/27/98 (Day 239): NOC data-set acquisition initiated.

10/31/98 (Day 304): Fault was detected by MBPCA in the first compression stage.

11/1/98 (Day 305): The compressor is stopped due to failure of several valves in the first compression stage. On investigation of the compressor internals, it was discovered that one of the inlet valves ruptured, and the resulting debris damaged the other five valves.

This fault developed within a few days. Consequently, the monitoring system alarm was late and was of little use. Figure 16 shows the status of the monitoring system display at

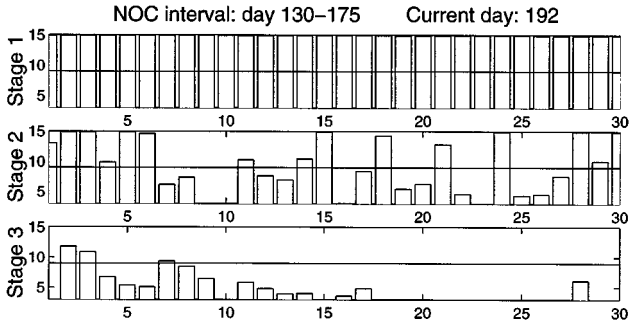

Figure 14. Monitoring system display status on the day the operators misinterpreted the source of the 8/16/98 fault.

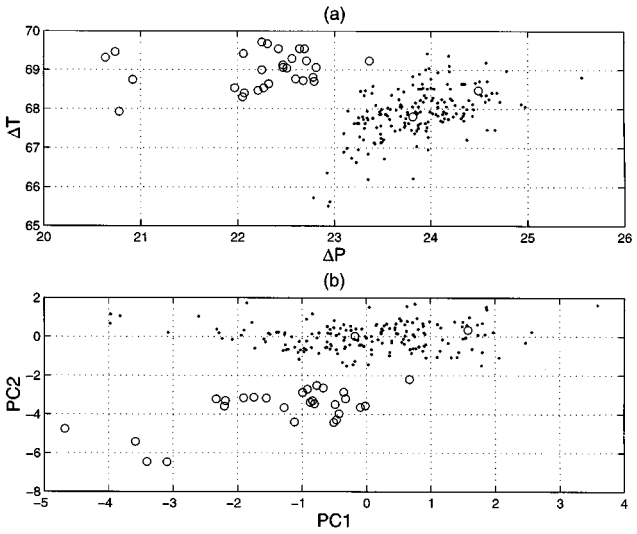

Figure 15. Data for the 8/16/98 fault.

Raw data (top); principal components (bottom): ┆—NOC, ○—faulty data.

the day of the fault detection, and Figure 17 describes the residuals and the scores of the data. It can be seen that the fault data are advancing in the first principal-component direction, which contains most of the variance, and thus the fault detection is delayed.

Analysis of Compressor Failure on the 3/12/99. The main events in the fault detection, in chronological order, were:

1/12/99: NOC data-set acquisition initiated.

3/12/99: The compressor is stopped due to outlet-valve failure in the third compression stage. The plant was shut down by the interlock safety system.

The monitoring system is neither capable nor designed to detect such rapidly developing faults, in this case, one that developed in seconds. The three different faults demonstrate that MBPCA carried out on the compressor data, which was collected every 6 h, can identify faults that develop slowly. However, faults that manifest themselves suddenly cannot be identified in advance. Figure 18 shows the compressor scores in the last two months before the 3/12/99 failure. Clearly, there is no significant difference between NOC and faulty

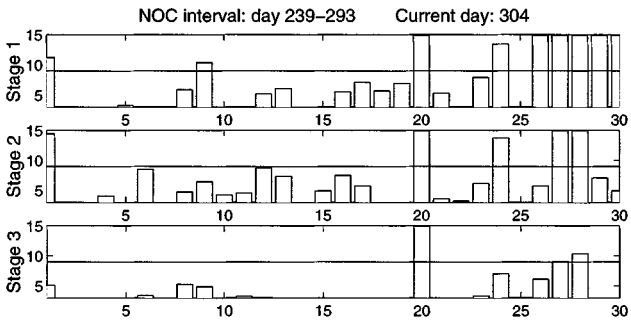

Figure 16. Monitoring system display at the day of the detection of the 11/1/98 fault.

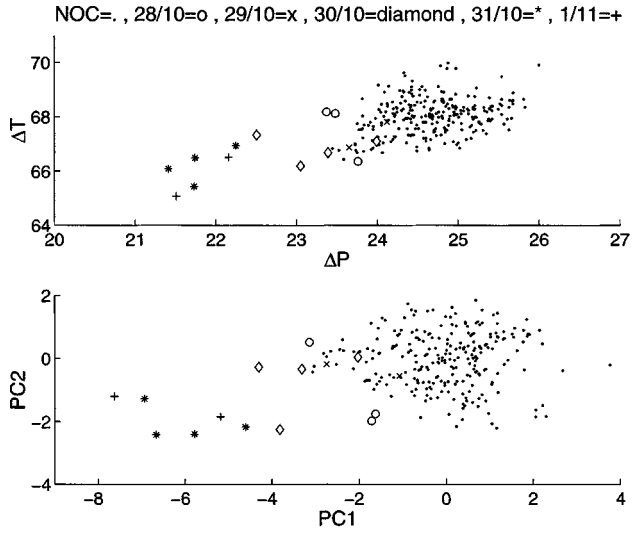

Figure 17. Raw data (top) and principal components (bottom) for the 11/1/98 fault.

scores. Either the fault has developed faster than the detection system's sample rate, or the measurements are not sufficient to isolate the fault in question.

Diagnosis of the potential fault detection

The fault-detection factor (F_{MB}) for the compressor can be computed analytically for each of the compression stages. Consider the case of the first stage, with variables scaled according to the estimated distribution of the pressures and the parameter k:

$$\Delta P = N(23.85,\ 0.5), \qquad k = N(2.85,\ 0.05), \qquad (37)$$

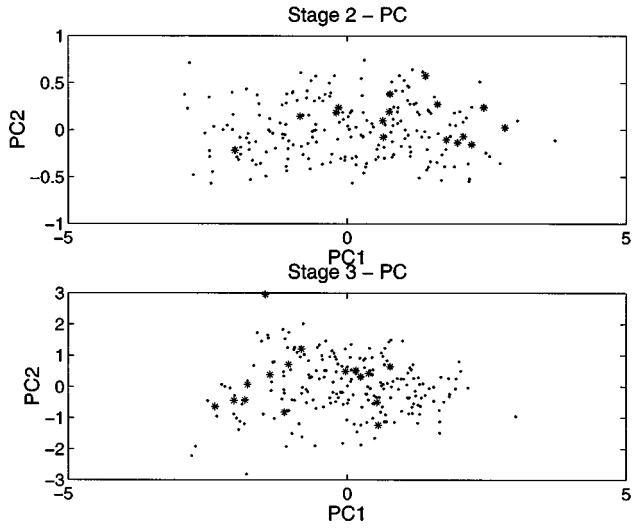

Figure 18. Compressor data for the last two months before the 3/12/99 fault (·—1/12/99–3/6/99; ∗—3/6/99–3/12/99).

where $N(\mu, \sigma)$ is a normal distribution, with μ the mean and σ the standard deviation. As a consequence of the distributions in Eq. 37, the expected values of the residuals $\Delta\tilde{P}$ and $\Delta\tilde{T}$ are

$$\Delta\tilde{P} = (\Delta P - 23.85)/0.5 \qquad (38)$$

$$\Delta\tilde{T} = \frac{k\Delta P}{\sigma(k\Delta P)} = \frac{k\Delta P - 2.85\cdot 23.85}{\sqrt{\left(\sigma^2(k)+\bar{k}^2\right)\left(\sigma^2(\Delta P)+\Delta\bar{P}^2\right)-\left(\bar{k}\cdot\Delta\bar{P}\right)^2}} = \frac{k\Delta P - 68}{1.85}. \qquad (39)$$

Since the principal component directions of a 2×2 correlation matrix are known and constant, the scores of the NOC data can be formulated and the limits of each score can be calculated according to $\pm 3\sigma$ confidence limits:

$$t_{1,\max} = \max\left(\frac{1}{\sqrt{2}}\left(\frac{\Delta P - 23.85}{0.5} + \frac{k\Delta P - 68}{1.85}\right)\right) = 5.2 \qquad (40)$$

$$t_{1,\min} = \min\left(\frac{1}{\sqrt{2}}\left(\frac{\Delta P - 23.85}{0.5} + \frac{k\Delta P - 68}{1.85}\right)\right) = -5.0 \qquad (41)$$

$$t_{2,\max} = \max\left(\frac{1}{\sqrt{2}}\left(\frac{\Delta P - 23.85}{0.5} - \frac{k\Delta P - 68}{1.85}\right)\right) = 1.9 \qquad (42)$$

$$t_{2,\min} = \min\left(\frac{1}{\sqrt{2}}\left(\frac{\Delta P - 23.85}{0.5} - \frac{k\Delta P - 68}{1.85}\right)\right) = -1.8 \qquad (43)$$

$$\Delta t_1 = 0.5(5.2 + 5.0) = 5.1 \qquad (44)$$

$$\Delta t_2 = 0.5(1.9 + 1.8) = 1.85. \qquad (45)$$

Thus, for the first compressor stage, the fault detection factor is

$$F_{MB} = \left\{\left[\frac{1}{\sqrt{2}\cdot 5.1}\left(\frac{\Delta P^d - 23.85}{0.5} + \frac{\Delta T^d - 68}{1.85}\right)\right]^2 + \left[\frac{1}{\sqrt{2}\cdot 1.85}\left(\frac{\Delta P^d - 23.85}{0.5} - \frac{\Delta T^d - 68}{1.85}\right)\right]^2\right\}^{1/2}. \qquad (46)$$

Figure 19 displays the calculated fault-detection factor for the first compression stage, for the case of the 8/16/98 fault. The predicted results indicate that for low uncertainty in the model (as is the case here), the fault detection is not compromised.

Since the temperatures and pressure measurements are assumed to be normally distributed, the T^2 statistic has been used to detect abnormal samples. The numerical results from Eqs. 40–43 apparently contradict this assumption, however, since the NOC limits are clearly not symmetrically placed. This is a consequence of accounting for uncertainty in k, which leads to a skewing of distribution of the principal com-

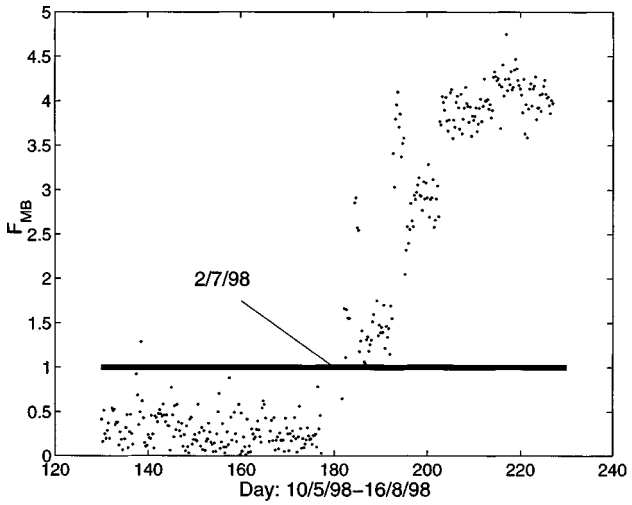

Figure 19. Calculated fault detection factor for the 8/16/98 fault.

ponent scores. In this application, the uncertainty is limited to a few percent only, leading to only a minor distortion, and thus justifying the use of the T^2 statistic. The use of KDE is recommended in cases with uncertainty that are more significant (Rotem and Lewin, 2000).

Conclusions

All data-based methods for failure diagnosis require the appropriate selection of the sampling rate, and the setting up of an appropriate method of false-alarm isolation. In contrast, any model-based method involves the construction of a model of sufficient accuracy to provide the desired performance, as well as its calibration and validation. Since MBPCA is a combination of the two, it requires all of these issues to be addressed. However, this article indicates that for the application considered, good performance can be achieved using a relatively simple model.

MBPCA can deal with batch processes, since the method is robust to a different batch length, and nonlinear processes, since it can handle the nonlinearity in the process, and it's relatively insensitive to changes in the operation point. The method is intended for processes whose dynamics can be described using first-principles models. This task might be problematic for large-scale processes, but in many situations, the formulation of such a model would be desirable in any case for other applications, such as control design, optimization, and sizing. Small-scale processes with simple mechanistic models and relatively monotonic dynamics as well suited for the application of MBPCA. For diagnosis of transient responses, the improved PCA methods introduced by Wachs and Lewin (1999) can be readily incorporated into the MBPCA framework.

Since uncertainty in the model affects the MBPCA data, the distribution of the scores are unknown and a nonparametric method is generally needed to estimate the confidence limits for normal operating conditions. The kernel-density estimator can be used for this purpose. It can be stated that in any case of model uncertainty, the MBPCA normal operation conditions limits should first be computed by KDE. The uncertainty in the model is a major disadvantage, since uncertainty in the model can have a crucial effect on the performance of the proposed method. Simulations and analytical tools should be used to estimate the effect of the uncertainty over the fault-detection resolution (Rotem and Lewin, 2000).

This article has demonstrated that MBPCA can be used to successfully monitor an ethylene compressor. The compressor is affected by a periodic disturbance, and thus it cannot be monitored adequately by standard PCA. The compressor model can be easily described using a simple model, and as has been demonstrated, the nonlinear components of the compressor data can be easily filtered. Although the compressor model is very simple, it demonstrates the ability of MBPCA to monitor industrial processes. The monitoring system described here accounts for some of the important decisions that must be considered by the plant operators in order to use the monitoring system effectively. For the compressor model, computing the effect of parametric uncertainty is straightforward, and the resulting expressions used to analyze the ability of the system to detect different faults.

Acknowledgments

This research was supported by Carmel Olefins Ltd. (COL). The authors are appreciative of the cooperation of COL personnel, and especially Dr. Michael Hoffer.

Notation

e = unexplained part of the data
k = T/P ratio constant
x = state variable vector
y = output data vector
y_M = expected data vector
σ = standard deviation

Superscripts and subscripts

d = fault data
T = matrix transpose
$\sim$ = standardized vectors
$\wedge$ = perturbation variables
D = raw data
M = expected (modeled) data

Literature Cited

Chen, J., A. Bandoni, and J. A. Romagnoli, "Robust PCA and Normal Region in Multivariate Statistical Process Monitoring," *AIChE J.*, **42**, 3563 (1996).

Dong, D., and T. J. McAvoy, "Batch Tracking via Nonlinear Principal Component Analysis," *AIChE J.*, **42**, 2199 (1996a).

Dong, D., and T. J. McAvoy, "Nonlinear PCA Based on Curves and Neural Network," *Comput. Chem. Eng.*, **20**, 65 (1996b).

Gazi, E., L. H. Ungar, and W. D. Seider, "Automatic Analysis of Monte-Carlo Simulations of Dynamic Chemical Plants," *Comput. Chem. Eng.*, **20**, S987 (1996).

Georgakis, C., B. Steadman, and V. Liotta, "Decentralized PCA Charts for Performance Assessment of Plant-Wide Control Structures," *Proc. Triennial World IFAC Congr.*, p. 5 (1996).

Hiden, H. G., M. J. Willis, M. T. Tham, and G. A. Montague, "Non-Linear Principal Components Analysis Using Genetic Programming," *Comput. Chem. Eng.*, **23**, 413 (1999).

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

Kosanovich, K. A., K. S. Dahl, and M. J. Piovoso, "Improved Process Understanding Using Multiway Principal Component Analysis," *Ind. Eng. Chem. Res.*, **35**, 138 (1996).

Kramer, M. A., ''Nonlinear PCA using Autoassociative Neural Networks,'' *AIChE J.*, **37**, 233 (1991).
Kresta, J. V., J. F. MacGregor, and T. E. Marlin, ''Multivariate Statistical Monitoring of Process Operating Performance,'' *Can. J. Chem. Eng.*, **69**, 35 (1991).
Martin, E. B., and A. J. Morris, ''Non-Parametric Confidence Bounds for Process Performance Monitoring Charts,'' *J. Process. Control*, **6**, 349 (1996).
Nomikos, P., and J. F. MacGregor, ''Monitoring Batch Processes Using Multiway PCA,'' *AIChE J.*, **40**, 1361 (1994).
Qin, S. J., ''Recursive PLS Algorithms for Adaptive Data Modeling,'' *Comput. Chem. Eng.*, **22**(4/5), 503 (1998).
Robertson, T., J. Chen, J. A. Romagnoli, and B. Newell, ''Intelligent Monitoring for Quality Control in Biological Nutrient Removal Wastewater Treatment Pilot Plant,'' *Proc. IFAC Symp. on Dynamics and Control of Process Systems*, Corfu, p. 590 (1998).
Rotem. Y., and D. R. Lewin, ''Assessing the Impact of Parametric Uncertainty on the Performance of Model-Based PCA,'' *Proc. ADCHEM 2000 Conf.*, Pisa (2000).
Rothwell, S. G., E. B. Martin, and A. J. Morris, ''Comparison of Methods for Handling Unequal Length Batches,'' *Proc. IFAC Symp. on Dynamics and Control of Process Systems*, Corfu, p. 66 (1998).
Scott, D. W., *Multivariate Density Estimation: Theory Practice and Visualization*, Wiley, New York (1992).
Wachs, A., and D. R. Lewin, ''Process Monitoring using Model-based PCA,'' *Proc. IFAC Symp. on Dynamics and Control of Process Systems*, Corfu, p. 86 (1998).
Wachs, A., and D. R. Lewin, ''Improved PCA Methods for Process Disturbance and Failure Identification,'' *AIChE J.*, **45**, 1688 (1999).
Wold, S., P. Geladi, and K. Esbensen, ''Multi-way Principal Components and PLS-Analysis,'' *J. Chemometrics*, **1**, 41 (1987).

Manuscript received Oct. 4, 1999, and revision received Mar. 15, 2000.