Application of Principal Component Analysis for Monitoring and Disturbance Detection of a Hydrotreating Process

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · AUGUST 2008

Impact Factor: 2.59 · DOI: 10.1021/ie0714605

Application of Principal Component Analysis for Monitoring and Disturbance Detection of a Hydrotreating Process

Stella Bezergianni* and Aggeliki Kalogianni

Chemical Process Engineering Research Institute, CPERI, Centre for Research & Technology Hellas, CERTH, Thermi-Thessaloniki, Greece

Principal component analysis (PCA) is a well-established technique for monitoring and disturbance detection of multivariate process, as it enables variability assessment through dimensionality reduction. PCA was applied to a hydroprocessing pilot plant to monitor the overall process variability. Contribution plots around points of increased variability were used to analyze process variability and its association with process variables. The methodology monitored successfully the set of 38 variables and diagnosed significant disturbances and their causes.

1. Introduction

Statistical process control (SPC) techniques are data-driven tools commonly employed for the examination of the overall variability in chemical processes. As large amounts of data are collected in the majority of complex chemical processes and since fundamental models are almost always too simplified for monitoring the dynamic behavior of such processes, these techniques become even more attractive. Due to the multivariate nature of chemical processes, multivariate SPC techniques offer easier and more effective monitoring capability than univariate SPC techniques such as Shewhart control charts, cumulative sum control charts, exponentially weighted moving average control charts, etc.[1−3]

In the early 1990s, the first applications of multivariate SPC techniques came out,[4] particularly partial least-squares (PLS) and principal component analysis (PCA). Since then, both multivariate SPC techniques were extended and enhanced, such as robust PCA.[5] Multiblock PLS[6] and multiway PCA[7] were the first approaches used for analyzing the variability of batch processes. To reduce the uncertainty between the beginning and the end of each batch, hierarchical PCA[8] was employed as an adaptive batch monitoring technique. The need for monitoring the process dynamics of multivariate systems was addressed by dynamic PCA,[9,10] model-based PCA,[11] and multiscale PCA.[12] Multivariate PCA and PLS techniques were also employed for process transitions of continuous processes as well as startups and restarts.[13] Another technique based on PCA was independent component analysis, which focuses on monitoring the variability of a selected set of independent components rather than of the principal components.[14] Multivariate SPC techniques have been further applied for fault diagnosis and disturbance isolation with promising results over the Tennessee Eastman plant simulator. Decentralized PCA statistical control charts[15] divided the process variables in groups associated with different sections of the process to effectively isolate the identified disturbances. Moreover, Fischer discriminant analysis was used to develop a criterion for diagnosing faults.[16]

Due to the multivariate, dynamic, and nonlinear nature of chemical processes, the efforts toward process monitoring and disturbance/fault detection and isolation continue. In this work, the classical PCA statistical process control methodology was applied to monitor the process variability of a pilot-scale hydroprocessing unit used primarily for catalyst evaluation. A typical experimental run on such a unit lasts between 1 and 2 months, and its success relies on the stability of the operating conditions that define the desired experimental protocols (specific feed rate, H_2/feed ratio, LHSV, reactor temperature, and pressure). The experimental runs are never identical as they are defined by different catalysts, different feedstockss and different experimental protocols. The main motivation for this work is the larger than expected variability and occurrence of frequent disturbances, which are characterizing the process, based on historical plant data. PCA statistical process control charts were used for monitoring the overall process, showing large amounts of variability dominating the process. These periods of increased process variability were further analyzed for detecting their source using contribution plots of the PCA model, which were able to locate the variables that included most of the process variability.

2. Methodology

2.1. Principal Component Analysis Control Charts. As it is well described in the literature,[17−19] the method of principal components is a linear transformation of the original variables x_i into a new set of variables y_i orthogonal to each other as

$$\mathbf{y} = \mathbf{V}^T\mathbf{x} \tag{1}$$

where $\mathbf{y}$ is a column vector known as scores and $\mathbf{x}$ is the data column vector centered by its mean and properly scaled. The matrix $\mathbf{V}$ is a matrix composed by the eigenvectors of the covariance matrix $\mathbf{S}$ of the original data matrix $\mathbf{X}$.

The eigenvalue decomposition of the covariance matrix $\mathbf{S}$ will give the transformation matrix $\mathbf{V}$, as presented in the next equation:

$$\mathbf{S} \overset{\text{eig}}{=} \mathbf{V}\Lambda\mathbf{V}^{-1} \tag{2}$$

and the matrix $\mathbf{V}$ can be used to transform the original variables into the scores $\mathbf{y}$, where $\mathbf{V}^{-1} = \mathbf{V}^T$. In order to scale the scores to have unit variance, a scaled transformation matrix $\mathbf{W}$ is defined:

$$\mathbf{W} = \mathbf{V}\Lambda^{-1/2} \tag{3}$$

In order to reconstruct the original variables from the principal components, the inverse of eq 1 can be used:

$$\hat{\mathbf{x}} = \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{y} \tag{4}$$

When the data are cross-correlated, the first few principal components can usually describe most of the variability within

* To whom correspondence should be addressed. Tel: +30-2310-498315. Fax: +30-2310-498380. E-mail. sbezerg@cperi.certh.gr.

10.1021/ie0714605 CCC: $40.75
Published on Web 08/16/2008

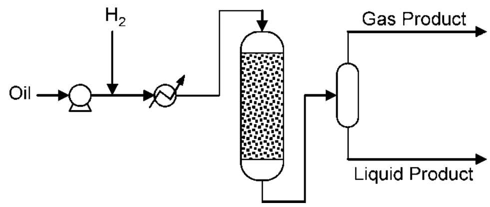

Figure 1. Simplified diagram of the hydroprocessing pilot plant (CPERI/CERTH).

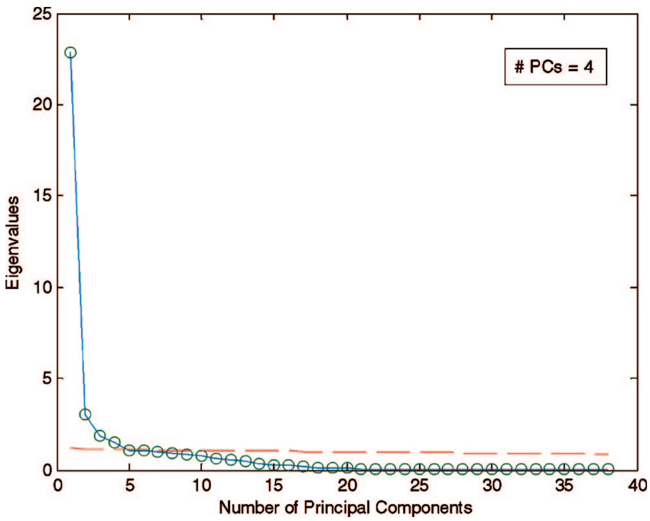

Figure 2. Parallel analysis of NOD set (38 variables) to determine number of principal components.

the data. Moreover, if the number of variables is reduced, then the computation of the PCA models becomes much faster since it involves matrices of smaller size. The eigenvalues are a significant measure of variability. The large eigenvalues show increased variability between the measurements, while small eigenvalues reveal the inherent correlation between the measurements. It is evident that the selection of the number of "significant" eigenvalues is of outmost importance for the PCA model.

There are several methods for determining the number of principal components that one should retain to compose a PCA model. Parallel analysis is a method widely studied in the literature[9,20] as it is both simple and efficient. It is a method in which the eigenvalues of the covariance matrix are compared in descending order with the ones of a matrix of the same dimensions that consists of random numbers. The method can be represented schematically by the plot of the eigenvalues of the covariance matrix in descending order and another curve depicting the eigenvalues of the independent data set. The point where the curves cross is taken to be the cutoff, which determines the number of principal components necessary to be retained. Figure 2 shows the schematic representation of parallel analysis on normal operating data (NOD) from the pilot plant under consideration, based on which only 4 out of 38 eigenvalues were found significant.

Retaining k principal components, the transformation matrix $\mathbf{W}$ becomes $\mathbf{W}_k$, and therefore

$$\mathbf{y}_k = \mathbf{W}_k^T\mathbf{x}_k = \mathbf{\Lambda}_k^{-1/2}\mathbf{V}_k\mathbf{x}_k \tag{5}$$

and similarly the original variables can be reconstructed from the scores through the following relationship:

$$\hat{\mathbf{x}}_k = \mathbf{W}_k(\mathbf{W}_k^T\mathbf{W}_k)^{-1}\mathbf{y}_k \tag{6}$$

where $\hat{\mathbf{x}}_k \in E^n$ (E^n denotes an nth dimensional space) while $\mathbf{y}_k \in E^k$ (E^k denotes a kth dimensional space). When the number of principal components and the scores $\mathbf{y}$ have been determined, the two statistical PCA process control charts can be constructed. The first one is based on Hotelling's T^2 statistics and is calculated as

$$T^2 = \mathbf{y}_k^T\mathbf{y}_k \tag{7}$$

representing the total amount of inherent variability that is inside the process. A significant change in its value is an indication of the existence of a disturbance in the process. The statistical distribution of its values follows the F-distribution.

The second PCA control chart is calculated from the sum of the squares of the residuals as follows:

$$Q = (\mathbf{x} - \hat{\mathbf{x}})^T(\mathbf{x} - \hat{\mathbf{x}}) = (\mathbf{x} - \mathbf{W}_k(\mathbf{W}_k^T\mathbf{W}_k)^{-1}\mathbf{y}_k)^T \times (\mathbf{x} - \mathbf{W}_k(\mathbf{W}_k^T\mathbf{W}_k)^{-1}\mathbf{y}_k) \tag{8}$$

a measure of the "goodness of fit" of the PCA model, showing the distance between the actual and the predicted data.

When the data matrix $\mathbf{X}$ consists of time shift duplicate vectors, the resulting model is denoted as a dynamic principal component analysis model (DPCA). In the other case of static vectors composing the data matrix $\mathbf{X}$, the resulting model is denoted as a static principal component analysis model or simply PCA.

2.2. Contribution Plots. While monitoring the process using the two PCA control charts, several disturbances can be detected. In order to find more information about these disturbances or the increased variability that they represent, one can locate where this additional process movement is localized. This way the variables, among which the increased variability is attributed, are identified. For batch processes where multiway PCA is used for process monitoring, diagnostic information has been extracted by the use of contribution plots.[7]

For continuous processes, the contribution plots have a simpler format. The diagnostic information can be deducted by creating contribution plots for the particular time steps that the excessive variability was observed. These contribution plots are a set of bar charts that show the correlation of the variability with the variables ($\mathbf{X}$) and scores ($\mathbf{Y}$) for this particular time step. Then, the scores with the highest value are examined by comparing their corresponding weight ($\mathbf{W}$) and the model residuals.

As shown in eq 7, the Hotelling's T^2 statistics is the sum of the squares of the scaled principal components. Therefore, for a new point in time, the T^2 statistics can be monitored as

$$T^2 = \sum_{r=1}^{k} y_r^2 \tag{9}$$

while the Q represents the sum of the squares of the residuals:

$$Q = \sum_{j=1}^{n} (x_j - \hat{x}_j)^2 \tag{10}$$

From eqs 12 and 13, it is clear that the percent contribution (%) to T^2 of each score r will be $100y_r^2/T^2$ and the percent contribution (%) of each variable j to Q will be respectively $100(x_j - \hat{x}_j)^2/Q$.

3. Application to a Hydroprocessing Pilot Plant

The monitoring and disturbance detection methodology described in section 2 was applied offline to historical data of a hydroprocessing unit of CPERI/CERTH. The process descrip-

Table 1. Process Variables Description

variable number	variable name	description	variable units
1	FT-21A	H_2 flow to reactor (primary)	scfh
2	FT-21B	H_2 flow to reactor (secondary)	scfh
3	LT-91	separator level	%
4	LV-91	separator level controller output	%
5	PT-101	RX outlet pressure	psig
6	PT-23	RX inlet pressure	psig
7	PV-81	outlet pressure controller output	%
8	TE-101	RX zone-1 outer temperature	°F
9	TE-102	RX zone-2 outer temperature	°F
10	TE-103	RX zone-3 outer temperature	°F
11	TE-104	RX zone-4 outer temperature	°F
12	TE-105	RX zone-5 outer temperature	°F
13	TE-106	RX zone-6 outer temperature	°F
14	TE-111	RX zone-1 inner temperature	°F
15	TE-112	RX zone-2 inner temperature	°F
16	TE-113	RX zone-3 inner temperature	°F
17	TE-114	RX zone-4 inner temperature	°F
18	TE-115	RX zone-5 inner temperature	°F
19	TE-116	RX zone-6 inner temperature	°F
20	TE-117	reactor product temperature	°F
21	TE-52	feed recycle temperature	°F
22	TE-53	feed to RX temperature	°F
23	TE-81	gas product temperature	°F
24	TE-91	separator temperature	°F
25	TE-92	liquid product to sample temperature	°F
26	TE-93	product sample heater temperature	°F
27	TIC-101	RX zone-1 temperature controller	%
28	TIC-102	RX zone-2 temperature controller	%
29	TIC-103	RX zone-3 temperature controller	%
30	TIC-104	RX zone-4 temperature controller	%
31	TIC-105	RX zone-5 temperature controller	%
32	TIC-106	RX zone-6 temperature controller	%
33	TIC-117	reactor product temperature controller	%
34	TIC-52	feed recycle temperature controller	%
35	TIC-53	feed to RX temperature controller	%
36	TIC-91	separator temperature controller	%
37	TIC-92	liquid product to sample temperature controller	%
38	TIC-93	product sample heater temperature controller	%

tion, the data collection, and PCA application to pilot plant data will be described in the following sections.

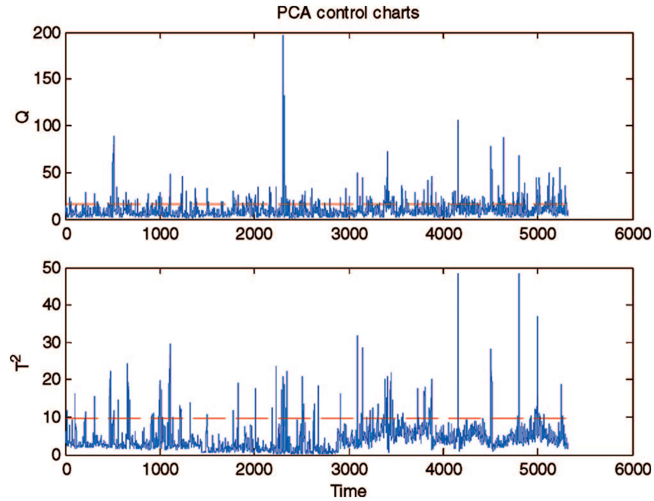

Figure 3. Principal component analysis models based on pilot plant NOD.

3.1. Process Description. The VB01 hydroprocessing pilot plant of CPERI/CERTH (Figure 1) was selected as an application platform of the proposed PCA monitoring methodology. This hydroprocessing unit was considered an excellent candidate since it combines the actual operating parameters (temperatures, pressures) and feedstocks of industrial hydroprocessing applications for the limiting but yet challenging size of pilot-scale units.

The hydroprocessing pilot plant consists of four sections: (a) feed, (b) reaction, (c) separation, and (d) product collection. There are a total of 38 process variables associated with the unit, and there are 17 control loops to maintain stable temperature, pressure, and flow rate within the reactor and feed/product lines (Table 1).

This hydroprocessing unit is used for hydrotreating, i.e., desulfurization/denitrogenation of refinery streams, as well as for hydrocracking, both continuous processes. The experimental data that were employed for PCA monitoring were of a single hydrocracking experimental run, which was conducted at a constant pressure and four different reactor temperatures throughout a period of 1 month. Within this month, several disturbances were observed in the pilot plant, including intentional disturbances and unforeseen ones. The main premise of this work was to apply the PCA monitoring and diagnosing tools to historical data, in order to validate the identified disturbances and to analyze their effects and potential causes.

3.2. Data Collection. As multivariate statistical control tools require as many process variables as possible, all process variables associated with the hydroprocessing pilot plant were utilized for the analysis. These 38 process variables included feed flow rates, reactor temperatures, feed and product line temperatures, and system pressures. Furthermore, the controlled variables of the automated control systems (valve openings %, electric actuators %) were also included in the analysis. However, set points and constant variables (i.e., liquid feed rate) were not included for the analysis, as their lack of variability would not add any additional information to the analysis. The process variables' data were collected every 1 min via the OSISOFT PI[21] data archiving and retrieval system, allowing no data compression.

For the statistical analysis, two different types of data sets are needed. The first type incorporates only a small amount of variability and serves as a representative of normal operating conditions, which is used to build the model for the analysis. The second type includes some variability and is used for

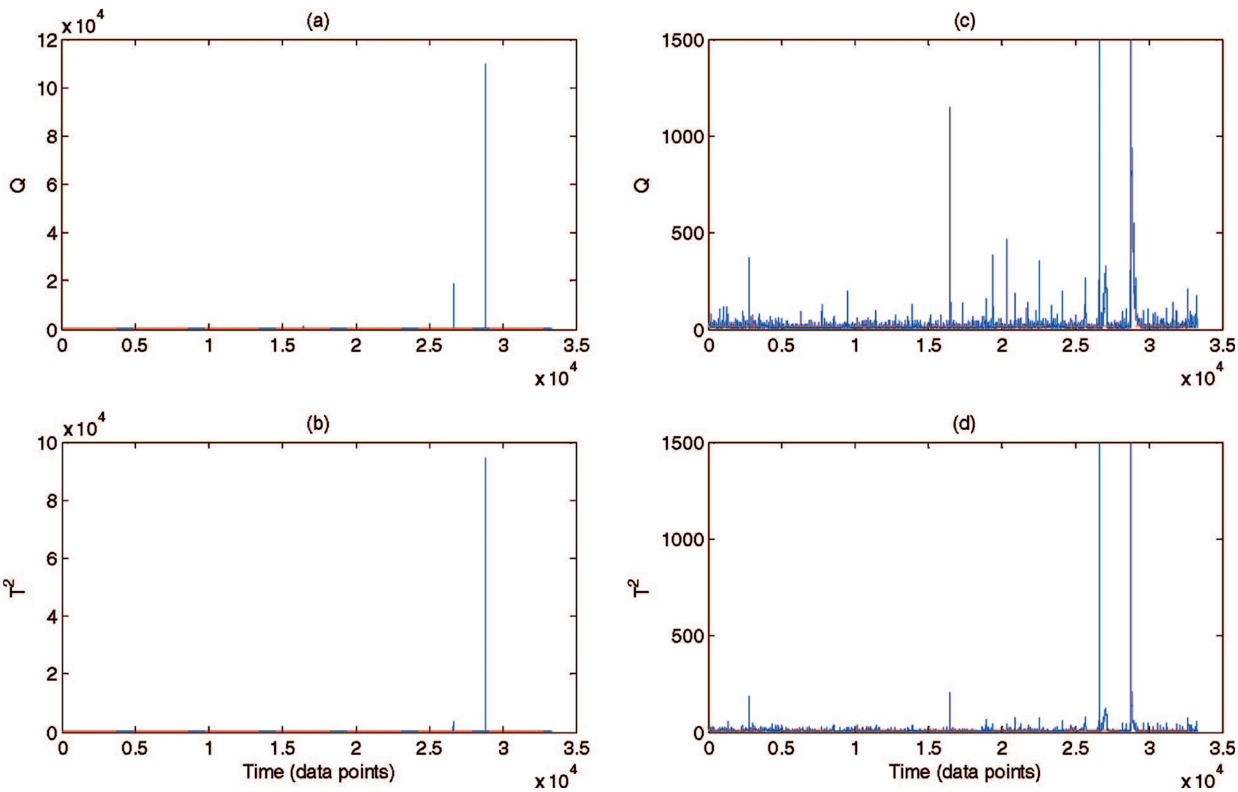

Figure 4. PCA control charts for VB-01 hydrocracking unit of (2−25 June, 2007).

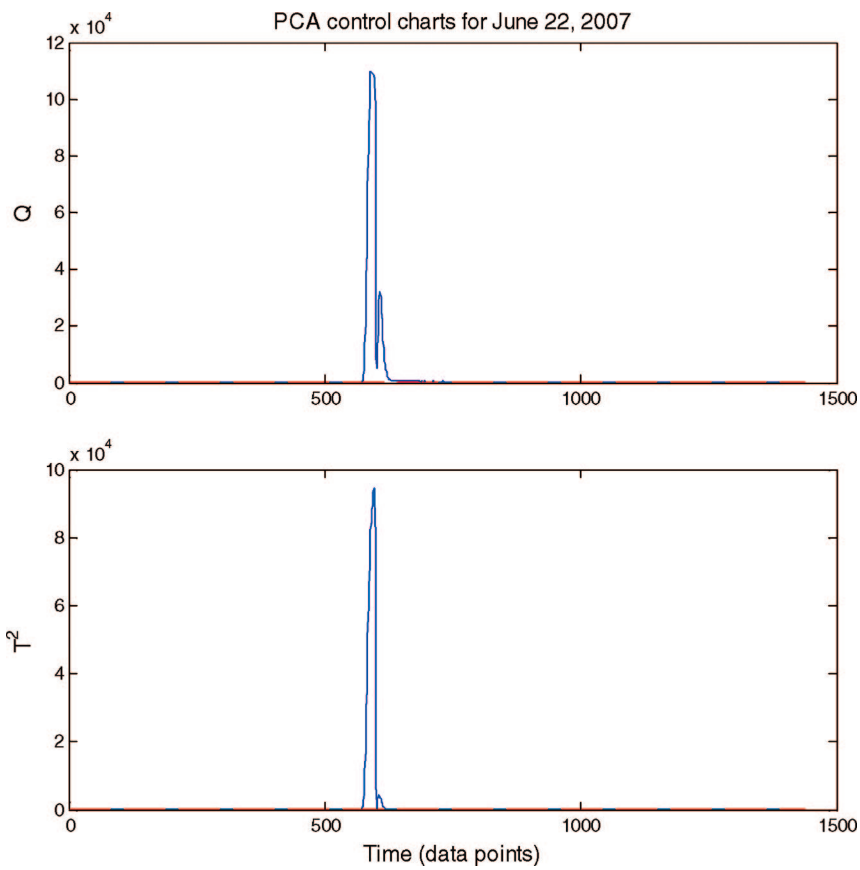

Figure 5. PCA control charts for June 22, 2007.

monitoring the process as well as detecting and isolating any occurring disturbances.

3.3. PCA Model Development. The first and crucial step for developing the PCA model is the selection of NOD used for model building. After careful analysis of the whole period under consideration (i.e., data from 2 to 25 June 2007; 33321 data points), four individual periods were chosen as good representatives of NOD. These were around 6, 8, 19, and 23 of

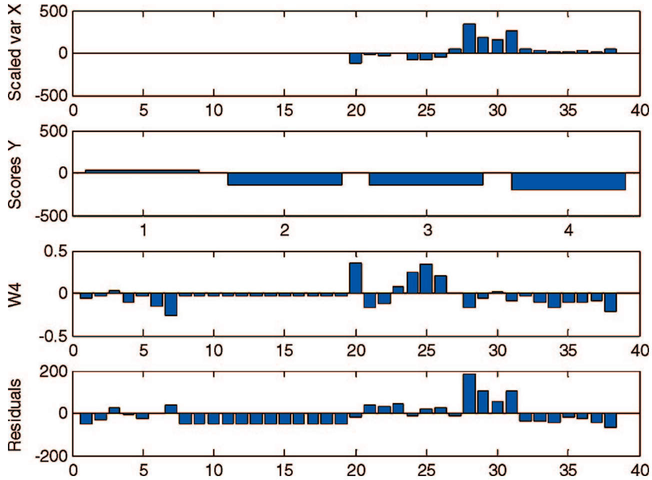

Figure 6. Contribution plots for $N = 589$ data point (June 22, 2007).

June 2007). In order for the four data sets to be considered as NOD sets, they have the following characteristics: (1) include some but not excessive variability; (2) include all different sets of operating conditions (mainly RX temperature) in order to be able to capture the different operating states; (3) exhibit no special mode of operation such as deliberate disturbances, which were imposed for maintenance purposes.

These four data sets exhibited normal process variability for all main process variables (system pressure, reactor temperatures), as observed by process engineers, and did not include any unexpected disturbances. The four individual NOD sets were superimposed forming a larger NOD set with 5317 data points that was used to build the model according to the methodology described in section 2.1.

Based on the eigenvalue decomposition of the NOD matrix described in eq 2, the eigenvalues were calculated, which were used in the next and crucial step of determining the principal components of variability. In Figure 2, the schematic representation of parallel analysis on NOD is depicted, according to which only 4 out of 38 eigenvalues were significant. These 4 eigenvalues are the principal components from which a PCA model can be constructed.

Once the number of principal components is estimated, eqs 5−8 provide the PCA model, which is expressed in the form of T^2 and Q statistical control charts. In Figure 3, the PCA statistical control charts of the NOD set are shown, indicating the normal amount of variability. As was declared in section 2.1, the control limits indicated with red dashed lines are the 95% confidence limits and will be the same for all the PCA statistical control charts that are used for monitoring the process variability. For this NOD, 95% of T^2 and 95% of Q were within the statistical control limits. Particularly the peaks of the T^2 control chart indicate the inherent process variability that lies within the NOD. This inherent variability is maintained within the PCA model, indicating the amount of variability that may be considered as "normal". Out of the total of 5317 data points of the NOD, 268 data points (5.0%) were outside the T^2 control limit and 516 data points (9.7%) were outside the Q control limit.

3.4. PCA Control Charts for Monitoring. Initially the PCA control charts for the whole data set were created, as shown in Figure 4a,b. From the overall PCA charts, two dominant disturbances appeared as the most interesting ones to study, which were around the periods of June 20 and 22, as the PCA statistical control charts revealed increased variability around these two days. As these two occupied a significant amount

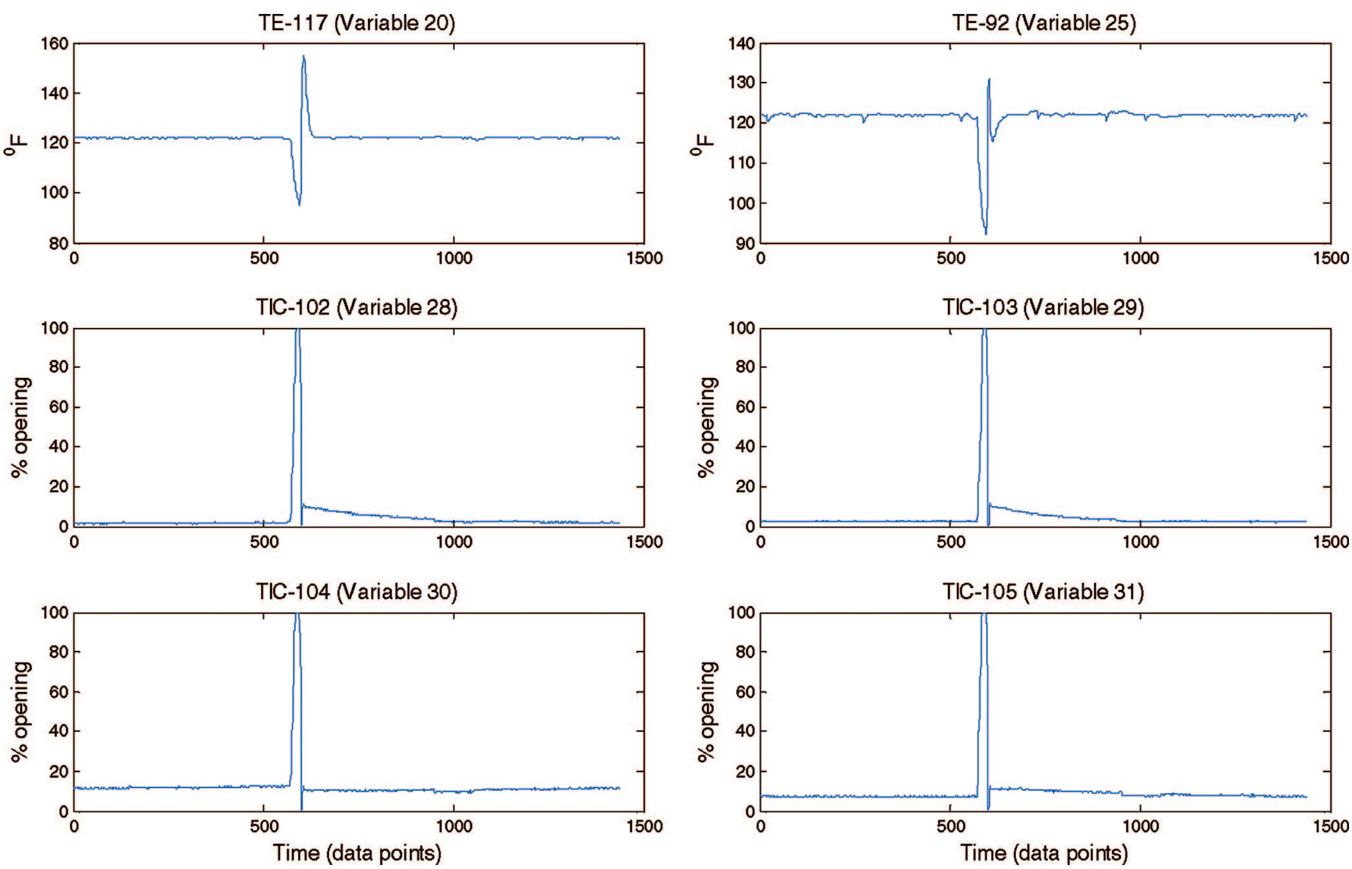

Figure 7. Variables 20, 25, 28, 29, 30, and 31 on June 22, 2007.

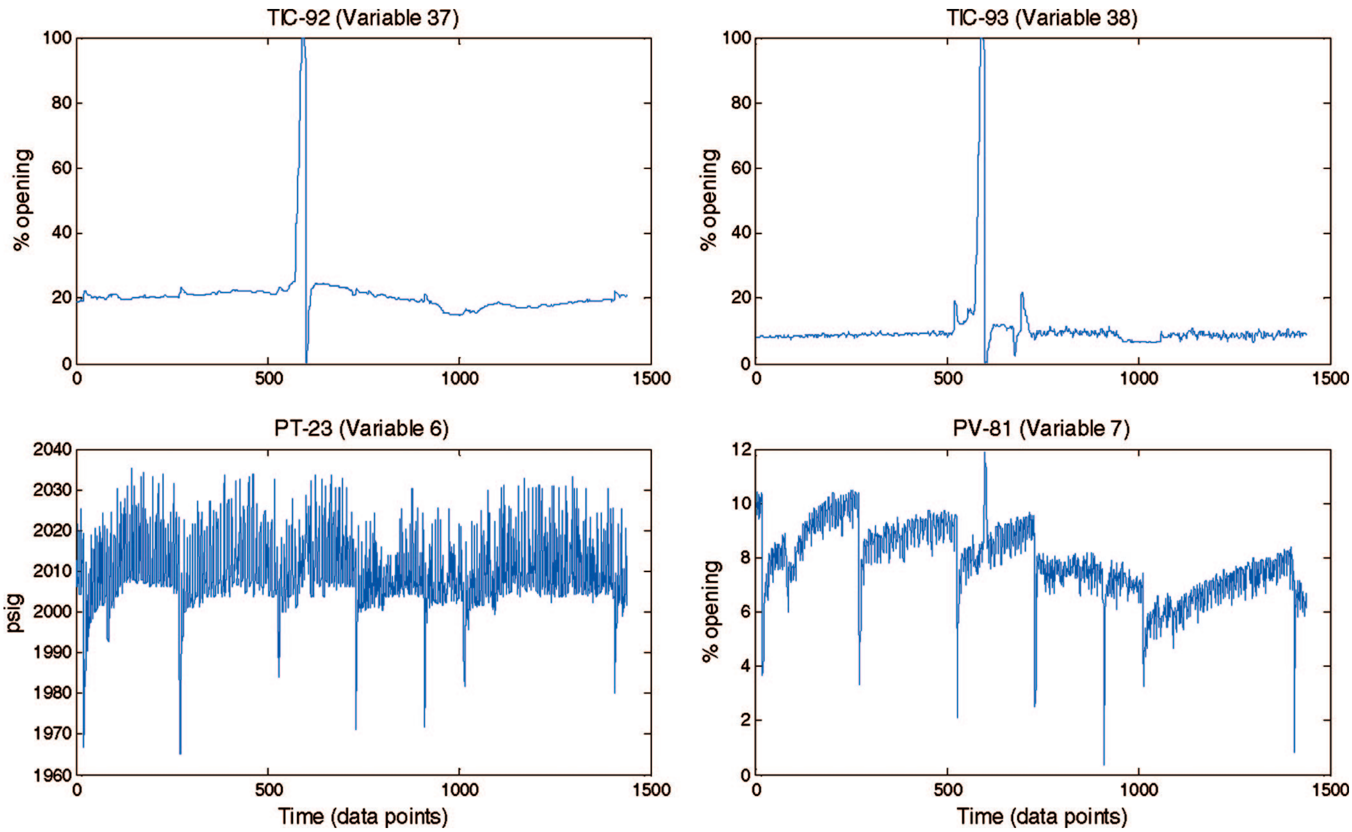

Figure 8. Variables 37, 38, 6, and 7 on June 22, 2007.

of variability, the scale of the PCA charts needed to be decreased to focus on smaller instances of increased variability. Therefore, a more detailed examination of the overall PCA control charts (Figure 4c,d) showed that there are other instances of increased variability, such as June 17. For the analysis that follows, three data sets around June 22, 20, and 17 were considered.

June 22, 2007. The PCA statistical control charts for this day are shown in Figure 5, which exhibited maximum variability at the $N = 589$ data points, indicating extraordinary process variability. The contribution plots for that particular data point, shown in Figure 6, were then created to further analyze the additional variability. As can be easily observed from the scaled X values in Figure 6, variables 28−31 showed the highest variability. Moreover, comparing the scores of the PCA associated with this excessive variability indicated that the fourth score accumulated most of this variability, as analyzed in the second chart of Figure 6. Comparing the weights of all variables to the fourth score, it was evident that variables 20 and 25 are mostly related with this score and therefore with the increased variability it accumulated. All six variables are associated with temperatures. Variables 28−31 are the temperature actuators in the intermediate reactor zones (TIC-102, TIC-103, TIC-104, TIC-105), while variable 20 is the reactor outlet temperature (TE-117) and variable 25 is the liquid product temperature at the outlet of the separator (TE-92).

These unscaled variables are plotted in Figure 7, where an abrupt change is observed for the four actuators and the two outlet temperatures. The sudden increase of up to 100%

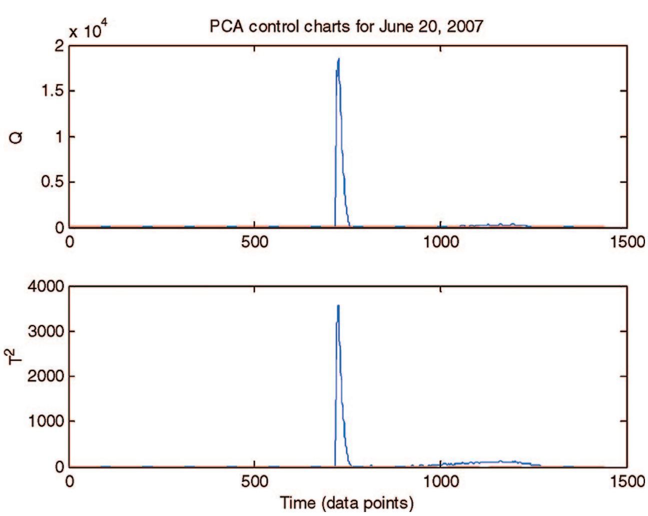

Figure 9. PCA control charts for June 20, 2007.

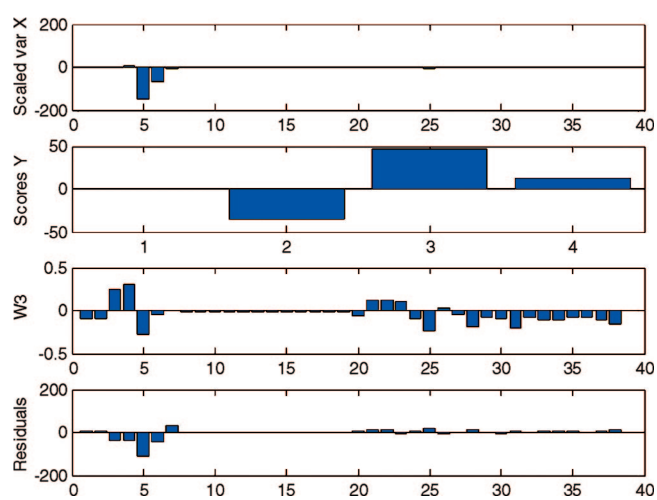

Figure 10. Contribution plots for $N = 726$ data point (June 20, 2007).

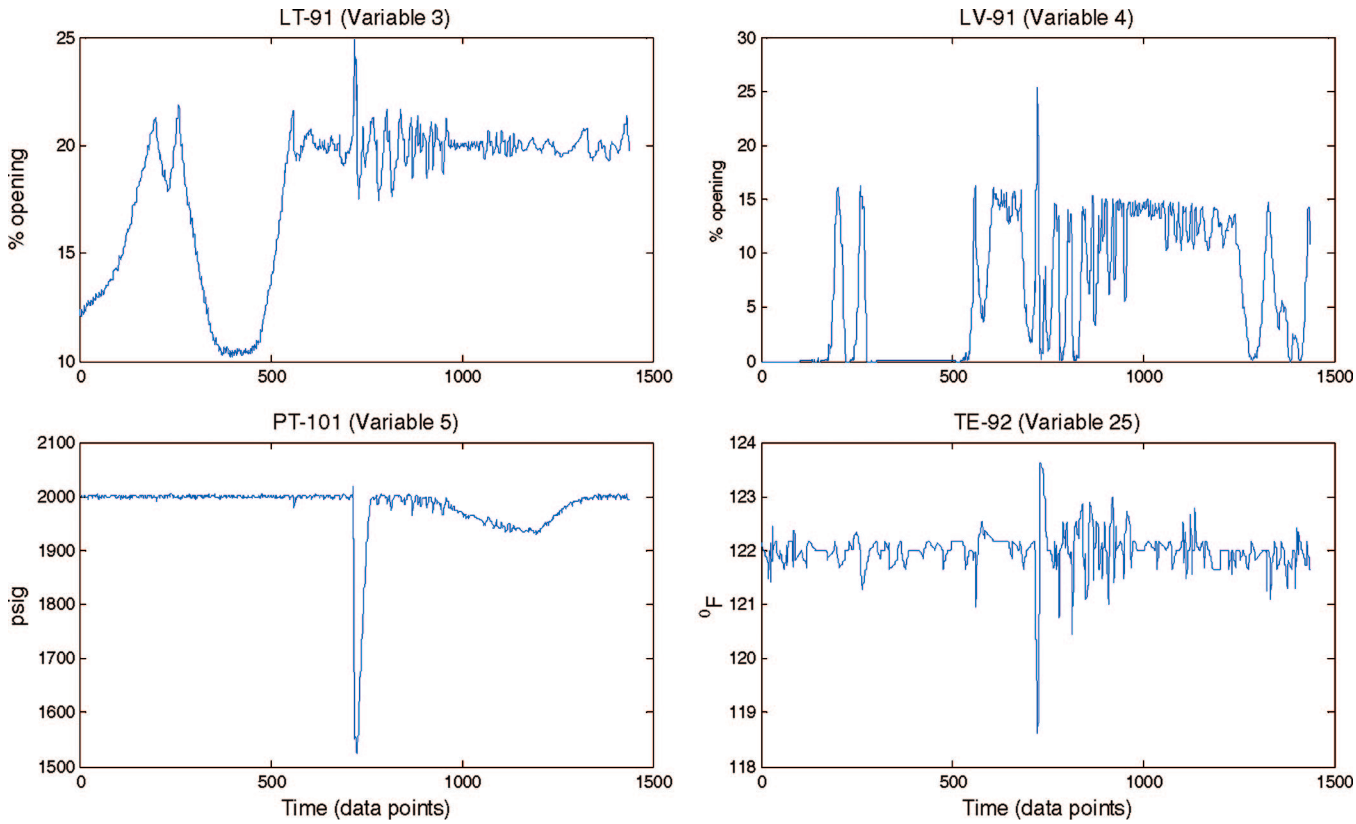

Figure 11. Variables 3, 4, 5, and 25 on June 20, 2007.

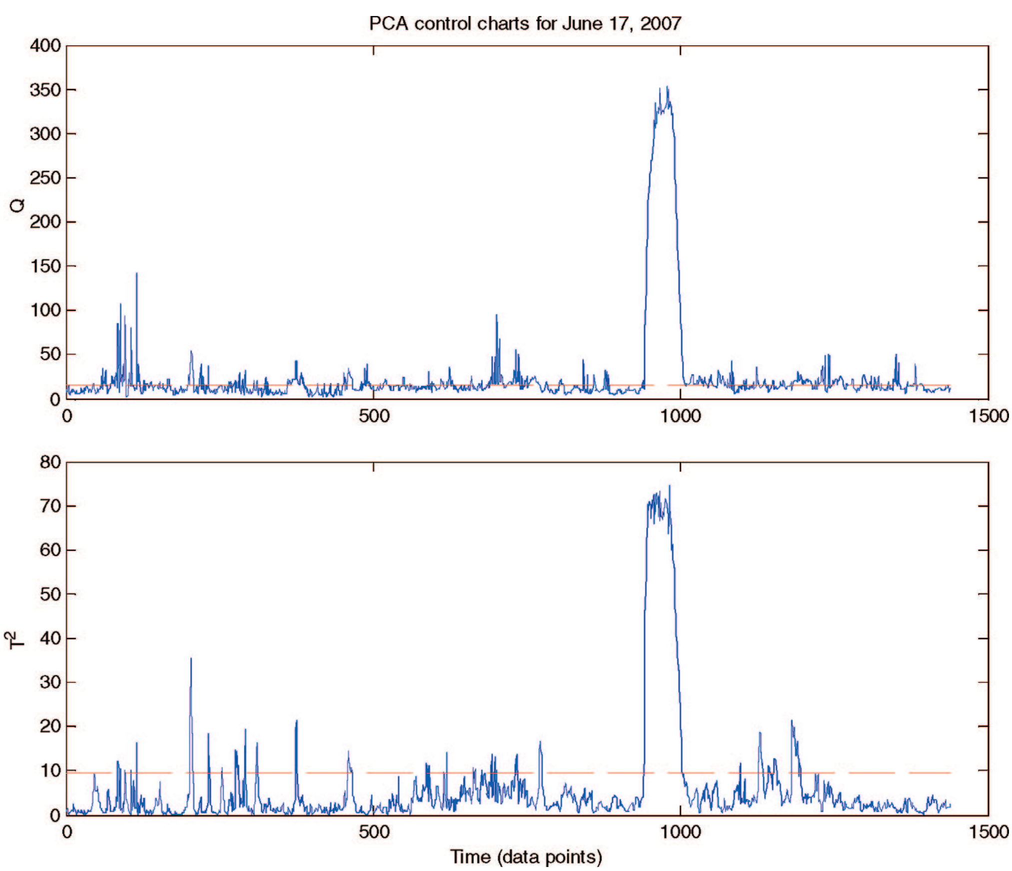

Figure 12. PCA control charts for June 17, 2007.

opening indicates either an erroneous temperature signal or an abnormal incident. Since this phenomenon appears for more that one temperature (all independently controlled), only the second possibility of an abnormal incident may explain this excessive variability. Indeed, this was the result of a fuse failure, which forced the actuators of all temperature

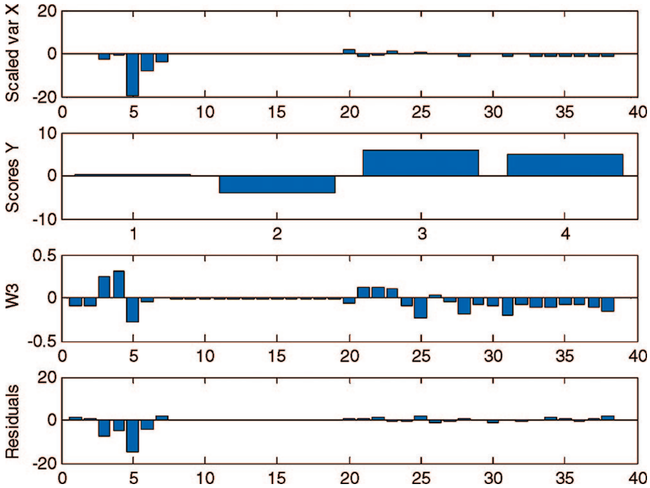

Figure 13. Contribution plots for $N = 983$ data point (June 17, 2007).

controllers to open 100%, which was immediately restored, and for that reason, the process shortly returned under normal operation.

Besides the aforementioned temperature-related variables, there are some other variables that are also significantly related with the fourth score, variables 37, 38, 6, and 7. Variables 37 and 38 are two other temperature actuators of the unit, while 6 and 7 are variables associated with the pressure of the unit, whose behavior was also affected by the fuse failure. As these four variables are independently plotted in Figure 8, it is shown that the disturbance around the 589th data point can be seen for variables 37, 38, and 7. However, variable 6 shows no appreciable change in its variability. However, this variable shows no weight on its contribution to the model residuals of the fourth score for the 589th data point. This variable is the system pressure, which often shows periods of increased variability, and therefore, such an alarm would actually be redundant.

Therefore the increased variability of June 22 was not caused by a process-related issue but rather by an imponderable factor such as the fuse failure, which led to simultaneous failure of all thermocouples and a general decrease of all the unit temperatures.

June 20, 2007. Another interesting behavior is captured in the PCA control charts of June 20 as captured in Figure 9, and particularly for the $N = 726$ data point. Further analysis of the $N = 726$ data point through contribution plots is performed (Figure 10). From the scaled X values comparison, variables 5 and 6 exhibited the highest variability. Variable 5 is the reactor outlet pressure actuator, which often exhibits increased variability as it is trying to control the challenging system pressure. Variable 6 is the separator outlet temperature, which may be directly affected by the pressure.

However, by examining the weights of the variables contributing to the third score associated with the increased variability under examination, it appeared that besides the pressure-related variables some other variables also have a significant weight. Particularly variables 3, 4, 5, and 25 also have a significant weight on the third score. For that reason, some of these variables are more carefully monitored via individual plots. In Figure 11, variables 3, 4, 5, and 25 are given

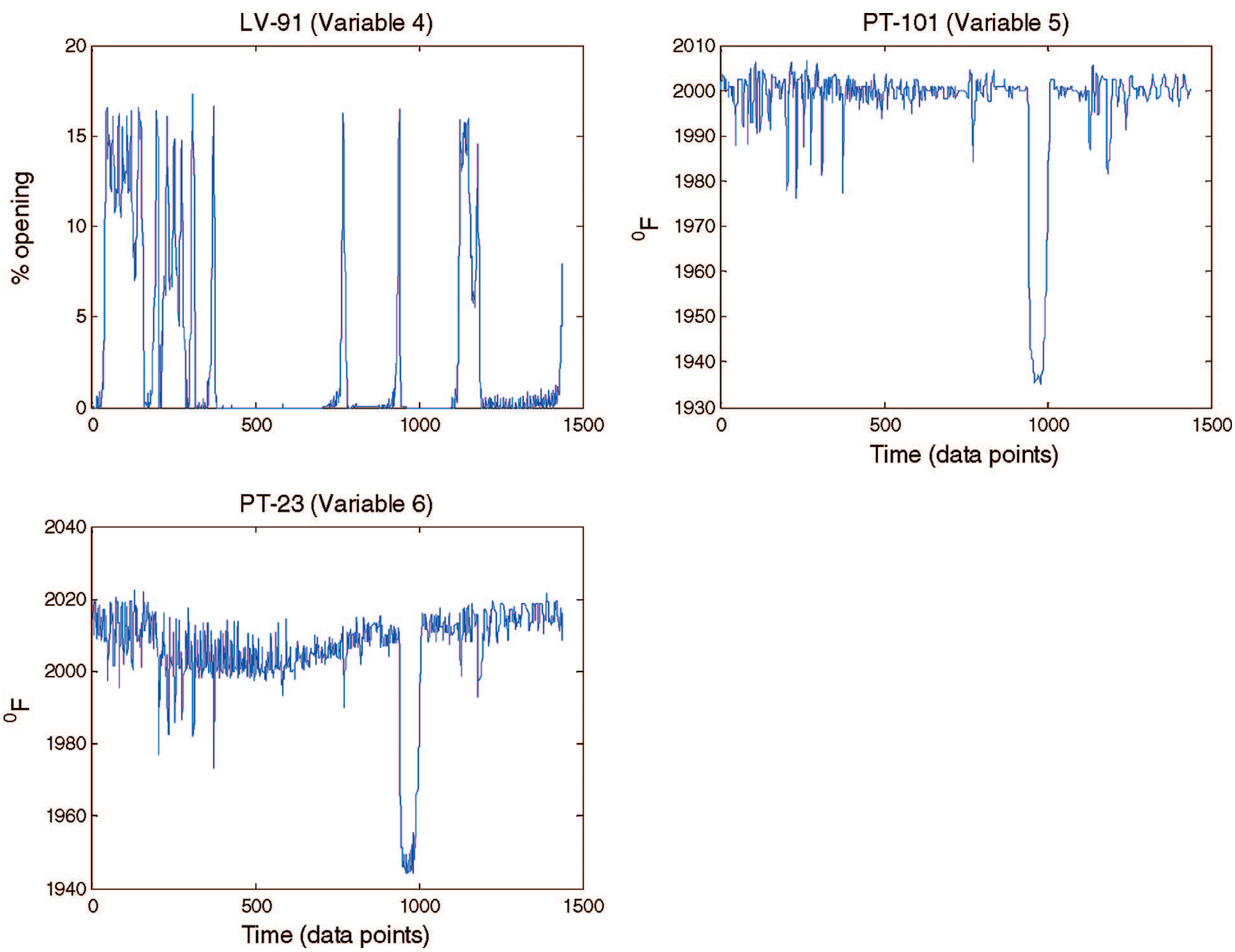

Figure 14. Variables 4, 5, and 6 on June 17, 2007.

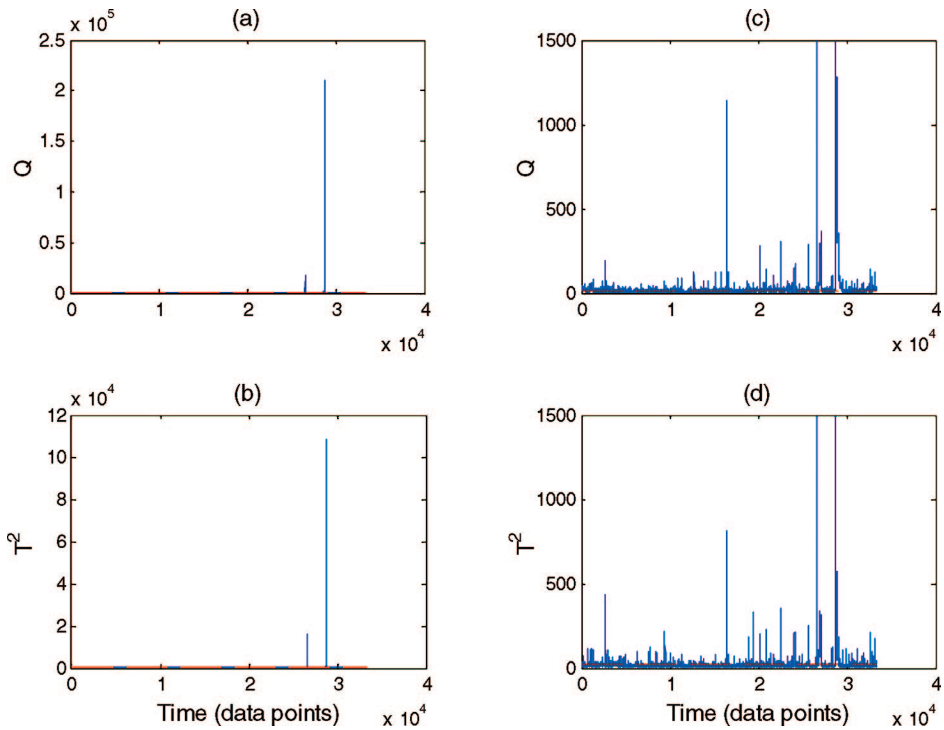

Figure 15. DPCA (first order) control charts for VB-01 hydrocracking unit of (June 2–25, 2007).

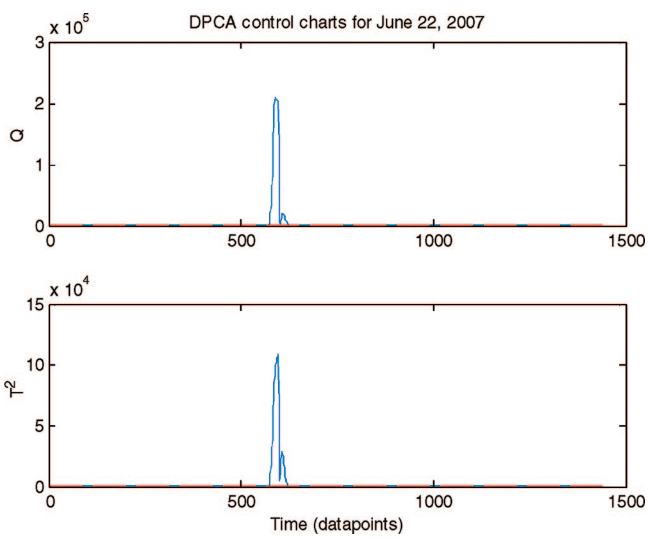

Figure 16. DPCA control charts for June 22, 2007.

showing and increased peak around the 726th data point. It should be noted that all these variables are strongly related to the system pressure (variable 5), which exhibits a sudden decrease.

The sudden pressure decrease is attributed to a deliberate opening of a trap placed after the separator and before the pressure control valve, which causes a sudden pressure drop. The pressure drop caused by this maintenance activity is responsible for a peak of the level of the separator, which forces the level control valve to open. At this point, it is worth noticing that the temperature variables (27–38) did not show any variability within the analysis given above, as they were unaffected by the disturbance. This fact further indicates that the pressure and the level disturbances do not affect the temperatures of the unit (with the exception of the separator liquid product temperature, while in the opposite case (see data analysis for June 22), the intense variability of the temperatures at various points of the unit appeared to have an impact on the pressure and level variability.

June 17, 2007. Apart from the two data sets that were studied as they were associated with the two highest peaks of the overall PCA control charts, another data set was examined. The PCA control charts of June 17 are shown in Figure 12, revealing an increased variability around the $N = 983$ data point. This increased variability was examined via contribution plots in Figure 13 for this data point. By comparing the scaled X values, it appears that once more the pressure-related variables 5 and 6 exhibited the highest variability, as was also verified by individually monitoring them (Figure 14).

As contribution plots indicated that the third weight was mostly related with the increased variability for the 983rd data point, its weights to the 38 variables indicated that variable 4 also may have a significant contribution. The three variables (4–6) are plotted in Figure 14 showing that for all three the overall increased variability can be related to its individual variability. However, in the case of variable 4, this variable showed peaks at different time instances as well, which may have been caused by more "normal" or mild phenomena, compared with the peak around the 983rd data point, which was associated with the pressure variability.

In this particular point, the intense disturbance attributed to the pressure variability was not due to any deliberate action. Failure in the control of the pressure could be one of the reasons of this increased variability. This occasional increased pressure variability is mostly attributed to the insufficient separation of gas–liquid in the separator that causes sweeping of liquid drops into the gas line, despite of the trap placed after the separator, causing controllability problems in the pressure control valve. These problems are likely to restrain the normal operation of the control valve, causing it to stick or block its outlet and consecutively failing to control the system pressure to the desired set point. Finally, it should be noted that this influence of

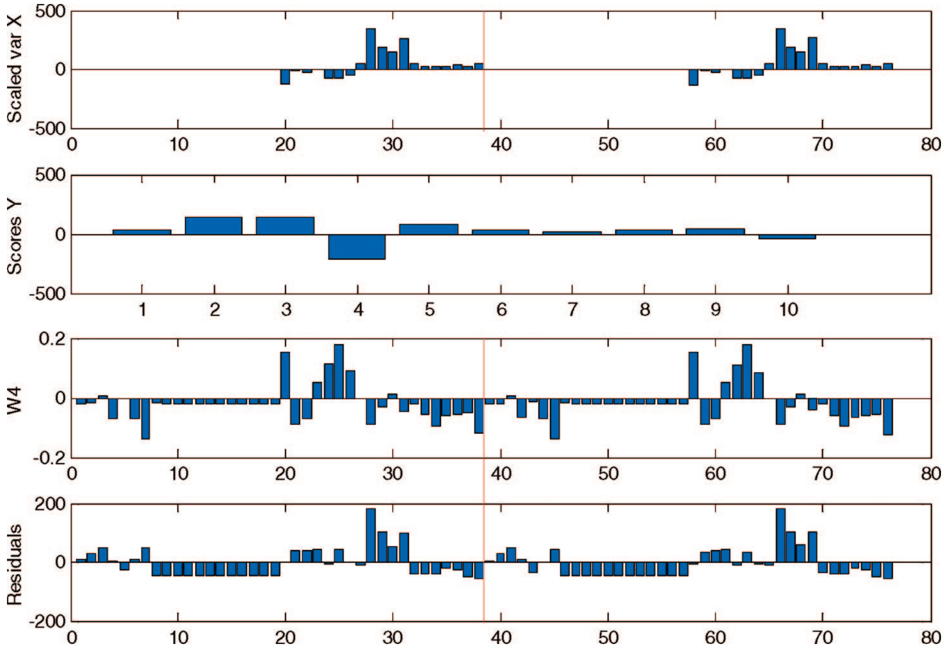

Figure 17. Contribution plots for $N = 589$ data point (June 22, 2007) created by DPCA.

pressure variability to the overall process variability was observed several times throughout the run length of this experiment.

3.5. DPCA Control Charts for Monitoring. Data time correlation is examined by dynamic PCA charts or DPCA.[9,10] The first-order DPCA control charts for the complete data set (June 2−25) were created and are presented in Figure 15a,b. As was observed from the standard static PCA control charts (Figure 4a,b), two dominant disturbances appeared around the same periods on June 20 and 22. A more detailed examination of the overall PCA control charts (Figure 15c,d) showed that there are other instances of increased variability, such as on June 17. The static PCA and DPCA charts show similar peaks, i.e., instances of increased process variability at the same time instances, indicating that no additional information is obtained from DPCA. The difference lies on the number of principal components, which is 4 for the PCA model and 10 for the DPCA of first order.

The DPCA statistical control charts for the day that exhibited the highest variability (June 22) are shown in Figure 16. The data point exhibiting maximum variability is the $N = 589$ data point, which is exactly the same point that the static PCA charts indicated (Figure 5).

Going one step further, the contribution plots for that particular data point are shown in Figure 17. The scaled X values, as well as the variable weights on the scores and residuals of the fourth score, show two identical set of values related to the 76 variables, indicating that there is no time-dependent correlation between these variables, i.e., sufficient sampling interval. This conclusion is reinforced by the fact that the contribution of each process variable as shown from the contribution plots of the DPCA model (Figure 17) is exactly the same as their contribution plots of PCA in Figure 6. Therefore, the DPCA add no additional information for this process.

4. Conclusions

Principal component analysis was employed to evaluate three cases of increased variability of a hydroprocessing pilot plant. The PCA model showed that hydroprocessing process monitoring is indeed a multidimensional problem and, therefore, needs multivariable methods to be achieved. From the total of 38 process variables, there were 4 principal components of variability that are sufficient to monitor the problem.

The PCA control charts combined with contribution plots, developed for data points of increased variability, appear to be a successful approach for monitoring disturbances and in some cases identifying their cause. Moreover, it was shown that the variability of certain variables and particularly of the ones associated with the system pressure affect significantly the overall process variability. Indeed pressure control is the primary problem of the overall operation of this pilot plant. On the other hand, temperature control is tight and failure to control it can only be attributed to independent events, such as the fuse failure that was diagnosed.

Finally, DPCA was also applied to the same data offering no additional information regarding the instances of increased variability or its causes.

Literature Cited

(1) Kano, M.; Hasebe, S.; Hashimoto, I.; Ohno, H. Evolution of multivariate statistical process control: Application of independent component analysis and external analysis. *Comput. Chem. Eng.* **2004**, *28*, 1157.

(2) Kourti, T.; MacGregor, J. F. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemom. Intell. Lab. Syst.* **1995**, *28*, 3.

(3) Martin, E. B.; Morris, A. J.; Papazoglou, M. C.; Kiparissides, C. Batch process monitoring for consistent production. *Comput. Chem. Eng.* **1996**, *20* (Suppl.1), S599.

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

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

(6) MacGregor, J. F.; Jaeckle, C.; Kiparissides, C.; Koutoudi, M. Process Monitoring and Diagnosis of Multiblock PLS Methods. *AIChE J.* **1994**, *40*, 826.

(7) Nomikos, P. Detection and Diagnosis of Abnormal Batch Operations Based on Multi-way Principal Component Analysis. *ISA Trans.* **1996**, *35*, 259.

(8) Rannar, S.; MacGregor, J. F.; Wold, S. Adaptive batch monitoring using hierarchical PCA. *Chemom. Intell. Lab. Syst.* **1998**, *41*, 73.

(9) Ku, W.; Storer, R. H.; Georgakis, C. Disturbance Detection and Isolation by Dynamic Principal Components Analysis. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 179.

(10) Tsung, F. Statistical monitoring and diagnosis of automatic controlled processes using dynamic PCA. *Int. J. Prod. Res.* **2000**, *38*, 625.

(11) Rotem, Y.; Wachs, A.; Lewin, D. R. Ethylene Compressor Monitoring Using Model-based PCA. *AIChE J.* **2000**, *46*, 1–825.

(12) Bakshi, B. R. Multiscale PCA with application to Multivariate Statistical Process Monitoring. *AIChE J.* **1998**, *44*, 1596.

(13) Duchesne, C.; Kourti, T.; MacGregor, J. F. Multivariate SPC for startups and grade transitions. *AIChE J.* **2002**, *48*, 2890.

(14) Albazzaz, H.; Wang, X. Z. Statistical process control charts for batch operations based on independent component analysis. *Ind. Eng. Chem. Res.* **2004**, *43*, 6731.

(15) Georgakis, C., Steadman, B., Liotta, V. *Decentralized PCA charts for performance assessment of Plant-Wide Control Structures*; IEEE Press: Piscataway, NJ, 1996; p 97.

(16) Chiang, L. H.; Russell, E. L.; Braatz, R. D. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 243.

(17) Jackson, J., E., *A user's guide to Principal Component Analysis*; John Wiley & Sons: New York, 1990.

(18) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*; John Wiley & Sons: New York, 1998.

(19) Chiang, L. H.; Russell, E. L.; Braatz, R. D. *Fault Detection and Diagnosis in Industrial Systems*; Springer-Verlag: London, 2001.

(20) Himes, D.; Georgakis, C.; Storer, R. H. Determination of the Number of Principal Components for Disturbance Detection and Isolation. In *Proceedings of American Control Conference*, Baltimore, MD, June 29–July 1, 1994; p 1279.

(21) http://www.osisoft.com/Products/PI%20System/.

Received for review October 29, 2007
Revised manuscript received May 27, 2008
Accepted July 7, 2008

IE0714605