

Statistical Fault Detection of Chemical Process - Comparative Studies

Majdi Mansouri^{1*}, Mohammed ZS², Raoudha Baklouti^{2,5}, Mohamed Nounou⁴, Hazem Nounou⁵, Ahmed Ben Hamida² and Nazmul Karim³

¹Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar

²Advanced Technologies for Medicine and Signals, National Engineering School of Sfax, Tunisia

³Chemical Engineering Department, Texas A&M University, College Station, TX 77843, USA

⁴Chemical Engineering Program, Texas A&M University at Qatar, Doha, Qatar

⁵Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar

Abstract

This paper addresses the statistical chemical process monitoring using improved principal component analysis (PCA). PCA-based fault-detection technique has been used successfully for monitoring systems with highly correlated variables. However, standard PCA-based detection charts, such as the Hotelling statistic, T^2 and the sum of squared residuals, SPE, or Q statistic, are not able to detect small or moderate events since they use only data from the most recent measurements. Different fault detection (FD) charts, namely generalized likelihood ratio test (GLRT), shewhart control chart and exponentially weighted moving average chart (EWMA) control chart have been shown to be among the most effective univariate fault detection methods and more suitable for detection small faults. The objective of this work is to improve the PCA-based fault detection by using more sophisticated FD charts to achieve further improvements and widen the applicability of the process monitoring techniques in practice. The PCA presented here is investigated as modeling algorithm in the phase of fault detection. The fault detection problem is addressed so that the data are first modeled using the PCA algorithm and then the faults are detected using FD chart. The detection stage is related to the evaluation of detection charts, which are declares the presence of the fault. Those charts are computed using the PCA-based residual. The fault detection performance is illustrated through a simulated continuously stirred tank reactor (CSTR) data. The results demonstrate the effectiveness of the PCA-based FD chart methods for detecting the single and the multiple sensor faults.

Keywords: CSTR process; Fault detection; Generalized likelihood ratio test; Principal component analysis; Shewhart; Exponentially weighted moving average; Cumulative sum

Introduction

Effective operation of various engineering systems requires tight monitoring of some of their key process variables. For example, detection of anomalies in chemical systems is crucial for their efficient application on a controlled continuous stirred tank reactor (CSTR). Also, detecting aberrations in chemical data helps the diagnosis of various diseases. The fault detection problem is an important process in process monitoring. Abnormal faults management mainly depends on diagnosis of the process faults and accurate fault detection.

Various fault detection techniques have been developed and utilized in practice. For example, statistical fault detection techniques that are based on hypothesis testing, such the generalized likelihood ratio test (GLRT), have been shown to be among the most effective univariate fault detection methods. Most practical processes, however, are multivariate, i.e., involve many variables that need to be monitored at the same time. In a previous research effort, we have developed Principal Component Analysis (PCA) and kernel PCA (KPCA)-based GLRT fault detection schemes [1,2], in which PCA and KPCA have been used as a modeling framework for fault detection. We have also, developed a recursive PCA and KPCA methods for modeling and fault detection problems to processes where online fault detection is needed [3]. In this work, we will focus on fault detection problem based on more sophisticated statistical approaches. Different multivariate fault detection techniques have been proposed for process monitoring of such systems: such as chemical, environmental, power, etc. Faults detection has been performed manually using data visualization tools [4], but these tools are time consuming for real-time detection in streaming data. PCA is among the most popular statistical methods used for modeling and faults detection problems, however, it provides linear combinations of variables that demonstrate major trends in data

set. In mathematical terms, PCA provides linear combinations of a set of measured variables that capture major trends in data set. Specifically, PCA yields orthogonal vectors of high energy contents in terms of the variance of the data.

The main indices used with PCA methods are Hotelling statistic, T^2 and the sum of squared residuals, SPE, or Q statistic. The T^2 statistic is a measure of the variation captured in the PCA model and the Q statistic is a measure of the percent variance not captured by the PCA model. In the current work, we address the fault detection problem, in which the data are modeled using the PCA method and the faults are identified using the fault detection charts. The FD charts include: Hotelling statistic, T^2 , Q statistic, generalized likelihood ratio test (GLRT), shewhart control chart and exponentially weighted moving average chart (EWMA) control chart. In fact, PCA model has been shown to be suitable to obtain an accurate principal component of a set of data. The PCA algorithm is applied to obtain the model and find the combinations of parameters that describe the major trends in a data set [5,6] and FD chart is used to detect the faults and both are applied to enhance the fault detection process. The Shewhart chart is a simple univariate monitoring chart that utilizes process data without the application of filters. The Shewhart chart is mainly able

***Corresponding author:** Majdi Mansouri, Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar, Tel: 97477734583; E-mail: majdi.mansouri@qatar.tamu.edu

Received February 16, 2016; **Accepted** February 28, 2016; **Published** February 29, 2016

Citation: Mansouri M, Mohammed ZS, Baklouti R, Nounou M, Nounou H, et al. (2016) Statistical Fault Detection of Chemical Process - Comparative Studies. J Chem Eng Process Technol 7: 282. doi:10.4172/2157-7048.1000282

Copyright: © 2016 Mansouri M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

to detect fairly large fault. Other univariate charts, such as CUSUM and EWMA, through the application of filters are able to smaller faults [7]. The CUSUM statistic assumes each process observation is of equal weightage, while the EWMA statistic assigns an exponential weightage to consecutive observations [8]. The advantage of the CUSUM and EWMA charts in the detection of smaller faults can be attributed to their extensive process memory. Although, the CUSUM and EWMA charts may be able to better detect smaller faults than the Shewhart chart, they cannot be used to detect a wide range of fault sizes, as they often need to be tuned to detect faults of different sizes.

Therefore, a more robust chart, such as the GLRT chart might be required for fault detection. GLRT has been proposed in [9] in order to monitor an adaptive system, which reaches three important problems; estimation, fault detection and magnitude compensation of jumps. GLRT is proposed for fault detection of different applications: geophysical signal segmentation [6], signals and dynamic systems [5], incident fault detection on freeways [9], missiles trajectory [10]. Hence, in the current work, we propose to benefit from the advantages of the GLRT in order to improve the fault detection task in the cases where the process model is not available. The fault detection performance is illustrated through a simulated continuously stirred tank reactor (CSTR) data. The results show the performance of the PCA-based FD chart methods for detecting the single and the multiple anomalies.

The rest of the paper is organized as the following. In Section 2, an introduction to PCA method is given. Then, the FD charts descriptions are presented in Section 3. After that, the PCA-based FD chart method used for fault detection which integrates PCA modeling and FD control chart, is presented in Section 4. Next, in Section 5, the PCA-based FD chart performances are studied through a simulated continuously stirred tank reactor data. At the end, the conclusions are presented in Section 6.

Description of Principal Component Analysis Methods

PCA is a linear dimensionality reduction modeling technique, which is very helpful when dealing with data sets having a high degree of cross correlation among the variables [11]. Let $X_i \in \mathbb{R}^m$ denotes the i -th sample vector representing m different variables or sensors. Also, assume there are n samples dedicated to each variable or sensor, and then the data can be represented as a matrix $X \in \mathbb{R}^{n \times m}$, where each column corresponds to a variable and each row corresponds to a sample. After scaling each variable to have a zero mean and unit variance, the X matrix can be expressed as the multiplication of two matrices, a score matrix S and a loading matrix W , through singular value decomposition (SVD), i.e.,

$$X = SW^T \quad (1)$$

Where $S = [s_1 \ s_2 \ \dots \ s_m] \in \mathbb{R}^{N \times m}$ is a transformed variables matrix, $s_i \in \mathbb{R}^N$ are the score vectors or PCs, and $W = [w_1 \ w_2 \ \dots \ w_m] \in \mathbb{R}^{m \times m}$ is an orthogonal vectors matrix $w_i \in \mathbb{R}^m$ which includes the eigenvectors associated with the covariance matrix of X , i.e., Σ , which is given by,

$$\Sigma = \frac{1}{N-1} X^T X = W \Lambda W^T \text{ with } W W^T = W^T W = I_N, \quad (2)$$

Where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_m)$ is a diagonal matrix containing the eigenvalues related to the m PCs, $\lambda_1 > \lambda_2 > \dots > \lambda_m$ and I_N is the identity matrix [12].

It is important to note that the PCA model yields same number

of principal components (PCs) as the number of original variables (m). For collinear process variables, however, a smaller number of PCs (l) are required to capture most of the variations in the data. The effectiveness of the PCA model depends on the number of retained PCs. Several methods for determining the optimum number of PCs have been proposed, which include the Scree plot [13], the cumulative percent variance (CPV), the cross validation [14], and the profile likelihood [15-17]. In this study, the cumulative percent variance method is utilized to estimate the optimum number of retained PCs, which can be computed as follows:

$$CPV(l) = \frac{\sum_{i=1}^l \lambda_i}{\text{trace}(\Sigma)} \times 100, \quad (3)$$

After determining the number of PCs (l), the data matrix X can be written as,

$$X = SW = [\hat{S} \ \tilde{S}] [\hat{W} \ \tilde{W}]^T, \quad (4)$$

Where $\hat{S} \in \mathbb{R}^{N \times l}$ and $\tilde{S} \in \mathbb{R}^{N \times (m-l)}$, are matrices of \hat{X} retained PCs and the \tilde{X} , ignored PCs, respectively, and the matrices $\hat{W} \in \mathbb{R}^{m \times l}$ and $\tilde{W} \in \mathbb{R}^{m \times (m-l)}$ are matrices of l retained eigenvectors and the $(m-l)$ ignored eigenvectors, respectively. Using Equation (4), the following can be written,

$$X = \hat{S} \hat{W}^T + \tilde{S} \tilde{W}^T = \overbrace{X \hat{W} \hat{W}^T}^{\hat{X}} + \overbrace{X (\tilde{W} \tilde{W}^T)}^R, \quad (5)$$

Where the matrix \hat{X} represents the modeled variation of X based on first l components, and the matrix R represents the residuals.

Fault Detection Charts

The PCA model is used for fault detection through one of the detection statistics (T^2 , Q , GLRT, Shewart and EWMA) which are presented next.

Hotelling's T^2 statistic

The Hotelling's T^2 statistic is a way of measuring the variation captured in the principal components at various time samples, and it is expressed as [17]:

$$T^2 = X^T \hat{W} \hat{\Lambda}^{-1} \hat{W}^T X, \quad (6)$$

where $\hat{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_l)$, is a diagonal matrix containing the eigenvalues related to the l retained PCs. For new real-time data, when the value of T^2 statistic exceeds the threshold, T^2_1 calculated as in [17], a fault is detected. The threshold number used for the T^2 statistic is computed as [17]

$$T^2_\alpha = \frac{l(N-l)}{N-l} F_{l, N-l, \alpha} \quad (7)$$

where α is the level of significance (α usually between 1% and 5%), N is the number of samples in data set, l is the number of retained PCs, and $F_{l, N-l}$ is the Fisher F distribution with l and $(N-l)$ degrees of freedom. These thresholds are computed using faultless data. When the number of observations, N , is high, the T^2 statistic threshold is approximated with a χ^2 distribution with l degrees of freedom, i.e., $T^2_\alpha = \chi^2_{l, \alpha}$.

Q statistic or squared prediction error (SPE)

Another fault detection index is the squared prediction error SPE or Q statistic, which is a measure of the amount of variations not captured by the PCA model [17]. It can be computed as the sum of squares of the residuals [18], i.e.,

$$Q = \tilde{X}^2 = (I - \hat{W} \hat{W}^T) X^2 \quad (8)$$

Where, $\tilde{X} = X - \hat{X} = (I - \hat{W} \hat{W}^T) X$

The monitored system, meanwhile, is accepted to be in normal operation if,

$$Q \leq Q_\alpha \quad (9)$$

The threshold Q_α used for the Q statistic can be computed as [12],

$$Q_\alpha = \varphi_1 \left[\frac{h_0 c_\alpha \sqrt{2\varphi_2}}{\varphi_1} + 1 + \frac{\varphi_2 h_0 (h_0 - 1)}{\varphi_1^2} \right] \quad (10)$$

Where $\varphi_{i(i=1,2,3)} = \sum_{j=i+1}^m \lambda_j^i$, $h_0 = 1 - \frac{2\varphi_1\varphi_3}{3\varphi_2^2}$, α is level of confidence and c_α is the value of the normal distribution. For new data, the Q statistic is computed and compared to the threshold Q_α [12]. When the confidence limit is violated, a fault is declared. It is important to note that the threshold value is computed based on the assumption that the measurements are independent and follow a multivariate normal distribution; therefore, the Q statistic is highly sensitive to modeling errors [19].

Shewhart chart

Walter Shewhart developed the Shewhart chart in the 1920s, while working for Bell Systems [20]. Shewhart believed continuous process monitoring carried out at different stages during a process could prove to be more economical and effective, as opposed to inspecting the final product [21]. The Shewhart chart is widely used in practice for process monitoring, mainly due to its relative simplicity, as opposed to other univariate control charts [22]. Shewhart charts have three distinct features [23]: Center Line (C), Upper Control Limit (UCL), and Lower Control Limit (LCL). The center line typically represents the targeted process mean. The Shewhart chart is designed based on the following main assumptions [24]: the presence of a moderate level of noise in the evaluated residuals, the residuals being independent (uncorrelated), and the fault-free residuals following a normal (Gaussian) distribution. Numerous variations of the Shewhart chart are available. However, the most popular chart monitors the sample mean (\bar{X}). This chart is occasionally coupled with either the range (R) or standard deviation (S) chart, which increases the robustness of the Shewhart chart against the variability in observations collected from different sensors monitoring a particular process variable [25]. The (R) and (S) charts are able to spot features in the trend of the data, that might not necessarily come forth with the use of the \bar{X} chart only. If observations are collected from multiple sensors (or if sampling is carried out), the following equations can be used to compute the sample mean:

$$\bar{x} = \sum_{j=1}^n \frac{X_{ij}}{n} \quad (11)$$

and

$$\bar{\bar{x}} = \sum_{i=1}^k \frac{X_i}{k} \quad (12)$$

Where, n and k represent the subgroup size and the number of subgroups, respectively. Sub-grouping is generally carried out if observations from multiple sensors monitoring a particular process variable are available, or if sampling is carried out. The approach presented here assumes that only a single sensor provides measurements, and single readings are used, and hence the R and S charts are not required. Shewhart computed the limits for the control chart as follows [25]:

$$UCL = \bar{\bar{x}} + L_n \quad (13)$$

And

$$LCL = \bar{\bar{x}} - L_n \quad (14)$$

Where, $\bar{\bar{x}}$ is the targeted process mean and L_n is the control width computed by the following equation [25]:

$$L_n = \frac{c\sigma}{\sqrt{n}} \quad (15)$$

Where, σ is the standard deviation of the fault-free data set and the constant can be computed using a nomogram [25]. The scale corresponds to the range where a given percentage of fault-free observations should lie. Analysis of a nomogram for different processes can be time-consuming, and therefore it is common practice to use the following equation to compute the control limits [26]:

$$UCL = \bar{\bar{x}} + 3\sigma \quad (16)$$

And

$$LCL = \bar{\bar{x}} - 3\sigma \quad (17)$$

For a fault-free data test that follows a normal distribution, the value of 3σ accounts for nearly 99.73% of all deviation, which makes it a popular choice for the limits of the Shewhart chart [27]. The conventional Shewhart chart is unable to detect relatively small faults, as it is only able to detect faults larger than three times the standard deviation of the fault-free data set [25]. This can be attributed to the fact that the Shewhart chart only considers the current process measurement when deciding the presence or absence of a fault, and thus has a very short memory as indicated in Figure 1a. The insensitivity of the Shewhart chart to faults with small magnitudes is even more evident when the data is contaminated with high levels of noise, as features get masked by measurement noise. Other univariate monitoring schemes through application of linear filters, possess a longer memory than the conventional Shewhart as they utilize additional information from previous observations. The linear filters help deal with the assumption of noise to an extent. The CUSUM chart takes into account all previous observations Figure 1b, while the EWMA chart applies an exponentially weighted average filter Figure 1c. Although, the other univariate charts may show an improved performance when compared the Shewhart chart, their performance is limited by the same assumptions. Hence, it is important to find an alternative that will help address these concerns.

The CUSUM chart is effective in detecting small faults in process data. However, the extensive memory of the CUSUM chart Figure 1b increases the possibility of false alarms as the CUSUM statistic takes extra observations to return to the fault-free steady state values. Therefore, an approach such as the EWMA chart, that utilizes an exponential filter might prove useful, and will be described next.

Exponentially weighted moving average (EWMA) chart

The exponentially weighted moving average chart (EWMA) chart was developed by Roberts in 1959 and was initially referred to as the Geometric Moving Average (GMA) chart [28]. Over time the GMA chart became popularly known as the EWMA chart [29]. Similar to the CUSUM chart, the EWMA chart is able to detect smaller faults shifts in the mean when compared to the Shewhart chart [30,31].

The EWMA statistic can be computed by [32]:

$$z_i = \lambda X_i + (1 - \lambda) z_{i-1} \quad (18)$$

Where, λ is the smoothing parameter (exponential filter) that can be assigned a value between 0 and 1. The smoothing parameter controls the memory of the process, i.e., a value closer to 0 placing less emphasis on more recent observations, and vice versa. The control limits for the EWMA chart can be computed using [7]:

$$UCL = \bar{\bar{x}} + L\sigma \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2i}]} \quad (19)$$

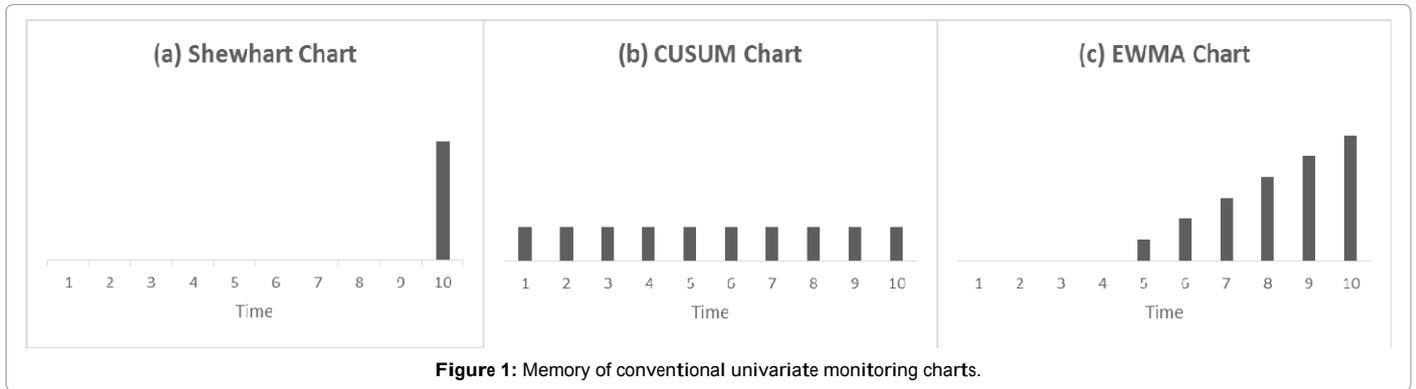


Figure 1: Memory of conventional univariate monitoring charts.

And

$$LCL = \bar{x} - L\sigma \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2i}]} \quad (20)$$

Where, L is defined as the control width of the EWMA chart. At steady state $[1 - (1-\lambda)^{2i}]$ simplifies to unity, and the following steady state values are obtained [7]:

$$UCL = \bar{x} + L\sigma \sqrt{\frac{\lambda}{2-\lambda}} \quad (21)$$

And

$$LCL = \bar{x} - L\sigma \sqrt{\frac{\lambda}{2-\lambda}} \quad (22)$$

Although, the CUSUM and EWMA charts are better able to detect smaller faults, they are not capable of detecting a wide range of fault sizes. Hence, a more robust approach such as the GLRT chart might be required and will be described next.

Generalized likelihood ratio test (GLRT)

The GLRT is a hypothesis testing technique which has been utilized successfully in model-based fault detection [5,9,10]. Let, $Y \in R^N$ be an observation vector formed by one of the two Gaussian distributions: $\mathcal{N}(0, \sigma^2 I_N)$ or $\mathcal{N}(\theta, \sigma^2 I_N)$, where θ is the mean vector (which is the value of the fault) and $\sigma^2 > 0$ is the variance (assumed to be known in this problem). The hypothesis test can be expressed as,

$$\begin{cases} \mathcal{H}_0 = \{Y \sim \mathcal{N}(0, \sigma^2 I_N)\}, & \text{(null hypothesis),} \\ \mathcal{H}_1 = \{Y \sim \mathcal{N}(\theta, \sigma^2 I_N)\}, & \text{(alternative hypothesis)} \end{cases} \quad (23)$$

Here, the GLRT method replaces the unknown parameter, θ , by its maximum likelihood estimate. This estimate is computed by maximizing the GLRT $\mathbb{T}(Y)$ as follows,

$$\begin{aligned} \mathbb{T}(Y) &= 2 \log \frac{\sup_{\theta \in R^N} f_{\theta}(Y)}{f_{\theta=0}(Y)} \\ &= 2 \log \left\{ \sup_{\theta} \exp \left\{ -\frac{Y - \theta_2^2}{2\sigma^2} \right\} / \exp \left\{ -\frac{Y_2^2}{2\sigma^2} \right\} \right\} \\ &= \frac{1}{\sigma^2} \left\{ \min_{\theta} Y - \theta_2^2 + Y_2^2 \right\} \\ &= \frac{1}{\sigma^2} \left\{ Y - \hat{\theta}_2^2 + Y_2^2 \right\} = \frac{1}{\sigma^2} \{Y_2^2\} \end{aligned} \quad (24)$$

Where $\hat{\theta} = \arg \min Y - \theta_2^2$ is the maximum likelihood estimate of θ , the probability density function of Y is $f_{\theta}(Y) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma^N} \exp \left\{ -\frac{1}{2\sigma^2} Y - \theta_2^2 \right\}$ and $\|\cdot\|_2$ represents the Euclidean norm. Because the GLRT utilizes the ratio of the distributions of the faulty and fault-free data, in the case of non-Gaussian variables, non-Gaussian distributions need to be used. It must be noted that, in the derivation shown above, maximizing the

likelihood function is equivalent to maximizing its natural logarithm since the logarithmic function is a monotonic one. The GLRT then decides between the hypotheses H_0 and H_1 as follows,

$$\delta(Y) = \begin{cases} \mathcal{H}_0 & \text{if } \mathbb{T}(Y) < t_{\alpha}, \\ \mathcal{H}_1 & \text{else.} \end{cases} \quad (25)$$

Here, the distribution of the decision function $\mathbb{T}(Y)$ under $T_{\alpha}^2 = X_{i,\alpha}^2$ allows designing a statistical test with a desired false alarm rate, α , where the threshold t_{α} is chosen to satisfy the following false alarm probability,

$$\mathbb{P}_0(\mathbb{T}(Y) \geq t_{\alpha}) = \alpha \quad (26)$$

Where $\mathbb{P}_0(\mathbf{A})$ represent the probability of an event A when Y is distributed according to the null hypothesis H_0 and α is the desired probability of the false alarm. Since Y is normally distributed (equation (23)), the statistics \mathbb{T} is distributed according to the X^2 law with (m-l) degrees of freedom. This law is central under H_0 and noncentral under H_1 with a parameter of non-centrality equal to: $\kappa_{\theta} = \frac{1}{\sigma^2} \theta_2^2$. Also, the power function of δ can be calculated as,

$$\beta_{\delta} = \mathbb{P}_0(\delta(Y) = \mathcal{H}_1) \quad (27)$$

To select an appropriate threshold for the GLRT statistic, its distribution needs to be determined. Since the noise is assumed to follow a Gaussian distribution, the test statistic will follow a chi-square distribution [33]. The normalized residual $\bar{R} \delta$ is distributed as:

$$\bar{R} \delta \sim \mathcal{N}(\theta, \sigma^2 I_N) \quad (28)$$

where $\theta=0$ under the null hypothesis (26). Then, the scaled test statistic is distributed as the non-central chi-square distribution as follows,

$$\mathbb{T} = \frac{1}{\sigma^2} \{Y_2^2\} \sim X_N^2 \quad (29)$$

with N degrees of freedom. Since the GLRT is applied online, the norm used in its statistic is computed using only the current data sample, and thus, the GLRT statistic follows a chi-square distribution with a degree of freedom equal 1.

Fault Detection Using PCA-based FD Chart Method

In this section, PCA is combined with FD chart to develop new fault detection with more sensitivity to small data faults. The PCA method is investigated here as modeling framework in the task of fault detection. The residuals of the response variables from PCA model can be assigned control limits. The proposed scheme can be used to detect the existence or lack of faults [34]. Under normal operating conditions (no faults), the residual of the monitored model is zero or close to zero when modeling measurement noise and uncertainties. However, in the presence of a fault the residuals differ significantly from zero,

showing the existence of a new state that can be clearly distinguished from the normal faultless working mode [35]. Here, FD chart is used to improve the process monitoring by using a more appropriate and accurate model. Due to the capacity of the FD chart to detect drifts with low severity in the data, this technique is appropriate for enhancing the detection of small or moderate faults. Thus, the PCA is used to create the model and find an accurate combinations of parameters which describe the major trends in a data set [5,6] and FD chart is used to detect the faults and both are utilized to improve faults detection process.

PCA-based FD chart process monitoring algorithm

Here, the FD chart is obtained using the residuals of the responses variables from the PCA model. Let the matrices X , \hat{X} and R be defined as follows: $X = [X_1 \ X_2 \ \dots \ X_m] \in \mathbb{R}^{N \times m}$, $R = [R_1 \ R_2 \ \dots \ R_m] \in \mathbb{R}^{N \times m}$ and $\hat{X} = [\hat{X}_1 \ \hat{X}_2 \ \dots \ \hat{X}_m] \in \mathbb{R}^{N \times m}$ let X_j , \hat{X}_j , and R_j be the j -th columns of the matrices, X , \hat{X} , and R , respectively. In the absence of a fault, the residual can be calculated as follows:

$$R = X - \hat{X}, \quad (30)$$

The difference between the observed value of the input variable, X , and the predicted value, \hat{X} , obtained from PCA model represent the residual of the input variable, $R = [R_1 \ R_2 \ \dots \ R_m] \in \mathbb{R}^{N \times m}$ which can be used as an indicator to detect a possible fault. Then, the FD chart decision function based on the residuals of the response variable can be computed using one of the fault detection chart described above using equation (6) for T^2 statistic, equation (8) for Q statistic, equation (11) for Shewhart, equation (18) for EWMA and equation (24) for GLRT statistic.

The developed PCA-based FD chart fault detection method can be implemented as described in Algorithm 1, and its performance is illustrated in the next section through its application to monitor the operation of a chemical reactor.

Algorithm 1: PCA-based FD fault detection algorithm

Input: $N \times m$ data matrix X , Confidence interval α

Output: FD statistic, FD threshold

- **Data preprocessing step:**

Standardize: computes data's mean and standard deviation, and standardize it

- **PCA running step:**

Compute the covariance matrix

Calculate the eigenvalues and eigenvectors and sort the eigenvalues in decreasing order

Compute the optimal number of principal components to be used using the CPV method

- **Compute the sum of approximate and residual matrices**

PCA testing step

Standardize the new data

Compute the FD chart decision function

Compute the FD chart control limits

Compute the FD chart statistic for the new data

Declare a fault when the FD chart decision function, exceeds the

control limits.

Next, we present the developed PCA-based FD chart process monitoring algorithm for fault detection of chemical process.

PCA-based Charts and Application to Fault Detection in Simulated CSTR Model

Next, the developed PCA-based FD chart algorithm presented is illustrated through its application on a controlled continuous stirred tank reactor (CSTR) in which a non-isothermal, irreversible first order reaction $A \rightarrow B$ takes place. Next, the CSTR model that is used for fault detection is described.

CSTR process description

The dynamic model for the non-isothermal CSTR can be given by [36,37],

$$\begin{aligned} \frac{\partial C_A}{\partial t} &= \frac{F}{V}(C_{A_0} - C_A) - k_0 e^{-E/RT} C_A \\ \frac{\partial T}{\partial t} &= \frac{F}{V}(T_0 - T) + \frac{(-\Delta H)k_0}{\rho C_p} e^{-E/RT} C_A - \frac{q}{V\rho C_p} \\ q &= \frac{aF_c^{b+1}}{F_c + \left(\frac{aF_c^b}{2\rho C_{pc}}\right)} (T - T_{cin}) \end{aligned} \quad (31)$$

Where k_0 is the reaction rate constant, E is the activation energy, C_A is the concentration of "A" in the inlet stream, C_B is the concentration of "B" in the exit stream, T is the temperature of the inlet stream, F is the flow rate in and out of the reactor, V is the reactor volume, T_i is the temperature of exit stream, T_{ci} is the temperature of the cooling fluid in the jacket, ΔH is the heat of reaction, U is the overall heat transfer coefficient, A is the area through which heat transfers from the reactor to the cooling jacket, and ρ and C_p are the density and heat capacity of the reactor contents and of all streams. Assuming a stoichiometric proportion of compounds "A" and "B" in the feed, one can assume that $C_B(t) = 2C_A(t)$. The outlet temperature (T) and the concentration (C_A) are controlled using proportional integral (PI) controllers by manipulating the inlet coolant flow rate (F_c) and the feed flow rate (F), respectively. The parameters of the PI controllers are as follows: $K_{c1} = -0.8$ and $\tau_i = 0.1$ for the temperature controller, and $K_{c2} = 2$ and $\tau_i = 0.1$ for the concentration controller.

Generation of dynamic data

In a practical setting, the data would be collected by changing the feed flow rate (which is chosen in this example to be the model input, i.e., F), and then measuring the state variables, i.e., the concentration and temperature as functions of time. Thus, the data are generated given some pre-defined model parameters.

The CSTR model parameters as well as other physical properties are shown in Table 1. The simulated CSTR is used to generate training and testing data sets by changing the set points of the concentration and temperature controllers in step-wise fashions. The process data used in training includes four variables, the coolant flow rate (F_c), the feed flow rate (F), the outlet concentration C_A , and the reactor outlet temperature T . Thus, the data matrix, which has 1000 rows and 4 columns, is used to construct the PCA model after scaling the variables.

Next, the performance of the developed PCA-based FD chart fault detection method is illustrated and compared to PCA through its two charts Q , T^2 . The comparison is assessed through three different cases studies representing three different types of faults. In the first case study, the sensor measuring the concentration of A (C_A) is assumed to

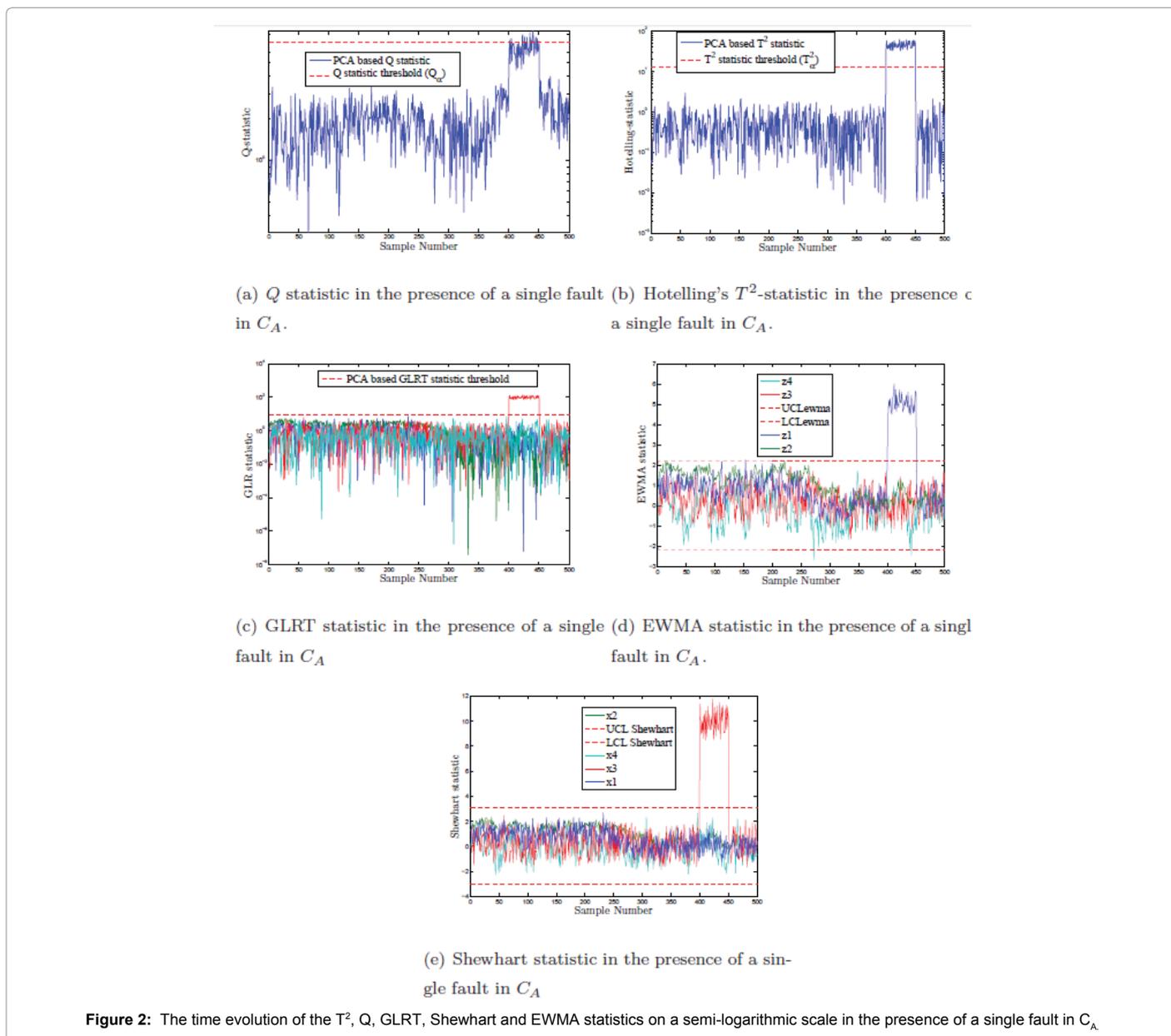
be faulty with single as well as multiple faults. In the second case study, similar faults (single and multiple) are introduced in the temperature of the reactor (T). In third case study, multiple faults are assumed to occur simultaneously in the concentration and temperature inside the reactor.

Case 1: Faults in the concentration C_A

The testing data used to evaluate the fault detection performances, which consist of 500 samples, are generated using the CSTR model described earlier. To simulate a single fault in the state variable C_A , an additive fault having a magnitude 20% of the total variation in C_A is introduced between samples 100 and 150. The results using the PCA-based Q technique Figure 2a show that it could successfully detect this single fault but with some false alarms. While, the performance of the

Parameter	Value	Parameter	Value
E(J/mol)	76534	V (l)	100
$-\Delta H$ (J/mol)	596619	ρ (g/l)	1000
k_0 (l/min.mol)	4.11×10^{13}	cp (J/g.K)	4.2
C_{A1} (mol/l)	1	T_1 (K)	250
T_1 (K)	350	UA (W.K)	5×10^4
R (J/mol.K)	8.31451		

Table 1: CSTR model parameters and physical properties.



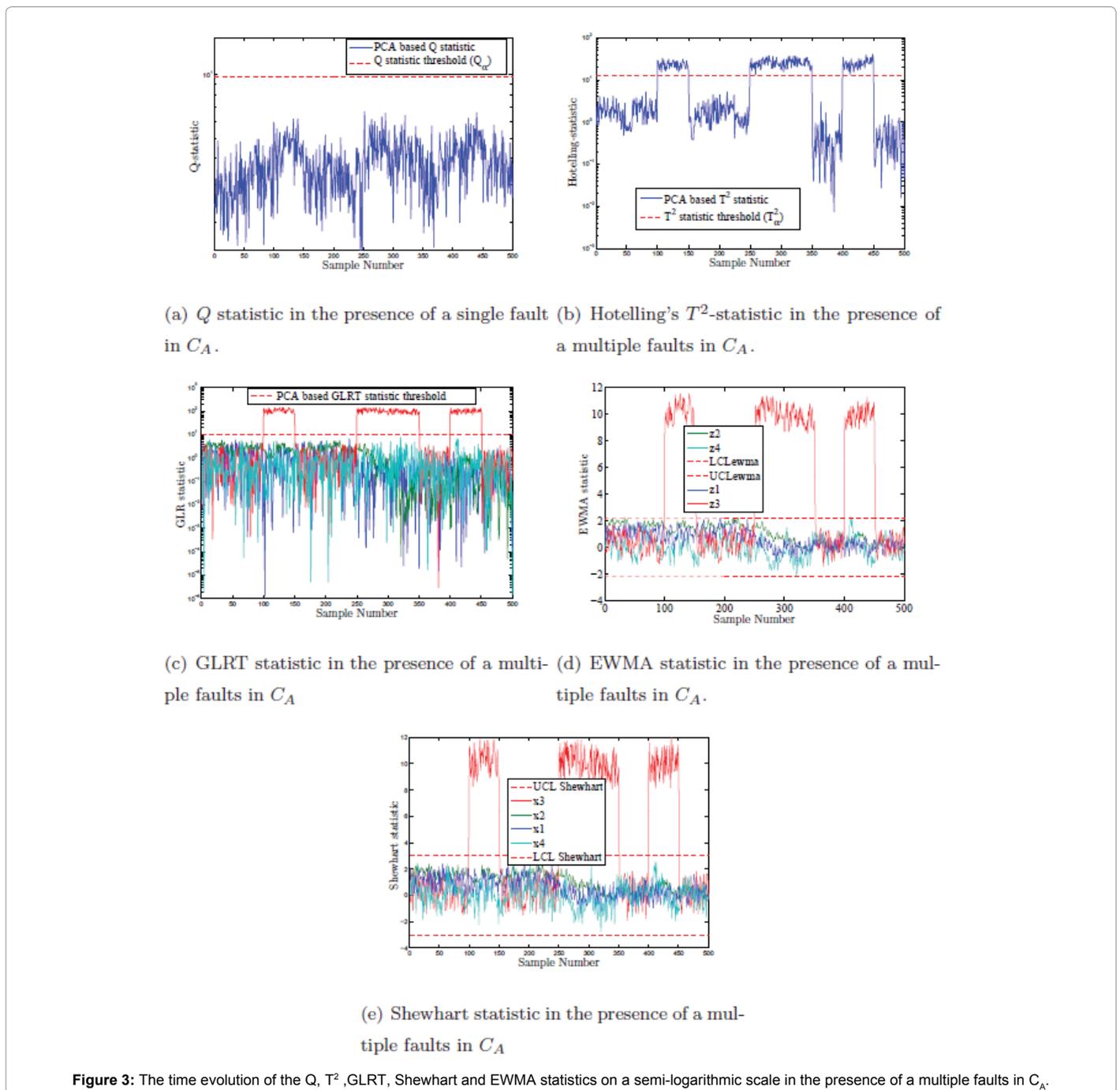
PCA-based T^2 , GLRT, Shewhart and EWMA methods, on the other hand, Figures 2b- 2e, shows that they could detect this single fault without any false alarms.

In the presence of a multiple faults in C_A , we can show from Figures 3a- 3e the results of the process monitoring of CSTR process using the PCA-based FD chart techniques. The PCA based Q technique cannot detect these faults Figure 3a. However, the PCA-based T^2 , Q, GLRT, Shewhart and EWMA methods can detect the faults effectively (as shown in Figures 3b-3e).

Case 2: Faults in the temperature T

In this case study, the sensor measuring the temperature T is assumed to be faulty with single as well as multiple faults. First, a single fault in the reactor temperature represented by a constant bias of amplitude equal 5% of the total variation in T is introduced between the sample numbers 100 and 150. Figures 4a- 4e show the ability of the PCA-based T^2 , Q, GLRT, Shewhart and EWMA methods to detect this additive fault, while the PCA-based Q technique results in some missed detections Figure 4b. However, the PCA-based T^2 method cannot detect this additive fault as shown in Figure 4b.

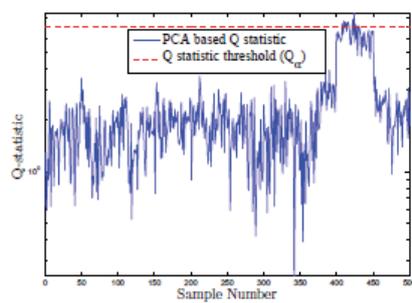
Case 3: Faults in the concentration C_A and temperature T



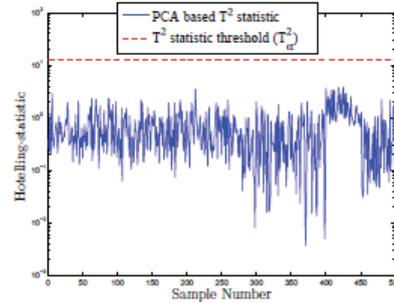
In this case study, simultaneous faults are introduced in both the concentration and temperature (each of which is represented by a bias of magnitude equal 20% of the variation in its corresponding variable). The results using the PCA-based T^2 , Q, GLRT, Shewhart and EWMA techniques for these multiple faults are shown in Figures 5a-5e. These results show that the PCA-based Q and PCA-based T^2 techniques could detect these multiple faults. The PCA-based GLRT, Shewhart and EWMA techniques, however, are capable to detect these faults without any false alarms as shown in Figures 5c-5e.

Conclusion

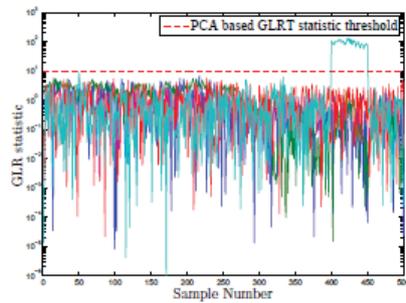
In this paper, principal component analysis (PCA)-based fault detection (FD) charts are used for fault detection. The FD charts include: Hotelling statistic, T^2 and Q statistic, generalized likelihood ratio test (GLRT), Shewhart control chart and exponentially weighted moving average chart (EWMA) control chart. The fault detection problem was addressed in which the data are modeled using the PCA method and the faults are identified using the fault detection charts. The PCA method is applied here as modeling framework in the phase of fault detection. The idea is to improve the FD control chart performance introducing modeling of the data using the PCA. The PCA-based FD chart fault detection performances are assessed through



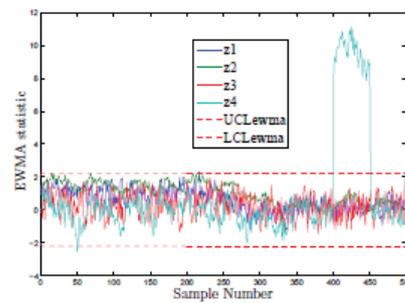
(a) Q statistic in the presence of a single fault in T .



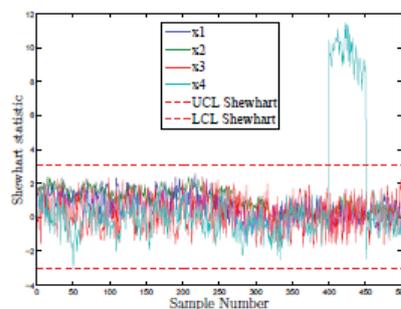
(b) Hotelling's T^2 -statistic in the presence of a single fault in T .



(c) GLRT statistic in the presence of a single fault in T .

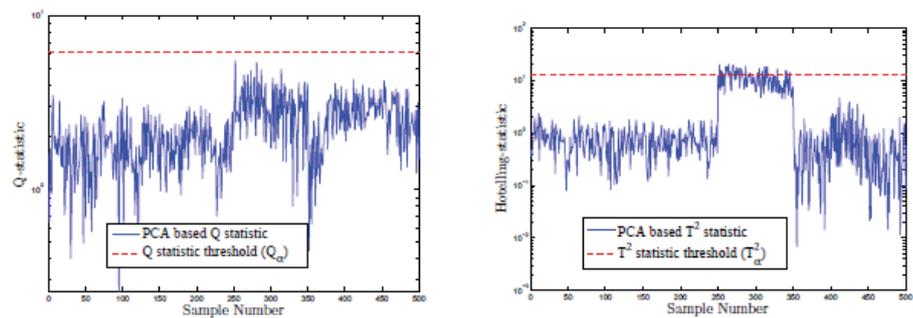


(d) EWMA statistic in the presence of a single fault in T .

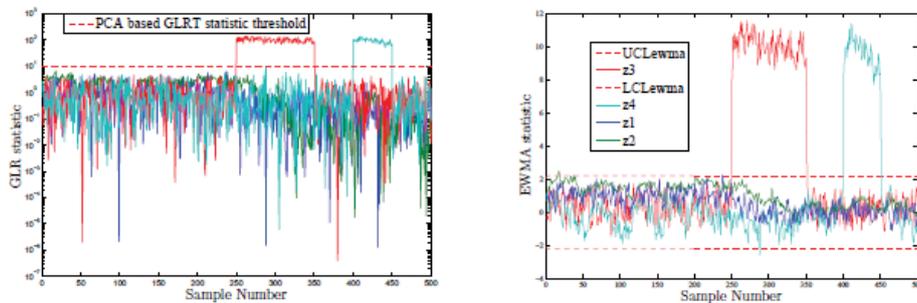


(e) Shewhart statistic in the presence of a single fault in T

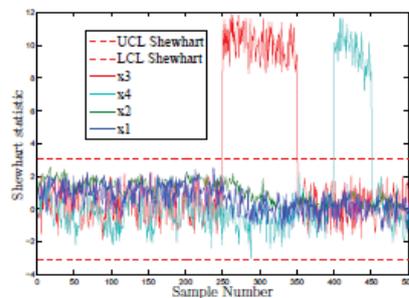
Figure 4: The time evolution of the T^2 , Q, GLRT, Shewhart and EWMA statistics on a semi-logarithmic scale in the presence of a single fault in T .



(a) Q statistic in the presence of simultaneous faults in C_A and T (b) Hotelling's T^2 -statistic in the presence of simultaneous faults in C_A and T .



(c) GLRT statistic in the presence of simultaneous faults in C_A and T . (d) EWMA statistic in the presence of simultaneous faults in C_A and T .



(e) Shewhart statistic in the presence of simultaneous faults in C_A and T .

Figure 5: The time evolution of the T^2 , Q , GLRT, Shewhart and EWMA statistics on a semi-logarithmic scale in the presence of simultaneous faults in C_A and T .

a simulated continuously stirred tank reactor (CSTR) data. The results show the performance of the PCA-based FD chart techniques over the conventional PCA through its two charts T^2 and Q for detecting a single and multiple sensor faults.

Acknowledgments

This work was made possible by NPRP grant NPRP7-1172-2-439 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

1. Mansouri M, Destain MF, Nounou H, Nounou M (2016) Enhanced monitoring of environmental processes. International Journal of Environmental Science and

Development 7: 525-531.

- Mansouri M, Nounou M, Nounou H, Karim N (2016) Kernel pca-based glrt for nonlinear fault detection of chemical processes. Journal of Loss Prevention in the Process Industries 40: 334-347.
- Baklouti R, Mansouri M, Nounou M, Nounou H, Hamida AB (2015) Iterated robust kernel fuzzy principal component analysis and application to fault detection. Journal of Computational Science.
- Mourad M, Bertrand-Krajewski JL (2002) A method for automatic validation of long time series of data in urban hydrology. Water Sci Technol 45: 263-270.
- Gustafsson F (1996) The marginalized likelihood ratio test for detecting abrupt changes. IEEE Transactions on Automatic Control 41: 66-78.

6. Nguyen D, Widrow B (1990) Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *IJCNN International Joint Conference, IEEE, USA*. pp: 21-26.
7. Montgomery DC (2005) *Introduction to statistical quality control*. John Wiley & Sons, New York.
8. Lowry CA, Woodall WH, Champ CW, Rigdon SE (1992) A multivariate exponentially weighted moving average control chart. *Technometrics* 34: 46-53.
9. Willsky AS, Chow E, Gershwin S, Greene C, Houpt P, et al. (1980) Dynamic model-based techniques for the detection of incidents on freeways. *IEEE Transactions on Automatic Control* 25: 347-360.
10. Dawdle JR, Willsky A, Gully SW (1982) Nonlinear generalized likelihood ratio algorithms for maneuver detection and estimation. *American Control Conference, IEEE Publishers, USA*. pp: 985-987.
11. Tipping ME, Bishop CM (1999) Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61: 611-622.
12. Jackson J, Mudholkar G (1979) Control procedures for residuals associated with principal component analysis. *Technometrics* 21: 341-349.
13. Zhu M, Ghodsi A (2006) Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics and Data Analysis* 51: 918-930.
14. Diana G, Tommasi C (2002) Cross-validation methods in principal component analysis: A comparison. *Statistical Methods and Applications* 11: 71-82.
15. Jolliffe I (2002) *Principal component analysis*. 2nd edn, Springer, Berlin.
16. Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24: 417-441.
17. Russell EL, Chiang LH, Braatz RD (2000) Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 51: 81-93.
18. Qin S (2003) Statistical process monitoring: Basics and beyond. *Journal of Chemometrics* 17: 480-502.
19. Benaicha A, Guerfel M, Boughila N, Benothman K (2010) New pca-based methodology for sensor fault detection and localization. *MOSIM'10, Hammamet, Tunisia*. pp: 10-12.
20. Shewhart W (1930) Economic quality control of manufactured product 1. *Bell System Technical Journal* 9: 364-389.
21. Shewhart WA, Deming WE (1939) Statistical method from the viewpoint of quality control. *The Graduate School, the Department of Agriculture, Washington DC*. p: 155.
22. Reynolds MR, Stoumbos ZG (2004) Should observations be grouped for effective process monitoring? *Journal of Quality Technology* 36: 343-366.
23. Cinar A, Palazoglu A, Kayihan F (2007) Chemical process performance evaluation. *CRC press*. p: 344.
24. Cinar A, Parulekar SJ, Undey C, Birol G (2003) Batch fermentation: modeling, monitoring, and control. *CRC Press*. p: 648.
25. Shewhart WA (1926) Quality control charts 1. *Bell System Technical Journal* 5: 593-603.
26. Montgomery DC, Runger GC (2010) *Applied statistics and probability for engineers*. John Wiley & Sons.
27. Burr JT (2004) *Elementary statistical quality control*. 2nd edn, CRC Press. p: 480.
28. Roberts SW (1959) Control chart tests based on geometric moving averages. *Technometrics* 1: 239-250.
29. Hunter S (1995) Just what does an ewma do. *ASQC Statistics Division Newsletter* 16: 4-12.
30. Carmo CV, Lopes LFD, Souza AM (2004) Comparative study of the performance of the cusum and ewma control charts. *Computers and Industrial Engineering* 46: 707-724.
31. Chao-Wen L, Reynolds MR (2001) Cusum charts for monitoring an auto correlated process. *Journal of Quality Technology* 33: 316.
32. Hunter JS (1986) The exponentially weighted moving average. *J Quality Technol* 18: 203-210.
33. Kay SM (1998) *Fundamentals of statistical signal processing: Detection theory*. Business and Economics, the University of California, USA. p: 672.
34. Kinnaert M (2003) Fault diagnosis based on analytical models for linear and nonlinear systems - a tutorial. *Proceedings of the 15th International Workshop on Principles of Diagnosis*, pp: 37-50.
35. Nyberg M, Nyberg CM (1999) *Model based fault diagnosis: Methods, theory, and automotive engine applications*. PhD Thesis.
36. Vajesta A, Schmitz R (1970) An experimental study of steady-state multiplicity and stability in an adiabatic stirred reactor. *AIChE Journal* 3: 410-419.
37. Mansouri MM, Nounou HN, Nounou MN, Datta AA (2014) State and parameter estimation for nonlinear biological phenomena modeled by s-systems. *Digital Signal Processing* 28: 1-17.