Chapter

Fault Detection of Single and Interval Valued Data Using Statistical Process Monitoring Techniques

Mohammed Ziyan Sheriff, Nour Basha,
Muhammad Nazmul Karim, Hazem Nounou
and Mohamed Nounou

Abstract

Principal component analysis (PCA) is a linear data analysis technique widely used for fault detection and isolation, data modeling, and noise filtration. PCA may be combined with statistical hypothesis testing methods, such as the generalized likelihood ratio (GLR) technique in order to detect faults. GLR functions by using the concept of maximum likelihood estimation (MLE) in order to maximize the detection rate for a fixed false alarm rate. The benchmark Tennessee Eastman Process (TEP) is used to examine the performance of the different techniques, and the results show that for processes that experience both shifts in the mean and/or variance, the best performance is achieved by independently monitoring the mean and variance using two separate GLR charts, rather than simultaneously monitoring them using a single chart. Moreover, single-valued data can be aggregated into interval form in order to provide a more robust model with improved fault detection performance using PCA and GLR. The TEP example is used once more in order to demonstrate the effectiveness of using of interval-valued data over single-valued data.

Keywords: principal component analysis, generalized likelihood ratio, hypothesis testing, fault detection, Tennessee Eastman Process, interval data

1. Introduction

Current technological advancements allow data to be collected from a number of different sources. The availability of abundant data collected from different sensors is beneficial, as they can be utilized in order to observe trends between and within different measured process variables. This allows process models to be developed in order to help identify if different processes or applications are behaving as expected [1]. Additionally, with industrial growth present in many developing countries, efficient process monitoring is essential for newer and more complex processes. Monitoring of these processes is required in order to ensure process safety, maintain product quality, increase economic benefits, and also to ensure that the process adheres to strict environmental regulation standards [2].

Statistical process monitoring methods can be classified into three broad categories: quantitative model based methods, qualitative model based methods, and process history based methods [3–5]. Quantitative model based methods require detailed knowledge of a process in order to construct a model that can be used for monitoring, for example, Kalman filters [3], while qualitative model based methods require the presence of process engineering experts in order to develop monitoring procedures or tasks, for example, fault trees [4]. In the absence of these two requirements, and due to the complexity of many processes that require monitoring, data-based techniques are often commonly used by the industry for various applications from drug design, to drinking water treatment [5–7].

Principal component analysis (PCA) is a powerful, linear data analysis technique widely used in research and industrial applications [8], for fault detection and isolation, data modeling and reconstruction, feature extraction, and noise filtration. PCA is useful for the extraction of dominant underlying information from a dataset, without any previous knowledge of the model. An example of the practical application of PCA has been discussed in [8], where data gathered from parallel sensors are used to quantify the quality of a given food sample. PCA is used to reduce the dimensionality of a dataset, whilst filtering out variability caused by noise [9]. The PCA model has been utilized in order to monitor a wide variety of processes, and has seen many extensions [10–13]. Two main fault detection statistics are typically utilized with a PCA model: Hotelling's T^2 statistic, and the Q statistic [10]. Variations captured by the principal component space are monitored using the T^2 statistic, while variations in the residual space are monitored using the Q statistic [14].

On the other hand, statistical hypothesis testing methods function by using statistical techniques in order to determine if observations collected from a given process follow the null hypothesis, that is, operating under normal operating conditions, or alternate hypothesis, that is, operating under abhorrent or faulty operating conditions [15]. These faults can be of different types, such as shifts in the mean, variance, or both. The generalized likelihood ratio (GLR) technique has received a lot of attention in process monitoring literature [10, 11, 13, 16]. The GLR method aims to maximize the detection rate for a fixed false alarm rate [15]. Therefore, an objective of this work is to provide a comparative review of the different GLR charts by utilizing examples such as the benchmark Tennessee Eastman Process (TEP) [17].

Data utilized in the construction of a PCA model may be of two types depending on the application being monitored: single-valued, and interval-valued. Single-valued data can be directly obtained from sensors measuring particular variables in a process, while interval-valued data is aggregated or artificially generated from batch single-valued measurements, thereby resulting in a range of possible measurement values for a given process variable at one time instant. The use of interval data in fault detection was originally introduced in order to reduce large datasets to a more manageable size [18], without compromising the integrity of the dataset. In addition, the use of interval data is beneficial because of its inherent ability to deal with missing values in samples, which may happen due to malfunctioning sensors or varying sampling frequencies between variables [19].

However, in cases where reducing the dataset may not be a viable option, due to a relatively limited sample size or sampling frequency, the use of interval data can be applied using a moving window aggregation method. This is also true of applications where batch process monitoring is not a viable option, thereby necessitating the need for real-time online monitoring of samples. The benchmark TEP example will be used once more in order to analyze the benefit of using moving window interval aggregation on the fault detection performance of PCA and GLR.

The rest of this chapter will be organized as follows. In Section 2, a more detailed introduction to PCA is provided along with a quick overview of the fault detection statistics used to examine the fault detection performance of the methods discussed in this paper. Section 3 will introduce hypothesis testing methods and the different GLR charts. In Section 4, the moving window interval aggregation method is explained, as well as its integration with PCA and GLR for the purposes of fault detection. Section 5 then presents illustrative examples using simulated synthetic data and TEP using a PCA-based GLR technique, used to demonstrate the effect that using GLR and interval data has on the fault detection performance. Conclusions are then presented in Section 6.

2. Principal component analysis (PCA)

Principal component analysis (PCA) is a linear dimensionality reduction tool used to reduce the number of variables in a dataset, whilst retaining most of the data's variability. PCA finds a new set of variables, called principal components, using a linear combination of the dataset's original cross-correlated variables [9]. The algorithm for PCA is summarized below.

2.1 PCA algorithm

Given a $n \times p$ classical training dataset X, where n is the number of sample rows and p is the number of variable columns, the PCA model is found as follows:

1. Find the correlation matrix R of X.

2. Find the column eigenvectors matrix P and the diagonal eigenvalues matrix Λ of R. Each eigenvector defines the linear combination coefficients used to find the principal components from the original variables, and each eigenvalue represents the amount of variance that its respective principal component covers in the dataset.

3. Retain l principal components that cover the minimum desired variability in the dataset, denoted as $\hat{P}$.

4. Find the predictive transformation matrix, $\hat{C} = \hat{P}\hat{P}^T$.

5. Find the residual transformation matrix, $\tilde{C} = 1 - \hat{C}$.

$\hat{C}$ is used to find the projection of the dataset onto the PCA model, and $\tilde{C}$ is used to find the amount of deviation of the dataset from its projection onto the PCA model, also known as the matrix of residuals. For more comprehensive details, please refer to [9, 19, 20].

The training dataset X defines the system under normal or optimal operating conditions, where there are no faults and the noise is minimal. Consequently, X is used to find the PCA model, defined using $\hat{C}$ and $\tilde{C}$ transformation matrices. The testing dataset S defines the system under unknown operating conditions, and it is monitored for faults using its respective residuals $\tilde{S} = S \cdot \tilde{C}$, as will be discussed later.

2.2 Fault detection statistics

Knowing the optimal number of eigenvectors or principal components to retain, fault detection is then carried out by evaluating the PCA model's residuals using any detection statistic. This section will focus on briefly introducing the two most well-known statistics in literature: The Q and T^2 statistics.

The Q-statistics of a $n \times p$ classical residual matrix $\tilde{X}$ is defined as [11]:

$$Q_x[i] = \sum_{j=1}^{p} \left(\tilde{X}_j[i]\right)^2 \tag{1}$$

Q_x is used to find the Q-threshold value γ, which defines the maximum possible value for a testing data's Q-statistic, denoted as Q_s, beyond which the sample will be declared as a fault [14, 19, 21]. The threshold is calculated using the empirical cumulative distribution function (CDF) of Q_x, which is an estimate of the true CDF of its discrete values.

The fault detection performance is tabulated by comparing Q_s with γ. If $Q_s[i]{>}\gamma$, then the ith sample is declared as faulty, otherwise it is normal. There are two metrics used for benchmarking each method: false alarm rate (FAR) and detection rate (DR).

FAR is the average percentage of samples that were wrongfully declared as faults. The detection rate is the average percentage of samples that were rightfully declared as faults. It is desirable to maximize DR, for a fixed FAR, in order to have a better fault detector.

Alternatively, the Hotelling T^2 statistic, which measures variations in the principal component space can be used, is computed as follows [22]:

$$T^2 = x^T \hat{P} \hat{\Lambda}^{-1} \hat{P}^T x, \tag{2}$$

where, $\hat{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_l)$, is a diagonal matrix that contains the eigenvalues that are associated with the l retained principal components The threshold for the T^2 statistic can be computed either computational or empirically [22]. The Q statistic is often utilized by authors instead of the T^2 statistic as it better able to detect smaller faults [10, 11].

3. Hypothesis testing methods

Hypothesis testing methods such as the generalized likelihood ratio (GLR), have received a lot of attention in recent literature [10, 13, 23]. Hypothesis testing methods utilize fundamental statistical theory in order to determine if given data conforms to a targeted distribution, that is, a null hypothesis, or deviates from this distribution, and follows an alternative distribution, that is, an alternate hypothesis [15]. In process monitoring terms, the parameters of the null and alternate hypotheses are defined using data from normal and abhorrent operating conditions, respectively [1].

3.1 Generalized likelihood ratio

The generalized likelihood ratio (GLR) technique defines the alternate hypotheses by parameters that can assume an infinite number of values, and is therefore called a composite hypothesis. An efficient point estimation method that utilizes the concept of maximum likelihood estimates (MLEs) is employed in order to estimate the required parameters.

The univariate GLR chart uses the concept of maximum likelihood estimates in order to maximize the detection rate for a fixed false alarm rate. The GLR process is accomplished through the following steps [15]:

1. The null and alternate hypotheses are defined, and their respective likelihood functions are derived.

2. Any unknown parameters in the alternate hypothesis are computed from the testing data using their MLEs, for example, the mean and/or variance.

3. The log likelihood ratio of the alternate to null hypotheses is then computed, and its maximum value is calculated, which maximizes the detection rate.

Univariate GLR charts can be designed based on the type of the fault that needs to be detected. Most processes experience shifts in the mean, and/or shifts in the variance, and three of these GLR charts will be explained next.

For the case when residuals are collected from processes under normal operating conditions, the likelihood function derived from a random normal distribution can be defined as follows [24]:

$$L\left(\infty, \mu_0, \sigma_0^2 | x_1, x_2, \dots, x_k\right) = (2\pi)^{-k/2}\left(\sigma_0^2\right)^{-k/2} \exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^{k}(x_i-\mu_0)^2\right) \tag{3}$$

where μ_0 and σ_0^2 mean and variance of the process variable measured under normal operating conditions respectively.

3.1.1 Univariate GLR chart for a shift in the mean

If a shift in the mean has occurred at time τ, from μ_0 to μ_1, the likelihood function of the alternate hypothesis is defined as follows [24]:

$$\begin{aligned} &L\left(\tau, \mu_1, \sigma_0^2 | x_1, x_2, \dots, x_k\right) \\ &\quad = (2\pi)^{-k/2}\left(\sigma_0^2\right)^{-k/2} \exp\left(-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^{\tau}(x_i-\mu_0)^2 + \sum_{i=\tau+1}^{k}(x_i-\mu_1)^2\right)\right) \end{aligned} \tag{4}$$

Since the magnitude of the new mean is unknown, its MLE can be computed using testing data as follow [24]:

$$\hat{\mu}_{1,\tau,k} = \frac{1}{(k-\tau)}\sum_{i=\tau+1}^{k} x_i. \tag{5}$$

The GLR statistic designed to specifically monitor a shift in the mean can now be computed by taking the log-likelihood ratio of (Eqs. (3) and (4)) [24]:

$$R_k = \max_{0 \le \tau < k} \frac{(k-\tau)}{2\sigma_0^2}\left(\hat{\mu}_{1,\tau,k}-\mu_0\right)^2. \tag{6}$$

The authors in [24] state that it is not necessary to store the entire length of previous historical data in order to compute the MLEs, but a window length of about 400 is sufficient to provide reliable results. Therefore, a window length of 400 was utilized throughout this work for all GLR charts.

3.1.2 Univariate GLR chart for a shift in the variance

If only a shift in the variance has occurred from at time τ, from σ_0^2 to σ_1^2, the alternate hypothesis for this case is defined as follows [25]:

$$L\left(\tau, \mu_0, \sigma_1^2 | x_{\tau+1}, x_2, \ldots, x_k\right) \\ = (2\pi)^{-k/2}\left(\sigma_1^2\right)^{-k/2} \exp\left(-\frac{1}{2\sigma_1^2}\left(\sum_{i=\tau+1}^{k}(x_i - \mu_0)^2\right)\right). \tag{7}$$

From a quality control standpoint we are only concerned with increases in variance, as larger variations imply that product is being manufactured with quality further away from the targeted amount, and since the magnitude of the new variance is unknown, its MLE can be computed using testing data as follows [25]:

$$\hat{\sigma}_{1,\tau,k}^2 = \max\left\{\sigma_0^2, \frac{1}{k-\tau}\sum_{i=\tau+1}^{k}(x_i - \mu_0)^2\right\}. \tag{8}$$

The GLR statistic designed to specifically monitor a shift in the variance can now be computed by taking the log-likelihood ratio of (Eqs. (3) and (7)) [25]:

$$R_k = \max_{0 \le \tau < k} \frac{k-\tau}{2}\left[\frac{\hat{\sigma}_{1,\tau,k}^2}{\sigma_0^2} - 1 - \ln\left(\frac{\hat{\sigma}_{1,\tau,k}^2}{\sigma_0^2}\right)\right] \tag{9}$$

3.1.3 Univariate GLR chart for a shift in the mean and/or variance

Since it is possible for most processes to experience both shifts in the mean and variance, a GLR statistic that is capable of detecting either type of shift can be designed. The likelihood function of the alternate hypothesis for this case is defined as follows [26]:

$$L\left(\tau, \mu_1, \sigma_1^2 | x_1, x_2, \ldots, x_k\right) \\ = (2\pi)^{-k/2}\left(\sigma_0^2\right)^{-\tau/2}\left(\sigma_1^2\right)^{-(k-\tau)/2} \exp\left(-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^{\tau}x_i - \mu_0\right)^2 - \frac{1}{2\sigma_1^2}\left(\sum_{i=\tau+1}^{k}x_i - \mu_1\right)^2\right). \tag{10}$$

The MLE of the mean can be computed from the testing data using (Eq. (5)). However, the variance now has to be computed utilizing the MLE for the mean as well [26]:

$$S_{\tau,k}^2 = \frac{1}{k-\tau}\sum_{i=\tau+1}^{k}\left(x_i - \hat{\mu}_{1,\tau,k}\right)^2. \tag{11}$$

As previously stated, from a quality control standpoint only an increase in the variance is of concern, and the MLE for the variance can be computed as follows [26]:

$$\hat{\sigma}^2{}_{1,\,\tau,\,k} = \max\{\sigma_0^2, \mathrm{S}^2_{\tau,\,k}\}. \tag{12}$$

If there are no shifts in the mean for testing data, the variance is computed as follows [26]:

$$S^2_{0,\,\tau,\,k} = \frac{1}{k-\tau}\sum_{i=\tau+1}^{k}(x_i-\mu_0)^2. \tag{13}$$

In this case, the GLR statistic designed to simultaneously monitor both shifts in the mean and variance, and can be computed by taking the log-likelihood ratio of (Eqs. (3) and (10)) resulting in the following equation [26]:

$$R_k = \max_{0\le\tau<k}\frac{k-\tau}{2}\left[\frac{S^2_{0,\,\tau,\,k}}{\sigma_0^2} - \frac{S^2_{\tau,\,k}}{\hat{\sigma}}{}^2_{1,\,\tau,\,k} - \ln\left(\frac{\hat{\sigma}^2_{1,\,\tau,\,k}}{\sigma_0^2}\right)\right] \tag{14}$$

It is important to note that for this particular GLR method, two parameters, that is, the mean and the variance have to be estimated using their MLE, since the type of shift is unknown.

3.1.4 Multivariate GLR chart for a shift in the mean

Since using a univariate GLR chart may not always be practical, Wang and Reynolds [27] introduce the multivariate GLR chart, designed to specifically monitor shifts in the process mean for multivariate applications. In this case, the GLR statistic is defined as follows:

$$R_k = \max_{\max(0,k-m)\le t<k}\left(\frac{k-t}{2}\left(\hat{\mu}_{1,t,k}-\mu_0\right)\cdot\sum\nolimits_0^{-1}\cdot\left(\hat{\mu}_{1,t,k}-\mu_0\right)\right) \tag{15}$$

Where μ_0 is the multivariate mean vector of the process under normal operating conditions, $\hat{\mu}_{1,t,k}$ is the MLE of a sustained process mean shift μ_1 at time index k over sample window of maximum length m, and $\sum_0$ is the process covariance matrix under normal conditions [27].

3.2 Fault detection using PCA-based GLR

The PCA method introduced in Section 2 is commonly utilized by many industries. Therefore, it is necessary to integrate the simplicity of the PCA method with the advantages brought forward by the GLR charts, so that it can be easily applied to

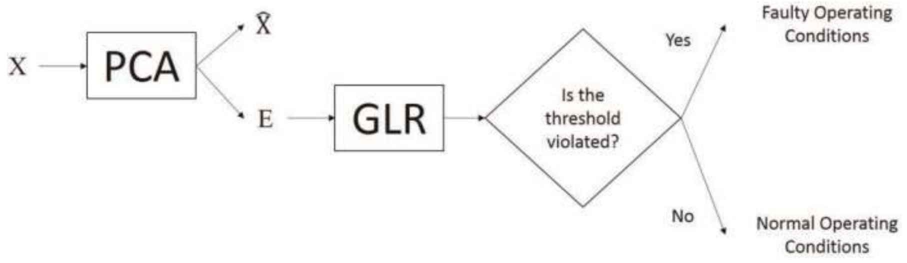
X
PCA
X̂
E
GLR
Is the threshold violated?
Yes
No
Faulty Operating Conditions
Normal Operating Conditions

Figure 1.
PCA-based GLR fault detection algorithm.

monitor processes online. **Figure 1** illustrates the fault detection algorithm utilized in this work.

PCA is utilized in order to model available data. The different GLR charts can then be applied on the residuals produced by the PCA model in order to determine if the process is operating under normal or faulty conditions. The fault detection threshold limits are obtained from an empirical distribution of the GLR statistic computed under normal operating conditions. The residual space is typically better able at detecting faults of smaller magnitude [10].

4. Moving window interval data aggregation

Data utilized in the construction of a PCA model may be of two types depending on the application being monitored: single-valued, and interval-valued. Single-valued data can be directly obtained from sensors measuring particular variables in a process, while interval-valued data is aggregated or artificially generated from batch single-valued measurements, thereby resulting in a range of possible measurement values for a given process variable at one time instant [18].

An interval is defined using a lower and upper bound, such as $[a, b]$, where $a \leq b$. In this work, interval data is generated by aggregating the single-valued samples in a dataset, such that the mean of each block of aggregated samples is defined as the interval center (c), and the standard deviation of each block of aggregated samples is defined as the interval radii (r). Consequently, the intervals can now be defined as $[c - r, c + r]$. Unlike the lower and upper bounds, the centers and radii are of particular importance because they can be used to represent unique characteristics of the classical samples from which they are generated [19].

Initially, the use of interval data is motivated by the need to quickly and efficiently monitor large datasets [28], in addition to its ability to deal with missing values without the need to remove entire samples. Generating intervals by aggregation is a form of batch processing, which may not always be ideal. The ability to monitor faults in real-time is typically much more desirable from a quality and safety standpoint. It also becomes impractical to use batch aggregation when discussing processes with a low sample size or low sampling frequency.

As a result, interval data aggregation must be adapted for real-time monitoring purposes. One way to do that would be to use a moving window aggregation technique, such that any observed sample is aggregated with previously gathered samples, if any, in the defined window size. This allows for the generation and processing of interval data in real-time, without the need to wait for multiple samples to be observed before processing.

As expected, however, this method suffers from some drawbacks relative to its batch aggregation counterpart. The moving window approach may cause smearing along the detection statistic, leading to higher false alarms and lower detection rates. This is especially true for large window sizes, as is the case for most methods which apply that approach. The problem can be mitigated by limiting the window size to reasonable limits, whilst also adjusting the threshold in order to meet the desired false alarm rates of the process.

4.1 Integration with PCA-based GLR

Interval principal component analysis (IPCA) methods are an extension to the classical PCA method, and they have been explored in literature for fault detection and isolation examples [29, 30]. In this work, three IPCA methods will be briefly introduced, before discussing our proposed method of integrating the moving window interval approach to the PCA-based GLR technique.

Centers IPCA (CIPCA) was introduced by Cazes et al. [31], where the idea was to only apply PCA to the matrix of interval centers. This method focuses on the variation between the intervals of a dataset, rather than the variations within them [18, 32]. Midpoint-Radii IPCA (MRIPCA) was developed by Lauro et al. [33–36], where PCA models are separately generated for the centers and radii matrices of the interval training dataset. Finally, the Symbolic Covariance IPCA (SCIPCA) method was introduced by Le-Rademacher et al. [18, 32] as a way to better represent the range and variability found in interval data.

In this paper, the integration of the moving window aggregation to PCA-based GLR will be as follows. After generating an interval sample for each single-valued sample, the single-valued matrices of interval centers and radii are extracted. The matrices are then concatenated along the variables dimension, so as to maintain the number of samples, but double the number of variables. This is similar to the MRIPCA method, except it avoids the need to apply PCA twice, eliminating any additional processing complexity.

5. Illustrative examples

This section evaluates the performance of the three PCA-based GLR charts described in Section 3, and the moving window aggregation method discussed in Section 4. The PCA-based GLR charts are evaluated under different fault scenarios, and this is done through two illustrative examples: a simulated synthetic data set, and the benchmark Tennessee Eastman Process (TEP). Three fault detection metrics are used to evaluate the performance of each univariate chart: missed DR (which is equal to 100-DR), FAR, and average out-of-control run length (ARL1). Finally, the moving window interval aggregation method, in tandem with the PCA-based multivariate GLR chart, are analyzed using the benchmark TEP process, and the results are tabulated and compared to the single-valued multivariate GLR chart.

5.1 Simulated synthetic data example

The purpose of this example is to utilize a simple linear model to compare and evaluate the performance of the difference PCA-based univariate GLR charts. The linear data set can be generated using the following model [37]:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} -0.3441 & 0.4815 & 0.6637 \\ -0.2313 & -0.5936 & 0.3545 \\ -0.5060 & 0.2495 & 0.0739 \\ -0.5552 & -0.2405 & -0.1123 \\ -0.3371 & 0.3822 & -0.6115 \\ -0.3877 & -0.3868 & -0.2045 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} + noise \qquad (16)$$

where, t_1, t_2, and t_3, are uniformly distributed random variables with ranges, $[0, 2]$, $[0, 1.6]$, and $[0, 1.2]$, respectively, while the noise follows a normal distribution with zero-mean and standard deviation of 0.2 [37].

The linear model is used to generate 6000 observations, split into training and testing data sets of 3000 observations each. The training data are used to train the PCA model, while the testing data are used to evaluate the performance of all techniques using three cases of faults: a shift in the mean, a shift in the variance, and a simultaneous shift in both.

Five charts are evaluated and compared: the PCA-based T^2 and Q charts, and the three different PCA-based univariate GLR charts. The faulty region is highlighted in light blue for all figures, and the fault detection threshold limits for all charts are represented by the red dotted line. For each case a Monte-Carlo simulation of 1000 realizations is carried out in order to obtain meaningful results, so that conclusions can be drawn.

5.1.1 Case 1: a shift in the mean

For this case, a shift in the mean of 1σ was introduced between observations 1501 and 3000 in x_1 in the testing data set. This fault size was chosen as most conventional techniques are unable to detect a fault of this magnitude. Faults of higher magnitude would likely provide misleading results and exaggerate the robustness of the method in question, leading to a biased comparison.

As can be seen through **Figure 2**, the T^2 and Q charts are unable to detect the entirety of the fault. In contrast, two GLR charts (**Figure 3a** and **c**), are able to detect most of the fault, while the GLR chart designed to monitor a shift in the variance (**Figure 3b**) could not detect that a shift in the mean was present.

Examining the summary of the fault detection results (**Table 1**), it can be observed that the GLR chart designed to monitor shifts in the mean (**Figure 3a**) provided the lowest missed DR and ARL_1 values, compared to all other charts.

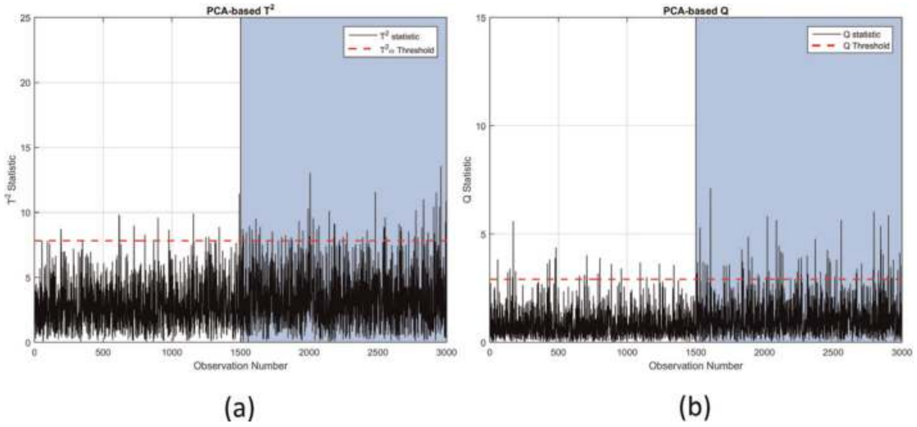

Figure 2.
PCA-based T^2 and Q charts (case 1).

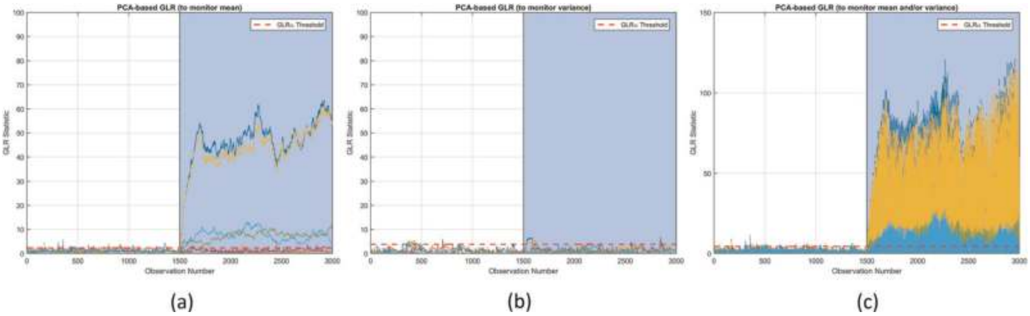

Figure 3.
PCA-based GLR charts (case 1).

The relatively high missed DR of the GLR chart designed to simultaneously monitor shifts in both the mean and variance (**Figure 3c**) can be attributed to the fact that two parameters need to be estimated from available data while maximizing the GLR statistic, thereby making it difficult to predict a shift in a single parameter as efficiently.

5.1.2 Case 2: a shift in the variance

For this case, an increase in the variance (double that of the training data) was introduced between observations 1501:3000 in x_1 in the testing data set. This shift in the variance is too small for detection by most conventional techniques.

As can be seen through **Figure 4**, the T^2 and Q charts are unable to detect the entirety of the fault. In contrast, two GLR charts (**Figure 5b** and **c**) were able to detect most of the fault, while the GLR chart designed to monitor a shift in the mean (**Figure 5a**) could not detect it as well. Examining the summary of the results (**Table 2**), it can be observed that the GLR chart designed to monitor a shift in the variance (**Figure 5b**) provided the lowest missed DR and ARL_1 values, compared to other charts.

5.1.3 Case 3: a shift in the mean and/or variance

For this case, a simultaneous shift in the mean of 1σ and an increase in the variance (double that of the training data) was introduced between observations 1501:3000 in x_1 in the testing data set.

	PCA-based T^2	**PCA-based Q**	**PCA-based GLR (to monitor mean)**	**PCA-based GLR (to monitor variance)**	**PCA-based GLR (to monitor mean and/or variance)**
Missed DR (%)	95.3	94.5	00.4	85.1	31.5
FAR (%)	05.2	05.5	05.3	05.8	04.6
ARL_1	20.1	16.6	04.8	81.8	05.0

Table 1.
Summary of fault detection results (case 1).

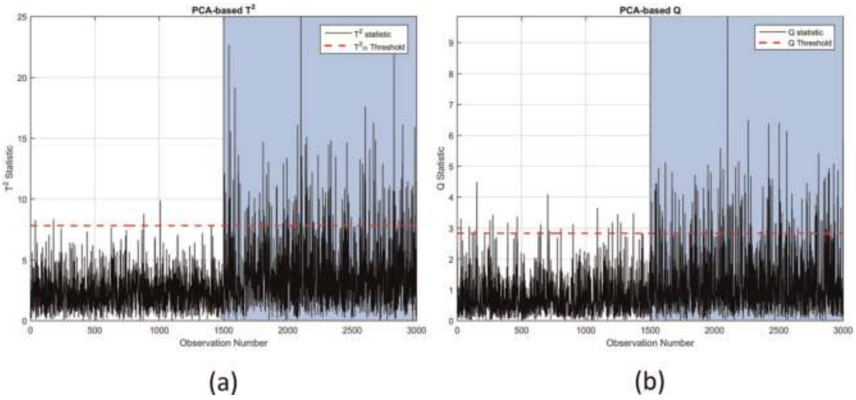

Figure 4.
PCA-based T^2 and Q charts (case 2).

As can be seen through **Figure 6**, the T^2 and Q charts are unable to detect the entirety of the fault once more. Although it might seem that all three GLR charts (**Figure 7**) are able to detect most of the fault, upon closer inspection of the results summarized in **Table 3**, it can be observed that the GLR charts designed to independently detect a shift in the mean (**Figure 7a**), and variance (**Figure 7b**), are able to provide significantly lower missed DR and ARL$_1$ values compared to the chart designed to monitors shifts in both (**Figure 7c**).

The main conclusion from this example is that if a process is expected to experience shifts in both the mean and/or variance, it is more beneficial to run the PCA-based GLR charts designed to independently monitor shifts in the mean and variance as two parallel charts, rather than utilizing the GLR chart designed to simultaneously monitor both. Based on this conclusion, only the former two GLR charts will be utilized for the next example.

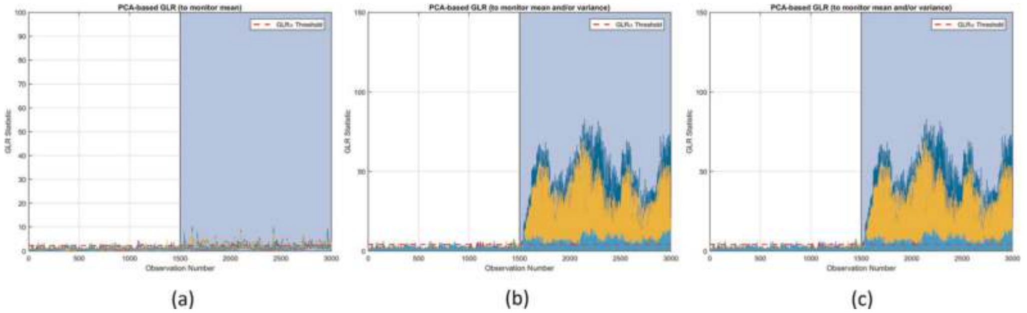

Figure 5.
PCA-based GLR charts (case 2).

	PCA-based T^2	PCA-based Q	PCA-based GLR (to monitor mean)	PCA-based GLR (to monitor variance)	PCA-based GLR (to monitor mean and/or variance)
Missed DR (%)	90.2	88.6	47.5	00.7	33.0
FAR (%)	05.3	05.4	05.0	04.8	04.8
ARL$_1$	10.1	8.3	07.9	04.5	05.6

Table 2.
Summary of fault detection results (case 2).

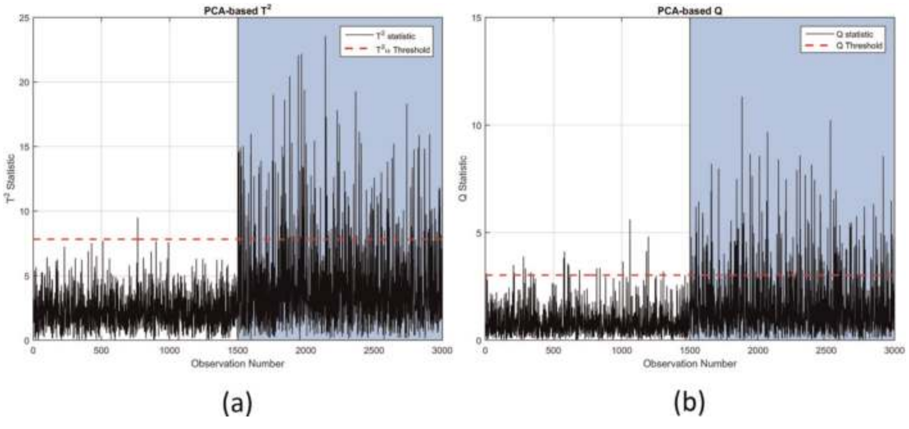

Figure 6.
PCA-based T^2 and Q charts (case 3).

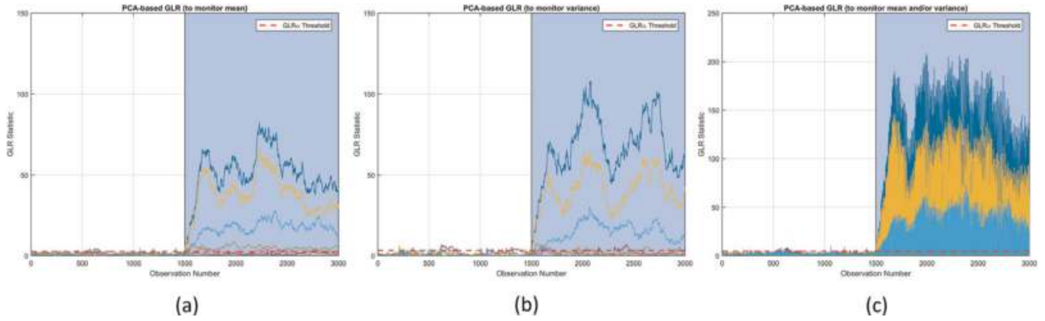

Figure 7.
PCA-based GLR charts (case 3).

	PCA-based T^2	PCA-based Q	PCA-based GLR (to monitor mean)	PCA-based GLR (to monitor variance)	PCA-based GLR (to monitor mean and/or variance)
Missed DR (%)	86.7	84.5	00.4	00.4	24.2
FAR (%)	05.2	05.2	04.9	05.3	05.5
ARL_1	07.5	06.0	03.2	03.9	04.9

Table 3.
Summary of fault detection results (case 3).

5.2 Tennessee Eastman Process (TEP)

In order to assess the feasibility of using two separate GLR charts to monitor shifts in the process mean and variance, their performance has to be evaluated using real data. Many authors utilize the Tennessee Eastman Process (TEP) in order to evaluate the performance of their techniques [17, 38, 39]. The Tennessee Eastman Process is a realistic simulation of an actual chemical process that consists of a reactor, condenser, stripper, compressor, and separator, and is widely accepted as a benchmark for fault detection [17].

The Tennessee Eastman Process contains a bank of pre-defined faults that can be utilized by authors in order to assess the performance of their developed fault detection algorithms. More information on the Tennessee Eastman Process, the process description, and the available bank of faults is available in literature [10, 17, 21, 38, 39].

Two fault scenarios will be examined in this work: IDV 3 and IDV 11 [39]. IDV 3 is a shift in the mean of the temperature of Feed D, while IDV 11 is random variation in the reactor cooling water inlet temperature [39]. These two fault scenarios were selected because the conventional techniques are unable to provide the best possible detection. For both scenarios, the fault is introduced after 800 observations of normal operation. The performance of four charts are evaluated: PCA-based T^2 and Q charts, and the PCA-based univariate GLR charts designed to independently monitor shifts in the mean and variance. The faulty region is highlighted in light blue in all figures.

5.2.1 IDV 3: a step fault in the mean of the temperature of feed D

For the case where there is a shift in the mean of the temperature of Feed D, the PCA-based T^2 and Q charts, and the PCA-based univariate GLR charts are illustrated in **Figures 8** and **9** respectively, and the fault detection results are summarized in **Table 4**.

From **Figure 8** it can be observed that the T^2 and Q charts are unable to detect the entirety of the fault, while the GLR chart designed to monitor shifts in the mean (**Figure 9a**) is able to detect the most of the fault, and provides the lowest missed DR (**Table 4**). Although, the T^2 chart returns a low ARL_1 value, it does not detect the fault efficiently, and the low ARL_1 value can be attributed to random noise.

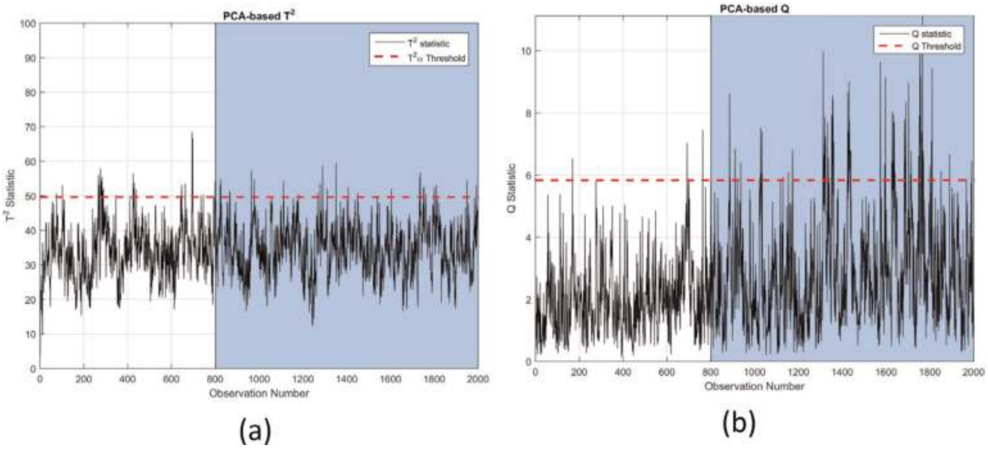

Figure 8.
PCA-based T^2 and Q charts (IDV 3).

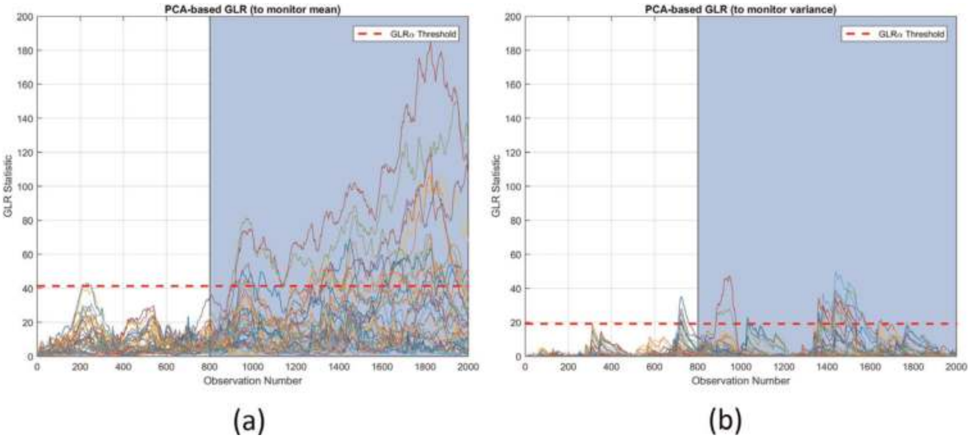

Figure 9.
PCA-based GLR charts (IDV 3).

	PCA-based T^2	PCA-based Q	PCA-based GLR (to monitor mean)	PCA-based GLR (to monitor variance)
Missed DR (%)	97.6	92.8	07.9	70.9
FAR (%)	04.8	04.5	05.0	05.4
ARL_1	02.0	86.0	84.0	84.00

Table 4.
Summary of fault detection results (IDV 3).

5.2.2 IDV 11: random variation in the reactor cooling water inlet temperature

For the case where there is random variation in the reactor cooling water inlet temperature, the T^2 and Q charts, and the GLR charts are illustrated in **Figures 10** and **11** respectively, and the fault detection results are summarized in **Table 5**.

Although it might seem like the T^2 and Q charts (**Figure 10**) are able to detect most of the fault, they still have higher missed DR than both GLR charts (**Figure 11**). The GLR chart designed to monitor shifts in the variance provides the lowest missed DR from the charts that were compared.

From this example we can conclude that the PCA-based GLR charts are able to provide improved fault detection results over the conventional PCA-based

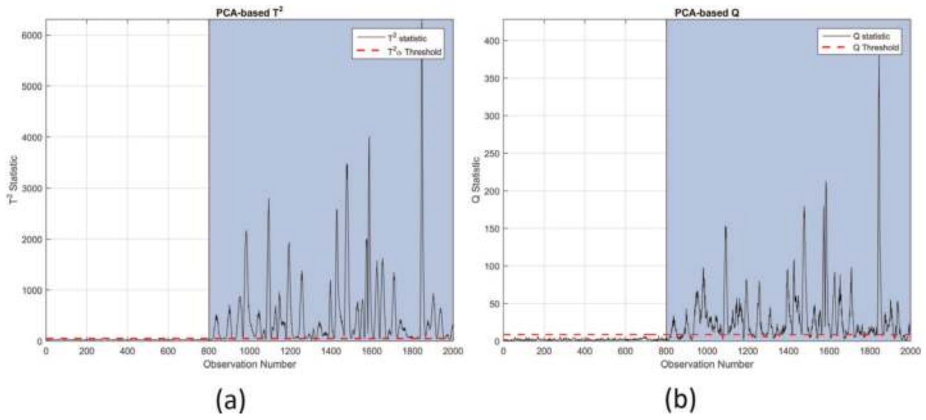

Figure 10.
PCA-based T^2 and Q charts (IDV 11).

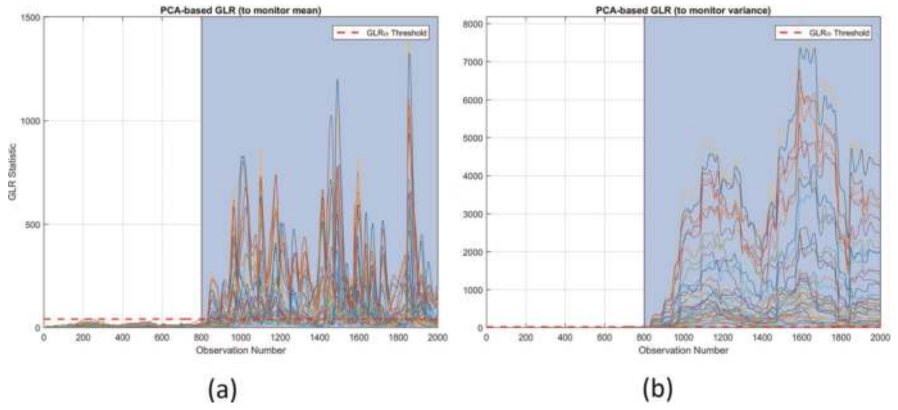

Figure 11.
PCA-based GLR charts (IDV 11).

	PCA-based T^2	PCA-based Q	PCA-based GLR (to monitor mean)	PCA-based GLR (to monitor variance)
Missed DR (%)	09.9	22.3	02.3	01.9
FAR (%)	05.1	05.0	05.0	05.4
ARL_1	20.0	24.0	28.0	24.0

Table 5.
Summary of fault detection results (IDV 11).

T^2 and Q charts. The improved results can be attributed to the use of MLEs to estimate the values of the unknown parameters used to maximize the GLR statistic, allowing for the best possible DR to be achieved for a fixed FAR. This example also demonstrates that the GLR charts can be easily designed and utilized to monitor chemical processes, such as the TEP.

5.2.3 IDV 3 and IDV 11: single-valued vs. interval-valued multivariate GLR chart

For the final case study, the moving window interval aggregation method is tested for the same fault scenarios tested previously for the TEP: IDV 3 and IDV 11. A smaller sample window size of 10 samples is used for the multivariate GLR chart

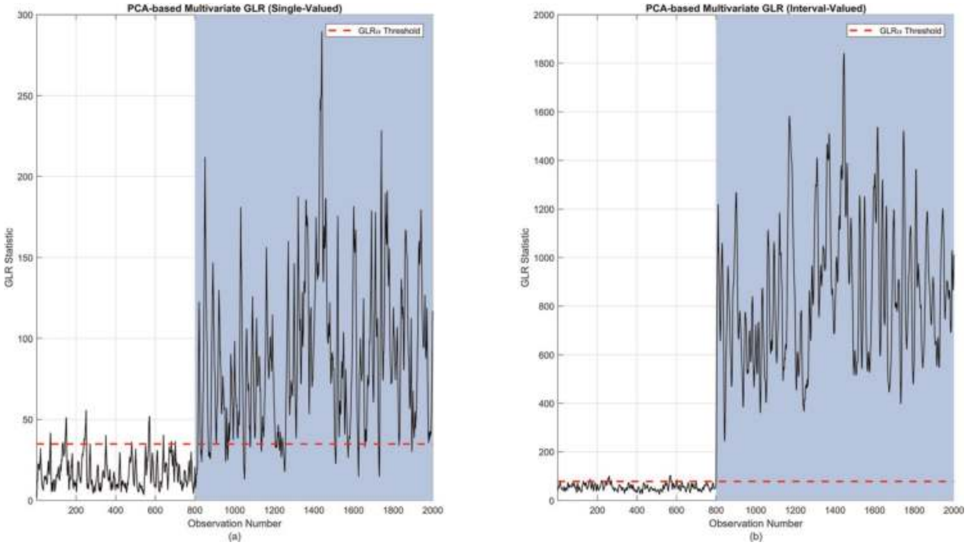

Figure 12.
PCA-based multivariate GLR charts (IDV 3).

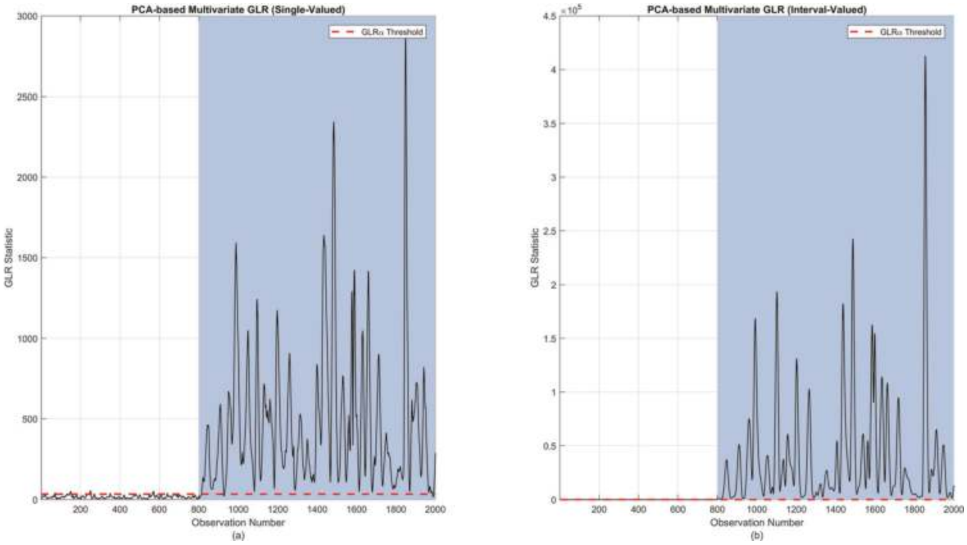

Figure 13.
PCA-based multivariate GLR charts (IDV 11).

	IDV 3 Single-valued multivariate GLR	IDV 3 Interval-valued multivariate GLR	IDV 11 Single-valued multivariate GLR	IDV 11 Interval-valued multivariate GLR
Missed DR (%)	15.1	00.0	02.0	00.0
FAR (%)	05.0	05.0	05.0	05.0

Table 6.
Summary of fault detection results (single vs. interval data) for α = 5%.

in order to highlight the difference between using single and interval-valued data more clearly.

The interval aggregation window size was set at 10 samples. The IDV 3 and IDV 11 scenarios for both data types are shown in **Figures 12** and **13**, and the metrics for each method are tabulated in **Table 6**.

There are two major observations to be made from the results. First, the use of the multivariate GLR chart allowed for a more stable FAR for all cases due to the presence of a single statistic to monitor for all variables, as opposed to the one for each variable when using the univariate GLR charts. Second, the missed DR when using interval data was significantly lower than that for single-valued data, reaching perfect performance levels of zero missed DR for both scenarios.

The latter observation is attributed to interval data, especially the method of generation, where the centers and radii are used as independent variables in the same dataset. This method of aggregation helps the PCA model account for shifts in the mean and variance respectively, similar to the univariate GLR chart outline in Section 3.1.3. However, it does so without the need to tune any extra parameters, due to the fact that a fault in the centers is likely to be caused by a shift in the mean, while a fault in the radii is likely to be caused by a shift in the variance.

6. Conclusions

In this chapter, the performance of GLR charts were compared to conventional fault detection statistics, specifically the Q and T^2 statistics, and the integration of interval-valued data into real-time process monitoring was explored. The performance of different PCA-based univariate GLR charts were examined using single-valued data through two illustrative examples: simulated synthetic data, and the Tennessee Eastman Process. The performance of the moving window interval aggregation method was evaluated alongside that of single-valued data for the multivariate GLR chart as well.

The results demonstrate that in order to monitor processes that may experience both shifts in the mean and/or variance, the best performance is achieved by implementing the two respective univariate GLR charts separately in parallel, rather than the single chart designed to simultaneously detect shifts in both, as the simultaneous estimation of two parameters is unable to provide the best possible fault detection performance. Moreover, the moving window interval aggregation method, when combined with the multivariate GLR chart, was able to provide a perfectly stable statistic, with an unwavering false alarm rate, in addition to the best possible performance in detecting shifts in the mean and variance for two scenarios of the Tennessee Eastman Process.

Acknowledgements

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. The statements herein are solely the responsibility of the authors.

Author details

Mohammed Ziyan Sheriff[1,2], Nour Basha[2], Muhammad Nazmul Karim[1], Hazem Nounou[3] and Mohamed Nounou[2]*

1 Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, USA

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

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

*Address all correspondence to: mohamed.nounou@qatar.tamu.edu

References

[1] Montgomery DC. Introduction to Statistical Quality Control. 7th ed. Hoboken, NJ: John Wiley and Sons; 2013

[2] Chakrabarty A, Mannan S, Cagin T. Multiscale Modeling for Process Safety Applications. 1st ed. Oxford, United Kingdom: Butterworth-Heinemann; 2015

[3] Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. Computers and Chemical Engineering. 2003;**27**:293-311

[4] Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. Computers and Chemical Engineering. 2003;**27**:313-326

[5] Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: Part III: Process history based methods. Computers and Chemical Engineering. 2003;**27**:327-346

[6] George JP, Chen Z, Shaw P. Fault detection of drinking water treatment process using PCA and Hotelling's T2 chart. International Journal of Computer and Information Engineering. 2009;**3**: 970-975

[7] Sanguansat P, editor. Principal Component Analysis: Multidisciplinary Applications. Rijeka: InTech; 2009. DOI: 10.5772/2694

[8] Sanguansat P, editor. Principal Component Analysis: Engineering Applications. Rijeka: InTech; 2012. DOI: 10.5772/2693

[9] Joliffe IT. Principal Component Analysis. 2nd ed. New York, NY: Springer-Verlag; 2002

[10] Sheriff MZ, Mansouri M, Karim MN, Nounou H, Nounou M. Fault detection using multiscale PCA-based moving window GLRT. Journal of Process Control. 2017;**54**:47-64. DOI: 10.1016/j.jprocont.2017.03.004

[11] Sheriff MZ, Botre C, Mansouri M, Nounou H, Nounou M, Karim MN. Process monitoring using data-based fault detection techniques: Comparative studies. In: Fault Diagnosis Detect. InTech; 2017. DOI: 10.5772/67347

[12] Mansouri M, Sheriff MZ, Baklouti R, Nounou M, Nounou H, Ben Hamida A, et al. Statistical fault detection of chemical process: Comparative studies. Journal of Chemical Engineering and Process Technology. 2016;**07**:1-10. DOI: 10.4172/2157-7048.1000282

[13] Botre C, Mansouri M, Nounou M, Nounou H, Karim MN. Kernel PLS-based GLRT method for fault detection of chemical processes. Journal of Loss Prevention in the Process Industries. 2016;**43**:212-224. DOI: 10.1016/j.jlp.2016.05.023

[14] Tharrault Y, Mourot G, Ragot J. Fault detection and isolation with robust principal component analysis. In: 2008 16th Mediterr. Conf. Control Autom. Vol. 18. 2008. pp. 429-442. DOI: 10.1109/MED.2008.4602224

[15] Montgomery DC, Runger GC. Applied Statistics and Probability for Engineers. 5th ed. Hoboken, NJ: John Wiley and Sons, Inc.; 2011

[16] Harrou F, Nounou MN, Nounou HN. Detecting abnormal ozone levels using PCA-based GLR hypothesis testing. In: Proc. 2013 IEEE Symp. Comput. Intell. Data Mining, CIDM 2013–2013 IEEE Symp. Ser. Comput. Intell. SSCI 2013; 2013. pp. 95-102. DOI: 10.1109/CIDM.2013.6597223

[17] Downs JJ, Vogel EF. A plant-wide industrial process control problem. Computers and Chemical Engineering. 1993;**17**:245-255. DOI: 10.1016/0098-1354(93)80018-I

[18] Le-Rademacher JG. Principal Component Analysis for Interval-Valued and Histogram-Valued Data and Likelihood Functions and some Maximum Likelihood Estimators for Symbolic Data. Athens, GA: University of Georgia; 2008

[19] Basha N. Interval Principal Component Analysis and its Application to Fault Detection and Data Classification. College Station, TX: Texas A&M University; 2018

[20] Strang G. Introduction to Linear Algebra. 5th ed. Wellesley, MA: Wellesley-Cambridge Press; 2016

[21] Russell EL, Chiang LH, Braatz RD. Fault Detection and Diagnosis in Industrial Systems. New York, NY: Springer-Verlag; 2001

[22] Hotelling H. Analysis of a complex of statistical variables into principal components. Journal of Education & Psychology. 1933;**24**:417-441. DOI: 10.1037/h0071325

[23] Harrou F, Nounou MN, Nounou HN, Madakyaru M. Statistical fault detection using PCA-based GLR hypothesis testing. Journal of Loss Prevention in the Process Industries. 2013;**26**:129-139. DOI: 10.1016/j.jlp.2012.10.003

[24] Reynolds MR, Lou JY. An evaluation of a GLR control chart for monitoring the process mean. Journal of Quality Technology. 2010;**42**:287-310

[25] Reynolds Jr MR, Lou J. A GLR control chart for monitoring the process variance. In: Lenz HJ, Schmid W, Wilrich, editors. Frontiers in Statistical Quality Control. New York, NY: Springer; 2012;**10**:3-17. DOI: 10.1007/978-3-7908-2846-7

[26] Reynolds MR, Lou J, Lee J, Wang S. The design of GLR control charts for monitoring the process mean and variance. Journal of Quality Technology. 2013;**45**:34-60

[27] Wang S, Reynolds MR. A GLR control chart for monitoring the mean vector of a multivariate normal process. Journal of Quality Technology. 2013;**45**: 18-33

[28] Billard L, Le-Rademacher J. Principal component analysis for interval data. Wiley Interdisciplinary Reviews: Computational Statistics. 2012; **4**:535-540. DOI: 10.1002/wics.1231

[29] Benaicha A, Guerfel M, Bougila K, Benothman N. New PCA-based methodology for sensor fault detection and localization. In: 8th Int. Conf. Model. Simul.; Hammamet: Tunisia; 2010

[30] Izem TA, Bougheloum W, Harkat MF, Djeghaba M. Fault detection and isolation using interval principal component analysis methods. IFAC-PapersOnLine. 2015;**48**:1402-1407. DOI: 10.1016/j.ifacol.2015.09.721

[31] Cazes P, Chouakria A, Diday E, Schektman Y. Extension de l'analyse en composantes principales à des données de type intervalle. Revue de Statistique Appliquée. 1997;**45**:5-24

[32] Le-Rademacher J, Billard L. Symbolic covariance principal component analysis and visualization for interval-valued data. Journal of Computational and Graphical Statistics. 2012;**21**:413-432. DOI: 10.1080/10618600.2012.679895

[33] Lauro CN, Palumbo F. Principal component analysis for non-precise data. In: New Developments in

Classification and Data Analysis. Berlin/Heidelberg: Springer-Verlag; n.d. pp. 173-184. DOI: 10.1007/3-540-27373-5_21

[34] Lauro CN, Palumbo F. Principal component analysis of interval data: A symbolic data analysis approach. Computational Statistics. 2000;**15**:73-87. DOI: 10.1007/s001800050038

[35] Lauro NC, Verde R, Irpino A. Principal component analysis of symbolic data described by intervals. In: Symbolic Data Analysis and the SODAS Software. Chichester, UK: John Wiley and Sons, Ltd; n.d. pp. 279-311. DOI: 10.1002/9780470723562.ch15

[36] Palumbo F, Lauro CN. A PCA for interval-valued data based on midpoints and radii. In: New Developments in Psychometrics. Japan, Tokyo: Springer; 2003. pp. 641-648. DOI: 10.1007/978-4-431-66996-8_74

[37] Alcala CF, Joe Qin S. Analysis and generalization of fault diagnosis methods for process monitoring. Journal of Process Control. 2011;**21**:322-330. DOI: 10.1016/j.jprocont.2010.10.005

[38] Lyman PR, Georgakis C. Plant-wide control of the Tennessee Eastman problem. Computers and Chemical Engineering. 1995;**19**:321-331. DOI: 10.1016/0098-1354(94)00057-U

[39] Yin S, Ding SX, Haghani A, Hao H, Zhang P. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. Journal of Process Control. 2012;**22**:1567-1581. DOI: 10.1016/j.jprocont.2012.06.009