# IMPROVED SHEWHART CHART USING MULTISCALE REPRESENTATION

A Thesis

by

MOHAMMED ZIYAN SHERIFF

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

| | |
|---|---|
| Chair of Committee, | Mohamed N. Nounou |
| Committee Members, | Hazem N. Nounou |
| | Konstantinos Kakosimos |
| Head of Department, | Nazmul Karim |

August 2015

Major Subject: Chemical Engineering

# ABSTRACT

Most univariate process monitoring techniques operate under three main assumptions, that the process residuals being evaluated are Gaussian, independent and contain a moderate level of noise. The performance of the conventional Shewhart chart, for example, is adversely affected when these assumptions are violated. Multiscale wavelet-based representation is a powerful data analysis tool that can help better satisfy these assumptions, i.e., decorrelate autocorrelated data, separate noise from features, and transform the data to better follow a Gaussian distribution at multiple scales.

This research focused on developing an algorithm to extend the conventional Shewhart chart using multiscale representation to enhance its performance. Through simulated synthetic data, the developed multiscale Shewhart chart showed improved performance (with lower missed detection and false alarm rates) than the conventional Shewhart chart.

The developed multiscale Shewhart chart was also applied to two real world applications, simulated distillation column data, and genomic copy number data, to illustrate the advantage of using the multiscale Shewhart chart for process monitoring over the conventional one.

# DEDICATION

I dedicate this thesis to my parents, Omar Khayyam Sheriff and Firzana Zubair, for their unconditional love and continued support. Thank you.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my committee chair, Dr. Mohamed Nounou, for his excellent vision and guidance through the course of my Master's at Texas A&M University. I would also like to thank my committee members, Dr. Hazem Nounou and Dr. Konstantinos Kakosimos for their valuable advice and feedback through the course of my research.

A special thanks also goes out to my friends and colleagues and the department faculty and staff of both campuses of Texas A&M University, in College Station and in Qatar, for making my undergraduate and graduate experiences truly memorable, and for providing a great learning environment, both inside and outside the classroom.

Finally, a note of appreciation to my parents and brother for their encouragement, support, and for being my motivation and inspiration.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

# 1. INTRODUCTION

Statistical Process Monitoring (SPM) is a vital component in many process industries to maintain product quality and to ensure a safe operation. Fault detection and diagnosis is a major function of SPM. Literature provides numerous classifications for fault detection and diagnosis methods. One way of classifying fault detection and diagnosis methods is as follows:

- Quantitative model-based methods [1].
- Qualitative model-based methods [2].
- Process history based methods [3].

Quantitative model-based methods involve residual generation through diagnostic observers, parity relations, Kalman filters, while qualitative model-based methods cover approaches such as Fault Trees, Qualitative Process Theory (QPT), Signed Directed Graphs (SDG) [1][2]. Process history based methods can be either qualitative or quantitative, and include methods such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Partial Least Squares (PLS) to name a few [3].

Most common fault detect techniques fall under process history based methods, and due to their heavy dependence on data, are often referred to as data-based methods as well. In this category of methods, historical process data are collected under fault-free conditions, and are then used to design fault detectors that can be used to monitor the process given new process measurements.

Data-based fault detection techniques rely on measuring key process variables, and observing their trends, which can be performed for a single process variable (using univariate techniques) or for the trends across multiple variables (using multivariate techniques) [4][5][6].

The Shewhart chart is a well-known univariate monitoring technique that uses the available process data without the application of filters. Shewhart charts are generally capable of detecting fairly large faults. Other univariate techniques include the Cumulative Sum (CUSUM), Moving Average (MA), and Exponentially Weighted Moving Average (EWMA) charts. These are univariate monitoring charts that utilize the application of linear filters on the available process data before fault detection [7][8][9][10][11][12]. These charts may be able to detect faults smaller than the conventional Shewhart chart since the filters used in these techniques provide a memory to the technique based on historical process data, and also because of the ability of these charts to reduce the noise content using the applied filters.

Linear filters are not very effective because they rely on defining a frequency threshold above which all features are considered noise. In practice, however, real process data can include important features that occur at high frequencies as well as noise that span a wide range of frequencies, such as colored (or autocorrelated) noise. Thus, linear filters are not well suited to handle practical process data. Non-linear filters, such as wavelet-based multiscale filtering, have shown promising results when dealing with real process data [13][14][15].

The presence of measurement noise in the data is not the only issue with most univariate fault detection methods, such as the Shewhart chart. These methods also rely on the assumptions that the process data are independent and Gaussian [5]. However, for practical reasons most process data violate one or more of these assumptions, which makes it essential to have more effective fault detection and diagnosis methods, especially under the violations of these assumptions. Wavelet-based representation of data has shown improved performance when these assumptions are violated. For example, the authors in [16] utilize wavelet-based multiscale Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) charts in order to monitor autocorrelated processes, and show an improvement in the false alarm rate with their multiscale approach. However, in practice missed detections are generally more critical. Literature currently lacks a review that examines the performance of a multiscale Shewhart chart, in terms of the missed detection and false alarm rates, for the violation of all three assumptions. Therefore, the main objective of this work is to utilize multiscale representation of data to improve the fault detection ability of the Shewhart chart. The Shewhart chart was primarily chosen for analysis due to its computational simplicity and widespread use. However, before discussing multiscale representation and how it can improve fault detection, it is important to discuss some of the commonly used univariate monitoring techniques as well as their advantages and limitations, as will be presented next.

## 1.1. Literature Review

In this section, descriptions of various univariate fault detection methods will be presented, mainly to highlight their strengths and limitations.

### *1.1.1. Conventional univariate monitoring techniques*

#### 1.1.1.1. Shewhart chart

The Shewhart chart was developed by Walter Shewhart in the 1920s, while working for Bell Systems [17]. Shewhart believed that inspecting the final product for quality was inefficient, and process monitoring when carried out at different stages could be more economical and effective [18]. The Shewhart chart is currently one of the most popular statistical quality control charts in practice. The relative computational simplicity of the Shewhart chart is what drives its popularity, as opposed to other control charts.

Shewhart charts have three distinct features [5]:

- Center Line (C) – Typically represents the targeted process mean,
- Upper Control Limit (UCL),
- Lower Control Limit (LCL).

The Shewhart chart is designed based on the following main assumptions [5]:

- The presence of a moderate noise level in the evaluated residuals.
- The residuals are independent (uncorrelated).
- The fault-free residuals are normally distributed (Gaussian).

Many variations of the Shewhart chart exist, but the most popular Shewhart chart monitors the sample mean $(\bar{x})$. This chart is occasionally coupled with either the range $(R)$ chart or the standard deviation $(S)$ chart, which increases the robustness of the Shewhart chart against the variability in observations collected from different sensors measuring a particular process variable. This is because of the added advantage of the $R$ and $S$ charts, being able to spot features in the trend of the data, that may not necessarily come forth with the use of the $\bar{x}$ chart only.

If observations from multiple sensors monitoring a particular process variable are available, the following equations can be used to compute the sample mean [19]:

$$\bar{x}_i = \sum_{j=1}^{n} \frac{x_{ij}}{n},$$

$$\bar{\bar{x}} = \sum_{i=1}^{k} \frac{\bar{x}_i}{k},$$

where, $n$ and $k$ represent the subgroup size and the number of subgroups, respectively.

Sub-grouping is generally carried out only if observations from multiple sensors measuring a single process variable are available. The approach presented in this work assumes that only a single sensor provides measurements, and therefore the raw data (single readings) are used, and $R$ and $S$ charts are not required.

Shewhart suggested the following limits for the Shewhart chart [20]:

$$UCL, LCL = \bar{\bar{x}} \pm L_n .$$

where, $\bar{\bar{x}}$ represents the targeted process mean and $L_n$ is computed using the following equation [20]:

$$L_n = \frac{c\sigma}{\sqrt{n}},$$

where, $\sigma$ represents the standard deviation of the fault-free data and the constant $c$ is computed using a nomogram [20]. The $c$ scale corresponds to the range where a given percentage of the observations lie. Analysis of a nomogram for different processes is arduous and time-consuming and it is common practice to use the following equation to compute the control limits [6]:

$$UCL, LCL = \bar{\bar{x}} + 3\sigma,$$

where, $\sigma$ is the standard deviation of the fault-free residuals.

For a set of normally distributed fault-free data, a value of $3\sigma$ accounts for nearly 99.73 % of all the deviation, which makes $3\sigma$ a popular choice for the limits of the Shewhart chart [21].

A major drawback for the conventional Shewhart chart is its inability to detect relatively small faults. In fact, it is only able to detect faults that are larger than three times the standard deviation of the fault-free data set [22].

This is because the Shewhart chart only considers the current process measurement when deciding on the presence or absence of a fault, and thus it has a very short memory as indicated in Figure 1. Also, the insensitivity of the Shewhart chart to faults with small magnitudes is more evident with high noise levels in the data because features get masked by measurement noise. Other fault detection methods possess a

longer memory than the Shewhart Chart by utilizing information from previous process observations. Some of these techniques are described next.

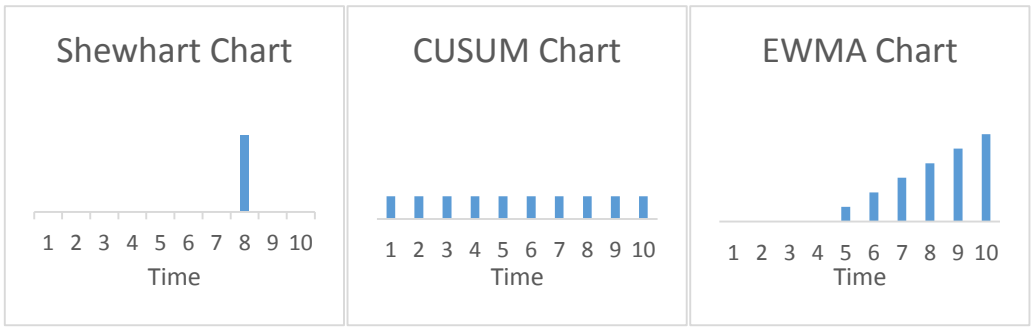


**Figure 1. Weightings used in various univariate fault detection statistics.**

### 1.1.1.2. CUSUM chart

Another univariate monitoring chart is the cumulative sum (CUSUM) chart, and like the Shewhart chart, numerous variations of the CUSUM chart exist. The most popular form of the CUSUM chart, however, is the two-sided CUSUM (Tabular) version. The positive and negative CUSUM statistics are calculated using the following equations [5]:

$$S_H(i) = \max\left[0, \bar{x}_i - (\mu_0 + K) + S_H(i-1)\right].$$

$$S_L(i) = \max\left[0, (\mu_0 - K) - \bar{x}_i + S_L(i-1)\right].$$

where, the parameter $K$ is defined as:

$$K = \frac{\Delta}{2},$$

and, $\Delta$ represents the size of the shift to be detected.

Control limits of $4\sigma$ or $5\sigma$ are suggested to provide a reasonable detection for a mean shift of around $1\sigma$ in the process data [6].

The CUSUM chart has a very long memory as it incorporates all previous observations in the decision making process (see Figure 1). Therefore, it is more capable of detecting smaller faults than the Shewhart chart, but can result in several false alarms, especially around sharp faults. As in the case of the Shewhart chart, the CUSUM chart also assumes that the evaluated residuals are independent and Gaussian, which is not always true in practice. The following section discusses the EWMA chart, which utilizes a different filter to deal with the extensive memory characteristic of the CUSUM chart.

#### 1.1.1.3. EWMA chart

The exponentially weighted moving average (EWMA) chart uses a statistic that gives exponential weighting to all previous measurements (see Figure 1). The EWMA statistic can be computed recursively as follows [6]:

$$z_i = \lambda x_i + (1-\lambda) z_{i-1},$$

and, its control limits can be computed using the following equation [7]:

$$UCL, LCL = \mu_0 \pm L\sigma \sqrt{\frac{\lambda}{(2-\lambda)}\left[1-(1-\lambda)^{2i}\right]},$$

where, $\lambda$ is a smoothing parameter than can be assigned a value between zero and one. A value close to zero would imply that less importance is given to the more recent observations. Adjusting the value of $\lambda$ helps change the memory of the EWMA chart to handle certain sizes of faults. However, as in the case of the Shewhart and

CUSUM charts, the EWMA chart also assumes that the evaluated residuals are independent and Gaussian.

## 1.2. Research Objectives

Most univariate monitoring charts operate under the assumptions that the evaluated residuals are independent, Gaussian and contain a moderate level of noise. Limited work has been conducted to effectively address the violation in assumptions of these charts. Multiscale representation of data has shown promise in dealing with violations in these assumptions, e.g., for the case of autocorrelated residuals [16]. Therefore, the main aim of this work is to utilize the advantages of multiscale representation of data to improve the effectiveness of the Shewhart Chart under violations of various assumptions. Specifically, the following objectives will be sought to achieve this aim:

- Assess the performance of the Shewhart chart under the violations of its assumptions, namely the high noise level, the presence of autocorrelation in the data, and their deviation from normality.
- Develop a multiscale Shewhart chart algorithm that improves its performance when compared to the conventional one.
- Use illustrative examples (both synthetic and real-world applications) to analyze the effectiveness of the developed multiscale Shewhart chart algorithm.
- Highlight some future research directions.

# 2. MONITORING USING SHEWHART CHARTS: ADVANTAGES AND DRAWBACKS

## 2.1. Introduction to Shewhart Charts

A brief introduction to the Shewhart chart and other conventional univariate monitoring techniques was provided in the previous section. The main reason for the Shewhart chart's popularity is its computational simplicity. Figure 2 displays how the Shewhart chart (or other univariate charts) can be applied when monitoring a process. First, data points (observations) are obtained from one (or more) sensors measuring a particular process variable. The data are then compared to a model prediction of the monitored process variable, which results in the residuals of this variable. These residuals are basically the difference between the measured variable and its set point (targeted process value).

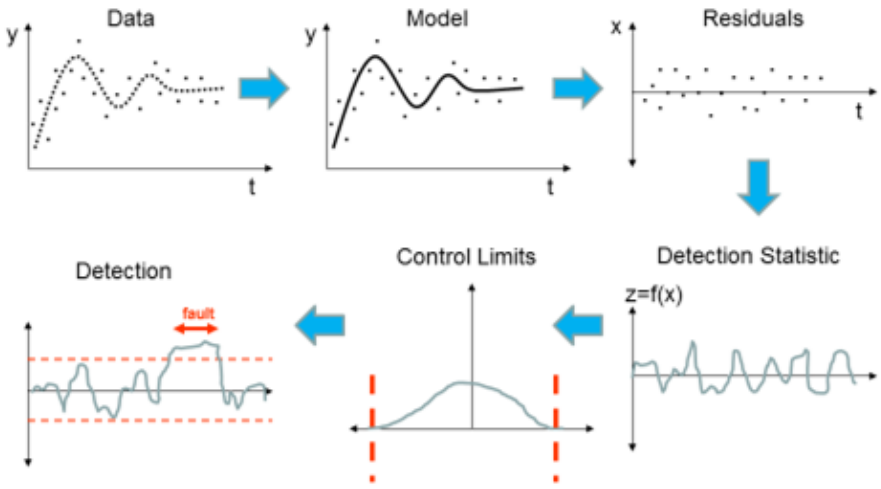


**Figure 2. Application of univariate charts for process monitoring.**

A detection statistic is then chosen, which is different for different monitoring techniques. After specifying the detection statistic, its distribution is determined to allow computing of the upper and lower control limits. These control limits are then used to detect faults in future process measurements.

For the Shewhart chart, which is the main focus of this research, the value of the residuals obtained at any time instant is the value of detection statistic. As stated previously, for Gaussian residuals, the upper and lower control limits can be computed as follows [6]:

$$UCL, LCL = \bar{\bar{x}} + 3\sigma ,$$

where, $\sigma$ is the standard deviation of the fault-free residuals.

An added advantage of the Shewhart chart is that it does not require any additional parameters to formulate the detection statistic, as well as the fault detection limits.

Figure 3 shows how the conventional Shewhart would look like in practice, with the red-dashed line representing the $3\sigma$ control limits. For 1024 normally distributed data samples, with a step fault of $3.5\sigma$ introduced between samples 501-600 (indicated by the region shaded in light blue), Figure 3 shows that the Shewhart chart is able to effectively detect a large section of the fault, with some missed detections and few false alarms.

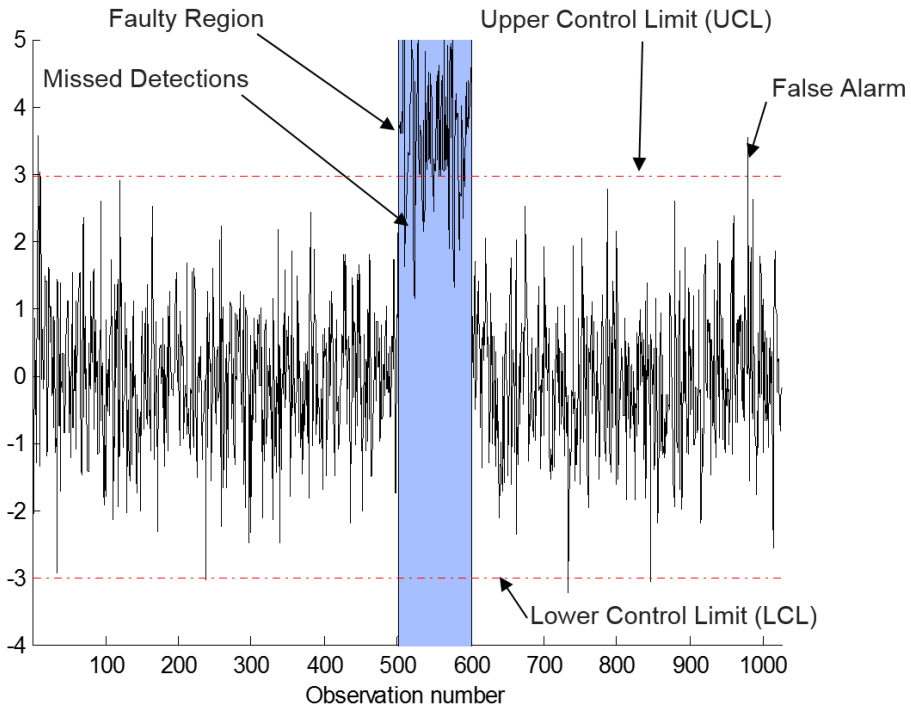


**Figure 3. Representation of the conventional Shewhart chart during process monitoring.**

The data used in Figure 3 assumes that none of the assumptions of the Shewhart chart are severely violated. However, as previously stated, the Shewhart chart operates under three main assumptions, which are the normality and independence of residuals and the presence of moderate noise level in the data. The next section examines the performance of the Shewhart chart under violation of these assumptions.

## 2.2. Assessing the Performance of the Shewhart Chart under Violation of Assumptions

### *2.2.1. Assessing the impact of high noise levels in the data on the performance of the Shewhart chart*

Most statistical monitoring techniques have to deal with noisy data, and this can affect the performance of any technique. High noise levels increase the difficulty in distinguishing between normal data and faults. In this section, the effect of measurement noise levels on the performance of the Shewhart chart will be examined.

The training data (of 500 observations generated from a zero mean Gaussian distribution) are used to compute the control limits of the Shewhart chart. The testing data are generated in a similar manner, and additive step faults with a magnitude of +2 and -2 are added between observations 201-250 and 401-450, respectively. The simulation is repeated for different values of noise standard deviation $(\sigma = 0.03$ to $\sigma=2)$, and for each value of $\sigma$ the performance of the Shewhart chart is assessed by examining the missed detection and false alarm rates. For each value of $\sigma$, a Monte Carlo simulation of 5000 realizations is performed to obtain meaningful results by averaging the missed detection and false alarm rates from all 5000 realizations.

The results of this simulation are illustrated in Figure 4, which shows that even though the noise level does not impact the false alarm rate, the missed detection rate can reach a 100% for very high levels of noise.

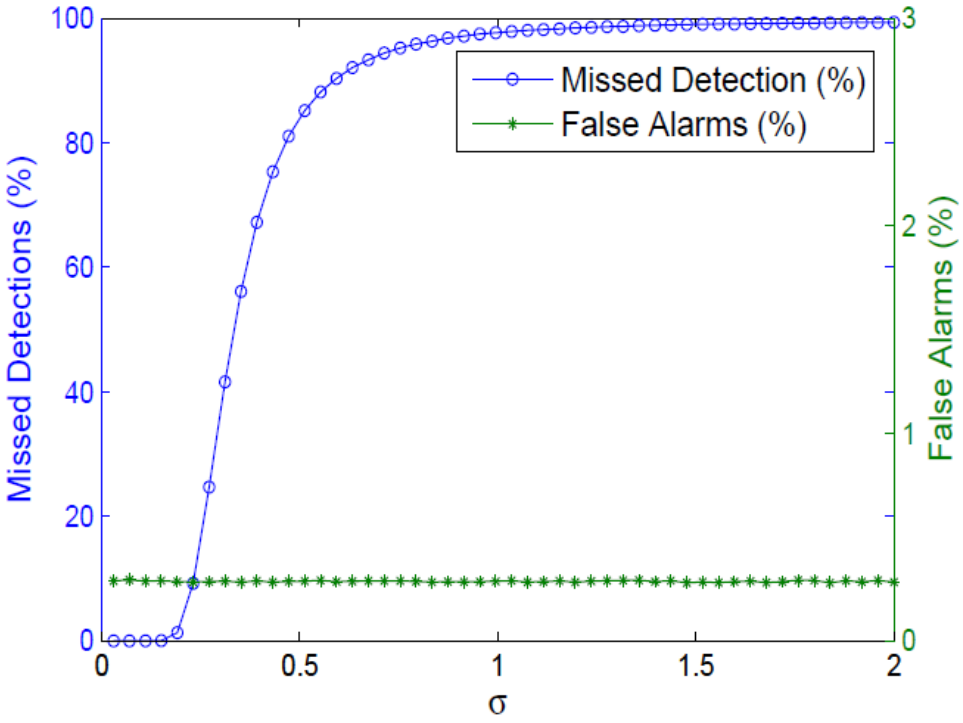


**Figure 4. Monte Carlo simulation results for the impact of noise on the performance of the conventional Shewhart chart.**

To understand the reason behind the deterioration in the performance of the Shewhart chart at high levels of noise, the charts for two noise levels, $\sigma = 0.8$ (moderate noise level) and $\sigma = 1.4$ (high noise level) are presented in Figure 5 and Figure 6, respectively. The red dashed lines represent the $3\sigma$ control limits that are used to identify faults, and the blue shaded areas represent the locations of the fault in the testing data.

These results clearly show that the deterioration in performance of the Shewhart chart is due to the fact that the noise starts to mask the faults as the noise level increases, hence reducing the fault detection capabilities of the Shewhart chart as indicated by the

increase in the missed detection rate. Additionally, it can be observed that the false alarm rate remains relatively constant for different noise levels, and this can be attributed to the fact that the control limits widen or contract depending on the amount of noise in the data, i.e., wider control limits for noisier data, and vice versa.

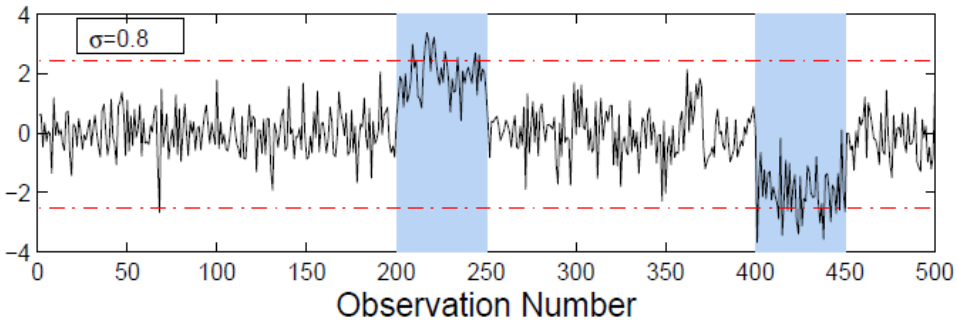


**Figure 5. Impact of noise level on the performance of the Shewhart chart, σ = 0.8.**

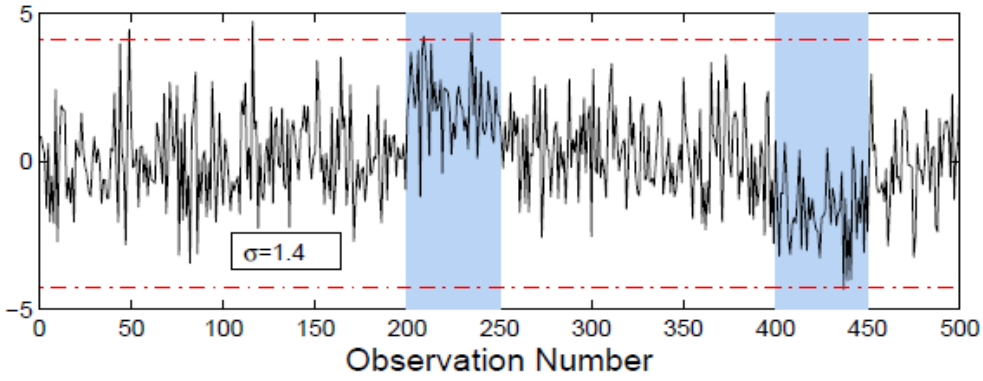


**Figure 6. Impact of noise level on the performance of the Shewhart chart, σ = 1.4.**

### *2.2.2. Assessing the impact of autocorrelation in the data on the performance of the Shewhart chart*

As previously stated, the Shewhart chart assumes that the residuals being evaluated are independent (uncorrelated). In practice, due to a variety of reasons, such as malfunctioning sensors or modeling errors, this assumption may not always be satisfied. In this section, the performance of the conventional Shewhart chart in the presence of autocorrelation in the data will be examined.

Most practical data, particularly chemical plants can be represented using stochastic time-series models, such as the autoregressive (AR) model. An AR model expresses the measured process data as a finite, linear sum of previous measurements, and random noise. The AR model of order $p$ has the following form [23]:

$\tilde{\quad} \quad \tilde{\quad} \quad \tilde{\quad} \quad \cdots \quad ,$

where, $a_i$ are the autoregressive coefficients for different previous measurements, $\tilde{\quad}$ is the series of deviations from $\mu$ (the targeted process variable), and $\xi_t$ is random noise, that is usually assumed to be Gaussian noise with zero mean and unit variance. A simple AR model with order one (AR(1)) can be used to represent autocorrelation in the data, and has the following form [23]:

$\tilde{\quad} \quad \tilde{\quad} \quad .$

To assess the effect of autocorrelation on the performance of the Shewhart chart, training data (consisting of 500 fault-free observations) are generated using an AR(1) model, and are then used to compute the $3\sigma$ control limits for the Shewhart chart. The

control limits are then used to detect faults in the testing data, which are also simulated using the same AR(1) model, but with an additive step fault (of a magnitude of 3) between observations 201-300.

The simulation is repeated for different values of $a$ (from 0.3 to 1) to represent different levels of autocorrelations. Smaller values of $a$ correspond to lower autocorrelation, while large values of $a$ correspond to higher autocorrelation. Again, a Monte Carlo simulation using 5000 realizations was carried out for each value of $a$ to ensure that statistically meaningful conclusions could be drawn and the results are illustrated in Figure 7.

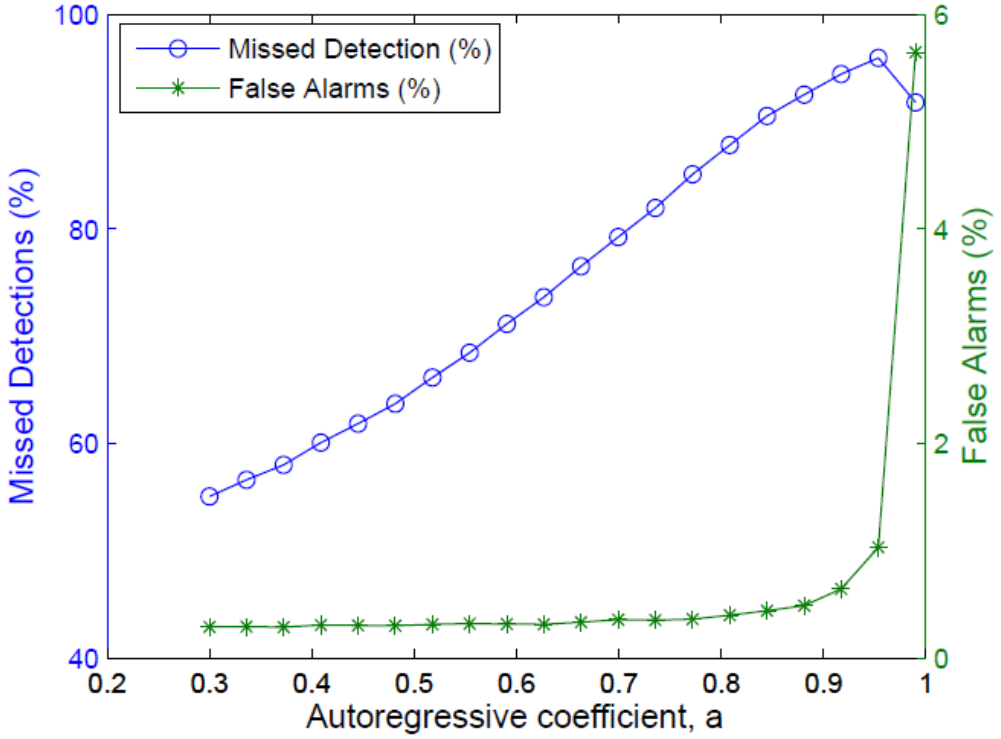


**Figure 7. Monte Carlo simulation results for the impact of autocorrelation on the performance of the conventional Shewhart chart.**

Figure 7 clearly shows that there is a deterioration in the performance of the conventional Shewhart chart for larger values of $a$, which correspond to greater autocorrelation in the observations. Although, the false alarm rate remains relatively constant (except at high values of autocorrelation), Figure 7 shows that the missed detection rate increases with an increase in the autocorrelation in the data, which indicates that the Shewhart chart is unable to efficiently detect the faults at high levels of autocorrelation.

To further illustrate the deterioration of performance of the Shewhart at high levels of autocorrelation, the time evolution of the Shewhart chart along with the corresponding autocorrelation functions, are displayed in Figure 8 and Figure 9 for the cases where $a = 0.3$ (moderate autocorrelation) and $a = 0.99$ (very high autocorrelation), respectively.

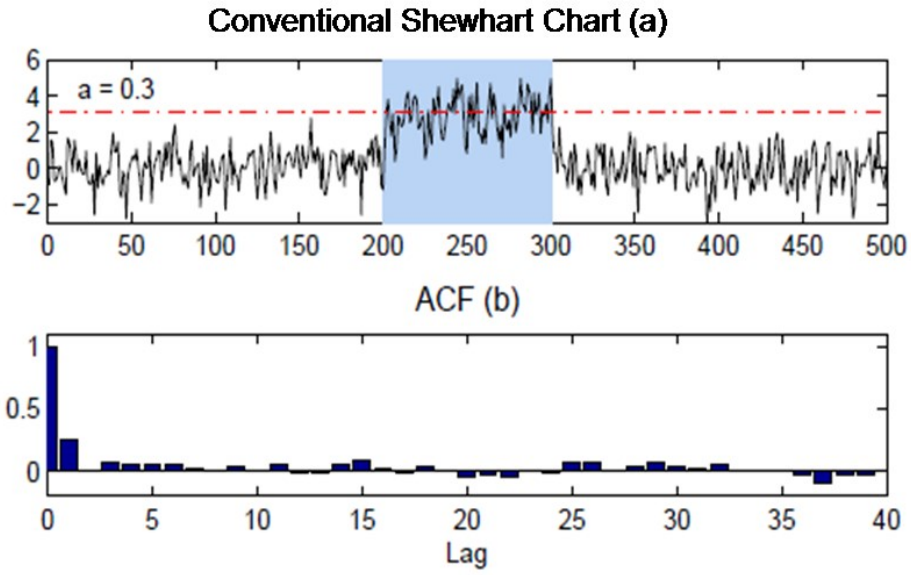


**Figure 8. Impact of autocorrelation on the performance of the Shewhart chart, $a = 0.3$.**

The results corresponding to $a = 0.3$ (shown in Figure 8), where the observations are moderately correlated show that Shewhart chart is able to detect the fault, but with a moderate missed detection rate. In the case of high level of autocorrelation, $a = 0.99$ (shown in Figure 9), however, the Shewhart chart is not able to effectively detect the presence of faults. These results clearly show that the presence of autocorrelation in the data greatly affects the performance of the conventional Shewhart chart.

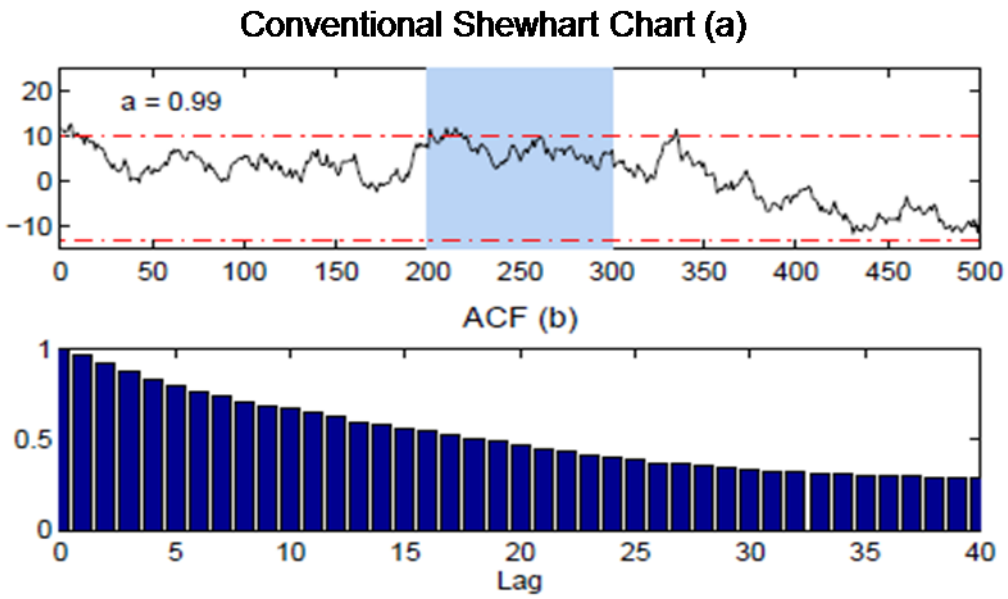


**Figure 9. Impact of autocorrelation on the performance of the Shewhart chart, $a = 0.99$.**

### *2.2.3. Assessing the impact of deviation from normality on the performance of the Shewhart chart*

The Shewhart chart also assumes that the evaluated residuals follow a Gaussian (normal) distribution. However, practical data may not always satisfy this assumption, which can degrade the performance of the Shewhart chart. In this section, the impact of violating the assumption of normality on the performance of the conventional Shewhart will be assessed.

There are numerous test and procedures that can be utilized to assess the normality of a particular distribution. Testing for univariate normality can be categorized as follow [24]:

- Graphical methods [25].
- Tests based on sample methods and moment-type tests [26][27].
- Other tests designed specifically to test for normality [28][29].

Graphical methods are important in various areas in the field of statistics, from fitting distributions to testing for goodness of fit, and they are able to provide a degree of detail that might not necessary come forth through a single statistic. Graphical methods can be either raw data plots or normal probability plots. Raw data plots include histogram plots, or stem-and-leaf plots. These plots provide a visual representation of the overall structure of a distribution. However, for a more accurate test for normality, normal probability plots are often preferred. Normal probability plots can be either normal quantile-quantile (Q-Q) plots or percent-percent (P-P) plots. To investigate the

deviation from normality of a particular distribution, it would be impractical to plot and examine a normal probability plot visually. Hence, a more practical approach is required.

Sample moment and moment type tests are frequently used to assess deviation from normality of distributions. Skewness is defined as the standardized third moment, and measures the degree of symmetry of a distribution, while kurtosis is defined as the standardized fourth central moment. More generally, kurtosis measures the peakedness and tail weight, and is often used in descriptive statistics as well as statistical inference literature [26]. A normal distribution is symmetrical, and thus the skewness value should be zero. A value for skewness greater than zero indicates that the distribution is skewed to the right, meaning that there are more observations on the right of the peak, and vice versa [27]. The kurtosis for normally distributed data equals three. A kurtosis value less than three means that the peak of the distribution is smaller than for a normal distribution, and vice versa. Even though sample moment and moment type tests are commonly used to assess the normality of a distribution, they do have drawbacks, i.e. the sample kurtosis can be very sensitive to outliers, and it is also possible for the statistic to rely heavily on only a few observations from the tail of the distribution, that could be either irrelevant or erroneous [30][31].

Other tests that were designed specifically to test for normality include sample entropy, Kullback-Leibler, relative entropy and similar metrics [28][29][32][33][34][35]. However, most of these tests only work under certain conditions. The Shapiro-Wilk metric is often considered to be one of the most powerful univariate normality tests, and hence, was chosen to quantify deviation from non-normality [24]. The power of a

particular test measures the ability of that test to reject the hypothesis (e.g., that a distribution is normal), when it is in fact not true.

For a given set of observations, $x_i$, that follow a normal distribution, the observations can be represented as follows [24][36]:

$$x_i = \mu + \sigma z_i .$$

The Wilk-Shapiro statistic can be computed using the following equation [24]:

$$W = \frac{b^2}{(n-1)s^2},$$

where, $s^2$ is the sample variance, and $n$ is the number of samples in the distribution. The constant $b$ is computed as follows [24]:

$$b = \mathbf{a'x} .$$

$b$ can be expanded as follows [24]:

$$b = \sum_{i=1}^{n/2} a_{n-i+1}\left(x_{n-i+1} - x_i\right),$$

$$a = \left(\mathbf{w'V}^{-1}\right) / \left(\mathbf{w'V}^{-1}\mathbf{V}^{-1}\mathbf{w}\right)^{1/2},\ \mathbf{a'a} = 1$$

where, the expected value of the $i$ th order is denoted by $\mathrm{E}\left(x_{(i)}\right) = w_i$, and $\mathrm{V}$ is the covariance matrix of the order statistics $x_{(1)},...,x_{(n)}$. For the vector $\mathbf{a}$, each element can be computed by $a_i = -a_{n-i+1}$.

An issue with the initial statistic that was formulated by Wilk and Shapiro was that the elements of $\mathbf{V}$ (and $a_i$) were known only for samples up to a size of 50 [36][37]. However, Royston and other works extend the portability of the Wilk-Shapiro test to

large samples [38][39][40][41]. The following transformation of W to normality is valid for sample sizes $7 \le n \le 2000$ [39]:

$$z = \left( \left(1 - W\right)^{\lambda} - \mu_y \right) / \sigma_y ,$$

where, the parameters $\lambda$, $\mu_y$, and $\sigma_y$ are provided by Royston [39]. Recent work has made it possible to extend the algorithm for sample sizes $3 \le n \le 5000$ [42][43]. Hence, the Shapiro-Wilk statistic was chosen to assess the impact of deviation from normality on the performance of the Shewhart chart. The Shapiro-Wilk metric varies from zero to one, a value closer to one implying that the distribution is closer to normality. This working range for the Shapiro-Wilk statistic (0 to 1) was another reason it was chosen over other statistics, to quantify the deviation from normality.

The chi-square distribution can be used to obtain distributions with varying degrees of non-normality and is represented by the following probability density function [21]:

$$f(x) = \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}},$$

where, $k$ represents the number of degrees of freedom, which can be altered to produce distributions of varying non-normality.

The training data (which consist of 500-fault free observations) are generated using a chi-square distribution in order to compute Shewhart chart control limits. The testing data (also of 500 fault-free observations) are generated using the same chi-square distribution, but with step faults added between observations 201-250 and 401-450, of

$\pm 3\sigma$, where $\sigma$ represents the standard deviation of the observations. Two faults in opposite directions were used to in order to observe the impact of the direction of the fault (positive or negative) on the fault detection performance, and also to ensure that the direction of the fault (relative to the skewness of the non-Gaussian data) does not affect the conclusions made out this assessment. The conventional Shewhart is then applied to detect these faults, and this process is repeated for different values of Shapiro-Wilk ranging from approximately 0.68 to 0.99. In order to obtain statistically valid conclusions, for each value of Shapiro-Wilk, a Monte-Carlo simulation is performed using 5000 realizations, and the results are displayed in Figure 10. These results clearly show that missed detection as well as false alarm rates increase (even though there is only a slight increase in the false alarm rate) as the Shapiro-Wilk decreases, i.e., as the observations deviate from normality.

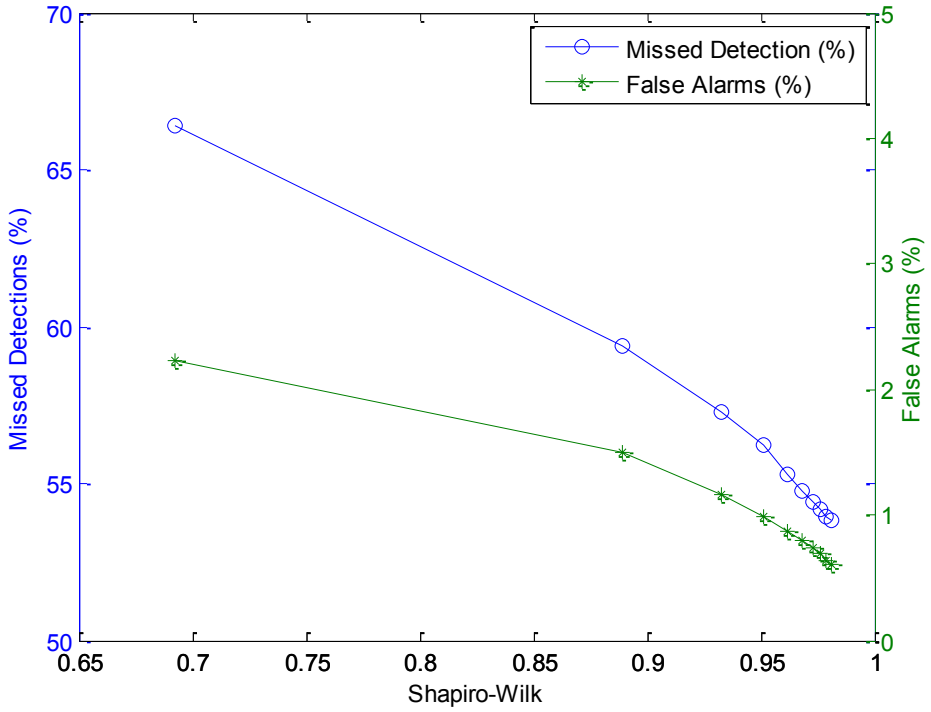


**Figure 10. Monte Carlo simulation results for the impact of deviation from normality on the performance of the conventional Shewhart chart.**

To further illustrate the impact of deviation from normality on the performance of the conventional Shewhart chart, the results of the Shewhart chart in the cases where the Shapiro-Wilk is 0.9867 (i.e., approximately Gaussian data), and in the case where the Shapiro-Wilk is 0.7351 (i.e., highly non-Gaussian data), along with their corresponding histograms are shown in Figure 11 and Figure 12, respectively. It can be clearly observed from these figures that as the observations deviate from normality, there is an increase in both the number of missed detections and false alarms.

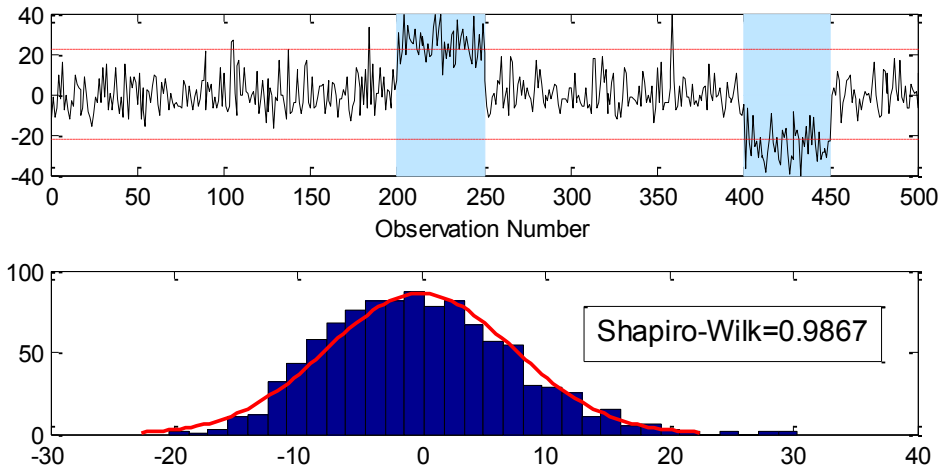


**Figure 11. Impact of deviation from normality on the performance of the Shewhart chart, Shapiro-Wilk=0.9867.**

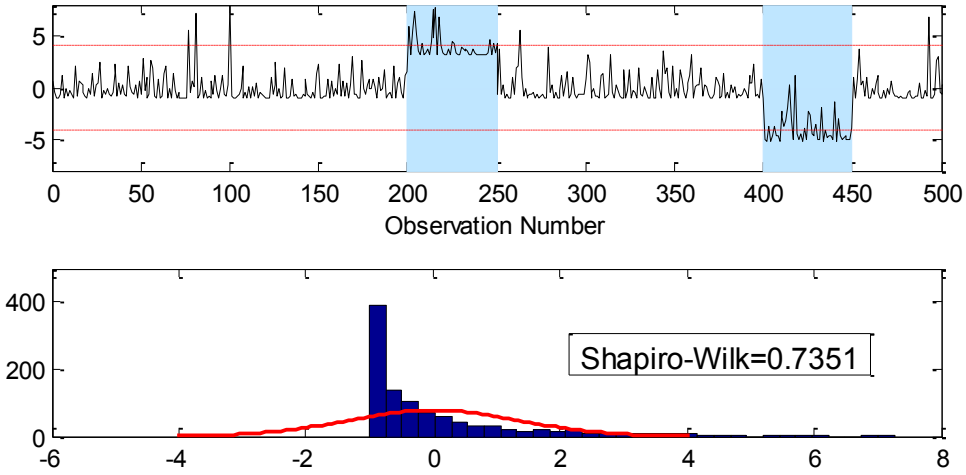


**Figure 12. Impact of deviation from normality on the performance of the Shewhart chart, Shapiro-Wilk=0.7351**

The results of the three assessments presented earlier on the effect of violating the underlying assumptions of the Shewhart chart clearly show that the violation of these assumptions adversely affect its performance. These limitations of the Shewhart chart can be addressed using multiscale representation. The next section presents a brief description of multiscale representation and how it can help satisfy the assumptions made by the Shewhart chart.

# 3. MULTISCALE REPRESENTATION AND ITS ADVANTAGES IN PROCESS MONITORING

In the previous section, it has been shown through simulated examples that the performance of the Shewhart chart deteriorates when the assumptions of normality, independence and moderate levels of noise (in the residuals) are violated. A data analysis approach that can be used to solve this problem and help deal with the limitations of the conventional Shewhart chart is multiscale representation of data. In this section, an introduction to multiscale wavelet-based representation of data will be presented, followed by a description of its advantages, and how it could potentially help address the deterioration in performance faced by the conventional Shewhart chart under violation of its underlying assumptions.

## 3.1. Introduction to Wavelet-Based Multiscale Representation of Data

Practical data usually possess multiscale features, i.e., contain features and noise that span a wide range in both the time and frequency domains. For example, a sharp change in the data spans a narrow range in the time domain, while spanning a much wider range in the frequency domain. Conversely, a slow change in the data may span a wide range in the time domain, but it may span a narrow range in the frequency domain [44]. It is also important to note that unlike white noise, that spans the entire frequency domain, autocorrelated noise can have varying contributions in the frequency domain.

Most commonly used univariate fault detection techniques operate at a single scale, as they are applied on time domain data, and thus usually they are unable to account for these multiscale characteristics. Wavelet-based multiscale representation of data is a powerful data-analysis, modeling, and feature extraction tool that has been widely used to help separate deterministic and stochastic features in real data [44][45].

Time-domain functions and data can be presented at multiple resolutions (scales) using wavelet and scaling basis functions, which have the following form [46][47]:

$$\theta(t) = \frac{1}{\sqrt{s}}\theta\left(\frac{t-u}{s}\right).$$

There are different families of basis functions that can be utilized to perform wavelet decomposition, such as the Daubechies and Haar basis functions [48][49][50][51].

Wavelet decomposition can be performed as follows. Given a time domain data set (signal), a coarser approximation of the signal (also called scaled signal) can be obtained by convoluting the original signal with a low pass filter $(h)$ (see Figure 13), that is derived from the scaling basis function which has the following form [52]:

$$\phi_{ij}(t) = \sqrt{2^{-j}}\phi\left(2^{-j}t - k\right),$$

where, $j$ and $k$ represent the discretized dilation and translation parameters, respectively.

The difference between the approximated signal and the original signal (also called detail signal) can be obtained by convoluting the original signal with a high pass

filter $(g)$ (see Figure 13), that is derived from the wavelet basis function, which has the following form [52]:

$$\psi_{ij}(t) = \sqrt{2^{-j}}\psi\left(2^{-j}t - k\right).$$

Repeating these approximations, the original signal can be represented as the sum of the last scaled signal and all detail signals, i.e. [52]:

$$x(t) = \sum_{k=1}^{n2^{-J}} a_{Jk}\phi_{Jk} + \sum_{j=1}^{J}\sum_{k=1}^{n2^{-J}} d_{jk}\psi_{jk}(t),$$

where, $J$ and $n$ represent the maximum possible decomposition depth and the length of the signal. Figure 13 provides a schematic diagram to illustrate the multiscale data representation procedure that was just described.

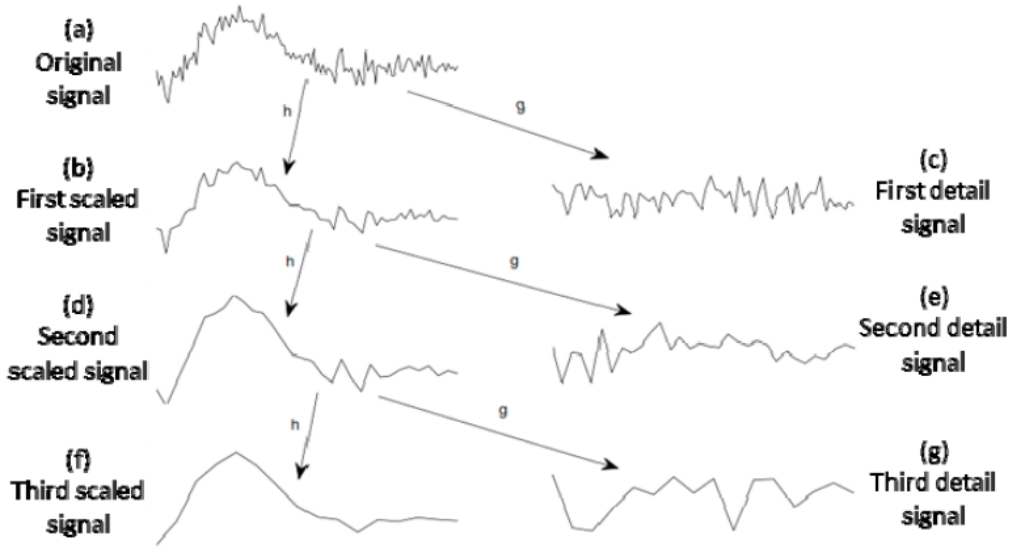


**Figure 13. Schematic diagram for multiscale representation of data.**

Wavelet-based multiscale representation of data has been utilized by various researchers in order to improve process monitoring. For example, the authors in [53]

have used multiscale representation to pre-filter data, and then use the pre-filtered data in process monitoring techniques such as Principal Component Analysis (PCA). Pre-filtering of the raw data, before process monitoring increases the effectiveness of the approach. However, pre-filtering may remove features in the data that are important for process monitoring. Also, multiscale representation of data has been used to develop a multiscale PCA (MSPCA) algorithm with improved performance over the conventional PCA methods [46]. This advantage is due to the fact that wavelet representation is an efficient noise-feature separation tool, which makes the MSPCA algorithm more sensitive to anomalies than the conventional PCA approach.

As shown in the previous sections, the presence of measurement noise, along with the limited abilities of the conventional univariate monitoring techniques to deal with autocorrelated and non-Gaussian data are challenges that need to be addressed. An effective tool that can be used to deal with these limitations is multiscale representation of data. The advantages of multiscale representation will be highlighted next.

## 3.2. Advantages of Multiscale Representation of Data

### *3.2.1. Noise-feature separation*

One of the main advantages of multiscale representation is its ability to separate noise from important features in the data. This is made possible by the successive application of high and low pass filters on the data, resulting in decomposing the data at multiple scales. This decomposition provides effective noise-feature separation as illustrated in Figure 13, and makes multiscale representation a more effective tool to filter or denoise measured process data compared to other commonly used linear filters, such as the exponentially weighted moving average (EWMA) filter.

The advantage of multiscale representation to separate noise from important features over the conventional linear filtering methods has been demonstrated through numerous applications, including filtering time-series genomic data [52].

In this research project, the noise-feature separation ability of multiscale representation will be exploited to improve the performances of the conventional Shewhart chart, to widen its applicability in practice.

### *3.2.2. Decorrelation of autocorrelated data*

It has been shown in Section 2.2.2 that one of the factors affecting the performance of the conventional Shewhart chart is the presence of autocorrelated observations, which can be due to a variety of reasons from modeling errors to malfunctioning sensors. An important advantage of multiscale representation is that

wavelet coefficients (detail signals) of autocorrelated data become approximately decorrelated at multiple scales [54].

The autocorrelation in a particular data set can be quantified using a simple statistic called the autocorrelation function (ACF). ACF quantifies the magnitude of the correlation between any two values from the data as a function of their separation [23]. It is a measure of the stochastic process memory in the time domain, and does not provide any information regarding the frequency fluctuation of the process [54]. A process contaminated with white noise has an autocorrelation function with a value of zero for all lags, except a value of unity at lag zero, and this indicates that the process is completely uncorrelated. On the other hand, correlated processes such as autoregressive moving average (ARMA) and autoregressive (AR) processes, have non-zero values at lags other than zero, and the ACF shows that there is autocorrelation in the data. However, as shown in Figure 14 the detail wavelet coefficients obtained from a correlated process are approximately decorrelated at multiple scales. The time domain data correspond to the following autoregressive AR(1) model:

$$x_t = ax_{t-1} + \varepsilon ,$$

where, $a = 0.7$, and $\varepsilon$ is white noise with zero mean and unit variance. The original time domain data are shown in Figure 14 (a), while the detail signals are shown in Figure 14 (c), (e), and (g). The second column in Figure 14 shows that although the ACF of the time-domain data (Figure 14 (b)) has non-zero values at several lags, which indicates that the observations are autocorrelated, the ACF of the detail signals are

approximately decorrelated (see Figure 14 (d), (f), and (h)). This observation can be attributed to the application of high-pass filters during wavelet decomposition [54].

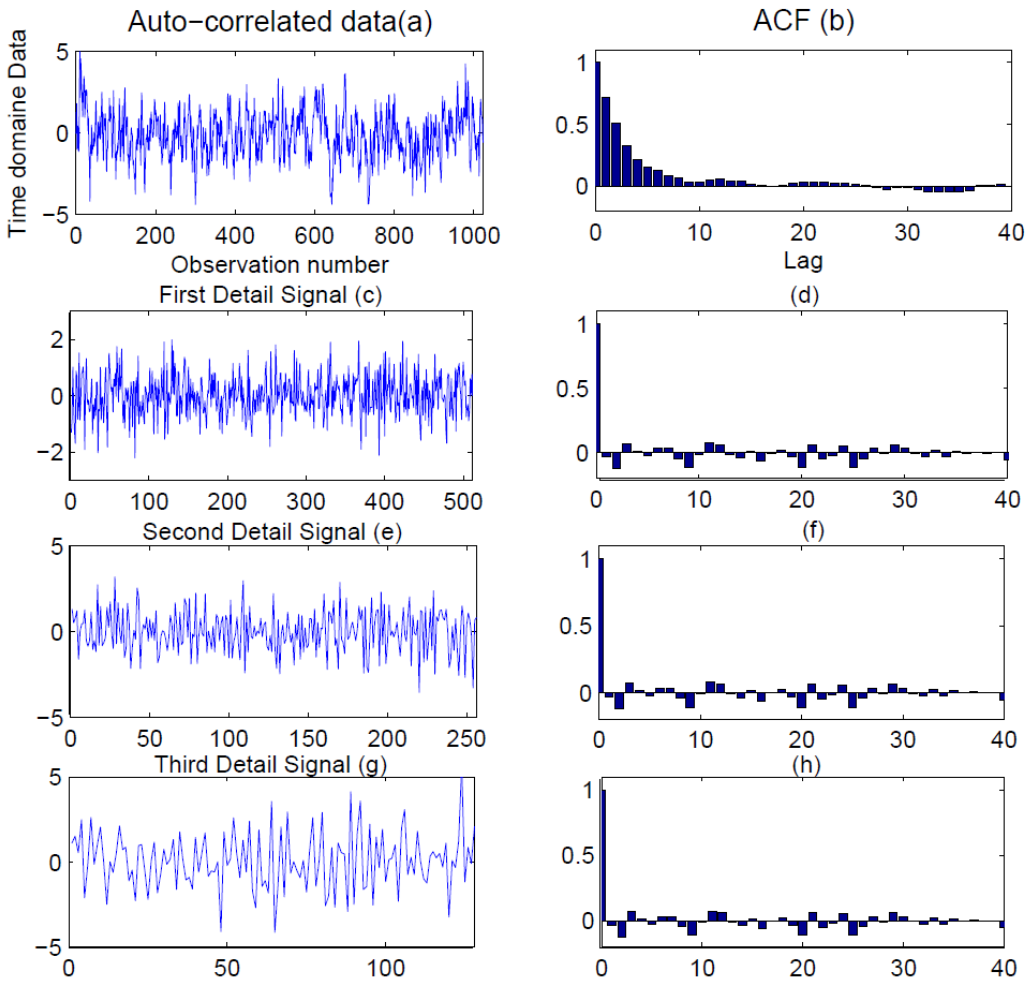


**Figure 14. Decorrelation of autocorrelated AR(1) data at multiple scales.**

### *3.2.3. Data are close to normal at multiple scales*

As shown in Section 2.2.3., another factor that affects the performance of the conventional Shewhart chart is the presence of non-Gaussian errors or noise in the measured observations. Non-Gaussian noise may be introduced by malfunctioning sensors that introduce skewed randomness or bias to the observations. Non-Gaussian errors can also be caused by modeling errors that leave non-modeled process variations in the model residuals. The previous section showed how wavelet decomposition results in detail signals that are approximately decorrelated at different scales. Fortunately, multiscale decomposition also helps solve the issue of non-Gaussian errors, as it helps makes the distribution of data more normal (Gaussian) at different scales.

Figure 15 shows that for chi-square distributed data, the distributions of details signals at multiple scales are much closer to normal than the original time domain data. Other distributions have also been used to validate this observation. Since the performance of the conventional Shewhart chart deteriorates when the data deviate away from normality, this feature of multiscale representation should help improve the performance of the Shewhart chart, especially when the data follow a non-Gaussian distribution.

In addition to the advantages illustrated and discussed earlier, multiscale representation of data is also able to handle non-stationary data, since the wavelet coefficients are stationary, and can therefore be used to represent both stationary and non-stationary data [15][46].

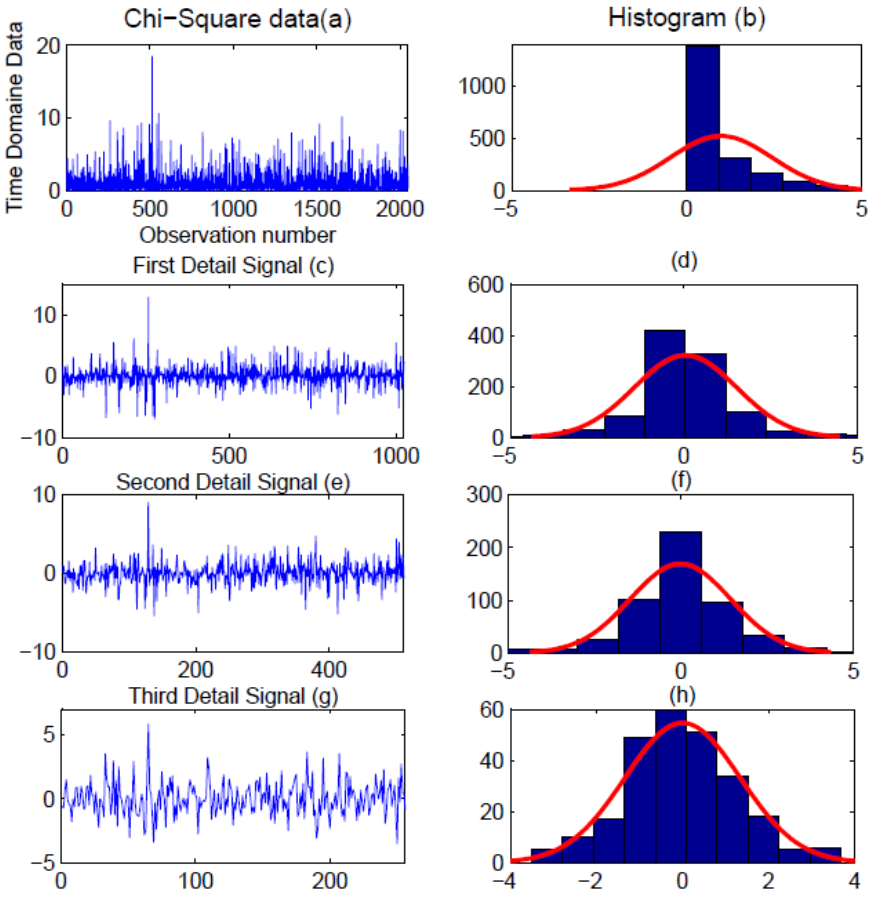


**Figure 15. Distribution of chi-square data at multiple scales.**

It is important to note that multiscale pre-filtering, i.e., filtering the data before applying any fault detection technique, may not fully exploit the advantages of multiscale representation. This is because performing multiscale filtering and fault detection as two separate tasks may remove features from the data which are important to the fault detection algorithm, or may introduce artifacts that were not originally in the data. Therefore, an algorithm that integrates fault detection with multiscale representation is needed to achieve the desired performance, which will be described next [55].

# 4. IMPROVED MONITORING USING MULTISCALE SHEWHART CHARTS

## 4.1. Multiscale Shewhart Charts

This section presents a multiscale Shewhart chart fault detection algorithm with an improved performance over the conventional Shewhart chart. The idea behind the multiscale Shewhart chart algorithm is to apply multiple fault detectors using the detail signals (at multiple scales) and the final scaled signal from the multiscale wavelet-based decomposition. The multiscale Shewhart fault detection algorithm consists of two phases as shown in Figure 16. In the first phase, fault-free training data are represented at multiple scales. Then, multiple fault detectors (i.e. Shewhart chart) are applied using detail signals at different scales, and the control limits at each scale are computed (which will be later used in the testing phase). These control limits are used to threshold the wavelet coefficients (detail signals) at each scale, and for each scale where any violation of the limits is detected, all wavelet coefficients at that scale are retained. Then, all retained detail signals are reconstructed, and fault detection (i.e. Shewhart chart) is applied again on the reconstructed data from which control limits are computed.

In the second phase, the testing data are also decomposed at multiple scales using the same wavelet filter used during the training phase. Then, the control limits computed at each scale from the training phase are applied to the detail signals of the testing data, and only the retained coefficients (after thresholding) are reconstructed to the time domain. Then, the control limits obtained from the reconstructed training data are applied to detect possible faults. This algorithm is schematically illustrated in Figure 16.

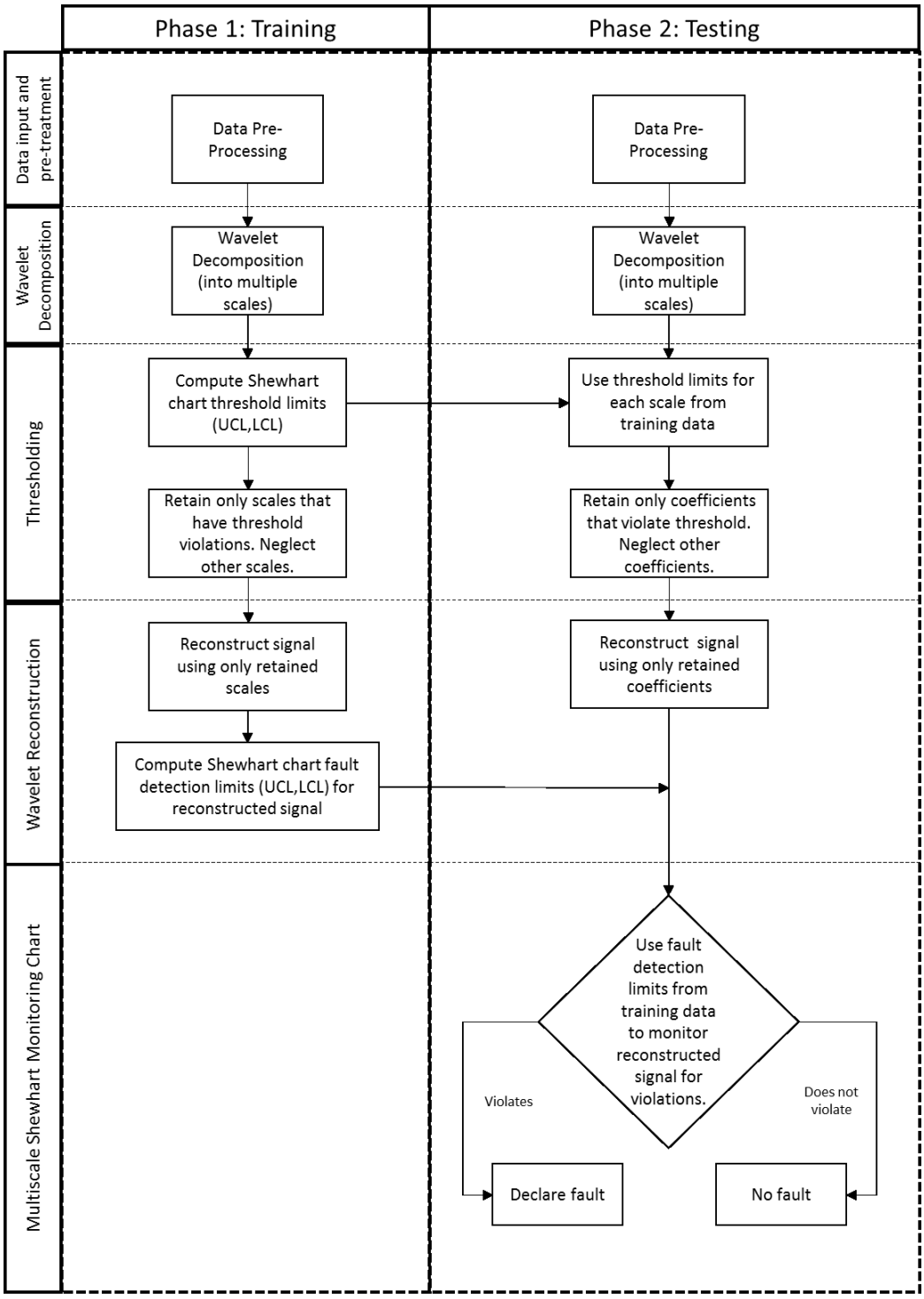


**Figure 16. A schematic diagram of the multiscale Shewhart chart fault detection algorithm [55].**

An important decision in the multiscale Shewhart algorithm is the choice of the decomposition depth or the number of scales used in wavelet decomposition. To examine the effect of decomposition depth on the fault detection performance, the following simulation is performed. It is important to note that for wavelet decomposition dyadic data is required.

A training data set (consisting of 512 observations) is generated using a zero mean Gaussian distribution), and are then used to compute the control limits of the Shewhart chart. The testing data set is generated in a similar manner, and additive step faults of $\pm 3\sigma$ between observations 201-250 and 401-450 are added, respectively, where $\sigma$ is the standard deviation of the data. The simulation is repeated for different decomposition depths, and for each decomposition depth, a Monte Carlo simulation of 5000 realizations is performed to obtain meaningful results. For each realization, the performance of the Shewhart chart is evaluated by examining the missed detection and false alarm rates. The mean values of missed detection and false alarm rates from all 5000 realizations are then computed at each decomposition depth. The results of the Monte Carlo simulation are displayed in Figure 17.

As Figure 17 shows, the minimum missed detection and false alarm rates occur at different scales. The optimum scale for the missed detection rate is the fourth, while the optimum scale for the false alarm rate is the first. In practice, a certain false alarm rate is generally acceptable, however, a high missed detection rate is less likely to be tolerated. Therefore, when deciding the decomposition depth to be used in the multiscale Shewhart chart fault detection algorithm, a compromise has to be made between the

acceptable missed detection and false alarm rates. As indicated by Figure 17, the ideal decomposition depth that satisfies these conditions is 4, and it is generally around half the maximum possible decomposition depth for a given set of observations.

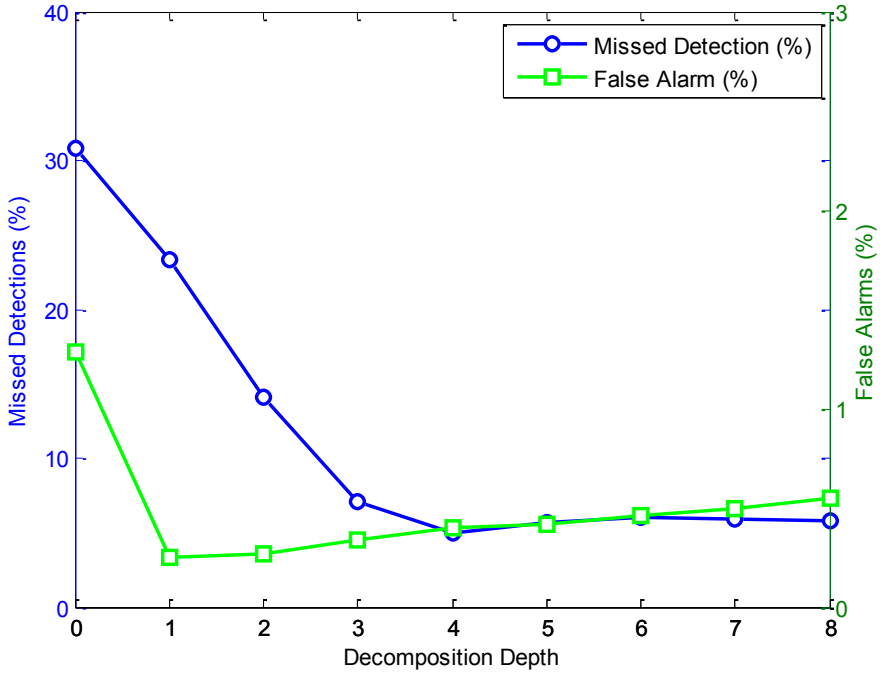


**Figure 17. Effect of decomposition depth on the performance of the multiscale Shewhart chart.**

Figure 17 also shows that the multiscale Shewhart chart provides an advantage over the conventional Shewhart chart. The conventional Shewhart chart provides a missed detection rate of around 30% (i.e., at decomposition depth = 0), and monitoring using the multiscale Shewhart chart is able to decrease the missed detection rate. However, after a decomposition depth of approximately 4, the percentage of missed detections remains relatively the same. This can be explained by examining what happens at the individual scales and the reconstruction of data set during the testing

phase. As the decomposition depth increases, the number of samples in the details signals of the training data decrease, which lead to tighter control limits. This plays an important role during the testing phase, when the detail signals at the deeper scales are all forced to have violations. This in turn forces all scales to be retained. Hence, decomposition to deeper scales does not provide any further advantages.

Monitoring using the developed multiscale Shewhart chart is also able to provide a reduction in the false alarm rate when compared to the conventional Shewhart chart, with a drop in false alarm rate from 17% (i.e., at decomposition depth = 0), to approximately 4% (at decomposition depth = 1). However, at deeper scales, as the decomposition depth increases, the percentage of false alarms increases. This can be attributed to Gibbs phenomenon, which is a common issue faced during wavelet reconstruction. Gibbs phenomenon is described as the production of artificial artifacts near discontinuities [56][57]. This results in the production of spikes in the data set, even if the actual observations are relatively constant. As the decomposition depth increases there are more scales that need to be reconstructed, which increases the potential for Gibbs phenomenon to take place, increasing the percentage of false alarms. When implementing the multiscale Shewhart chart it is important to choose a decomposition depth that minimizes both the percentage of missed detections and false alarms, which is approximately half the maximum possible decomposition depth. The choice of decomposition depth is therefore dependent on the amount of noise in the data. In this work, a decomposition depth of half the maximum possible depth will be used.

The performance of the multiscale Shewhart chart and its advantage over the conventional Shewhart is demonstrated through several illustrative examples in the next section.

## 4.2. Illustrative Examples (Simulated Synthetic Data)

In this section, the performances of both the conventional and multiscale Shewhart charts will be compared using synthetic data in order to illustrate the advantages of the multiscale method over the conventional one, especially when the data are non-Gaussian, autocorrelated, or contaminated with high levels of noise. As done previously, the performance of conventional and multiscale charts are evaluated and compared with respect to the missed detection and false alarm rates, and Monte Carlo simulations using 5000 realizations are performed for each comparison in order to make conclusions that are statistically valid. It is important to note that in this comparative study, the same data sets used in Section 2.2 are used.

### *4.2.1. Comparison of performance using data with different levels of noise*

The data used in this comparison are uncorrelated (independent observations) and normal (Gaussian). The training data (of 500 observations generated from a zero mean Gaussian distribution) are used to compute the control limits of the Shewhart chart. The testing data are generated in a similar manner, and additive step faults with a magnitude of +2 and -2 are added between observations 201-250 and 401-450, respectively. The simulation is repeated for different values of noise standard deviation

$(\sigma = 0.03$ to $\sigma$=2), and for value of $\sigma$ the performance of the Shewhart chart is assessed by examining the missed detection and false alarm rates. For each value of $\sigma$, a Monte Carlo simulation of 5000 realizations is performed to obtain meaningful results by averaging the missed detection and false alarm rates from all 5000 realizations.

Figure 18 shows that when the conventional and multiscale Shewhart charts are compared at different noise levels, the multiscale Shewhart chart does provide an advantage. The missed detection and false alarm rates are represented by the blue and green curves, respectively, while the conventional and multiscale techniques are represented by the thin and thick curves, respectively.

Even though the false alarm rates are comparable for both methods, the missed detection rates for the multiscale method are consistently lower than the conventional method, which is apparent at high levels of noise. This is because of the noise-feature separation provided by the multiscale approach, which allows better identification of faults.

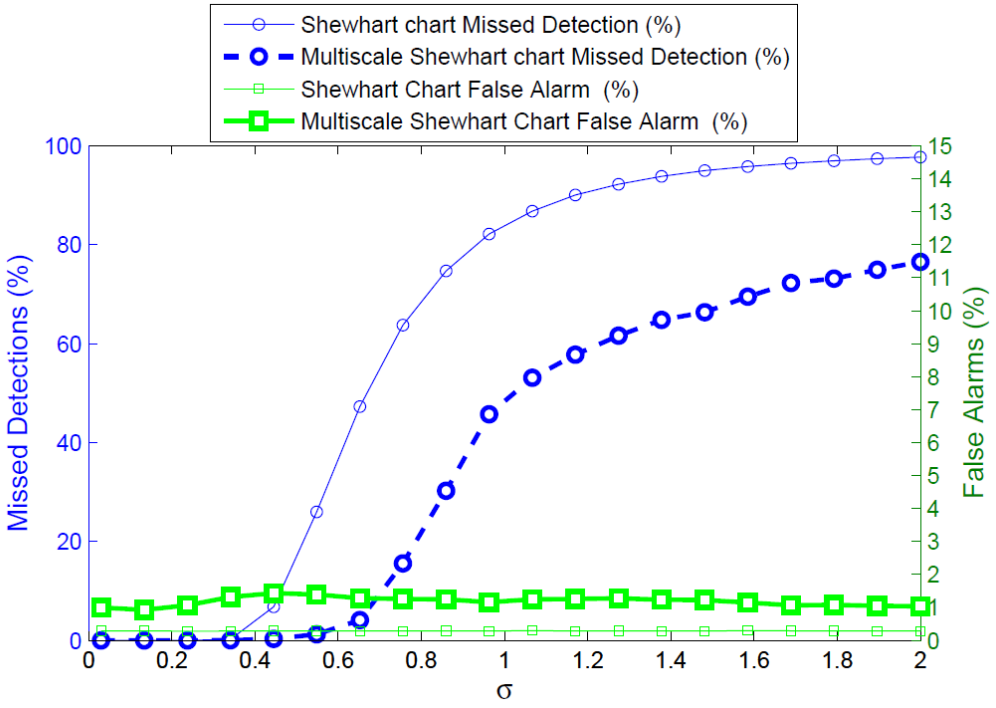


**Figure 18. Comparison between the performances of the conventional and multiscale Shewhart monitoring charts for the case of noisy data.**

Figure 19 demonstrates the advantage of the Multiscale Shewhart chart over the conventional Shewhart chart, where the value of noise standard deviation is 1.2 (moderate noise level). Furthermore, the false alarm in the multiscale Shewhart chart around observation 170, which is not present on the conventional Shewhart chart can be explained by Gibbs phenomenon. This also explains the slightly higher false alarm rate of the multiscale Shewhart chart, when compared to the conventional Shewhart chart in Figure 18.

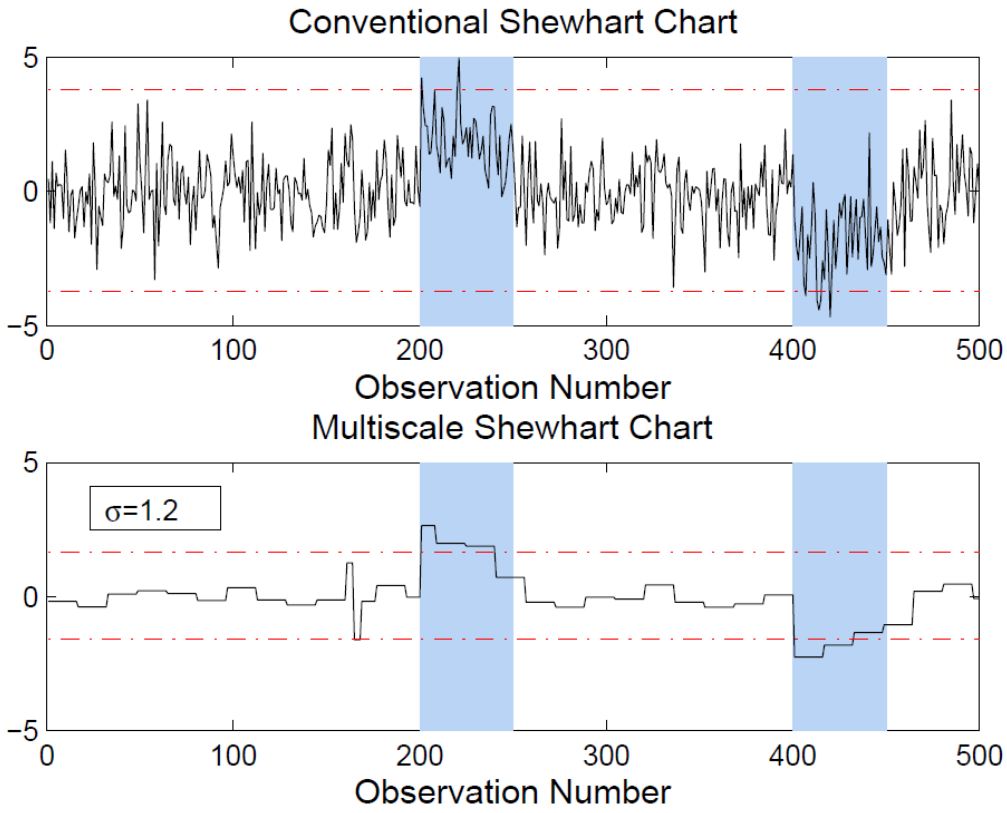


**Figure 19. Impact of noise on the conventional and multiscale Shewhart charts, σ = 1.2.**

### *4.2.2. Comparison of performance using autocorrelated data*

The data used for this comparison are generated using an autocorrelated AR(1) model. Training training data (consisting of 500 fault-free observations) are generated using the AR(1) model, and are then used to compute the $3\sigma$ control limits for the Shewhart chart. The control limits are then used to detect faults in the testing data, which are also simulated using the same AR(1) model, but with an additive step fault (of a magnitude of 3) between observations 201-300. The simulation is repeated for different values of $a_1$ (from 0.3 to 1) to examine the effect of autocorrelation. Again, a Monte Carlo simulation using 5000 realizations was carried out for each value of $a_1$ to ensure that statistically meaningful conclusions could be drawn and the results are illustrated in Figure 20.

Although, the false alarm rates of the two methods are comparable, Figure 20 shows that the multiscale method outperforms the conventional method as it provides significantly lower missed detection rates. This simulation clearly shows the robustness of the developed multiscale method to the presence of autocorrelation in the observations. This advantage is because multiscale representation of data causes autocorrelated data to become approximately decorrelated at multiple scales, and thus better satisfies the assumption of independence made in Shewhart charts.

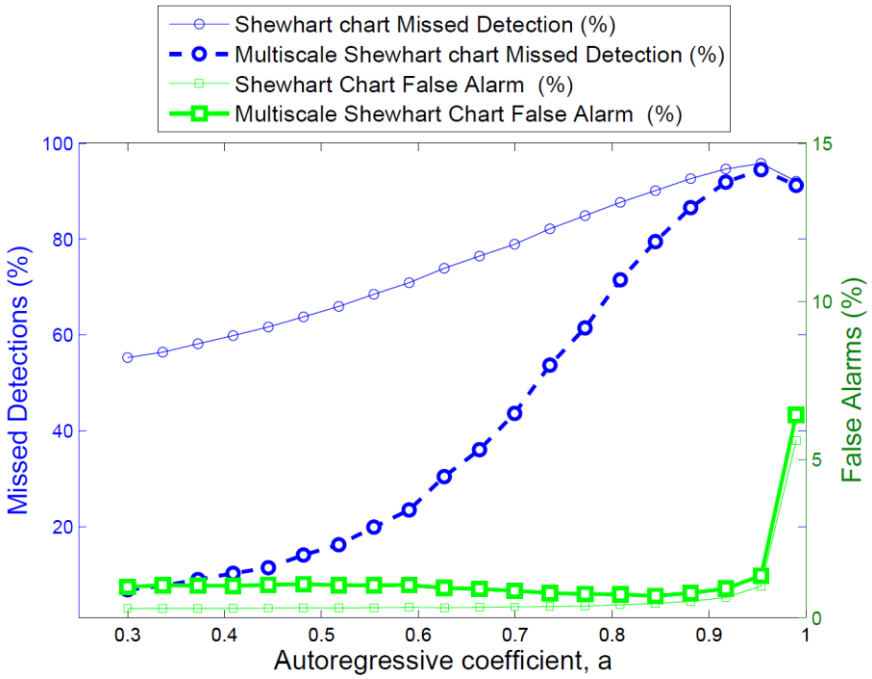


**Figure 20. Comparison between the performances of the conventional and multiscale Shewhart monitoring charts for the case of autocorrelated data.**

Figure 21 also illustrates the advantage of the multiscale Shewhart chart over the conventional Shewhart chart, where the autoregressive coefficient is 0.3 (moderate autocorrelation).

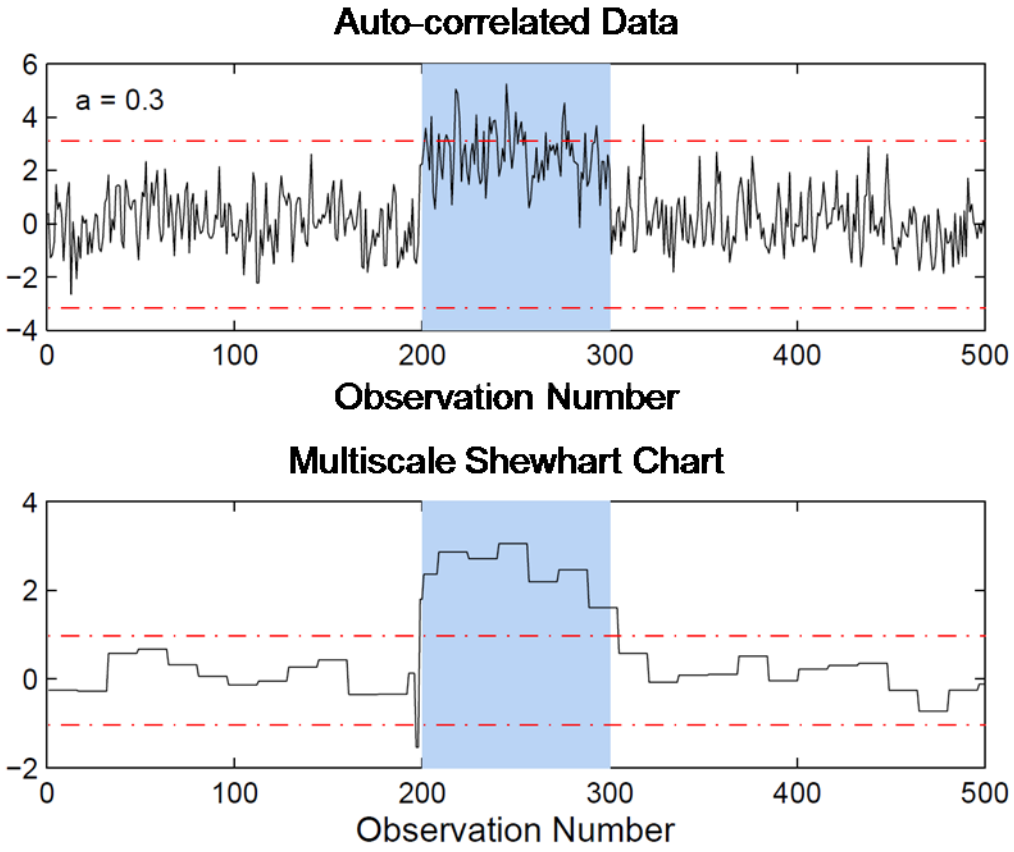


**Figure 21. Impact of autocorrelation on the conventional and multiscale Shewhart chart, $a$ = 0.3.**

Similar to Section 4.2.1, the slightly higher false alarm rate of the multiscale Shewhart chart when compared to the conventional Shewhart chart (see Figure 20), can be explained by Gibbs phenomenon. This effect is also observed in the multiscale Shewhart chart in Figure 21 (around observations 197-198 and 301-304).

### *4.2.3. Comparison of performance using non-Gaussian (chi-square) data*

The data from the same chi-square distribution that was used to examine the deterioration of performance of the conventional Shewhart chart, was used to compare the relative performances of the conventional and multiscale Shewhart charts. The training data (which consist of 500-fault free observations) are generated in order to compute Shewhart chart control limits. The testing data (also of 500 fault-free observations) are generated using the same chi-square distribution, but with step faults added between observations 201-250 and 401-450, of $\pm 3\sigma$, where $\sigma$ represents the standard deviation of the observations. Both the conventional and multiscale Shewhart charts are then applied to detect these faults, and this process is repeated for different values of Shapiro-Wilk ranging from approximately 0.68 to 0.99. In order to obtain statistically valid conclusions, for each value of Shapiro-Wilk, a Monte-Carlo simulation is performed using 5000 realizations, and the results are displayed in Figure 22.

These results show that there is a deterioration of performance in both the conventional and multiscale Shewhart charts at lower values of Shapiro-Wilk (i.e., when the observations follow a distribution farther away from normality). However, it should be noted that the multiscale Shewhart chart provides lower missed detection and false alarm rates, when compared to the conventional Shewhart chart. These advantages are mainly because the distributions of the detail signals (of a non-Gaussian data set) become nearly Gaussian at multiple scales. This advantage of the multiscale Shewhart chart is also illustrated by Figure 23, for the case when the Shapiro-Wilk statistic is 0.9427 (nearly Gaussian).

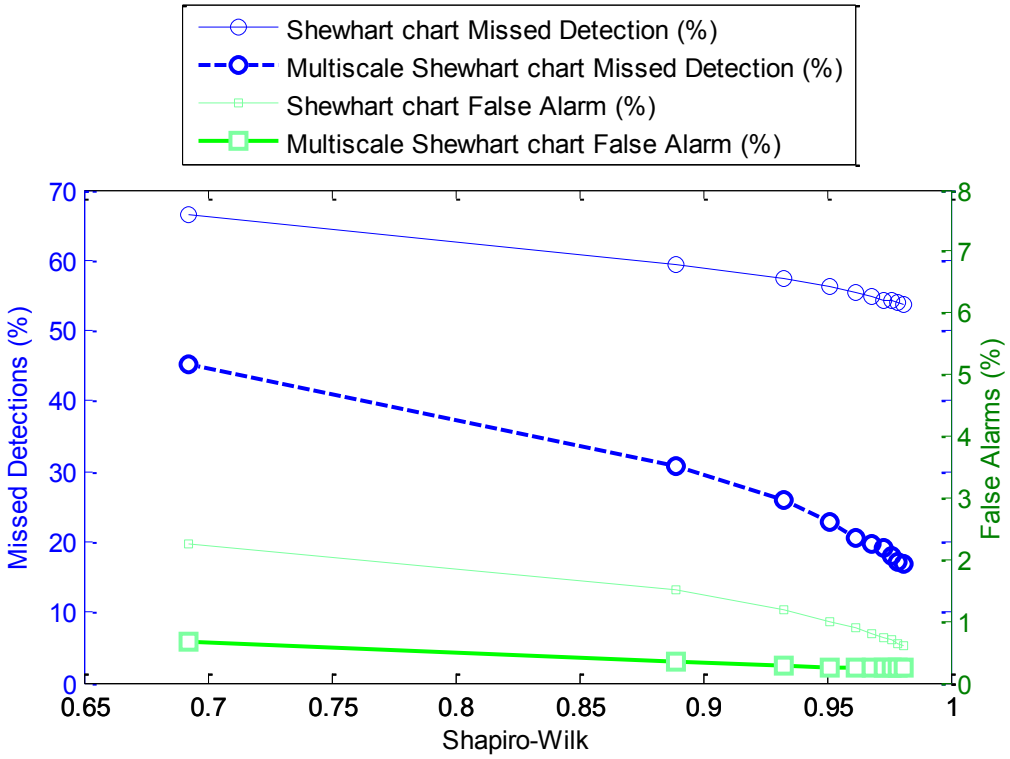


**Figure 22. Comparison between the performances of the conventional and multiscale Shewhart chart monitoring charts for the case of Non-Gaussian data.**

As observed from Figure 22 and Figure 23, the multiscale Shewhart chart is able to perform better than the conventional Shewhart chart in detecting a step fault with a magnitude of $\pm 3\sigma$, when the observations deviate from normality. Similar Monte Carlo simulations are performed using fault sizes of $\pm 2.5\sigma$ and $\pm 3.5\sigma$, and the results for three step sizes are presented in Figure 24.

There are a few conclusions can be drawn from the results that were obtained. In all three cases, it can be noted that when the observations are approximately normal, that the multiscale Shewhart chart is able to outperform the conventional Shewhart chart. The

larger the magnitude of the fault, the more accurately the multiscale Shewhart chart can detect these large faults (see Figure 24 (b) and (c)) regardless of the degree of non-normality.

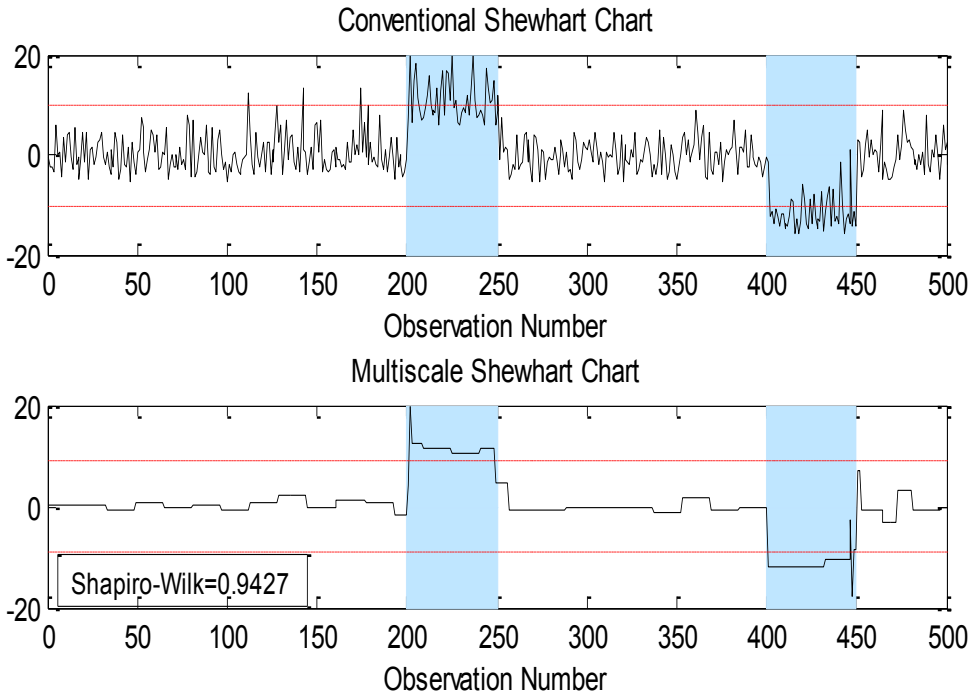


**Figure 23. Impact of deviation from non-normality on the performances of the conventional and multiscale Shewhart charts, Shapiro-Wilk=0.9427.**

However, when the magnitude of the step fault is smaller, the missed detection rates obtained by the conventional and multiscale Shewhart charts are comparable for different degrees of non-normality (see Figure 24 (a)). In this case the multiscale Shewhart chart does not show much of an advantage over the conventional Shewhart chart, and both charts miss a majority of the fault.

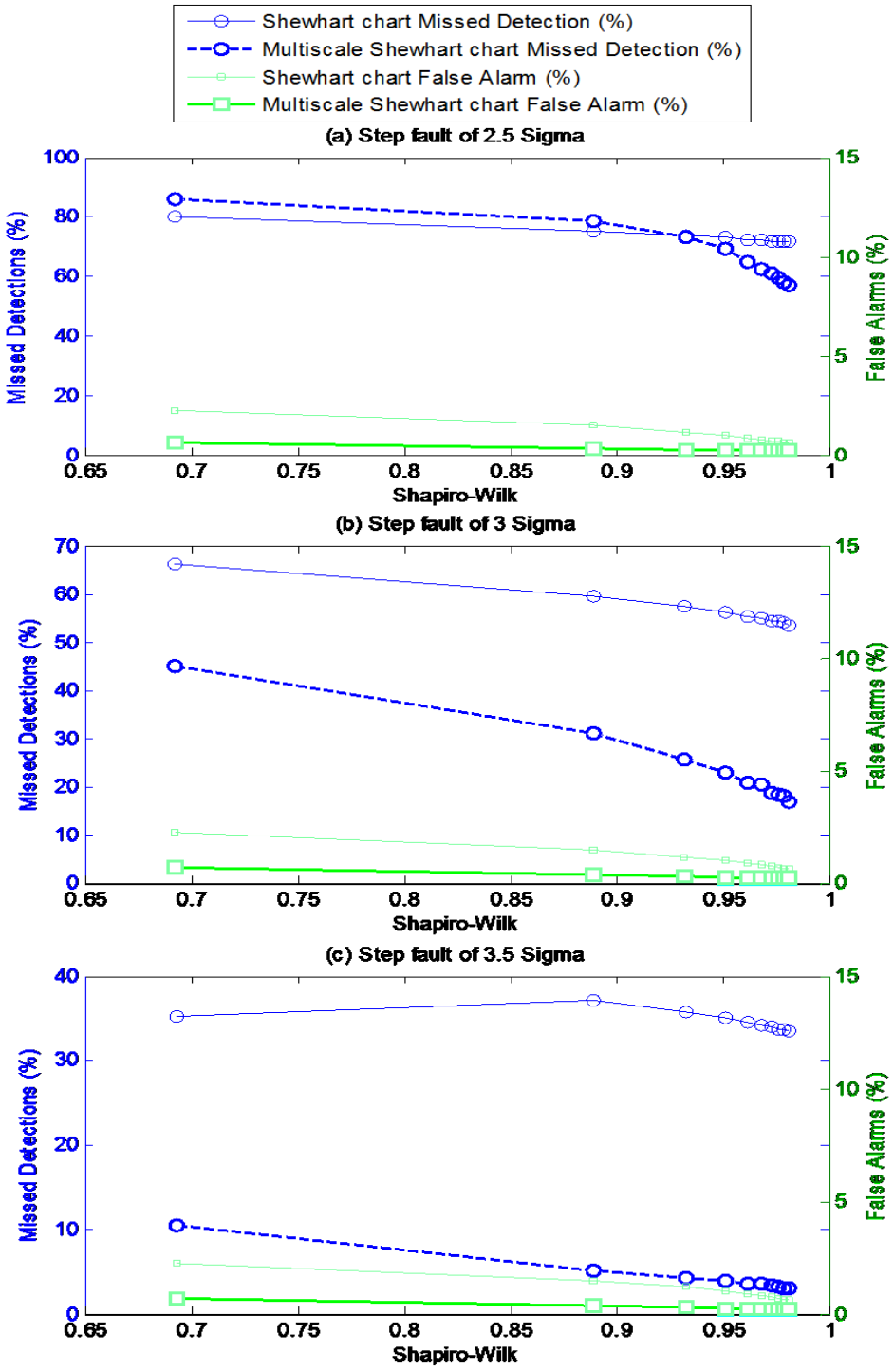


**Figure 24. Comparison between the performances of the conventional and multiscale Shewhart chart monitoring charts for the case of Non-Gaussian data (for step faults of different magnitudes).**

These results are further illustrated by Figure 25 and Figure 26, that display the performance of the conventional and multiscale Shewhart charts for step faults of $\pm 2.5\sigma$ and $\pm 3.5\sigma$, respectively. Figure 25 corresponds to point in Figure 24 (a), when the distribution is approximately normal (and with a smaller fault), showing that there is only a slight advantage with the multiscale Shewhart chart when compared to the conventional Shewhart chart. As stated before, this is due to the small size of the fault.

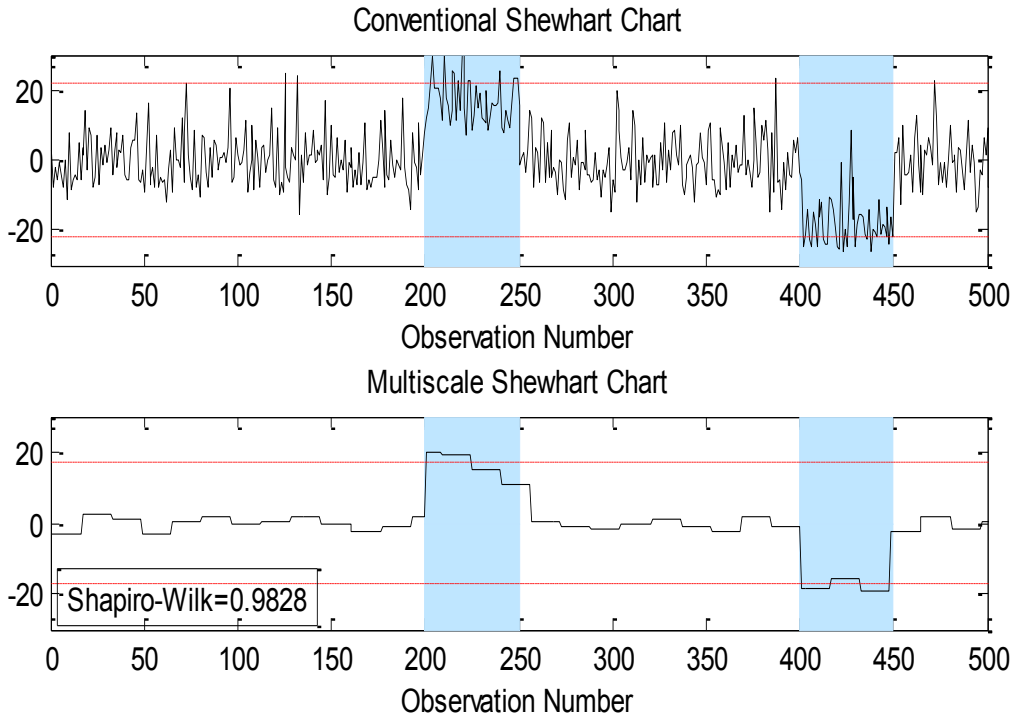


**Figure 25. Comparison in performance of the conventional and multiscale Shewhart chart for a step fault of magnitude σ = 2.5.**

Figure 26 corresponds to the point on Figure 24 (c), when the distribution is approximately normal (and with a larger fault), showing that the multiscale Shewhart chart performs better than the conventional Shewhart chart for faults larger than $3\sigma$.

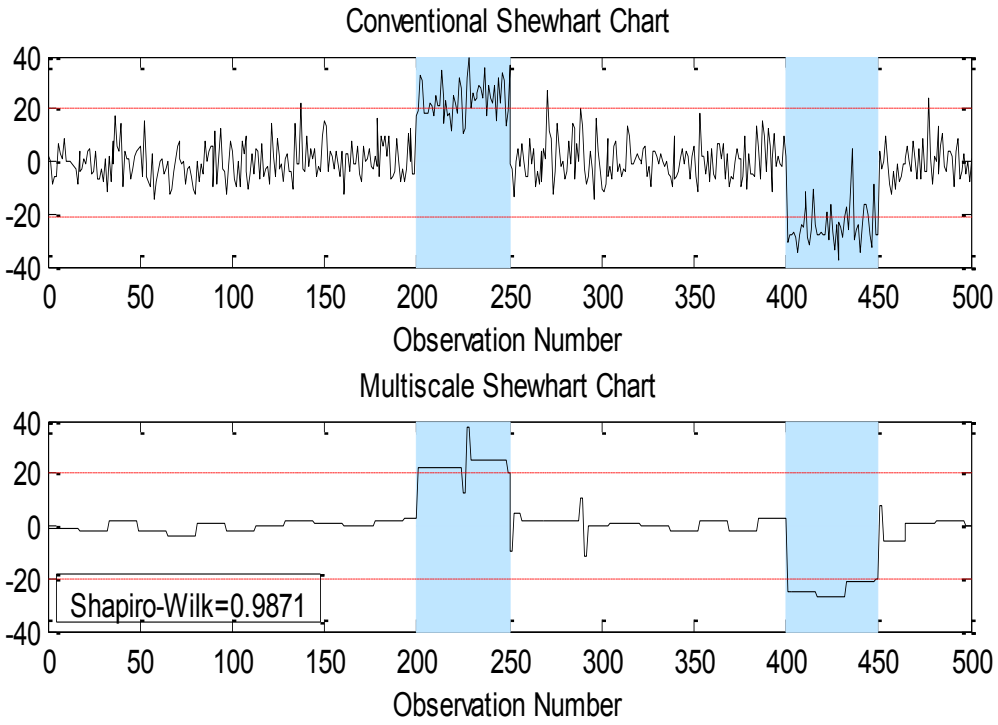


**Figure 26. Comparison in performance of the conventional and multiscale Shewhart chart for a step fault of magnitude σ = 3.5.**

# 5. APPLICATIONS OF MULTISCALE SHEWHART CHARTS

Synthetic data provide a useful means of examining the performance of different monitoring techniques. However, to assess the actual effectiveness of any monitoring technique, data from real applications need to be used. In this section, simulated distillation column data and real genomic copy number data are used to compare the performances of the conventional and multiscale Shewhart charts.

## 5.1. Monitoring using a Simulated Distillation Column

Separation processes are a critical component of many industries and distillation columns play a vital role. The operation of distillation columns need to be monitored in order to ensure that the quality meets the sought standards.

In this example, the conventional and multiscale Shewhart charts are compared through their utilization to monitor the operation of a simulated distillation column. To achieve this, the dynamic operation of a distillation column consisting of 32 theoretical layers (containing a total condenser and a reboiler), was simulated using Aspen Tech 7.2. The feed stream enters the column at stage 16, as a saturated liquid (flow rate of 1 kmol/s), at 322 K and composition 40 mole % propane and 60 mole % isobutene. The nominal operating conditions of the distillation column are provided in [58]. The dynamic data are generated by perturbing the feed and reflux flow rates from their nominal operational values. First, step changes (with magnitudes of $\pm 2\%$) in the feed flow rate around its nominal value are introduced. The process is first allowed to settle to

a new steady state before similar step changes (with magnitudes of $\pm 2\%$) in the reflux flow rate around its nominal normal operational condition are introduced. A date set consisting of 1024 observations is generated. The dynamic compositions and temperature data obtained from this simulation are assumed to be free of noise, and are then contaminated with zero mean Gaussian noise. From fault detection point of view, the objective of this simulated example is to monitor the composition of propane in the distillate stream, $x_D$.

As illustrated in Section 2.1, the available process observations need to be compared to their values to compute the residuals that can be used for monitoring purposes. Using the available observations from the simulated distillation column, a Partial-Least-Squares (PLS) regression model was constructed and used to obtain the residuals.

The residuals obtained were then split into two sets (512 samples each) for training and testing. The training data set was used to compute the control limits to be applied on the testing data. To illustrate the advantage of the developed multiscale Shewhart technique over the conventional Shewhart chart, two step faults of magnitudes $\pm 2.5\sigma$ were added to the residuals between observations 51-100 and 401-450, respectively (where $\sigma$ is the standard deviation of the samples in the data set).

Figure 27 displays the results that were obtained for both the conventional Shewhart chart (see Figure 27 (a)) and the multiscale Shewhart chart (see Figure 27 (b)). The red line represent the $3\sigma$ control limits of the Shewhart chart. As demonstrated by the Figure 27, the multiscale Shewhart chart is able to better detect the fault, with a

significantly lower missed detection rate than the conventional Shewhart chart. Figure 27 also shows that the multiscale Shewhart chart provides a lower false alarm rate than the conventional Shewhart chart.

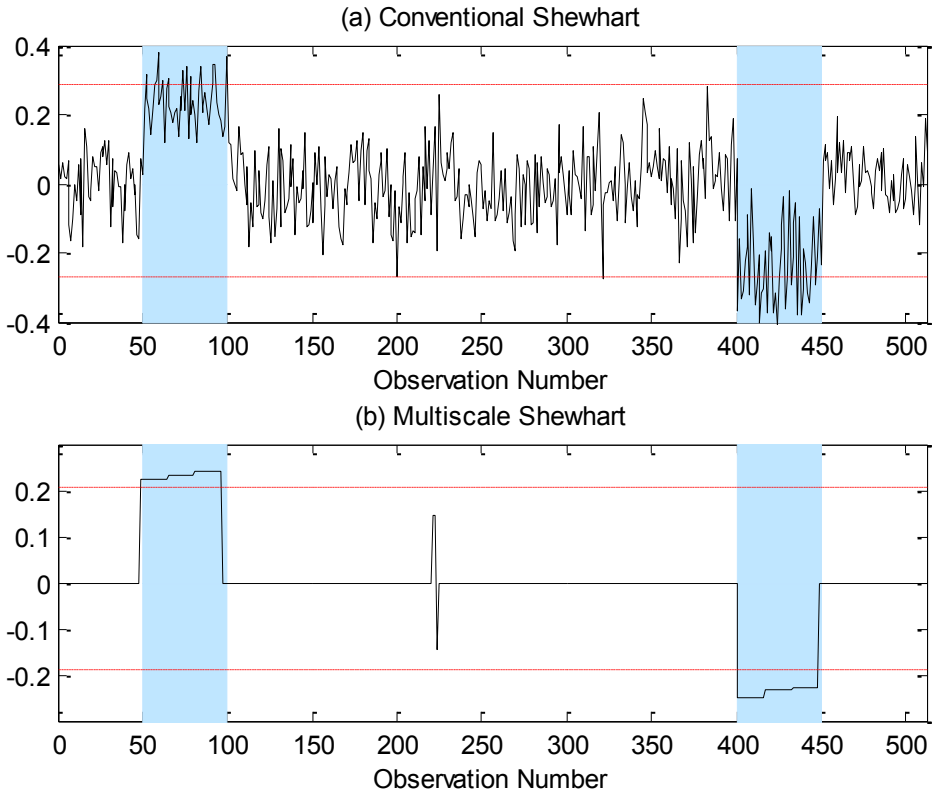


**Figure 27. Comparison in performance of the conventional and multiscale Shewhart chart for a step fault of magnitude σ = 2.5 in the residuals.**

The same simulation was repeated for a step fault of $\pm 3.5\sigma$, added to the residual between observations 51-100 and 401-450, respectively, and Figure 28 illustrate the results that were obtained. Although, the conventional Shewhart (see Figure 28 (a)) now performs better than its performance in the case of a smaller fault (see Figure 27 (a)), the missed detection rate is still relatively high when compared to the multiscale

Shewhart chart (that achieves nearly 100% detection). Figure 28 again illustrates that the multiscale Shewhart chart is able to provide a lower false alarm rate than the conventional Shewhart chart.

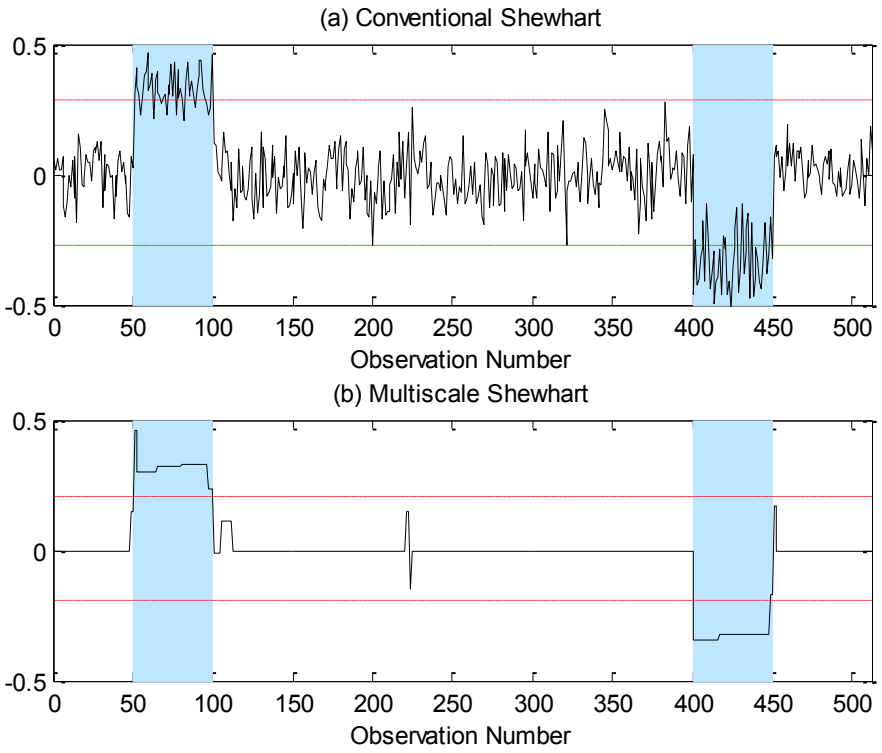


**Figure 28. Comparison in performance of the conventional and multiscale Shewhart chart for a step fault of magnitude σ = 3.5 in the residuals.**

As illustrated through this simulated distillation column example, the multiscale Shewhart chart is able to perform better than the conventional Shewhart chart for two magnitudes of step faults, as it is able to better detect the fault in both scenarios. Therefore, implementing the multiscale Shewhart chart for process monitoring in chemical industries could prove to be advantageous.

### 5.2. Detection of Aberration in Copy Number Data

Copy number aberrations are often present in cancer genomes, and other genomic diseases [52][59][60]. However, many of these aberrations are random, and go undetected [61]. Microarray Comparative Genomic Hybridization (CGH) is an experimental approach that is employed in order to scan DNA copy number data. This process allows for measurement of relative changes at thousands of DNA loci.

During the process, test DNA sequences are hybridized using a normal DNA sequence. The fluorescence intensities obtained out of a tested DNA sequence is compared to those obtained from a control (healthy) sequence, and the ratio of the two intensities is computed. When log2 of the ratio is zero, this indicates that there is no difference between the normal and tested sequence. However, if log2 of the ratio is either below or above zero, this indicates that there is an anomaly in the copy number data (which could be either amplification or deletion of DNA).

The objective of applying fault detection to monitor aberrations in copy number data is to identify points along the DNA sequence that are statistically different from zero, which indicates that the genome is not normal. This information could be of particular importance to medical professionals to help diagnose various diseases using genetic copy number data [52].

One major issue of copy number data is that they are very noisy, where the signal-to-noise ratio can go as lower as unity [59]. As demonstrated through Section 2.2, this affects the performance of most fault detection techniques. Another issue faced while monitoring copy number data is the availability of small number of data samples.

The noise-feature separation advantage of the developed multiscale Shewhart chart can help overcome these issues. It is also possible that genomic copy number data can be autocorrelated or non-Gaussian, and the developed multiscale Shewhart chart can deal with these challenges as well.

Hence, the aim of this application is to improve the fault detection performance using the developed multiscale extensions of the Shewhart chart, in order to better detect anomalies in genetic copy number data.

To illustrate how the developed multiscale Shewhart chart can be used for monitoring copy number data for aberrations, real genetic copy number data are used. The copy number data used represents Bacterial Artificial Chromosome (BAC) array data on 15 fibroblast cells [60]. Two chromosomes (9 and 14) were used for this example, and these data sets are freely available [60].

The number of available copy number data samples are generally small. Chromosome 9 only has 106 observations available. In order to implement process monitoring techniques, the available data needs to be split into training and testing data sets. Therefore, we decided to use the observations closest to zero as the training data set, while the testing data set is comprised of the whole set of available observations.

Figure 29 displays the results that were obtained, for both the conventional Shewhart chart (see Figure 29 (a)), and the multiscale Shewhart chart (see Figure 29 (b)). Figure 29 clearly shows that the available data are very noisy.

As demonstrated in Figure 29, the multiscale Shewhart chart is able to detect the fault better than the conventional Shewhart chart. It is important to note that the precise

location of the fault is unknown when both techniques are implemented. The conventional Shewhart chart has one potential missed detection and another observation nearly going undetected, along with one false alarm. The fault-free and faulty regions on the multiscale Shewhart chart can be distinguished more clearly (with no missed detections or false alarms) than the conventional Shewhart chart.

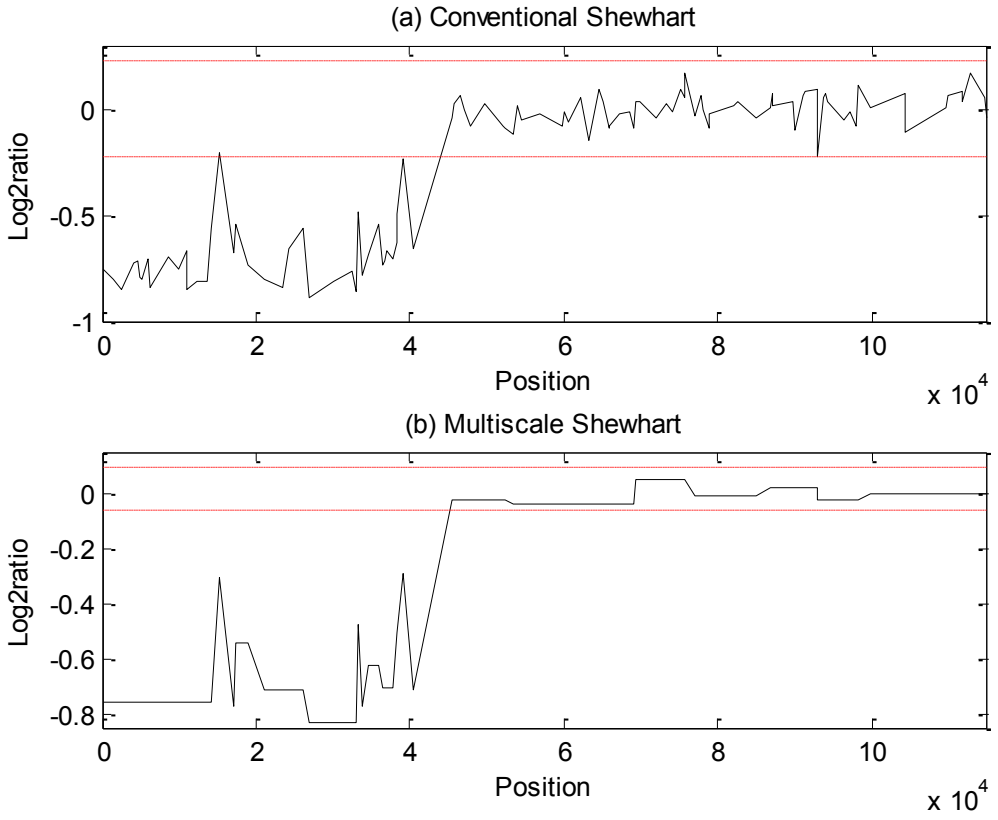


**Figure 29. Comparison in performance of the conventional and multiscale Shewhart chart for chromosome 9.**

The number of observations for chromosome 14 is also very small, with only 75 observations available. Unlike the previous example implemented to monitor the fault in Chromosome 9, the limited number of observations available to be used for fault

detection gives rise to a problem with regards to the monitoring technique. One important part of the training phase of the multiscale Shewhart chart algorithm is that there are violations of the threshold limits at certain scales. However, due to the scarcity of observations, this may lead to none of the threshold limits at any scale being violated. After reconstruction to the time domain, this would result in very narrow control limits being produced, that cannot be used for efficient fault detection. A solution to this issue would be to decrease the width of the threshold at multiple scales, which can be done by altering $L$ (the control width of the Shewhart chart).

Figure 30 displays the results that were obtained, for both the conventional Shewhart chart (see Figure 30 (a)), and multiscale Shewhart chart (see Figure 30 (b)). As in the case of Chromosome 9, the available data for chromosome 14 can be observed to be very noisy.

As illustrated in Figure 30, the multiscale Shewhart chart is able to detect the fault better than the conventional Shewhart chart. The conventional Shewhart chart has three false alarms. Similar to chromosome 9, the fault-free and faulty regions on the multiscale Shewhart chart for chromosome 14 can be distinguished more clearly (with no missed detections or false alarms) than the conventional Shewhart chart.

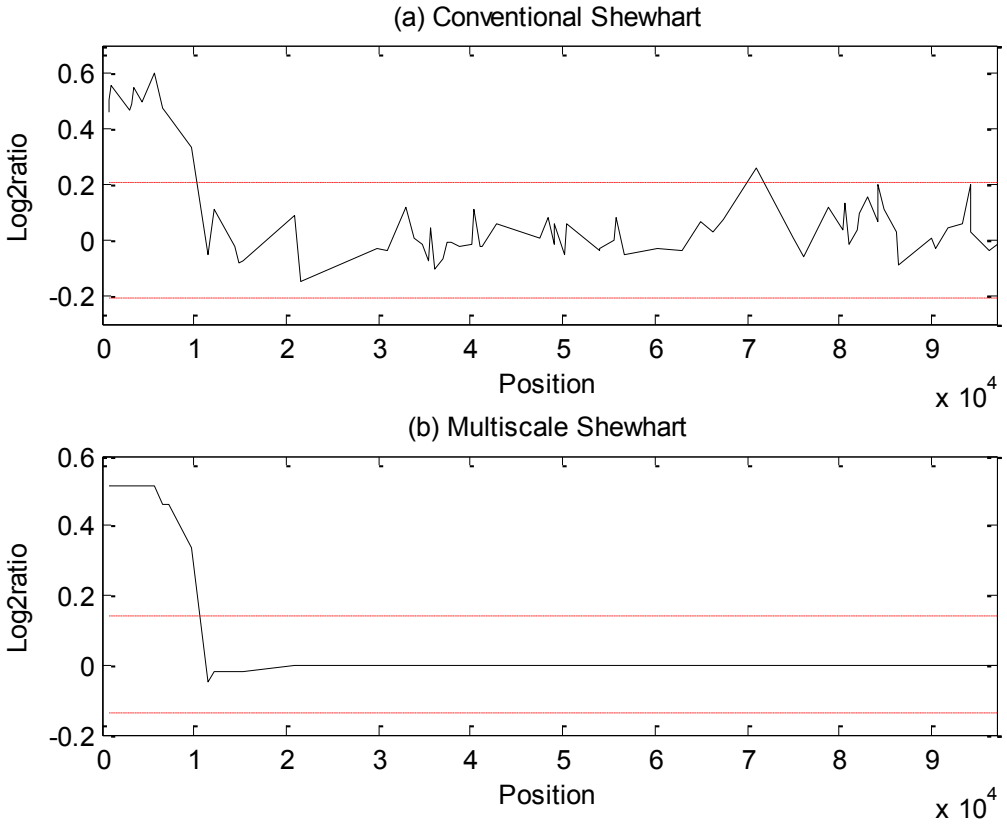


**Figure 30. Comparison in performance of the conventional and multiscale Shewhart chart for chromosome 14.**

These copy number examples clearly show that the multiscale Shewhart chart is able to perform better (with no missed detection or false alarms) than the conventional Shewhart chart for both chromosome data sets. Hence, implementing the multiscale Shewhart chart to monitor genomic copy number data could be beneficial.

# 6. CONCLUSIONS AND FUTURE DIRECTIONS

## 6.1. Concluding Remarks

Most univariate process monitoring techniques are known to operate under three main assumptions, that the process residuals being evaluated contain a moderate level of noise, are independent (uncorrelated), and follow a normal distribution (when fault-free). The performance of the conventional Shewhart chart was adversely affected when these assumptions are violated, as demonstrated by illustrative examples for all three violations.

The impact of high noise levels and autocorrelation in the residuals was severe, as they led to missed detection rates of approximately 100 % at extreme values. The impact of deviation from normality was less critical, since the missed detection rate only reached approximately 70%, for a reasonably sized fault. It is important to note that the larger the size of the fault, the easier it is for the fault to be detected.

Multiscale-wavelet based representation is a powerful data analysis tool that could help better satisfy these assumptions when applied at multiple scales (on the details signals), i.e., separate noise from features, decorrelated autocorrelated time data, and transform the data to better follow a normal distribution.

This research then focused on developing an algorithm to extend the conventional Shewhart chart using multiscale representation to deal with the impact of violation of its underlying assumptions in order to enhance its performance.

The same simulated synthetic data that was used to assess the impact of violation of assumptions of the conventional Shewhart chart was then used to comparatively assess the performance of both charts in order to illustrate the ability of the multiscale Shewhart chart to deal with the violation of its assumptions.

The multiscale Shewhart chart was able to provide a reduction in the missed detection rate of approximately 30 %, over a wide range of noise levels, with a comparable false alarm rate. The multiscale Shewhart chart was also able provide a reduction in the missed detection rate of nearly 40%, over a wide range of autocorrelation (except at high values of autoregressive coefficient), with a comparable false alarm rate. The slightly higher false alarm rate for the multiscale approach when assessing the violation of both of these assumptions can be attributed mainly to Gibbs phenomenon. For the violation of the assumption of normality, the multiscale Shewhart chart is able to provide a reduction in the missed detection rate of almost 25%, along with a lower false alarm rate as well. When compared to the violation of the other two assumptions, the multiscale Shewhart chart is able to provide better false alarm rates than the conventional Shewhart chart when dealing with the violation in assumption of normality, because the advantage brought about by multiscale representation in this case outweighs the effect (disadvantage) presented by Gibbs phenomenon.

The choice of decomposition depth was an important factor in the implementation of the developed multiscale technique, and after evaluating the missed detection and false alarm rates for a range of decomposition depths it was concluded that

the optimal decomposition depth would generally be approximately half of the maximum possible decomposition depth for a given set of observations.

The developed multiscale Shewhart chart was then applied to two real world applications, to monitor the distillate feed of a simulated distillation column, and to identify potential faulty genomic copy number data, to illustrate the advantage of using the multiscale Shewhart chart for process monitoring over the conventional one.

When used to monitor the distillate feed of the simulated distillation column for two difference fault sizes ($2.5\sigma$ and $3.5\sigma$), the multiscale approach is able to detect both faults with fewer missed detections and false alarms than then conventional Shewhart chart. It is important to note that the conventional Shewhart chart is only able to detect fairly large faults.

When used for fault detection in genomic copy number data, the multiscale Shewhart chart is able to better detect faults than the conventional Shewhart chart, with fewer missed detections and false alarms.

Although, the developed multiscale Shewhart chart does demonstrate a significantly improved performance over the conventional Shewhart, especially when the assumptions of the conventional Shewhart chart are violated, there are many directions for future research, which are suggested next.

## 6.2. Future Directions

In this work, a multiscale Shewhart chart is developed to improve monitoring of univariate processes. Numerous possible research directions can be explored to make further enhancements. Some of these directions are summarized below:

- Extend the developed technique to other univariate monitoring techniques, e.g., CUSUM, EWMA.
- Extend the developed technique to deal with multivariate processes by developing multiscale multivariate monitoring techniques.
- Extend the developed technique to handle online processes, where monitoring is required to be performed as data are collected.
- Assess the performance of the developed techniques for different types of faults (other than a step fault), such as drifts or change in the variance.

# REFERENCES


[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 27, pp. 293–311, 2003.

[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies," *Comput. Chem. Eng.*, vol. 27, pp. 313–326, 2003.

[3] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part III: Process history based methods," *Comput. Chem. Eng.*, vol. 27, pp. 327–346, 2003.

[4] A. Cinar and C. Undey, "Statistical process and controller performance monitoring: A tutorial on current methods and future directions," *Proc. Am. Control Conf.*, no. June, pp. 2625–2639, 1999.

[5] A. Cinar, A. Palazoglu, and F. Kayihan, *Chemical Process Performance Evaluation*, 1st ed. Boca Raton, FL: CRC Press, 2007.

[6] D. C. Montgomery, *Introduction to Statistical Quality Control*, 7th ed. Hoboken, NJ: John Wiley & Sons, 2013.

[7] S. W. Roberts, “Control chart tests based on geometric moving averages,” *Technometrics*, vol. 1, no. 3, pp. 239–250, Aug. 1959.

[8] J. S. Hunter, “The exponentially weighted moving average,” *J. Qual. Technol.*, vol. 18, no. 4, pp. 203–210, 1986.

[9] S. Crowder, “Design of exponentially weighted moving average schemes,” *J. Qual. Technol.*, vol. 21, no. 3, pp. 155–162, 1989.

[10] J. M. Lucas and M. S. Saccucci, “Exponentially weighted moving average control schemes: Properties and enhancements,” *Technometrics*, vol. 32, no. 1, pp. 1–12, Feb. 1990.

[11] D. M. Hawkins and D. H. Olwell, *Cumulative Sum Charts and Charting for Quality Improvement*, 1st ed. New York, NY: Springer-Verlag, 1998.

[12] E. S. Page, “Control charts with warning lines,” *Biometrika*, vol. 42, no. 1–2, pp. 243–257, 1955.

[13] B. Bakshi, “Multiscale analysis and modeling using wavelets,” *J. Chemom.*, vol. 434, no. January, pp. 415–434, 1999.

[14] H. Aradhye, B. Bakshi, R. Strauss, and J. Davis, “Multiscale SPC using wavelets: theoretical analysis and properties,” *AIChE J.*, vol. 49, no. 4, pp. 939–958, Apr. 2003.

[15] R. Ganesan, T. K. Das, and V. Venkataraman, "Wavelet-based multiscale statistical process monitoring: A literature review," *IIE Trans.*, vol. 36, no. 9, pp. 787–806, Sep. 2004.

[16] H. Guo, K. Paynabar, and J. Jin, "Multiscale monitoring of autocorrelated processes using wavelets analysis," *IIE Trans.*, vol. 44, no. 4, pp. 312–326, 2012.

[17] W. A. Shewhart, "Economic quality control of manufactured product," *Bell Syst. Tech. J.*, vol. 9, no. 2, pp. 364–389, 1930.

[18] W. A. Shewhart, *Statistical Method from the Viewpoint of Quality Control*. Lancaster, PA: Lancaster Press, Inc., 1939.

[19] E. L. Grant and R. S. Leavenworth, *Statistical Quality Control*, 6th ed. New York, NY: McGraw-Hill, 1988.

[20] W. A. Shewhart, "Quality control charts," *Bell Syst. Tech. J.*, vol. 5, no. 4, pp. 593–603, 1926.

[21] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 5th ed. Hoboken, NJ: John Wiley & Sons, Inc., 2011.

[22] I. W. Burr, *Elementary Statistical Quality Control*, 1st ed. New York, NY: Marcel Dekker, Inc., 1979.

[23] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. Hoboken, NJ: John Wiley & Sons, Inc., 2008.

[24] H. C. J. Thode, *Testing for Normality*, 1st ed. New York, NY: CRC Press, 2002.

[25] Z. Lu, M. Li, and W. Zhao, "Normality of ethernet traffic at large time scales," *Math. Probl. Eng.*, vol. 2013, no. 1, pp. 1–7, 2013.

[26] K. P. Balanda and H. L. Macgillivray, "Kurtosis: A critical review," *Am. Stat.*, vol. 42, no. 2, pp. 111–119, 1988.

[27] H. M. Park, "Univariate analysis and normality test using SAS, STATA, and SPSS," Bloomington, 2008.

[28] O. Vasicek, "A test for normality based on sample entropy," *J. R. Stat. Soc. Ser. B*, vol. 38, no. 1, pp. 54–59, 1976.

[29] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[30] P. J. Huber, "Projection pursuit," *Ann. Stat.*, vol. 13, no. 2, pp. 435–475, 1985.

[31] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, Jun. 2000.

[32] S. Park, "A goodness-of-fit test for normality based on the sample entropy of order statistics," *Stat. Probab. Lett.*, vol. 44, pp. 359–363, 1999.

[33] D. Ramirez, J. Via, I. Santamaria, and P. Crespo, “Entropy and Kullback-Leibler divergence estimation based on Szego’s theorem,” in *17th European Signal Processing Conference (EUSIPCO 2009)*, 2009, pp. 2470–2474.

[34] E. Seier, “Comparison of tests for univariate normality tests for normality,” *Interstat*, no. 20, pp. 1–17, 2002.

[35] M. A. Stephens, “EDF statistics for goodness of fit and some comparisons,” *J. Am. Stat. Assoc.*, vol. 69, no. 347, pp. 730–737, 1974.

[36] S. S. Shapiro and M. B. Wilk, “An analysis of variance test for normality (complete samples),” *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[37] A. E. Sarhan and B. G. Greenberg, “Estimation of location and scale parameters by order statistics from singly and doubly censored samples,” *Ann. Stat.*, vol. 27, no. 2, pp. 427–451, 1956.

[38] Royal Statistical Society. Series C (Applied Statistics), “The W test for normality,” *Royston, J. P.*, vol. 31, no. 2, pp. 176–180, 1982.

[39] J. P. Royston, “An extension of Shapiro and Wilk’s W test for normality to large samples,” *J. R. Stat. Soc. Ser. C-Applied Stat.*, vol. 31, no. 2, pp. 115–124, 1982.

[40] P. Royston, “Approximating the Shapiro-Wilk W-test for non-normality,” *Stat. Comput.*, vol. 2, pp. 117–119, 1992.

[41] P. Royston, “A toolkit for testing for non-normality in complete and censored samples,” *J. R. Stat. Soc. Ser. D (The Stat.*, vol. 42, no. 1, pp. 37–43, 1993.

[42] P. Royston, “Remark AS R94: A remark on algorithm AS 181: The W-test for normality,” *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 44, no. 4, pp. 547–551, 1995.

[43] N. M. Razali and Y. B. Wah, “Power comparisons of Shapiro-Wilk, Kolmogorov, Lilliefors and Anderson-Darling tests,” *J. Stat. Model. Anal.*, vol. 2, no. 1, pp. 21–33, 2011.

[44] M. N. Nounou and H. N. Nounou, “Reduced noise effect in nonlinear model estimation using multiscale representation,” *Model. Simul. Eng.*, vol. 2010, pp. 1–8, 2010.

[45] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. II, no. 7, pp. 674–693, 1989.

[46] B. Bakshi, “Multiscale PCA with application to multivariate statistical process monitoring,” *AIChE J.*, vol. 44, no. 7, pp. 1596–1610, Jul. 1998.

[47] J. Davis, M. Piovoso, K. A. Hoo, and B. R. Bakshi, “Process data analysis and interpretation,” *Adv. Chem. Eng.*, vol. 25, pp. 1–103, 1999.

[48] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Commun. pure Appl. Math.*, vol. XLI, pp. 909–996, 1988.

[49] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Library of Congress Cataloging-in-Publication Data, 1992.

[50] S. Zhou, B. Sun, and J. Shi, “An SPC monitoring system for cycle-based waveform signals using Haar transform,” *Autom. Sci. Eng. IEEE Trans.*, vol. 3, no. 1, pp. 60–72, Jan. 2006.

[51] A. Haar, “Zur theorie der orthogonalen funktionensysteme,” *Math. Ann.*, vol. 69, no. 3, pp. 331–371, Sep. 1910.

[52] M. N. Nounou, H. N. Nounou, N. Meskin, a Datta, and E. R. Dougherty, “Multiscale denoising of biological data: a comparative analysis.,” *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 9, no. 5, pp. 1539–44, 2012.

[53] K. A. Kosanovich and M. J. Piovoso, “PCA of wavelet transformed process data for monitoring,” *Intell. Data Anal.*, vol. 1, no. 1–4, pp. 85–99, 1997.

[54] M. N. Nounou and B. R. Bakshi, “Multiscale methods for denoising and compression,” in *Wavelets in Chemistry*, B.Walczak, Ed. Elsevier Science B.V., 2000, pp. 119–150.

[55] M. Z. Sheriff, F. Harrou, and M. Nounou, “Univariate process monitoring using multiscale Shewhart charts,” in *2014 International Conference on Control, Decision and Information Technologies (CoDIT)*, 2014, pp. 435–440.

[56] E. J. Candès and F. Guo, “New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction,” *Signal Processing*, vol. 82, no. 11, pp. 1519–1543, 2002.

[57] M. Li and S. Ghosal, “Fast translation invariant multiscale image denoising fast translation invariant multiscale image denoising,” Raleigh, NC.

[58] M. Madakyaru, M. Nounou, and H. Nounou, “Linear inferential modeling: theoretical perspectives, extensions, and comparative analysis,” *Intell. Control Autom.*, vol. 3, no. November, pp. 376–389, 2012.

[59] A. K. Alqallafand and A. H. Tewfik, “DNA copy number detection and sigma filter,” in *IEEE International Workshop Genomic Signal Processing and Statistics*, 2007, pp. 1–4.

[60] A. M. Snijders, N. Nowak, R. Segraves, S. Blackwood, N. Brown, J. Conroy, G. Hamilton, A. K. Hindle, B. Huey, K. Kimura, S. Law, K. Myambo, J. Palmer, B. Ylstra, J. P. Yue, J. W. Gray, A. N. Jain, D. Pinkel, and D. G. Albertson, “Assembly of microarrays for genome-wide measurement of DNA copy number.,” *Nat. Genet.*, vol. 29, no. November, pp. 263–264, 2001.

[61] H. T. Wu, I. Hajirasouliha, and B. J. Raphael, “Detecting independent and recurrent copy number aberrations using interval graphs,” *Bioinformatics*, vol. 30, pp. 195–203, 2014.