

**ENHANCED MONITORING USING MULTISCALE
EXPONENTIALLY WEIGHTED MOVING AVERAGE
CONTROL CHARTS**

A Thesis

by

MD. ALAMGIR MOJIBUL HAQUE

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
Co-chair of Committee,	Hazem N. Nounou
Committee Members,	Luc Vechot
Head of Department,	M. Nazmul Karim

August 2016

Major Subject: Chemical Engineering

Copyright 2016 Md. Alamgir Mojibul Haque

ABSTRACT

The exponentially weighted moving average (EWMA) method is a widely used univariate process monitoring technique. This conventional EWMA technique is normally designed to optimize the out of control average run length (ARL_1) specific to a fixed in control average run length (ARL_0). This design procedure of EWMA technique is based on some assumptions – the evaluated process residuals are Gaussian, independent and contain moderate level of noise. Violation of these assumptions may adversely affect its fault detection abilities. Wavelet based multiscale representation of data is a powerful data analysis tool and has inherent properties that can help deal with these violations of assumptions, which thus improve the performance of EWMA through satisfying its assumptions.

The main purpose of this work is to develop a multiscale EWMA technique with improved performance over the conventional technique and establish a design procedure for this method to optimize its parameters by minimizing the out of control average run length for different fault sizes and using a specified in control average run length assuming that the residuals are contaminated with zero mean Gaussian noise.

Through several comparative studies using Monte Carlo simulations, it has been shown that the multiscale EWMA technique provides a better performance over the conventional method. Multiscale EWMA is shown to provide smaller ARL_1 and missed detection rate with a slightly higher false alarm rate compared to the conventional EWMA technique not only when both the techniques are designed to perform optimally but also

when data violate the assumptions of the EWMA chart. The advantages of the multiscale EWMA method over the conventional method are also illustrated through their application to monitor a simulated distillation column.

DEDICATION

I dedicate this thesis to my parents, Md. Samsul Haque and Nurjahan Begum, for their inspiration and unconditional love. I would also like to dedicate this work to all my teachers who taught to me to understand the true meaning of being educated.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my committee chair, Dr. Mohamed Nounou, for his insightful ideas and valuable guidance which had helped me overcoming the difficult phases of my Master's thesis at Texas A&M University. Without his persistent help and continuous inspiration, this dissertation work would not have been possible.

I would also like to thank my committee co-chair Dr. Hazem Nounou and committee member Dr. Luc Vechot for their important advice and feedback through the course of my research.

I am also grateful to Dr. Majdi Mansouri and Md. Ziyen Sheriff for helping me in completing the thesis by providing various resources. In addition, I am deeply obliged to Texas A&M University at Qatar for providing me the opportunity to do research.

And last but not the least, I acknowledge the motivation and inspiration given to me by my parents.

NOMENCLATURE

SPM	Statistical Process Monitoring
CUSUM	Cumulative Sum
MA	Moving Average
EWMA	Exponentially Weighted Moving Average
PLS	Partial Least Squares
PCA	Principal Component Analysis
CVA	Canonical Variate State Space
SPC	Statistical Process Control
UCL	Upper Control Limit
LCL	Lower Control Limit
ARL_1	Out of Control Average Run Length
ARL_0	In Control Average Run Length
AR	Autoregressive
SW	Shapiro-Wilk
ACF	Autocorrelation Function

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
NOMENCLATURE.....	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES.....	ix
LIST OF TABLES	xiii
1. INTRODUCTION.....	1
1.1. Literature review	4
1.1.1. Conventional univariate monitoring techniques.....	5
1.1.2. Indicators for monitoring process performance	11
1.2. Research objectives	14
2. MONITORING USING EWMA CHARTS	16
2.1. Design procedure of the conventional EWMA technique.....	16
2.2. Assessing the performance of the EWMA chart under violation of assumptions	19
2.2.1. Assessing the impact of high noise levels in the data on the performance of the EWMA chart.....	19
2.2.2. Assessing the impact of autocorrelation in the data on the performance of EWMA chart	23
2.2.3. Assessing the impact of deviation from normality in the data on the performance of the EWMA chart.....	27
3. WAVELET BASED MULTISCALE REPRESENTATION	33
3.1. Introduction to wavelet based multiscale representation.....	33
3.2. Advantages of multiscale representation of data	35
3.2.1. Noise feature separation	35
3.2.2. Decorrelation of autocorrelated data	36
3.2.3. Data are closer to normal at multiple scales	38

	Page
4. WAVELET BASED MULTISCALE EWMA CHART.....	41
4.1. Process monitoring using Multiscale EWMA Chart.....	41
4.2. Design procedure of optimizing the parameters of the multiscale EWMA fault detection technique	43
4.2.1. Optimizing the multiscale EWMA parameters	44
4.2.2. Design steps for the multiscale EWMA technique.....	49
5. PERFORMANCE COMPARISON BETWEEN THE CONVENTIONAL AND MULTISCALE EWMA CHARTS	51
5.1. Comparison between the performance of the conventional and multiscale EWMA techniques under no violations of the EWMA assumptions	51
5.2. Comparison between the performances of the conventional and multiscale EWMA techniques under violations of the EWMA assumptions	55
5.2.1. Comparison of performance using data with different level of noise	56
5.2.2. Comparison of performance using autocorrelated data.....	58
5.2.3. Comparison of performance using non-Gaussian (chi-square) data	61
6. APPLICATION OF THE MULTISCALE EWMA CHART	65
7. CONCLUSIONS AND FUTURE DIRECTIONS	69
7.1. Concluding remarks.....	69
7.2. Future directions	72
REFERENCES	74

LIST OF FIGURES

	Page
Figure 1: A schematic representation of the weightings used to compute the detection statistics used in various univariate SPC charts.	7
Figure 2: Schematic representation of false alarm and missed detection.	12
Figure 3: Schematic representation of average run length.....	13
Figure 4: Trade-off between false alarm rate and missed detection rate.	14
Figure 5: Combination of λ and L for ARL_0 500.....	17
Figure 6: Optimal λ for different fault size for ARL_0 500.....	18
Figure 7: Impact of noise level on ARL_1 values for the conventional EWMA chart. ...	20
Figure 8: Impact of noise level on the false alarm and missed detection rates of the conventional EWMA chart.	21
Figure 9: EWMA statistics for different noise levels.	22
Figure 10: Impact of autocorrelation on ARL_1 for the conventional EWMA chart.	25
Figure 11: Impact of autocorrelation on the false alarm and missed detection rates for the conventional EWMA chart.	25
Figure 12: Impact of autocorrelation on the performance of EWMA chart for an autoregressive coefficient value of 0.3.	26
Figure 13: Impact of autocorrelation on the performance of EWMA chart for an autoregressive coefficient value of 0.9.	26
Figure 14: Impact of deviation from normality on ARL_1 for the conventional EWMA chart.	30
Figure 15: Impact of deviation from normality on the false alarm and missed detection rates for the conventional EWMA chart.	30
Figure 16: Impact of deviation from normality on the performance of EWMA chart, $SW = 0.6953$	31
Figure 17: Impact of deviation from normality on the performance of EWMA chart, $SW\ stat = 0.9881$	32

	Page
Figure 18: Schematic diagram for multiscale representation of data.....	35
Figure 19: Decorrelation of autocorrelated data at multiple scales.	37
Figure 20: Distribution of chi-squared data at multiple scales.	39
Figure 21: A schematic diagram of the multiscale EWMA fault detection algorithm.	42
Figure 22: Combination of λ and L for in control average run length 500.	45
Figure 23: 3-D plot of the optimization of parameters for fault size 2.5σ for multiscale EWMA chart for ARL_0 value of 500.	46
Figure 24: Optimal λ 's for the multiscale EWMA chart for different fault sizes for $ARL_0 = 500$	46
Figure 25: Effect of decomposition depth on the ARL_1 of the multiscale EWMA chart.	48
Figure 26: Effect of decomposition depth on false alarm and missed detection rates of the multiscale EWMA chart.	49
Figure 27: Performance comparison between the conventional and multiscale EWMA charts in terms of ARL_1 for different fault sizes.	52
Figure 28: Performance comparison between the conventional and multiscale EWMA charts in terms of false alarm rate for different fault sizes.	53
Figure 29: Performance comparison between the conventional and multiscale EWMA charts in terms of missed detection rate for different fault sizes.	54
Figure 30: Trade-off between false alarm and missed detection rate for the multiscale EWMA chart.	55
Figure 31: Performance comparison between the multiscale and conventional EWMA technique in terms of ARL_1 for different level of noise in the data.	57
Figure 32: Performance comparison between the multiscale and conventional EWMA techniques in terms of false alarm and missed detection rate for different level of noise in the data.	57
Figure 33: Detection statistics for the conventional and multiscale methods in the case where the noise standard deviation equals 1.5.	58

Figure 34: Performance comparison between the multiscale and conventional EWMA technique in terms of ARL_1 for different level of autocorrelation in the data.	59
Figure 35: Performance comparison between the multiscale and conventional EWMA technique in terms of false alarm and missed detection rates for different level of autocorrelation in the data.	60
Figure 36: Detection statistic for the conventional and multiscale methods in the case where the autoregressive coefficient equal to 0.5.	61
Figure 37: Performance comparison between the multiscale and conventional EWMA technique in terms of ARL_1 for non-Gaussian data (Chi-square).....	62
Figure 38: Performance comparison between the multiscale and conventional EWMA technique in terms of false alarm and missed detection rates for non-Gaussian data (Chi-square).....	63
Figure 39: Detection statistics for the conventional and multiscale methods in the case where the Shapiro Wilk equals 0.69.	64
Figure 40: Comparing the performances of the conventional and multiscale EWMA chart for a step fault of magnitude $\pm 2\sigma$ in the residuals of simulated distillation column data.	67
Figure 41: Comparing the performances of the conventional and multiscale EWMA chart for a step fault of magnitude $\pm\sigma$ in the residuals of simulated distillation column data.	68

LIST OF TABLES

	Page
Table 1: Comparison of optimum values of L and λ with those obtained from literature.....	19

1. INTRODUCTION

Process monitoring is essential for proper operation of various engineering systems. The goal of statistical process monitoring (SPM) is to detect the occurrence of faults (fault detection) and the nature of operational changes that cause a process to deviate from its desired target (fault diagnosis). Thus, process monitoring involves two main tasks: fault detection and fault diagnosis. There are various types of fault detection techniques, which will be discussed later. However, fault detection needs to be followed by fault diagnosis that aims at locating the root cause of the process change and enables the process operators to take necessary actions to correct the situation (process recovery), thereby returning the process back to its desired operation. In this way, the product quality can be maintained and safe operation can be assured. In this work, improving the task of fault detection is addressed.

The term fault here is generally defined as a shift from the target value of a variable or a calculated parameter associated with a process [1]. Fault detection methods can be classified into three categories [2]:

- I. Model-based methods
- II. Data based methods
- III. Knowledge based methods

In model based approaches, measured data of a process variable are compared with a model. This model is obtained from basic understanding about the process and is expressed in terms of mathematical relationships between the process inputs and outputs

[3][4]. Diagnostic observers, parity relations, Kalman filters and parameter estimation techniques are some of the frequently used model based approaches. The effectiveness of the model based methods depends on the accuracy of the process model. But in practice, it can be very difficult task to derive an accurate model, especially if the process involves a large number of inputs and outputs.

Knowledge based methods provide an alternative approach to the model based approach in the case of complex processes with incomplete knowledge or when analytical models are not available. Examples of knowledge based methods include causal analysis and expert systems [2][5][6].

Data based methods, on the other hand, rely on the availability of process data, from which process information can be extracted and used for fault detection and diagnosis [7][8][9]. In this category of methods, historical process data are collected by measuring key process variables under fault free conditions which are then used to construct an empirical model. This empirical model is later used to find out the residuals, which quantify the difference between the observed value of a variable and the expected value of that variable predicted by the process model. These residuals are then evaluated to monitor the process.

Several data based techniques can be found in the literature. These techniques are divided into two different classes depending on the number of variables that are being monitored: univariate and multivariate techniques[10]. Univariate process monitoring techniques are used to monitor a single variable, while multivariate techniques are used to monitor multiple process variables. The univariate techniques include the Shewhart,

cumulative sum (CUSUM) chart, moving average (MA) and the exponentially weighted moving average (EWMA) chart. Multivariate monitoring techniques, on the other hand, include partial least square (PLS) [11], principal component analysis (PCA) [12], canonical variate state space (CVA) [13], and others. The scope of this work is limited to the univariate SPM techniques.

Among the univariate techniques, Shewhart chart is one of the simplest charts. It evaluates the raw residuals, i.e., it doesn't use any filter to process the residuals. The Shewhart chart has been shown capable of detecting large faults or large shifts in the mean of a process variable. This is because the Shewhart chart only considers the most recent data sample in fault detection, which makes it not very sensitive to small changes in the data. Other univariate techniques apply linear filters on the residuals to improve their sensitivities to small shifts, such as the CUSUM and EWMA techniques. Another advantage of these filters is that they reduce the noise content in the data. However, it is known that linear filters are not very effective in removing noise from real data because they operate at a single scale, i.e., they work at a fixed scale or frequency and discard all features in the data that are above a certain frequency level [14][15]. In practice, however, process data are multiscale in nature due to the changes that occur in the process at different times and different frequencies. So the mismatch between the nature of the measured data and the nature of the linear filters makes the traditional statistical process control (SPC) methods inappropriate to deal with practical data. Non-linear filters like wavelet based multiscale filtering have shown much promise in dealing with real data [15][16] [17][18].

Another limitation associated with the existing univariate fault detection methods is that these methods assume the measured data are independent and normally distributed (Gaussian). Real process data, however, don't usually follow these assumptions, which deteriorates the performance of these conventional monitoring methods. Multiscale wavelet based method has showed inherent ability to deal with those assumptions [15], which helps improve the effectiveness of process monitoring, especially for correlated data [19][20]. Therefore, the objective of this work is to utilize the advantages of multiscale wavelet based representation to improve the monitoring performance of the EWMA control chart, especially for the detection of faults with small magnitudes. A design procedure for optimizing the EWMA parameters (which is based on minimizing the out of control average run length) will be developed. Also, the performance of the developed multiscale EWMA technique will be assessed and compared with its conventional time domain counterpart in the cases where the data are autocorrelated and non-Gaussian. The different techniques will be compared using three performance indices, which include the missed detection rate, false alarm rate, and the average run length.

In the next sections, a review on some of the popular univariate control charts will be presented, followed by a description of the main research objectives of this work.

1.1. Literature review

In this section, descriptions of the commonly used univariate control charts are presented along with, their advantages and limitations.

1.1.1. Conventional univariate monitoring techniques

1.1.1.1. Shewhart chart

Walter Shewhart first developed Shewhart chart in the 1920s [21]. It was intended to monitor quality of a manufacturing process at different stages. It is widely used in the field of statistical quality control because of its computational simplicity which makes it easy to implement. Only three features are needed to design a Shewhart chart – a center line (C) or a mean value, the upper control limit (UCL) and the lower control limit (LCL) [22].

The shewhart chart can be of different types depending on the parameters that are being monitored. To monitor the average level of a process variable, the mean chart (\bar{x}) is used. On the other hand, the Range chart or standard deviation (S) chart is used to monitor the sample process variation or spread. Shewhart chart is usually used to monitor the sample mean. In some occasions, the mean chart is coupled with the range chart or S chart when robustness against the variability in the observations is required. The simultaneous use of both the mean and range chart ensures the capture of almost all important features hidden in the data as only one chart may not be able to do that.

The sample mean of a particular process variable, x can be computed by the following equations [22]:

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

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

where, n and k represent the subgroup size and the number of subgroups respectively. The number of subgroups usually represents the number of sensors that measure a particular variable. So, the subgroup size k is equal to 1 when only one sensor measures the variable and then \bar{x}_i equals to \bar{x} . When a single variable is being monitored by multiple sensors, then sub-grouping is required.

The upper and lower control limits are defined as follows:

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

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

where, σ represents the standard deviation of the priori data when there are no faults in the data and c is a constant which is computed using a nomogram. When a process sample observation falls outside the control limits, then this indicates that the process mean has shifted from the target value, \bar{x} .

So, it is evident that the control limits have to be chosen carefully, which depend on the L_n value. Generally a value of 3σ is used for the parameter L_n [23] because finding accurate values of c from nomogram for different processes is very difficult. Similar equations have been used to compute the center line, upper and lower control limits for the R and S chart [24].

As indicated earlier, the Shewhart chart can't detect relatively small faults, so it is not advisable to use it in fault detection when small deviations from the process mean are expected. In fact, the Shewhart chart is only able to detect faults larger than three times the standard deviation of the original signal [25], which is a major drawback of the Shewhart chart. This inability to detect small mean shifts is due to the short memory of

the Shewhart chart as it only considers the current process measurement. Other control charts have longer memory as they average past samples to compute their detection statistics. The weighing used to compute the detection statistic for various charts are shown in Figure 1.

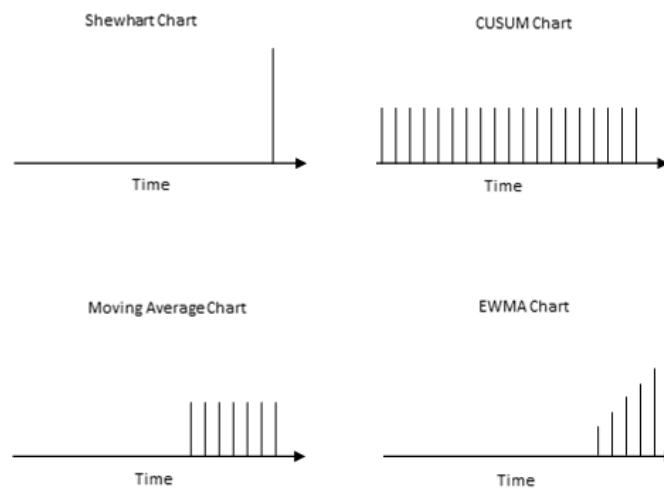


Figure 1: A schematic representation of the weightings used to compute the detection statistics used in various univariate SPC charts.

1.1.1.2. CUSUM chart

An effective alternative to the Shewhart chart is the cumulative sum (CUSUM) chart. It was first introduced by Page in 1954 [26]. The design of the conventional CUSUM chart involves computing the CUSUM statistic which is defined by the following equation [27]:

$$S_i = \sum_{j=1}^i (x_j - \mu_0)$$

This quantity S_i is plotted against the sample number i where μ_0 is the mean value of the process variable under fault free conditions. The CUSUM statistic can also be computed recursively as follows:

$$S_i = (x_i - \mu_0) + S_{i-1}$$

When several observations are available at each time sample, then the observation x_j is replaced by average of all the observations at a particular sample time, \bar{x}_j .

To detect a particular fault in the process, the one-sided CUSUM charts are used by plotting the following statistic [27]:

$$S_i = \max[0, \bar{x}_i - (\mu_0 + K)]$$

where, K is the reference value to detect a shift in the mean of size Δ . K is defined as follows:

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

When S_i exceeds a decision interval H , then it is assumed that mean of the process variable has shifted from the targeted value by a margin of Δ . The value of H can be computed as:

$$H = \frac{d\Delta}{2} \text{ where } d = \left(\frac{2}{\delta^2}\right) \ln\left(\frac{1-\beta}{\alpha}\right) \text{ and } \delta = \frac{\Delta}{\sigma_{\bar{x}}}$$

where, α and β are the type I and type II error probabilities respectively and $\sigma_{\bar{x}}$ is the standard deviation of the process variable.

The most popular form of the CUSUM chart, however, is the two-sided CUSUM chart. The positive and negative CUSUM statistics are calculated as follows:

$$S_{H(i)} = \max[0, \bar{x}_i - (\mu_0 + K) + S_{H(i-1)}]$$

$$S_{L(i)} = \max[0, (\mu_0 - K) - \bar{x}_i + S_{L(i-1)}]$$

When $S_{H(i)}$ and $S_{L(i)}$ exceed the decision interval or the control limits, then it is assumed that the process is not in control. Computation of the control limits require knowledge about probability density function of the distribution of process variable, which is usually hard to obtain. So, a practical experience is important in designing a CUSUM chart. Control limits of 4σ or 5σ are suggested to provide a reasonable detection for a mean shift of around 1σ in the process data [23].

The CUSUM chart can perform better than the Shewhart chart in detecting relatively small mean shifts although it can result in more false alarms or type I errors.

1.1.1.3. EWMA chart

The exponentially weighted moving average chart was first introduced in the literature by Roberts in 1959 [28]. Since then it has been widely used as forecasting tool [29] and also as a tool for process monitoring and diagnosis [30]. The EWMA control scheme is easy to implement. The design of EWMA control scheme includes computation of the EWMA statistic and the upper and lower control limits.

The EWMA statistic can be computed as follows

$$Z_i = \lambda x_i + (1 - \lambda)Z_{i-1}, \quad 0 < \lambda \leq 1$$

where, λ is called a smoothing parameter, which changes the memory of the detection statistic.

EWMA is also known as a geometric moving average method, because the EWMA statistic can also be written as a moving average of the current and past observations as follows:

$$Z_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i Z_0$$

The above equation shows that the weights assigned to past observations decrease exponentially, giving the name of the EWMA technique. The upper and lower control limits are defined in terms of the standard deviation of the EWMA statistic and are computed as follows:

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

where, μ is the standard deviation of the observations. The last term in the bracket quickly converges to zero as the number of observation increases and thus the control limits can be computed as follows:

$$UCL, LCL = \mu_0 \pm L\sigma \sqrt{\frac{\lambda}{(2 - \lambda)}}$$

Whenever, the control statistic falls outside the range of the control limits, the process is considered to be out of control. To use the EWMA method, the choice of the smoothing parameter needs to be made carefully. Generally, a value in the range of 0.2 to 0.3 is found reasonable [27]. However, in practice, the optimum choice of the smoothing parameter depends on the size of the mean shift to be detected. For large mean shifts, large values of λ are needed, while smaller values of λ are needed to detect smaller mean shifts

more quickly [31][32]. This can be attributed to the fact that, when λ equals 1, the value of the EWMA statistic only uses the most recent observation. This makes the EWMA chart equivalent to the Shewhart chart, which is only capable of detecting large faults. On the other hand, for very small values of λ , the EWMA method becomes similar to the CUSUM, which is more capable of detecting smaller shifts.

Furthermore, all the fault detection charts described earlier rely on some assumptions, which include:

- There is a moderate level of noise in the data.
- The process data are independent, i.e., uncorrelated
- The process data follow a normal or Gaussian distribution.

Practical data, however, don't usually satisfy these assumptions. As a result, the performance of these control chart deteriorates. Several indicators are used to analyze the performance of a control chart. Those indicators are described in the next section.

1.1.2. Indicators for monitoring process performance

The most commonly used indicators for assessing the performance of process monitoring performance include the out of control average run length (ARL_1) and false alarm rate. In this work, these two indicators will be used along with the missed detection rate, which quantifies the effectiveness of detection achieved by the fault detection method.

False alarm, which is also known as a type I error, represents the case where the SPC declares the presence of a fault when in reality there isn't any fault in the process. In

other words, control chart shows an out of control signal when the process is actually in control.

Missed detection, on the other hand, means the SPC chart fails to detect a fault when actually a fault exists in the process, which is also a type II error.

Missed detection and false alarms are illustrated in the Figure 2, in which the highlighted area represents a fault region, while the all other areas are fault free. Therefore, points in the fault free region that fall outside the control limits are false alarms, while any point inside the fault region, but within the control limits, is a missed detection.

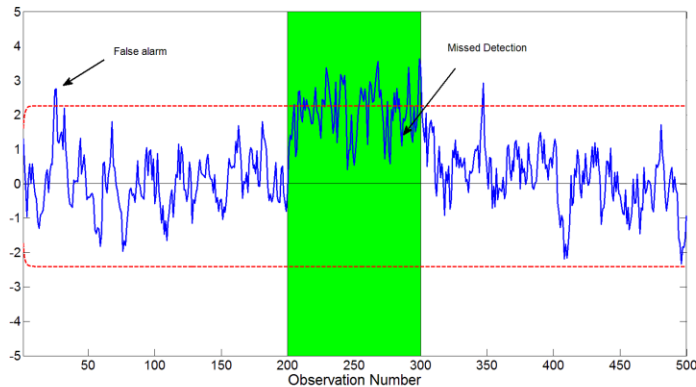


Figure 2: Schematic representation of false alarm and missed detection.

The average run length, on the other hand, is the average number of samples a fault detection method takes before it declares the presence of a fault. Average run length can be used to characterize both types of error, I and II. The in control average run length (ARL_0) is the average number of observations a control chart takes to show an out of

control signal when the process is in control. ARL_0 corresponds to a type I error. On the other hand, out of control average run length (ARL_1) represents the average number of observations that a control chart takes to declare a fault after a fault occurs, which corresponds to a type II error. These average run lengths are illustrated in figure 3.

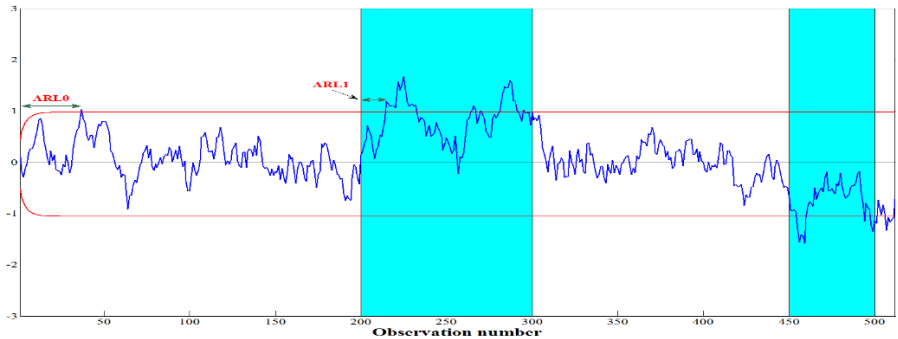


Figure 3: Schematic representation of average run length

For a good control chart performance, all three indicators (ARL_1 , missed detection rate and false alarm rate) need to be as small as possible. But in practice, however, a control chart that is designed to respond quickly to certain changes in the process mean value will become sensitive to high frequency effects or noise. As a result, false alarm rate during normal operation will increase [33]. On the other hand, if one wants to reduce the false alarm rate by expanding the control width this can eventually increase the missed detection rate and out of control average run length. So this means that there is a trade-off between the false alarm rate and missed detection rate. This is illustrated in the following

figure where control width and smoothing parameter of EWMA chart has been varied for a fixed fault size of 1σ .

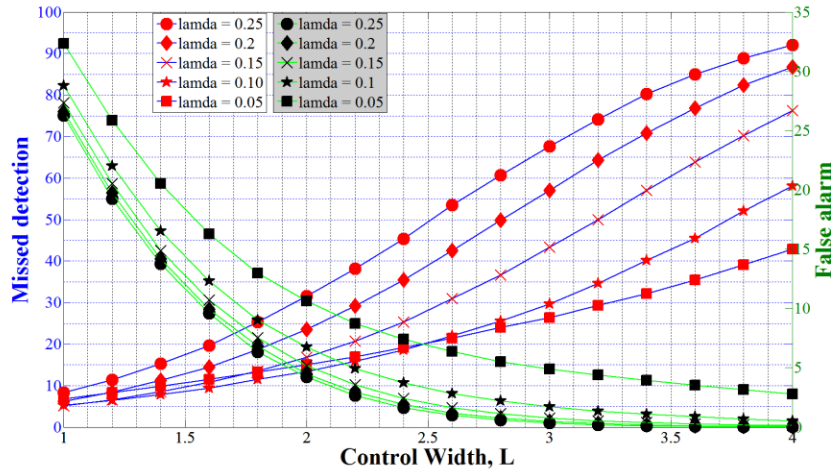


Figure 4: Trade-off between false alarm rate and missed detection rate.

Figure 4 shows that, it's not possible to decrease all the indicators at the same time, so it is very important to prioritize the indicators before designing the EWMA chart. By doing so, one can make sure that the control chart gives optimum performance in terms of a selected indicator. The selection of the indicator normally depends on the process requirement.

1.2. Research objectives

As discussed earlier, violating the assumptions of the conventional univariate monitoring techniques (such as EWMA) degrades their performances. Multiscale representation has inherent abilities to deal with those assumptions and thus can help

improve the effectiveness of these techniques. Therefore, the main objective of this work is to utilize the advantages of multiscale representation of data to improve the fault detection abilities of the EWMA control chart, especially under violation of its assumptions, i.e., when the data have large noise content, autocorrelation, or non-Gaussian distribution. Specifically, the following objectives will be sought in this work:

- Assess the performance of EWMA under the violation of its assumptions
- Develop a multiscale EWMA fault detection method that combines the advantages of multiscale representation and those of the EWMA technique
- Develop a design procedure for optimizing the parameters of the multiscale EWMA technique based on ARL_1
- Compare the performances of the multiscale EWMA and the conventional EWMA techniques using their optimum parameters
- Comparing the performances of the multiscale and conventional EWMA techniques under the violation of the main assumptions
- Provide possible directions for future research work

2. MONITORING USING EWMA CHARTS

While the Shewhart chart takes into account only the current data sample to evaluate the process performance, the CUSUM and EWMA charts consider a weighted sum of past observations. The CUSUM chart gives equal weight to all past observations, while the EWMA chart gives more importance to the more recent observations [27]. Both of the CUSUM and EWMA charts perform almost equally in detecting small mean shift but the EWMA chart is somewhat easier to set up and operate. Moreover, since the EWMA statistic is weighted average of all past and current observations, it is less sensitive to the normality assumption [23]. For these reasons, the EWMA chart has been chosen to study for this work as a model of conventional univariate technique.

2.1. Design procedure of the conventional EWMA technique

EWMA technique has been studied extensively by many researchers, and its properties and design procedures are well established [34][31][35]. The following design procedure of the conventional EWMA technique based on ARL_1 has been developed [35]:

- Choose an acceptable value of ARL_0 , in control average run length
- Specify the minimum fault size that needs to be detected as quickly as possible, and determine the value of λ which produces the lowest ARL_1 for that specific fault size.
- Find the value of the control width L , which along with the value of λ (found from previous step) provides the required ARL_0 value.

Plots that can be used to find these optimum EWMA L and λ combinations for different ARL_0 values are available in [35]. In this work, an attempt is made to reproduce those curves by simulation so that similar plots can be constructed for the multiscale EWMA technique in later stages.

To illustrate how to reproduce these plots, an ARL_0 value of 500 has been selected as an example. Then, training fault free data consisting of 8192 zero mean Gaussian observations having a unit standard deviation are used to find out different combination of λ and L values such that each combination gives an in control average run length (ARL_0) of 500. A Monte Carlo simulation of 5000 realizations is used for each combination to be sure that these combinations in fact give the specified ARL_0 . The following figure is constructed using all of these combinations.

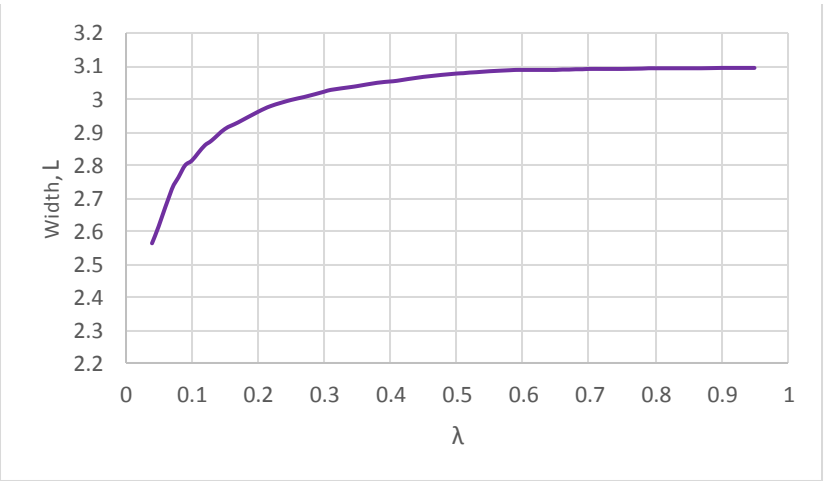


Figure 5: Combination of λ and L for ARL_0 500.

To select the optimum combination that gives the lowest ARL_1 , faulty testing data having the same length as the training data and a fault of size equal to the standard deviation of the data (i.e., 1σ) are generated. Then the EWMA chart is applied to the testing data using all the combinations of λ and L values shown in Figure 5 to see which combination gives the lowest ARL_1 value. This combination of L and λ values is the optimum for a fault size of 1σ . A Monte Carlo simulation is used for each combination to get statistically meaningful results. The same procedure is repeated for different fault sizes, which provides the results shown in Figure 6.

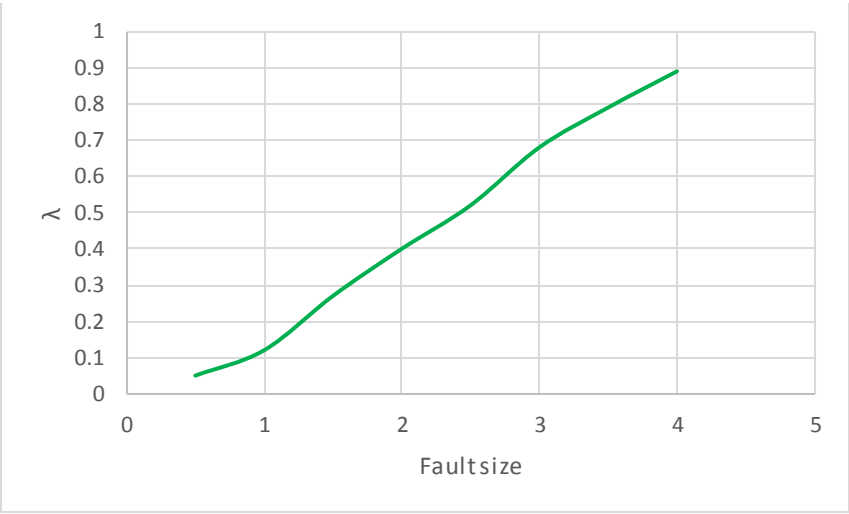


Figure 6: Optimal λ for different fault size for ARL_0 500.

Optimum values of λ and L for different fault sizes found by simulation are compared with the values obtained from Crowder[35] in Table 1 below. The values are sufficiently close which validate the simulations in this work.

Table 1: Comparison of optimum values of L and λ with those obtained from literature.

Fault size	ARL ₀	Optimum width, L		Optimum λ	
		Simulated value	Reported in Crowder [35]	Simulated value	Reported in Crowder [35]
0.5	500	2.615	2.61	0.05	0.05
1.0		2.861	2.87	0.12	0.13
1.5		3.007	3.0	0.27	0.25
2.0		3.054	3.04	0.4	0.37
2.5		3.081	3.08	0.52	0.52
3.0		3.091	3.09	0.68	0.68
3.5		3.094	3.09	0.79	0.8
4.0		3.095	3.09	0.89	0.89

Having verified the design procedure for the conventional EWMA technique based on lowest ARL₁ values for different fault sizes, EWMA method can be used to assess how it performs under violation of its assumptions. The assessment will be performed in terms of all three indicators, ARL₁, false alarm rate and missed detection rate.

2.2. Assessing the performance of the EWMA chart under violation of assumptions

2.2.1. Assessing the impact of high noise levels in the data on the performance of the EWMA chart

In this section, the impact of different measurement noise levels on the performance of EWMA technique proposed by the previous section will be assessed. For that purpose, training data (without fault) consisting of 8192 samples are generated. These data have zero mean and, unit variance and follow a Gaussian distribution. These training data are used to compute the control limits using a EWMA chart. Then, testing data having

the same length are generated and faults of magnitudes of ± 1 are introduced between samples 2000-3000 and 4500-5000. The control limits obtained from the training data are applied on the EWMA statistics computed using the testing data to detect the fault and to compute the three indicators- ARL_1 , false alarm rate and missed detection rate. A Monte Carlo simulation of 5000 realizations is performed to get statistically meaningful results. This whole simulation is then repeated for different values of noise ranging from 0.03 to 2 times standard deviation (σ) of the data.

The results of this Monte Carlo simulation, which are shown in Figures 7 and 8 show that- the ARL_1 and missed detection rate increase, while the false alarm rate remains relatively constant with a slight decrease at higher noise levels.

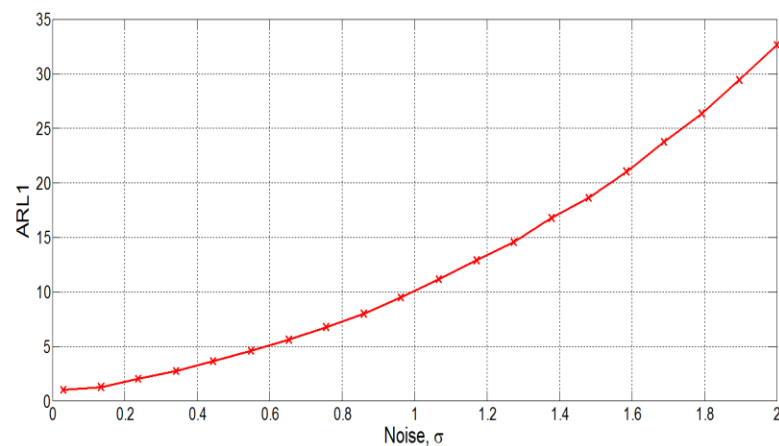


Figure 7: Impact of noise level on ARL_1 values for the conventional EWMA chart.

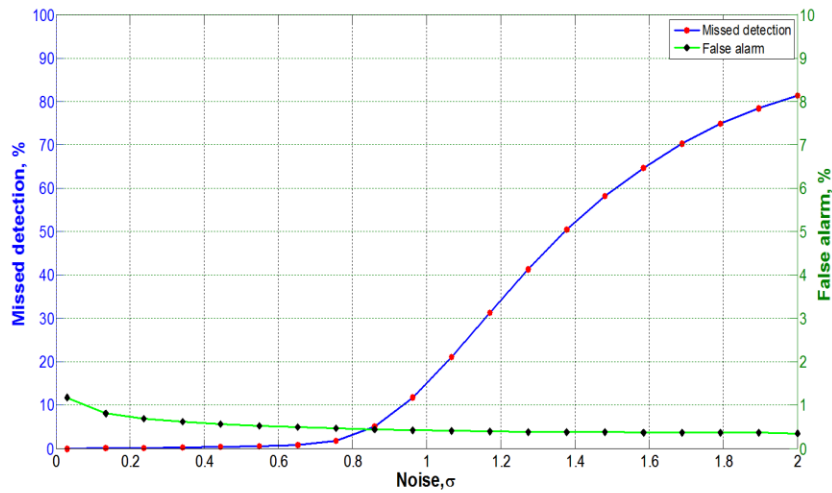


Figure 8: Impact of noise level on the false alarm and missed detection rates of the conventional EWMA chart.

To get proper understanding of what actually happens when the noise level increases, one can take a look at the EWMA control chart of an individual realization. The EWMA statistic of the testing data along with its control limits are shown for different noise levels ($\sigma = 0.5, 1, \text{ and } 1.5$) in Figure 9. These figures show that at a very high noise level (e.g., $\sigma = 1.5$), the EWMA chart fails to detect the fault most of the time. This is due to the fact the high noise level masks the fault and makes it harder to properly detect the fault, which results in higher missed detection rates and ARL_1 .

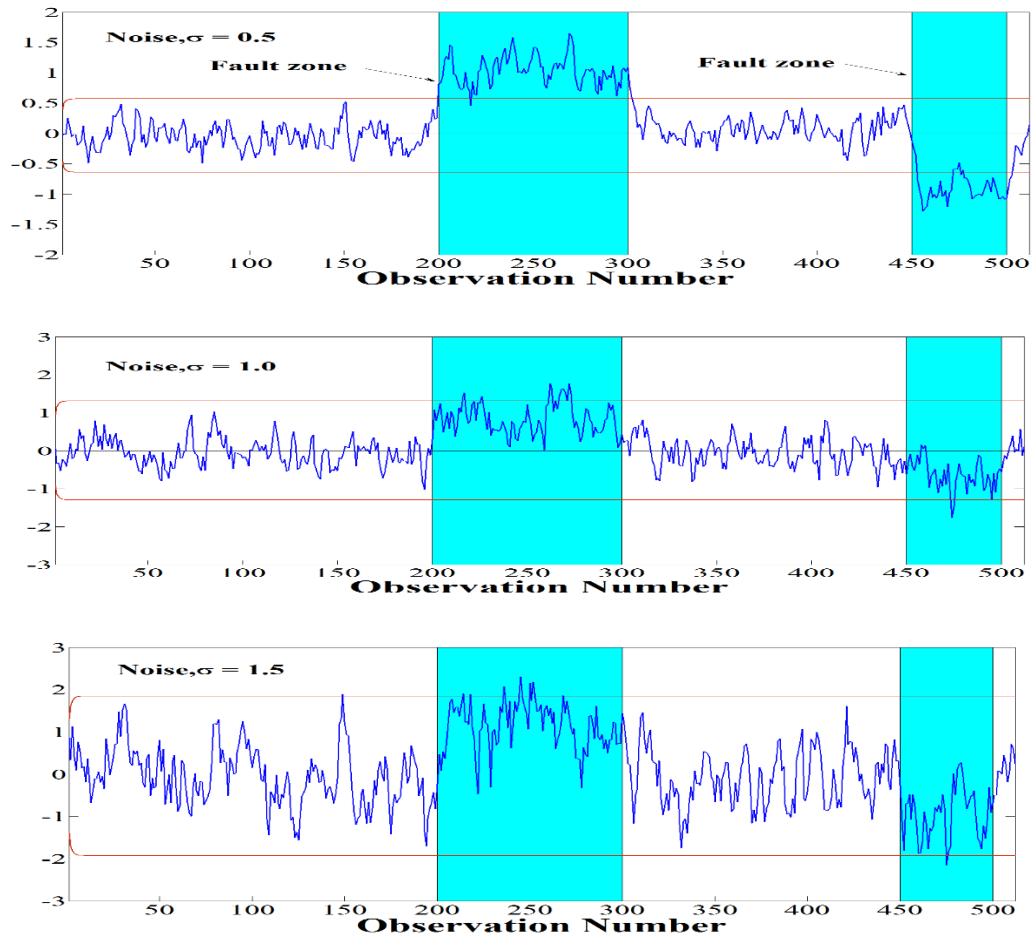


Figure 9: EWMA statistics for different noise levels.

The false alarm rate, on the other hand, remains relatively constant for different noise levels because the control chart itself adjusts the control width by widening and shrinking depending on the level of noise in the data. The initial relative high rate of false alarm for low noise level is due to the combination of two reasons – narrow control width and inertia effect. The inertia effect normally occurs for low values of λ . In this simulation low value of λ (obtained from the plot) is used because fault size is relatively low and it is known that low values of λ can detect small faults quickly. Low value of λ means it gives

low weight to the new data. As a result, when a shift occurs, the EWMA statistic takes some time to detect this sudden shift. In Figure 9, for a noise level of 0.5σ , when the process returns to the normal value from a negative shift, the EWMA chart fails to track this change instantaneously, which causes few false alarms. For high noise levels, this inertia effect gets nullified by the widened control width.

2.2.2. Assessing the impact of autocorrelation in the data on the performance of EWMA chart

Autocorrelation is the presence of correlation between a data samples and previous samples. The run length properties of traditional EWMA technique can be affected by the presence of autocorrelation in the data. For example, the ARL_0 value can be much smaller than what it is designed for when the data have positive correlation [36][37]. Since the design procedure for conventional EWMA chart assumes independent samples, its performance can degrade in the case of autocorrelated data. In this section, performance of EWMA chart is assessed in the presence of autocorrelation.

Autocorrelation in the data can be quantified using various models. A commonly used model is the autoregressive (AR) model, in which the data are represented by a linear sum of previous measurements and random noise. An AR model of order p is defined as follows [38]:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + \xi_t$$

where, ϕ_i are the autoregressive coefficients for different lagged measurements, \tilde{z}_t is the deviation from the targeted process variable μ and ξ_t is random noise which is usually assumed to be a random Gaussian variable with zero mean and unit variance. For this

work, a simple AR (1) model is used to simulate autocorrelated data, and is assumed to have the following form:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \xi_t$$

To assess the effect of autocorrelation, training and testing data, both of 8192 observations, are generated using an AR (1) model, and a fault of $\pm 1\sigma$ size is introduced on the testing data between samples 2000-3000 and samples 4500-5000. The control limits calculated using the training data, which are then applied on the testing data statistics to find out the ARL_1 , false alarm rate and missed detection rate. A Monte Carlo simulation of 5000 realizations is performed to have meaningful results. This same simulation is repeated for different autoregressive coefficients ranging from 0.1 to 1. Smaller values of coefficient mean lower autocorrelation and bigger values represent higher autocorrelation. The results of these simulations are illustrated in Figures 10 and 11.

Figure 10 shows that, out of control average run length remains almost constant at a value of 10 for a wide range of autoregressive coefficient, ϕ_1 . This value of ARL_1 is actually equal to the optimum ARL_1 value designed to detect a fault of size of 1 in presence of no autocorrelation in the data. Figure 10 also shows that ARL_1 increases at relatively very high values of autoregressive coefficients. Such high level of autocorrelation is not very common in practice, so it can be concluded that the effect of autocorrelation on ARL_1 is not significant in practice.

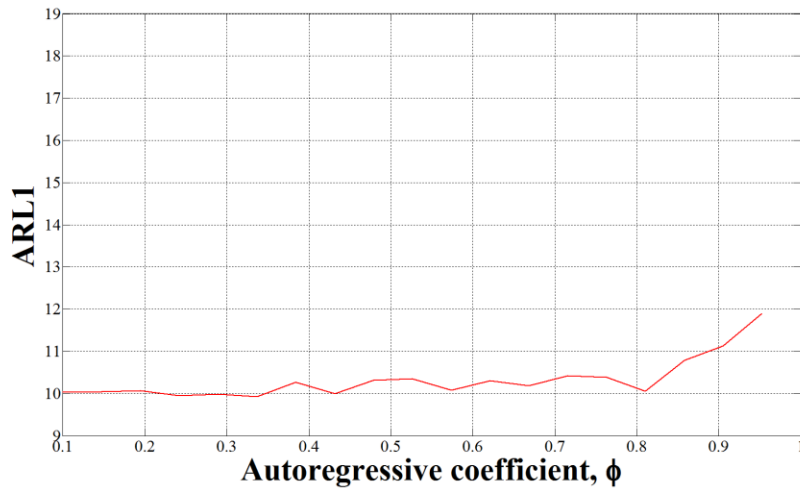


Figure 10: Impact of autocorrelation on ARL1 for the conventional EWMA chart.

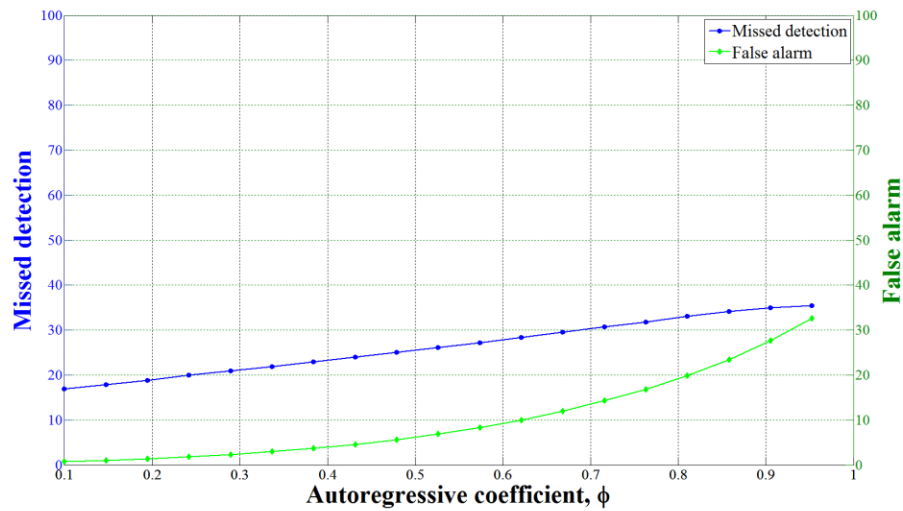


Figure 11: Impact of autocorrelation on the false alarm and missed detection rates for the conventional EWMA chart.

On the other hand, the false alarm rate and missed detection rate increase gradually for larger autoregressive coefficients as shown in Figure 11. To further illustrate the

deterioration in performance of the EWMA chart in the presence of autocorrelation in the data, the EWMA control chart of individual realizations for two different values of ϕ_1 are shown in Figures 12 and 13. These Figures clearly show that the performance of EWMA deteriorates at larger autocorrelation levels.

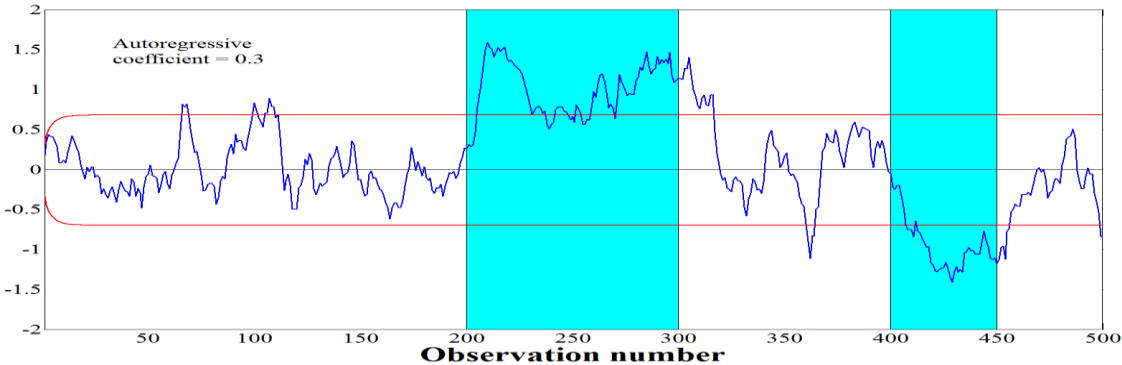


Figure 12: Impact of autocorrelation on the performance of EWMA chart for an autoregressive coefficient value of 0.3.

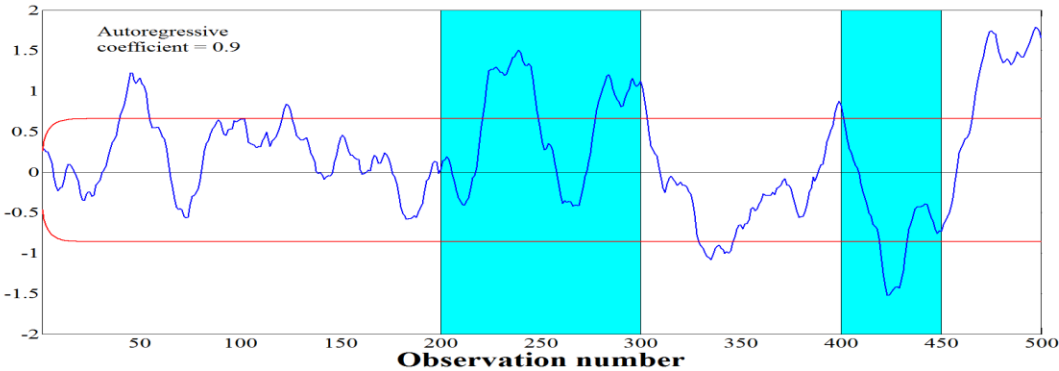


Figure 13: Impact of autocorrelation on the performance of EWMA chart for an autoregressive coefficient value of 0.9.

2.2.3. Assessing the impact of deviation from normality in the data on the performance of the EWMA chart

The conventional EWMA chart assumes that residuals follow a normal or Gaussian distribution. However, this assumptions may not always be true. In this section, the impact of deviation from normality in the data on the performance of EWMA chart is assessed.

To simulate non-Gaussian data, several distributions can be used, which include the chi-square distribution, the lognormal distribution, the gamma distribution, the Weibull distribution etc. In this work, a chi-square distribution with varying degrees of non-normality is used to simulate non-Gaussian data.

The chi-square distribution is a special case of the gamma distribution [39]. The probability density function of a gamma distributed random variable, x , is defined as follows:

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \text{ for } x > 0$$

where, $\lambda > 0$ and $r > 0$. This distribution becomes chi-square distribution when $\lambda = \frac{1}{2}$ and $r = \frac{k}{2}$. So, the probability density function of the chi-square distribution has the following form:

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

where, k represents the degrees of freedom value. Different degree of non-normality can be produced by varying this degree of freedom value. The degree of non-normality in data can be measured in various ways, which can be categorized as follows [40]:

- Graphical methods
- Moment type tests methods
- Other tests designed specifically to test for normality

Graphical methods include plots of either the raw data or the plot of probability density functions of the data. Example of raw data plots are histogram, stem and leaf plots, box plots (skeletal). On the other hand, probability plots include normal quantile plots (Q-Q), percentile plots (P-P) etc. These graphical methods visualize differences between the empirical distribution and the theoretical distribution like a normal distribution of the data and are not convenient to investigate the deviation from normality of the process data.

Moment type tests include skewness and kurtosis tests and are frequently used to quantify the degree of normality of a particular distribution of data. Skewness roughly check normality by measuring the degree of symmetry of a distribution. A normally distributed data is symmetrical. So, skewness of a normally distributed data is zero. A deviation from this zero value, either positive or negative, represents a skewed distribution with long tail to right or left side of the distribution respectively [41]. The skewness coefficients bigger than 1 or less than -1 indicate fair amount of skewness and thus deviation from normality. Kurtosis measures the peakedness of the distribution or heaviness of the tail. If a variable is normally distributed then its kurtosis value is 3. A kurtosis value of less than 3 indicates a thicker tailed and lower peaked distribution

compared to a normal distribution [42]. Though this sample moment type tests are commonly used, they are not adopted in this work to measure the deviation from normality. Because both the kurtosis and skewness are sensitive to outliers. These outliers can cause inaccurate values of kurtosis or skewness, which consist of a very little portion of the data sample and may result from measurement errors.

Besides the graphical method and moment type tests method, various other methods are available to test normality and include sample entropy, Kullback-Leiber, relative entropy and similar metrics [43][44][45]. In this work, the Shapiro-Wilk test is used to measure the degree of non-normality which also falls in this category and is a powerful univariate normality test. The Shapiro-Wilk metric value has a range between 0 and 1. Values closer to 1 means the data follow a distribution closer to normal, while values closer to 0 represents distributions that are further away from normality.

In this simulation, which is intended to assess the effect of deviation from normality on the performance of EWMA, training and testing data sets consisting of 8192 observations each, are generated using a chi-square distribution, and faults having magnitudes of $\pm 1\sigma$ are introduced in the testing data between samples 2000-3000 and samples 4500-5000. The control limits are computed using the training fault free data, and are then used to evaluate the faulty testing data using the three indicators - ARL_1 , false alarm rate and missed detection rate. This simulation is performed for different degrees of freedom values which correspond to different Shapiro-Wilk statistics, i.e., different degrees of non-normality. A Monte Carlo simulation of 5000 realizations is performed to obtain statistically meaningful results, which are shown in Figures 14 and 15.

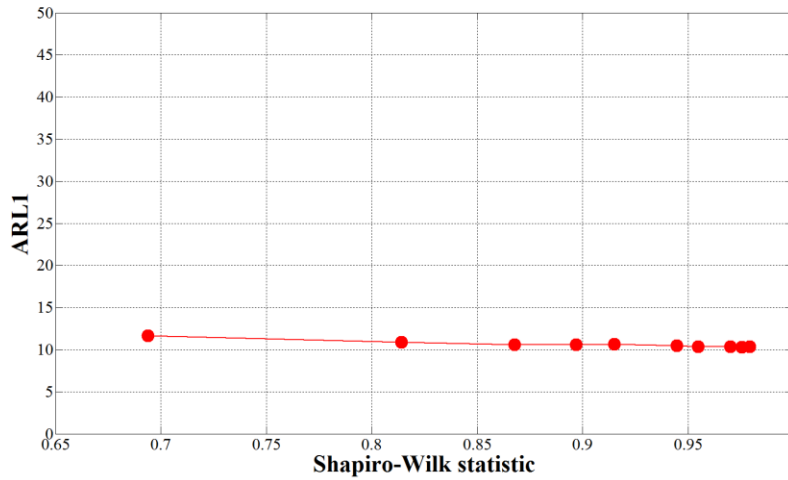


Figure 14: Impact of deviation from normality on ARL_1 for the conventional EWMA chart.

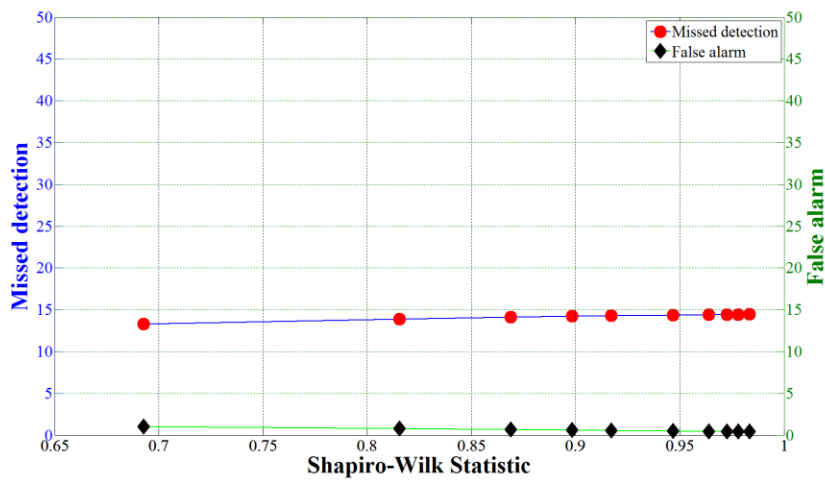


Figure 15: Impact of deviation from normality on the false alarm and missed detection rates for the conventional EWMA chart.

It is seen from Figures 14 and 15 that all three indicators, ARL_1 , false alarm rate and missed detection rate, do not change significantly with the Shapiro-Wilk statistic. These indicators remain almost constant even though the data deviate from normality.

These simulation results validate the fact that the EWMA technique is insensitive to non-normality in the data [27].

To explain these results, individual EWMA realizations for two different values of Shapiro-Wilk statistics are investigated. Figures 16 and 17 show the EWMA control charts along with their histogram for Shapiro-Wilk statistics 0.6953 and 0.9881 respectively.

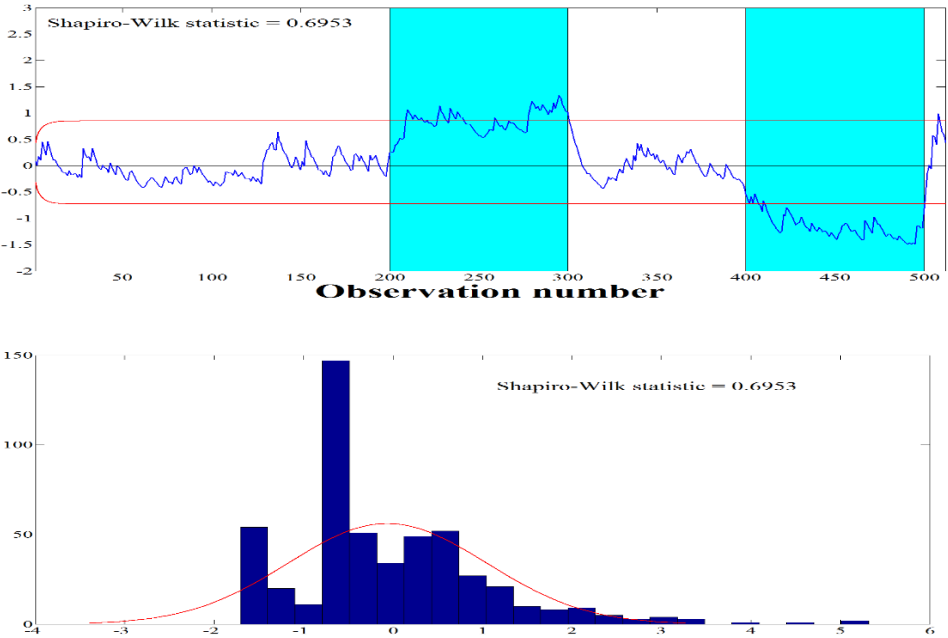


Figure 16: Impact of deviation from normality on the performance of EWMA chart, SW = 0.6953.

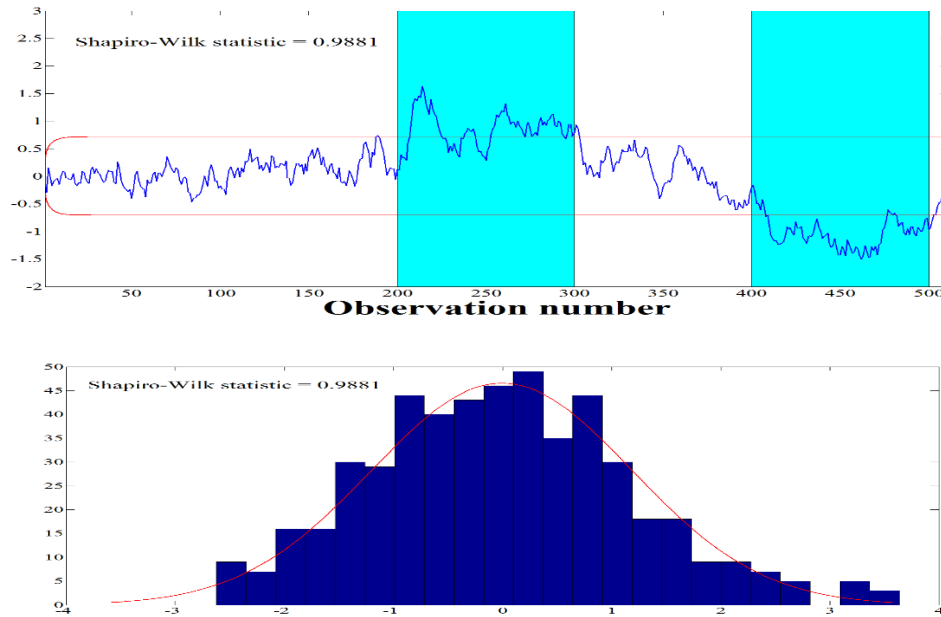


Figure 17: Impact of deviation from normality on the performance of EWMA chart, SW stat = 0.9881.

From Figure 16 and 17, it is clear that there is not much difference in the plots of EWMA statistics for the two cases even though the histograms show that the two data sets have different degrees of non-normality.

In summary, it can be seen from the simulations performed to assess the impact of violating the normality, independence, and noise content assumptions made in the development of the EWMA fault detection method that these violations in the independence and noise level assumptions can seriously degrade its performance. In this work, wavelet based multiscale representation of data will be used to deal with these assumptions and thus to help improve the performance of EWMA. More information about multiscale representation and its advantages for process monitoring will be presented in the next chapter.

3. WAVELET BASED MULTISCALE REPRESENTATION

In this chapter, wavelet based multiscale representation of data will be introduced, followed by a discussion on its advantages for process monitoring that can help improve the performance of EWMA.

3.1. Introduction to wavelet based multiscale representation

Real process data or signals are normally a combination of various features such as measurement noise, process disturbances, process dynamics, and faults etc. These features usually contain varying contributions over time and frequency. For instance, a step fault in a signal is localized in time domain, but spans a wide range in the frequency domain, while correlated noise spans a wide range in the time domain but a small range in the frequency domain. So, effective feature extraction from such data requires representing the data at both time and frequency, which can be achieved by decomposing the data at multiple scales using wavelets.

Multiscale decomposition algorithm was first developed by Mallat in 1989 [46], in which a signal is represented at multiple resolutions by expressing the data as a weighted sum of the orthonormal basis functions, called wavelets and scaling functions, that have the following form [47][48]:

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

where, s and u represent the dilation and translation parameters respectively and θ is the mother wavelet. A number of basis functions are available which can be used as wavelet

functions in multiscale decomposition, such as, the Daubechies and Haar basis functions [49][50].

In wavelet based multiscale data decomposition, low pass and high pass filters are applied on the data. For example, applying a low pass filter on the original data provides a coarser approximation of the data, which is called the first scaled signal (see Figure 18). The low pass filter is derived from a scaling function of the form:

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

where, j and k represent the discretized dilation and translation parameters respectively. The difference between the first scaled signal and the original data (called first detail signal, see Figure 18) can be computed by applying a high pass filter that is derived from a wavelet function of the form:

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

Repeating the application of the low pass and high pass filter provides scaled and detailed signal at various levels, which correspond to different frequencies. After applying the low pass and high pass filters, the original signal or data can be expressed as the sum of the last scaled signal and all detail signals from all scales, which can be mathematically expressed as follows [51]:

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

where, J and n represent the maximum possible decomposition depth and the length of the signal. This multiscale data representation procedure has been illustrated in Figure 18 [52].

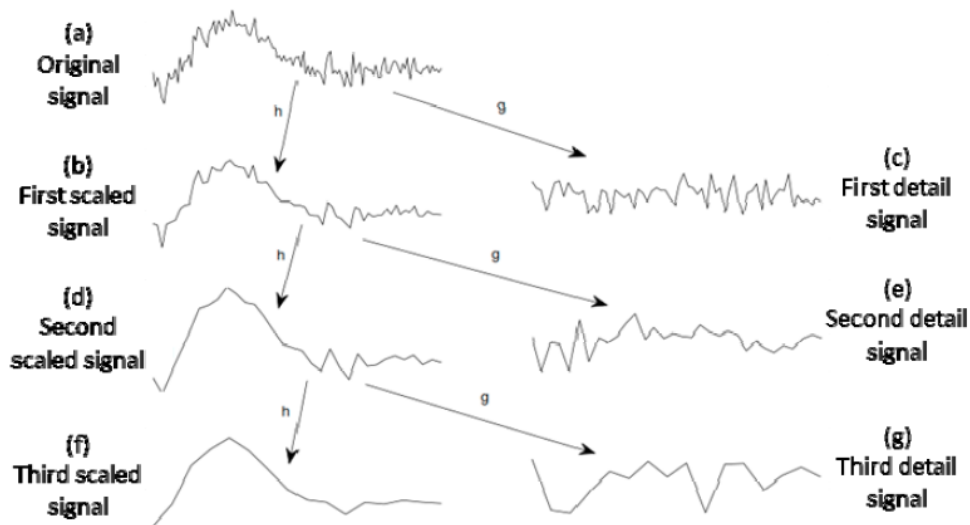


Figure 18: Schematic diagram for multiscale representation of data.

Wavelet based multiscale decomposition is an effective data analysis tool that has been widely used in various applications including physical, medical, engineering and social sciences. It has also been found useful in improving the effectiveness of various fault detection methods such as PCA by developing a multiscale PCA monitoring method [48][53], which has been used in practice to improve monitoring wastewater treatment processes [54]. Some of the advantages of multiscale representation in process monitoring are discussed next.

3.2. Advantages of multiscale representation of data

3.2.1. Noise feature separation

Multiscale representation has the ability to separate noise from important features in the data. When data is decomposed at multiple scales by passing through low pass and

high pass filters, noise is effectively separated from the important features. Random noise in a signal normally are present over all the coefficients, while deterministic features in the data are captured in a few, but relatively large coefficients. The important features in the data are usually captured by the last scaled signals as well as any large wavelet coefficient (in the detail signals), while other small wavelet coefficients usually correspond to noise [14][55]. Thus, multiscale provides an effective method for noise-feature separation as shown in 18.

This advantage of noise-feature separation of multiscale representation of data has been used effectively for various application such as filtering time series genomic data [56].

3.2.2. Decorrelation of autocorrelated data

Another advantage of multiscale wavelet based representation is that the wavelet coefficients of detail signals at different scales become approximately decorrelated even though the original data are autocorrelated [51].

To demonstrate the effect of multiscale decomposition on the level of autocorrelation in the detail signals at multiple scales, the autocorrelation function (ACF) is used. The ACF quantifies the magnitude of the correlation between data samples as a function of their separation [38]. Thus, it measures the memory of stochastic processes [51]. For uncorrelated data, the ACF shows zero values for all lags (time difference between two samples) except for lag zero, where it shows a value of unity. On the other hand, for correlated processes, ACF shows non-zero values for lags other than zero.

Figure 19 shows the detail signals obtained from a correlated signal (that represents an AR(1) model with an autoregressive model parameter of 0.7) and the corresponding ACF's at different scales. Figure 19 clearly shows that even though the time-domain data are autocorrelated (where its ACF has non-zero values at lags other than zero), the detail signals are approximately decorrelated.

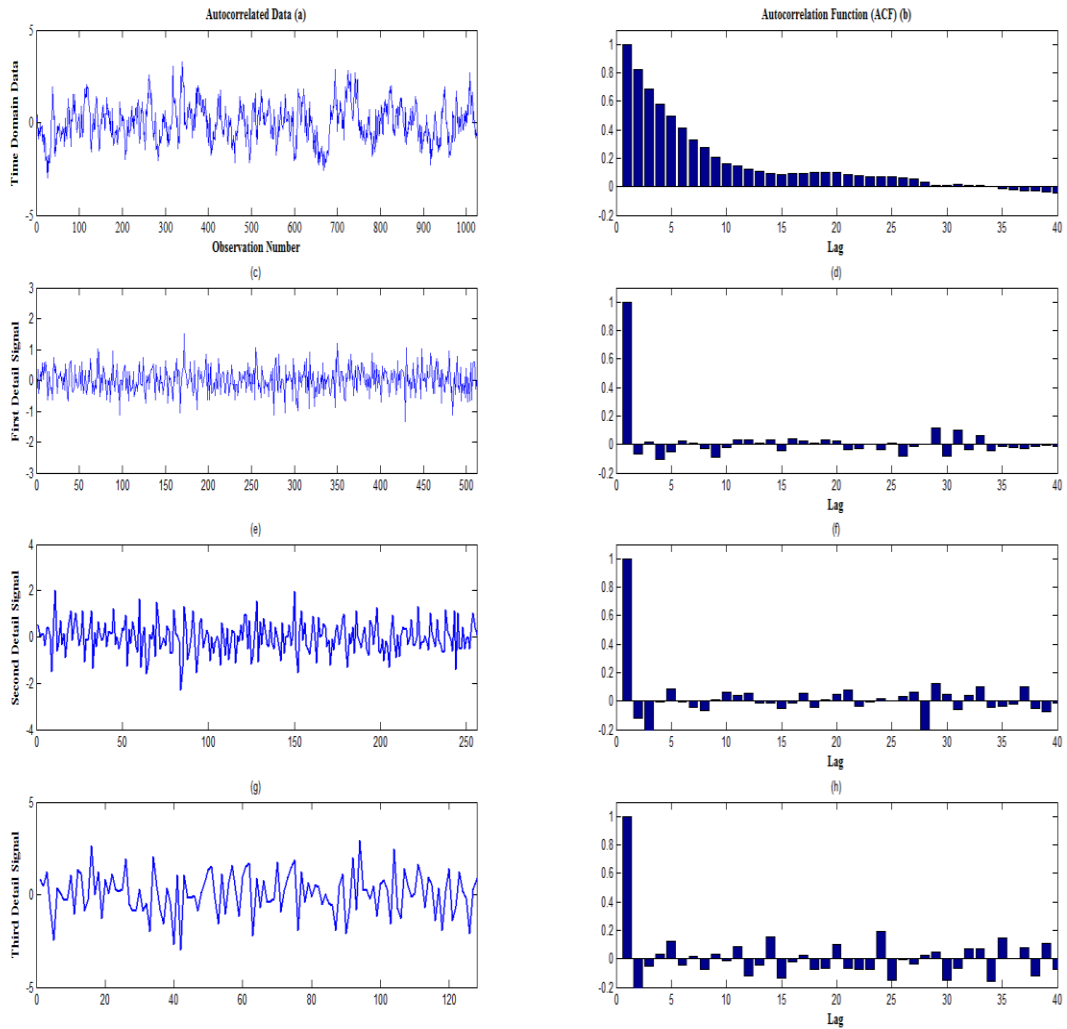


Figure 19: Decorrelation of autocorrelated data at multiple scales.

3.2.3. Data are closer to normal at multiple scales

Multiscale wavelet decomposition also makes the distribution of data closer to normal or Gaussian at multiple scales even if the original data follow non-normal distribution. Even though the effect of distribution on the performance of EWMA is not significant, transforming the data to be closer to normal helps satisfy its assumption better.

To show the advantage of multiscale representation in providing detail signals that are closer to normal at multiple scales, histograms of a chi-square distributed signal as well as its detail signals at multiple scales are shown in Figure 20. The original time-domain data have a Shapiro-Wilk statistic of 0.6953, which means it has a high degree of non-normality. Figure 20 shows that, as the decomposition depth increases, the detailed signals become more and more Gaussian.

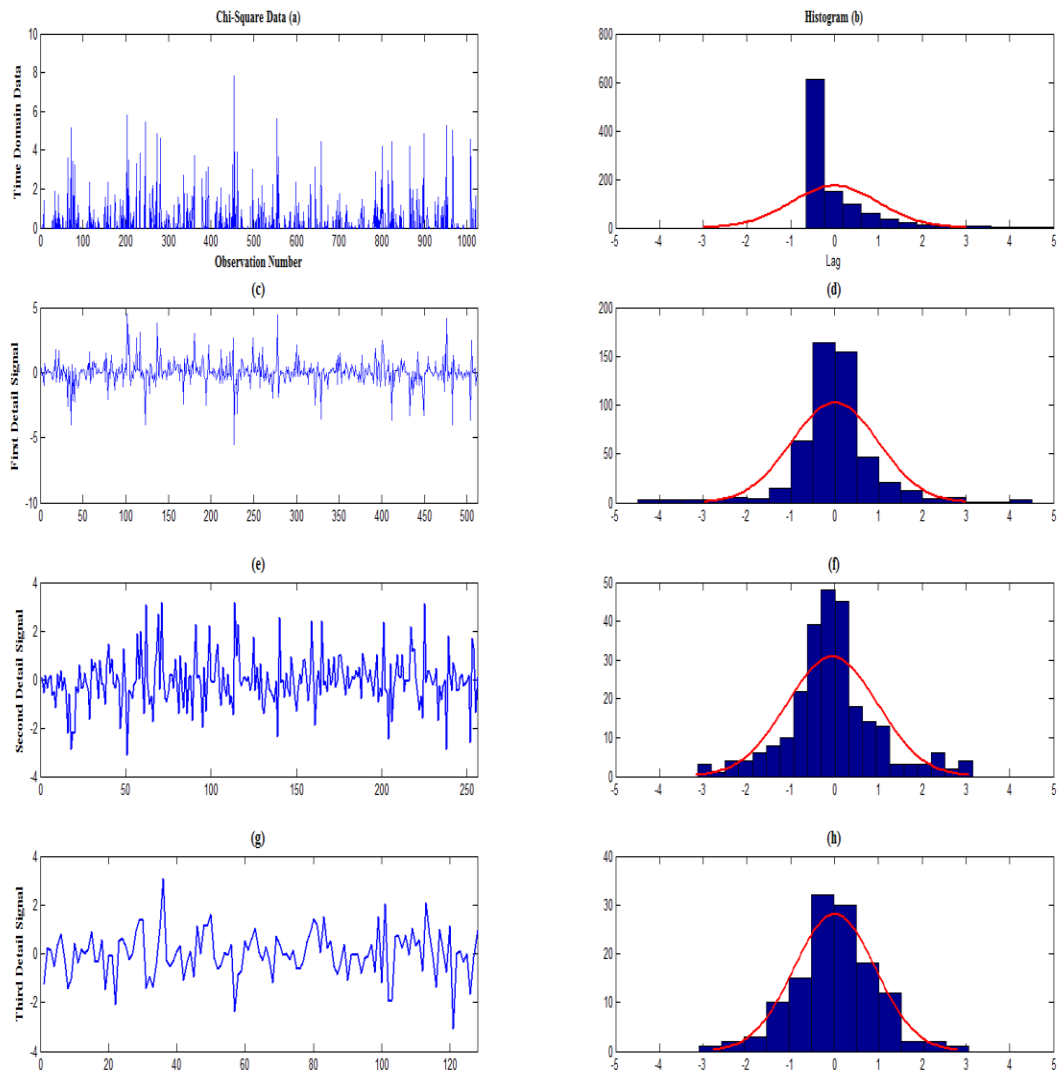


Figure 20: Distribution of chi-squared data at multiple scales.

So, wavelet based multiscale decomposition helps transform the data to be closer to normal at multiples scales despite their distribution in the time-domain doesn't follow normal distribution.

The above described advantages of multiscale representation clearly show that they can help satisfy the independence, normality, and noise level assumptions made by

various univariate fault detection methods, such as EWMA which is the method of interest in this work. Thus, utilizing multiscale representation should provide improvements to the performance of EWMA. However, performing multiscale fault detection in two separate steps (multiple filtering and then fault detection) may not provide the sought improvements since filtering may remove features that are important for fault detection. Thus, an algorithm that integrates fault detection using EWMA and multiscale representation is needed [52]. A multiscale EWMA fault detection method is presented in the next chapter.

4. WAVELET BASED MULTISCALE EWMA CHART

In this chapter, an algorithm for wavelet based multiscale EWMA fault detection will be described. Then, a design procedure for optimizing the parameters of the multiscale EWMA technique based on the lowest ARL_1 value will be presented.

4.1. Process monitoring using Multiscale EWMA Chart

The proposed multiscale EWMA monitoring technique consists of two phases [52] as shown in the Figure 21. In the first phase, fault free training data are normalized so that they have zero mean and unit variance, and are then decomposed at multiple scale using wavelet based multiscale decomposition. Then, the EWMA chart is applied to the detail signals at different scales as well as to the last scaled signal, and the control limits are computed at all scales. These control limits are used then to threshold the wavelet coefficients of detail signals. If any wavelet coefficient violates the control limits at a certain scale, all the wavelet coefficients at that scale are retained. If no violation of the limits occur at a certain scale, then all wavelet coefficients at that scale are ignored. The retained detail signals and the last scaled signals are then reconstructed to get the final reconstructed signal. Finally, EWMA is applied on the reconstructed signal to obtain the final multiscale EWMA detection statistic and the control limits.

In the second phase, the testing data are decomposed at multiple scales using the same wavelet filters used in the training phase after normalizing the data using the same mean standard deviation obtained in training. The control limits obtained from the training phase are then applied to the detailed signals of the testing data at the respective scales

and also to the last scaled signal. At any scale, the wavelet coefficients that violate the control limits are retained while other that don't violate the limits are ignored. Then, a reconstructed signal from all the retained coefficients is obtained. Finally, the previously obtained control limits from reconstructed training data are then applied on the EWMA statistic of the reconstructed testing data to detect possible faults. This algorithm is illustrated schematically in Figure 21.

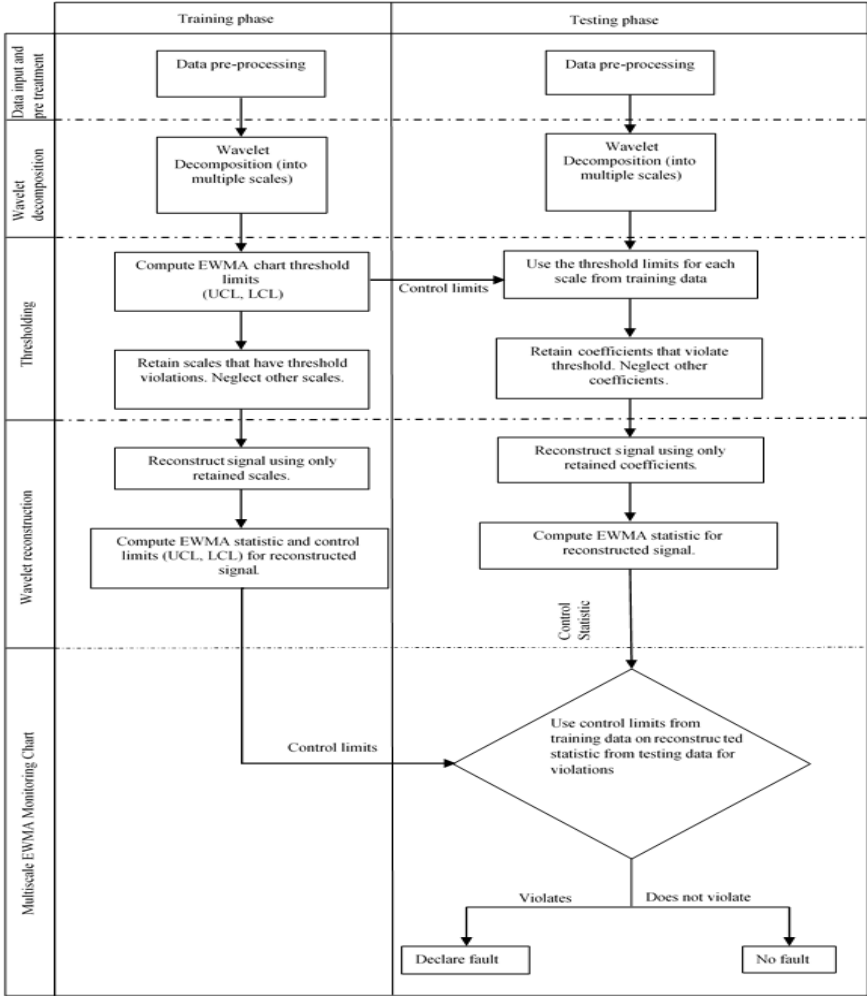


Figure 21: A schematic diagram of the multiscale EWMA fault detection algorithm.

The multiscale EWMA fault detection algorithm presented in Figure 21 provides a general framework for the implementation of multiscale EWMA, but it does not provide a strategy for selecting its parameters, namely the smoothing parameter, λ , and the control width, L . Applying the multiscale EWMA algorithm without optimizing these parameters may not provide better performance compared to the well-designed conventional EWMA technique for detecting specific fault sizes. In the next section, a procedure for selecting the optimum parameters used in the multiscale EWMA fault detection technique will be presented

4.2. Design procedure of optimizing the parameters of the multiscale EWMA fault detection technique

As indicated earlier, it is not possible to minimize all the monitoring performance indicators – ARL_1 , false alarm and missed detection rate at the same time for the conventional EWMA fault detection method, which is also true for the multiscale EWMA technique. So, to establish a design procedure for the multiscale EWMA technique, first of all, a method for optimizing its parameters based on an indicator needs to be established. The parameters optimized to provide lower ARL_1 values, will not give lower false alarm rates and vice versa. In this work, ARL_1 is used as a design basis to optimize the multiscale EWMA parameters and to establish a design procedure for its implementation. Selection of the ARL_1 as a design basis allows faster detection of faults, which is desirable in online process monitoring.

In this section, a procedure of optimizing the parameters of the multiscale EWMA technique for a fixed ARL_0 value will be discussed and then a design procedure of the implementation of the multiscale EWMA technique will be established.

4.2.1. Optimizing the multiscale EWMA parameters

The conventional EWMA fault detection method is associated with two parameters – control width, L and smoothing parameter, λ . The multiscale EWMA methods, however, brings in another parameter, which is the multiscale decomposition depth. For establishing a design procedure for implementing the multiscale EWMA method, all three parameters needed to be optimized to give a minimum ARL_1 value, which is the used design criteria.

To optimize the parameters – L and λ , a fault free training data consisting of 8192 Gaussian samples having a zero mean and unit standard deviation are generated. The decomposition depth is selected to be 6 (around half the maximum decomposition depth, which is 13 since $2^{13} = 8192$). Then, the multiscale EWMA algorithm is used to find out different combinations of λ and L values such that each combination gives an in control average run length (ARL_0) of 500. A Monte Carlo simulation of 5000 realizations is used for each combination to make sure that these combinations in fact give the specified ARL_0 . To find out these combinations, first of all, a fixed value of λ is assumed, and then different values of L are used to compute the ARL_0 value. Among all of these combinations, the pair of L and λ that results in an in control average run length value of 500 is stored. This process is repeated for different values of λ , and in each case the value of L that results in an in-control average run length of 500 is stored. These obtained combinations of L and λ are used to generate the plots shown in Figure 22.

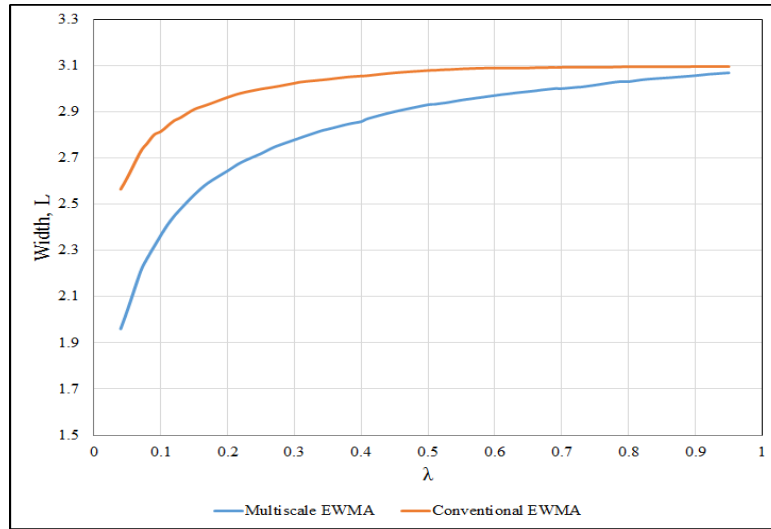


Figure 22: Combination of λ and L for in control average run length 500.

Now, to select the optimum combination that gives the lowest ARL_1 , faulty testing data having the same length as the training data and a fault of size equal to the standard deviation of the data (i.e., σ) between samples 2000 to 3000 and a fault of size equal to -1σ between samples 4500 to 5000 are generated. The size of the fault ($\pm 2.5\sigma$) is selected here randomly, the same process will be repeated later for other fault sizes. Then, the multiscale EWMA algorithm is applied to the testing data using all the combinations of λ and L values shown in Figure 22 to see which combination gives the lowest ARL_1 value, which is the optimum combination for a fault size of σ . A Monte Carlo simulation is used for each combination to get statistically meaningful results. The results of this optimization process are shown in Figure 23 that shows that optimum L and λ combination that minimize the ARL_1 criterion for a fault size of 2.5σ .

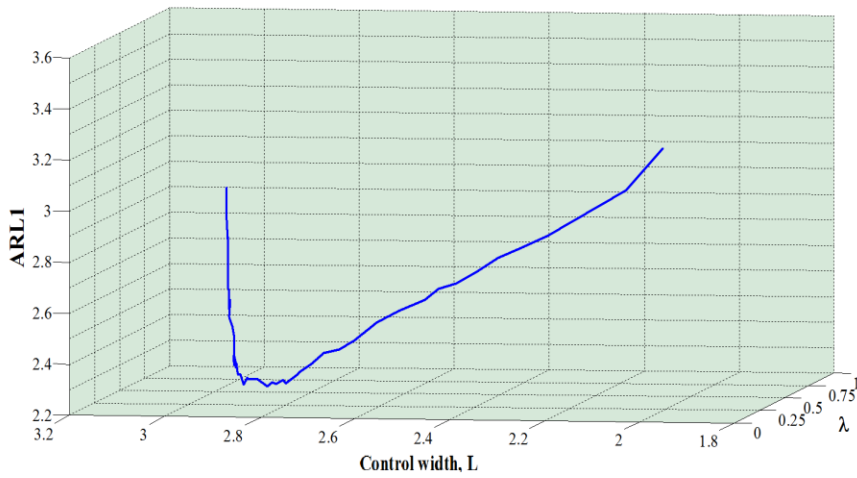


Figure 23: 3-D plot of the optimization of parameters for fault size 2.5σ for multiscale EWMA chart for ARL_0 value of 500.

The same procedure is repeated for different fault sizes, which provides the results shown in Figure 24, which is used to find out the optimum value of λ for different fault sizes.

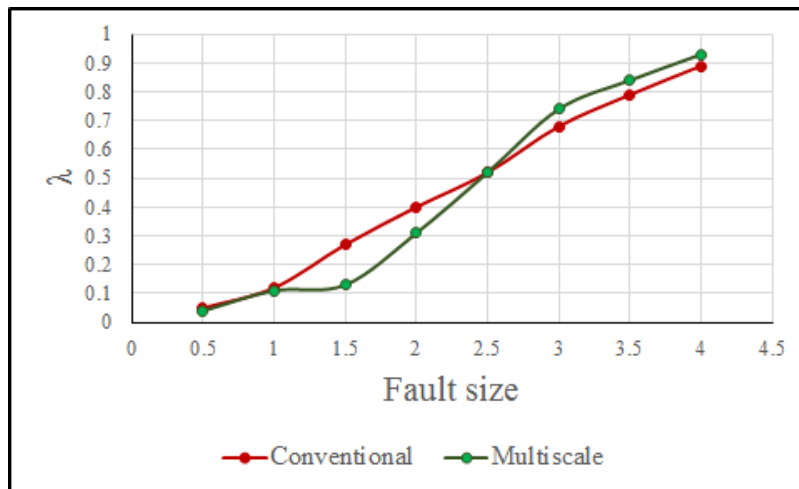


Figure 24: Optimal λ 's for the multiscale EWMA chart for different fault sizes for $ARL_0 = 500$.

As the optimum values of L and λ are available now for different fault sizes, the next step is to optimize the decomposition depth. Normally, noise can be reduced more by increasing the decomposition level. However, a larger decomposition depth will cause delay in detecting a fault in the time domain [20]. So selecting the optimum decomposition level is important to get full advantage of the multiscale algorithm.

To examine the effect of decomposition depth on the performance of the multiscale EWMA technique, the following simulation study is performed. In this simulation, a training data set consisting of 8192 observations that follow zero mean Gaussian distribution is generated. Then the multiscale EWMA algorithm is applied on the training data using decomposition depth of one and the control limits for all detailed as well as for the reconstructed signal are computed. Then, multiscale EWMA is applied on a testing data set that is generated the same way as the training data set but with two faults having magnitudes of $\pm 1\sigma$ between samples 2000-3000 and samples from 4500-5000. The control limits obtained from the training data are used to detect faults in the testing data, and the ARL_1 , false alarm rate and missed detection rate are computed. The same simulation is repeated using different decomposition depths ranging from 1 -10. A Monte Carlo simulation using 5000 realization is performed for this study and the results are shown in Figures 25 and 26.

Figure 25 shows that, the ARL_1 value decreases sharply from a decomposition depth of 0 (which corresponds to the time domain or conventional EWMA) to a depth of 1. This is an indication that multiscale EWMA method outperforms the conventional EWMA method with respect to ARL_1 . Then, the ARL_1 reaches its lowest value at

decomposition depth of 6 and then it increases again for larger decomposition depths. These results show that increasing the decomposition depth improves the ARL_1 (i.e., improves the speed of detection, but up to a certain depth beyond which the speed of detection deteriorates).

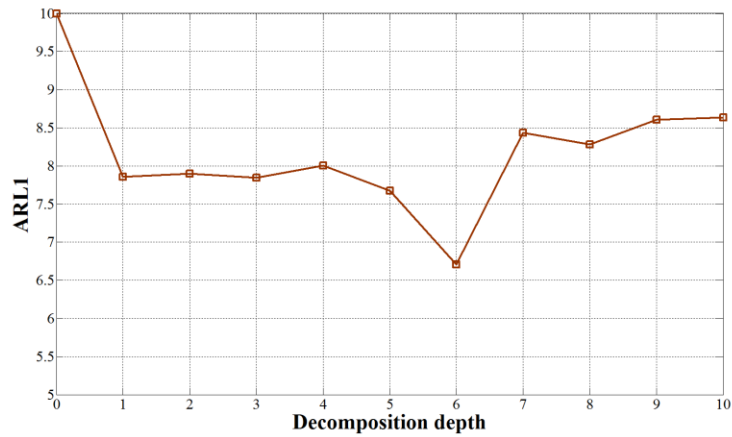


Figure 25: Effect of decomposition depth on the ARL_1 of the multiscale EWMA chart.

The effect of decomposition depth on the false alarm and missed detection rates, on the other hand, is illustrated in Figure 26, which shows that multiscale EWMA provides improvement in terms of the missed detection rate and a slight increase in the false alarm rate with respect to the conventional method. Figure 26 also shows that both missed detection and false alarm rates are not significantly affected by the decomposition depth. Thus, Figures 25 and 26 show that the multiscale EWMA algorithm provides its best performance at a decomposition depth around half the maximum decomposition depth

(which is 6 in this simulated study). It is important to note that, applying multiscale EWMA, as in any multiscale technique, requires a dyadic data set.

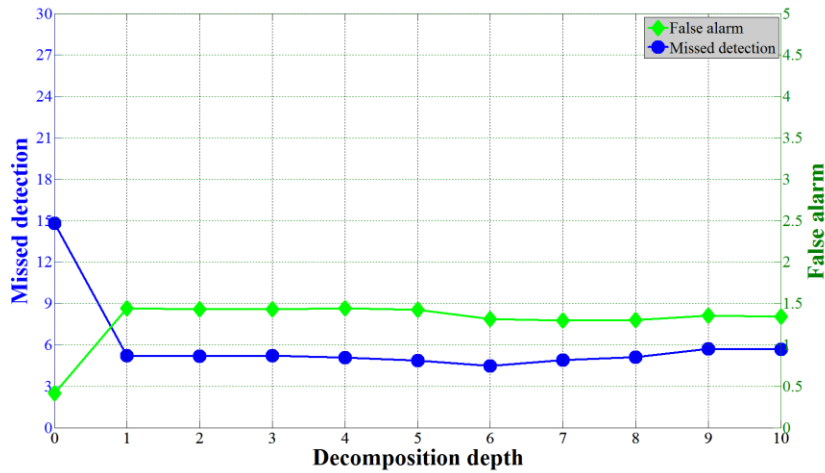


Figure 26: Effect of decomposition depth on false alarm and missed detection rates of the multiscale EWMA chart.

With the help of these simulated results, a design procedure for the multiscale EWMA can be established, which will be presented in the next section.

4.2.2. Design steps for the multiscale EWMA technique

The design procedure of the multiscale EWMA technique consists of the following steps:

- Choose an acceptable value of in control average run length (ARL_0), which is taken as 500, in this work
- Select the decomposition depth based on the length of the data (half the maximum decomposition depth is recommended) -data has to be dyadic

- Specify the minimum fault size that needs to be detected as quickly as possible, and determine the value of λ which provides the lowest ARL_1 for that specific fault size. (using Figure 24)

- Find the value of the control width L , which along with the value of λ (found from previous step) provides the required ARL_0 value. (using Figure 22)

In the next chapter, the performance of the multiscale EWMA described earlier will be compared with that of the conventional method from different perspectives to see which one gives better performance in fault detection.

5. PERFORMANCE COMPARISON BETWEEN THE CONVENTIONAL AND MULTISCALE EWMA CHARTS

Both the conventional and multiscale EWMA techniques are designed to detect faults in the shortest possible time by minimizing the out-of-control average run length (ARL_1). In this chapter, the performances of the two techniques will be compared using their optimally designed parameters. The two techniques will be compared using all three indices, ARL_1 , missed detection rate and false alarm rate. First, a comparison will be performed at different sizes of mean shifts when the assumptions of the conventional EWMA method are satisfied, which include the normality and independence of the residuals. Then, the performance will be compared when these assumptions are violated.

5.1. Comparison between the performance of the conventional and multiscale EWMA techniques under no violations of the EWMA assumptions

In this section, the performances of the conventional and multiscale EWMA techniques are compared at different sizes of mean shifts and using their optimal parameters to make sure that each technique provides its best performance. To perform this comparison, training data consisting of 8192 independent zero mean Gaussian observations are generated. The testing data are also generated in a similar manner but with additive faults of magnitude 1 between samples 2000-3000 and of magnitude -1 between samples 4500 -5000. Then, the conventional EWMA technique is applied using L and λ values of 2.861 and 0.12, respectively, which are the optimum values to detect

fault of size 1. Similarly, the multiscale EWMA technique is applied on the same data with its own optimum parameter values, which are 2.41, 0.11 and 6 for L , λ and the depth, respectively. All three indices, ARL_1 , false alarm rate and missed detection rate are computed for both techniques and a Monte Carlo simulation of 5000 realization is performed to get meaningful results. Then, the same procedure is repeated for different fault sizes ranging from 0.5 to 4, and the simulation results are shown in Figures 27, 28 and 29.

Figure 27 shows that, for smaller fault sizes, the multiscale EWMA detects the faults significantly quicker than the conventional EWMA technique. For example, for a fault of size 0.5, the multiscale EWMA has an ARL_1 value of 17.496 which is much smaller than the ARL_1 value of the conventional EWMA, which is 27.731. However, for larger faults, both the techniques perform almost equally.

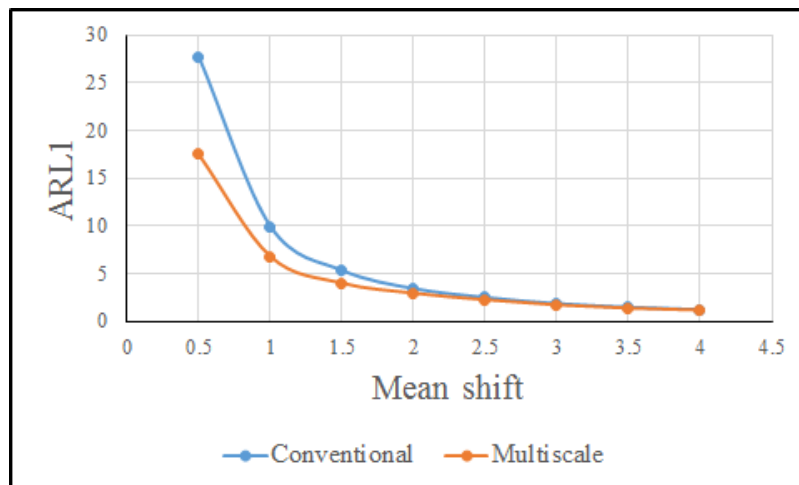


Figure 27: Performance comparison between the conventional and multiscale EWMA charts in terms of ARL_1 for different fault sizes.

Figure 28, on the other hand, which shows the comparison results in terms of false alarm rate, shows that the multiscale EWMA technique results in a slight increase in the false alarm rate compared to the conventional method, especially for small shift sizes. However, the difference between the false alarm rates for the two techniques are not too high and the maximum false alarm rate that the multiscale technique produces (for fault size of 0.5) is equal to 4.3%. These relatively higher false alarm rates are expected though as the technique is designed to give a minimum ARL_1 which can increase the false alarm rate (see section 1.1.2).

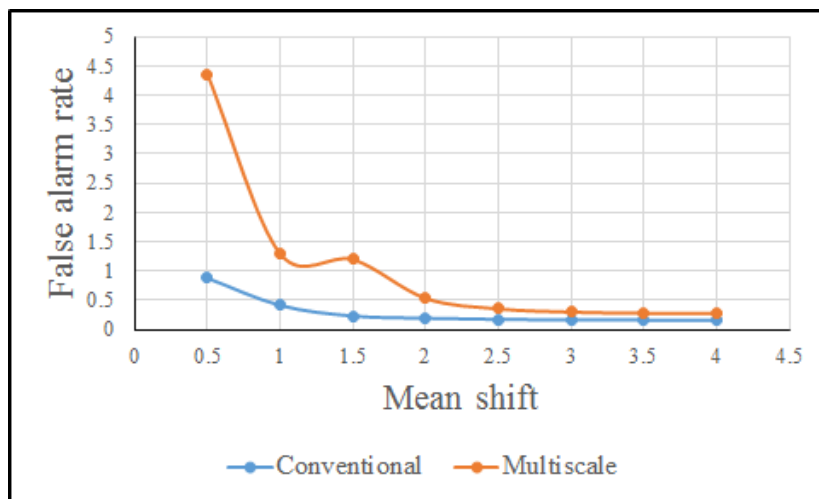


Figure 28: Performance comparison between the conventional and multiscale EWMA charts in terms of false alarm rate for different fault sizes.

The comparison results between the conventional and multiscale EWMA techniques in terms of missed detection rate are shown in Figure 29, which shows that the multiscale EWMA technique provides improved performance over the conventional one

for most of the fault sizes except for very high fault sizes (for fault sizes of 3 – 4), where both the techniques perform almost equally.

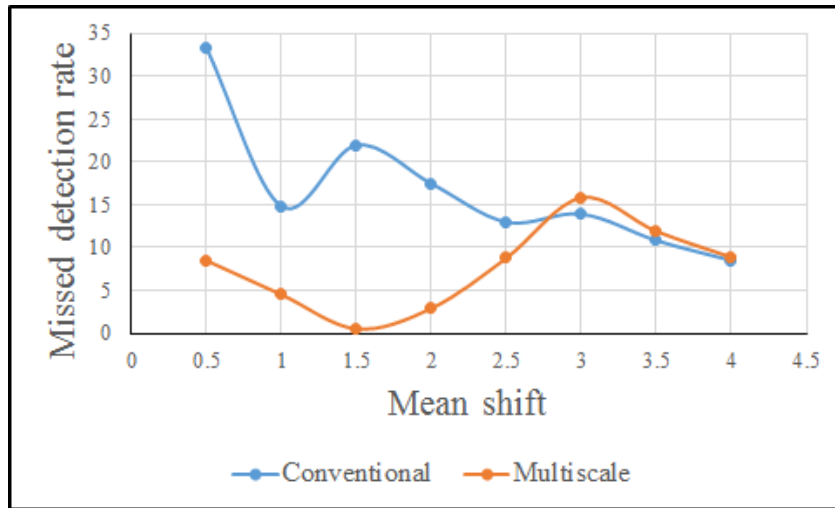


Figure 29: Performance comparison between the conventional and multiscale EWMA charts in terms of missed detection rate for different fault sizes.

To explain the nonlinear behavior of the plots shown in Figure 29, it is important to understand how the false alarm and missed detection rates change for different values of L and λ for the multiscale EWMA technique. The effect of changing the control width (L) and smoothing parameter (λ) on the missed detection and false alarm rates are illustrated in Figure 30 for L values ranging from 1-4 and λ values ranging from 0.05-0.25. In this analysis, the noise level and fault sizes are kept fixed at 1 and ± 1 respectively. Figure 30 clearly shows a nonlinear behavior for the missed detection and false alarm rates at different L and λ values, which explains the nonlinear behavior observed in Figure 29.

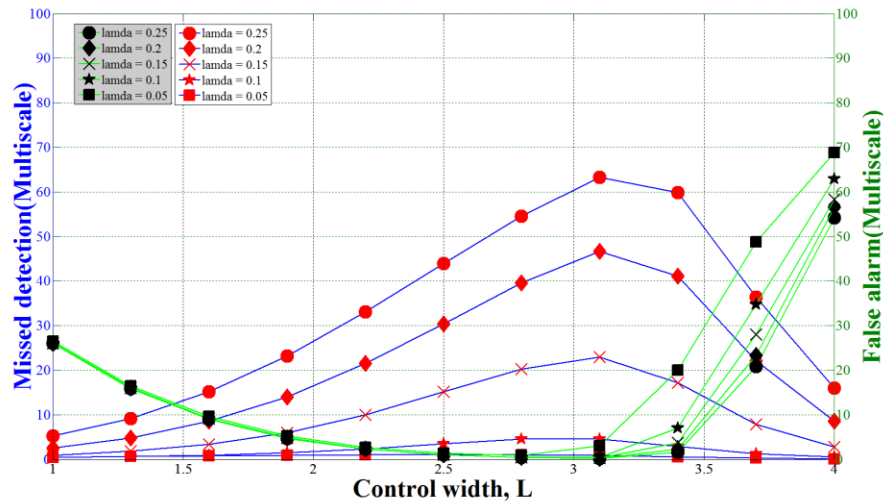


Figure 30: Trade-off between false alarm and missed detection rate for the multiscale EWMA chart.

In summary, Figures 28 and 29 show that although both the multiscale and conventional EWMA technique are designed to optimize the ARL_1 criterion, the multiscale EWMA provides improvement in the missed detection rate, besides ARL_1 . So, even though the false alarm rate increases slightly, it is evident that the multiscale algorithm outperforms its conventional time domain EWMA.

5.2. Comparison between the performances of the conventional and multiscale EWMA techniques under violations of the EWMA assumptions

In this section, the performance of the multiscale EWMA chart will be assessed and compared with that of the conventional EWMA method under violation of EWMA chart assumptions. To compare the performances of the two EWMA techniques, the same data that were used in section 2.2 to assess the effect of noise level, autocorrelation and deviation from normality will be used here. The optimum parameters are used for the

respective techniques, which correspond to the fault size introduced in the data, which is 1σ for all the cases.

5.2.1. Comparison of performance using data with different level of noise

In this study, the effect of the noise level on the performances of the conventional and multiscale EWMA methods is assessed. Here, the fault size is fixed at 1 and the noise standard deviation is varied from 0.01-2, and then in each case the ARL_1 , missed detection rate and false alarm rate are computed, and the results are shown in Figures 31 and 32. Figure 31 shows that- the multiscale EWMA chart gives better performance in terms of ARL_1 value than the conventional method, especially at larger noise levels. The missed detection and false alarm rates, on other hand, are illustrated in Figure 32, which shows that the multiscale EWMA technique outperforms the conventional method with respect to the missed detection rate, while the false alarm rates are comparable for both techniques with a slight advantages in favor of the conventional method.

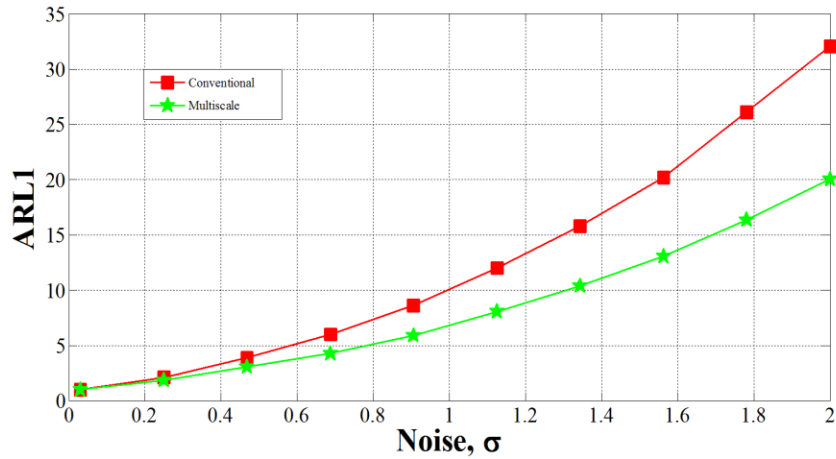


Figure 31: Performance comparison between the multiscale and conventional EWMA technique in terms of ARL₁ for different level of noise in the data.

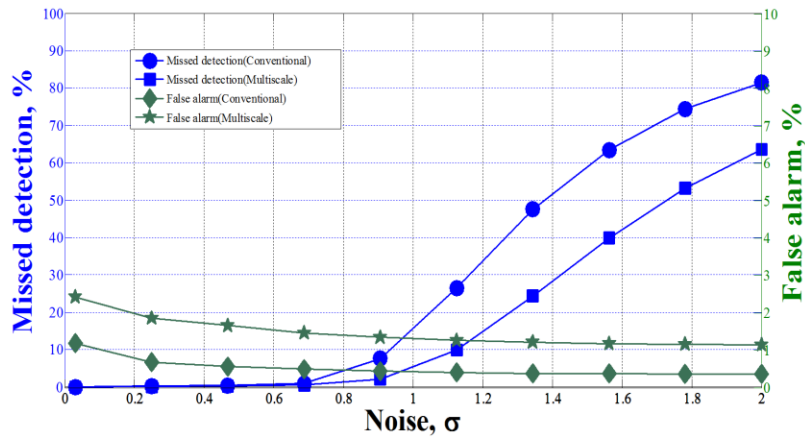


Figure 32: Performance comparison between the multiscale and conventional EWMA techniques in terms of false alarm and missed detection rate for different level of noise in the data.

To better explain the relative performances of the conventional and multiscale EWMA techniques, the detection statistics used in both methods are shown in Figure 33. It shows that the multiscale EWMA technique helps remove noise from the data through the

application of EWMA at multiple scales, which narrows the control limits of the final reconstructed signal. This helps provide a quicker detection of faults and a smaller missed detection rate, but at the expense of a slight increase in the false alarm rate compared to the conventional technique. Therefore, the multiscale EWMA method provides a better performance than the conventional method, especially at higher noise levels.

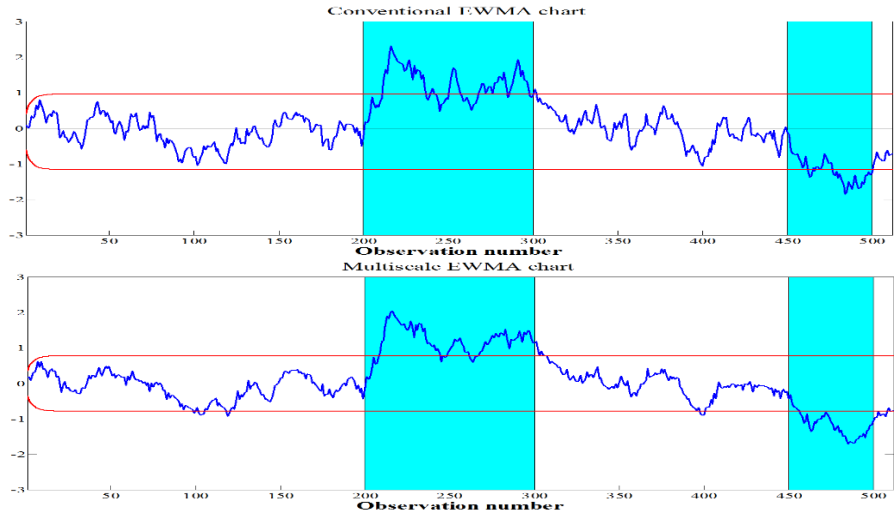


Figure 33: Detection statistics for the conventional and multiscale methods in the case where the noise standard deviation equals 1.5.

5.2.2. Comparison of performance using autocorrelated data

In this section, the conventional and multiscale techniques are compared through their application on autocorrelated data, which are generated using an autoregressive model, AR(1). The value of the autoregressive parameter, is varied between 0.1 and 0.95 and at each value the ARL_1 , false alarm rate, and missed detection rate are computed which

are shown in Figures 34 and 35. Figure 34 shows that- the multiscale EWMA method consistently provides smaller ARL_1 values compared to the conventional EWMA almost for all values of the autoregressive coefficients. Figure 34 also shows that for the multiscale EWMA method, the ARL_1 value remains almost constant except very high level of autocorrelation, where it still provides a quicker detection than the conventional EWMA technique.

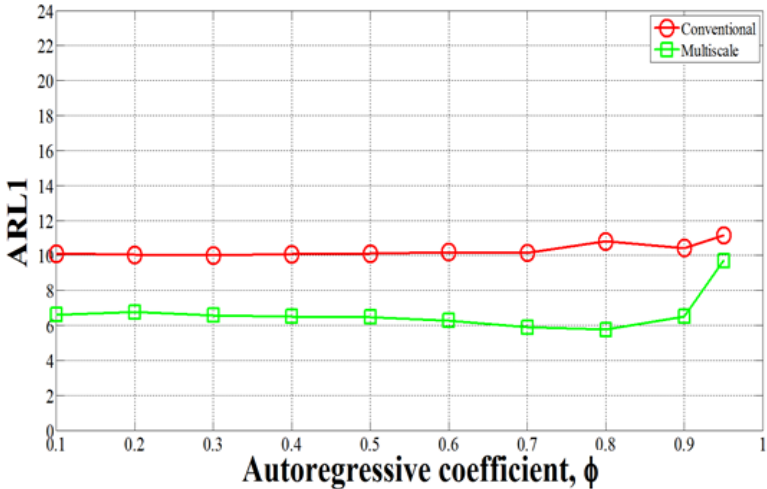


Figure 34: Performance comparison between the multiscale and conventional EWMA technique in terms of ARL_1 for different level of autocorrelation in the data.

Figure 35, on the hand, shows that the multiscale EWMA technique provides smaller missed detection than the conventional method for all levels of autocorrelation. The false alarm rate for the conventional method is smaller than that of the multiscale

technique, especially at high level of autocorrelation, i.e., large autoregressive model parameter.

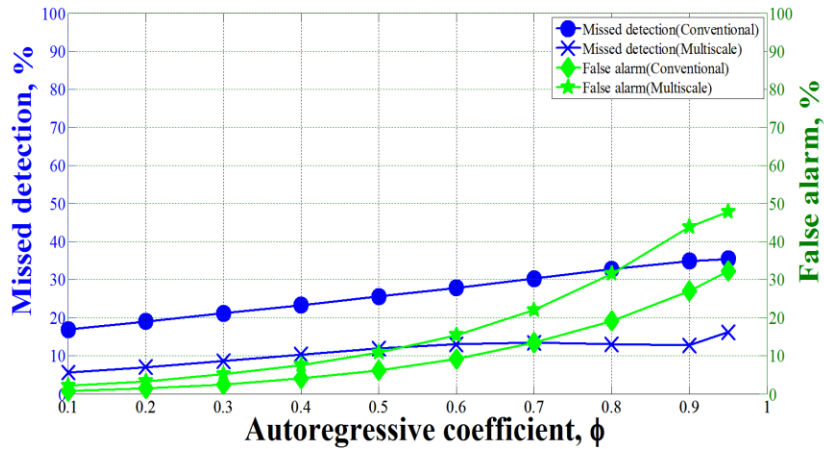


Figure 35: Performance comparison between the multiscale and conventional EWMA technique in terms of false alarm and missed detection rates for different level of autocorrelation in the data.

The advantages of the multiscale EWMA chart can be better illustrated by comparing the detection statistics for both techniques, which are illustrated in Figure 36. It shows that, again narrower control limits are obtained using the multiscale EWMA chart, which helps achieve a smaller ARL_1 value and missed detection rate, but with a relative increase in the false alarm rate, especially at high levels of autocorrelation. In summary, the multiscale EWMA method outperforms its conventional counterpart using autocorrelated data.

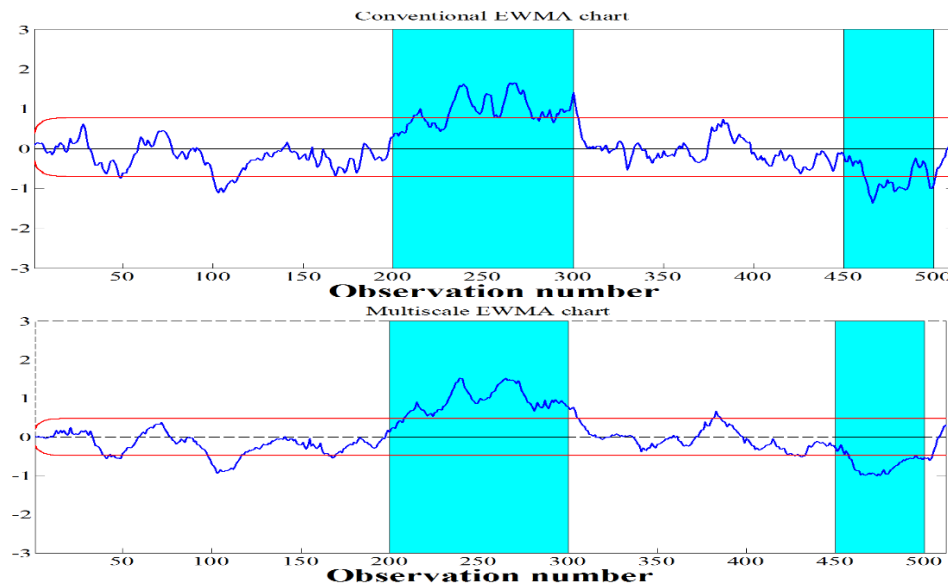


Figure 36: Detection statistic for the conventional and multiscale methods in the case where the autoregressive coefficient equal to 0.5.

5.2.3. Comparison of performance using non-Gaussian (chi-square) data

In this section, the effect of deviation from normality is examined for the multiscale and conventional EWMA methods through their application on chi-square data at different degrees of non-normality, i.e., different values of Shapiro Wilk ranging from 0.69 to 1. The results of the assessment are shown in Figures 37 and 38. Figure 37 shows that, for both techniques the ARL_1 remains almost constant at different Shapiro Wilk values, and that the multiscale technique consistently provides smaller ARL_1 values (i.e., quicker detection) irrespective of the degree of non-normality.

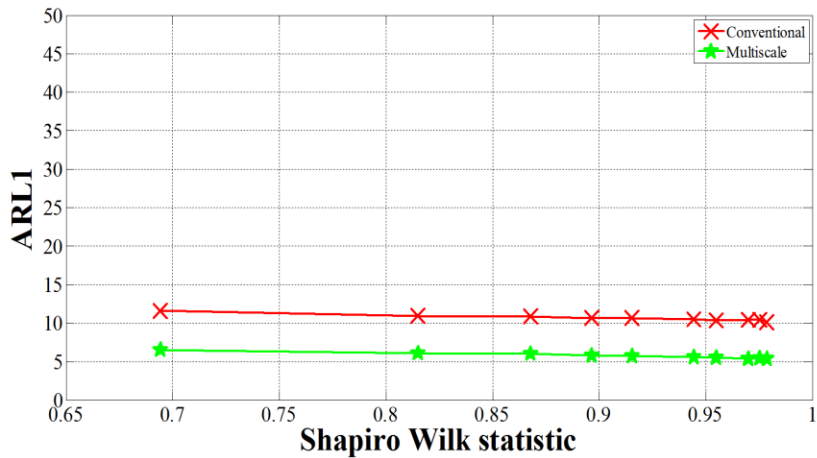


Figure 37: Performance comparison between the multiscale and conventional EWMA technique in terms of ARL1 for non-Gaussian data (Chi-square).

Similarly, the false alarm and missed detection rates also remain almost constant for both the techniques at different degrees of non-normality (see Figure 38). However, the multiscale EWMA method provides smaller missed detection rates but higher false alarm rates.

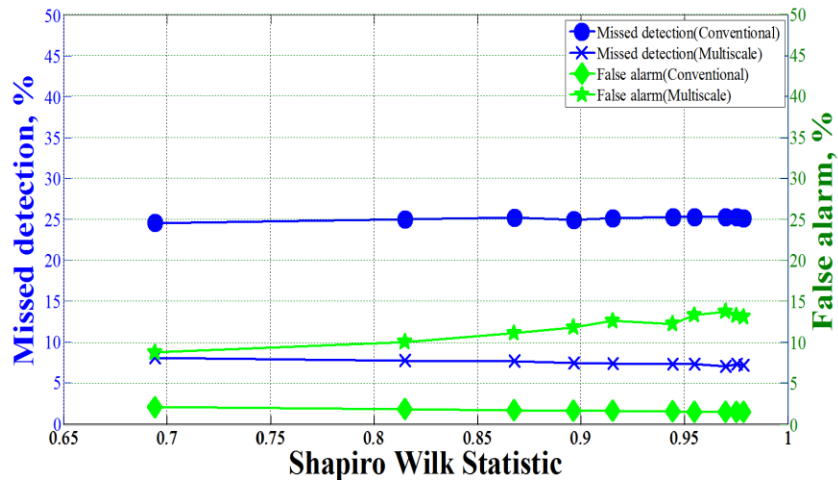


Figure 38: Performance comparison between the multiscale and conventional EWMA technique in terms of false alarm and missed detection rates for non-Gaussian data (Chi-square).

The performances of the conventional and multiscale EWMA techniques can also be compared by comparing their detection statistics, which are shown in Figure 39, which shows that, the multiscale EWMA technique provides faster detection by shrinking the control limits which result in higher false alarm rates.

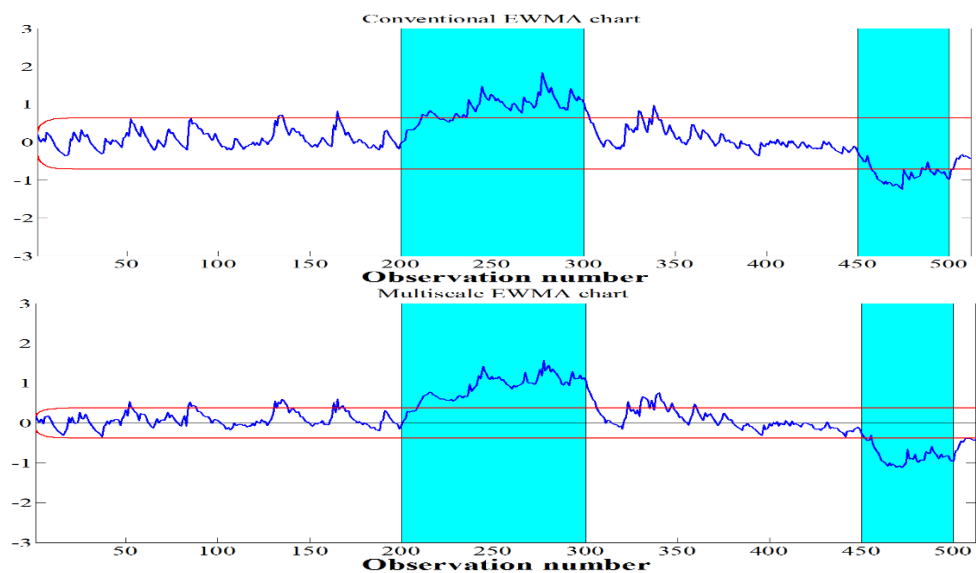


Figure 39: Detection statistics for the conventional and multiscale methods in the case where the Shapiro Wilk equals 0.69.

In summary, the multiscale EWMA technique outperforms the conventional EWMA technique not only when both the techniques are designed to perform optimally but also when data violate the assumptions of the EWMA chart. In the next chapter, the performances of both techniques will be evaluated using real application data.

6. APPLICATION OF THE MULTISCALE EWMA CHART

In this chapter, real application data will be used to assess the effectiveness of the multiscale EWMA method as a monitoring technique. To do so, simulated distillation column data will be used.

To simulate real process data, a distillation column consisting of 32 theoretical stages with a total condenser and a reboiler was simulated using Aspen Tech 7.2. The feed stream, which enters the column in a saturated liquid form at stage 16 with a mass flow rate of 1kg/mole.s and a temperature of 322 K, has compositions of 40 mole % propane and 60 mole % isobutene. The nominal operating conditions of the distillation column can be found in [57]. Dynamic data of the distillation column are generated by changing the feed and reflux flow rates from their nominal operating values, which is done by introducing step changes of magnitudes $\pm 2\%$ in the feed and reflux flow rates. After each step change, the process is given sufficient time to settle to a new steady state after which another step change is introduced. After introducing all step changes and reaching to a final steady state, 1024 observations that represent the process behavior are generated. All simulated data are assumed to be noise free, and therefore, the simulated data are contaminated with zero mean Gaussian noise to represent measurement errors.

The objective of this example is to compare the performances of the conventional and multiscale EWMA methods through monitoring the composition of propane in the distillate stream using faults of different magnitudes. Since both techniques require evaluating residuals, a partial least squares (PLS) model is developed to predict the

residuals of the propane composition using temperature data at various trays of the column as well as the feed and reflux rates.

The residuals obtained by this process are then divided into two set – training and testing, each consisting of 512 samples. Two step faults of magnitude $\pm 2\sigma$ are then introduced to the residuals between samples 201-250 and samples 401-450 of the testing data respectively, where σ is the standard deviation of the residuals. Then, the control limits are computed from the training data by using the corresponding optimum parameters for a fault size of 2σ of both techniques. Those control limits are then used to compute ARL_1 , false alarm rate and missed detection rates for both the techniques using the testing data.

Figure 40, which shows the detection statistics for both the conventional and multiscale EWMA techniques, clearly shows that the multiscale EWMA technique provides a smaller missed detection rate (5%) compared to one obtained using the conventional technique (33%). However, the false alarm rate obtained using the multiscale technique (3.9%) is a little larger than what is obtained using the conventional technique (0.6%). On the other hand, the multiscale EWMA provides a little smaller ARL_1 value (4) compared to the one obtained using the conventional technique (5).

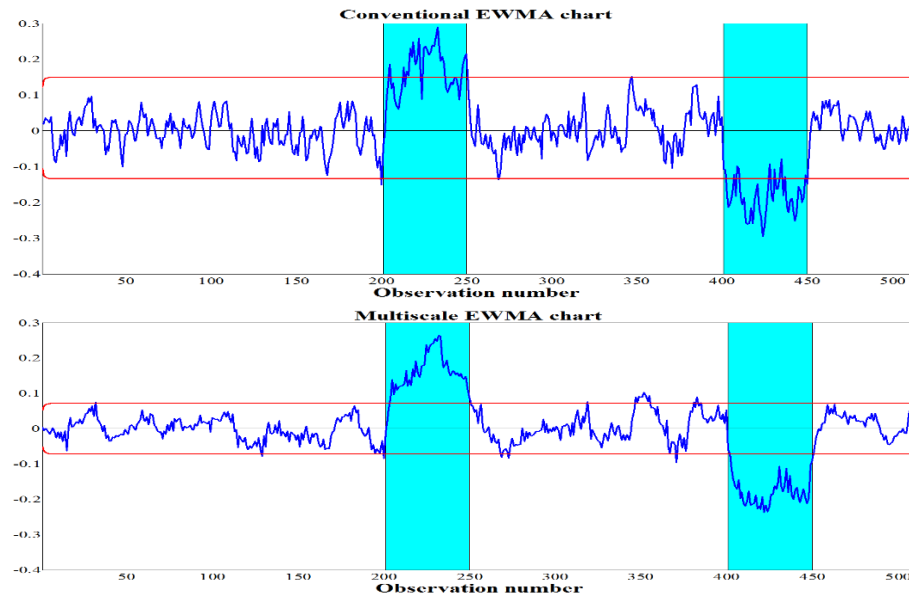


Figure 40: Comparing the performances of the conventional and multiscale EWMA chart for a step fault of magnitude $\pm 2\sigma$ in the residuals of simulated distillation column data.

A similar comparison is performed using smaller fault sizes ($\pm\sigma$), and the results are shown in Figure 41. The detection statistics shown in Figure 41 show that the missed detection rate for the multiscale EWMA is smaller (19%) than the conventional EWMA (39%). The multiscale EWMA method provides relatively faster detection ($ARL_1=16$) than the conventional method ($ARL_1=19$). The false alarm rate is higher for the multiscale technique (6%) than for the conventional technique (1%).

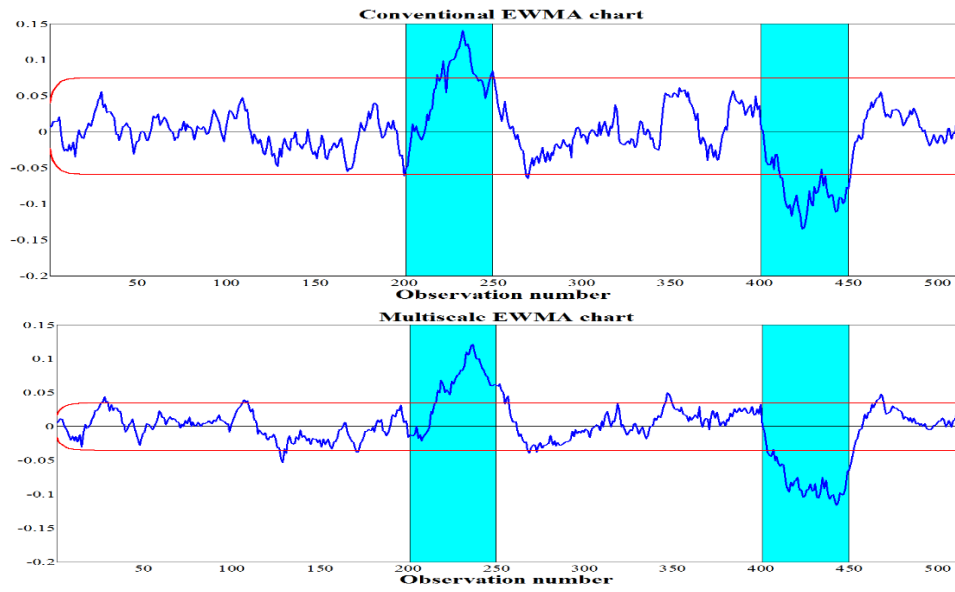


Figure 41: Comparing the performances of the conventional and multiscale EWMA chart for a step fault of magnitude $\pm\sigma$ in the residuals of simulated distillation column data.

7. CONCLUSIONS AND FUTURE DIRECTIONS

7.1. Concluding remarks

Most univariate process monitoring techniques, such as EWMA, operate under three main assumptions – the residuals obtained from process data contain a moderate level of noise, are independent or uncorrelated, and follow a Gaussian distribution. The available optimized values of the EWMA parameters (L and λ), which are obtained by minimizing the out of control average run length (ARL_1) for different fault sizes, are obtained by making these assumptions. In this work, it has been demonstrated by illustrative examples that, when the noise level increases, the performance of the optimally designed EWMA technique deteriorates. For example, for a noise level of 2σ , the missed detection rate and ARL_1 increases to 80% and 33, respectively. Similarly, the effect of autocorrelation can be large, especially at high levels of autocorrelation. However, the deviation from normality does not seem to affect the performance of the EWMA technique.

Wavelet based multiscale representation can help satisfy these assumptions as it has been shown that multiscale representation provides a good separation of noise from important features in the data, and that data become more independent and, closer to normal at multiple scales. Therefore, multiscale wavelet based representation is used in this work to improve the monitoring performance of the EWMA method.

To utilize the advantages of wavelet based multiscale representation to improve the performance of EWMA, an algorithm that integrates the EWMA chart with multiscale

representation is developed. Also, a procedure for selecting the optimum parameters of multiscale EWMA technique which include L , λ and the decomposition depth for detecting different fault sizes as quickly as possible, is designed. This selection of parameters (L and λ) varies depending on the in control average run length (ARL_0), the minimum fault size that needs to be detected and the type of residuals. Residuals can be contaminated with different levels of noise, can be autocorrelated, or non-Gaussian. however, in this work, the data are assumed to follow a zero mean Gaussian distribution, and the design is based on minimizing the out of control average run length ARL_1 for an in control average run length 500.

Then, the conventional and multiscale techniques are compared using the three indicators (ARL_1 , false alarm rate and missed detection rate) for different fault sizes when the optimum parameters for both techniques are used. The multiscale EWMA demonstrates better performance over the conventional EWMA in terms of ARL_1 and missed detection rate by providing smaller values of these indicators, especially for smaller fault sizes. The conventional EWMA method is known for its ability to detect small faults quickly, however, the multiscale EWMA technique improves the detection speed even more. In terms of false alarm rate, the multiscale EWMA provides slightly higher values than the conventional EWMA technique but overall this small disadvantage gets outweighed by the improvement in the ARL_1 value and the missed detection rate.

The impact of violating the assumption of the conventional EWMA method on the performances of both techniques is then studied by comparing their performances under each violation. It has been shown that for noisy, autocorrelated and non-normally

distributed data, the multiscale EWMA provides smaller values for ARL_1 and missed detection rate but higher values for false alarm rate compared to the conventional EWMA. When the data contain moderate to high levels of noise, the multiscale EWMA provides more than 30% reduction in ARL_1 value and 25-35% reduction in missed detection rate at the expense of 1-2.5% overall false alarm rate.

On the other hand, the multiscale EWMA provides almost 30% reduction in ARL_1 value over the conventional EWMA technique for wide levels of autocorrelation (almost for all values of the autoregressive coefficient when an AR(1) model is used to generate autocorrelated data). The multiscale EWMA technique also reduces the missed detection rate by half compared to the conventional EWMA method for all values of the autoregressive coefficient. However, there is significant increase in the false alarm rate of the multiscale EWMA over the conventional technique at high levels of autocorrelation (when the autoregressive coefficient is larger than 0.6).

The deviation from normality in the residuals doesn't seem to affect the performances of both the multiscale and conventional EWMA. All the three indices remain almost constant for different values of the Shapiro-Wilk statistic (the measurement of the degree in the data). The multiscale EWMA, once again, provides a better performance in terms of the ARL_1 value and the missed detection rate by reducing their values to almost half of the values obtained using the conventional technique. The false alarm rate obtained by the multiscale EWMA technique is higher than the corresponding value obtained by the conventional technique, and reaches a maximum value of 14% compared to around 2% for the conventional method.

The performance of the developed multiscale EWMA chart in practice is also illustrated through its application to monitor the distillate composition of a simulated distillation column. The performance of the multiscale EWMA is measured for two different cases when two step faults of different sizes (σ and 2σ) are added on the residuals. In both cases, multiscale EWMA provides faster detection (smaller ARL_1), lower missed detection rates and higher false alarm rates.

It is clear through various comparative studies, that the multiscale EWMA method can provide a significant improvement in fault detection over the conventional EWMA method. This will also pave the way for further investigations to improve the effectiveness of fault detection even more, some of the suggested future research directions are described next.

7.2. Future directions

In this work, the developed multiscale EWMA technique has been shown to outperform the conventional EWMA technique. These results will lead to further research in the area as described below:

- In this thesis, a method for determining the optimum multiscale EWMA parameters was developed for an in control average run length (ARL_0) of 500. Extensions of such approach using other ARL_0 are needed.
- The multiscale EWMA method developed in this work is univariate, a multivariate version of EWMA has been developed. Therefore, developing a multiscale multivariate EWMA method can extend the advantages of multiscale EWMA to multivariate processes.

- The developed multiscale EWMA method is batch because multiscale representation used requires the entire data set a priori. Therefore, an online extension of the multiscale EWMA method is needed to allow monitoring online processes.

- The developed multiscale EWMA method relies on minimizing the ARL_1 value for optimizing its parameters, which leads to a faster detection. However, using a different performance metric (which can be a combination of ARL_1 , missed detection rate, and false alarm rate) may lead to a better performance with respect to all indicators.

REFERENCES

- [1] P. Seifried, "Fault Detection and Diagnosis in Chemical and Petrochemical Processes," *Chemie Ing. Tech.*, vol. 51, no. 7, pp. 766–766, Jul. 1979.
- [2] E. L. Russell, L. H. Chiang, and R. D. Braatz, *Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes*, 1st ed. London: Springer London, 2000.
- [3] R. Isermann, *Fault-Diagnosis Systems: An Introduction From Fault Detection to Fault Tolerance*. Berlin; New York: Springer, 2006.
- [4] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A Review of Process Fault Detection and Diagnosis," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–311, 2003.
- [5] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, "A Review of Process Fault Detection and Diagnosis: Part II: Qualitative Models and Search Strategies," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 313–326, 2003.
- [6] P. M. Frank, "Fault Diagnosis in Dynamic Systems Using Analytical Knowledge-Based Redundancy – A Survey and Some New Results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.
- [7] S. J. Qin, "Survey on Data-Driven Industrial Process Monitoring and Diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, 2012.
- [8] V. Venkatasubramanian, "A Review of Process Fault Detection and Diagnosis: Part III: Process History Based Methods," *Comput. Chem. ...*, vol. 27, pp. 293–311, 2003.

- [9] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A Review on Basic Data-Driven Approaches for Industrial Process Monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6414–6428, 2014.
- [10] A. Cinar and C. Undey, "Statistical Process and Controller Performance Monitoring: A tutorial on Current Methods and Future Directions," *Proc. 1999 Am. Control Conf.*, vol. 4, no. June, pp. 2625–2639, 1999.
- [11] G. Li, S. J. Qin, and D. Zhou, "Geometric Properties of Partial Least Squares for Process Monitoring," *Automatica*, vol. 46, no. 1, pp. 204–210, 2010.
- [12] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002.
- [13] A. Negiz and A. Cinar, "Statistical Monitoring of Multivariable Dynamic Processes with State-Space Models," *Aiche J.*, vol. 43, no. 8, pp. 2002–2020, 1997.
- [14] M. N. Nounou and B. R. Bakshi, "On-Line Multiscale Filtering of Random and Gross Errors Without Process Models," *AICHE J.*, vol. 45, no. 5, pp. 1041–1058, 1999.
- [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, 2004.
- [16] B. R. Bakshi, B. R. Bakshi, S. Top, and S. Top, "Multiscale Statistical Process Monitoring and Diagnosis of Univariate and Multivariate Processes," *Control*, pp. 1–6, 1996.
- [17] B. R. Bakshi, "Multiscale Analysis and Modeling Using Wavelet," vol. 434, no.

- January, pp. 415–434, 1999.
- [18] H. B. Aradhye, B. R. Bakshi, R. A. Strauss, and J. F. Davis, “Multiscale SPC Using wavelets: Theoretical Analysis and Properties,” *AIChE J.*, vol. 49, no. 4, pp. 939–958, 2003.
- [19] A. Ikononopoulos and A. Endou, “Wavelet Application in Process Monitoring,” *Nucl. Technol.*, vol. 125, no. 2–3, pp. 225–233, 1999.
- [20] 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.
- [21] W. A. Shewhart, “Economic Quality Control of Manufactured Product,” *Bell Syst. Tech. J.*, vol. 9, pp. 364–389, 1930.
- [22] E. L. Grant and R. S. Leavenworth, *Statistical Quality Control*, 6th ed. New York: McGraw-Hill, 1988.
- [23] D. C. Montgomery, *Introduction to Statistical Quality Control.*, 7th ed. Hoboken, New Jersey: John Wiley & Sons, 2013.
- [24] A. Cinar, S. J. Parulekar, C. Under, and G. Birol, *Batch Fermentation : Modeling, Monitoring, and Control*, 1st ed. New York: CRC Press., 2003.
- [25] I. W. Burr, *Elementary Statistical Quality Control*, 1st ed. New York: M. Dekker, 1979.
- [26] E. S. Page, “Continuous Inspection Schemes,” *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [27] A. Cinar, A. Palazoglu, and F. Kayihan, *Chemical Process Performance*

Evaluation. Boca Raton, FL: CRC Press, 2007.

- [28] S. W. Roberts, "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, vol. 1, no. 3, pp. 239–250, 1959.
- [29] J. S. Hunter, "The Exponentially Weighted Moving Average," *J. Qual. Technol.*, vol. 18, no. 4, pp. 203–210, 1986.
- [30] S. Crowder and M. Hamilton, "An EWMA for Monitoring a Process Standard Deviation," *J. Qual. Technol.*, vol. 24, no. 1, pp. 12–21, 1992.
- [31] S. V Crowder, "A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts," *Technometrics*, vol. 29, no. 4, pp. 401–407, 1987.
- [32] J. M. Lucas and M. S. Saccucci, "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, vol. 32, no. 1, pp. 1–12, 1990.
- [33] A. S. Willsky, "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, vol. 12, no. 6, pp. 601–611, 1976.
- [34] J. M. Lucas and M. S. Saccucci, "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, vol. 32, no. 1, p. 1, 1990.
- [35] S. V. Crowder, "Design of Exponentially Weighted Moving Average Schemes," *J. Qual. Technol.*, vol. 21, no. 3, pp. 155–162, 1989.
- [36] R. A. Johnson and M. Bagshaw, "The Effect of Serial Correlation on the Performance of CUSUM Tests," *Technometrics*, vol. 16, no. 1, pp. 103–112,

1974.

- [37] A. V. Vasilopoulos and A. P. Stamboulis, “Modification of Control Chart Limits in the Presence of Data Correlation,” *J. Qual. Technol.*, vol. 10, pp. 20–30, 1978.
- [38] G. E. P. Box, *Time series analysis : Forecasting and Control*, 4th ed. Hoboken, NJ: John Wiley, 2008.
- [39] D. C. Montgomery, *Applied Statistics and Probability for Engineers*, 6th ed. Hoboken, NJ: John Wiley and Sons, Inc, 2014.
- [40] H. C. Thode, *Testing for Normality*, 1st ed. New York: Marcel Dekker, 2002.
- [41] H. M. Park, “Univariate Analysis and Normality Test Using SAS, Stata, and SPSS.,” *Indiana Univ. Inf. Technol. Serv.*, pp. 1–41, 2008.
- [42] K. P. Balanda and H. L. MacGillivray, “Kurtosis: A Critical Review,” *Am. Stat.*, vol. 42, no. 2, pp. 111–119, 1988.
- [43] O. Vasicek, “A Test for Normality Based on Sample Entropy,” *J. R. Soc. Ser. B(Methodological)*, vol. 38, no. 1, pp. 54–59, 1976.
- [44] D. Ramírez, J. Vía, I. Santamaría, and P. Crespo, “Entropy and Kullback-Leibler Divergence Estimation Based on Szego’s Theorem,” in *Signal Processing Conference, 2009 17th European*, 2009, pp. 2470–2474.
- [45] E. Seier, “Comparison of Tests for Univariate Normality,” *Interstat*, vol. 1, pp. 1–17, 2002.
- [46] S. G. Mallat, “Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$,” *Trans. Am. Math. Soc.*, vol. 315, no. 1, pp. 69–87, 1989.
- [47] S. G. Mallat, “A Theory for Multiresolution Signal Decomposition: The Wavelet

- Representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.
- [48] B. R. Bakshi, “Multiscale PCA with Application to Multivariate Statistical Process Monitoring,” *AIChE J.*, vol. 44, no. 7, pp. 1596–1610, 1998.
- [49] I. Daubechies, “Orthonormal Bases of Compactly Supported Wavelets,” *Commun. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, 1988.
- [50] S. Zhou, B. Sun, and J. Shi, “A SPC Monitoring System for Cycle-Based Waveform Signals Using Haar Transform,” *IEEE Trans. Autom. Sci. Eng.*, vol. 3, no. 1, pp. 60–72, 2006.
- [51] 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.
- [52] M. Z. Sheriff, F. Harrou, and M. Nounou, “Univariate Process Monitoring Using Multiscale Shewhart Charts,” in *International Conference on Control, Decision and Information Technologies*, 2014.
- [53] N. Lu, F. Wang, and F. Gao, “Combination Method of Principal Component and Wavelet Analysis for Multivariate Process Monitoring and Fault Diagnosis,” *Ind. Eng. Chem. Res.*, vol. 42, no. 18, pp. 4198–4207, 2003.
- [54] C. Rosen and J. A. Lennox, “Multivariate and Multiscale Monitoring of Wastewater Treatment Operation,” *Water Res.*, vol. 35, no. 14, pp. 3402–3410, 2001.
- [55] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, “Wavelet

- Shrinkage : Asymptopia?,” *J. R. Stat. Soc. Ser. B(Methodological)*, vol. 57, no. 2, pp. 301–369, 1995.
- [56] 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. Bioinforma.*, vol. 95, pp. 1539–1544, 2012.
- [57] M. Madakyaru, M. N. Nounou, and H. N. Nounou, “Linear Inferential Modeling: Theoretical Perspectives, Extensions, and Comparative Analysis,” *Intell. Control Autom.*, vol. 03, no. 04, p. 376, 2012.