REAL-TIME DETECTION OF WAVE PROFILE CHANGES

by

BEHNAM TAVAKKOL

B.S., Sharif University of Technology, Iran, 2010

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Industrial and Manufacturing Systems Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Approved by:

Major Professor
Shing I Chang

Abstract

This research studies a few methodologies for real-time detection of wave profile changes. In regular profile monitoring, change detection takes place at the end of time period when a complete profile is available. In real-time change detection of profiles, a potential profile change takes place between the beginning and the end of the time period. The decision involves the identification whether a process is in control or out of control before the entire profile is generated. In this regard, five proposed methodologies were developed and tested in this thesis.

Earthquake waves, manufacturing processes, and heart beat rate are a few examples of profiles with different natures that the proposed methodologies can be applied to. Water temperature profiles generated durning a curing process are considered as an example in this study. Successful implementation of the proposed work on these profiles would cause saving great amounts of time and money.

Five methods are studied for monitoring the water control process of a curing process. The first four proposed methodologies are based on an univariate approach where the statistic used for process monitoring is the enclosed area between the profiles and their fitted cutting lines. A multivariate approach is also proposed.

A simulation study is also conducted when the best method is chosen based on it performance and simplicity of operations. Various types of acceptable and unacceptable profiles are simulated for the best proposed method identified in the preliminary study. The best method has a satisfactory performance in detecting the changes in the unacceptable profiles. In addition, the false alarm rate in identifying acceptable profiles as bad profiles is lower than 10%.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Dr. Shing I. Chang for all his supports, understanding, guidance and encouragement throughout my research. I am thankful to Dr. Stanley Lee and Dr. Kun Chen for agreeing to participate as my committee members and all their supports throughout this research. Moreover, I would like to thank all my friends for their various assistances during the period of my studies at Kansas State University. The last but not the least, my greatest thanks goes to my parents without whose support I would not have been able to finish this.

Chapter 1-Introduction

1.1 Problem Statement

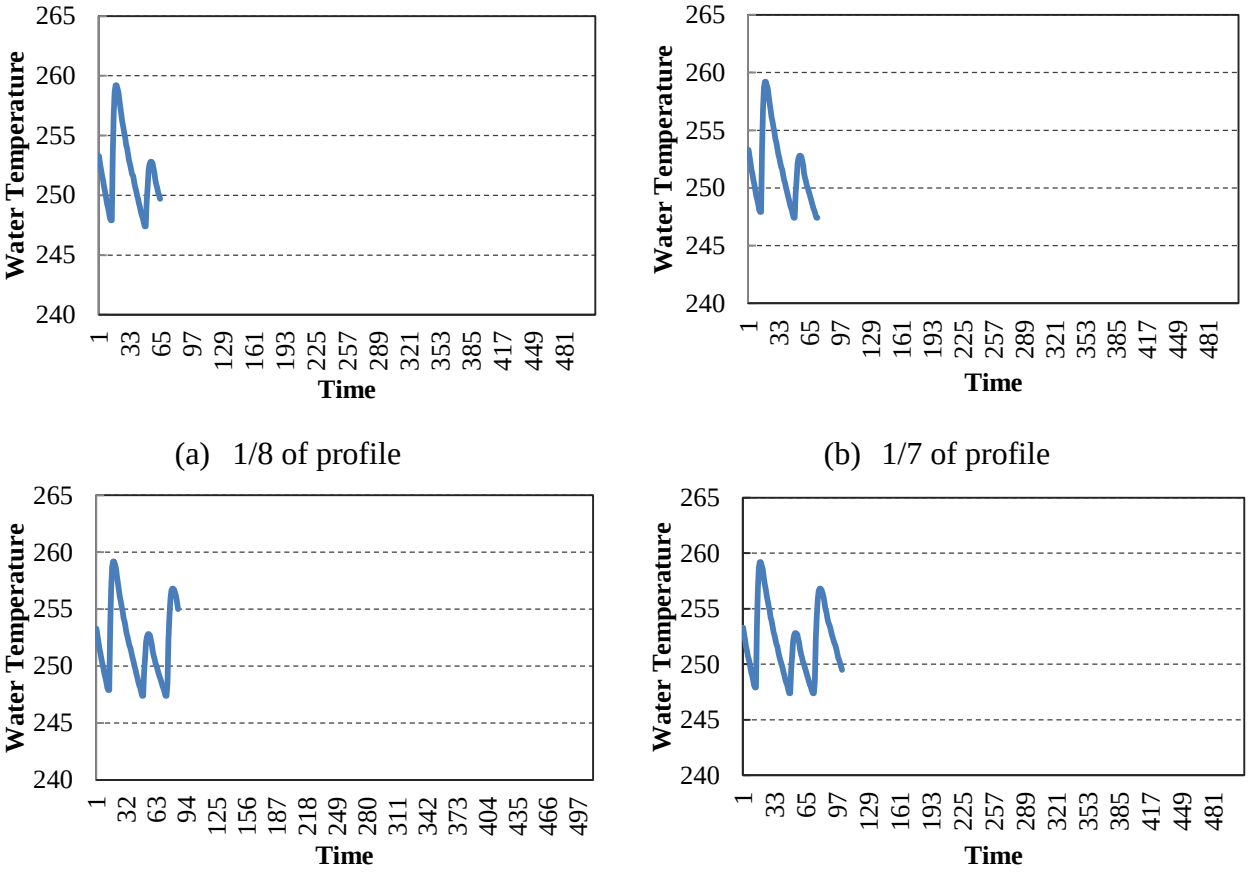
Profile monitoring has drawn much attention in the field of Quality Engineering in recent years. Profiles are time series over time or space. There are many characteristics that are captured in form of profiles. Earthquake waves, heart beat waves, temperature changes over time and many others are all examples of profiles. Comparing to univariate statistical process monitoring, profile analysis demands different tools and techniques. Profiles may have different shapes that depending on those, different regression models could be fitted to them. In this regard, Zhu, et al. (2010) and Zou, et al. (2007) studied the linear profiles while Williams, et al. (2007) and Ding, et al.(2006) studied nonlinear profiles. Examples of the research conducted about linear and nonlinear profiles will be brought in chapter two.

In profile analysis, the decision about the quality of a profile is usually made at the end of the period when a profile is completely generated. Most of the research conducted in the field of profile monitoring is based on this approach. In this study, a new approach in the profile monitoring area is proposed to detect profile changes based on real-time data feed. It would be extremely beneficiary to detect an irregular profile before the entire profile is generated. This would be of a great importance in many different fields. For example, in field of earthquake detection, as soon as a wave is detected as abnormal, appropriate warnings would be given and thus, the intensity of undesired consequences might get reduced. Manufacturing is another field that can take advantage of this approach. Producing a product may consist of several process parameters measured in form of profiles. Similar to the earthquake example, it is important to detect profile changes as soon as possible, which could save product cost through defect preventions.

It is of great importance to catch an out-of-control situation as fast as possible but there are some limits as well. It would be perfect if we were able to analyze the profiles continuously, but this requires extensive amount of data calculations and analyses. An alternative is to divide the entire profile into finite number of portions and then do the analysis at the end of each portion. Depending on the shape of profiles, different numbers of portions on a profile may be

required. Having more portions does not always translate into better detection performance. The optimal number of portions and locations of the points depends on each profile application. If there are many change patterns or many change locations that distinguish the bad profiles from the good ones, it may require more portions in general. Otherwise, a small number of portions may be adequate.

Figure 1.1 Completion of a water temperature profile

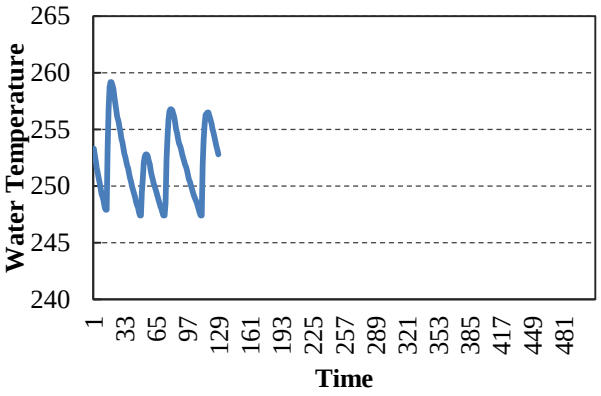


(a) 1/8 of profile

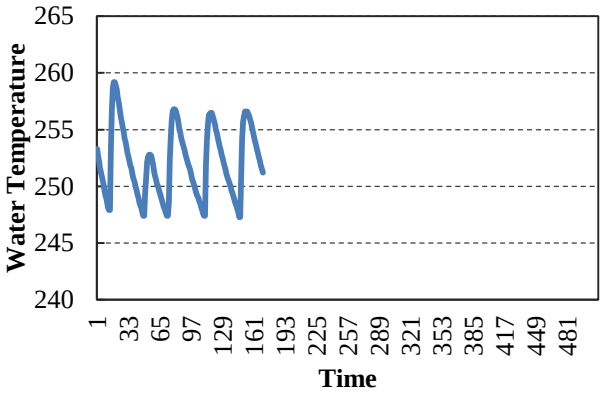
(b) 1/7 of profile

(c) 1/6 of profile

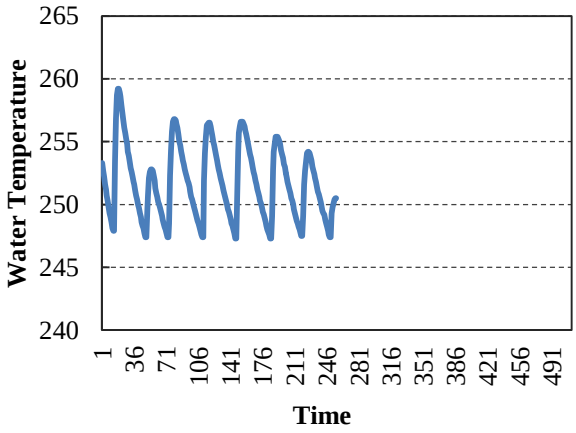
(d) 1/5 of profile



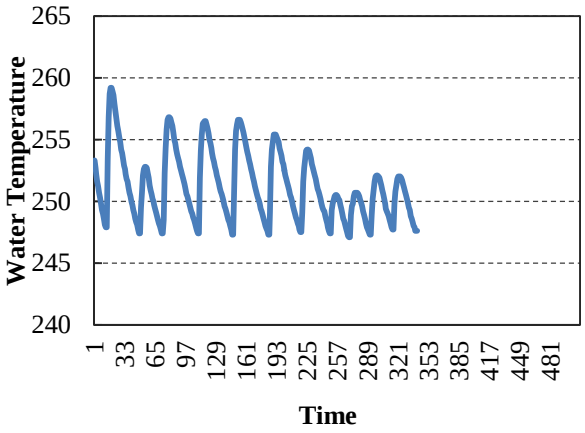
(e) 1/4 of profile



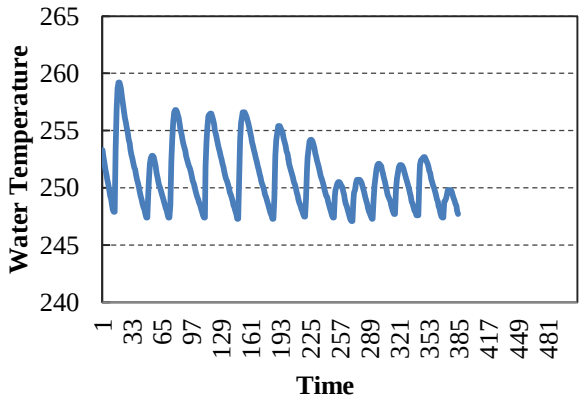
(f) 1/3 of profile



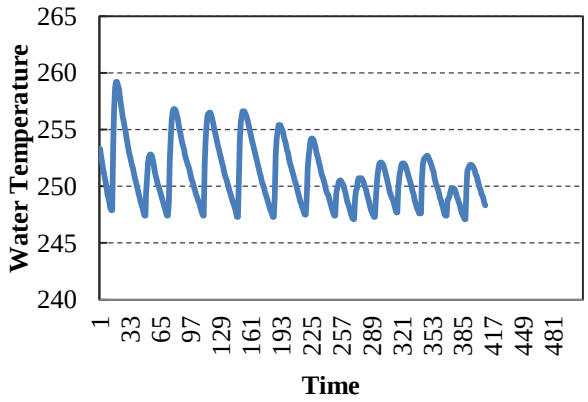
(g) 1/2 of profile



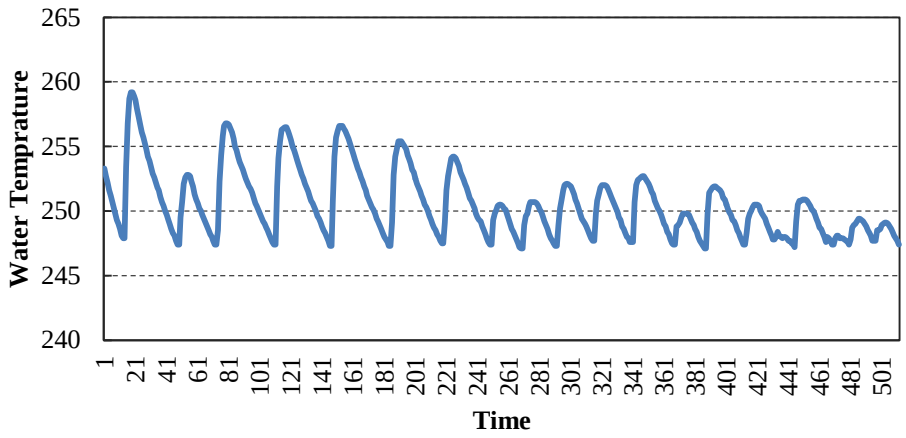
(h) 2/3 of profile



(i) 3/4 of profile



(j) 4/5 of profile



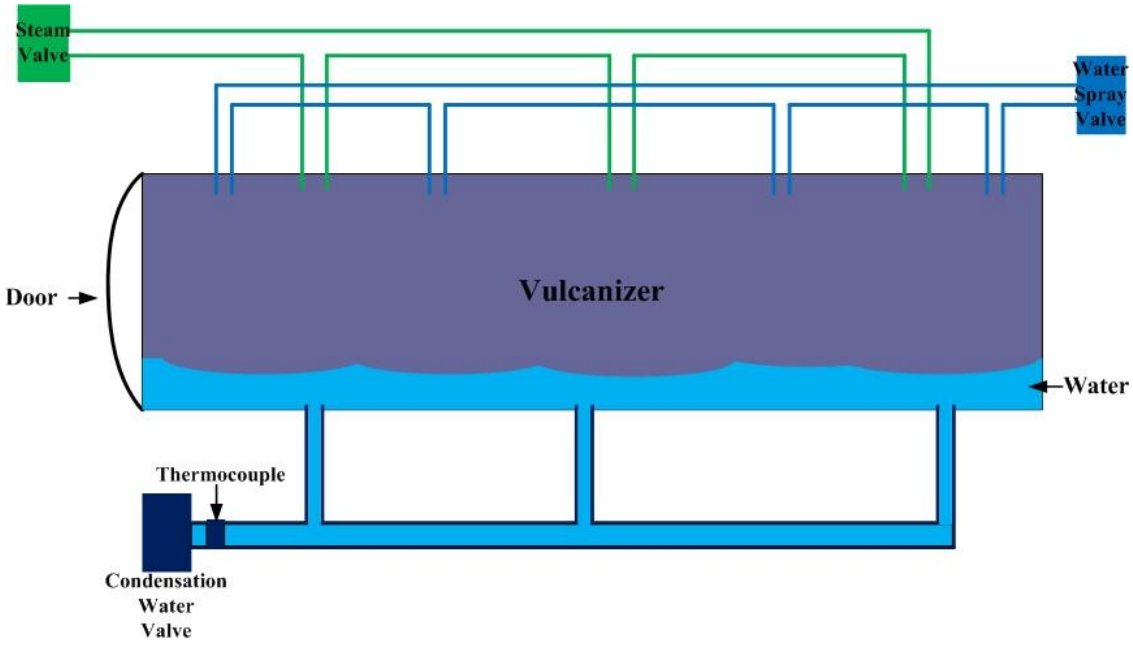
(k) Complete Profile

The aim of this thesis is to develop a s framework for detecting changes in profiles based on real-time data feed. Several methodologies will be studied to accomplish this goal. Figure 1.1-a through Figure 1.1-k display a water temperature profile divided into eleven portions. This profile is obtained from a curing process that will be illustrated as a case study in the following section. The complete profile is displayed in Figure 1.1-k. Considering the shape and pattern of this profile, eleven portions are considered for change detection analyses. The obtained profiles from this case study will be used throughout this thesis for testing and validating the proposed methodologies.

1.2 Case Study: The Curing Process

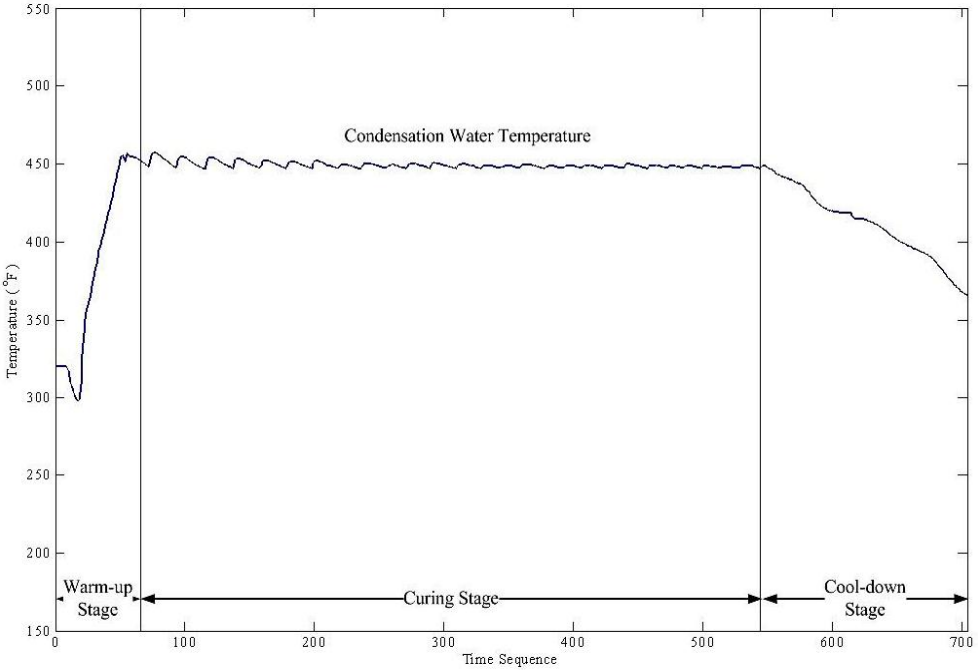
This case study is about one of the processes in producing high-pressure hose. This process is called curing. A curing process takes place in a sealed heat chamber called an autoclave or vulcanizer (vulcanization 2010). High-pressure hose products in reels are loaded into the vulcanizer. Then the vulcanizer is heated according to a curing recipe to reach a set temperature for a fixed amount of time. The housing that contains a vulcanizer is often called vessel. Figure 1.2 shows a schematic diagram of a vessel of a typical vulcanizer.

Figure 1.2 A schematic diagram of a vessel of a typical vulcanizer from Chang et al. **(Chang, Tsai and Lin)**



A release valve and a thermocouple are located at the bottom of the vessel. During a curing cycle, when high-pressure hose products are cured over time, moisture in the products gradually releases into the bottom of the vessel. This moisture accumulates into condensation water that pools at the bottom of a vessel. The condensation water temperature readings collected over time from the thermocouple form a profile that reflects proper on-and-off functions of the water-release valve. This profile consists of three stages: the warm-up stage, the curing stage, and the cool-down stage. Figure 1.2 displays these three stages of the condensation water temperature's profile.

Figure 1.3 Condensation water temperature profile of a typical curing cycle from Chang, et al. (2011)



The release valve is triggered when the water temperature drops below a pre-defined setting. Once the water is released, the water temperature at the valve will rise again and another cycle begins.

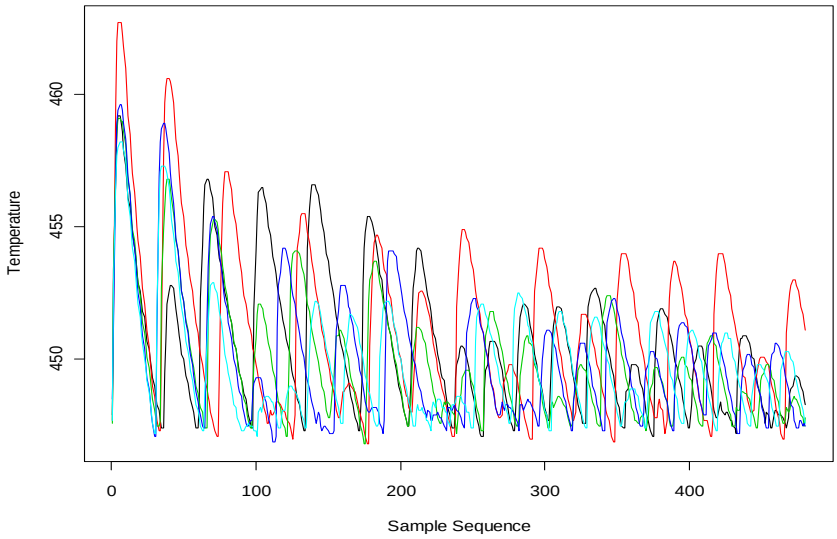
If the water valve operates normally, then it opens and closes routinely to keep the water level at the bottom of the vessel at an acceptable level. If the water-release valve operates infrequently, reels of hose products may immerse in accumulated water to cause either cosmetic or functional damage. If the water-release valve operates too frequently or opens at all times, it wastes energy due to escaping heat.

Most activities of the condensation water temperature take place during the curing stage. Therefore, the focus of this research is on this stage. Chang, et al. (2011) studied the implementation of a SPC to the curing stage of the condensation water temperature profile. Their aim was to detect profile changes after a complete profile was obtained while in this research the proposed SPC tool is implemented on a partial profile based on real-time data feed.

If the process is in control, it means that the water release valve is operating normally. Otherwise, the process is out of control, which means that the valve is either operating infrequently or too frequently. One of the research goals is to detect possible out-of-control operations as soon as possible before a complete profile is obtained.

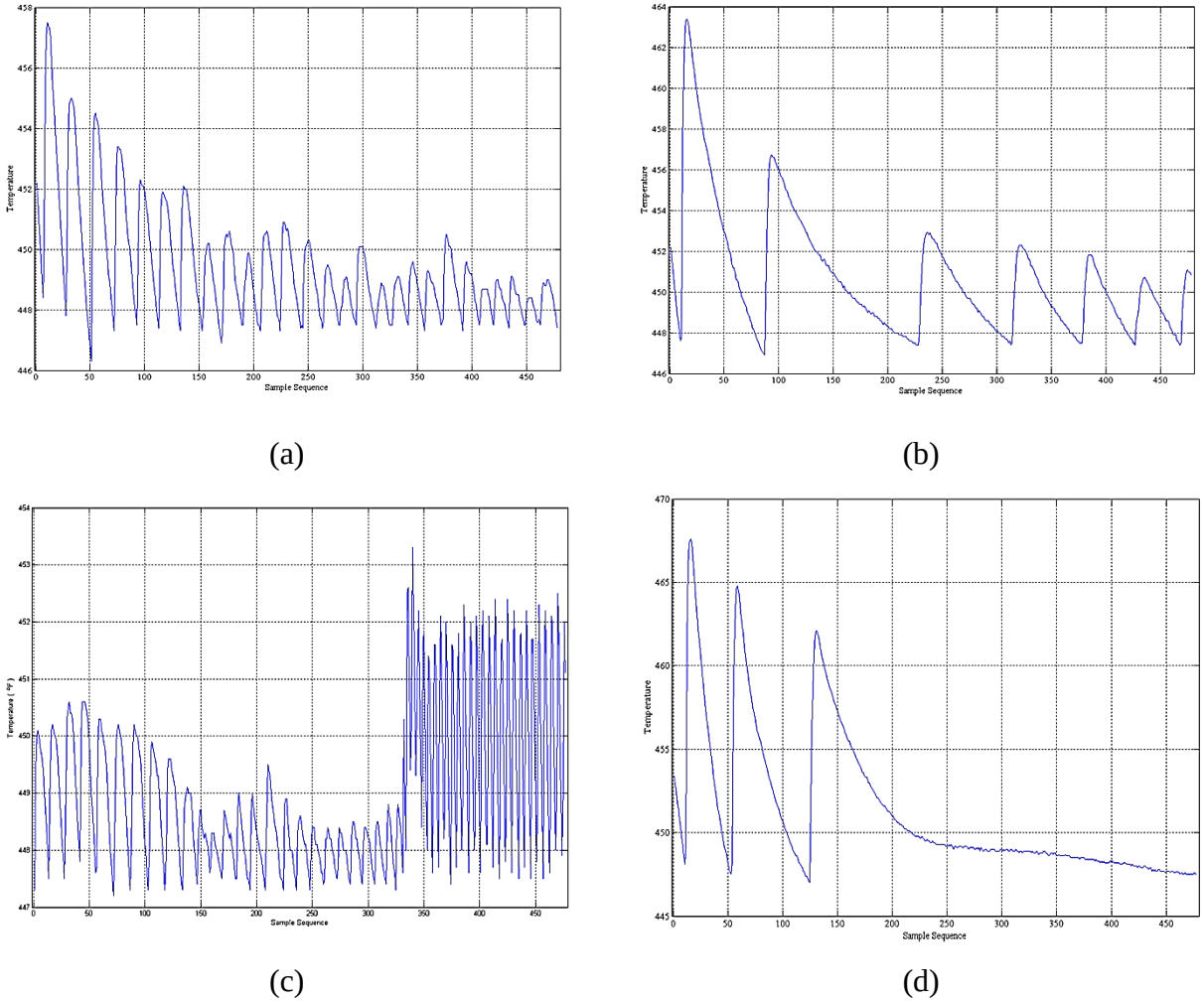
This case study presents a major challenge. There is no standard form as in-control form of the water temperature profile. In other words, there is no gold standard profile that we would be able to compare each profile with for process monitoring purposes. Figure 1.4 shows five different water temperature profiles that are all in control during the curing stage.

Figure 1.4 Five in-control wave profiles from Chang, et al. (2011)



As it's shown in Figure 1.4, the five in-control wave profiles do not have the same frequency and amplitude. Instead of finding a gold (standard) profile, quality engineers classify a set of profiles as acceptable or unacceptable. For example, the profiles of Figure 1.5-a and Figure 1.5-b are classified as acceptable and profiles of Figure 1.5-c and Figure 1.5-d are classified as unacceptable. One of main causes for the differences is the load of products that may contain different moisture contents Chang, et al. (2011).

Figure 1.5 Patterns of condensation water temperature profiles. (a) and (b) are acceptable , while (c) and (d) are unacceptable



(a) (b)

(c) (d)

1.3 Thesis Overview

This thesis contains the following chapters. In this chapter the problem statement was defined. A case study that will be used as a sample to test the proposed methodologies was also illustrated. In chapter two, a brief history of research related to profile analysis and process monitoring will be provided. Chapter three will include a few proposed methodologies for real-time profile change detection. The performance of each methodology in the illustrated case study will also be studied in the same chapter.

Chapter four will provide discussions and analyses of selecting the best approach introduced in chapter three. This chapter will also include a simulation study based on the best method selected. Finally, chapter five will contain conclusions and future research.

Chapter 2-Literature Review

Profile monitoring is the use of control charts for cases in which the quality of a process or product can be characterized by a functional relationship between a response variable and one or more explanatory variables [Woodall, 2007]. Profiles occur in many areas. A few examples of profiles such as earthquake waves were illustrated in the previous chapter. In terms of modeling approaches for profile monitoring, several categories are obtained. Woodall (2007) specifies linear regression, multiple and polynomial regression, non-linear regression, mixed models and the use of wavelets as a few categories in this regard. Extensive amounts of research have been conducted in all of the mentioned categories. A review of the most conducted current research will be summarized in the following categories: linear regression, nonlinear regression and mixed models.

2.1 Linear Regression

Linear regression approach is the one that is applied to the profiles that linear models are the best fit to them. Those profiles can be called as linear profiles. In the category of linear regression or in other words monitoring the linear profiles, the following researches have been conducted.

Mahmoud, et al. (2006) proposed a change point approach based on the segmented regression technique for testing the constancy of the regression parameters in a linear profile data set. Zhu, et al. (2010) proposed a Shewhart control chart for monitoring slopes of linear profiles. Zou, et al. (2007) proposed a multivariate exponentially weighted moving average monitoring scheme for linear profiles. Gupta, et al. (2006) compared the performance of two Phase II monitoring schemes for linear profiles. One of them is called NIST and is based on classical calibration method monitoring the deviations from the regression line and the other one is called KMW and is based on individually monitoring the parameters of the linear profile. Their simulation study revealed that the NIST performs poorly in comparison with the KMW method. Hosseinifard, et al. (2010) and Hosseinifard, et al. (2011) proposed a feed forward neural network to detect and classify drift shifts in linear profiles. In the field of quality control most

research focuses on developing control charts to monitor the product characteristics or process monitoring but monitoring the measurement gauges is also important because their performance affects the obtained results of measurements. In this regard, Chang, et al. (2006) developed a Shewhart chart to monitor the linearity between two measurement gauges.

2.2 Nonlinear Regression

Nonlinear regression approach is the one that nonlinear models provide the best fits. Those profiles can also be called as nonlinear profiles. In the category of nonlinear profiles, Chang, et al. (2010) proposed a framework to monitor nonlinear profiles. Their framework is able to identify mean shifts or shape changes of profiles. They first apply Discrete Wavelet Transformation to remove noise from the profiles and then use B-splines to generate critical points to define the shape of profiles. Their methodology also allows users to define number of segments that they would like to divide the profile into. When the studied process is determined to be out of control, the distance difference statistic for each segment provides diagnostic information. Hotelling T^2 chart is also used as the multivariate control chart to be fed by the proposed statistics (the same one used in this thesis). As a further analysis, in case of detecting out-of-control profiles, decomposition method could be applied to the T^2 statistics.

In another research, Ding, et al. (2006) conducted the Phase I analysis for monitoring nonlinear profiles in manufacturing processes. They introduced the high dimensionality and data contamination as the challenging components to the Phase I analysis of nonlinear profiles. They presented a two-component strategy to overcome those challenges: First, a data-reduction component that projects the original data into lower subspace while preserving the data-clustering structure and second, a data-separation technique that can detect single and multiple shifts as well as outliers in the data. Shiau, et al. (2009) proposed a method for monitoring the nonlinear profiles with random effects by nonparametric regression. They used the technique of principal components analysis for analyzing the covariance structure of the profiles. Based on the principal components scores they proposed a monitoring scheme. Kazemzadeh et al. (2008) developed three methods for monitoring polynomial profiles in Phase I. These three methods are called the Change Point Approach, F-Approach and, the T^2 Statistics. They also developed a method based on likelihood ratio test to identify the location of shifts. Williams, et al. (2007)

extended the use of T^2 control chart to monitor the coefficients resulting from a parametric nonlinear regression model fit to profile data.

2.3 Mixed Models

A mixed model is a statistical model containing both fixed effects and random effects. Mixed models are in several types. Linear mixed models and nonlinear mixed models are the most often studied ones.

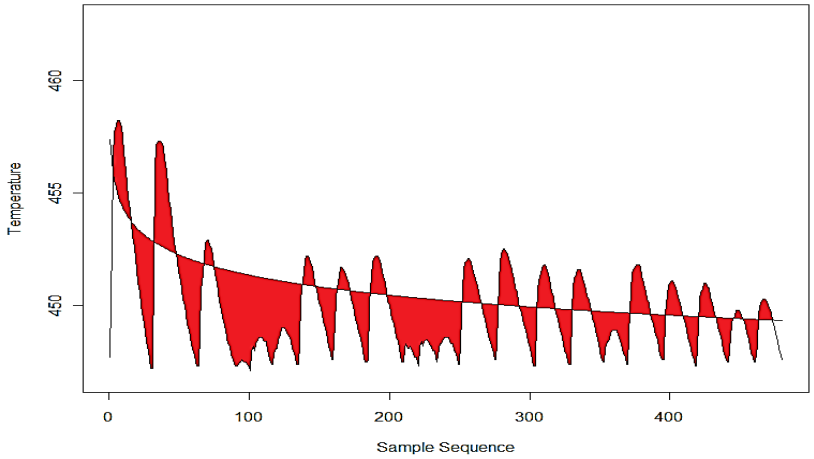
Jensen et al. (2008) proposed a method of fitting the profiles for data where the within-profile measurements are correlated with each other, thus relaxing the assumption of independent errors. They did so by fitting a linear mixed model (LMM), which allows accounting for the correlation within profiles. The LMM also allows considering the profiles as a random sample of profiles from a common population distribution, which may be a more realistic assumption than assuming that the profiles are completely independent of each other. To relax the restriction of uncorrelated measurements within a profile, Jensen et al. (2006) proposed the use of nonlinear mixed models to monitor the nonlinear profiles in order to account for the correlation structure.

There are also approaches consisting of mixed parametric and nonparametric models. Abdel-Salam et al. (2012) proposed a semi parametric mixed model approach to Phase I profile monitoring. Recently, in the absence of an obvious parametric (P) model, nonparametric (NP) methods have been employed in the profile monitoring context. For situations where a P model is adequate over part of the data but inadequate of other parts, Abdel-Salam et al. (2012) proposed a semi parametric procedure that combines both P and NP profile fits. They referred to their semi parametric procedure as mixed model robust profile monitoring (MMRPM). For each approach (P, NP, and MMRPM) they proposed a version of Hotelling T^2 statistic for use in Phase I analysis to determine unusual profiles based on the estimated random effects and obtain the corresponding control limits. Their simulation results showed that their MMRPM method performed well in making decisions regarding outlying profiles when compared to methods based on a misspecified P model or based on NP regression. They applied all three methods to the automobile engine data of Amiri et al. (2009) and found that the NP and the MMRPM methods indicated signals that did not occur in a P approach.

2.4 Background of The Current Research

In a research conducted by Chang, et al. (2011), the dataset from the curing process illustrated in section 1.2 was used. Different types of models such as 2nd order polynomial, 3rd order polynomial, B-spline and, exponential decay were fitted to some wave profiles. The exponential-decay function was used as the best function to fit the wave profiles. The sum of areas generated from the enclosure between the exponential-decay cutting line and the wave profile was used as a measure for further implementations. The wave profile and the fitted exponential decay function are reflected in Figure 2.1.

Figure 2.1 The enclosed area between the fitted exponential decay function and the profile



They denoted the sum of areas from m in-control samples of wave profiles as x_j'. A polynomial model in equation (2.1) was proposed to account for the number of peaks and valleys that were different from cycle to cycle.

$$log(x_j') = \beta_0 + \beta_1 z_j + \beta_2 z_j^2 + \epsilon_j, j = 1,2, \dots, m \quad (2.1)$$

where z_j is the number of waves of the jth profile.

A second-order polynomial model was then fit to remove noise due to the number of waves in each profile. Standardized residuals of this regression model form the quality characteristic for control charting:

$$e = log(x_i') - (\widehat{\beta_0} + \widehat{\beta_1} z_i + \widehat{\beta_2} z_i^2) \quad (2.2)$$

The standardized residuals should have been independent and normally distributed. The normality assumption was met. Moreover, an EWMA filter was applied to remove the

autocorrelation. The obtained statistic was then used for control charting purposes. IX control chart was used as an appropriate tool to monitor the residuals. Finally, their method could satisfactorily detect the time that the valve malfunction begins.

We have used the same data set in order to conduct a new research on the wave profiles. Once again, the research goal is to detect a profile change as fast as possible before a curing process ends. It means that we would like to identify an out-of-control wave pattern during, not after, a curing cycle.

2.5 Control Charting Tools

In this section, a few control charting tools that are popular in the field of Quality Engineering and were used in this research are introduced. Those are respectively IX control charts, EWMA charts and, Hotteling T^2. But before that, a brief introduction to the concept of control charts is brought: According to Montgomery (2008), the control chart is a graphical display of a quality characteristic that has been measured or computed from a sample versus the sample number or time. The chart includes three horizontal lines called as center line, upper control limit and, lower control limit. If the process is in-control, all of the sample points should fall between the upper and lower control limits. As long as this happens, the process is assumed to be in control, and no action is necessary. However, a point that plots outside of the control limits is interpreted as an evidence that the process is out of control, and an investigation and a corrective action are required to find and eliminate the assignable cause or causes responsible for this anomaly.

We may give a general model for a control chart. Let ω be a sample statistic that measures some quality characteristic of interest, and suppose that the mean of ω is μ_ω and the standard deviation of ω is σ_ω. Then the central line and the two control limits become:

$$\begin{aligned} UCL &= \mu_\omega + L\sigma_\omega \\ Center\ line &= \mu_\omega \\ LCL &= \mu_\omega - L\sigma_\omega \end{aligned} \tag{2.3}$$

where L is the distance of the control limits from the center line, expressed in standard deviation units. This general theory of control charts and control charts developed according to these principals are called Shewhart control charts.

2.5.1 EWMA Control Charts

The Shewhart control chart is a good tool for identifying the relatively big shifts in the mean of the process but it performs relatively poorly in detecting the small shifts. One of the good alternatives to the Shewhart control chart in detecting small shifts is Exponentially Weighted Moving Average control chart or EWMA chart. This control chart was introduced by Roberts (1959).

The exponentially weighted moving average is defined as

$$z_i = \lambda x_i + (1-\lambda) z_{i-1} \tag{2.4}$$

where $0 < \lambda \leq 1$ is a constant and the starting value (required with the first sample at i=1) is the process target, so that

$$z_0 = \mu_0 \tag{2.5}$$

Sometimes the average of preliminary data is used as the starting value of the EWMA, so that

$$z_0 = \bar{x} \tag{2.6}$$

The control limits of the EWMA are computed according to the equations (2.7) and (2.8):

$$UCL = \mu_0 + L\sigma_0 \sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]} \tag{2.7}$$

$$\text{CL} = \mu_0$$

$$LCL = \mu_0 - L\sigma_0 \sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]} \tag{2.8}$$

The term $\left[1-(1-\lambda)^{2i}\right]$ in equations (2.7) and (2.8) approaches unity as i gets larger. This means that after the EWMA control chart has been running for several time periods, the control limits will approach steady state values given by

$$UCL_t = \mu_t + L\sigma_t \sqrt{\frac{\lambda}{2-\lambda}} \tag{2.9}$$

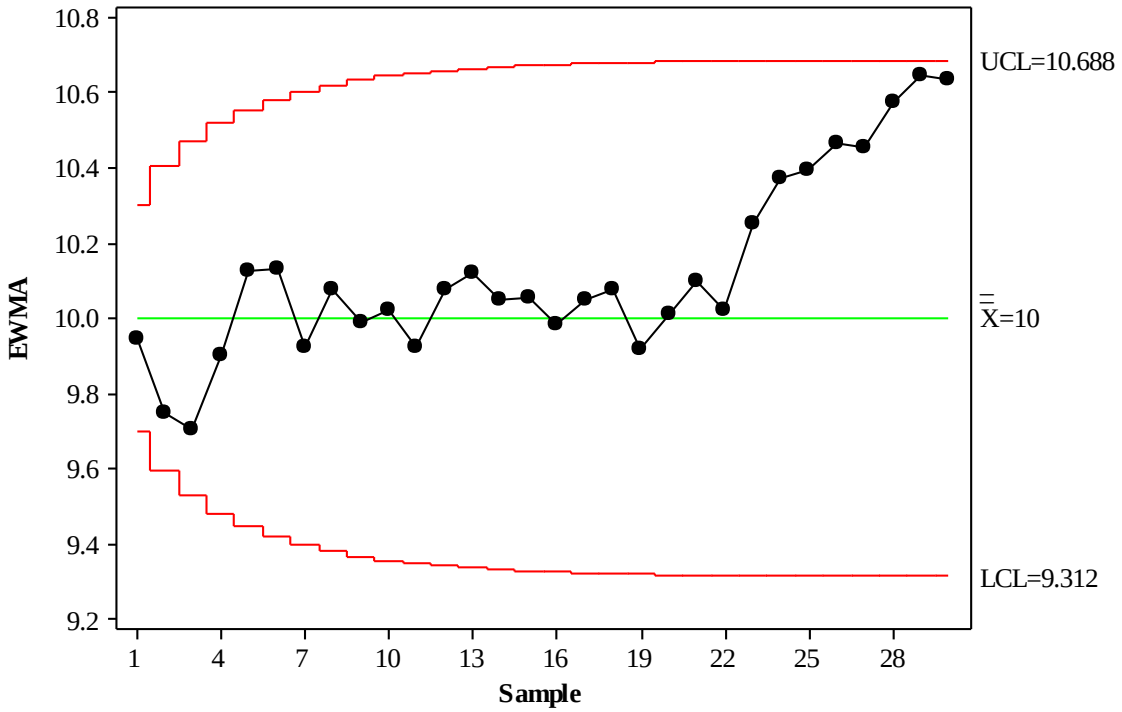
$$CL = \mu_t$$

$$LCL_t = \mu_t - L\sigma_t \sqrt{\frac{\lambda}{2-\lambda}} \tag{2.10}$$

Figure 2.2 displays an EWMA chart obtained from the data in Example 9.2 of the Montgomery (2008). This Figure is appropriately showing the control limits' approach towards the steady-

state values. It is also noticeable that the control limits at the beginning have smaller values. Moreover, all of the plotted points have fallen between the control limits. Thus, the process is assumed to be in-control.

Figure 2.2 EWMA control chart of Example 9.2 in Montgomery 2008



2.5.2 IX Control Charts

IX is one of the control charts that are used for monitoring the individual observations or in other words, situations in which the sample size is equal to one. Montgomery (2008) mentioned five examples for the situations in which IX chart are usable:

1. Automated inspection and measurement technology is used, and every unit manufactured is analyzed so there is no basis for rational subgrouping.
2. Data comes available relatively slowly, and it is inconvenient to allow sample sizes of $n > 1$ to accumulate before analysis. The long interval between observations will cause problems with rational subgrouping. This occurs frequently in both manufacturing and nonmanufacturing situations.

3. Repeat measurements on the process differ only because of laboratory or analysis error, as in many chemical processes.
4. Multiple measurements are taken on the same unit of product, such as measuring oxide thickness at several different locations on a wafer in semiconductor manufacturing.
5. In process plants, such as papermaking, measurements on the same parameter such as coating thickness across the roll will differ very little and produce a standard deviation that is much too small if the objective is to control coating thickness along the roll.

The main reason for the IX control charts to be used in this research is the one mentioned in example one. In this research, every profile or every segment in a profile is important and should be analyzed and none of the statistics of one profile can be combined with the ones of other ones. Thus, the samples sizes of greater than one are out of our interest and sample size of one will be considered for different analyses.

IX control charts consist of two different charts: Individual value chart and Moving range chart. The first one monitors the process mean while the second, monitors the variability. If a point plots outside of the control limits of each of these two charts, the process is assumed to be out-of-control and further investigation is needed. Equations (2.11) and (2.12) are the control limits of the individual value chart and moving range chart respectively.

$$\begin{aligned} UCL &= \bar{\bar{X}} + 3\frac{\overline{MR}}{d_2} \\ CL &= \bar{\bar{X}} \\ LCL &= \bar{\bar{X}} - 3\frac{\overline{MR}}{d_2} \end{aligned} \tag{2.11}$$

$$\begin{aligned} UCL &= D_4\overline{MR} \\ LCL &= D_3\overline{MR} \end{aligned} \tag{2.12}$$

Moving range at each point i is defined as $MR_i = |x_i - x_{i-1}|$. In many applications, the moving range two successive observations is used as the basis of estimating the process variability.

Later on in this research, having n=2, values of d_2, D_3 and D_4 are set as 1.128, 0 and 3.267 respectively.

2.5.3 Hotteling T^2 Control Charts

The Hotteling T^2 control chart is used for monitoring the multivariate processes. The fundamentals of the Hotteling T^2 are explained as follows. The Hotteling T^2 distance also called as Hotelling T^2 is a measure that accounts for the covariance structure of a multivariate normal distribution. The T^2 distance is a constant multiplied by a quadratic form. This quadratic form is obtained by multiplying the following three quantities:

1. The vector of deviations between the observations and the mean μ, which is expressed by $(X-\mu)'$,
2. The inverse of the covariance matrix, S^{-1},
3. The vector of the deviations, $(X-\mu)$.

It should be mentioned that for independent variables, the covariance matrix is a diagonal matrix and T^2 becomes proportional to the sum of squared standardized variables. In general, the higher the T^2 value, the more distant is the observation from the mean. The formula for computing the T^2 is:

$$T^2 = n(X-\mu)\, S^{-1}(X-\mu)' \tag{2.13}$$

The constant n is the sample size from which the covariance matrix was estimated. The T^2 distances lend themselves readily to graphical displays and as a result the T^2-chart is the most popular among the multivariate control charts. If we replace μ with $\bar{\bar{X}}$ in the equation (2.13), the obtained statistics from the new equation is used in the Hotelling T^2 control chart.

The Phase I control limits for the T^2 control chart are given by:

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$
$$LCL = 0 \tag{2.14}$$

In Phase II, when the chart is used for monitoring future processes, the control limits are as follows:

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$
$$LCL = 0 \tag{2.15}$$

In summary, in this chapter, a brief review of the profile analysis and the research in this area was introduced. Then, a few tools that are to be used in this research and are going to be

shown in the next chapters are introduced. As the subject of this research is new and there is no literature in real-time profile prediction, most of the literature here was related to the profile analysis in general.

Chapter 3-Proposed Methodologies

In this chapter, five methodologies for detecting the real-time changes of wave profiles are proposed and the results of applying each method on the water temperature data set are also obtained. These five methods include: Filtering then Standardizing, Standardizing then Filtering, Regression Approach, D method and finally, Multivariate analysis. Before that, a brief explanation of the process of selecting the profiles and the primary approach of obtaining the areas after fitting models to the profiles is brought.

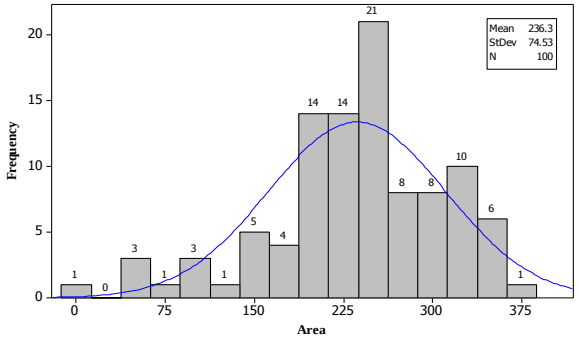
3.1 Data Selection and Preparation

The proposed methodologies in this research were all tested on the dataset obtained from the curing process demonstrated in chapter one. Out of the 188 wave profiles of the curing process obtained from March to April of 2011, 183 were selected because some profiles were obtained from abandoned curing cycles. The curing process takes a fixed time unit for which we will assign a number 100% cured or simply 1 without loss of generality. We divided this time frame into eleven portions represented by the fractions: 1/8, 1/7, 1/6, 1/5, 1/4, 1/3, 1/2, 2/3, 3/4, 4/5, and 1. For example, $\frac{1}{2}$ represents that a half of the wave profile or the data from the beginning to the midway point is obtained so far. As it was discussed in the previous chapter, the number of portions should be selected according to the pattern and location of the changes in bad profiles which distinguish them from the good profiles. In this regard, it seemed that eleven portions were enough for the analyses. However, other number of portions could be tested too to be able to obtain more general results. For this research, only eleven portions were tried. In the next step, 100 profiles were selected for Phase I analyses. The exponential-decay function was fitted to these 100 profiles. The sums of the areas generated from the enclosure between the exponential-decay cutting lines and the wave profiles were calculated for them. This calculation was done 11 times due to 11 different portions of the wave profile mentioned earlier. For example for the portion "½", the exponential-decay was fitted to the half of the original wave profile and then the sums of the areas were calculated.

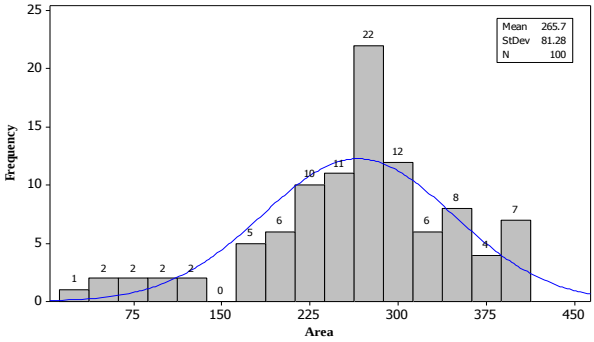
Figure 3-1-a through Figure 3-1-k show 11 histograms of frequency of wave profiles. Each histogram displays the number of wave profiles within different ranges of sum of areas for

its corresponding portion of data. In addition, Figure 3-1-a through Figure 3-1-k show the Anderson-Darling test results. The Anderson-Darling tests validate the normality assumptions.

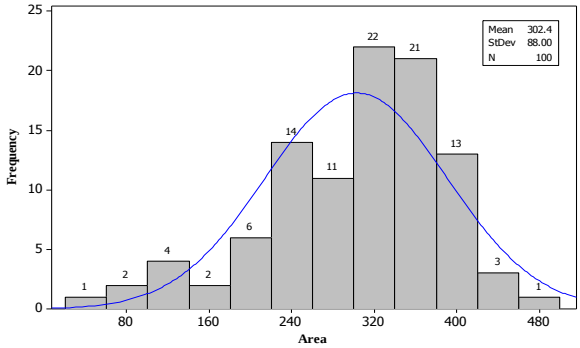
Figure 3-1 Frequency of wave profiles having different portions of profiles



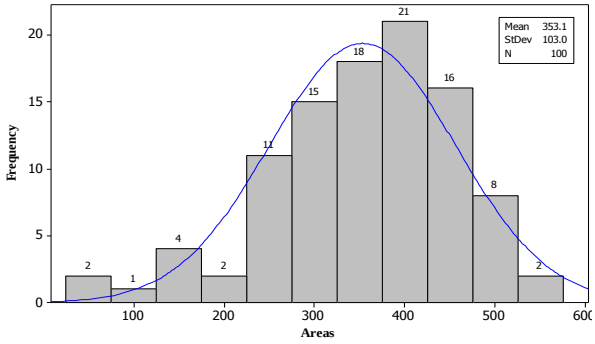
(a) 1/8 of profile



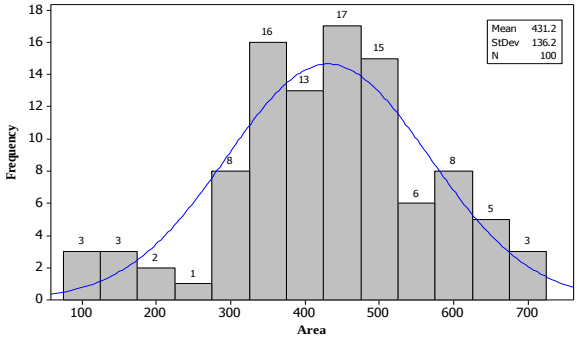
(b) 1/7 of profile



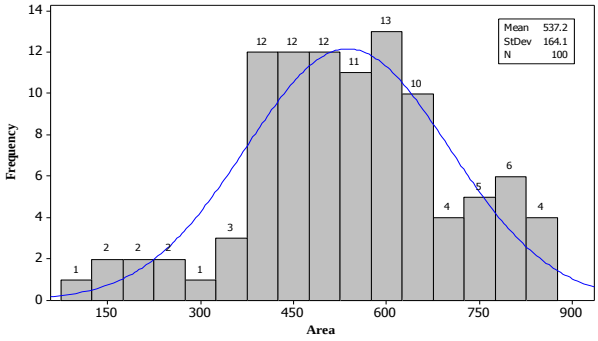
(c) 1/6 of profile



(d) 1/5 of profile

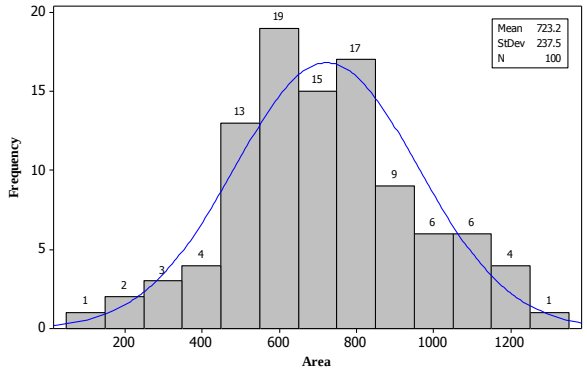


(e) 1/4 of profile

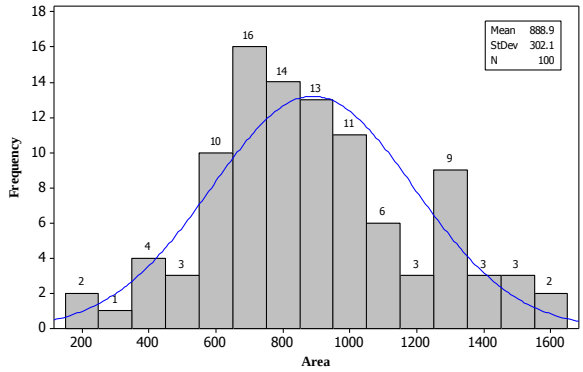


(f) 1/3 of profile

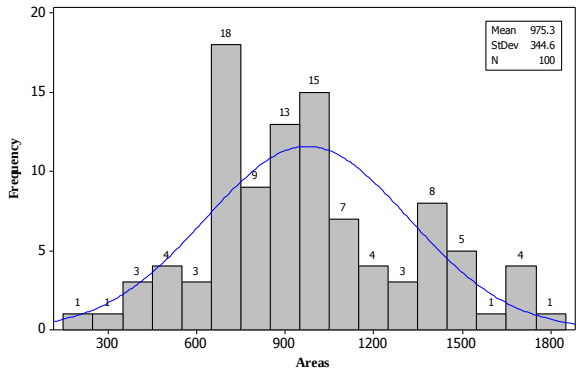
Figure 3-2 Frequency of wave profiles having different portions of profiles



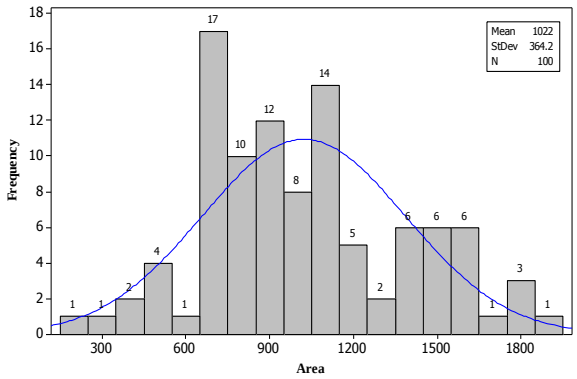
(g) 1/2 of profile



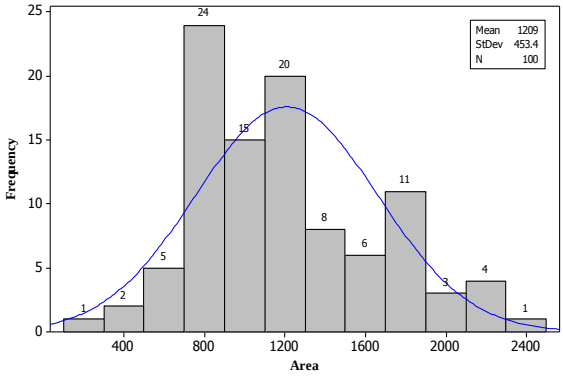
(h) 2/3 of profile



(i) 3/4 of profile



(j) 4/5 of profile



(k) all profile

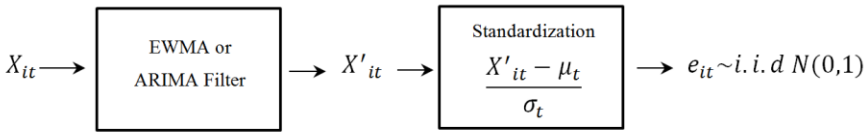
According to the normal distributions fitted to the 11 histograms and the Anderson-Darling test results, we concluded that the sums of areas for all the 11 portions do not follow normal distributions. This may be inferred because the p-values obtained by the Anderson-

Darling test are much less than the 0.05 value. Thus, we rejected the null hypothesis, which was the normality of the data. However, normality of the sums of areas for the 11 portions was not considered as a necessary assumption through different steps of this research.

3.2 First Method (Filtering then Standardizing)

A scheme of all the steps in the first proposed methodology is displayed in Figure 3-3. We considered variable X_{it} as the accumulated area for the profile i from the starting point up to the point t. The time index t could gain different values but in the special case of curing process we only defined eleven values of 1/8, 1/7, 1/6, 1/5, 1/4, 1/3, 1/2, 2/3, 3/4, 4/5, and 1 for it. It is clear that for any profile i, different values of X_{it} at different times were dependent and correlated. In order to remove the correlation from X_{it} we needed to apply a filter. Figure 3-4 shows a graph of autocorrelation function for X_{it} for a randomly selected profile i from the Phase I profiles of the curing process case before applying the filter. Figure 3-5 shows the partial autocorrelation for X_{it} for that profile.

Figure 3-3 A scheme of the first proposed methodology



According to Figure 3-4 and Figure 3-5 we could conclude that the best type of filter to remove the autocorrelation in this case would be an AR(1) filter. This was because of having a single spike at the beginning of the autocorrelation graph. Figure 3-6 and Figure 3-7 show the autocorrelation function and partial autocorrelation function for X_{it} for the randomly selected profile i after applying the AR(1) filter to the same previous profile. Variable X_{it} after applying the filter was named X'_{it}. As it is seen from Figure 3-6 and Figure 3-7 autocorrelation was totally removed from this specific profile after applying the filter. This shows that the filter did its job perfectly. The same procedure was applied to all of the profiles and X'_{it} variables were obtained.

Figure 3-4 Autocorrelation Function for $\boldsymbol{X_{it}}$ for a randomly selected Phase I profile i

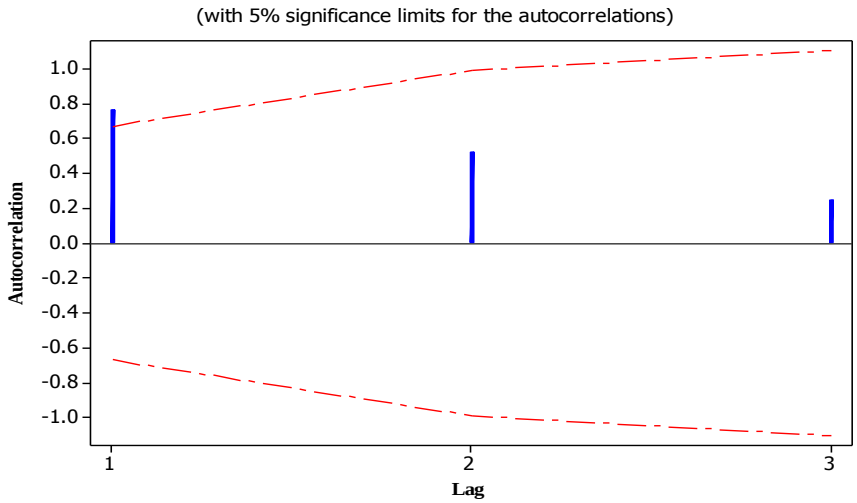


Figure 3-5 Partial Autocorrelation Function for $\boldsymbol{X_{it}}$ for a randomly selected Phase I profile i

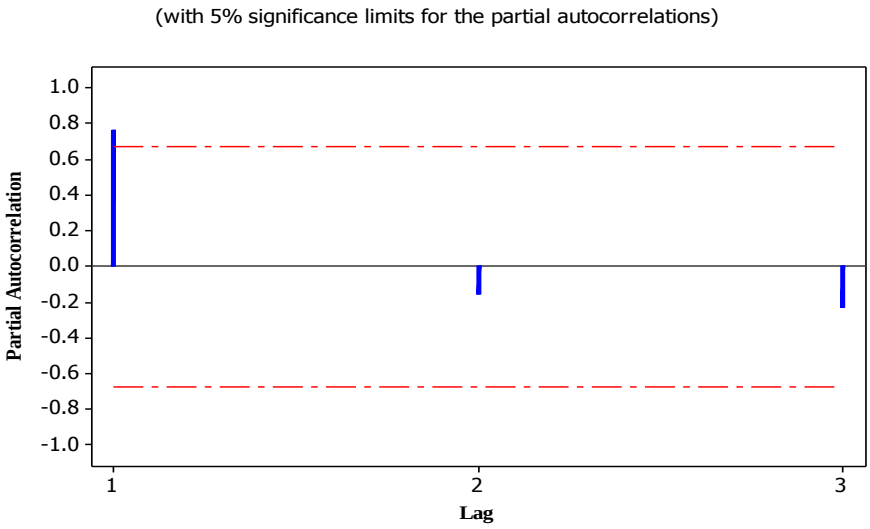


Figure 3-6 Autocorrelation Function for X'_{it} for one randomly selected profile

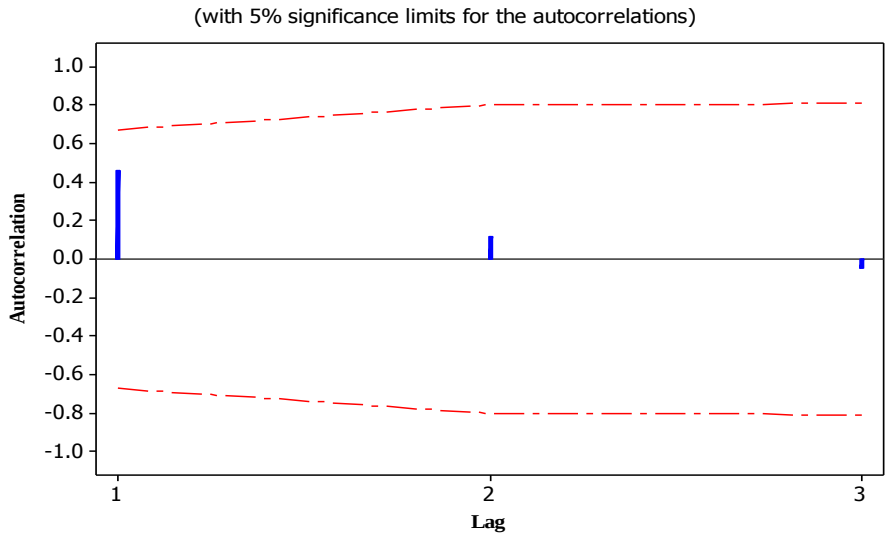
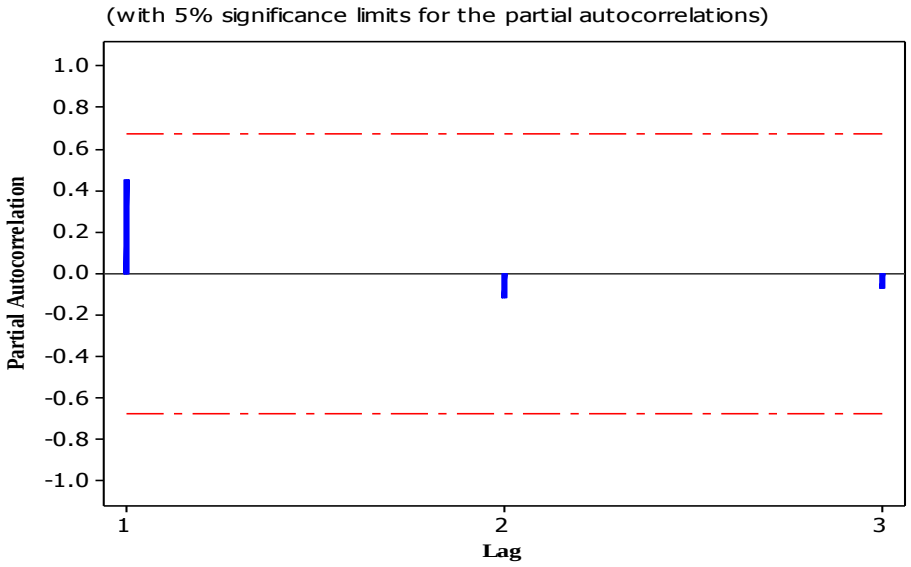


Figure 3-7 Partial Autocorrelation Function for X'_{it} for one randomly selected profile



The next step was standardization. Standardization was conducted according to the following formula:

$$\frac{X'_{it} - \mu_t}{\sigma_t} = e_{it} \tag{3.1}$$

where X'_{it} was the area after the filtering for profile i at time t, μ_t was the mean of the filtered areas of 100 phase I profiles at time t , and σ_t was the standard deviation of the filtered areas of 100 Phase I profiles at time t.

Again, in the example of curing process, t obtained values of 1/8, 1/7, 1/6, 1/5, 1/4, 1/3, 1/2, 2/3, 3/4, 4/5, and 1. The output of the above formula was called e_{it} and was supposed to follow $N(0,1)$. The next step was to apply a control chart on each profile. This means applying a control chart on values of e_t at different t's.

A good and useful control chart for monitoring this process would be IX control chart. IX is one of the control charts which are used for monitoring the individual observations. Since we were not interested in grouping the e_{it}'s of profile i, we would like to study them as individual observations. As we knew that e_{it}'s of profile i were independent and not correlated, IX control charts were applicable for monitoring them. Control limits for the IX control were defined in equations (2.11) and (2.12).

Having n=2, values of d_2, D_3 and D_4 are set as 1.128, 0 and 3.267 respectively. An important aspect in setting the control limits is the consistency between them in all of the profiles. Regularly, the control limits for each profile are set individually using the information of that particular profile and according to abovementioned formulas. In order to have consistent control limits for the Phase I profiles, one way was to calculate the mean and standard deviation of e_{it} values for each profile and then consider the mean of the means of all of the Phase I profiles and also the mean of the standard deviations of all of the Phase I profiles. The obtained statistics could then be used as the input values for setting the consistent control limits.

In this research, the control limits were set as -3, 0, and 3 for the individual value chart and also for the moving range chart they were set as 0, 1.128, and 3.686. The reason for this was to obtain values for the mean of the means and for the mean of the standard deviations that were approximately equal to 0 and 1.

Figure 3-8 shows one of the detected out of control profiles of Phase I of the curing process by using the IX control chart. Figure 3-9 shows the related IX control chart. As it is seen from the Figure 3-9, this profile was identified as out of control from the beginning observation. In addition, later on it signaled out again a little after middle of the process.

Figure 3-8 One correctly identified out of control profile

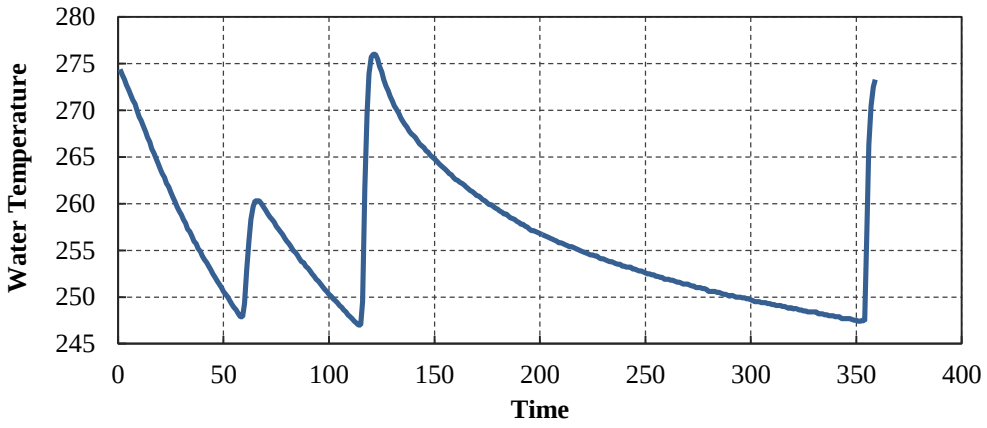
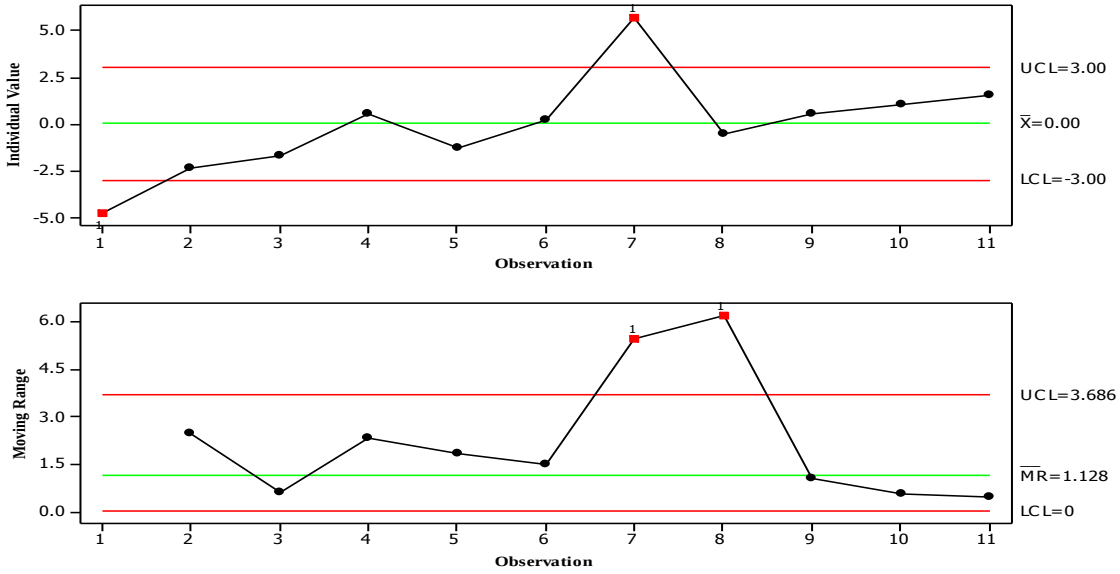


Figure 3-9 IX control chart for one correctly identified out of control profile



Further investigation in later parts of this research revealed that a few profiles which this profile was also among them, should not have been selected as Phase I profiles. This is due to the aim of the Phase I which was to train the system with in-control profiles.

Totally, there were three out of control profiles in Phase I profiles with the same shape as the one in Figure 3-8. IX control chart could successfully identify two of them. One of them which is shown in Figure 3-10 could not be identified. Figure 3-11 shows the IX control chart for that profile.

Figure 3-10 An unidentified out-of-control profile using the first method

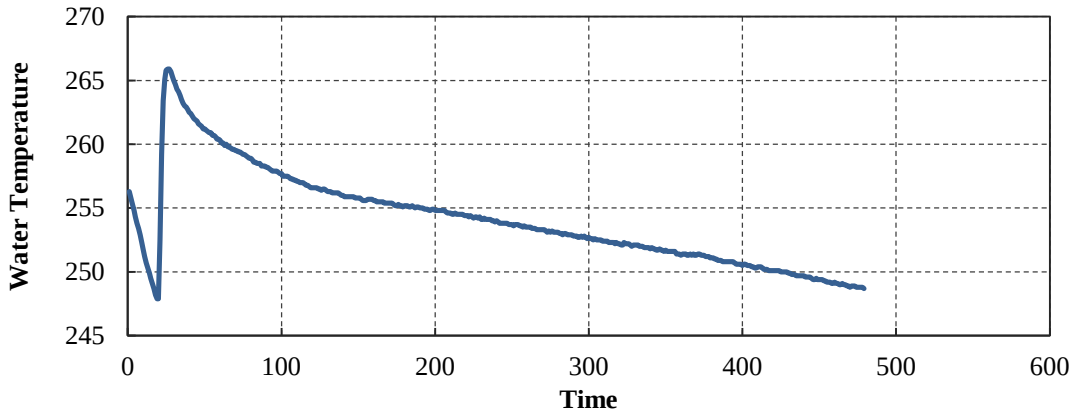
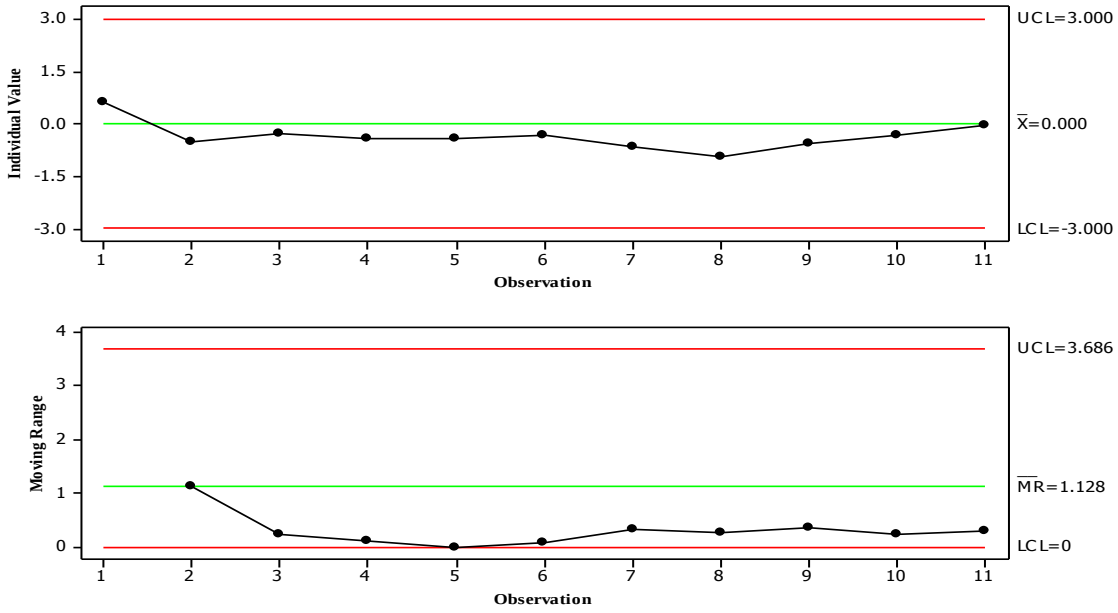


Figure 3-11 IX control chart for one unidentified out of control profile using the first method



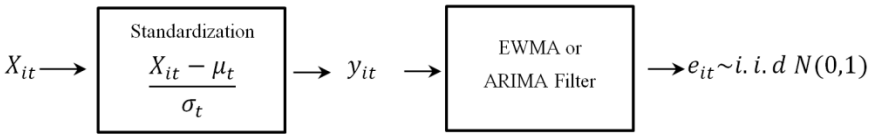
This probably happened because of the unusual shape of that profile. It spiked at the beginning but descended towards the end. This caused that the enclosed area obtained after fitting an exponential decay function be a relatively small value and did not differ much from the ones for good profiles. There were also a few more out-of-control cases with different shapes, which the IX chart was unable to recognize. In summary, having eight out of control profiles still in Phase I, IX control chart could identify two of them correctly and missed the 6 remaining. In addition, it identified eight in-control profiles as out of control ones incorrectly. As a rework, all of the undesirable and doubtful profiles were removed from the Phase I, and the analysis was

conducted again. The removed profiles were used again as inputs for Phase II. The goal then was to test if the proposed methodology was able to identify all of these profiles as out-of-control. It turned out that only two of them could be detected. Therefore, it was revealed that although it was correct to train the control chart with only in-control profiles, in this particular case of curing process, there would not be much difference in the obtained results if we would consider those few profiles in Phase I. Having poor performance in detecting the out-of-control profiles in companion with the large number of false alarms, led us to try other methods for getting better results.

3.3 Second Method (Standardizing then Filtering)

Now the second method that is standardizing the data then applying the filter is illustrated. Figure 3-12 displays a scheme of the second methodology. In this regard, at first, following the below formula the data at each of the eleven time points were standardized.

Figure 3-12 A scheme of the second proposed methodology



$$y_{it} = \frac{X_{it} - \mu_t}{\sigma_t} \tag{3.2}$$

For equation (3.2): $t = \frac{1}{8}, \frac{1}{7}, \ldots, 1$, $i = 1,2, \ldots, 100$, X_{it} was the original data of area for profile i at time t and y_{it} was the standardized variable. μ_t and σ_t were the mean and standard deviation of X_{it}'s of 100 Phase I profiles. Figure 3-13and Figure 3-14 show the ACF and PACF charts for a randomly selected Phase I profile. As it is seen from these Figures, there was correlation between the data in this profile.

Figure 3-13 ACF chart for a randomly selected Phase I profile

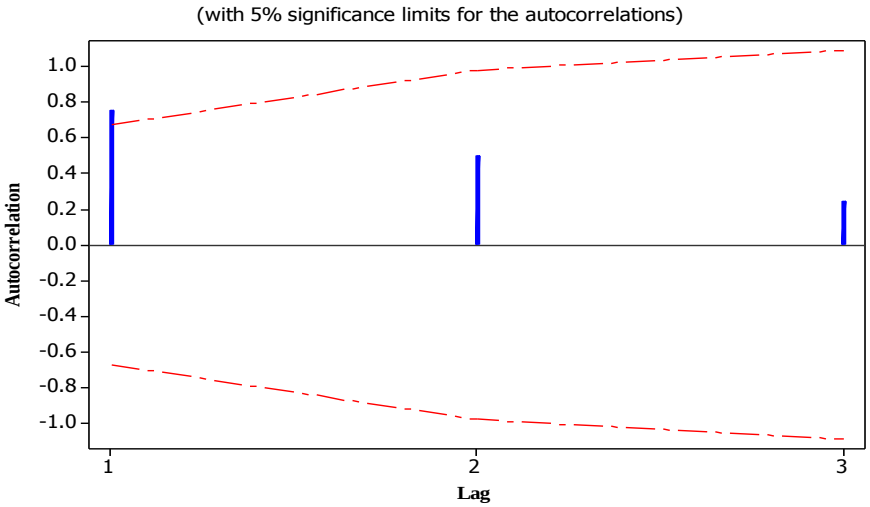
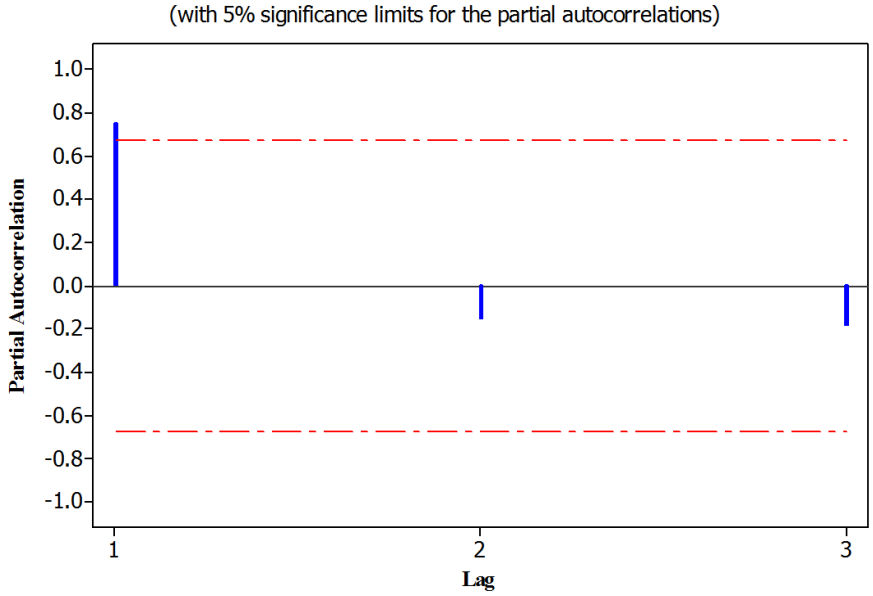


Figure 3-14 Partial autocorrelation function chart for the randomly selected Phase I profile



The ACF and PACF charts recommended AR1 as an appropriate model to filter the correlation with. The first spike in the autocorrelation was the main reason of choosing this type of filter. Thus, the AR1 model was applied to the y_{it}'s. Figure 3-15 and Figure 3-16 show the autocorrelation function and partial autocorrelation function after applying the AR1 filter for a randomly selected profile. As it is seen from these Figures, autocorrelation was removed from the profiles at different times.

Figure 3-15 ACF chart for a randomly selected Phase I profile

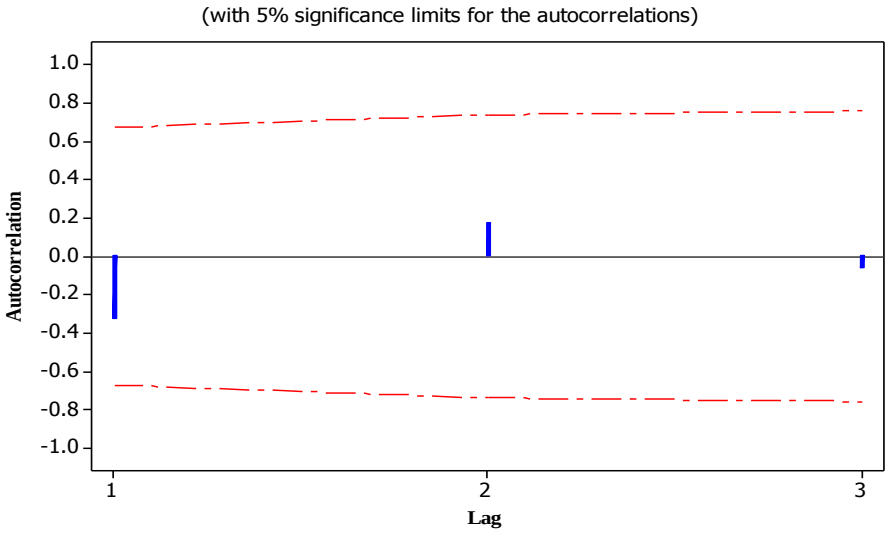
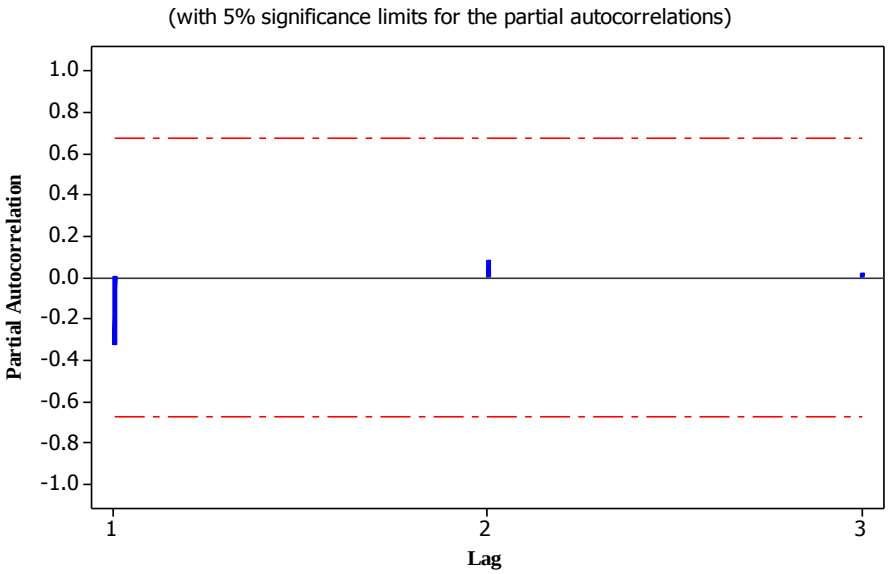


Figure 3-16 PACF chart



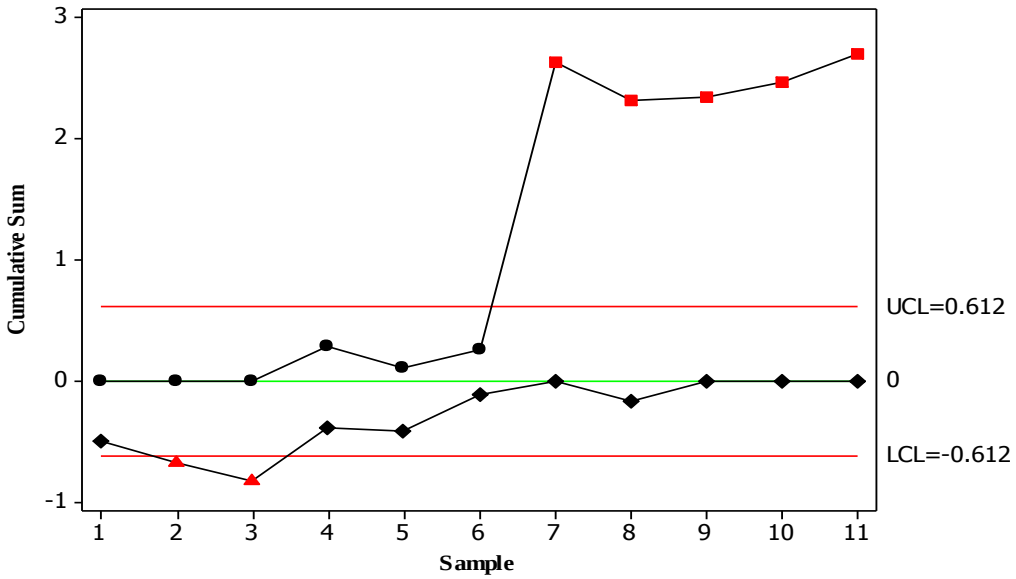
Now the appropriate data was ready for further analysis and applying the control charts. Two series of CUSUM charts were tried. The first one was called the CUSUM on mean, which was applying the CUSUM on the latest residuals and the second one was called the CUSUM on variance, which was applying of the CUSUM on the output of following formulas:

$$z_{it} = \frac{e_{it} - \mu_i}{\sigma_i} \tag{3.3}$$

$$v_{it} = \frac{\sqrt{|z_{it}|} - 0.822}{0.349}$$

We had to find appropriate fixed control limits to apply on all of the Phase I profiles. Finding those limits was crucial. As we know, CUSUM control limits depend on the amount of variation that we would like to detect. For the CUSUM on mean chart, we set the variation as the average of the standard deviations of all the 100 profiles in Phase I. This equaled to 0.153015. Figure 3-17 shows the CUSUM on mean for one of the correctly identified out of control profiles. For the Phase I, using CUSUM on mean method, and having the control limits of UCL=0.612 and LCL=-0.612, four out of eight real out of control profiles were identified correctly. In addition, this method identified six profiles incorrectly. In other words, it gave six false alarms.

Figure 3-17 The CUSUM on mean for profile 86, a real out of control profile

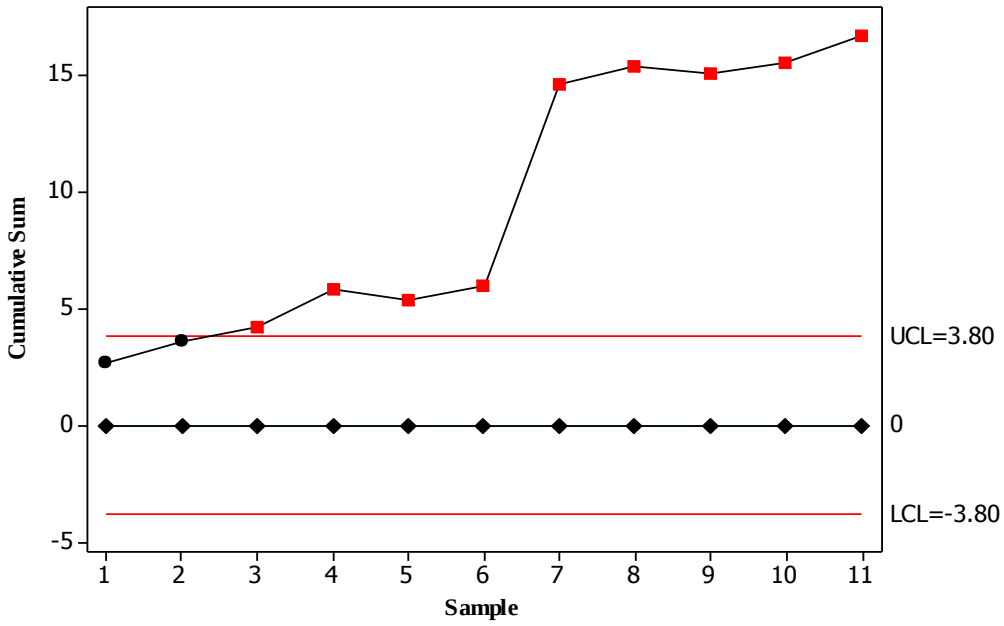


For the CUSUM on variance chart, we set the variation as the average of the standard deviations of all the 100 profiles in Phase I. This equaled to 0.9511. Figure 3-18 shows the CUSUM on variance for one of the correctly identified out of control profiles. For the Phase I, using CUSUM on mean method, and having the control limits of UCL=3.805 and LCL=-3.805,

four out of eight real out of control profiles were identified correctly. In addition, this method gave seventeen false alarms.

It turned out that this control chart (CUSUM on variance) was also not applicable for the purpose of this research that was prediction of the profile. Because it used the values of z_{it} that were calculated having the mean and standard deviation of complete profiles while they were not available at the times before the end of profile.

Figure 3-18 The CUSUM on variance for profile 86, a real out of control profile

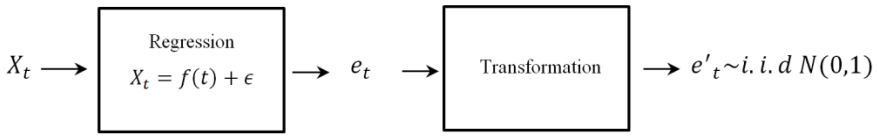


The same as the first method, the results were not satisfactory enough to consider this method as a major method for monitoring the performance of each of the profiles. Thus, there was still a need to try other methods.

3.4 Third Method (Regression Model)

The regression approach is now demonstrated here. A scheme displaying different steps of this methodology is brought in Figure 3-19. A simple linear model was fitted to the 100 Phase I profiles. Since in the curing process case study each profile consisted of 11 points by itself, the X matrix would be a 1100*2 matrix.

Figure 3-19 A scheme of the third proposed methodology



$$X = \begin{bmatrix} 1 & 1/8 \\ 1 & 1/7 \\ 1 & 1/6 \\ 1 & 1/5 \\ 1 & 1/4 \\ 1 & 1/3 \\ 1 & 1/2 \\ 1 & 2/3 \\ 1 & 3/4 \\ 1 & 4/5 \\ 1 & 1 \\ . & . \\ 1 & 1 \end{bmatrix}_{1100\times 2} \tag{3.4}$$

Plot of the obtained residuals after fitting the first order model versus the estimates of the areas ($\hat{x}$) is brought in Figure 3-20.

As it is seen from the Figure 3-20 the variance of the residuals increased tremendously as the $\hat{x}$ increased. In order to get the residuals with a relatively more fixed variance, the Box-Cox transformation method was applied on the original data of the areas. Figure 3-21 shows the Box-Cox transformation results. The λ=0.1 was picked as the best value for transformation. Figure 3-22 shows the residuals after fitting the linear model on the transformed values of y versus the estimates of the transformed values of y.

Figure 3-20 Plot of the obtained residuals versus the estimates of the areas

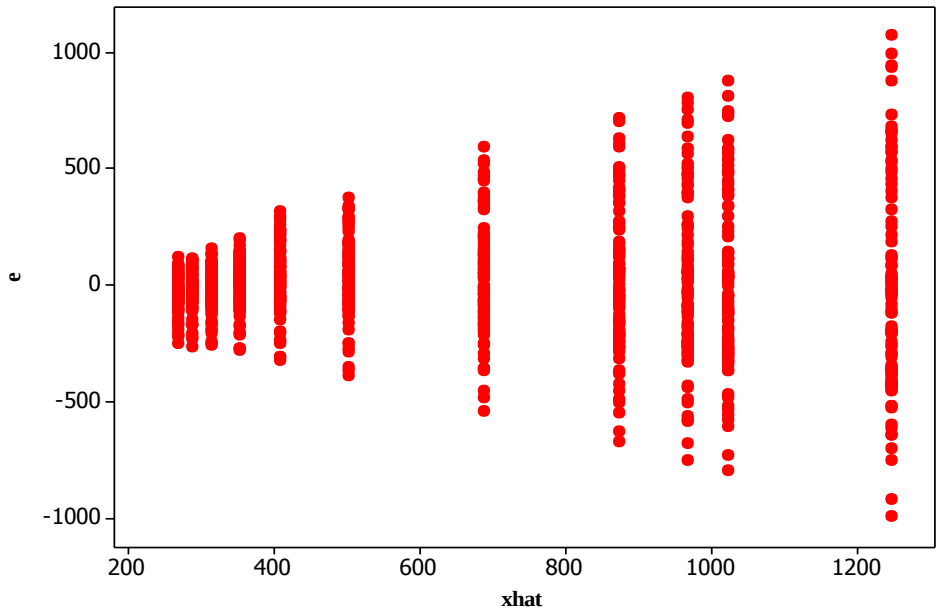


Figure 3-21 The Box-Cox Transformation

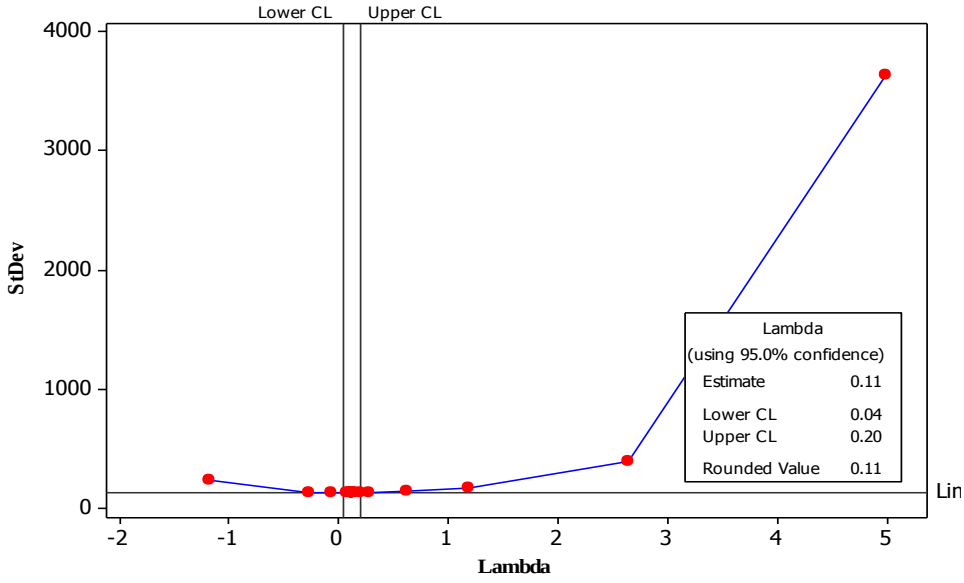
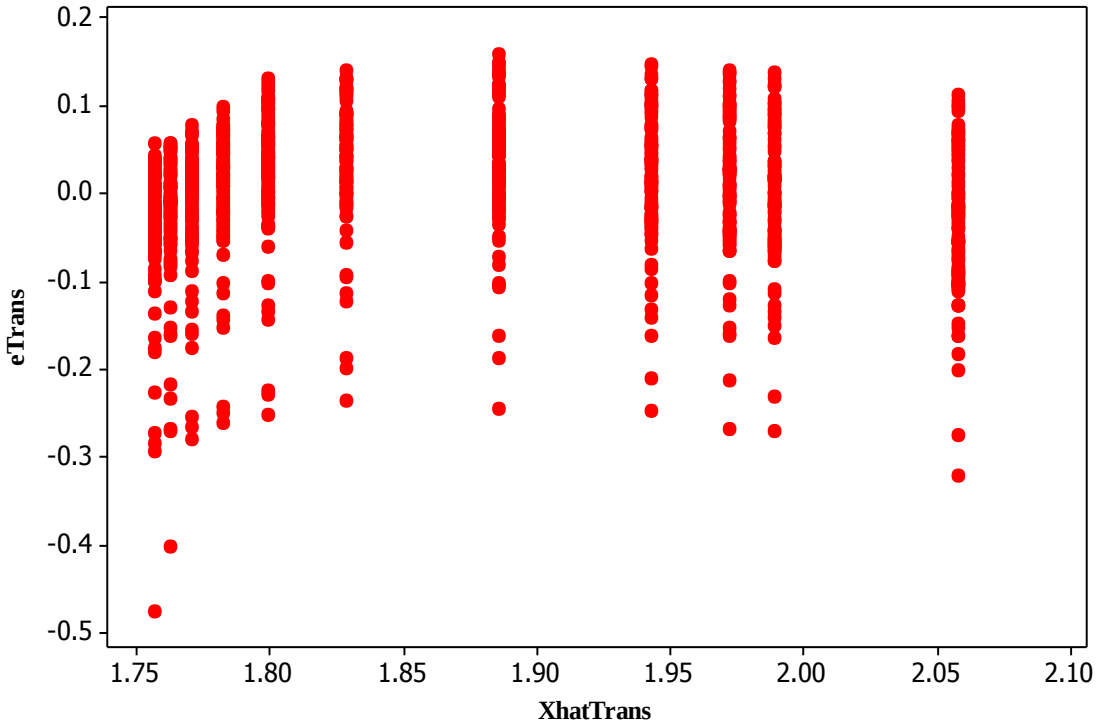


Figure 3-22 Plot of the obtained residuals after transformation versus the estimates of the transformed areas



By comparing Figure 3-20 with Figure 3-22 it is seen that the variety in variance of the residuals over the time after transformation was much less than the one before the transformation. In addition, before the transformation the residuals used to fall in the bonds of ranging from -1000 to 1000 but after transformation they fell in the intervals of -0.5 to 0.2. Figure 3-22 shows that the residuals derived after fitting the linear model and without applying the transformation to the *y* values were not following a normal distribution but had the mean of zero. Figure 3-23 shows that the residuals derived after fitting the linear model to the transformed y values also did not follow normal distribution but had mean of zero. This may have been because of fitting a linear model to the data. If we applied a second-order model, we may have gotten the normal distribution.

Figure 3-23 Probability plot of residuals without transformation

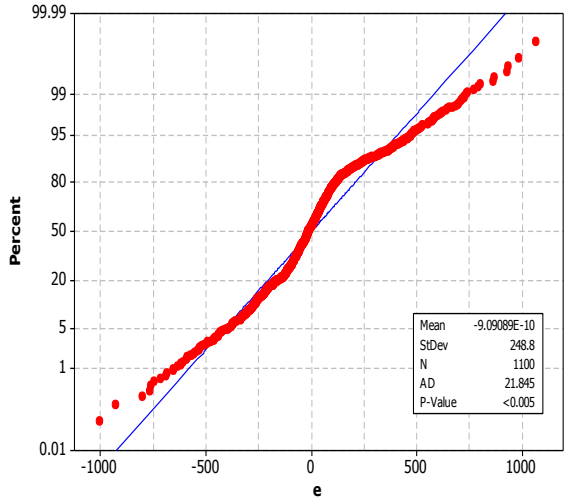
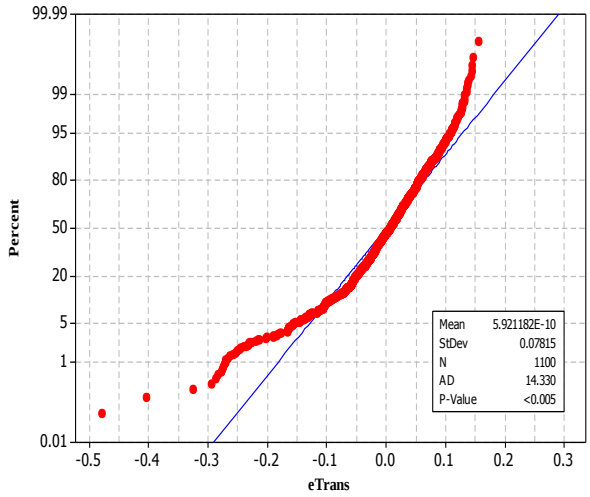


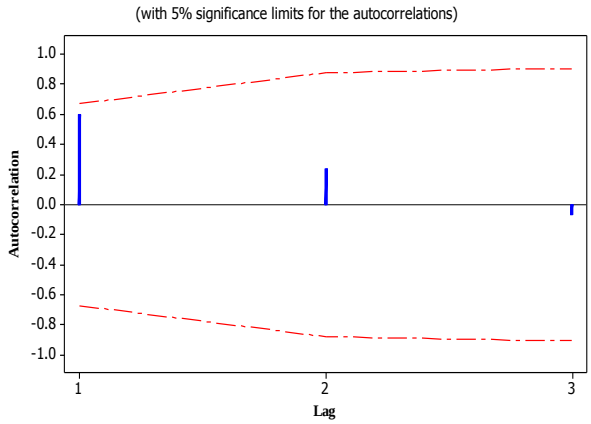
Figure 3-24 Probability plot of residuals with transformation



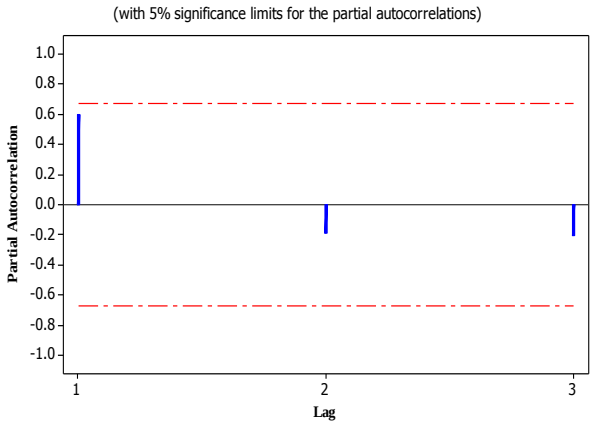
Thus according to the Figure 3-23 residuals followed $N(0,0.078)$. Further analysis may have been including standardizing these residuals and converting them to follow $N(0,1)$. Figure 3-25-a through Figure 3-25-g show Autocorrelation and Partial Autocorrelation functions on three of the 100 Phase I profiles. These functions were applied to the residuals obtained after fitting the regression model to the transformed y's. As it can be seen from these Figures, the residuals were not correlated.

Since there was no correlation between the residuals of the model, there were ready to be fed to control charts. Two types of control charts, IX chart and EWMA, were tested. Figure 3-26 and Figure 3-27 show IX control charts for two randomly selected profiles. As it is seen from these Figures, there was a similar pattern in the shape of the two charts. They both started rising from the beginning, reached to their peak somewhere in the middle and went down until the last point. This was more likely due to the weakness of the linear model for this problem. Further analysis revealed that this pattern was similar between all of the profiles. In summary, lacking arbitrary patterns for the plotting points in this method made it biased and undesirable.

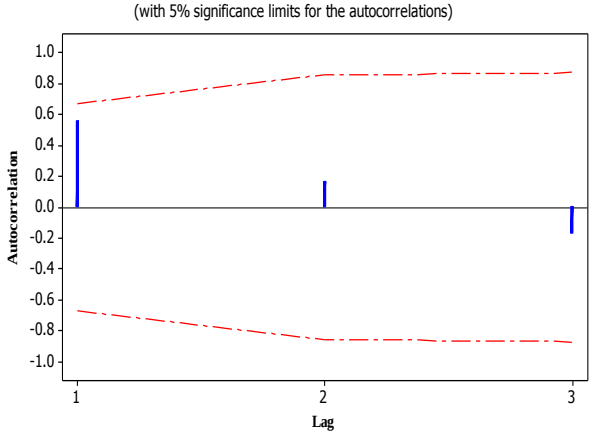
Figure 3-25 ACF and PACF charts for three Phase I profiles



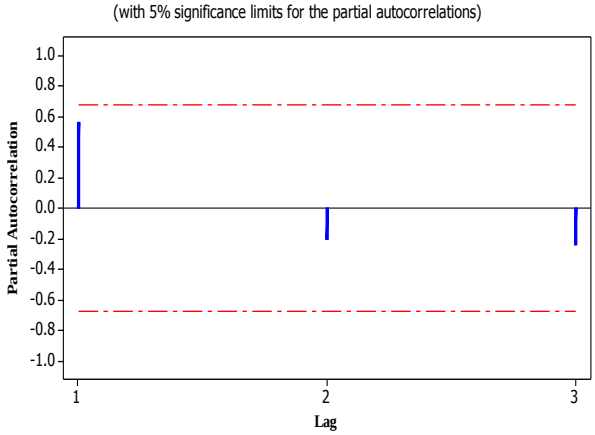
(a).ACF for Profile 1



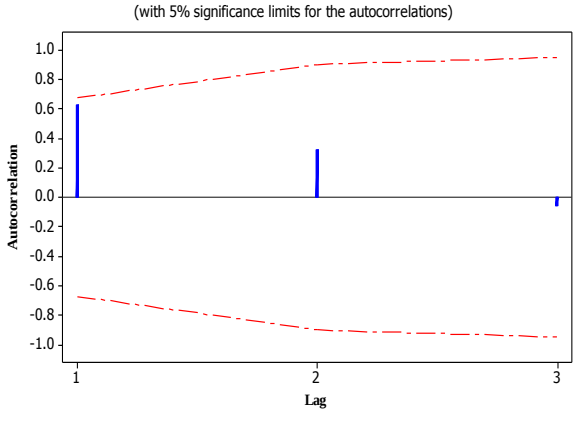
(b) PACF for Profile 1



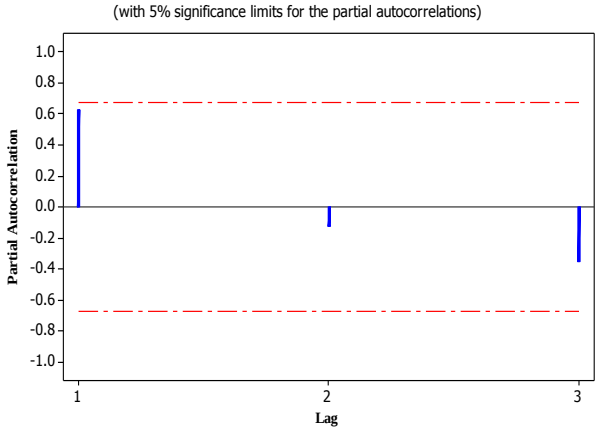
(c) ACF for Profile 2



(d) PACF for Profile 2



(f) ACF for Profile 100



(g) PACF for Profile 100

Figure 3-26 IX control chart for the first randomly selected profile

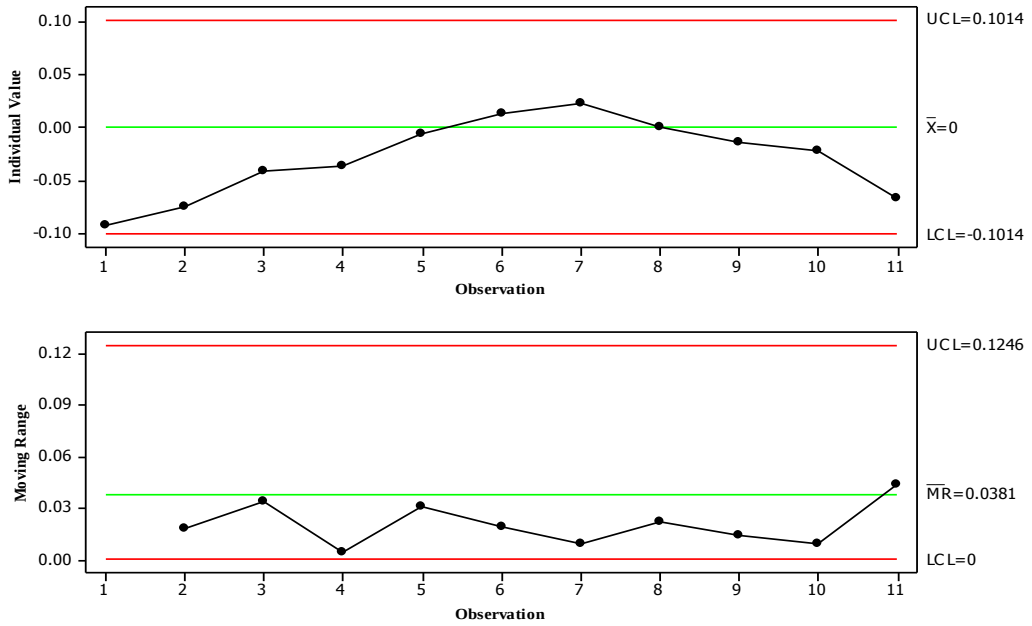
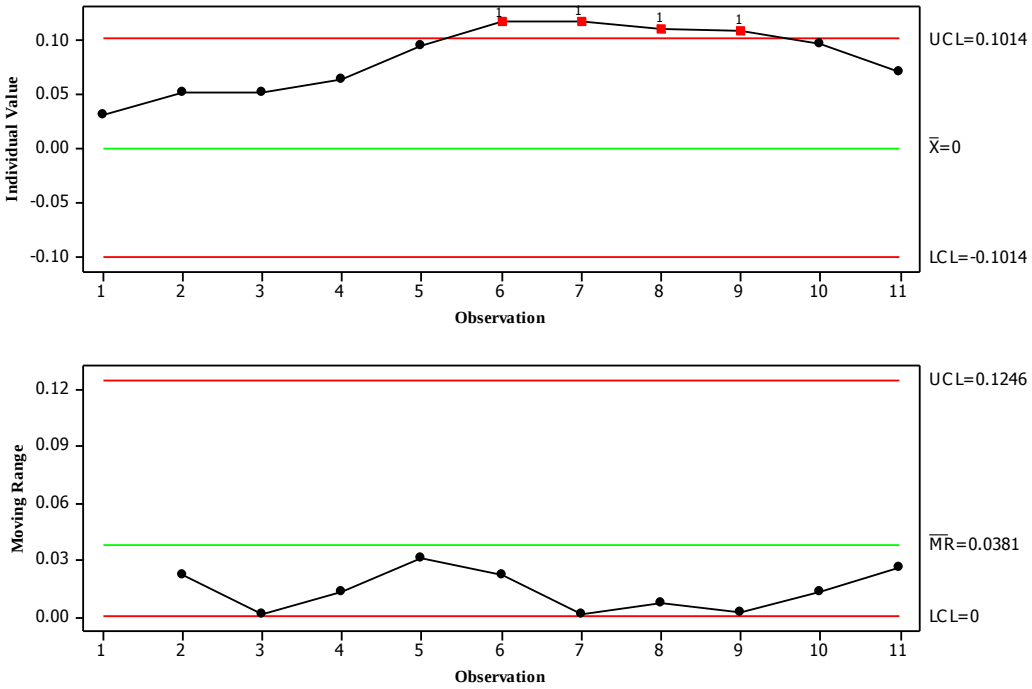


Figure 3-27 IX control chart for the second randomly selected profile



EWMA charts were also applied to the residuals of the model. A brief explanation of EWMA charts was brought in the second chapter. Meanwhile, Equations (2.7) and (2.8) provide

the corresponding formulae. We chose L=2.7 and λ=0.1. These values were the typical ones in using EWMA control charts. Generally, larger values of λ to have looser control limits. Figure 3-28 and Figure 3-29 show the EWMA charts for two randomly selected profiles. The same as IX control chart there was common pattern in the shape of the EWMA control chart. The result suggested that the linear regression model did not perform well for this problem.

We knew that the second profile was an out of control one but as it is seen from the Figure 3-29, the EWMA chart is not able to identify it.

Figure 3-28 EWMA control chart for the first randomly selected profile

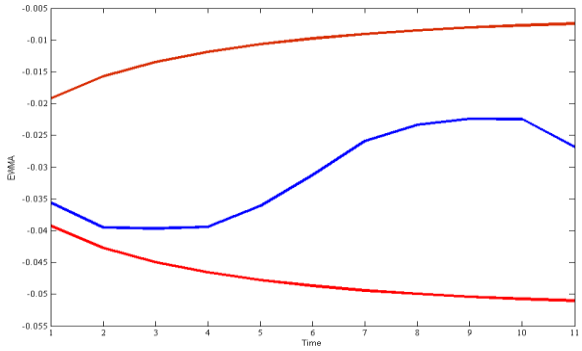
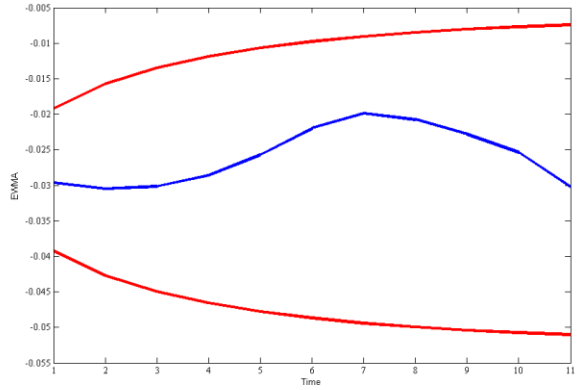


Figure 3-29 EWMA control chart for the second randomly selected profile

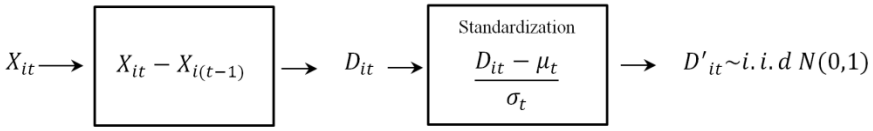


It turned out that this method had a substantial lack of functionality which may have been due to linearity of the fitted regression model. To make this up, other methods were tried that will be explained in the following sections of this chapter.

3.5 Fourth Method (Segmenting then Standardizing)

The fourth proposed method is demonstrated as follows. A scheme containing different steps of this method is also brought in Figure 3-30. Again, X_{it} was considered as the area for the profile i from the beginning of the process up to the time t. Depending on how many segments we liked to divide the process into, we would have different numbers in index t. Although the ideal case was to have continuous t, a more realistic way was to implement it discretely. D_{it} was defined as the difference of the enclosed areas for profile i between time t-1 and time t. Thus D_{it} was calculated according to the following formula:

Figure 3-30 A scheme of the fourth proposed methodology



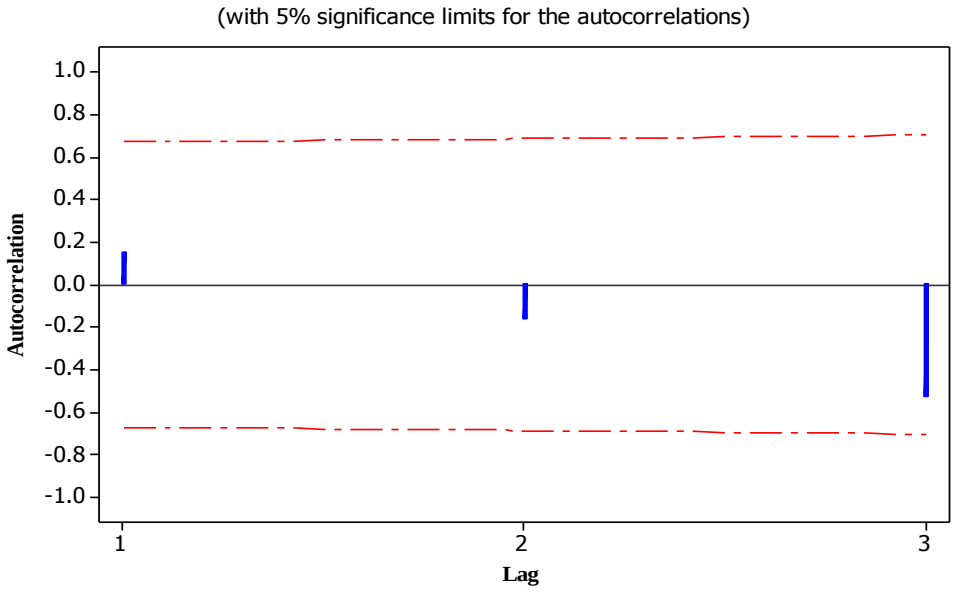
$$D_{it} = X_{it} - X_{i(t-1)} \tag{3.5}$$

The next step in this analysis was standardization of D_{it} . This was done according to the below formula:

$$D'_{it} = \frac{D_{it} - \mu_t}{\sigma_t} \tag{3.6}$$

where; $t = \frac{1}{8}, \frac{1}{7}, \ldots, 1$, μ_t was the average of D_{it}'s of 100 Phase I profiles at time t, and σ_t was the standard deviation of D_{it}'s of 100 Phase I profiles at time t. The output of the above formula was called D'_{it} and was used for further analysis. The advantage of using this approach was that D_{it}'s were not correlated because of the way the D_{it} was defined. This meant that there was no need to apply any type of filter to remove the autocorrelation. Figure 3-31 shows the autocorrelation function for one of the Phase I profiles using the D_{it} approach. As it is obvious from this Figure, D_{it} values were not correlated together. This implied that they were ready to be fed to control charts.

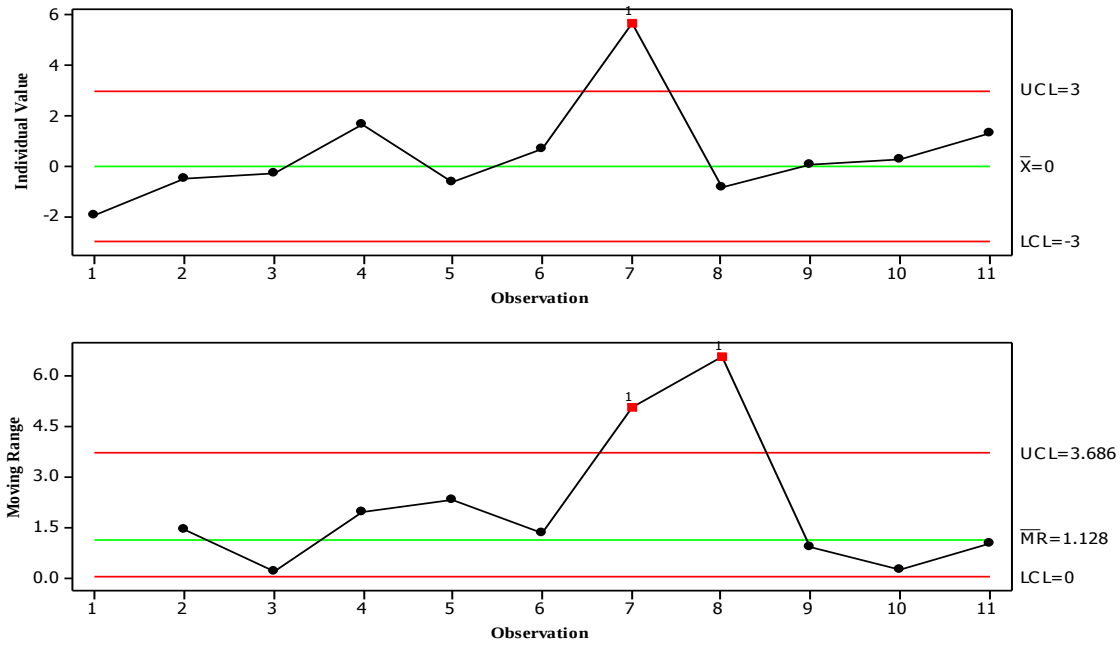
Figure 3-31 Autocorrelation Function for one of the profiles



Since we were not interested in grouping the D'_{it}'s, we liked to study them as individual observations. As we knew that D'_{it}'s were independent and not correlated, IX control chart was applicable for monitoring them. Control limits for the IX control charts were defined in equations (2.6) and (2.12). Again, having n=2, values of d_2, D_3 and D_4 are set as 1.128, 0 and 3.267 respectively.

The same as before, in this research, the consistent control limits for IX control charts were considered. The control limits were set as -3, 0 and, 3 for the individual value chart and also for the moving range chart they were set as 0, 1.128, and, 3.686. Figure 3-32 shows an IX control chart for one of the Phase I profiles. That profile was already displayed in Figure 3-8.

Figure 3-32 IX control chart for one of the correctly identified out of control profiles

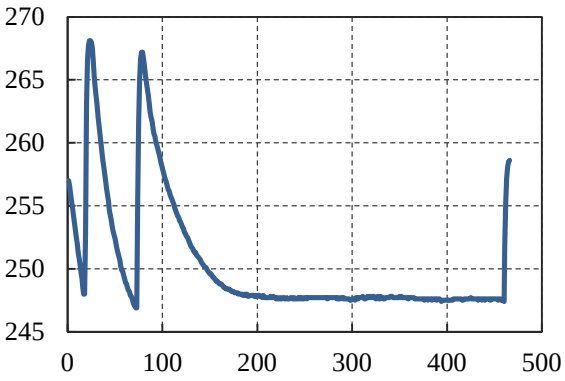


Using IX control chart, in the Phase I, three out of eight real out of control profiles were identified correctly. In addition, this method identified six profiles incorrectly. In other words, it gave six false alarms. These results were still not satisfactory enough but D method seemed to be a better method compared to the other three methods.

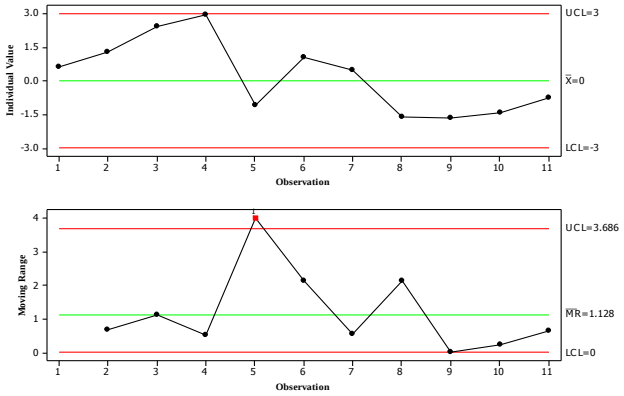
The results obtained from this method were pretty much the same as the ones obtained from the first and second methods but utilization of this method was much easier than the rest and the level of computations was also much lower. For the three previous methods Phase I analyses were only conducted and according to the unsatisfactory results in that phase, Phase II analyses were not conducted anymore, but since D method was identified as the best one among the four, Phase II analysis was conducted for it too, to have a better verification of its performance.

Yet, in order to validate the performance of D method as the best method, further analysis is needed. Simulation analysis which will be conducted in chapter 4, includes the implementation of D method on simulated profiles. Possible satisfactory results of that analysis, will be used to validate the D method.

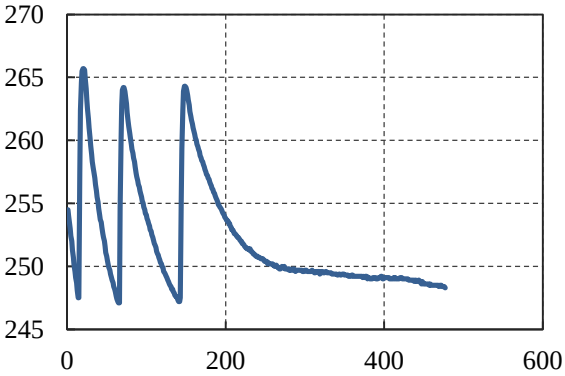
Figure 3-33 Three of correctly detected out of control Phase II profiles with their IX control charts



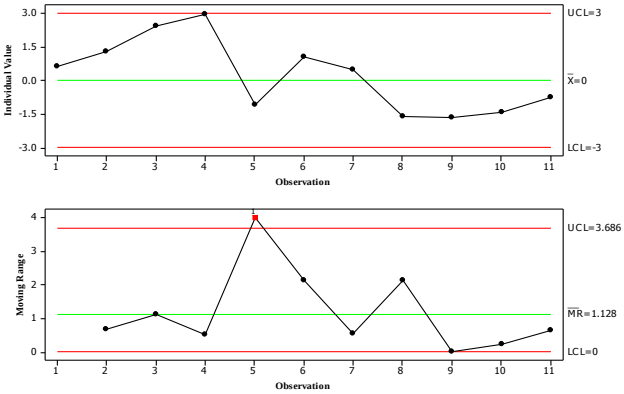
(a).Profile 43



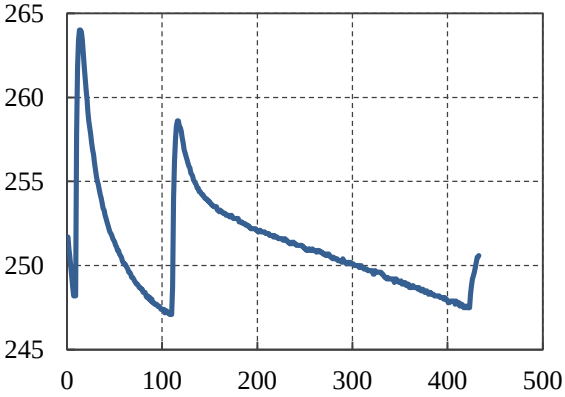
(b) IX control chart for Profile 43



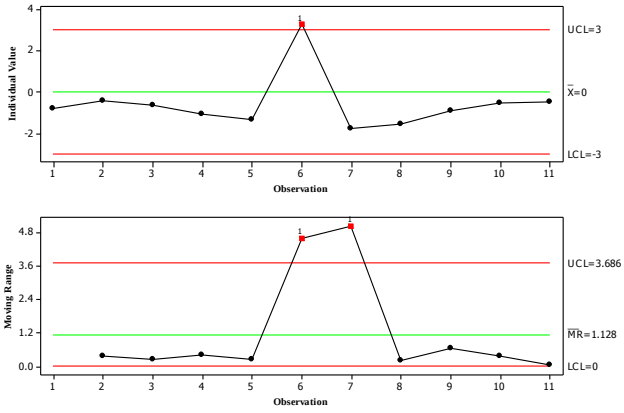
(c) Profile 53



(d) IX control chart for Profile 50



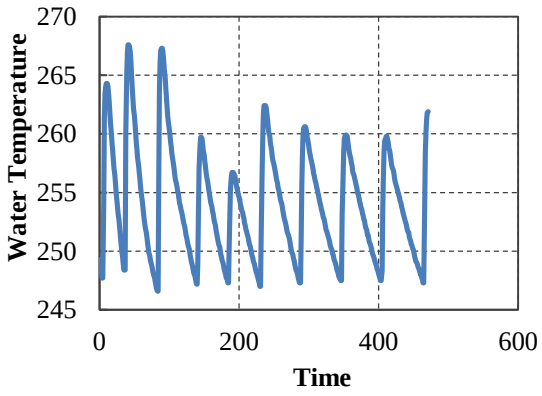
(e) Profile 65



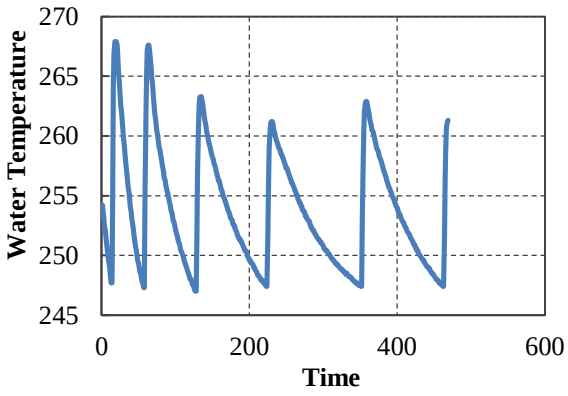
(f) IX control chart for Profile 65

For the Phase II analysis, following the same approach illustrated for the first method, all of the undesirable or in other words out-of-control profiles were removed from the Phase I profiles. Technically, in the Phase I, the system is trained and will be ready for the Phase II. It turned out that this method was able to identify all of the five bad profiles in Phase II. There were also six false alarms in this Phase. Figure 3-33 shows the three correctly identified bad profiles in Phase II besides their IX control charts. It was seen that IX control charts signal at points $\frac{1}{4}, \frac{1}{4}$ and $, \frac{1}{3}$. This meant that they were able to identify these bad profiles in less than half of the whole process time. Figure 3-34 shows three of the profiles with false alarms. Although the waves in these profiles had large amplitudes; they had relatively higher frequencies compared to the real out of control profiles.

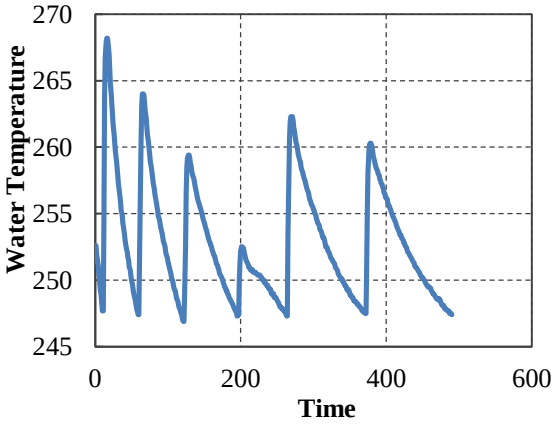
Figure 3-34 Three profiles with false alarms and their IX control charts



(a).Profile 19



(b) Profile 32



(c) Profile 21

3.6 Fifth Method (Multivariate Analysis)

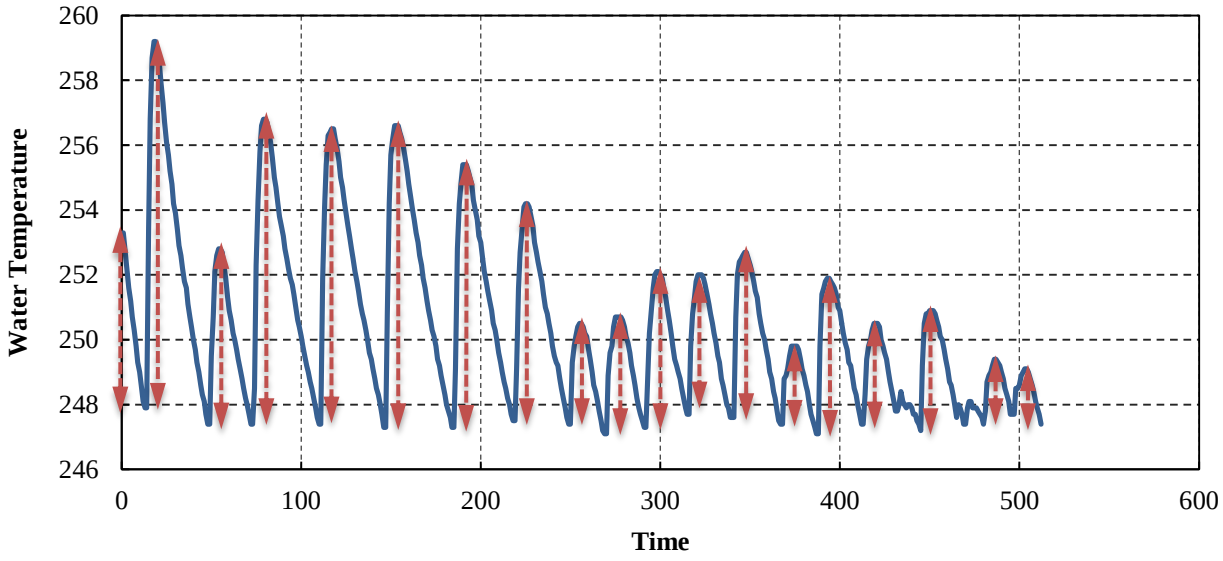
So far, all of the proposed methods utilized univariate approaches. The studied variable in all of the four previous methods was based on the enclosed area between the wave profile and the fitted cutting line. In this method, a multivariate analysis was adopted. Figure 3-35 displays a scheme of this methodology. However, the use of multivariate analysis demanded extensive amount of computation. This was because of the need to apply multivariate analysis on each of the portions that a profile was divided into. Considering this fact, a multivariate analysis was only conducted for the complete profile. Future research is needed in order to have an easier application of multivariate analysis on other portions. In this research, two types of multivariate analyses were conducted. The first analysis was with three variables and the second one was with five variables. Each of them will be explained in the following sections.

Figure 3-35 A scheme of the fifth proposed methodology

3.6.1. Three Variables Analysis

In this section, three variables analysis is illustrated. The first considered variable was the mean of the magnitudes of the differences between the consecutive peaks and valleys (X_1). The second variable was the standard deviation of the magnitudes of the differences between the consecutive peaks and valleys (X_2), and, the third variable was the number of peaks and valleys for each profile (X_3). These three variables formed a vector called X with the elements of X_1, X_2 and X_3. Figure 3-36 illustrates these variables. The dashed double arrow lines display the magnitudes of the differences between the consecutive peaks and valleys.

Figure 3-36 Dashed double arrow lines as the magnitudes of the differences between the consecutive peaks and valleys for a regular curing process profile



Note that having 183 profiles, 100 of them were used for Phase I, and 83 of them were used for Phase II. Desired multivariate analysis was to apply a Hotelling T^2 control chart using the three defined variables. Since observing the performance of each individual profile was of interest, the Hotelling T^2 control chart for the individual observations was utilized. The test statistic for the Hotelling T^2 method is as follows:

$$T^2 = (X - \bar{X})' S^{-1} (X - \bar{X}) \tag{3.7}$$

where;

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$
$$S = \frac{1}{n-1} \sum_{i=1}^{n} (X_j - \bar{X})(X_j - \bar{X})' \tag{3.8}$$

Considering the 100 Phase I profiles, the control limits could be defined as:

$$Phase\ I: UCL = \frac{(n-1)^2}{n} \beta_{\alpha, \frac{p}{2}, \frac{n-p-1}{2}} \tag{3.9}$$

where n is the number of profiles, p is the number of variables which in this case was equal to 3 and α is the type I error that in most of the Hotelling T^2 analysis is set as 0.0027.

Having the abovementioned information the control limits and the test statistic values for the 100 Phase I profiles were obtained.

$$UCL = \frac{(100-1)^2}{100}\ \beta_{0.0027,\frac{3}{2},\frac{100-3-1}{2}}$$
$$UCL_{\alpha=0.0027} = 13.3733 \tag{3.10}$$

Table 3-1 shows the results of conducting the Hotelling T^2 analysis considering different values of α. Second column of this Table is for α=0.0027. As it is seen from there, in this case, there were three type II errors. In other words, four out of eight Phase I out of control profiles were identified.

After removing the seven out of control profiles from the Phase I, n would be equal to 93. Thus, Phase II control limit was calculated according to the following:

$$Phase\ II: UCL = \frac{p(n+1)(n-1)}{n(n-p)} F_{\alpha,p,n-p}$$
$$UCL = \frac{3(93+1)(93-1)}{93(93-3)} F_{0.0027,3,93-3} = 15.7466 \tag{3.11}$$

Again, the second column of Table 3-1 reflects the desired statistics. In the Phase II, two out of five real out of control profiles were identified correctly. This meant having three type II errors. The five real out of control profiles' Figures are brought later on.

In summary, none of the reflected values were satisfactory enough. The best-case scenario may have been the one with α=0.05 but as it is seen from the Table 3-1, the number of false alarms were a lot. α=0.05 means ARL_0=20, or in other words means having a false alarm every 20 observations which was not desirable. In addition, still, one out of control case profile could not have been identified in this case.

These results suggested the consideration of more variables in the model. Three variables may have not been enough for identification of the out of control cases. Therefore, two more variables were added to the original three.

Table 3-1 Type I and Type II Errors in different phases for the three variables analysis

	α=0.002 ARL_0=500		α=0.0027 ARL_0=370.3		α=0.004 ARL_0=250		α=0.0375 ARL_0=26.66		α=0.05 ARL_0=20	
	Type I Error	Type II Error	Type I Error	Type II Error	Type I Error	Type II Error	Type I Error	Type II Error	Type I Error	Type II Error
Phase I	0	0.5	0	0.375	0	0.375	0.021	0.25	0.021	0.25
Phase II	0.012	0.6	0.025	0.6	0.025	0.6	0.076	0.6	0.012	1
UCL-Phase I	13.93288		13.37397		12.63406		8.221624		7.624556	
UCL-Phase II	16.51756		15.74663		14.74206		9.106114		8.387128	

3.6.2. Five Variables Analysis

Another analysis was conducted in which five variables were considered. Those five variables were Mean of the magnitudes of differences between consecutive peaks and valleys (X_1), Standard deviation of the magnitudes of differences between consecutive peaks and valleys (X_2), Mean of the time differences between the consecutive peaks and valleys (X_3), Standard deviation of the time differences between consecutive peaks and valleys (X_4), and, the number of peaks and valleys (X_5). These five variables formed a vector called X with the elements of X_1, X_2, X_3, X_4, X_5 . Figure 3-37 helps for better understanding of the added variables. The test statistics for the Hotelling T^2 method are the same as equations (3.7) and (3.8).

where, in this case:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix} \tag{3.12}$$

Considering the 100 Phase I profiles the control limits could have been defined according to equation (3.10). where in this case, p is equal to 5, n is equal to 100 and α was set as 0.0027. Thus, having the needed values the control limit was calculated.

$$UCL = \frac{(100-1)^2}{100}\,\beta_{0.0027,\frac{5}{2},\frac{100-5-1}{2}} = 17.0161 \tag{3.13}$$

Again, similar to the three variables analysis, after removing the seven real out of control profiles from the Phase I, n would be equal to 93. Thus, for the Phase II, using equation (3.11), the control limit was:

$$UCL = \frac{5(93+1)(93-1)}{93(93-5)}\,F_{0.0027,5,93-5} = 20.9851$$

Table 3-2 Type I and Type II errors in different phases for the five variables analysis

α=0.0027	Type I Error	Type II Error
Phase I	0.01	0.25
Phase II	0.064	0

Figure 3-37 Dashed double arrow lines as the magnitudes of the differences between the consecutive peaks and valleys and, dotted double arrow lines as the time differences between the consecutive peaks and valleys for a regular curing process profile

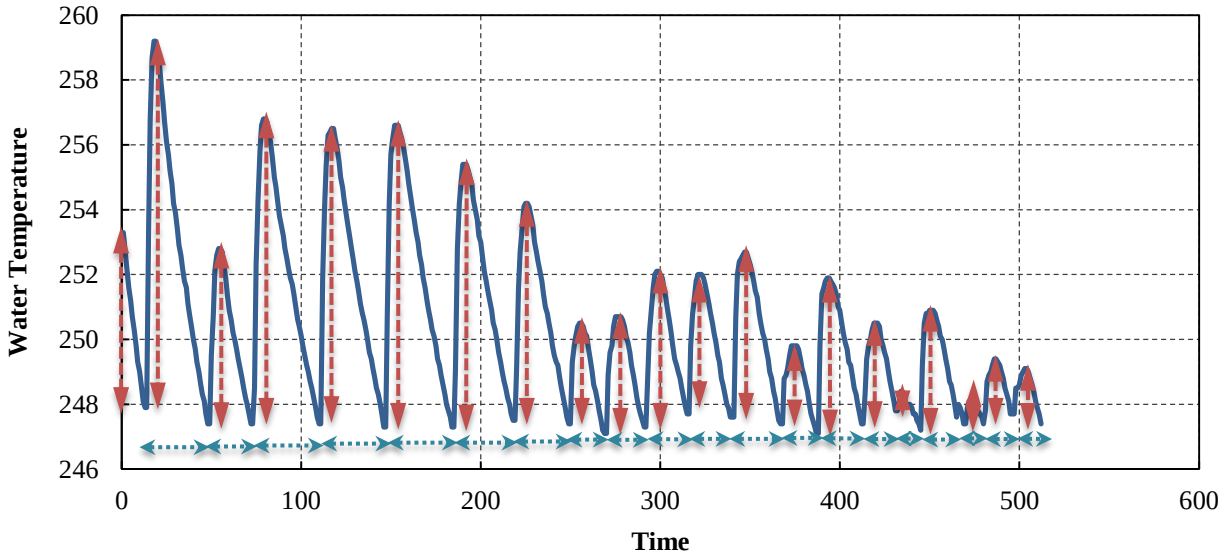
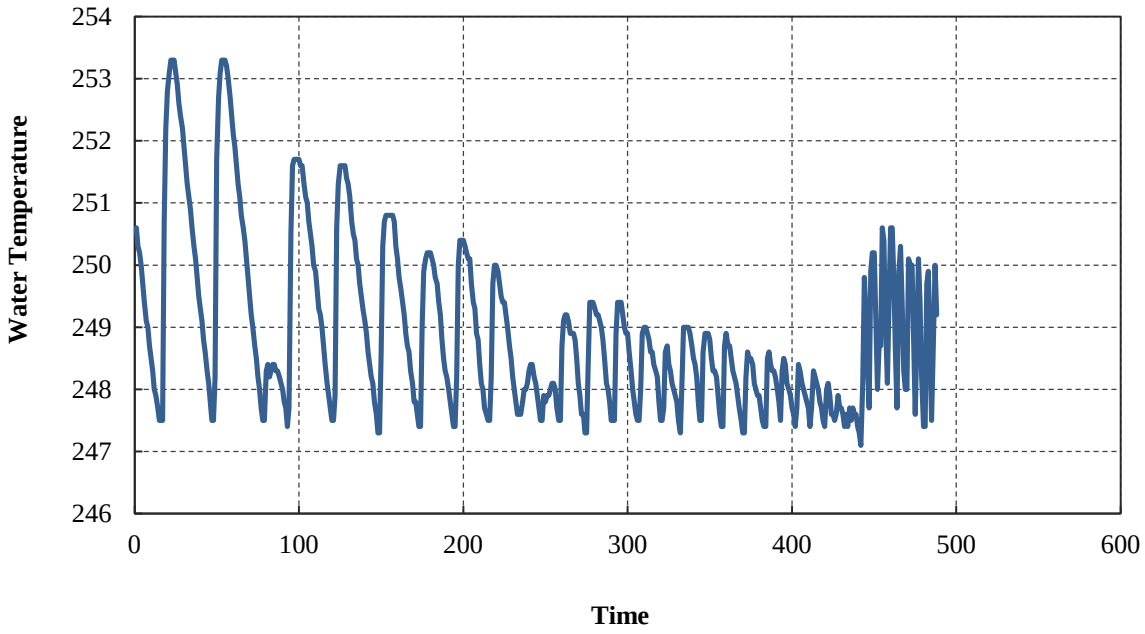


Table 3-2 reflects the statistics for the five variables Hotelling T^2 chart's performance. The five variables method did a relatively good job in Phase I with identifying five of the seven out of control profiles.

In order to investigate the shortcoming of the

Figure 3-38 Profile 34

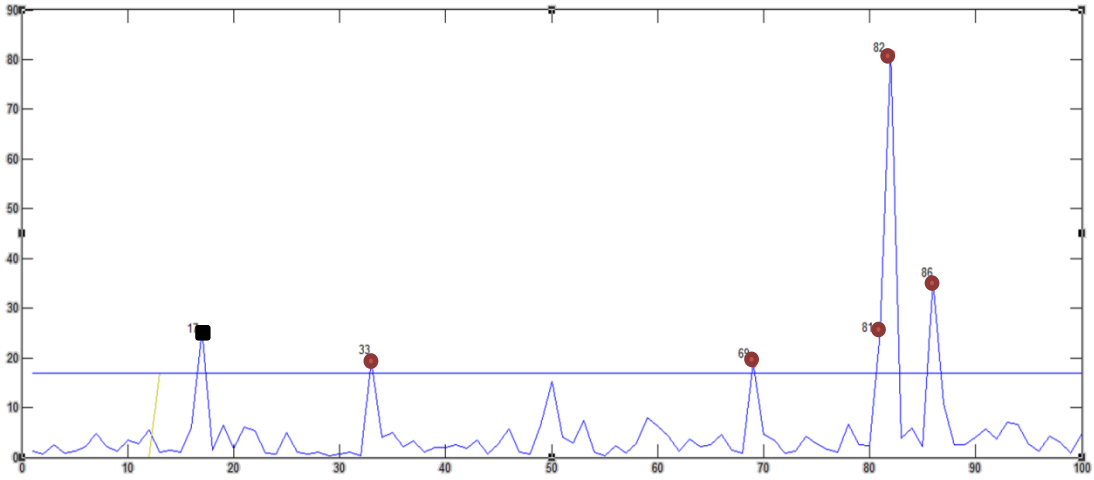


Finally, Figure 3-42 displays the Hotelling $T2$ control chart for the Phase II of the process. Again, the points highlighted in red demonstrate the real out of control profiles and the points highlighted in black demonstrate the false alarms. The statistics mentioned in Table 3-2 can be verified here. Five red (or circle) points demonstrate five correctly identified out of control profiles and five black (or square) points demonstrate five false alarms.

Figure 3-40 shows the five real Phase II out-of-control profiles. Although the Phase II statistics were not too good, they were not too bad either. Figure 3-41 shows the five false alarms' profiles. It is up to the opinion of expert to identify if all of these were false alarms or not

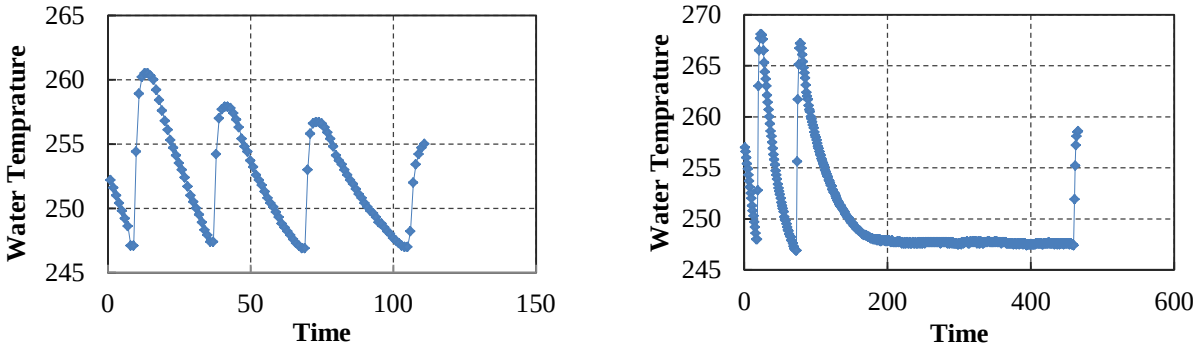
but it seemed that profile 64, in Figure 3-41(e) could be noted as an out of control profile. If this was the case, the method would have a better performance with identifying 6 out of control profiles and 4 wrong signals.

Figure 3-39 Hotelling T^2 control chart for the Phase I of the process

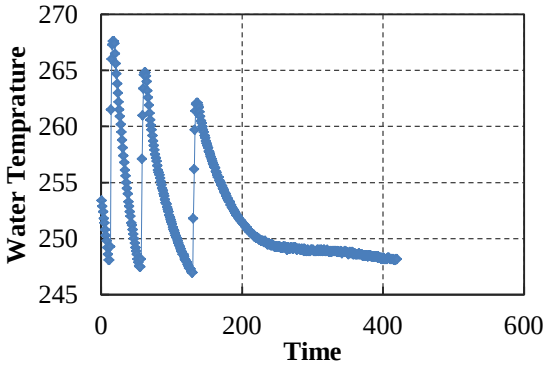


Finally, Figure 3-42 displays the Hotelling T^2 control chart for the Phase II of the process. Again, the points highlighted in red demonstrate the real out of control profiles and the points highlighted in black demonstrate the false alarms. The statistics mentioned in Table 3-2 can be verified here. Five red (or circle) points demonstrate five correctly identified out of control profiles and five black (or square) points demonstrate five false alarms.

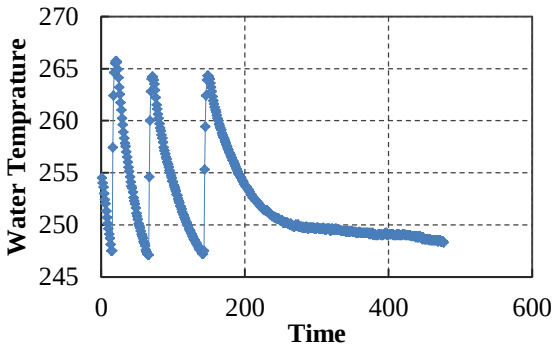
Figure 3-40 The five real Phase II out of control profiles



(a)-Profile 42

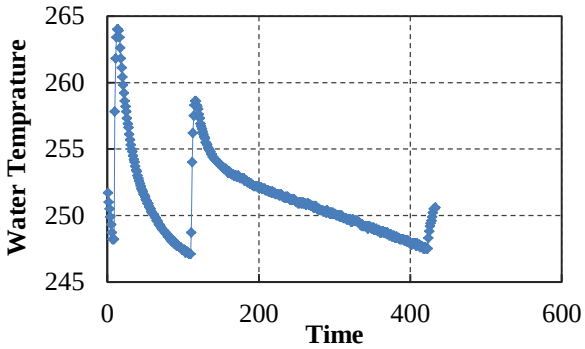


(b)-Profile 43



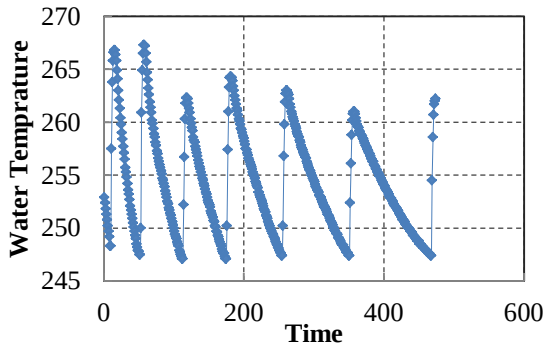
(c)-Profile 47

(d)-Profile 50

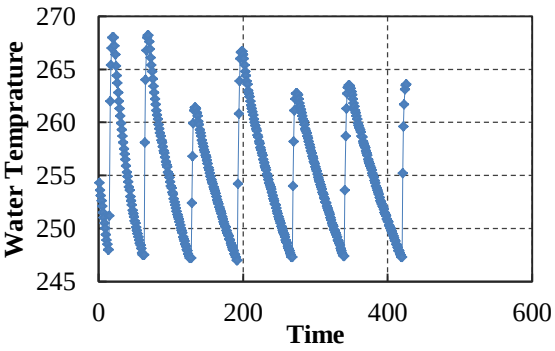


(e)-Profile 65

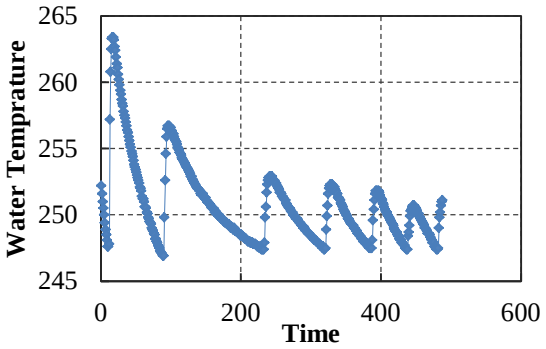
Figure 3-41 The five Phase II profiles, identified incorrectly as out of control



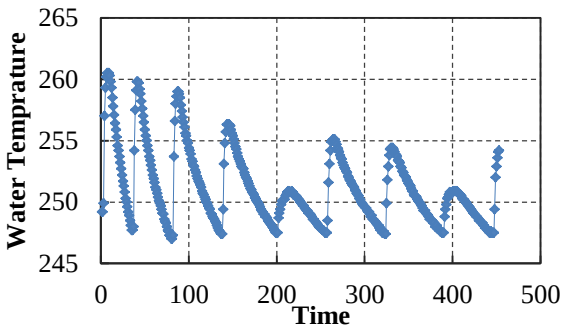
(a)-Profile 1



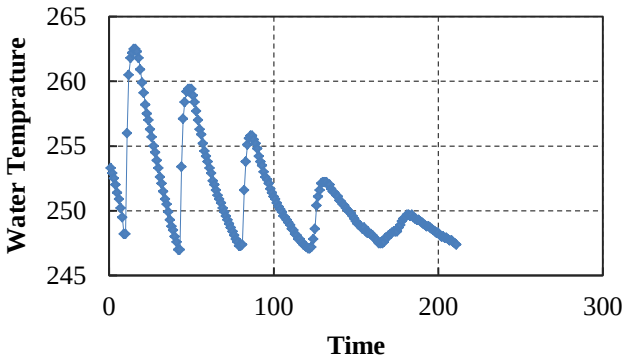
(b)-Profile 2



(c)-Profile 40



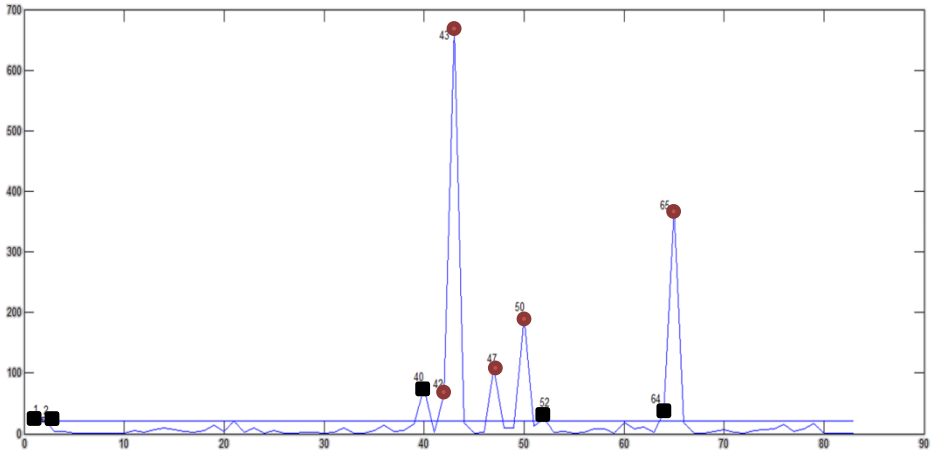
(d)-Profile 52



(e)-Profile 64

In summary, it turned out that the multivariate analysis does a more satisfactory job compared than the univariate analysis in order to identify the out of control profiles. Further research, revealed that considering five variables for the model would better capture the profiles' characteristics compared to considering the three variables.

Figure 3-42 Phase II-Hotelling T^2 Chart-Five variables method



Chapter 4-Simulation Analysis

The fourth method discussed in the previous chapter was selected as the best one due to its simplicity and performance. A simulation study will be conducted to better evaluate the performance of this method. Simulated profiles will be established for various in-control and out-of-control scenarios. Based on the in-control profiles, we are able to train the control charts better in Phase I with more desirable profiles and thus obtain a more accurate evaluation of the performance of the studied method in Phase II.

4.1 The Studied Profiles

Three types of profiles are simulated. First type of profiles, are those that were deemed goods when the process is in control. These good profiles are the ones that have a descending shape and exponential decay function is chosen as the best fit to them. An example of real-world profiles of this type is shown in Figure 4-1-a while a simulated profile is shown in Figure 4-1-b. Second type of profiles, are the bad 1 profiles with inadequate amount of peaks and valleys within them and bigger magnitudes of difference between peaks and valleys. An example of the production profiles is shown in Figure 4-2-a. A simulated one is shown in Figure 4-2-b. Finally, third type of profiles, are bad 2 profiles that look similar to the good profiles except for their last segments with a sudden shift. An example of these profiles is shown in Figure 4-3-a, of which a simulated one is shown in Figure 4-3-b.

4.2 Simulation Models

In this section the mathematical functions for three types of simulated profiles are listed. These functions are basically the ones that are used for modeling vibration systems. Because of the similarity between the water temperature profiles of the curing process and vibration systems, these functions were utilized. All three functions utilize pretty much the same parameters but with different settings.

4.2.1. Function for Simulated Good Profiles

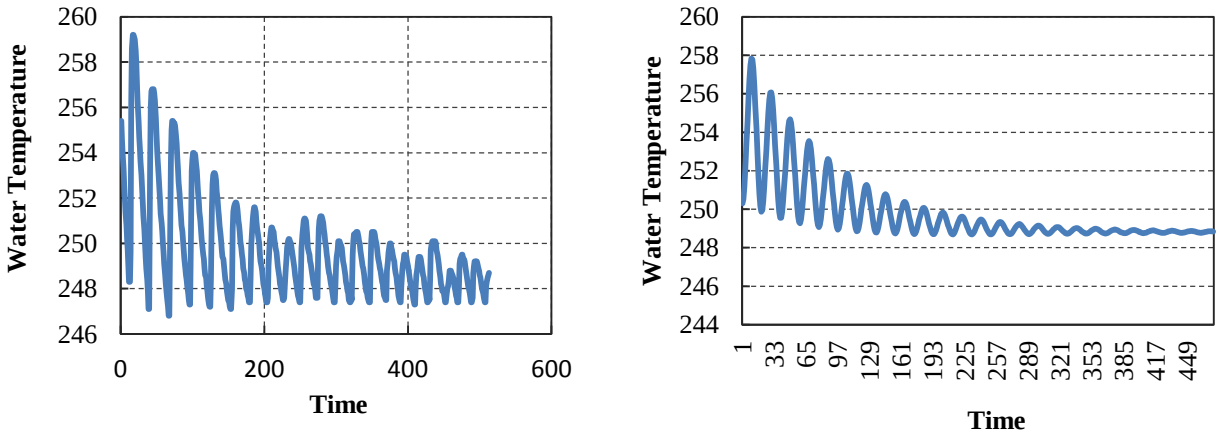
At first, the function for simulating the good profiles is illustrated. These profiles have a descending shape and exponential-decay function. This function is reflected in Equation (2.7):

$$x(t) = fe^{-nt}cos(mt) + C(1 - \cos(mt)) - \frac{C(\frac{t}{T} - sin(mt))}{mT} \tag{4.1}$$

$$y(t) = f + 2A\frac{(x(t) - \frac{\max(x(t))+\min(x(t))}{2})}{\frac{\max(x(t))-\min(x(t))}{2}}, \qquad t = 0, \dots, T \tag{4.2}$$

where $x(t)$ is the water temperature at time t, $y(t)$ is the scaled water temperature at time t, f is the scaling parameter of the water temperature, A is the amplitude parameter. e^{-nt} is the damping factor where n determines speed of the damping, m is the controller parameter of frequency of the wave, t accounts for time, and T is the time at the end of the profile. Intervals of 0.05 were used for t in this model. In addition, T was set as 23.7 so that all of the profiles consisted of the same number of points of 475.

Figure 4-1 (a) An example of the good or the so-called "typical" profiles, (b) An example of the simulated good or the so-called "typical" profiles



(a) (b)

As it is seen, equation (4.1) consists of three parts. The first part takes care of the damping behavior of the water temperature wave, while the second and the third parts take care of the exponential behavior of the wave. Constant C is a used for calibration purposes in the second and third parts and is set equal to 1000 as the result of try and error.

Equation (4.2) displays the scaling function. Outcomes of the equation (4.1) are in a range different than the ones for real profiles. In order to make the range of the simulated water temperatures similar to the real profiles equation (4.2) was utilized.

Figure 4-1-b displays an example of the good simulated profiles. Comparing Figure 4-1-b to Figure 4-1-a, the simulated profiles were similar to the real profiles.

4.2.1.1 Range of the defined variables

In order to get more similar profiles to the real first type profiles, the appropriate ranges of the parameters had to be set. These parameters were as follows:

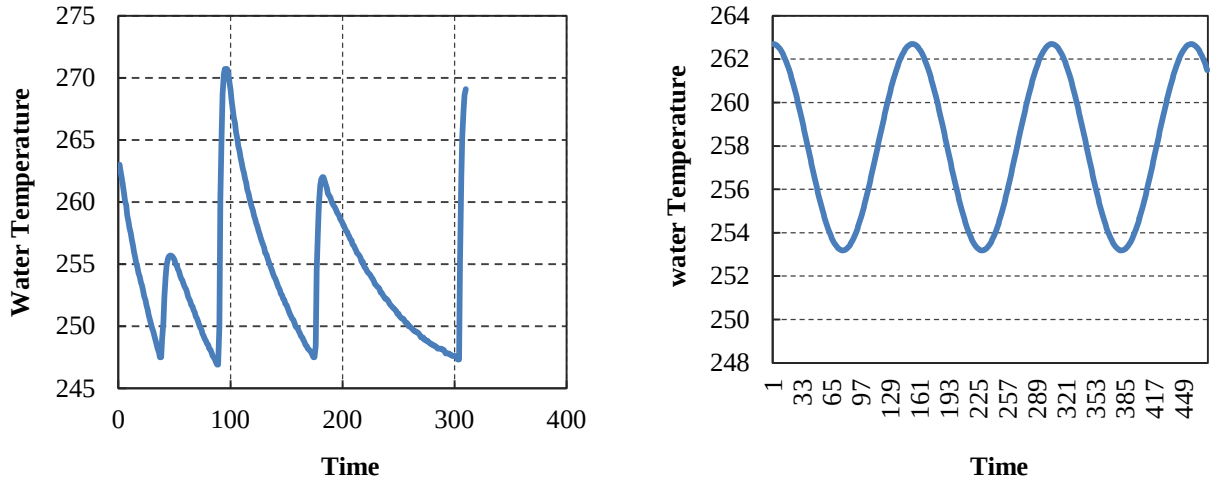
$$A \sim Normal(10, 0.7)$$
$$f \sim Normal(260,2)$$
$$n \sim Normal(0.2,0.005)$$
$$m \sim Normal(5,1)$$

This range of variables provides profiles with the number of peaks and valleys in the range of 1 to 40.

4.2.2. Function for the Bad 1 Profiles

The second profiles are the ones with low numbers of peaks and valleys and high magnitudes of differences between the values of the peaks and valleys. As it can be seen from Figure 4-2-a, a real profile of bad 1 type may have been both ascending and descending within itself. Another characteristic of this type of profiles is that they do not damp. The simulation function proposed in equations (4.3) and (4.4) does not capture the first characteristic but it does the second one. Figure 4-2-b displays a bad 1 simulated profile. By comparing this Figure with Figure 4-2-a, it is seen that the simulated profile could represent the second type profiles reasonably well.

Figure 4-2 (a) An example of the type 1 bad profiles, (b) An example of the simulated type 1 bad profiles



(a) (b)

Equation (4.3) displays the utilized function for simulating the second type profiles:

$$x(t) = f\cos(mt) \tag{4.3}$$

$$y(t) = f + 2A\frac{(x(t) - \frac{\max(x(t))+\min(x(t))}{2})}{\frac{\max(x(t))-\min(x(t))}{2}}, t = 0, \dots, T \tag{4.4}$$

where again, $x(t)$ is the water temperature at time t, $y(t)$ is the scaled water temperature at time t, f is the scaling parameter of the water temperature, A is the amplitude parameter. m is the controller parameter of frequency of the wave, t accounts for time, and T is the time at the end of the profile. Intervals of 0.05 were used for t in this model again. Moreover, T was set as 23.7 so that all of the profiles consisted of the same number of points of 475.

Unlike equation (4.1), equation (4.3) only consists of one part. The second and third part of the equation (4.1) which were for the modeling of the exponential behavior of the profiles, were removed. The exponential portion of the first part was also removed here since the damping is not desired.

4.2.2.1 Range of the defined variables

Similar to what was done for the good simulated profiles, the appropriate ranges of the parameters had to be set for the bad 1 profiles too. These parameters were as follows:

$$A \sim Normal(10, 0.7)$$
$$f \sim Normal(260,2)$$
$$n \sim Normal(1.1,0.2)$$
$$m \sim Normal(0.5,0.2)$$

This range of variables provides profiles with the number of peaks and valleys in the range of 1 to 10.

4.2.3 Function for the Bad 2 Profiles

The bad 2 profiles were the ones that were similar to the first type profiles except for their last segments that they had a sudden shift. Figure 4-3-a displays a profile of this kind. The mathematical function for simulating this type of profiles is much the same as the one used for simulating the first type profiles. Equation (4.5) displays the function for simulating the first part of the profile which is similar to the good profiles. The first part begins when the time is zero and last until the time t_i where the shift happens. Equation (4.6) displays the utilized function for simulating the shift part of the profile.

$$x(t) = fe^{-nt}cos(mt) + C(1 - \cos(mt)) - \frac{C(\frac{t}{T} - sin(mt))}{mT}, \qquad t = 0, \dots, t_i \tag{4.5}$$

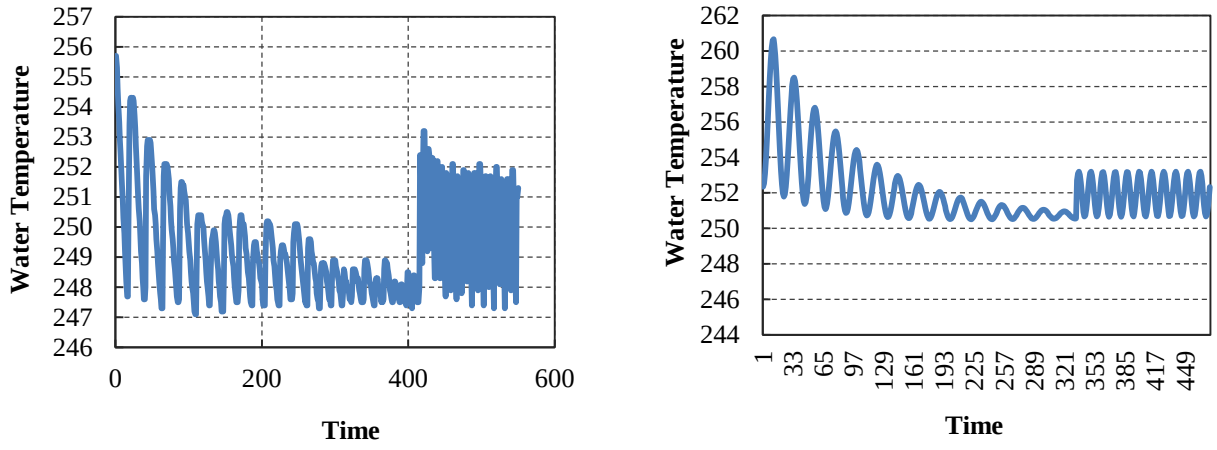
$$x(t) = \frac{3}{4}f + A_1 \cos(m_2 t), \qquad t = t_{i+1}, \dots, T \tag{4.6}$$

$$y(t) = f + 2A_2 \frac{(x(t) - \frac{\max(x(t)) + \min(x(t))}{2})}{\frac{\max(x(t)) - \min(x(t))}{2}}, \qquad t = 0, \dots, T \tag{4.7}$$

where the same as before, $x(t)$ is the water temperature at time t, $y(t)$ is the scaled water temperature at time t, f is the scaling parameter of the water temperature, A_1 and A_2 are the

amplitude parameters for respectively the first and second parts of the profile. e^{-nt} is the damping factor where n determines speed of the damping, m and m_2 are the controller parameters of frequencies of the waves for the respectively first and second parts of the profile, C is a constant used for calibration purposes in the second and third parts and was set equal to 1000 as the result of try and error, t accounts for time, T is the time at the end of the profile and t_i is the time that shift happens. The same as the two previous models, intervals of 0.05 were used for t in this model. In addition, T was set as 23.7 so that all of the profiles consist of the same number of points of 475. A random function was utilized to generate the t_i values from the interval of $t = [12.45,17.45]$ .The same as before, Equation (4.7) was used for scaling purposes.

Figure 4-3 (a) An example of the type 2 bad profiles, (b) An example of the simulated type 2 bad profiles



(a) (b)

Figure 4-3-b displays a bad 2 simulated profile. Similarity between this Figure and Figure 4-3-a, demonstrates that the simulated profiles were much the same as the real type three profiles.

4.2.3.1 Range of the defined variables

The appropriate ranges of the parameters for simulating the bad 2 profiles were as follows:

$$A_1 \sim Normal(195, 0.45)$$
$$A_2 \sim Normal(10, 0.7)$$
$$f \sim Normal(260,2)$$
$$n \sim Normal(0.2,0.005)$$
$$m_1 \sim Normal(5,1)$$
$$m_2 \sim Normal(10,1)$$

4.3 Design of Experiments

The purpose of the simulation study is to evaluate the performance of the best proposed methodology, as it was mentioned before, the D method, on the simulated profiles. In this regard, three experiments were designed. First of all, 150 first type profiles or in other words in-control profiles were generated as Phase I profiles to be used in all of the conducted experiments. The same as before, these profiles were created to be used for training purposes. Table 4-1 displays a summary of the three conducted experiments.

Table 4-1 Summary of the three designed experiments

	First Experiment	Second Experiment	Third Experiment
Phase I profiles	150	150	150
Phase II good profiles	1000	80% of 1000	90% of 1000
Phase II bad 1 profiles	1000	10% 0f 1000	5% 0f 1000
Phase II bad 2 profiles	1000	10% of 1000	5% of 1000

4.3.1 First Experiment-Three Batches of 1000 Profiles for Good, Bad 1 and Bad2 Profiles

In the first experiment a batch of 1000 good profiles in addition to 1000 second type profiles (called bad 1) and 1000 third type profiles (bad 2) were generated for Phase II analysis. The goal was to evaluate the performance of the D method, on detecting the bad 1 and the bad 2 profiles as out of control profiles and the good profiles as in control profiles after training the system with the Phase I profiles. Clearly, as low as possible the number of false alarms were in evaluating the first batch of good profiles, and also as high as possible the speed of the change detection was in evaluating the bad 1 and bad 2 profiles, the better performance of the D method was.

Using the IX control chart as the monitoring tool again, the fourth method's performance was evaluated. By utilizing equations (2.11) and (2.12) and having n=2, again, values of d_2, D_3 and D_4 were set as 1.128, 0 and 3.267 respectively. This gives the control limits of -3 and 3 for the individual value chart and 0, 1.128 and 3.686 for the moving range chart.

4.3.1.1 Simulation Outcomes

Out of the 1000 Phase II good profiles, 927 profiles were detected as in-control correctly. The other 73 profiles had false alarms. In other words, the proposed method had the false alarm rate of 0.073. For the 1000 bad 1 profiles, the proposed method could identify all of them correctly from the beginning. In other words 100 percent of the bad 1 profiles was detected as out-of-control from t=$\frac{1}{8}$. The reason for this performance was because of the large difference in shape of the bad 1 profiles compared with the first type profiles from the beginning point. The enclosed area between the bad 1 profiles and their exponential fits would gain higher values even from the very beginning portions of the profile, compared to the ones for the good profiles. This would cause the early change detection easy.

The simulation results of the 1000 bad 2 profiles also had the satisfactory detection rate of 100 percent and the average first detection time for them was $\frac{15}{22}$ or 0.6818 and the standard deviation of that was equal to 0.123. These statistics are also reflected in Table 4-2.

Table 4-2 Average and standard deviation of first detection time in the first experiment

	Bad 1 Profiles	Bad 2 Profiles
Average of first detection time	0.125	0.681
Standard deviation of first detection time	0	0.123

The simulation results of the 1000 bad 2 profiles also had the satisfactory detection rate of 100 percent and the average first alarm time for them was $\frac{15}{22}$ or 0.6818 and the standard deviation of that was equal to 0.123. These statistics are also reflected in Table 4-2.

Figure 4-4 Histogram of percentage of the first detection time for the 1000 bad 2 profiles in the first experiment

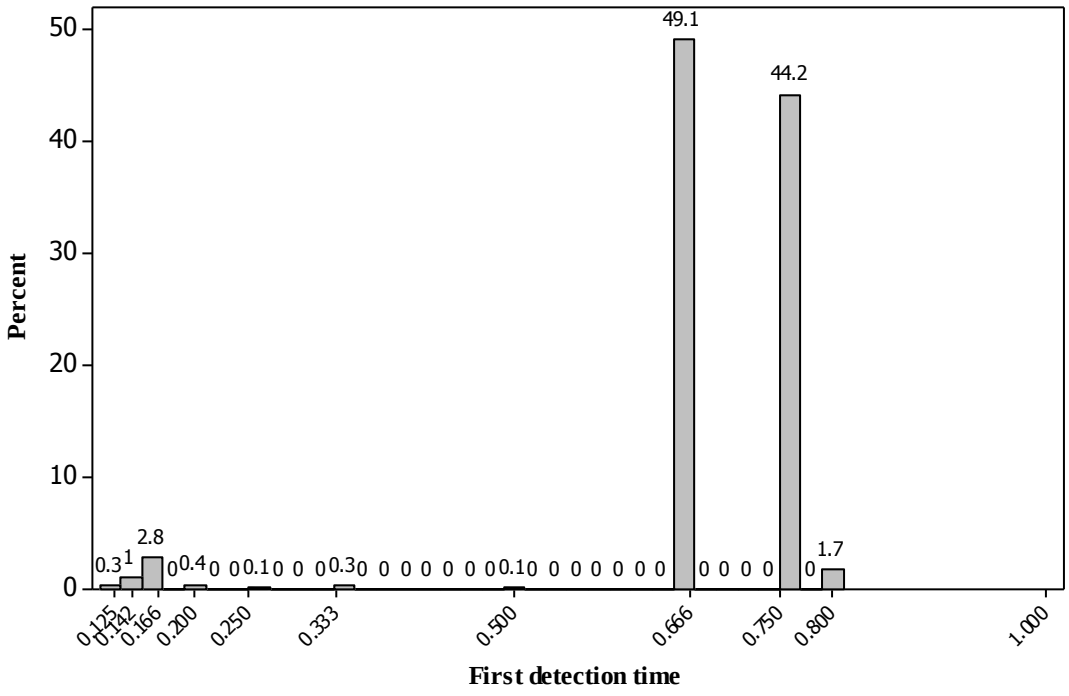


Figure 4-4 displays a histogram of the first detection time of the 1000 bad 2 profiles. The eleven values on x axis represent the eleven portions that a profile is divided into. As it is seen from this histogram, majority of the changes are detected at the relatively last portions. This is due to the type of the changes in the third type profiles that happens at the last segments. The proposed D method correctly identifies the shifts toward to end of the profile segments.

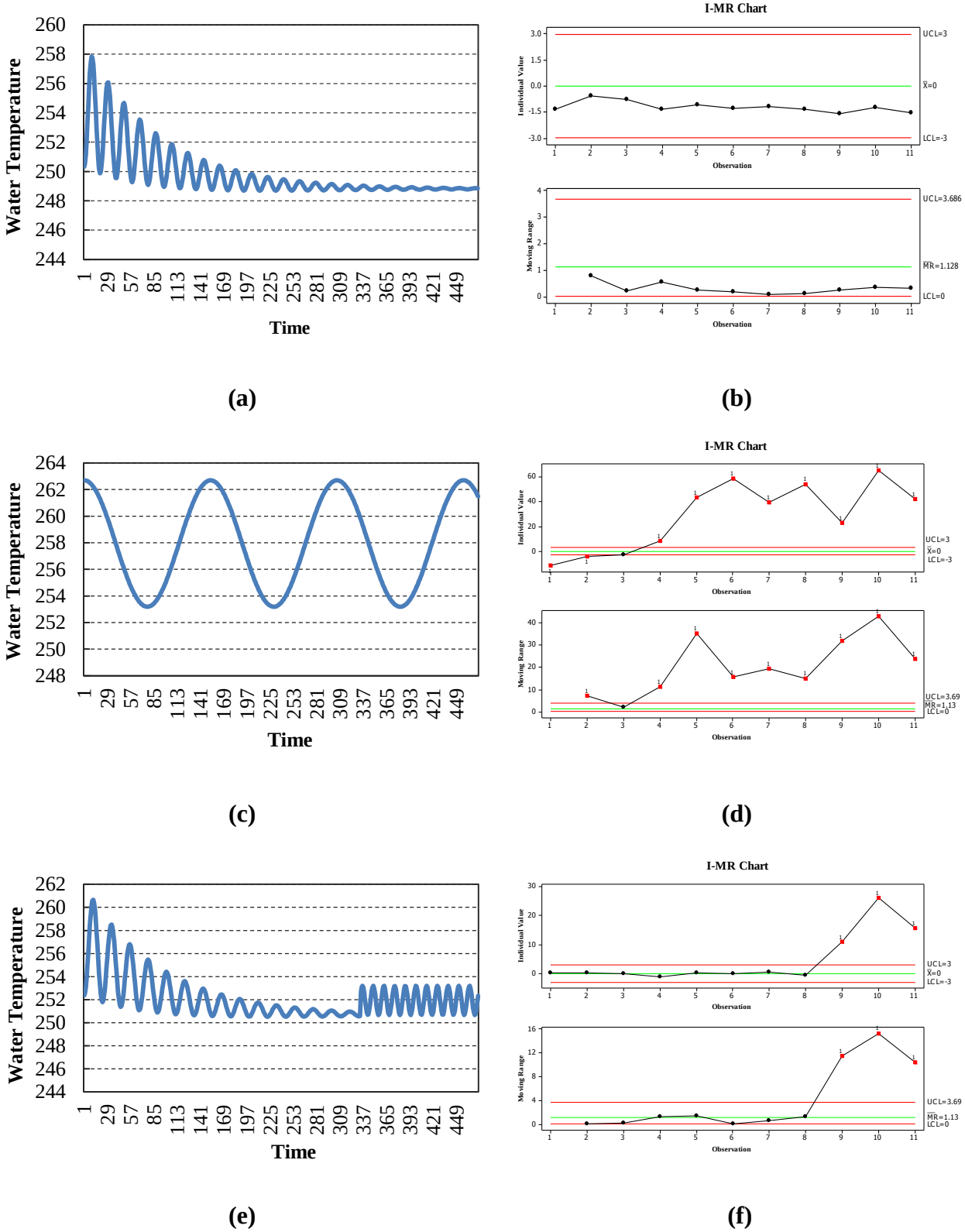
Table 4-3 shows the statistics of Type I and Type II errors. Type I error happens in the case that the profile is truly in-control but is detected as out-of-control. On the other hand, Type II error happens in the cases that the studied profiles are truly a bad one but cannot be detected. As low as possible these two errors are, the more satisfactory the performance of the proposed approach is. By looking again at Table 4-3, three cells are found as blank. Two of them are meant to show the Type I errors and one of them is for displaying the Type II error. This is because the Type I error is not meaningful for the cases that are already out-of-control and Type II error is not applicable for the cases that are in-control. Other reflected statistics in this Table (Type I error for the typical profiles and Type II error for bad profiles), were satisfactory for the proposed methodology on the simulated profiles. Type II error of zero means that there is no bad profile identified as good ones. In other words, it means that all of the bad profiles are detected. Type I error for the typical profiles is also low and demonstrates that the false alarm rose less than 10 percent of the cases.

Table 4-3 Type I and Type II errors in the first experiment

	Typical Profiles (good profiles)	Bad 1 Profiles	Bad 2 Profiles
Type I Error	0.073	-	-
Type II Error	-	0	0

Figure 4-5-a, Figure 4-5-c and Figure 4-5-e illustrate three examples of the three types of simulated profiles while Figure 4-5-b, Figure 4-5-d and Figure 4-5-f reflect their related IX control charts. Figure 4-5-b shows that the IX and MR charts identify that the process is in control correctly. Figure 4-5-d shows that the related profile is detected as out of control from the beginning meaning t= $\frac{1}{8}$. Finally, Figure 4-5-f displays that the related profile is detected as an out-of-control one since t= $\frac{3}{4}$.

Figure 4-5 Examples of the three types of profiles with their related IX control charts



4.3.2 Second Experiment-A Batch of 1000 Consisting 80%, 10% and 10% for Good, Bad 1 and Bad 2 profiles

In the second experiment, a mixture of the three types of profiles was generated. This mixture was created using an algorithm to create 80% of the good profiles, 10% of the bad 1 profiles and 10% of the bad 2 profiles. Again, the goal was to evaluate the performance of the D method, on detecting the bad 1 and the bad 2 profiles as out of control profiles and the good profiles as in control profiles after training the system with the initial 150 Phase I profiles.

Using the IX control chart as the monitoring tool again, the fourth method's performance was evaluated. By utilizing equations (2.11) and (2.12) and having n=2, again, values of d_2, D_3 and D_4 were set as 1.128, 0 and 3.267 respectively. This gives the control limits of -3 and 3 for the individual value chart and 0, 1.128 and 3.686 for the moving range chart.

4.3.2.1 Simulation Outcomes

Using a random algorithm to create the three types of profiles, out of 1000 generated profiles, 781 were good profiles, 111 were bad 1 profiles and 108 were bad 2 profiles. 46 false alarms rose for the first type profiles. The methodology is also able to detect all of the bad or in other words the bad 1 and bad 2 profiles. Table 4-4 displays the average and standard deviation of first time detection time of the bad 1 and the bad 2 profiles in the second experiment. As it is seen from this Table all of the bad 1 profiles are detected from the beginning portions. This is similar to the results of the first experiment and similar reasons are applicable. In summary, the obtained results for the bad 2 profiles are much the same as the ones in the first experiment.

Table 4-4 Average and standard deviation of first detection time in the second experiment

	Bad 1 Profiles	Bad 2 Profiles
Average of first detection time	0.125	0.666
Standard deviation of first detection time	0	0.122

Figure 4-6-Histogram of percentage of the first detection time for the 10 percent bad 2 profiles in the second experiment

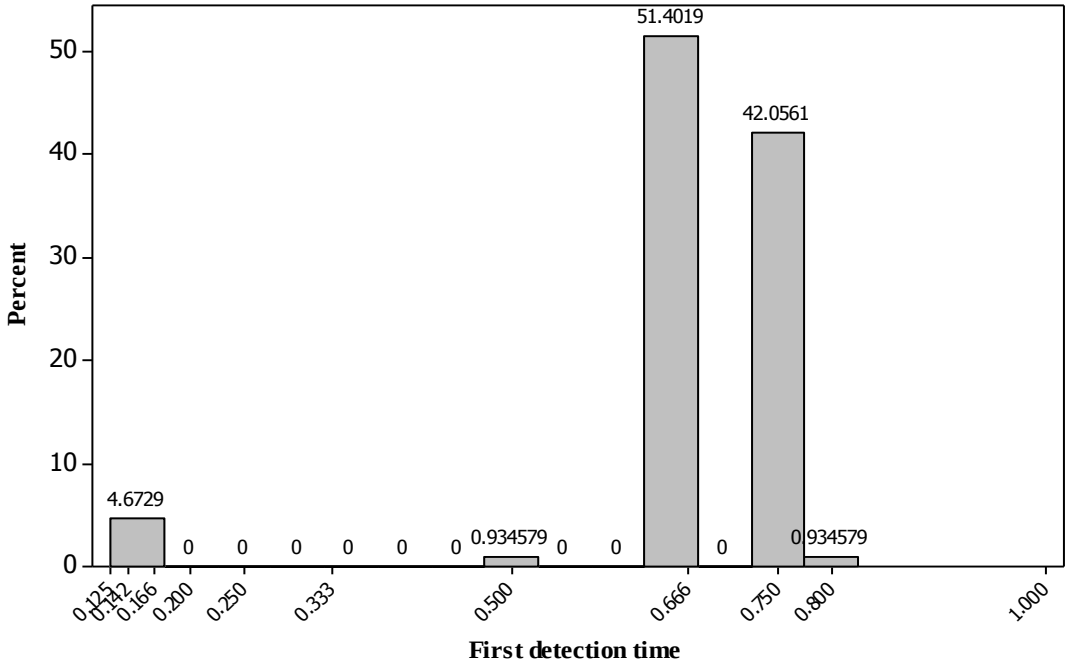


Figure 4-6 displays a histogram of the first alarm time for the 10 percent bad 2 profiles. The eleven values on x axis represent the eleven portions that a profile is divided into. As it is seen from this histogram, majority of the changes are again detected at the relatively last portions. As it was explained for the first experiment, this is due to the type of the changes in the bad 2 profiles that happens at the last segments.

Table 4-5 Type I and Type II errors in the second experiment

	Typical Profiles (good profiles)	Bad 1 Profiles	Bad 2 Profiles
Type I Error	0.058	-	-
Type II Error	-	0	0

Table 4-5 shows the statistics of Type I and Type II errors. Three cells are found as blank in this Table. Two of them are meant to show the Type I errors and one of them is for

displaying the Type II error. As it was explained for the first experiment, this is because the Type I error is not meaningful for the cases that are already out-of-control and Type II error is not applicable for the cases that are in-control. Other reflected statistics in this Table (Type I error for the typical profiles and Type II error for bad profiles), is satisfactory for the proposed methodology on the simulated profiles. Type II error of equal to zero means that there are no bad profiles identified as good ones, or in other words, it means that all of the bad profiles are detected. Type I error for the typical profiles is also low and demonstrates that the false alarm rose almost 5 percent of the cases.

4.3.3 Third Experiment- A Batch of 1000 Consisting 90%, 5% and 5% for Good, Bad 1 and Bad 2 Profiles

Finally, as the last experiment, a mixture of the three types of profiles consisting 90% of good, 5% of bad 1 and 5% of the bad 2 was generated. Again, the goal was to evaluate the performance of the D method, on detecting the bad 1 and the bad 2 profiles as out of control profiles and the good profiles as in control profiles after training the system with the initial 150 Phase I profiles.

Using the IX control chart as the monitoring tool again, the fourth method's performance was evaluated. By utilizing equations (2.11) and (2.12) and having n=2, again, values of d_2, D_3 and D_4 were set as 1.128, 0 and 3.267 respectively. This gives the control limits of -3 and 3 for the individual value chart and 0, 1.128 and 3.686 for the moving range chart.

4.3.3.1 Simulation Outcomes

Using a random algorithm to create the three types of profiles, out of 1000 generated profiles, 891 were good profiles, 57 were bad 1 profiles and 52 were bad 2 profiles. Sixty two false alarms rose for the first type profiles. The methodology is also able to detect all of the bad or in other words the bad 1 and bad 2 profiles. Table 4-6 displays the average and standard deviation of first time detection of the bad 1 and bad 2 profiles in the second experiment. As it is seen from this table all of the bad 1 profiles are detected from the beginning portions. This is again similar to the results of the first and second experiments and similar reasons are applicable.

In summary, the results for the third type (bad 2) profiles are also much the same as the ones in the first experiment.

Table 4-6 Average and standard deviation of first detection time in the third experiment

	Bad 1 Profiles	Bad 2 Profiles
Average of first detection time	0.125	0.666
Standard deviation of first detection time	0	0.153

Figure 4-7 Histogram of the first detection time for the 5 percent bad 2 profiles in the third experiment

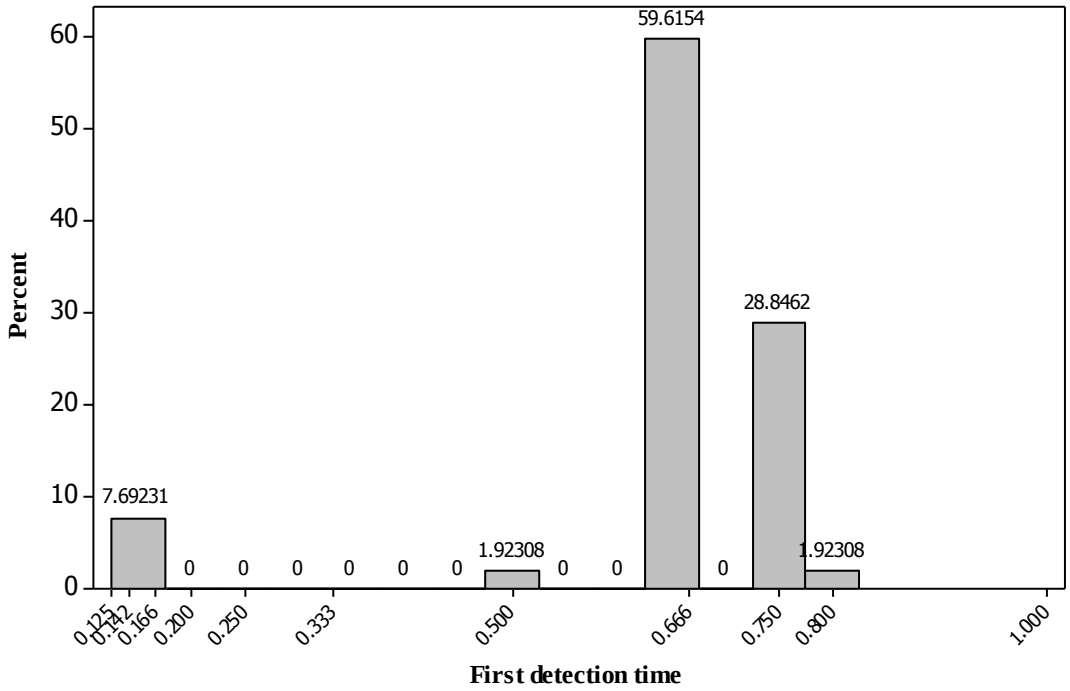


Figure 4-7 displays a histogram of the first detection time for the 5 percent bad 2 profiles. The eleven values on x axis represent the eleven portions that a profile is divided into. As it is seen from this histogram, majority of the changes are again detected at the relatively last portions. As it was explained for the first and second experiments, this is due to the type of the

changes in the third type profiles that happens at the last segments and makes it really difficult to call the changes at the beginning portions.

Table 4-7 Type I and Type II errors in the second experiment

	Typical Profiles (good profiles)	Bad 1 Profiles	Bad 2 Profiles
Type I Error	0.069	-	-
Type II Error	-	0	0

Table 4-7 shows the statistics of Type I and Type II errors. Three cells are found as blank in this Table. Two of them are meant to show the Type I errors and one of them is for displaying the Type II error. As it was explained for the first experiment too, this is because the Type I error is not meaningful for the cases that are already out-of-control and Type II error is not applicable for the cases that are in-control. Other reflected statistics in this Table (Type I error for the typical profiles and Type II error for bad profiles), is satisfactory for the proposed methodology on the simulated profiles. Type II error of equal to zero means that there are no bad profiles identified as good ones, or in other words, it means that all of the bad profiles are detected. Type I error for the typical profiles is also low and demonstrates that the false alarm is raised less than 10 percent of the cases.

Chapter 5 - Conclusions

5.1 Summary of The Research

When quality characteristics are in form of profiles, different sets of tools and methods are needed to be applied on them for statistical process control. Depending on what regression models can better fit to profiles, they are assigned into different categories. Linear and nonlinear profiles are the most common types of profiles in this regard. Noticeable amounts of research have been conducted in the field of profile monitoring. A common aspect of them is that profile monitoring depends on complete profiles. This means that a process is determined to be in control or out of control when the entire profile information is obtained.

In this research, the process monitoring task based on profiles is based on real-time data feed before an entire profile information is obtained. In other words, considering profiles as time series over time, the goal is to predict the status of the profile at the end of time interval, in earlier stages. For example, when the studied quality characteristic is for instance earthquake waves or heart beat rate, it is extremely beneficial to have early detections. In these cases, early change detection could prevent disasters. Thus, great number of lives and amounts of money could be saved. Many processes have quality characteristics in form of profiles as well. As soon as a defective process is detected, time and money could be saved. This may also help save the product under production due to early mal-process detection. The curing process studied through this thesis provides such an example.

In this thesis, five methodologies were proposed for detecting change in wave profiles based on real-time data feed. Four of the five methodologies utilized a single variable and one of them utilized multiple variables. The four univariate methods were called Filtering then standardizing, Standardizing then filtering, Regression analysis, and D method, respectively. The fifth method was called the Multivariate approach.

The first four methods were based on the research conducted by Chang et al. (2011). Their utilized variable was the enclosed area between a wave profile and its exponential-decay function fit. The water temperature profiles were obtained from a curing process of producing hoses to validate their proposed methodologies. The same dataset was utilized to validate the proposed methodologies in this thesis too.

Performances of all the five methods were evaluated on the water temperature profiles. Although the multivariate method had the best performance in monitoring the complete profiles; it was not selected because of the extensive amount of computations needed for monitoring real-time profile data. Instead, the D method was selected as the best method due to its similar performance to the first three proposed methods, the ease of use, and its fewer implementation steps.

D method included the following steps: determining the desired decision times (t's), fitting the exponential decay function to the obtained profiles at time t, calculating the values of enclosed areas between the profiles and the fitted functions from the beginning until time t, calculating the difference of the enclosed areas between time t-1 and time t for each profile, and standardizing the obtained statistics for each profile using the mean and standard deviation of the obtained statistics from the Phase I profiles. Finally, the obtained statistics from these procedures for each of the profiles are fed to the control charts for monitoring them.

In the case study, one hundred and eighty three water temperature profiles were collected to test the methodologies on. From this number of profiles, one hundred profiles were selected for Phase I and eighty three profiles were selected for Phase II analyses. Although 183 water temperature profiles were beneficiary for preliminary evaluation of the proposed methods. It is necessary to conduct a simulation study to study the proposed D method.

Simulations provide various controlled out-of-control scenarios and adequate numbers of profiles for fair comparisons. Three types of profiles in the simulation study include one type of good or typical profiles and two types of bad profiles. Three sets of experiments were conducted using the simulated profiles.

In the first experiment, having 150 good profiles as Phase I profiles, three batches of 1000 good, bad1 and bad2 profiles were created for Phase II analyses and D method have a successful performance with Type I error of 0.073 and Type II error of 0. In addition, all bad 1 profiles can be detected at 0.125 of a complete profile and the bad 2 profiles could be detected in average of 0.681 of a complete profile.

In the second experiment, having the same 150 good profiles for Phase I, a batch of 1000 profiles consisting 80% of good profiles, 10% of bad 1 and 10% of bad2 profiles was created. D method could successfully detect all of the bad profiles again. Type I error was 0.058 and Type

II error 0. The bad 1 profiles could all be detected at 0.125 of a complete profile and the bad 2 profiles could be detected in average of 0.666 of a complete profile.

In the third experiment, again, having the same 150 good profiles for Phase I, a batch of 1000 profiles consisting 90% of good profiles, 5% of bad 1 profiles and 5% of bad 2 profiles was created. D method could again successfully detect all of the bad profiles. Type I error was 0.069 and Type II error 0. The bad 1 profiles could all be detected at 0.125 of a complete profile and the bad 2 profiles could be detected in average of 0.666 of a complete profile.

In summary, it turned out that the proposed D method was able to have a satisfactory performance in detecting changes in simulated wave profiles based on real-time data feed.

5.2 Future Research

In the curing case study in this research, a profile is divided into eleven portions. This means that there were eleven decision points to check if the change in profile was detected or not. This choice seemed to be appropriate due to the shape of the curing process's water temperature profiles. The proposed methodology should be generalized to the cases having other numbers of portions and the obtained results should be compared with the ones of this research.

Further research might also be needed in investigating an algorithm that gives users the opportunity to diagnose a process at any times before a profile is completely generated. Monitoring the process at each data-collection time unit demands huge amounts of computations and is not feasible. This future research strikes a balance between monitoring at each data point and monitoring at fixed points when users suspect that an out-of-control process is pending. The multivariate analysis approach can also be utilized for change detections with real-time data feed. Although the results from the analysis having the complete profiles were better than the univariate methods in the case study, this methodology was not explored further due to its huge amount of computations and its lack of real-time detection capability. Further research is needed to overcome this computational obstacle and enhance its real-time detection functionality.

Finally, the proposed methodologies were tested on the water temperature profiles obtained from the curing process of hose production. Other types of profiles, with different shapes might be needed to further validate the proposed methodologies.

References

Abdel-Salam, Abdel-Salam G., Jeffrey B. Birch and Willis A. Jensen. "A Semiparqametric Mixed Model Approach to Phase I Profile Monitoring." (2012).

Amiri, Amirhossein, Willis A. Jensen and Reza Baradaran Kazemzadeh. "A Case Study on Monitoring Polynomial Profiles in the Automative Industry." 26.5 (2009).

Chang, Shing I and Srikanth Yadama. "Statistical Process Control for Monitoring Non-linear Profiles Using Wavelet Filtering and B-Spline Approximation." *International Journal of Production Research* (2010): 1049-1068.

Chang, Shing I, et al. "An SPC Implementation on Wave Profiles When No Gold Standard Profile is Available: A Case Study." *Submitted to JQT* (2011).

Chang, Tee-Chin and Fah-Fatt Gan. "Monitoring Linearity of Measurement Gauges." *Journal of Statistical Computation and Simulation* (2006): 889-911.

Chen, W. F., Saleeb, A. F.,. *Constitutive equation for engineering materials,*. Vols. 1-elasticity and modeling. New York: Wiley Inerscience, 1982.

Ding, Yu, Li Zeng and Shiyu Zhou. "Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes." *ASQ;38,3* (2006): 199-216.

Gupta, S, D.C Montgomery and W.H Woodall. "Performance Evaluation of Two Methods for Online Monitoring of Linear Calibration Profiles." *International Journal of Production Research* (2006): 1927-1942.

Hosseinifard, S.Z and M Abdollahian. "A Supervised Learning Method in Monitoring Linear Profile." *Seventh International Conference on International Technology*. 2010.

Hosseinifard, S.Z, M Abdollahian and P Zeephongsekul. "Application of Artificial Neural Networks in Linear Profile Monitoring." *Expert Systems with Applications* (2011): 4920-4928.

Jensen, A. Willis, Jeffrey B. Birch and William H. Woodall. "Monitoring Correlation Within Linear Profiles Using Mixed Models." (2008).

Jensen, Willis A and B. Jeffrey Birch. "Profile Monitoring via Nonlinear Mixed Models." (2006).

Kazemzadeh, R.B, R Noorossana and A Amiri. "Phase I Monitoring of Polynomial Profiles." *Communications in Statistics--Theory and Methods* (2008): 1671-1686.

Mahmoud, Mahmoud A., et al. "A Change Point Method for Linear Profile Data." *Quality and Reliability Engineering International* (2006): 247-268.

Montgomery, Douglas C. *Introduction to Statistical Quality Control*. Wiley, 2008.

Riva, P., Cohn, M. Z.,. "Engineering Approach to Nonlinear Analysis of Concrete Structures." *Journal of Structural Division, ASCE* 116 (1990): 2162-2186.

Roberts, S. W. "Control Chart Tests Based on Geometric Moving Average." 1 (1959).

Shiau, Jyh-Jen Horng, et al. "Monitoring Nonlinear Profiles with Random Effecrs by Nonparametric Regression." *Communications in Statistics--Theory and Methods* (2009): 1664-1679.

Williams, James D., William H. Woodall and Jeffrey B. Birch. "Statistical Monitoring of Nonlinear Product and Process Quality Profiles." *Quality and Reliability Engineering International;23* (2007): 925-941.

Woodall, William H. "Current Research on Profile Monitoring." 17.3 (2007).

Zhu, Junjia and Dennis K. J. Lin. "Monitoring the Slopes of Linear Profiles." *Quality Engineering* (2010): 1-12.

Zou, Changliang, Fugee Tsung and Zhaojon Wang. "Monitoring General Linear Profiles Using Multivariate Exponentially Weighted Moving Average Schemes." *Technometrics;49,4* (2007): 395-408.