Multivariate Bayesian Process Control

by

Zhijian Yin

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Mechanical and Industrial Engineering
University of Toronto

© Copyright by Zhijian Yin (2008)
Multivariate Bayesian Process Control

Doctor of Philosophy, 2008

Zhijian Yin

Graduate Department of Mechanical and Industrial Engineering
University of Toronto

Abstract
Multivariate control charts are valuable tools for multivariate statistical process control (MSPC) used to monitor industrial processes and to detect abnormal process behavior. It has been shown in the literature that Bayesian control charts are optimal tools to control the process compared with the non-Bayesian charts. To use any control chart, three control chart parameters must be specified, namely the sample size, the sampling interval and the control limit. Traditionally, control chart design is based on its statistical performance. Recently, industrial practitioners and academic researchers have increasingly recognized the cost benefits obtained by applying the economically designed control charts to quality control, equipment condition monitoring, and maintenance decision-making. The primary objective of this research is to design multivariate Bayesian control charts (MVBCH) both for quality control and conditional-based maintenance (CBM) applications.

Although considerable research has been done to develop MSPC tools under the assumption that the observations are independent, little attention has been given to the development of MSPC tools for monitoring multivariate autocorrelated processes. In this research, we compare the
performance of the squared predication error (SPE) chart using a vector autoregressive moving average with exogenous variables (VARMAX) model and a partial least squares (PLS) model for a multivariate autocorrelated process. The study shows that the use of SPE control charts based on the VARMAX model allows rapid detection of process disturbances while reducing false alarms.

Next, the economic and economic-statistical design of a MVBCH for quality control considering the control limit policy proved to be optimal by Makis(2007) is developed. The computational results illustrate that the MVBCH performs considerably better than the MEWMA chart, especially for smaller mean shifts. Sensitivity analyses further explore the impact of the misspecified out-of-control mean on the actual average cost. Finally, design of a MVBCH for CBM applications is considered using the same control limit policy structure and including an observable failure state. Optimization models for the economic and economic statistical design of the MVBCH for a 3 state CBM model are developed and comparison results show that the MVBCH performs better than recently developed CBM Chi-square chart.
Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Viliam Makis, for his guidance, encouragement and patience throughout my Ph.D. program. I am grateful for the opportunity to work under his excellent guidance. I would also like to thank the other members of my supervisory and examination committee Professors Mireille Broucke, Radu Craiu, Roy Kwon, David Martell, and Baris Balcioglu for their reviews of my work and invaluable feedback.

Also, my sincere thanks go to the staff in the Department of Mechanical and Industrial Engineering, especially Brenda Fung and Lorna Wong.

The financial support from Natural Sciences and Engineering Research Council of Canada (NSERC), Imperial Oil, Syncrude Oil and the University of Toronto Fellowship is gratefully acknowledged.

Many thanks go to the graduate students in our QRM lab and the department, Jianmou Wu, Qiang Miao, Judy Geng, Mohamed Wahab Mohamed Ismail, Stephen Stoyan, Michael Jong Kim, Ming Yang, Lilac Wang, Marshal Wang, Jun Liu, Zhong Ma for their help and friendship.

My deep appreciation goes to my parents for always giving me the opportunity to pursue whatever I was interested in. My sister has also been a tremendous source of encouragement and support. The thesis is dedicated to my beloved husband for his love, support, understanding and encouragement.
## Contents

**Chapter 1 Introduction and Thesis Outline** ................................................................. 1

1.1 Overview .................................................................................................................. 1

1.2 Research Objectives and Contributions ................................................................. 3

**Chapter 2 Application of VARMAX to Activated Sludge Process** ............................... 8

2.1 Introduction ............................................................................................................. 8

2.2 Modelling and SPC for Autocorrelated Process ....................................................... 10

2.2.1 Multivariate Time Series Modelling ................................................................. 10

2.2.2 Literature Review of Multivariate Control Charts .............................................. 12

2.2.2.1 Multivariate Shewhart Charts ...................................................................... 13

2.2.2.2 Multivariate EWMA Charts ........................................................................ 14

2.2.2.3 Multivariate CUSUM Charts ....................................................................... 14

2.2.3 Monitoring Multivariate Processes by Using SPE .............................................. 17

2.3 VARMAX Model for Activated Sludge Process ..................................................... 18

2.3.1 Activated Sludge Process .................................................................................. 18

2.3.2 Model Order Selection ....................................................................................... 22

2.3.3 Model Checking ............................................................................................... 23

2.4 VARMAX and PLS Modeling Comparison ............................................................ 25

2.4.1 Partial Least Squares ....................................................................................... 25

2.4.2 Model Fitting Comparisons of VARMAX and PLS ........................................... 26

2.4.3 Performance Comparisons of VARMAX and PLS ........................................... 28
List of Tables

Table 2.1 Model order selection statistics (BIC) for different values of $p$ and $q$ ......................... 22
Table 2.2 Model order selection statistics (BIC) for different values of $s$ .................................... 22
Table 2.3 Residual cross-correlation matrices .............................................................................. 24
Table 2.4 The performance of the VARMAX model and the PLS model ................................. 27
Table 3.1 Cost comparison of MEWMA and Bayesian charts for Set A ................................. 61
Table 3.2 Cost comparison of MEWMA and Bayesian charts for Set B ................................. 62
Table 3.3 Economic statistical design of Bayesian chart for Set A .............................................. 62
Table 3.4 Results of sensitivity analysis for Set A ($d^2_{act} = d^2_i$) ........................................... 64
Table 3.5 Results of sensitivity analysis for Set B ($d^2_{act} = d^2_i$) ........................................... 64
Table 3.6 Results of sensitivity analysis for Set A ($\mu_{act} = k\mu_i$) ........................................... 65
Table 4.1 Simulated data and control chart statistics ................................................................. 75
Table 4.2 Simulation standard errors for selected simulation runs for Case 1 .................. 110
Table 4.3 Comparison of ECL between the cost model and simulation method .................. 110
Table 4.4 Comparison of ECost between the cost model and simulation method .............. 110
Table 4.5 Comparison of results between Bayesian chart and Chi-square chart ............... 113
Table 4.6 Results for economic statistical design of Bayesian control chart ....................... 114
List of Figures

Figure 2.1 A typical layout of a wastewater treatment process .................................................... 19
Figure 2.2 Layout of the ASWWTP - USP ................................................................................... 21
Figure 2.3 The order selection of the PLS model ........................................................................ 26
Figure 2.4 SPE chart using VARMAX model when the process is in control ............................. 28
Figure 2.5 SPE chart using PLS model when the process is in control ........................................ 29
Figure 2.6 SPE chart using VARMAX model when the process is out of control ....................... 30
Figure 2.7 SPE chart using PLS model when the process is out of control ............................... 30
Figure 3.1 An illustration of the Bayesian control chart ............................................................... 40
Figure 3.2 Process is stopped when $mh < t < mh + nE$ ............................................................. 43
Figure 3.3 Process is stopped when $mh + nE \leq t$ ..................................................................... 43
Figure 4.1 The multivariate Bayesian control chart for the example ........................................... 76
Figure 4.2 The multivariate Chi-square control chart for the example ........................................ 77
Figure 4.3 Condition-based maintenance policy .......................................................................... 79
Figure 4.4 Results for economic and economic statistical design ................................................ 114
Chapter 1 Introduction and Thesis Outline

1.1 Overview

As the competition in all industrial markets continues to grow, industrial engineers now face the problem of adapting to increasing demands for high quality products or services under limited resources. In order to do so and keep manufacturing systems running smoothly, a company must invest significant efforts into quality control and maintenance. Statistical process control (SPC) has been used to control quality for more than five decades and has proven to be extremely effective in helping companies meet the competition challenges. In order to provide a prompt response to a modern computerized process with a large quantity of multivariate data collected on line, multivariate SPC (MSPC) methods have attracted considerable attention by researchers in the past decades. The Primary MSPC function is to identify any assignable causes that result in a shift in the process mean. An assignable cause may be due to operator errors, improperly controlled machines or defective input material. A process is said to be out of control when such an assignable cause is present. A process is in control when only common causes of variation are present. In order to detect if the process is in control or out of control, it is necessary to routinely monitor the process to detect an abnormal process mean shift. Based on the assumption that observations are uncorrelated and normally distributed, a multivariate control chart can be applied to monitor any abnormal changes in a production system causing in a shift in the process mean. Although the covariance matrix of a process may remain the same, a large number of
nonconforming products will be produced if the magnitude of the process mean shift is unacceptable. It may be useful to define exactly what a nonconforming product is. A product is nonconforming when it fails to meet one or more of its specifications. The magnitude of a process mean shift is often called the shift size of the process mean and is usually measured in the process standard deviation units in univariate cases. In multivariate situations, the shift size of the process mean can be measured by Mahalanobis distance (M-distance) between the in-control mean and the out-of-control mean. When the shifts are relatively small, they cannot be easily identified because of process variation.

Multivariate control charts are valuable tools for MSPC to monitor the shifts in a process mean vector. A $T^2$ chart was first proposed by Hotelling (1947), see also Alt (1984). Later, following the success of the $T^2$ chart, more sophisticated methods were proposed, such as a multivariate exponentially weighted moving average (MEWMA) control chart and several multivariate cumulative sum (CUSUM) procedures. The MEWMA and multivariate CUSUM control charts utilize a series of observations rather than simply taking the current sample used in the $T^2$ chart. The MEWMA and CUSUM charts are more sensitive to smaller mean shifts and have the ability to detect the shifts more quickly than the $T^2$ chart. Note that most of multivariate control charts assume that the observations collected from the process are not autocorrelated and follow a multivariate normal distribution.

Another type of a control chart is a Bayesian control chart, which plots the posterior probability that the process is out of control given the process history. In 1952, Girshick and Rubin first presented a Bayesian approach to introduce an economic model to quality control. Their work
was followed by Bather (1963), Taylor (1965, 1967), Eckles (1968), Ross (1971), Carter (1972), and White (1977). Several univariate Bayesian $\bar{X}$ charts for finite production runs have been presented by Tagaras (1994, 1996) and Tagaras and Nikolaidis (2002) and a Bayesian attribute control chart was presented by Calabrese in 1995. Makis (2006a and 2007) presented a theoretical development of a multivariate Bayesian control chart for finite and infinite production runs. It has been shown in the literature that the Bayesian approach is optimal to control the process compared with the non-Bayesian charts. To determine the structure of the optimal policy under partial observations, Makis (2006a and 2007) considered fixed sample size and sampling interval and proved that a control limit policy is optimal for both finite and infinite horizon problems. Makis (2006b), presented the structure of the optimal control policy to maintain an $n$-state system with a continuous observation process.

1.2 Research Objectives and Contributions

Considerable research has been done to develop MSPC tools under the assumption that the observations from a multivariate process are independent. In the presence of autocorrelation, however, the rate of control chart false alarms tends to increase. One of the well accepted methods to monitor autocorrelated processes is a residual approach (Alwan and Roberts (1988), Shu et al. (2002)). The general idea of this residual approach is to use a time series technique to model the process and then apply a control chart to the residuals of the model. Since the residuals are approximately uncorrelated and normally distributed when the process is in control, traditional SPC tools can therefore be used to monitor such residuals. However, little attention has been given to the development of MSPC tools that can be used to monitor the multivariate autocorrelated processes.
The first objective of this research is to apply a vector autoregressive moving average with exogenous variables (VARMAX) model to an autocorrelated process in order to monitor such process by a control chart. More specifically, a VARMAX model is fitted to the in-control historical process data offline and residuals can be calculated based on the collected observation and the VARMAX model. Since the residuals should approximately follow normal distribution and be independent, a multivariate control chart can be applied to the residuals. In Chapter 2, we build a VARMAX model using the data obtained from the Activated Sludge Wastewater Treatment Plant - University of Sao Paulo (ASWWTP-USP) simulator when the process is in control and compare the behavior of the calculated SPEs in an out-of-control situation with the SPE obtained from a partial least squares (PLS) model. The VARMAX model order is selected using Schwarz’s Bayesian criterion to model the activated sludge process simulated by the ASWWTP - USP simulator. We find that the VARMAX model performs considerably better in both in-control and out-of-control situations. When the process is in control, the number of false alarms on the SPE chart based on the PLS model is much higher than on the SPE chart based on the VARMAX model.

As discussed earlier, the use of control charts is an effective and efficient way for companies to maintain or improve product quality. To use a control chart, three control chart parameters must be specified, namely the sample size, the sampling interval and the control limit. Traditionally, control chart design is based on its statistical performance for quality control. Probabilities of Type I and II errors have been calculated to measure the statistical performance of a simple control chart, such as a Chi-square chart. Type I error occurs when a sample point is outside the
control limits when the process is in control and type II error occurs when a sample point is between the control limits when the process is out of control. Average run lengths (ARLs) need to be used to measure the statistical performance of more complex control charts, such as a MEWMA chart. Given the statistical performance requirements when the process is in control and out of control region, the control chart parameters can be determined. The design of control charts can be based also on cost minimization and such design is called the economic design of control charts. Since the first economic design of a control chart (Duncan (1956)), many researchers have studied the economic designs of other control charts. Later, economic-statistical designs for $\bar{X}$ and R charts were proposed by Saniga whose work was motivated by Woodall’s (1985, 1986) criticisms on the possibility of poor statistical performance of economically-designed control charts. The economic statistical design of a univariate EWMA chart by Montgomery et al. (1995) was extended to the multivariate case by Linderman and Love (2000).

Industrial practitioners and academic researchers have increasingly recognized the economic benefits obtained by applying SPC techniques to maintenance decision-making. The efforts can be found in the literature (e.g. Tagaras (1988), Cassady et al. (2000), Linderman et al. (2005) and Ivy and Nembhard (2005)). One common assumption in these papers is that there are two system states, an in-control state and an out-of-control state. This is a reasonable assumption for quality control applications. However, such an assumption is usually not appropriate for maintenance modeling. Maintenance management aims to avoid costly system failures, so a failure state should be considered in maintenance models. Wu (2006) considered an observable failure state in his contribution to the economic and economic statistical design of the multivariate Chi-square chart for condition-based maintenance (CBM) applications, and similar development of the
The MEWMA chart can be found in Liu (2006). It is shown in Makis (2007 and 2006a) that the multivariate Bayesian chart has better performance than the multivariate Chi-square and MEWMA charts. Hence, further study of the design of the Bayesian chart is very valuable. In this research, we will design a multivariate Bayesian control chart considering the class of control limit policies proved to be optimal by Makis (2006a, 2006b, 2007) for both quality control and a 3 state CBM model, i.e. we want to find the control limit, the sample size, and the sampling interval minimizing the long-run expected average cost.

The second objective of this research is to present the economic and economic-statistical design of a multivariate Bayesian control chart for quality control. In Chapter 3, we develop an optimization model for the economically-designed multivariate Bayesian control chart to determine the optimal sample size, sampling interval, and control limit minimizing the total average cost of the process. Our approach has some similarities with the Lorenzen and Vance’s approach, but it differs in the way in which the expected cycle cost and cycle length are derived. Our approach is more suitable for the economic design of an advanced control chart, such as the Bayesian control chart. A Markov chain approach is used to analyze the ARL behavior of the Bayesian chart and to derive the formulas for the expected cycle length and cost. The closed form of the expected cost per time unit is obtained and used as the objective function of the optimization model for the economically-designed Bayesian chart. Statistical constraints are added for the economic statistical design of the chart to overcome the weakness of the pure economical design of the chart. The performance comparison of the Bayesian chart and MEWMA chart illustrates that the Bayesian chart performs considerably better than the
MEWMA chart, especially for smaller shifts of the process mean. Sensitivity analyses further explore the effect of the misspecified out-of-control mean on the actual average cost.

The third objective of this research is to design a multivariate Bayesian control chart for CBM applications by considering the same control limit policy structure and including an observable failure state. In Chapter 4, a multivariate Bayesian control chart is designed for CBM applications and illustrated by a numerical example. Optimization models for the economic and economic statistical design of the control chart are developed to determine the optimal control chart parameters that minimize the total expected average maintenance cost. Comparison results of the Bayesian chart and the CBM Chi-square chart show that the Bayesian chart performs better than the CBM Chi-square chart and the savings are higher for stricter statistical requirements.
Chapter 2 Application of VARMAX to Activated Sludge Process

2.1. Introduction

MSPC methods techniques can be used to monitor all variables simultaneously, including the process variables $X$ and the quality variables $Y$. MSPC can provide a prompt response to a modern computerized production process with a large quantity of complex information collected online. One of the most important objectives of SPC is to monitor the process in order to verify whether the process is operated under normal condition. During the past decades, the research in this area has been very active, and a large number of publications have covered a variety of industrial processes, such as chemical, manufacturing, and pharmaceutical processes. Two extensive review papers in MSPC were written by MacGregor and Kourti (1995) and Qin (2003).

Multivariate control charts are useful tools for MSPC. To monitor the process mean vector, the Chi-square and $T^2$ charts were first proposed by Hotelling (1947) (see also Alt 1984). Later, with the success of the $T^2$ chart, more sophisticated methods were proposed, such as a MEWMA control chart and several multivariate CUSUM procedures. The MEWMA and multivariate CUSUM control charts utilize a series of observations rather than simply taking the current sample used in the $T^2$ chart. These additional observations allow the MEWMA and CUSUM
charts to be more sensitive to smaller mean shifts and have the ability to detect the shifts more quickly than the $T^2$ chart. The traditional SPC charts are based on the assumption that the process observations are uncorrelated. However, because the time interval between two subsequent samples can be small in modern industry, such assumption is often violated in collected real data that exhibits both autocorrelation and cross-correlation and hence, the traditional control charts are not directly applicable.

Two useful practical tools to develop process models for MSPC are the principal component analysis (PCA) and the PLS method, also called projection to latent structures. These MSPC methods have attracted considerable attention from many researchers who have considered MSPC schemes by using $T^2$ and $Q$ charts for monitoring continuous and batch processes (see e.g. MacGregor & Kourti, 1995; Kesavan, Lee, Saucedo & Krishnagopalan, 2000; Simoglou, Martin & Morris, 2000; Lee, Yoo & Lee, 2004). PCA is often used for dimensionality reduction of a large number of variables assuming the subsequent samples are independent. PLS has been found extensively useful when the process data as well as the quality data are available and need to be considered. However, the statistical properties of the resulting parameter estimates obtained from these methods have not been well researched. Moreover, the traditional $T^2$ chart or $T^2_\delta$ chart based on principal components obtained after applying PCA or PLS modeling are still not suitable for process monitoring due to serial correlation.
Many researchers have studied SPC methods for univariate autocorrelated processes and applied state space models for multivariate autocorrelated processes. A residual approach for univariate autocorrelated process monitoring was proposed in Alwan and Roberts (1988). When the process is in control, the residuals behave as a univariate white noise and therefore a traditional control chart can be applied to monitor the residuals. Pan and Jarrett (2004) extended this approach to a multivariate process monitoring and mentioned that the vector autoregressive (VAR) models are at least as good as, and often better than, the state space model in their paper. Östermark and Saxén (1996) built a VARMAX model for a blast process. However, very few papers have considered multivariate time series models for MSPC.

In this chapter, a VARMAX model will be used to model the multivariate time series processes. A Q control chart based on the SPE calculated from one-step-ahead forecasting errors will be applied to processes monitoring. This methodology will be applied to model and monitor the nitrate concentrations in the anoxic and aerobic zones of the activated sludge process simulated by the ASWWTP-USP simulator and then compared with a PLS model.

2.2 Modelling and SPC for Autocorrelated Process

2.2.1 Multivariate Time Series Modelling

A VARMAX model is a VARMA model with exogenous input variables which are represented by ‘X’. Let \( \{ y_t \} \) denote a \( k \)-dimensional vector time series of endogenous variables and \( \{ x_t \} \)
denote an $n$-dimensional vector time series of exogenous variables. Assuming that the mean vectors have been subtracted from the initial observation vectors, the VARMAX($p, q, s$) model can be expressed as follows

\[ y_t + \sum_{j=1}^{p} \Phi_j y_{t-j} = \sum_{j=0}^{s} \Theta^*_j x_{t-j} + \epsilon_t + \sum_{j=1}^{q} \Theta_j \epsilon_{t-j} \]  

(2.1)

where $\Phi_j$ and $\Theta_j$ are $k \times k$ matrices, $\Theta^*_j$ is a $k \times n$ matrix, and $\epsilon_t$ is a vector white noise process with mean vector $\theta$ and covariance matrix $\Sigma = E(\epsilon_t \epsilon_t')$.

When the multivariate time series follows a pure VAR model, the method of least squares can be applied to estimate the model parameters efficiently. However, when the model is not a pure AR model, maximum likelihood method and its extensions need to be chosen to estimate the parameters. Currently, the procedure of VARMAX provided by SAS/ETS cannot obtain the estimation results within a reasonable time for mixed vector time series models. In this study, VARMAX model coefficients were estimated using exact maximum likelihood, which has been implemented in $E^4$, a Matlab Toolbox written by Terceiro et al. (2000).

The higher model order may improve the model fitting, because of a larger number of introduced parameters. However, it can cause model overfitting and an overfitted model may bring very poor prediction results. Akaike Information Criterion (AIC) can be used to choose the most appropriate model order. An alternative approach is to use Schwarz’s Bayesian Criterion (SBC), also known as Schwarz’s Bayesian Information Criterion (BIC), which is defined as
BIC = \(-2\log(L) + r\log(N)\) \hspace{1cm} (2.2)

where \(L\) is the maximum value of the likelihood function, \(r\) is the number of parameters estimated by maximum likelihood in the VARMAX model, and \(N\) is the sample size. The simulation results in Takimoto (2001) show that AIC might overestimate the order as the sample size increases. Takimoto and Hosoya (2004) mentioned that the order determined by BIC is more concentrated around the true order. Therefore, BIC is used in this study to compare various models fitted by the exact maximum likelihood and the model that has a minimum value of BIC is chosen as the best choice.

### 2.2.2 Literature Review of Multivariate Control Charts

Simply using univariate control charts to monitor cross-correlated quality variables individually may cause high false alarm rate and fail to signal when the process is out of control. Multivariate control charts can capture the relationships between the variables and thus overcome the disadvantage of univariate control charts. Several multivariate control charts, such \(T^2\), MEWMA and MCUSUM charts, have been presented in the literature. A review in this area can be found in Lowry and Montgomery (1995). A brief literature review of the multivariate control charts is given in this section. Given a process defined by \(q\) quality characteristics that follows a multivariate normal distribution \(N_q(\mu_0, \Sigma)\) when the process is in control and follows \(N_q(\mu_1, \Sigma)\) when the process is out of control, a multivariate control chart is designed to monitor the mean shift. \(y_i\) denotes the \(i^{th}\) observation from the process.
2.2.2.1 Multivariate Shewhart Charts

A natural extension to the univariate Shewhart chart is presented in Hotelling (1947). Alt (1984) provided more detailed discussion of the chart. The statistic of the multivariate Shewhart chart can be computed as follows,

\[ \chi_i^2 = (\mathbf{y}_i - \mathbf{\mu}_0)^\top \mathbf{\Sigma}^{-1} (\mathbf{y}_i - \mathbf{\mu}_0), \quad i = 1, 2, \ldots \]  

(2.3)

One can prove that the statistic has a central Chi-square distribution with q degrees of freedom when \( \mathbf{y} \sim N_q(\mathbf{\mu}_0, \mathbf{\Sigma}) \). When \( \mathbf{y} \sim N_q(\mathbf{\mu}_1, \mathbf{\Sigma}) \), it follows a noncentral chi-square distribution with the q degrees of freedom and the noncentrality parameter \( d^2 \), where \( d^2 = (\mathbf{\mu}_1 - \mathbf{\mu}_0)^\top \mathbf{\Sigma}^{-1} (\mathbf{\mu}_1 - \mathbf{\mu}_0) \).

\( d^2 \) is also known as the M-distance between \( \mathbf{\mu}_1 \) and \( \mathbf{\mu}_0 \). Since the statistics follow a chi-square distribution, this Shewhart chart is called a Chi-square chart.

In the case that the covariance matrix \( \mathbf{\Sigma} \) is unknown, it needs to be estimated from a sample of n observations as \( \mathbf{S} = (n-1)^{-1} (\mathbf{Y} - \overline{\mathbf{y}} \cdot \mathbf{1}_{\text{vec}})(\mathbf{Y} - \overline{\mathbf{y}} \cdot \mathbf{1}_{\text{vec}})^\top \), where \( \mathbf{Y} \) is a \( q \times n \) observation matrix and \( \overline{\mathbf{y}} \) is the \( q \times 1 \) sample mean vector. The statistic of the Hotelling chart is then given by

\[ T_i^2 = (\mathbf{y} - \mathbf{\mu}_0)^\top \mathbf{S}^{-1} (\mathbf{y} - \mathbf{\mu}_0), \quad i = 1, 2, \ldots \]  

(2.4)

This chart is called as a \( T^2 \) control chart. Given the probability of type I error, \( \alpha \), the upper control limit of the \( T^2 \) control chart can be calculated as follows,

\[ T^2_{UCL} = \frac{(n-1)(n+q)}{n(n-q)} F_{d}(q, n-q) \]  

(2.5)
where $F_{\alpha}(q, n-q)$ is the upper $\alpha$ percentage point of a F-distribution with degrees of freedom $q$ and $n-q$.

### 2.2.2.2 Multivariate EWMA Charts

Lowry et al. (1992) developed a multivariate extension of the univariate EWMA control chart introduced by Roberts (1959). The statistic of the multivariate EWMA control chart is defined as

$$Z_i = R_y i + (1-R)Z_{i-1}, \ i = 1, 2, \ldots,$$

where $Z_0 = 0$ and $R = \text{diag}(r_1, r_2, \ldots, r_q)$, $0 < r_j \leq 1$, $j = 1, 2, \ldots, q$. If $r_1 = r_2 = \ldots = r_q = r$, the statistic can be simplified as

$$Z_i = r y_i + (1-r)Z_{i-1}, \ i = 1, 2, \ldots,$$

The chart gives an out-of-control signal when

$$Q_i = Z_i^T \Sigma_{Z_i}^{-1} Z_i > h$$

where $h$ is the upper control limit and $\Sigma_{Z_i}$ is the covariance matrix of $Z_i$. Note that there is no lower control limit and the statistic $Q_i$ is nonnegative since $\Sigma_{Z_i}$ is semi-positive definite matrix. The control limit $h$ can be chosen by statistical design based on ARLs requirement or economic design of the control chart. Prabhu and Runger (1997) modified the Markov chain approach by Brook and Evans (1972) to study the average run length performance of the MEWMA control chart.

### 2.2.2.3 Multivariate CUSUM Charts
After Woodall and Ncube (1985) suggested using simultaneous univariate CUSUM charts to monitor each of the \( q \) quality characteristics, several multivariate CUSUM procedures have been proposed by Healy (1987), Crosier (1988), Pignatiello and Runger (1990), and Runger and Testik (2004). In this section, the multivariate CUSUM charts by these authors will be briefly reviewed.

Healy (1987) presented a multivariate CUSUM control chart that signals when

\[
S_i = \max(S_{i-1} + \mathbf{a}'(\mathbf{y}_i - \mu_0) - 0.5d, 0) > h, \quad i = 1,2,\ldots,
\]

(2.8)

where \( \mathbf{a}' = \frac{(\mu_1 - \mu_0)'\Sigma^{-1}}{[(\mu_1 - \mu_0)'\Sigma^{-1}(\mu_1 - \mu_0)]^{1/2}} \), \( d = [(\mu_1 - \mu_0)'\Sigma^{-1}(\mu_1 - \mu_0)]^{1/2} \), which is the M-distance between \( \mu_0 \) and \( \mu_1 \), and \( h \) is the upper control limit. Note that \( \mathbf{a}'(\mathbf{y}_i - \mu_0) \) has a standard normal distribution when the process is in control. When the process is out of control, \( \mathbf{a}'(\mathbf{y}_i - \mu_0) \) has a normal distribution with mean \( D \) and variance 1. This procedure gives good ARL performance with correct estimation of the out-of-control mean vector.

Crosier (1988) proposed two multivariate CUSUM schemes. The statistic of the multivariate CUSUM scheme that has better ARL performance is defined as

\[
Q_i = [\mathbf{s}_i'\Sigma^{-1}\mathbf{s}_i]^{1/2}, \quad i = 1,2,\ldots,
\]

(2.9)

where

\[
\mathbf{s}_i = \begin{cases} 
0, & \text{if } C_i \leq k \\
(s_{i-1} + \mathbf{y}_i - \mu_0)(1 - k / C_i), & \text{if } C_i > k
\end{cases} \quad \text{for } i = 1,2,\ldots,
\]

and \( C_i = [((s_{i-1} + \mathbf{y}_i - \mu_0)')\Sigma^{-1}(s_{i-1} + \mathbf{y}_i - \mu_0)]^{1/2} \).

The multivariate CUSUM scheme signals when \( Q_i > h \), where \( h \) is the control limit.
Crosier suggests to choose \( k = d/2 \) in order to minimize the out of control ARL for the selected value of \( d \) and for a given on-target ARL.

Pignatiello and Runger (1990) developed another multivariate CUSUM scheme (MC1). Their best-performing control chart is based on the vector of cumulative sums \( C_i = \sum_{j=-n+1}^{i} (y_j - \mu_0) \) for \( i = 1, 2, \ldots \). Then, \( \frac{1}{n_i} C_i = \left( \frac{1}{n_i} \sum_{j=-n+1}^{i} y_j \right) - \mu_0 \), which represents the difference between the accumulated sample average and the target value for the mean. The MC1 statistic can be constructed as

\[
MC1 = \max \left\| C_i \right\| - kn_i, 0 \right\|, \quad (2.10)
\]

where \( n_i = \begin{cases} n_{i-1} + 1 & \text{if } MC_{i-1} > 0 \\ 1 & \text{otherwise} \end{cases} \) and \( \left\| C_i \right\| = \left[ C_i^\top \Sigma^{-1} C_i \right]^{1/2} \) which can be used as a measurement of the distance of the estimate of the mean of the process from the target mean for the process. The MC1 scheme signals when \( MC1_i > h \), where \( h \) is the upper control limit. \( k \) can be chosen to be equal to half of the distance of \( \mu_0 \) from \( \mu_1 \). The authors believe that alternate choices of \( k \) may improve the performance of the MC1 chart. Note that the ARL of the MC1 scheme cannot be modeled by the Markov chain approach. A Monte Carlo simulation was used by Pignatiello and Runger to evaluate the ARL performance of this scheme. In their paper, the correlated quality variables are first transformed to uncorrelated variables with identity covariance matrix by PCA.
Runger and Testik (2004) presented several multivariate extensions of the CUSUM chart. The one with best ARL performance among these extensions is called MCONE. This procedure is based on a generalized likelihood ratio (GLR) method. The chart signals if there exists a value of $\tau$ such that

$$
\sqrt{\left[\sum_{i=\tau}^{t} (y_i - \mu_0')\right] \Sigma^{-1} \left[\sum_{i=\tau}^{t} (y_i - \mu_0)\right]} > \frac{c}{d} + \frac{d}{2}(t-\tau+1)
$$

(2.11)

where $d$ is the M-distance between $\mu_0$ and $\mu_t$ and $c$ needs to be determined to calculate the control limit. Note that this is V-mask type of control chart in which the control limits are not constant and depend on the time when a sample is collected.

### 2.2.3 Monitoring Multivariate Processes by Using SPE

The first step of MSPC is to fit the process by a suitable model based on the historical data obtained from normal operating condition. After a model has been fitted to the in-control data, the next step is to choose a multivariate control chart to detect if any assignable causes occur in the process. The squared prediction error denoted by $\text{SPE}_{y_{new}}$, also known as Q statistic (Jackson, 1991), for any new observation $y_{new}$ is defined as,

$$
\text{SPE}_{y_{new}} = \sum_{j} (y_{new,j} - \hat{y}_{new,j})^2 = (y_{new} - \hat{y}_{new})^T(y_{new} - \hat{y}_{new})
$$

(2.12)

where $\hat{y}_{new}$ is predicted by the fitted model, such as a VARMAX model. Note that $\text{SPE}_{y_{new}}$ is the distance between $y_{new}$ and $\hat{y}_{new}$. $\text{SPE}_{y_{new}}$ is plotted on a control chart, which is called as SPE control chart. Small value of $\text{SPE}_{y_{new}}$ indicates that the current observation can be well predicted by the fitted model. In other word, small value of $\text{SPE}_{y_{new}}$ can indicate that the process is still in
control. Box (1954) proposed an approximate way to calculate the upper control limit with a significance level $\alpha$ for the SPE control chart. The upper control limit can be calculated by

$$UCL_{\text{SPE}} = g \chi^2_{h, \alpha},$$  \hspace{1cm} (2.13)

where $g = \theta_2 / \theta_1$ and $h = \theta_1^2 / \theta_2$. $\theta_1$ and $\theta_2$ are equal to $\sum_{i=1}^{k} \lambda_i$ and $\sum_{i=1}^{k} \lambda_i^2$, respectively. Here, $\lambda_i$ is the eigenvalue of the covariance matrix $\Sigma$ of the residuals.

2.3 VARMAX Model for Activated Sludge Process

2.3.1 Activated Sludge Process

Wastewater is usually processed by several different treatments in a modern wastewater treatment plant, before it can be released back to the nature. In Figure 2.1, a typical scheme for a wastewater treatment plant is given. There are four sections of the wastewater treatment plant in Figure 2.1.
The activated sludge process (ASP) is an important process in a wastewater treatment plant which incorporates a large number of biological, physiological and biochemical processes. Advanced control schemes for ASP have been studied intensively to meet environmental regulation requirements. However, not much research has been done on using MSPC tools to monitor the wastewater treatment process. Li and Qin (2001) identified a direct dynamic principle component analysis (DPCA) model and indirect DPCA model for a wastewater treatment process. In their indirect DPCA model, an errors-in-variables subspace model
identification algorithm was used to obtain the estimate of the DPCA model. Sotomayor et al. (2003) applied subspace-based algorithms to model an activated sludge process. Lee et al. (2004) used independent component analysis (ICA) for process monitoring. ICA is a multivariate analysis method to decompose multi-dimensional data linearly into factors or components which are statistically independent. Although the ICA method can extract useful information from observations, it is not able to determine the order of the dynamic system.

In this chapter, the VARMAX model is used to model the continuous-flow pre-denitrifying ASP, simulated by the ASWWTP-USP (Activated Sludge Wastewater Treatment Plant – University of São Paulo) benchmark (Sotomayor et al. 2001), and to monitor this process. The layout of the reactors used in the simulator is shown in Fig. 2.2. Based on the Activated Sludge Model No. 1, this simulation benchmark includes the processes of organic matter removal, nitrification and denitrification of domestic effluents. Eliminating the carbonaceous organic matter is the main goal in the wastewater. However, recycling nitrogen should be considered as well when the ammonia and ammonium in the sewage are treated. The biological recovery of nitrogen occurs in two phases, the nitrification, where autotrophic bacteria under aerobic conditions convert ammonium into nitrate, and the denitrification, where heterotrophic bacteria under anoxic conditions use organic compounds as a reducing agent to convert nitrate into nitrogen gas. In the simulator, the denitrification phase takes place in the anoxic zone and it is necessary to promote a recirculation of the water rich in nitrate, which is generated in the aerobic zones of the bioreactor,
back into the anoxic zone for denitrification.

The set of process variables selected in this study includes the influent flow $Q_{in}$, the internal recycle flow rate $Q_{int}$, the external carbon dosage $Q_{ext}$, the influent readily biodegradable substrate $S_{S_{in}}$, the influent ammonium concentration $S_{N_{H_{in}}}$, and the influent nitrate concentration $S_{N_{O_{in}}}$. The quality variables are the nitrate concentration in the anoxic zone $S_{N_{O_{1}}}$ and in the second aerobic zone $S_{N_{O_{3}}}$. The in-control historic data were obtained at a sampling rate of 0.16hr in the total period of 176hrs to fit a VARMAX model offline. Note that the data from the first 64hrs were discarded in order to eliminate the effects of the initial settings. The control of nitrate Sno in the anoxic zone and aerobic zone is the main focus in this research. The process variable set is $X = [Q_{in}, Q_{int}, Q_{ext}, S_{N_{H_{in}}}, S_{N_{O_{in}}}, S_{S_{in}}]$ and the quality variable set is $Y = [S_{N_{O_{1}}}, S_{N_{O_{3}}}]$. The data was pre-processed by subtracting the mean and then divided by the standard deviation from each column of the data, i.e. mean zero and variance one.

Figure 2.2 Layout of the ASWWTP - USP
2.3.2 Model Order Selection

The VARMAX($p$, $q$, $s$) model coefficients were estimated by E4 for each combination of $p = 0, 1, 2, 3, 4, 5$, $q = 0, 1, 2, 3, 4$ and $s = 0$. Table 2.1 presents the values of BIC for values of $p$ equal to 0, 1, 2, 3, 4 and $q$ equal to 0, 1, 2, 3, 4, 5. In Table 2.1, the VARMAX(5, 3, 0) has the minimum BIC value, -2.9808, and BICs of the VARMAX(4, 2, 0) and VARMAX(3, 3, 0) are -2.9588 and -2.9747, respectively, which are near to the minimum of BICs. In order to choose a suitable order of $s$, different value of $s$, the BICs are calculated for VARMAX(5, 3, $s$), VARMAX(4, 2, $s$) and VARMAX(3, 3, $s$) with $s = 0$ and the results are shown in Table 2.2. From Table 2.2, the VARMAX(3, 3, 1) should be chosen since it has the minimum value of BIC, -2.9825.

Table 2.1 Model order selection statistics (BIC) for different values of $p$ and $q$

<table>
<thead>
<tr>
<th>The order of $p$</th>
<th>The order of $q$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.9207</td>
<td>-2.5003</td>
<td>-2.6396</td>
<td>-2.6659</td>
<td>-2.6684</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-2.7507</td>
<td>-2.8087</td>
<td>-2.8876</td>
<td>-2.9212</td>
<td>-2.8831</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-2.7656</td>
<td>-2.9021</td>
<td>-2.8876</td>
<td>-2.9212</td>
<td>-2.8831</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-2.7519</td>
<td>-2.9004</td>
<td>-2.9588</td>
<td>-2.9517</td>
<td>-2.9609</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-2.7469</td>
<td>-2.8720</td>
<td>-2.9414</td>
<td>-2.9808</td>
<td>-2.9629</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2 Model order selection statistics (BIC) for different values of $s$

<table>
<thead>
<tr>
<th>VARMAX($p$, $q$, $s$)</th>
<th>$s$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3, 3, $s$)</td>
<td>-2.9747</td>
<td>-2.9825</td>
<td></td>
</tr>
<tr>
<td>(4, 2, $s$)</td>
<td>-2.9588</td>
<td>-2.9297</td>
<td></td>
</tr>
<tr>
<td>(5, 3, $s$)</td>
<td>-2.9808</td>
<td>-2.9458</td>
<td></td>
</tr>
</tbody>
</table>

The VARMAX(3, 3, 1) model estimated by the exact maximum likelihood is shown as follows:
\[
y_t = \hat{\Theta}_0^* x_t + \hat{\Theta}_1^* x_{t-1} + \epsilon_t + \hat{\Theta}_2^* \epsilon_{t-1} + \hat{\Theta}_3^* \epsilon_{t-2} + \hat{\Theta}_4^* \epsilon_{t-3} - (\hat{\Phi}_1^* y_{t-1} + \hat{\Phi}_2^* y_{t-2} + \hat{\Phi}_3^* y_{t-3})
\]  (2.14)

where the coefficient matrixes with estimated standard deviations in brackets are

\[
\hat{\Phi}_1 = \begin{bmatrix}
-1.7143 & 0.0031 \\
(0.1811) & (0.1963)
\end{bmatrix},
\hat{\Phi}_2 = \begin{bmatrix}
0.9104 & 0.0558 \\
(0.2748) & (0.3132)
\end{bmatrix},
\hat{\Phi}_3 = \begin{bmatrix}
-0.1435 & -0.0771 \\
(0.1118) & (0.1366)
\end{bmatrix},
\hat{\Theta}_1 = \begin{bmatrix}
0.5186 & -1.4857 \\
(0.0797) & (0.1997)
\end{bmatrix},
\hat{\Theta}_2 = \begin{bmatrix}
-0.7848 & 0.2535 \\
(0.1126) & (0.3193)
\end{bmatrix},
\hat{\Theta}_3 = \begin{bmatrix}
0.3009 & 0.2274 \\
(0.0516) & (0.1364)
\end{bmatrix},
\hat{\Phi}_0^* = \begin{bmatrix}
0.0107 & -0.1227 & 0.0235 & -0.0135 & 0.0013 & 0.0280 \\
(0.0100) & (0.0163) & (0.0062) & (0.0114) & (0.0145) & (0.0103)
\end{bmatrix},
\hat{\Phi}_1^* = \begin{bmatrix}
-0.005 & 0.0345 & -0.0041 & 0.0065 & 0.0159 & 0.0002 \\
(0.0106) & (0.0112) & (0.0061) & (0.0080) & (0.0091) & (0.0090)
\end{bmatrix}
\]

with \( \hat{\Sigma} = \)

\[
\begin{bmatrix}
0.0564 & 0.0191 \\
(0.0028) & (0.0010)
\end{bmatrix},
0.0191 & 0.0085 \\
(0.0010) & (0.0004)
\]

### 2.3.3 Model Checking

In this section, the residuals from the fitted VARMAX(3, 3, 1) model are examined to check
model adequacy. As discussed before, one of the assumptions of the VARMAX model is that the residuals $\varepsilon_t$ should approximately be a vector white noise process. Thus, most of the elements of residual correlation matrices should be within the two standard error limits, $\pm 2/\sqrt{N} = \pm 0.0707$.

Table 2.3 Residual cross-correlation matrices

<table>
<thead>
<tr>
<th>Lag 1-5</th>
<th>-0.02</th>
<th>-0.02</th>
<th>-0.00</th>
<th>-0.02</th>
<th>-0.03</th>
<th>-0.06</th>
<th>-0.05</th>
<th>-0.01</th>
<th>-0.01</th>
<th>0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.00</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.07</td>
<td>-0.02</td>
<td>-0.00</td>
<td>-0.05</td>
</tr>
<tr>
<td>Lag 6-10</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.04</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.00</td>
<td>-0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td>Lag 11-15</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.00</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
<td>-0.00</td>
<td>0.01</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>Lag 16-20</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.05</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.07</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.08</td>
<td>-0.07</td>
<td>0.05</td>
<td>0.08</td>
<td>-0.06</td>
<td>-0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
</tbody>
</table>

(b) Residual cross-correlation matrices in term of indicator symbols

<table>
<thead>
<tr>
<th>Lag 1-5</th>
<th>. . . . . . . . . . . . . . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag 6-10</td>
<td>. . . . . . . . . . . . . . . .</td>
</tr>
<tr>
<td>Lag 11-15</td>
<td>. . . . . . . . . . . . . . . .</td>
</tr>
<tr>
<td>Lag 16-20</td>
<td>. . . . . . . . . . . . . . . +</td>
</tr>
</tbody>
</table>

(c) Pattern of correlations for each element in the matrix

<table>
<thead>
<tr>
<th></th>
<th>Sno1</th>
<th>Sno3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sno1</td>
<td>...........</td>
<td>...........+..</td>
</tr>
<tr>
<td>Sno3</td>
<td>...........-..</td>
<td>...........+..</td>
</tr>
</tbody>
</table>

The residual correlation matrices for lags 1 through 20 are displayed in Table 2.3. Table 2.3 also shows the indicator symbols, first used by Tiao and Box (1981), for residual correlations. “+”
indicates that a value is greater than $+2/\sqrt{N}$, “−” for a value less than $-2/\sqrt{N}$, and “.” for a value between the limits. It can be seen that only one residual autocorrelation at lag 17 is below the lower limit and two autocorrelations at lag 18 are beyond the upper limit. Examination of the residuals from the fitted VARMAX(3, 3, 1) model gives no indication of inadequacy of the model, and therefore this model can be accepted as an adequate representation for the simulated activated sludge process under normal condition.

2.4 VARMAX and PLS Modeling Comparison

2.4.1 Partial Least Squares

The partial least squares method was first presented by Wold in the 1960’s for econometric applications. Since then, PLS has become a very useful method to identify, monitor and control dynamic processes which have a controllable variable data matrix $X_{mn}$ and a quality variable data matrix $Y_{mk}$. To fully describe the PLS regression, two linear models are considered:

\[
\begin{align*}
Y &= TQ + E \\
X &= TP + F
\end{align*}
\]  

(2.15)

where $T$ is the $m\times A$ factor score matrix, $Q$ is a matrix of regression coefficients (loadings) for $T$, $P$ is $n\times A$ factor loading matrix, and $E$ and $F$ are noise terms. Here, $T = XW$, where $W$ is called a weight matrix, which is computed so that each column of $W$ maximizes the covariance between responses and the corresponding factor scores. The loading matrix $Q$ is obtained by the general least squares procedure for the regression of $Y$ on $T$. After $Q$ is computed, the prediction model
for $Y$ is available, $Y = XWQ + E$. Höskuldsson (1988) presented a classical way to perform the PLS procedure. In this algorithm, the first score vector $t_1 = w_1^T x$, where $w_1$ is the eigenvector with the largest eigenvalue of the matrix $X^T Y Y^T X$. Then $X$ is replaced by $X - t_1 p_1^T$, where $p_1 = X t_1 / t_1^T t_1$, and the same procedure is repeated until all the scores or latent vectors are calculated.

The adequate number of latent variables in the PLS model is chosen by means of cross-validation. The optimal order has the minimum of mean square error of prediction (RMSPcv). It can be seen from Fig. 2.3 that the model with four factors yields better RMSPcv than the model with three factors, but the difference is very slight. Therefore, the model with three factors is acceptable.

**Figure 2.3** The order selection of the PLS model

![Figure 2.3 The order selection of the PLS model](image)

### 2.4.2 Model Fitting Comparisons of VARMAX and PLS
In this section, two performance criteria, mean relative squared error (MRSE) and mean variance-accounted-for (MVAF), are used to measure the modeling fitting performance of the VARMAX model and the PLS. The two performance criteria can be calculated as follows:

\[
MRSE = \frac{1}{t} \sum_{i=1}^{t} \sqrt{\frac{1}{N} \sum_{j=1}^{N} \hat{e}_{ij}^2} \times 100\% \tag{2.16}
\]

\[
MVAF = \frac{1}{t} \sum_{i=1}^{t} \left( 1 - \frac{\text{var}(\hat{e}_i)}{\text{var}(y_i)} \right) \times 100\% \tag{2.17}
\]

where \( \hat{e}_i \) is the residual between the \( i \)th observation and its prediction.

<table>
<thead>
<tr>
<th>Model</th>
<th>MRSE (%)</th>
<th>MVAF (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARMAX</td>
<td>9.1383</td>
<td>96.8411</td>
</tr>
<tr>
<td>PLS</td>
<td>80.0530</td>
<td>42.1390</td>
</tr>
</tbody>
</table>

From Table 2.4, we can see that MRSEs of the fitted VARMAX model and PLS model are 9.1383 and 80.0530, respectively, and the MVAF of the VARMAX model is more than twice as the one of the PLS model. From these two criteria, the VARMAX model fits the in-control process much better than the PLS model. Furthermore, the MRSE value from the best choice of the subspace-based algorithms used by Sotomayor et al. (2003) is 31.8091, and the MVAF value is 89.9037. By comparing the values in Table 2.4 with their results, the VARMAX model seems to produce the best model in terms of identification, and the subspace-based models in the paper by Sotomayor et al. (2003) are better than the PLS model.
2.4.3 Performance Comparisons of VARMAX and PLS

In the previous sections, the VARMAX model and the PLS model have been developed to describe the normal operating condition for the activated sludge process. In this section, SPEs are first calculated based on the VARMAX model and the PLS model. The control charts of the SPEs are then set up with the probability control limits at significance level $\alpha = 0.01$.

2.4.3.1 In-control Situation

When the process is in control, the residuals behave approximately as a white noise process and the calculated SPE values should be small. However, if an assignable cause occurs, the mean of residuals will shift from the in-control residual mean vector and the SPEs will increase considerably.
In Fig. 2.4, it can be seen that by using the VARMAX model, there are only five false alarms in 800 in-control samples. However, the number of false alarms in Fig. 2.5 increases to 12 by using the PLS model, which is more than twice when compared with the VARMAX model.

2.4.3.2 Out-of-control Situation

An out-of-control case was simulated by introducing a step change after the 112th sample was taken. Since a control scheme was used in the simulator, the disturbance could first be handled by the controller before quality variables went out of control. From Fig. 2.6, it is can be seen that the SPE chart using the VARMAX model quickly signaled the out-of-control condition. The SPE chart using the PLS model shown in Fig. 2.7 also indicated the disturbance, but there were too many false alarms during the time the process was in control.
Figure 2.6 SPE chart using VARMAX model when the process is out of control

Figure 2.7 SPE chart using PLS model when the process is out of control
In summary, the VARMAX model performed considerably better in both in-control and out-of-control situations. When the process is in control, the number of false alarms on the SPE chart based on the PLS model is much higher than on the chart based on the VARMAX model. When the process is out of control, the control chart based on the VARMAX model can give much fewer false alarms than the chart based on the PLS model although both of the SPE charts can quickly detect an out-of-control situation.
Chapter 3 Economic and Economic-Statistical Design of Multivariate Bayesian Control Chart for MSPC

3.1. Introduction

Multivariate control charts are valuable tools for MSPC to monitor the process mean shifts. The observations collected from the process are assumed to be independent and follow a multivariate normal distribution. However, in an autocorrelated process, such assumptions will be violated and thus the methodology discussed in previous chapter need be applied first before any multivariate control charts can be used to monitor the process. More specifically, a VARMAX model is fitted to in-control historical process data offline and residuals are calculated based on the collected observations using the VARMAX. Since the residuals should follow approximately a normal distribution and be independent, a multivariate control chart can be applied to the residuals. After the control charts assumptions have been satisfied, the next key issue is to specify control chart parameters including the sample size, the sampling interval and the control limit. One way to design control chart parameters is solely based on the statistical performance requirements of a control chart for both in control and out of control situations (see Woodall 1985 for a Shewhart chart and a CUSUM chart, Jones 2002 for EWMA control chart). The statistical performance of a control chart can be measured by the average run length (ARL). The ARLs of the Shewhart-type charts can be easily calculated from the distribution of the control
statistic. For instance, the ARL of a $\bar{X}$ chart equals $1/\alpha$ for the in-control situation. Here $\alpha$ is the probability of the type I error. Besides the simulation techniques and the integral equation method (Champ and Rigdon 1991, Rigdon 1995a, 1995b), a Markov chain approximation has been widely used for the calculations of ARLs of univariate EWMA and CUSUM control charts (Brook and Evans 1972, Lucas and Saccucci 1990) and multivariate control charts (Runger and Prabhu 1996, Yin and Makis 2006).

Unlike the statistical design approach, the economic design approach selects control chart parameters based on an economic criterion. The basic idea is to formulate an optimization model with control chart parameters as decision variables and to employ suitable optimization techniques to find the optimal solution for the model. Note that the resulting problem is often a nonlinear optimization problem. The economic design model of a $\bar{X}$ control chart was first developed by Duncan (1956). He assumed that the occurrence of a shift caused by the assignable cause follows a Poisson process. It was followed by the contributions for $\bar{X}$ charts by Goel et al. (1968), Duncan (1971), Gibra (1971), and Knappenberger and Grandage (1969), for CUSUM charts by Taylor (1968), and Goel and Wu (1973), and for $p$ charts by Ladany (1973). Lorenzen and Vance (1986) generalized the Duncan’s model and presented a unified approach for determining optimal parameter for univariate control charts. Some extensive review papers on the economic design were written by Montgomery (1980), Svoboda (1987), Ho and Case (1994), and Keats et al. (1997). Motivated by Woodall’s (1985, 1986) criticisms on the possibility of poor statistical performance of economically-designed control charts, Saniga (1989) proposed economic-statistical design for $\bar{X}$ and R charts, which is a constrained optimization problem subject to the requirements on ARLs in both in-control and out-of-control situations. The
economic statistical design of a univariate EWMA chart by Montgomery et al. (1995) was extended to the multivariate case by Linderman and Love (2000).

Another type of a control chart is a Bayesian control chart that plots the posterior probability that the process is out of control given the process history. In 1952, Girshick and Rubin first presented a Bayesian approach to quality control, followed by Bather (1963), Taylor (1965, 1967), Eckles (1968), Ross (1971), Carter (1972), and White (1977). It has been shown in the literature, e.g. Taylor (1965), that the Bayesian approach is an optimal way to determine the chart parameters compared with the non-Bayesian charts.

Several univariate Bayesian $\bar{X}$ charts for finite production runs have been presented by Tagaras (1994, 1996) and Tagaras and Nikolaidis (2002). Tagaras (1994, 1996) provided the algorithms based on dynamic programming to calculate the Bayesian control chart parameters. Some comparison results for univariate Bayesian $\bar{X}$ charts can be found in Tagaras and Nikolaidis (2002). Based on the theory of Partially Observable Markov Decision Processes (POMDP), Calabrese (1995) presented an optimal policy structure for a univariate Bayesian attribute control chart with a fixed sample size and sampling interval for finite production runs. Makis (2006a and 2007) presented a theoretical development of a multivariate Bayesian control chart with fixed sample size and sampling interval for finite and infinite production runs. He proved that a simple control limit policy structure is optimal and provided computational algorithms to calculate the optimal control limit.
In this chapter, we will design an optimization model for the economical design of the multivariate Bayesian control chart considering the control limit policy proved to be optimal by Makis (2007). The objective is to find the optimal control chart parameters, namely the sample size, sampling interval and control limit, to minimize the total cost of the process. The model considered in this research has a cost structure similar to the one considered in Lorenzen and Vance’s model. However, not only the definition of a renewal cycle but also the methods to derive the formulas for the expected cycle cost and cycle time are different. The objective function for the economically-designed Bayesian chart is the long run expected average cost that should be minimized. Statistical constraints are added for the economic statistical design of the chart. The computational results to compare the performance of the multivariate Bayesian control chart and the EWMA chart are included as well.

3.2 Multivariate Bayesian Control Chart

In this section, the multivariate Bayesian control chart designed by Makis (2007) is reviewed. The Bayesian chart monitors a process defined by $q$ quality characteristics. Let $y_m$ denote the sample collected at time $mh$ for $m=1, 2, \ldots$ and $h$ is the sampling interval. Assume that the time between the occurrence of two consequent assignable causes is exponentially distributed with the mean $1/\lambda$. Without loss of generality, the mean vector $\mu_0$ is assumed to be a zero vector.

Assume that $y_m$ follows a multivariate normal distribution with the mean vector $\mu_0$ and the covariance matrix $\Sigma$ when the process is in control. An assignable cause shifts the process mean from $\mu_0$ to $\mu_1$ with the unchanged covariance matrix. The shift size can be measured by the M-distance between $\mu_1$ and $\mu_0$ where the M-distance is defined as $d_1 = [(\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0)]^{1/2}$. 
Let $P_t$ denote the probability that the process is out of control at time $t$ given the observations up to time $t$. By using Bayes’ theorem, Makis (2007) proved that $P_t$ can be updated by

$$P_t = 1 - (1 - P_{mh})e^{-\lambda(r-mh)}, \text{ for } mh < t < (m+1)h$$

$$P_{mh} = \frac{1 - e^{-\lambda h}(1 - P_{(m-1)h})}{1 - e^{-\lambda h}(1 - e^{\lambda h} + Z_m^2/2)}, \text{ for } m = 1, 2, \ldots,'$$

(3.1)

where $Z_m = 2 \sum_{j=1}^{n} (y_j^m - \mu_0)^T \Sigma^{-1} (\mu_0 - \mu_1)$, is a normal random variable. Specifically, $Z_m$ follows $N(0, 4nd_i^2)$ when the process is in control and $Z_m$ follows $N(-2nd_i^2, 4nd_i^2)$ when it is out of control.

After the $m^{th}$ sample is collected, the $P_{(m+1)h}$ is calculated by (3.1). If the posterior probability $P_{(m+1)h}$ is greater than the upper control limit $P^*$, the process will be stopped and a search for an assignable cause will be triggered. If the process is not stopped while the process is out of control, a large quality-related cost per unit time will occur. It is shown by Makis (2007) that a simple control limit structure is an optimal policy for the multivariate Bayesian control chart.

### 3.3 Model Formulation

#### 3.3.1 Model Description

Suppose the Bayesian control chart described in Section 3.2 is used to monitor the same process as described in the previous section. Assume the process begins in an in-control state and it stays in the in-control state until an assignable cause occurs. It is assumed that the process is stopped during the search of an assignable cause and it will be repaired and brought back to the in-control state if it was in an out-of-control state. A sample of $n$ units is collected from the process at
equally spaced sampling epochs with a constant time interval $h$. The time to sample and chart one item on the Bayesian chart is $E$. As soon as a point falls beyond the control limit, the process is stopped to start searching for an assignable cause. If an assignable cause is found, it means that the process is out of control and the repair actions are needed to bring the process back to its initial state. Otherwise, the process is resumed without any intervention. It is also assumed that when the process shifts to the out of control state, it cannot return to the in-control state automatically without intervention.

### 3.3.2 Model Parameters

In order to compare the performance of the Bayesian chart with a EWMA chart later in the thesis, the process parameters, same as the ones in Lorenzen and Vance’s model, are considered. The following notation is used in the cost model:

**Input variables:**

- $T_0 =$ expected search time when false alarm occurs.
- $T_i =$ expected search time to discover an assignable cause ($> T_0$).
- $T_r =$ expected time to repair the process.
- $C_0 =$ quality cost per unit time while the process is in-control.
- $C_1 =$ quality cost per unit time while the process is out-of-control ($> C_0$).
- $Y =$ cost per false alarm.
- $W =$ cost to locate the assignable cause and repair the process.
- $a =$ fixed cost per sample
- $b =$ cost per unit sampled.
Output (calculated) variables

\( n = \) the sample size.

\( h = \) the sampling interval

\( P^* = \) the control limit

Other notation

\( CL = \) the cycle length

\( E(CL) = \) the expected cycle length

\( \text{Cost} = \) the cycle cost

\( E(\text{Cost}) = \) the expected cycle cost

It is assumed that \( nE \) is less than a sampling interval \( h \) and \( T_0 \leq T_i \) in this study to simplify the problem without losing too much generality.

3.3.3 Modification of Bayesian Control Chart

Since we consider a delay \( nE \) to collect and process a sample, the posterior probability \( P_{mh} \) cannot be calculated until time \( mh+nE \) although the sample was taken at time \( mh \). However, the posterior probability \( P_t \) has evolved since time \( mh \). Thus, \( P_{mh+nE} \) instead of \( P_{mh} \) should be plotted on the Bayesian chart. The following lemma is provided for updating \( P_{mh+nE} \) when the delay is considered.

**Lemma 1.** Given \( P_0, P_{mh+nE} \) can be calculated as follows:

\[
P_{mh+nE} = 1 - e^{-\lambda nE} + e^{-\lambda nE} \frac{1 - e^{-\lambda h + \lambda nE} (1 - P_{(m-1)h+nE})}{1 - e^{-\lambda h + \lambda nE} (1 - P_{(m-1)h+nE}) (1 - \exp((nd_i^2 + Z_m) / 2))},
\]

\[(3.2)\]
for $m = 1, 2, \ldots$

where $d_i$ is the M-distance between the out-of-control mean vector $\mu_i$ and the in-control mean vector $\mu_0$, and $Z_m = 2\sum_{j=1}^{n} (y_j^m - \mu_0)^T \Sigma^{-1}(\mu_0 - \mu_i)$.

**Proof.**

For $t > 0$ and given $P_0$, Makis (2007) proved that $P_t$ can be obtained as follows:

\[
P_{mh} = \frac{1 - e^{-\lambda h} (1 - P_{(m-1)h})}{1 - e^{-\lambda h} (1 - P_{(m-1)h})(1 - \exp((nd_i^2 + Z_m)/2))}, \text{ for } m = 1, 2, \ldots, (3.3)
\]

\[
P_t = 1 - e^{-\lambda (t - mh)} (1 - P_{mh}), \text{ for } mh < t < (m + 1)h (3.4)
\]

From (3.4), we have,

\[
P_{mh+nE} = 1 - e^{-\lambda (mh+nE - mh)} (1 - P_{mh})
\]

\[
= 1 - e^{-\lambda nE} (1 - P_{mh}) \quad (3.5)
\]

and $P_{(m-1)h} = 1 - e^{\lambda nE} (1 - P_{(m-1)h+nE}) (3.6)$

Using (3.3), (3.5) and (3.6), we have,

\[
P_{mh+nE} = 1 - e^{-\lambda nE} \left[ 1 - \frac{1 - e^{-\lambda h} \left( e^{\lambda nE} (1 - P_{(m-1)h+nE}) \right)}{1 - e^{-\lambda h} \left( e^{\lambda nE} (1 - P_{(m-1)h+nE}) \right)(1 - \exp((nd_i^2 + Z_m)/2))} \right] \]

\[
= 1 - e^{-\lambda nE} + e^{-\lambda nE} \frac{1 - e^{-\lambda h + \lambda nE} (1 - P_{(m-1)h+nE})}{1 - e^{-\lambda h + \lambda nE} (1 - P_{(m-1)h+nE})(1 - \exp((nd_i^2 + Z_m)/2))}
\]

In this research, it is assumed that $P_0 = 0$. Under this assumption, note that

\[
P_{nE} = 1 - e^{-\lambda nE} (1 - P_0) = 1 - e^{-\lambda nE}.
\]
3.4 Economic Design and Economic-Statistical Design of Multivariate Bayesian Control Chart for MSPC

In this research, the multivariate Bayesian control chart parameters $n$, $h$, and $P^*$ are considered to minimize the expected cost per time unit. In Lorenzen-Vance cost model, a cycle is defined as the time until the process is repaired and brought back to the initial state. In their model, average run lengths (ARLs) need to be calculated to obtain the expected average cost. For a simple control chart, such as $\bar{X}$ charts, the ARLs do not depend on the initial value of the chart statistic. However, for more complex control charts, such as the EWMA chart and the Bayesian chart, the
ARLs are not independent of the initial value of the chart statistic. Therefore, we define a cycle differently to avoid using ARLs to express the expected average cycle cost. Here a cycle includes the time until an alarm occurs, plus the search time if it is a false alarm, or plus the search and repair time if the assignable cause occurs. Because of the assumption that the in-control time follows a memoryless exponential distribution, the process can be treated as being in the initial state after it is resumed and therefore the process is a renewal reward process. Based on the renewal theory, the expected cost per time unit is:

\[
E(\text{Cycle cost}) = \frac{E(\text{Cost})}{E(\text{CL})}
\]

From the definition of \( P^* \) and the assumption of \( nE < h \), two constraints need to be added. Therefore, the optimization model for the economic design of the Bayesian chart can be formulated as

\[
\min z(n, h, P^*) = \frac{E(\text{Cost})}{E(\text{CL})} \tag{3.8}
\]

\[
st. \quad 0 \leq P^* \leq 1
\]
\[
\quad nE < h
\]
\[
\quad n \text{ is a positive integer}
\]

Suitable statistical constraints should be added to the above model to formulate the optimization model for the economic-statistical design of the chart. ARLs can be used to measure the statistical performance of the Bayesian chart. Further discussion will be given in later section about how to use ARLs to express such statistical constraints.

3.4.1 Expected Cycle Length
In this section, the formula for the expected cycle length is derived. Conditioned on the in-control time $T$, we have

$$E(\text{CL}) = \int_0^{+\infty} E(\text{CL}|T=t) \lambda e^{-\lambda t} dt$$

$$= \int_0^{+\infty} \left\{ \sum_{i=1}^{t/h} E(\text{CL}|T=t, \text{alarm occurs at } ith \text{ sample}) \cdot P(\text{an alarm occurs at } ith \text{ sample}) + \sum_{i=[t/h]+1}^{+\infty} E(\text{CL}|T=t, \text{alarm occurs at } ith \text{ sample}) \cdot P(\text{an alarm occurs at } ith \text{ sample}) \right\} \lambda e^{-\lambda t} dt$$

In order to deal with the integer part of $t/h$, the interval $[0, +\infty)$ is divided into $[0, h), [h, 2h), [2h, 3h), \ldots$, and so on. Then, the expected cycle length can be calculated as:

$$E(\text{CL}) = \sum_{m=0}^{+\infty} \left\{ \sum_{i=m+1}^{(m+1)h} E(\text{CL}|T=t) \lambda e^{-\lambda t} dt \right\}$$

$$= \sum_{m=0}^{+\infty} \left\{ \sum_{i=m+1}^{(m+1)h} \left\{ \int \left( \sum_{i=1}^{\infty} E(\text{CL}|T=t, 1st A = i) \cdot P(1st A = i | T = t) \right) dt \right\} \lambda e^{-\lambda t} dt \right\} \lambda e^{-\lambda t} dt$$

where $P(1st A = i | T = t)$ denotes the probability that the first alarm occurs at time $ih+nE$ given that the process shifts from in-control to out-of-control at time $t$.

Let $CL_{-1} = \int_{mh}^{(m+1)h} \sum_{i=1}^{\infty} \left\{ E(\text{CL}|T=t, 1st A = i) \cdot P(1st A = i | T = t) \right\} \lambda e^{-\lambda t} dt$ and $CL_{-2} = \int_{mh}^{(m+1)h} \sum_{i=m+1}^{\infty} \left\{ E(\text{CL}|T=t, 1st A = i) \cdot P(1st A = i | T = t) \right\} \lambda e^{-\lambda t} dt$.

From (3.9), we have

$$E(\text{CL}) = \sum_{m=0}^{+\infty} (CL_{-1} + CL_{-2}) \quad (3.10)$$

Due to the delay to collect and process a sample, it is possible for a chart signal to become a true alarm although the statistic is updated by a sample which is taken from the in-control process. More specifically, given that an assignable cause occurs between $mh$ and $mh+nE$, a search for an assignable cause takes place if the chart signals at the $m^{th}$ sampling epoch. In the situation
described by Fig. 3.2, the process is out of control when the process is stopped by an alarm. It is called a true alarm if an alarm occurs when the process is out of control. Note that the real state of the process can be known only by performing a full inspection of the process. Otherwise, it is a false alarm. Since the process in Fig. 3.2 is out of control, the chart signal will be validated as a true alarm by a full inspection.

Figure 3.2 Process is stopped when $mh < t < mh + nE$

When a true alarm occurs and an assignable cause is found, the cycle length is equal to the sum of the time until the alarm occurs, the time to discover the assignable cause ($T_i$), and the time to repair the process ($T_2$). Thus, when the alarm occurs at the $m^{th}$ sample and $mh \leq t \leq mh + nE$, the cycle length is equal to

$$mh + nE + T_i + T_2.$$  \hspace{1cm} (3.11)

Figure 3.3 Process is stopped when $mh + nE \leq t$
In the situation described by Fig. 3.3, the process is still in control when the process is stopped by an alarm. In this case, the cycle length includes the time until the alarm occurs and the time to identify the false alarm \( T_0 \). Thus, when the alarm occurs at the \( m \)th sample and \( mh + nE \leq t \leq (m+1)h \), the cycle length equals

\[
mh + nE + T_0.
\] (3.12)

Similarly, for \( i \geq m + 1 \) and \( mh \leq t \leq (m+1)h \), the cycle length equals

\[
ih + nE + T_1 + T_2,
\] (3.13)

and for \( i < m \) and \( mh \leq t \leq (m+1)h \), the cycle length equals

\[
ih + nE + T_0
\] (3.14)

Given that the shift occurs between \( mh \) and \( (m+1)h \), from (3.12) - (3.14) we have

\[
CL_1 = \int_{mh}^{(m+1)h} \left\{ \sum_{i=1}^{m} \{ E(CL|T = t, \text{alarm at ith sample}) \cdot P(1st A = i | T = t) \} \right\} \lambda e^{-\lambda t} dt
\]

\[
= \int_{mh}^{(m+1)h} \sum_{i=1}^{m-1} \{ E(CL|T = t, \text{alarm at ith sample}) \cdot P(1st A = i | T = t) \} \lambda e^{-\lambda t} dt
\]

\[
+ \int_{mh}^{(m+1)h} \{ E(CL|T = t, \text{alarm at mth sample}) \cdot P(1st A = i | T = t) \} \lambda e^{-\lambda t} dt
\]

\[
= \int_{mh}^{(m+1)h} \sum_{i=1}^{m-1} \left\{ (ih + nE + T_0) \cdot P(1st A = i | T = t) \right\} \lambda e^{-\lambda t} dt
\]

\[
+ \int_{mh}^{(m+1)h} \{ (mh + nE + T_1 + T_2) \cdot P(1st A = m | T = t) \} \lambda e^{-\lambda t} dt
\]

\[
+ \int_{mh}^{(m+1)h} \{ (mh + nE + T_0) \cdot P(1st A = m | T = t) \} \lambda e^{-\lambda t} dt
\]

and

\[
CL_2 = \int_{mh}^{(m+1)h} \left\{ \sum_{i=m+1}^{+\infty} \{ (ih + nE + T_1 + T_2) \cdot P(1st A = i | T = t) \} \right\} \lambda e^{-\lambda t} dt
\] (3.15)

After computing the definite integrals in (3.15) and (3.16), the simplified formulas for \( CL_1 \) and \( CL_2 \) are:
\[ CL_1 = e^{-\lambda mh} \left[ (1 - e^{-\lambda h}) \sum_{i=m}^{\infty} \left\{ e^{-\lambda mh} i P(\text{1st A} = i \mid T = t, t \in [mh, (m+1)h]) \right\} \right. \]
\[ \left. + (nE + T_0) \sum_{i=1}^{m} P(\text{1st A} = i \mid T = t, t \in [mh, (m+1)h]) \right) \]
\[ + (1 - e^{-\lambda nh}) P(\text{1st A} = m \mid T = t, t \in [mh, (m+1)h]) \left( T_1 + T_2 - T_0 \right) \]

and

\[ CL_2 = e^{-\lambda mh} (1 - e^{-\lambda h}) \sum_{i=m+1}^{\infty} \left\{ (ih + nE + T_1 + T_2) \cdot P(\text{1st A} = i \mid T = t, t \in [mh, (m+1)h]) \right\} \]

Put (3.17) and (3.18) into (3.10), we have,

\[ E(CL) = nE + T_0 + (1 - e^{-\lambda h})h \sum_{m=0}^{\infty} \left\{ e^{-\lambda mh} E(\text{samples to signal}|T=t, t \in [mh, (m+1)h]) \right\} \]
\[ + (1 - e^{-\lambda nh})(T_1 + T_2 - T_0) \sum_{m=0}^{\infty} \left\{ e^{-\lambda mh} P(\text{1st A} = m \mid T = t, t \in [mh, (m+1)h]) \right\} \]
\[ + (1 - e^{-\lambda h}) (T_1 + T_2 - T_0) \sum_{m=0}^{\infty} \left\{ e^{-\lambda mh} P(\text{1st A} \geq m+1 \mid T = t, t \in [mh, (m+1)h]) \right\} \]

where \( E(\text{samples to signal}|T=t, t \in [mh, (m+1)h]) \) is the expected number of samples taken before the first alarm occurs, \( P(\text{1st A} = m \mid T = t, t \in [mh, (m+1)h]) \) is the probability that the first alarm occurs at the \( m \text{th} \) sample, and \( P(\text{1st A} \geq m+1 \mid T = t, t \in [mh, (m+1)h]) \) is the probability that the first alarm occurs at or after the \((m+1)\text{th}\) sample, given that the process mean shifts between \( mh \) and \((m+1)h\).

### 3.4.2 Expected Cycle Cost

Similar to the previous derivation, the expected cycle cost can be calculated as follows,

\[ E(\text{Cost}) = \sum_{m=0}^{\infty} \int_{mh}^{(m+1)h} E(\text{Cost}|T=t) \lambda e^{-\lambda t} dt \]
\[ = \sum_{m=0}^{\infty} \int_{mh}^{(m+1)h} \left\{ \sum_{i=1}^{m} \{ E(\text{Cost}|T=t, \text{1st A} = i) \cdot P(\text{1st A} = i \mid T = t) \} \right\} \lambda e^{-\lambda t} dt \]
Let
\[ \text{Cost}_1 = \int_{m_{th}}^{(m+1)h} \left( \sum_{i=1}^{m} \left\{ E(Cost|T = t, 1st A = i) \cdot P(1st A = i | T = t) \right\} \right) \lambda e^{-\lambda t} dt \]
and
\[ \text{Cost}_2 = \int_{m_{th}}^{(m+1)h} \left( \sum_{i=m+1}^{\infty} \left\{ E(Cost|T = t, 1st A = i) \cdot P(1st A = i | T = t) \right\} \right) \lambda e^{-\lambda t} dt \]

Thus,
\[ E(Cost) = \sum_{m=0}^{\infty} (\text{Cost}_1 + \text{Cost}_2) \quad (3.21) \]

In the situation described by Fig. 3.2, when an alarm occurs after the process shifts, the cycle cost is equal to the sum of the sampling cost until the alarm occurs, the cost \( W \) to locate the assignable cause and repair the process, the nonconformity cost \( C_0 \) during the in-control period and the nonconformity cost \( C_1 \) during the out-of-control period. Thus, when the alarm occurs at the \( m \)th sample and \( mh \leq t \leq mh + nE \), the cycle cost is
\[ tC_0 + (mh + nE - t)C_1 + (a + bn)m + W. \quad (3.22) \]

In the situation shown in Fig. 3.3, the process is still in control when the process is stopped by an alarm at the \( m \)th sample. In this case, the cycle cost includes the sampling cost until the alarm occurs, the cost to investigate the false alarm \( Y \), and the nonconformity cost \( C_0 \) during the in-control period. Thus, when an alarm occurs at the \( m \)th sample and \( mh + nE \leq t \leq (m + 1)h \), the cycle cost equals
\[ (mh + nE)C_0 + (a + bn)m + Y. \quad (3.23) \]

Similarly, for \( i \geq m + 1 \) and \( mh \leq t \leq (m + 1)h \), the cycle cost equals
\[ tC_0 + (ih + nE - t)C_i + (a + bn)i + W, \]  

(3.24)

and for \( i < m \) and \( mh \leq t \leq (m+1)h \), the cycle cost equals

\[ (ih + nE)C_0 + (a + bn)i + Y \]  

(3.25)

Given that the shift occurs between \( mh \) and \( (m+1)h \), from (3.22) - (3.25) we have

**Cost \_1**

\[
\begin{align*}
\text{Cost}_1 &= \int_{mh}^{(m+1)h} \left\{ \sum_{i=1}^{m} \left\{ E(\text{Cost}|T = t, \text{alarm at ith sample}) \cdot P(1stA = i | T = t) \right\} \lambda e^{-\lambda t} dt \right. \\
&= \int_{mh}^{(m+1)h} \left\{ \sum_{i=1}^{m} \left\{ E(\text{Cost}|T = t, \text{alarm at ith sample}) \cdot P(1stA = i | T = t) \right\} \lambda e^{-\lambda t} dt \right. \\
&+ \int_{mh}^{(m+1)h} \left\{ E(\text{Cost}|T = t, \text{alarm at mth sample}) \cdot P(1stA = m | T = t) \right\} \lambda e^{-\lambda t} dt \\
&= \int_{mh}^{(m+1)h} \left\{ \sum_{i=1}^{m} \left\{ ((ih + nE)C_0 + (a + bn)i + Y) \cdot P(1stA = i | T = t) \right\} \lambda e^{-\lambda t} dt \right. \\
&+ \int_{mh}^{(m+1)h} \left\{ ((mh + nE)C_0 + (a + bn)m + Y) \cdot P(1stA = m | T = t) \right\} \lambda e^{-\lambda t} dt \\
&= e^{-\lambda mh} \left\{ (hC_0 + a + bn) \sum_{i=1}^{m} i P(1stA = i | T = t, t \in [mh,(m+1)h]) \right. \\
&\left. + (nEC_0 + Y) \sum_{i=1}^{m} i P(1stA = i | T = t, t \in [mh,(m+1)h]) \right\} \\
&+ e^{-\lambda mh} P(1stA = m | T = t, t \in [mh,(m+1)h]) \left\{ (C_0 \frac{1}{h} - C_1 \frac{1}{h} + W - Y)(1 - e^{-hnE}) \right. \\
&\left. + nE(C_1 - C_0) \right\} \\
\end{align*}
\]

(3.26)

and

**Cost \_2**

\[
\begin{align*}
\text{Cost}_2 &= \int_{mh}^{(m+1)h} \left\{ \sum_{i=m+1}^{\infty} \left\{ tC_0 + (ih + nE - t)C_1 + (a + bn)i + W \right\} \cdot P(1stA = m | T = t, t \in [mh,(m+1)h]) \right\} \lambda e^{-\lambda t} dt \\
&= e^{-\lambda mh} \left\{ (C_0 - C_1)[(mh + \frac{1}{h}) - ((m+1)h + \frac{1}{h})e^{-\lambda h}] \right. \\
&\left. + (1 - e^{-\lambda h})(nEC_1 + W) \right\} \sum_{i=m+1}^{\infty} P(1stA = m | T = t, t \in [mh,(m+1)h]) \\
&+ e^{-\lambda mh} (1 - e^{-\lambda h})(hC_1 + a + bn) \sum_{i=m+1}^{\infty} i P(1stA = m | T = t, t \in [mh,(m+1)h]) \\
\end{align*}
\]

(3.27)
From (3.21), (3.26) and (3.27), we have

\[
E(\text{Cost}) = nEC_0 + Y + (1 - e^{-\lambda h})(hC_0 + a + bn) \sum_{m=0}^{\infty} e^{-\lambda mh} E(\text{samples to signal}|T=t, t \in [mh,(m+1)h])
\]

\[
+ \{(1 - e^{-\lambda nE})(C_0 \frac{1}{\tau} - C_1 \frac{1}{\tau} + W - Y) + nE(C_1 - C_0)\}
\]

\[
\times \sum_{m=0}^{\infty} e^{-\lambda mh} P(1st A = m | T = t, t \in [mh,(m+1)h])
\]

\[
+ (1 - e^{-\lambda h})h(C_0 - C_1) \sum_{m=0}^{\infty} \left\{ m e^{-\lambda mh} P(1st A \geq m+1 | T=t, t \in [mh,(m+1)h]) \right\}
\]

\[
\frac{C_0 - C_1}{h} \left( (1 - e^{-\lambda h})(nEC_1 - nEC_0 - Y + W) \right)
\]

\[
\sum_{m=0}^{\infty} e^{-\lambda mh} P(1st A \geq m+1 | T=t, t \in [mh,(m+1)h])
\]

\[
+ (1 - e^{-\lambda h})h(C_1 - C_0) \sum_{m=0}^{\infty} \left\{ e^{-\lambda mh} \sum_{i=m+1}^{\infty} i P(1st A = i | T=t, t \in [mh,(m+1)h]) \right\}
\]

\[
. (3.28)
\]

3.5 A Markov Chain Model for the Multivariate Bayesian Control Chart

In order to further simplify (3.19) and (3.28), we need to know how to calculate the following terms

a) \( E(\text{Number of samples to signal}|T=t, t \in [mh,(m+1)h]) \)

b) \( P(1st A = m | T = t, t \in [mh,(m+1)h]) \)

c) \( P(1st A \geq m+1 | T = t, t \in [mh,(m+1)h]) \)

In this section, a Markov chain approximation method is used to derive formulas to calculate the above expected value and probabilities. In order to do so, the interval \([0, 1]\) is partitioned into \(l+1\) intervals to approximate the continuous random variable \( P_{mh+nE} \). From Markov chain theory, \( \{P_{mh+nE}\} \) can be approximated by a discrete state space Markov chain model. It is defined that the Markov chain \( \{X_m\} \) is in transient state \( i \) when \( \frac{(i-1)p^2}{T} \leq P_{mh+nE} < \frac{i^2p^2}{T} \) for \( 1 \leq i \leq l \), and \( X_m \) is in
absorbing state $l+1$ when $P^* \leq P_{mh+nE} < 1$. Since state $l+1$ is the only absorbing state, $P_{i(l+1)} = 0$ for $1 \leq j \leq l$ and $P_{i(l+1)} = 1 - \sum_{j=1}^{l} P_{ij}$.

### 3.5.1 The Transition Matrix for In-control Situation

From Lemma 1, for $1 \leq i, j \leq l$

$$P_{ij}(in) = P(X_m = j|X_1, \ldots, X_{(m-1)} = i, \text{process is in control})$$

$$= P\left( \frac{(j-1)}{l}P^* \leq 1 - e^{-\lambda nE} + e^{-\lambda nE} \frac{1 - e^{-\lambda h + \lambda nE}(1 - P_{(m-1)h+nE})}{1 - e^{-\lambda h + \lambda nE}(1 - P_{(m-1)h+nE})(1 - \exp((nd_i^2 + Z_m)/2))} < \frac{jP^*}{l} \mid X_{(m-1)} = i \right)$$

The middle value $\frac{(2i-1)P^*}{2l}$ of the interval $[\frac{(i-1)P^*}{l}, \frac{iP^*}{l}]$ is used to approximate $P_{mh+nE}$ and let

$$a_i = e^{-\lambda h + \lambda nE}(1 - P_{(m-1)h+nE}) \approx e^{-\lambda h + \lambda nE}(1 - \frac{(2i-1)P^*}{2l}).$$

Then, for $1 \leq i \leq l, 1 < j \leq l$,

$$P_{ij}(in) = P\left( \frac{1 - a_i}{2\ln(1 - \frac{e^{\lambda nE} jP^* / l + 1 - e^{\lambda nE}}{a_i}) - nd_i^2} < Z_m < \frac{1 - a_i}{2\ln(1 - \frac{e^{\lambda nE} (j-1)P^* / l + 1 - e^{\lambda nE}}{a_i}) - nd_i^2} \right)$$

Since $Z_m$ follows $N(0, 4nd_i^2)$ when the process is in control, we have,
\[ P_{ij}(in) = \Phi \left( \frac{1-a_i}{2d_i \sqrt{n}} \cdot \frac{1-\lambda_{nE}^i (j-1)P^*/(l+1-\lambda_{nE}^i)}{a_i} - nd_i^2 \right) \]

\[ -\Phi \left( \frac{1-a_i}{2d_i \sqrt{n}} \cdot \frac{1-\lambda_{nE}^i (j-1)P^*/(l+1-\lambda_{nE}^i)}{a_i} - nd_i^2 \right) \]

\text{(3.29)}

For \( j = 1, \)

\[ P_{1i}(in) = P(X_m = 1|X_1, \cdots, X_{(m-1)} = i) \]

\[ = P \left( 0 \leq 1 - \lambda_{nE}^1 + \lambda_{nE}^1 \cdot \frac{1-\lambda_{nE}^{m-1} (1-P_{(m-1)h+nE}^*) (1-\exp((nd_i^2 + Z_m)/2))}{1-\lambda_{nE}^{m-1} (1-P_{(m-1)h+nE}^*) (1-\exp((nd_i^2 + Z_m)/2))} < \frac{P^*}{l} \right) \]

\[ = 1 - \Phi \left( \frac{1-a_i}{2d_i \sqrt{n}} \cdot \frac{1-\lambda_{nE}^i (j-1)P^*/(l+1-\lambda_{nE}^i)}{a_i} - nd_i^2 \right) \]

\text{(3.30)}

3.5.2 The Transition Matrix for Out-of-control Situation

As mentioned in previous section, the optimal control limit \( P^* \) can be determined by the economic design or economic statistical design of the Bayesian chart given the out-of-control mean \( \mu_1 \). Assume the actual out-of-control mean is \( \mu_{act} \), which may be different from \( \mu_1 \).
Lemma 2. \( Z_m \) follows \( N(-2nd_{i2}^2, 4nd_i^2) \) where \( d_{i2}^2 = (\mu_{act} - \mu_o)^T \Sigma^{-1} (\mu_i - \mu_o) \) and
\[
d_i^2 = (\mu_i - \mu_o)^T \Sigma^{-1} (\mu_i - \mu_o)
\]
if the actual process out-of-control mean is \( \mu_{act} \).

Proof.

As a linear combination of normally distributed \( y_j^m \), \( Z_m \) is normally distributed based on the properties of a multivariate normal distribution. Furthermore,
\[
E(Z_m) = E \left[ 2 \sum_{j=1}^n (y_j^m - \mu_o)^T \Sigma^{-1} (\mu_o - \mu_i) \right]
\]
\[
= 2nE \left[ (y_j^m - \mu_o)^T \right] \Sigma^{-1} (\mu_o - \mu_i)
\]
\[
= 2n(\mu_{act} - \mu_o)^T \Sigma^{-1} (\mu_o - \mu_i)
\]
\[
= -2nd_{i2}^2
\]

\[
Var[Z_m] = Var \left[ 2 \sum_{j=1}^n (y_j^m - \mu_o)^T \Sigma^{-1} (\mu_o - \mu_i) \right]
\]
\[
= 4 \sum_{j=1}^n Var \left[ (y_j^m - \mu_o)^T \Sigma^{-1} (\mu_o - \mu_i) \right]
\]
\[
= 4nVar \left[ (y^m - \mu_o)^T \Sigma^{-1} (\mu_o - \mu_i) \right]
\]

It is known that \( EAx = A\mu \) and \( CovAx = A\Sigma A' \) for any \( A \in R^{mxn} \) and \( x \in R^n \) (for reference, to see Johnson and Wichern (1992)). Thus,
\[
Var[Z_m] = 4nVar \left[ (y^m - \mu_o)^T \Sigma^{-1} (\mu_o - \mu_i) \right]
\]
\[
= 4n(\mu_o - \mu_i)^T \Sigma^{-1} \Sigma \Sigma^{-1} (\mu_o - \mu_i)
\]
\[
= 4n(\mu_i - \mu_o)^T \Sigma^{-1} (\mu_i - \mu_o)
\]
\[
= 4nd_i^2
\]

Therefore, when the off-target mean is \( \mu_{act} \), \( Z_m \) follows \( N(-2nd_{i2}^2, 4nd_i^2) \). Note that \( d_{i2}^2 = d_i^2 \) if \( \mu_{act} = \mu_i \). Therefore, \( Z_m \) follows \( N(-2nd_i^2, 4nd_i^2) \) when the actual out-of-control mean is the same as assumed \( \mu_i \). □
Now, we are ready to apply Lemma 2 to calculate the transition matrix when the process is out of control.

For $1 \leq i \leq l, 1 < j \leq l$,

$$P_{ij}(\text{out}) = \Phi \left\{ \frac{1 - a_i}{2 \ln(1 - \frac{1 - e^{2\ln E} (j-1)P^*/l + 1 - e^{2\ln E}}{a_i}) - nd_i^2 + 2nd_{i2}} \right\},$$

$$- \Phi \left\{ \frac{1 - a_i}{2 \ln(1 - \frac{1 - e^{2\ln E} jP^*/l + 1 - e^{2\ln E}}{a_i}) - nd_i^2 + 2nd_{i2}} \right\},$$

and for $j=1$,

$$P_{ii}(\text{out}) = 1 - \Phi \left\{ \frac{1 - a_i}{2 \ln(1 - \frac{1 - e^{2\ln E} P^*/l + 1 - e^{2\ln E}}{a_i}) - nd_i^2 + 2nd_{i2}} \right\}. \quad (3.32)$$

### 3.5.3 Several Lemmas

The following lemmas are proven to calculate the unknown term a) to c). Define a $l$ dimensional column $\mathbf{1} = (1, \ldots, 1)^T$ and let an $1 \times l$ row vector $\mathbf{J}(i)$ have a one in the place corresponding to the starting state $i$ and zeros in the other places, for $i=1, \ldots, l$. Let $\mathbf{P}(in) = [P_{ij}(in)]_{i=1}^{l}$ be the matrix consisting of the first $l$ columns of the transition matrix.
Let $P(\text{1st Alarm} = j \mid T=t, t \in [mh,(m+1)h], X(0) = i)$ denote the probability that an alarm first occurs on the $j$th sample.

**Lemma 3.** For $1 \leq j \leq m$ and $m \geq 0$, given the initial state $i$,

$$P(\text{1st Alarm} = j \mid T=t, t \in [mh,(m+1)h], X(0) = i) = J(i) \cdot P^{j-1}(in) \cdot v,$$

where $v = (I - P(in))1$.

**Proof:**

Let $\lambda$ denote the absorbing state $l+1$. Let $\Delta P^{(j)}_{i,\lambda}(m,in)$ denote the probability that the Markov chain is in the state $\lambda$ at the $j$th sampling epoch without having visited the state $\lambda$ up to the $j$th sample given the initial state $X(0) = i$ and the transient part of the transition matrix $P(in)$.

Define $P^{(j)}(i,k) = \left( \begin{array}{ccc} P(i_1,\ldots,i_j) \\ i_{j+1} \end{array} \right)$, $v = (v_1,\ldots,v_j)^T$, and $\Delta P^{(j)}_{\lambda}(m,in) = (\Delta P^{(j)}_{\lambda}(m,in),\ldots,\Delta P^{(j)}_{\lambda}(m,in))^T$.

For $1 \leq j \leq m, i = 1,\ldots,l$

$$P(\text{1st Alarm} = j \mid T=t, t \in [mh,(m+1)h], X(0) = i) = P(X(j) = l+1, \text{not in state } l+1 \text{ before the } j\text{th sample} \mid T=t, t \in [mh,(m+1)h], X(0) = i)$$

$$= \Delta P^{(j)}_{\lambda}(m,in)$$

For $j=1$,

$$\Delta P^{(1)}_{\lambda}(m,in) = P^{(i+1)}(in) = v,$$

(3.33)

from which it follows that $\Delta P^{(1)}_{\lambda}(m,in) = P^{(0)}(in) \cdot v$, (3.34)

where $P^{(0)}(in) = I$.

For $2 \leq j \leq m$, 53
\[ \Delta P_{i\rightarrow a}^{(j)}(m, in) \]

\[ = P(X(j) = l + 1, \text{not in state } l + 1 \text{ before the } j^{th} \text{ sample} \mid T=t, t \in [mh, (m+1)h], X(0) = i) \]

\[ = \sum_{k \neq a} \left\{ P(X(j) = l + 1, \text{not in state } l + 1 \text{ before} \mid T=t, t \in [mh, (m+1)h], X(0) = i, X(1) = k) \right\} \]

\[ = \sum_{k \neq a} P_k(in) \cdot \Delta P_{i\rightarrow a}^{(j-1)}(m, in) \]

\[ \text{(3.35)} \]

In the matrix form, \[ \Delta P_{a}^{(j)}(m, in) = P(in) \cdot \Delta P_{a}^{(j-1)}(m, in). \]

\[ \text{(3.36)} \]

Next, by mathematical induction method we show \[ \Delta P_{a}^{(j)}(m, in) = P^{(j-1)}(in) \cdot v \] for \( 2 \leq j \leq m \).

From (3.35), we know that \[ \Delta P_{i\rightarrow a}^{(2)}(m, in) = \sum_{k \neq a} P_k(in) \cdot \Delta P_{i\rightarrow a}^{(2-1)}(m, in) = \sum_{k=1}^{l} P_{ik}(in) \cdot v_k = P_{i}(in) \cdot v, \]

where \( P_{i}(in) \) represents the \( i^{th} \) row in the matrix \( P(in) \).

Hence, \[ \Delta P_{a}^{(j)}(m, in) = P^{(j-1)}(in) \cdot v \] is true for \( j=2 \).

Assume \[ \Delta P_{a}^{(j)}(m, in) = P^{(j-1)}(in) \cdot v \] is true for \( j=n \), then \[ \Delta P_{a}^{(n)}(m, in) = P^{(n-1)}(in) \cdot v. \]

From (3.36), \[ \Delta P_{a}^{(n+1)}(m, in) = P(in) \cdot \Delta P_{a}^{(n+1-1)}(m, in) = P(in) \cdot P^{(n-1)}(in) \cdot v = P^{(n)}(in) \cdot v \]

Hence, \[ \Delta P_{a}^{(j)}(m, in) = P^{(j-1)}(in) \cdot v \] is true for \( 2 \leq j \leq m \). Combining this conclusion with (3.34), we have \[ \Delta P_{a}^{(j)}(m, in) = P^{(j-1)}(in) \cdot v \] for \( 1 \leq j \leq m \). If the initial state is state \( i \), the probability is \[ J(i) \cdot P^{j-1}(in) \cdot v. \]

Let \( P(1st \text{ Alarm} = j \mid T=t, t \in [mh, (m+1)h], X(0) = i) \) denote the probability that an alarm first occurs on the \( j^{th} \) sample.
Lemma 4. For \( j \geq m + 1 \) and \( m \geq 0 \), given the initial state \( i \)

\[
P(\text{1st Alarm} = j \mid T=t, t \in [mh,(m+1)h], X(0) = i) = J(i) \cdot P^{(m)}(in) \cdot P^{(j-m-1)}(out) \cdot w,
\]

where \( w = (I - P(out))1 \).

Proof:

Let \( \Delta P_{i,a}^{(j)}(m, out) \) denote the probability that the Markov chain will visit the state \( a \) the first time on the \( j^{th} \) sample given the initial state \( X(0) = i \) and the transient transition matrix \( P(out) \).

Define \( w = (w_1, \ldots, w_i)^T \), and \( \Delta P_{i,a}^{(j)}(m, out) = (\Delta P_{i,a}^{(j)}(m, out), \ldots, \Delta P_{i,a}^{(j)}(m, out))^T \).

Similar to the proof for Lemma 2, \( \Delta P_{i,a}^{(j)}(m, out) = P^{(j-1)}(out)w \) and

\[
\Delta P_{i,a}^{(j)}(m, out) = \sum_{k=1}^{j} P^{(j-k)}(out) \cdot w_k
\]

For \( j \geq m + 1 \),

\[
P(\text{1st Alarm} = j \mid T=t, t \in [mh,(m+1)h], X(0) = i) = P(X(j) = l + 1, \text{not in state } l + 1 \text{ before the } j^{th} \text{ sample} \mid T=t, t \in [mh,(m+1)h], X(0) = i)
\]

\[
= \sum_{g \neq l+1} \left\{ P(X(j) = l + 1, \text{not in state } l + 1 \text{ before} \mid T=t, t \in [mh,(m+1)h], X(0) = i, X(m) = g) \right\} \cdot P(X(m) = g \mid T=t, t \in [mh,(m+1)h], X(0) = i)
\]

\[
P(X(m) = g \mid T=t, t \in [mh,(m+1)h], X(0) = i) \text{ is the probability that the process starting in initial state } i \text{ will be in state } g \text{ after } m \text{ transitions, when the process is in control. Thus,}
\]

\[
P(X(m) = g \mid T=t, t \in [mh,(m+1)h], X(0) = i) = P_{lg}^{(m)}(in).
\]

Hence,

\[
P(\text{1st Alarm} = j \mid T=t, t \in [mh,(m+1)h], X(0) = i) = \sum_{g=1}^{j} P_{lg}^{(m)}(in) \cdot \Delta P_{g,a}^{(j-m)}(m, out)
\]
\[= \sum_{g=1}^{l} P_{ig}^{(m)}(\text{in}) \cdot \sum_{k=1}^{l} P_{jk}^{(j-m-1)}(\text{out}) \cdot w_k \]

\[= \sum_{g=1}^{l} P_{ig}^{(m)}(\text{in}) \cdot P_{g}^{(j-m-1)}(\text{out}) \cdot w \]

\[= P_{i}^{(m)}(\text{in}) \cdot P^{(j-m-1)}(\text{out}) \cdot w \]

3.5.4 Calculation of ARLs for the Multivariate Bayesian Control Chart

Let \( T_j \) denote the expected time that the Markov chain stays in transient state \( j \) starting from initial state \( i \) for \( i, j = 1, 2, \cdots, l \). One can prove that the matrix \( T \) of the expected times in transient states is equal to \( (I - P_r)^{-1} \), where \( P_r \) is the matrix which only contains the transition probabilities for the transient states. Let \( \text{ARL}_{in}(i) \) denote the expectation of the number of transitions needed from initial state \( i \) to absorbing state \( l+1 \) when the mean vector of the process is on target, and \( \text{ARL}_{out}(i) \) denote the average run length when the process is out of control. It can be shown that \( \text{ARL}_{in}(i) \) or \( \text{ARL}_{out}(i) \) equal to the sum of the total time the Markov chain spends in transient states starting in initial state \( i \). Thus, \( \text{ARL}_{in}(i) = \sum_{j=1}^{l} T_{ij}(in) \) and \( \text{ARL}_{out}(i) = \sum_{j=1}^{l} T_{ij}(out) \).

Assume that the initial state of the statistic is 1 when the process starts. According to the study in (Yin and Makis, 2006), \( \text{ARL}_{in}(l) \) is the smallest value of \( \text{ARL}_{in}(i) \) and it can be used as the worst-state ARL when the process is in control. The largest value of \( \text{ARL}_{out}(i) \), which is \( \text{ARL}_{out}(1) \), is chosen for the worst-state ARL when the process is out of control. It is because
that smaller value of ARL\(_{in}(i)\) means more false alarms and larger value of ARL\(_{out}(i)\) means more delay for the chart to detect the shift. Let ARL\(_{in}\) denote the minimum value of ARL\(_{in}(i)\) and ARL\(_{out}\) denote maximum value of ARL\(_{out}(i)\). In order to guarantee the requirements of the statistical performance of the control chart, the worst-state ARLs are used in the statistical constraints in this research.

### 3.6 Closed-form Formulas for Expected Cycle Length and Expected Cycle Cost

In this section, the closed-form formulas for expected cycle length and expected cycle cost will be derived. Since it is assumed that \(P_0 = 0\), the initial state of the Markov chain \(\{X_m\}\) is state 1. Let \(J\) denote \(J(1)\).

#### 3.6.1 Expected Cycle Length and Expected Cycle Cost

From Lemma 3 and Lemma 4, the unknown terms in (3.19) and (3.28) can be derived as follows,

\[
P(1st A \geq m + 1 | T = t, t \in [mh, (m + 1)h])
= \sum_{j=m+1}^{\infty} P(1st A = j | T = t, t \in [mh, (m + 1)h])
= \sum_{j=m+1}^{\infty} J \cdot P^{(m)}(in) \cdot P^{(j-m-1)}(out) \cdot w
\]

\[
= J \cdot P^{(m)}(in) \cdot \left( \sum_{j=m+1}^{\infty} P^{(j-m-1)}(out) \right) \cdot w
\]

\[
= J \cdot P^{(m)}(in) \cdot \left( \sum_{j=0}^{\infty} P^{j}(out) \right) \cdot w
\]

\[
= J \cdot P^{(m)}(in) \cdot (I - P(out))^{-1} \cdot (I - P^{\infty}(out)) \cdot w
\]

\[
= J \cdot P^{(m)}(in) \cdot (I - P(out))^{-1} \cdot (I - P(out))1
\]

\[
= J \cdot P^{(m)}(in) \cdot 1
\]

\[
P(1st A \leq m | T = t, t \in [mh, (m + 1)h]) = J \cdot (I - P^{\infty}(in)) \cdot 1,
\]

\[
(3.37)
\]

\[
(3.38)
\]
\[ P(1st\ A = m \mid T = t, t \in [mh, (m + 1)h]) = J \cdot P^{m-1}(in)v, \quad (3.39) \]

and

\[ \sum_{j=m+1}^{\infty} jP^{j-m-1}(out) = (I - P(out))^{-2} + m(I - P(out))^{-1}. \quad (3.40) \]

From (3.37)-(3.40), we have:

\begin{align*}
E(\text{No. of samples to signal} \mid T=t, t \in [mh, (m + 1)h]) &= \sum_{j=1}^{\infty} jP(1st\ A = j \mid T=t, t \in [mh, (m + 1)h]) \\
&= J \left( \sum_{j=1}^{m} jP^{(j-1)}(in) \right) \cdot v + J \cdot P^{(m)}(in) \left( \sum_{j=m+1}^{\infty} jP^{(j-m-1)}(out) \right) \cdot w \\
&= J \left[ (I - P(in))^{-1} - P^{m}(in)(I - P(in))^{-1} + P^{(m)}(in)(I - P(out))^{-1} \right] \cdot 1
\end{align*}

To simplify the final formulas for the expected cycle length and cost, the following notation is used.

\begin{align*}
D_1 &= \sum_{m=0}^{\infty} \left\{ e^{-\lambda mh} P(1st\ A = m \mid T = t, t \in [mh, (m + 1)h]) \right\} = J \cdot (I - e^{-2h} P(in))^{-1} \cdot e^{-2h} \cdot v \\
D_2 &= \sum_{m=0}^{\infty} \left\{ e^{-\lambda mh} P(1st\ A \geq m + 1 \mid T = t, t \in [mh, (m + 1)h]) \right\} = J \cdot (I - e^{-2h} P(in))^{-1} \cdot 1 \\
D_3 &= \sum_{m=0}^{\infty} \left\{ e^{-\lambda mh} E(\text{samples to signal} \mid T=t, t \in [mh, (m + 1)h]) \right\} \\
&= J \cdot \left\{ (1 - e^{-2h})^{-1} [I - P(in)]^{-1} + [I - e^{-2h} P(in)]^{-1} \left\{ [I - P(out)]^{-1} - [I - P(in)]^{-1} \right\} \right\} \cdot 1 \\
D_4 &= \sum_{m=0}^{\infty} \left\{ me^{-\lambda mh} P(1st\ A \geq m + 1 \mid T = t, t \in [mh, (m + 1)h]) \right\} = J \cdot (I - e^{-2h} P(in))^{-2} \cdot e^{-2h} P(in) \cdot 1 \\
D_5 &= \sum_{m=0}^{\infty} \left\{ e^{-\lambda mh} \sum_{j=m+1}^{\infty} jP(1st\ A = j \mid T = t, t \in [mh, (m + 1)h]) \right\} \\
&= J \cdot \left\{ (1 - e^{-2h} P(in))^{-1} \left\{ (I - P(out))^{-1} + (1 - e^{-2h} P(in))^{-1} \cdot e^{-2h} P(in) \right\} \right\} \cdot 1
\end{align*}

Finally, \( E(\text{CL}) \) and \( E(\text{Cost}) \) can be written as:

\begin{align*}
E(\text{CL}) &= nE + T_0 + (1 - e^{-2h})hD_3 + (1 - e^{-\lambda nE})(T_1 + T_2 - T_0)D_1 + (1 - e^{-2h})(T_1 + T_2 - T_0)D_2 \quad (3.41)
\end{align*}
\[ E(\text{Cost}) = nEC_0 + Y + (1 - e^{-\lambda h})(hC_0 + a + bn)D_1 + \left\{ (1 - e^{-\lambda nE})(C_0 + \frac{1}{h} - C_1 + \frac{1}{h} + W - Y) + nE(C_1 - C_0) \right\} D_1 \]

\[ + (1 - e^{-\lambda h})h(C_0 - C_1)D_2 + \left\{ \frac{1}{h} - (h + \frac{1}{h})e^{-\lambda h} \right\} D_2 + (1 - e^{-\lambda h})h(C_1 - C_0)D_3 \]

### 3.6.2. Optimization Models for Economic and Economic Statistical Design

In the above section, the expected cycle length and cycle cost are expressed in a closed-form. The optimization problem for the economic design of the multivariate Bayesian control chart can now be formulated as

\[
\min z(n, h, P^*) = \frac{\text{Equation (3.42)}}{\text{Equation (3.41)}}
\]

subject to:

- \( 0 \leq P^* \leq 1 \)
- \( nE < h \)
- \( n \) is a positive integer

This is a mixed nonlinear constrained optimization problem. The basic idea to find the optimal solution is to solve the optimization problem considering the continuous variables for every possible value of the discrete variable, the sample size \( n \). The sample size \( n \) can be any integer between 1 to 20. Since an unconstrained problem is easier to solve than a constrained problem, the problem is first transferred into an unconstrained problem. Let \( h = x_1^2 + nE \) and \( P^* = 0.5 \sin(x_2) + 0.5 \), and then the model (3.43) can be transformed to an unconstrained Optimization toolbox in MATLAB is used to find the optimal solution of the transformed problem. More specifically, fminsearch, a function in the MATLAB optimization toolbox, is used. Fminsearch follows the Nelder-Mead algorithm. The Nelder-Mead method is also known as simplex method. This method does not require any gradient calculation for the objective
function. However, it is time-consuming to converge to the optimal point, thus it is usually suitable only for low dimension problems. The Nelder-Mead algorithm showed computational stability when it was used to optimize the problem (3.43) for two sets of parameters with various sample sizes. The computational results will be shown in the following section.

The main criticism of the economic design of a control chart is that the chart may have poor statistical performance. To overcome this weakness of the economic design of a control chart, a model for economic-statistical design of the Bayesian chart is presented. Two constraints for ARL_{in} and ARL_{out} are added to the pure economic design of the Bayesian chart. Thus, the optimization model for the economic statistical design of the multivariate Bayesian control chart can now be formulated as

\[
\min z(n, h, P^*) = \frac{\text{Equation(3.42)}}{\text{Equation(3.41)}}
\]

subject to \( 0 \leq P^* \leq 1 \)

\( nE < h \)

\( \text{ARL}_{\text{in}} \geq \text{ARL}_{L} \)

\( \text{ARL}_{\text{out}} \leq \text{ARL}_{U} \)

\( n \) is a positive integer

The calculation of the ARL_{in} and ARL_{out} has been shown in section 3.5.4. Since two additional nonlinear constraints are added, the approach to transfer the constrained nonlinear model for the pure economic design to unconstrained nonlinear model cannot be applied to (3.44). The function of fmincon in Matlab is used to obtain the optimal solution for the constrained optimization model in (3.44).
3.7 Computational Results

In this section, the optimal cost of the Bayesian chart is compared with the optimal cost of the MEWMA chart based on their economic designs, using the following two sets of parameters (A and B) considered previously by Montgomery et al. (1995) and by Linderman and Love (2000a, 2000b):

**Set A**: \( \theta = 0.01, \ a = 0.5, \ b = 0.1, \ E = 0.05, \ T_1 = 2, \ T_2 = 2, \ C_0 = 10, \ C_1 = 100, \ Y = 50, \ W = 25. \)

**Set B**: \( \theta = 0.05, \ a = 5.0, \ b = 1, \ E = 0.5, \ T_1 = 2, \ T_2 = 2, \ C_0 = 10, \ C_1 = 100, \ Y = 50, \ W = 25. \)

3.7.1 Comparison of Multivariate Bayesian and MEWMA Charts

In this section, the optimal economic designs for the Bayesian chart were obtained for both Set A and Set B. Details about the method and MATLAB programs have been given in the previous sections. The economic designs of the MEWMA chart for both sets were calculated by modifying the inputs of the program published in Linderman and Love (2000b) under the assumptions made in this research. The notation used in Table 3.1 to 3.2 is as follows. \( E_n, E_h, E_L, r \) and \( E_{cost} \) are the optimal values of the sample size, the sampling interval, the control limit, the exponential weight and the optimal average cost of the MEWMA chart, respectively. \( B_n, B_h, B_L \) and \( B_{cost} \) are the optimal values of the sample size, the sampling interval, the control limit and the average cost of the Bayesian chart, respectively. \( SCost \) is the average of the simulated cycle cost from 50,000 simulation runs. \( \%_1 \) is the percentage cost difference between the Bayesian chart and MEWMA chart. \( \%_2 \) is the percentage cost difference between the model and simulation results for the Bayesian chart.

<table>
<thead>
<tr>
<th>( d_i^2 )</th>
<th>( E_n )</th>
<th>( E_h )</th>
<th>( E_L )</th>
<th>( r )</th>
<th>( E_{cost} )</th>
<th>( B_n )</th>
<th>( B_h )</th>
<th>( B_L )</th>
<th>( B_{cost} )</th>
<th>( %_1 )</th>
<th>( SCost )</th>
<th>( %_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>20</td>
<td>2.147</td>
<td>9.654</td>
<td>0.88</td>
<td>14.02</td>
<td>8</td>
<td>0.994</td>
<td>0.399</td>
<td>13.357</td>
<td>4.964</td>
<td>13.368</td>
<td>0.082</td>
</tr>
</tbody>
</table>
Table 3.2 Cost comparison of MEWMA and Bayesian charts for Set B

<table>
<thead>
<tr>
<th>$d_i^2$</th>
<th>$E_n$</th>
<th>$E_h$</th>
<th>$E_L$</th>
<th>$r$</th>
<th>Ecost</th>
<th>$B_n$</th>
<th>$B_h$</th>
<th>$B_L$</th>
<th>Bcost</th>
<th>%_1</th>
<th>Scost</th>
<th>%_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>5.350</td>
<td>0.059</td>
<td>0.623</td>
<td>27.353</td>
<td>2</td>
<td>2.622</td>
<td>0.156</td>
<td>25.099</td>
<td>8.980</td>
<td>25.175</td>
<td>0.303</td>
</tr>
<tr>
<td>1.0</td>
<td>1</td>
<td>5.086</td>
<td>0.234</td>
<td>0.512</td>
<td>27.468</td>
<td>4</td>
<td>2.447</td>
<td>0.136</td>
<td>25.031</td>
<td>9.736</td>
<td>24.982</td>
<td>-0.196</td>
</tr>
<tr>
<td>2.0</td>
<td>3</td>
<td>2.342</td>
<td>5.405</td>
<td>0.875</td>
<td>26.363</td>
<td>5</td>
<td>2.588</td>
<td>0.131</td>
<td>25.373</td>
<td>3.902</td>
<td>25.420</td>
<td>0.185</td>
</tr>
<tr>
<td>3.0</td>
<td>2</td>
<td>2.214</td>
<td>5.430</td>
<td>0.873</td>
<td>24.488</td>
<td>2</td>
<td>1.286</td>
<td>0.813</td>
<td>23.494</td>
<td>4.231</td>
<td>23.540</td>
<td>0.196</td>
</tr>
</tbody>
</table>

The columns labeled as %_2 in Table 3.1 and 3.2 give a comparison of the model and simulation. The differences range from -0.034 percent to 0.303 percent, which are very small. It means that the differences can be treated as simulation errors. Thus, the cost model is verified by the simulation results. In Table 3.1, the savings by using the Bayesian control chart over the MEWMA chart range from 1.961 percent to 4.964 percent for the four values of $d_i^2$. Much higher savings are obtained by the Bayesian control chart over the MEWMA for set B, ranging from 3.902 to 9.736 percent. It is instructive to notice that the Bayesian chart can bring more substantial savings when $d_i^2$ is smaller. It indicates that the Bayesian chart performs considerably better than the MEWMA chart for smaller shifts of the process mean.

Table 3.3 Economic statistical design of Bayesian chart for Set A

<table>
<thead>
<tr>
<th>$d_i^2$</th>
<th>B_n</th>
<th>B_h</th>
<th>B_L</th>
<th>Bcost Stat</th>
<th>ARL_{in}</th>
<th>ARL_{out}</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>8</td>
<td>0.945</td>
<td>0.399</td>
<td>13.359</td>
<td>202.109</td>
<td>2.808</td>
<td>0.017</td>
</tr>
<tr>
<td>1.0</td>
<td>8</td>
<td>1.271</td>
<td>0.399</td>
<td>12.502</td>
<td>237.788</td>
<td>1.675</td>
<td>0.010</td>
</tr>
<tr>
<td>2.0</td>
<td>6</td>
<td>1.416</td>
<td>0.300</td>
<td>11.887</td>
<td>242.167</td>
<td>1.264</td>
<td>0</td>
</tr>
<tr>
<td>3.0</td>
<td>5</td>
<td>1.415</td>
<td>0.264</td>
<td>11.627</td>
<td>342.939</td>
<td>1.149</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.3 shows the computational results for the economic statistical design of the Bayesian chart with the constraints $ARL_{in} \geq 200$ and $ARL_{out} \leq 10$ for Set A. The notation B_n, B_h and B_L in Table 3.3 is the same as in Table 3.1 and 3.2. Bcost_stat denotes the optimal values of the expected average cost of the economic-statistically designed Bayesian chart. % is the
percentage difference between the economic design and the economic statistical design for the Bayesian chart. It can be seen that the cost only increases 0.017% and 0.01% with the additional statistical constraints for \( d_i^2 = 0.5, 1.0 \), respectively. There is no cost penalty for \( d_i^2 = 2.0, 3.0 \) since the Bayesian chart designed considering only the economic requirements has met the statistical requirements. One may notice that the optimal solutions are slightly different from the ones in pure economic design for \( d_i^2 = 2.0, 3.0 \). It may be due to different optimization search techniques used in these two optimization models. It also shows the flatness of the region near the optimal solution.

3.7.2 Sensitivity Analysis

In previous sections, the detailed procedures of the economic and economic statistical design of the Bayesian control chart have been presented. In most of practical applications, it is necessary to be aware of how sensitive the expected average cost is to the changes of the actual process parameters. In this section, we will study the effect on the expected average cost of the misspecified out-of-control process mean vector \( \mathbf{\mu}_{\text{act}} \). Assume all chart parameters, such as \( \mathbf{P}^*, n \) and \( h \), have been determined by assuming the out-of-control mean is \( \mathbf{\mu}_0 \). From Lemma 2, one can see that the performance of the chart is affected only by the actual value of \( d_{i2}^2 \), where

\[
d_{i2}^2 = (\mathbf{\mu}_{\text{act}} - \mathbf{\mu}_0)^T \Sigma^{-1}(\mathbf{\mu}_{\text{act}} - \mathbf{\mu}_0).
\]

Let

\[
d_{\text{act}}^2 = [(\mathbf{\mu}_{\text{act}} - \mathbf{\mu}_0)^T \Sigma^{-1}(\mathbf{\mu}_{\text{act}} - \mathbf{\mu}_0)]^{1/2},
\]

which is the M-distance between \( \mathbf{\mu}_{\text{act}} \) and \( \mathbf{\mu}_0 \). Note that \( d_{\text{act}} \) represents the actual shift size while \( d_i \) is the critical shift size that the chart is designed to detect. The results of the actual average costs in Table 3.4-3.6 were obtained from a MATLAB simulation program. The total number of simulation runs for each case was 50000.
First, we investigate the sensitivity of the economic design of a Bayesian chart with the misspecified directions of $\mu_{\text{act}}$ under the assumption that $d_{\text{act}}^2 = d_1^2$. This assumption indicates that the shift size is assumed to be estimated correctly. To demonstrate the impact of misspecified direction of $\mu_{\text{act}}$ with correct estimation of the shift size on the average cost of the process, we run the simulation program for both Set A and Set B for $d_1^2 = 0.5, 1.0, 2.0, 3.0$.

Table 3.4 and Table 3.5 show that there is a fairly small cost penalty with the ratio of $d_{\text{act}}^2 / d_1^2$ near to 1. However, the cost penalty increases when the ratio decreases. It is also noticeable that the change of cost is pretty small when $d_1^2 = 0.5, 1.0, 2.0$ for Set B which contains larger cost parameters.

**Table 3.4 Results of sensitivity analysis for Set A ($d_{\text{act}}^2 = d_1^2$)**

<table>
<thead>
<tr>
<th>$\frac{d_{\text{act}}^2}{d_1^2}$</th>
<th>$d_1^2 = 0.5$</th>
<th>$d_1^2 = 1.0$</th>
<th>$d_1^2 = 2.0$</th>
<th>$d_1^2 = 3.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. cost</td>
<td>Change (%)</td>
<td>Av. cost</td>
<td>Change (%)</td>
<td>Av. cost</td>
</tr>
<tr>
<td>0.5</td>
<td>17.837</td>
<td>34.135</td>
<td>16.671</td>
<td>33.109</td>
</tr>
<tr>
<td>0.9</td>
<td>13.714</td>
<td>3.135</td>
<td>12.804</td>
<td>2.233</td>
</tr>
<tr>
<td>1</td>
<td>13.298</td>
<td>0.000</td>
<td>12.524</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Table 3.5 Results of sensitivity analysis for Set B ($d_{\text{act}}^2 = d_1^2$)**

<table>
<thead>
<tr>
<th>$\frac{d_{\text{act}}^2}{d_1^2}$</th>
<th>$d_1^2 = 0.5$</th>
<th>$d_1^2 = 1.0$</th>
<th>$d_1^2 = 2.0$</th>
<th>$d_1^2 = 3.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. cost</td>
<td>Change (%)</td>
<td>Av. cost</td>
<td>Change (%)</td>
<td>Av. cost</td>
</tr>
<tr>
<td>0.5</td>
<td>26.842</td>
<td>6.544</td>
<td>26.833</td>
<td>6.516</td>
</tr>
<tr>
<td>0.6</td>
<td>26.319</td>
<td>4.468</td>
<td>26.640</td>
<td>5.749</td>
</tr>
<tr>
<td>0.7</td>
<td>26.367</td>
<td>4.660</td>
<td>25.965</td>
<td>3.072</td>
</tr>
<tr>
<td>0.8</td>
<td>25.881</td>
<td>2.729</td>
<td>25.595</td>
<td>1.604</td>
</tr>
<tr>
<td>0.9</td>
<td>25.494</td>
<td>1.196</td>
<td>25.302</td>
<td>0.440</td>
</tr>
<tr>
<td>1</td>
<td>25.193</td>
<td>0.000</td>
<td>25.191</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Secondly, we investigate the sensitivity of the economic design of a Bayesian chart with misspecified shift sizes under the assumption that $\mu_{act} = k\mu_1$, where $k$ is any positive real number. To demonstrate the impact of a misspecified shift size with correct estimation of the direction of $\mu_1$ on the average cost of the process, the actual average costs were simulated for $d_{act}^2 = 0.5, 1.0, 2.0,$ and $3.0$ for each $d_i^2 = 0.5, 1.0, 2.0, 3.0$ for set A. In order to do so, we need to know the actual mean of $Z_m$, which is $-2nd_{12}^2$ from Lemma 2, given $d_{act}^2$. With the assumption that $\mu_0$ is a zero vector, it can be shown that $d_{12}^2$ is equal to $d_i d_{act}$. In Table 3.6, $d_i^2$ is the M-distance between the assumed $\mu_1$ and $\mu_0$, and $d_{act}^2$ is the M-distance between the actual $\mu_1$ and $\mu_0$. The notation Av. cost in Table 3.6 represents the actual cost for the actual $d_{act}^2$ by using the Bayesian chart with the optimal parameters calculated based on the assumed $d_i^2$. For example, the optimal Bayesian chart parameters based on $d_i^2 = 0.5$ are $n = 8$, $h = 0.994$, and $P^* = 0.399$. Using these parameters, the actual cost for $d_{act}^2 = 1.0$ was simulated and the simulation result is 12.621 in Table 3.6.

<table>
<thead>
<tr>
<th>$d_{act}^2$</th>
<th>$d_i^2 = 0.5$</th>
<th>$d_i^2 = 1.0$</th>
<th>$d_i^2 = 2.0$</th>
<th>$d_i^2 = 3.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. cost</td>
<td>Change (%)</td>
<td>Av. cost</td>
<td>Change (%)</td>
<td>Av. cost</td>
</tr>
<tr>
<td>0.5</td>
<td>13.298</td>
<td>0.000</td>
<td>13.807</td>
<td>3.825</td>
</tr>
<tr>
<td>1.0</td>
<td>12.621</td>
<td>0.777</td>
<td>12.524</td>
<td>0.000</td>
</tr>
<tr>
<td>2.0</td>
<td>12.224</td>
<td>2.989</td>
<td>11.961</td>
<td>0.772</td>
</tr>
<tr>
<td>3.0</td>
<td>12.129</td>
<td>4.156</td>
<td>11.870</td>
<td>1.936</td>
</tr>
</tbody>
</table>

Table 3.6 shows that the cost penalty is lower with underestimated shift sizes than the penalty with overestimated shift sizes. It is important to point out that the percent change in the average
cost is very large when the actual M-distance is much smaller than the estimated M-distance. For example, the percent change in the average cost is 32% when $d_{act}^2 = 0.5$ with the assumed $d_1^2 = 3.0$. In other words, the average cost would increase significantly if the system is monitored using the control chart with the parameters that are chosen optimally for highly overestimated M-distance.
Chapter 4 Multivariate Bayesian Control Chart for CBM and its Economic and Economic Statistical Design

4.1 Introduction

Maintenance is a broad term that encapsulates a variety of activities and decisions, such as tests, inspections, repairs and replacements, in order to maintain or restore a system to a functional condition. Over the past two decades, maintenance has evolved from simple corrective maintenance to age-based maintenance, inspection-based maintenance, and eventually to CBM. The idea of CBM is that preventive maintenance can be triggered by monitoring condition-related variables. CBM modelling, as the leading edge of maintenance research, has attracted considerable attentions over recent years. It has been shown that CBM can improve the performance of the system, reduce system downtime and eventually bring significant economic gains for organizations when integrated into their maintenance program.

As its name implies, CBM decides whether maintenance is necessary by taking the actual condition of equipment or system into account. In order to determine the actual condition of system, a sample is usually collected from the system at regular intervals. However, in a real situation the system condition is only partially observable through sampling or inspection. For example, sensors, such as accelerometers, can be installed to collect vibration data in order to assess the state of the gearbox indirectly. Obviously, the actual condition of the gearbox is
known only after the machine has been shut down and examined by technicians. The observation process is stochastically related to the hidden deterioration process defining the system state. Because failures occur and are observable in most real systems, an observable failure state should be included in CBM optimization models. Some CBM optimization models considering a failure state for systems with a univariate or discrete value observation processes can be found in the literature (see Makis and Jardine, 1992; Jensen and Hsu, 1993; Makis and Jiang, 2003; Gupta and Lawsirirat, 2006).

There is a close connection between CBM and statistical process control. Industrial practitioners and academic researchers have increasingly recognized the economic benefits obtained by applying SPC techniques to maintenance decision-making. Some efforts can be found in the literature. For example, Tagaras (1988) first proposed a cost model to integrate process control and maintenance operations. Preventive maintenance is performed regularly if no assignable cause is confirmed. Cassady et al. in 2000 applied $\bar{X}$ control chart to perform preventive maintenance in addition to an age-replacement preventive maintenance policy by considering an equipment failure which is assumed not to cause any process failures. More development can be found in Linderman et al. (2005) and Ivy and Nembhard (2005). One common assumption in these efforts is to assume that there are two system states, an in-control and out-of-control state. This is a reasonable assumption for quality control applications because SPC uses statistical techniques, such as control charts, to only indicate the occurrence of assignable causes in a process. However, such an assumption is not appropriate for maintenance modeling because to avoid costly system failures is one of the indispensable concerns of maintenance management. Wu (2006) and Liu (2006) considered an observable failure state in their contributions to the
economic and economic statistical designs of multivariate Chi-square chart and MEWMA chart for CBM applications.

As mentioned in the previous chapter, Makis (2006b and 2007) proved that a simple control limit structure is optimal when a multivariate Bayesian control chart is applied in statistical process control. By considering the same simple control limit policy structure and including an observable failure state, the multivariate Bayesian control chart is designed for CBM applications and illustrated by a numerical example in section 4.2. Optimization models for economic and economic statistical design of the control chart are developed to determine the optimal control chart parameters to minimize the total expected average maintenance cost in section 4.3 to section 4.6. Computational results are shown to compare the performance of the Bayesian chart and a Chi-square chart in section 4.7.

4.2 The Multivariate Bayesian Control Chart for CBM Application

It is assumed that a system can be in one of two operational states, a normal state and a warning state, or in a failure state. Only the failure state is observable. The observation process is represented by q normally distributed characteristics. A sample from the observation process is obtained at regular sampling epochs. The sample is used to calculate a statistic which is plotted on the multivariate Bayesian control chart. Here, the statistic of the multivariate Bayesian control chart is the posterior probability that the system is in the warning state. The following maintenance policy is applied: the system is replaced when it fails or when an alarm from the Bayesian chart indicates that the system is in the warning state.
4.2.1 Problem Statement and Assumptions

We consider a system which can be in one of three states \{1, 2, 3\}. States 1 and 2 are operational states and state 3 is the failure state, which is observable. State 1 is not worse than state 2. In the following text of this chapter, state 1 is also referred to as a normal state and state 2 as a warning state. The operational states are not observable. The state process \{X_t, t \in R_+\} is a continuous-time homogeneous Markov chain with the state space \{1, 2, 3\}. Let \(S_x = \{1,2\}\) and \(\overline{S_x} = \{1,2,3\}\).

Preventive replacement can be performed to avoid costly failures. When the system fails, failure replacement follows. Let \(\xi\) denote the observable failure time of the system. It is assumed that the system is renewed to the normal state after replacement. To facilitate preventive replacement decision making, the system is monitored at times \(h, 2h, \ldots\), and \(\{Y_1, Y_2, \ldots\}\) represents the observation process, where \(Y_i\) is the information collected at time \(ih, i = 1, 2, \ldots\).

Let \(y_j = (y_{j1}, y_{j2}, \ldots, y_{jq})^T\) denote the \(j\)th observation \(q\)-dimensional vector in a size \(n\) sample obtained at the \(i\)th inspection epoch. When \(X_{ih} = 1\), the observation vector \(y_j\) follows multivariate normal distribution \(N_q(\mu_1, \Sigma)\) and when \(X_{ih} = 2\), it follows \(N_q(\mu_2, \Sigma)\). The M-distance between \(\mu_1\) and \(\mu_2\) is denoted by \(d = [(\mu_2 - \mu_1)^T \Sigma^{-1} (\mu_2 - \mu_1)]^{1/2}\). Let \(Y_i = \Delta \in R^{nxq}\) indicate the occurrence of failure and define \(\overline{S_y} = \{R^{nxq}, \Delta\}\). The length of time that the system remains in an operational state before making a transition into a different state is exponentially distributed. Let \(\lambda_1\) denote the transition rate from state 1 to state 2, \(\lambda_2\) be the transition rate from state 2 to state 3, and \(\lambda_3\) be the transition rate from state 1 to state 3.

4.2.2 Proposed Multivariate Bayesian Procedure
For $t \geq 0$, let $P_t(i)$ denote the probability that the process is in state $i$ at time $t$ given the observations up to time $t$, which can be expressed as follows:

$$P_t(i) = P(X_t = i \mid Y_\tau, \ldots, Y_{\lfloor t/h \rfloor h}, I_{\lfloor \xi > t \rfloor})$$

The probability $P_{mh}(2)$, $m = 1, 2, \ldots$, that the system is in the warning state at time $mh$ given the system history up to time $mh$ is plotted on the multivariate Bayesian control chart for maintenance decision-making. Let $\bar{P}_t(i) = P(X_t = i \mid Y_\tau, \ldots, Y_{\lfloor t/h \rfloor h}, I_{\lfloor \xi > t/h \rfloor h})$ be the probability that the process is in state $i$ at time $t$ given the observable information up to time $\lfloor t/h \rfloor h$. The instantaneous transition rate $Q$-matrix for the continuous time Markov chain process is,

$$Q = \{q_{ij}\} = \begin{bmatrix} -(\lambda_i + \lambda_3) & \lambda_1 \\ 0 & -\lambda_2 \\ 0 & 0 \end{bmatrix}$$

The sojourn time in state $i$ follows an exponential distribution with mean $1/v_i$ where $v_i = -q_{ii}$ for any $i \in S_\xi$. The transition matrix for the embedded discrete-time Markov chain has the form as follows,

$$P = \{p_{ij}\} = \begin{bmatrix} 0 & \frac{\lambda_i}{(\lambda_i + \lambda_3)} \\ \frac{\lambda_i}{(\lambda_i + \lambda_3)} & 0 \\ 0 & 0 \end{bmatrix}$$

where $p_{ij} = P(X_{mh} = j \mid X_{(m-1)h} = i)$ for $m=1, 2, \ldots$.

By solving the Kolmogorov’s backward differential equations, the transition probability matrix for the continuous-time Markov chain is given by
where \( r^*_j(t) = P(X_{(m-1)h+t} = j \mid X_{(m-1)h} = i) \) for \( m = 1, 2, \ldots \). This matches the results in Wu and Makis (2007).

Obviously, at sampling time \( mh \), \( P_{mh}(i) = P_{mh}(i) \). The formula to update \( P_{mh}(2) \) is given by

**Theorem 1** as follows.

**Theorem 1.** For \( m \geq 0 \) and given \( P_{0}(i), i = 1, 2, \ldots \), \( P_{mh}(2) \) can be updated as follows:

\[
P_{mh}(2) = \frac{\overline{P}_{mh}(2)}{\exp\left(\frac{1}{2}(nd^2 + Z_m)\right) \cdot \overline{P}_{mh}(1) + \overline{P}_{mh}(2)},
\]

where \( \overline{P}_{mh}(1) = e^{-(\lambda_1 + \lambda_2)h} \cdot (1 - P_{(m-1)h}(2)) \) and

\[
\overline{P}_{mh}(2) = \frac{\lambda_1 (e^{-(\lambda_1 + \lambda_2)h} - e^{-(\lambda_1 + \lambda_2)h})}{\lambda_1 + \lambda_3 - \lambda_2} \cdot (1 - P_{(m-1)h}(2)) + e^{-(\lambda_1 + \lambda_2)h} \cdot P_{(m-1)h}(2).
\]

Here,

\[
Z_m = \sum_{j=1}^{n} (\mathbf{y}_j^m - \mathbf{\mu}_1)^T \Sigma^{-1} (\mathbf{\mu}_1 - \mathbf{\mu}_2)
\]

is \( N(0, 4nd^2) \) when the system is in state 1 and

\[
Z_m \sim N(-2nd^2, 4nd^2)
\]

when it is in state 2.

Proof:

For \((m-1)h < t < mh\), \( m \geq 1 \),

\[
\overline{P}_{i}(t) = \begin{cases}
  r^*_i(t - (m-1)h) \cdot \overline{P}_{(m-1)h}(1) + r_2(t - (m-1)h) \cdot \overline{P}_{(m-1)h}(2) & \text{for } i \in S_x \\
  r^*_i(t - (m-1)h) \cdot \overline{P}_{(m-1)h}(1) + r_2(t - (m-1)h) \cdot \overline{P}_{(m-1)h}(2) + \overline{P}_{(m-1)h}(3) & \text{for } i = 3
\end{cases}
\]

(4.1)
Note that $P_i(3)$ can be calculated by subtracting $P_i(1) + P_i(2)$ from 1, since $\sum_{i=1}^{3} P_i(i) = 1$.

From the definition of $P_i(i)$, $P_{mh}(i) = P(X_{mh} = i \mid Y_h, \ldots, Y_{(m-1)h}, I_{[x_h > (m-1)h]})$

For $t = mh$, $m \geq 1$ and given $y_j^m$,

\[
P_{mh}(1) = \frac{\prod_{j=1}^{n} f(y_j^m \mid \mu_1, \Sigma) \cdot P_{mh}(1)}{\prod_{j=1}^{n} f(y_j^m \mid \mu_1, \Sigma) \cdot P_{mh}(1) + \prod_{j=1}^{n} f(y_j^m \mid \mu_2, \Sigma) \cdot P_{mh}(2)}
\]

\[
= \frac{P_{mh}(1) + \exp(-\frac{1}{2}(nd^2 + 2 \sum_{j=1}^{n} (y_j^m - \mu_1)^T \Sigma^{-1}(\mu_1 - \mu_2))) \cdot P_{mh}(2)}{P_{mh}(1) + \exp(-\frac{1}{2}(nd^2 + Z_m)) \cdot P_{mh}(2)}
\]

and

\[
P_{mh}(2) = \frac{\prod_{j=1}^{n} f(y_j^m \mid \mu_2, \Sigma) \cdot P_{mh}(2)}{\prod_{j=1}^{n} f(y_j^m \mid \mu_1, \Sigma) \cdot P_{mh}(1) + \prod_{j=1}^{n} f(y_j^m \mid \mu_2, \Sigma) \cdot P_{mh}(2)}
\]

\[
= \frac{\exp(\frac{1}{2}(nd^2 + 2 \sum_{j=1}^{n} (y_j^m - \mu_1)^T \Sigma^{-1}(\mu_1 - \mu_2)) \cdot P_{mh}(1) + P_{mh}(2)}{\exp(\frac{1}{2}(nd^2 + Z_m)) \cdot P_{mh}(1) + P_{mh}(2)}
\]

where $P_{mh}(i) = 0$ for $i = 3$.

From (4.1), we have

\[
P_{mh}(1) = e^{-(\lambda_1 + \lambda_2)h} \cdot P_{(m-1)h}(1) = e^{-(\lambda_1 + \lambda_3)h} \cdot (1 - P_{(m-1)h}(2))
\]

and

\[
P_{mh}(2) = \frac{\lambda_1(e^{-(\lambda_2 + \lambda_3)h} - e^{-(\lambda_2 + \lambda_3)h})}{\lambda_1 + \lambda_3 - \lambda_2} \cdot P_{(m-1)h}(1) + e^{-(\lambda_2 + \lambda_3)h} \cdot P_{(m-1)h}(2)
\]

\[
= \frac{\lambda_1(e^{-(\lambda_2 + \lambda_3)h} - e^{-(\lambda_1 + \lambda_3)h})}{\lambda_1 + \lambda_3 - \lambda_2} \cdot (1 - P_{(m-1)h}(2)) + e^{-(\lambda_2 + \lambda_3)h} \cdot P_{(m-1)h}(2)
\]

73
$Z_m$ has been proven to follow $N(0,4n^2)$ when the system is in state 1 and to follow $N(-2n^2,4n^2)$ when it is in state 2 (See the proof in Makis (2007)).

For $(m-1)h < t < mh$ and $m \geq 1$,

if $\xi > t$,

$$P_i(i) = \begin{cases} 
\frac{\bar{P}_i(i)}{\bar{P}_i(1) + \bar{P}_i(2)} & \text{for } i \in S_x \\
0 & \text{for } i = 3
\end{cases}$$

otherwise,

$$P_i(i) = \begin{cases} 
0 & \text{for } i \in S_x \\
1 & \text{for } i = 3
\end{cases}.$$  \hspace{1cm} (4.3)

### 4.2.3 An Illustrative Example for the Implementation of the Multivariate Bayesian Control Chart

Consider a system described by a three dimensional observation vector, which is $N_3(\mu_1, \Sigma)$ when the system is in a normal state and $N_3(\mu_2, \Sigma)$ when the system is in a warning state. Suppose the mean vectors and the covariance matrix are given as

$$\mu_1 = [0,0,0]^T, \quad \mu_2 = [2,2,4]^T \quad \text{and} \quad \Sigma = \begin{bmatrix} 2.0 & 0.5 & 1.2 \\ 0.5 & 1.0 & 0.8 \\ 1.2 & 0.8 & 1.5 \end{bmatrix}.$$  

The transition rates are $\lambda_1 = 0.1$, $\lambda_2 = 0.15$ and $\lambda_3 = 0.01$ for the system deterioration process.

The sample size $n$ is 1 and a sample is taken every one time unit ($h=1$). The control limit for the multivariate Bayesian control chart can be selected according to the statistical performance requirements, such as the average run length (ARL) when the system is in state 1. Using the
assumed parameters given above, a MATLAB program was written to generate a control limit $P^*$ that would satisfy the statistical requirement $\text{ARL}_1=200$. It was found that $P^* = 0.67$.

Twenty observations simulated for the system described above and the multivariate Bayesian control chart statistics are shown in Table 4.1. The system started from state 1 and moved to the state 2 between sample 16 and sample 17. The system failure occurred after sample 18 was taken. Assume that $P_0(2) = 0$ and $P_0(1) = 1$. Given the first observation $y^1=[-0.9960, -1.2014, -1.2699]^T$, $P_\hat{h}^1(1)$, $P_\hat{h}^2(2)$, $Z_1$ and $P_h$ are calculated as follows:

$$P_\hat{h}^1(1) = e^{-(0.1+0.01)} \cdot (1 - 0) = 0.8958$$

$$P_\hat{h}^2(2) = \frac{0.1(e^{-0.15} - e^{-0.11})}{0.1 + 0.01 - 0.15} \cdot (1 - 0) + e^{-0.15} \cdot 0 = 0.0878$$

$$Z_1 = 2 \times ([-0.9960, -1.2014, -1.2699] - [0, 0, 0]) = \begin{bmatrix} 2.0 & 0.5 & 1.2 \end{bmatrix}^{-1} \begin{bmatrix} 0.5 & 1.0 & 0.8 \\ 1.2 & 0.8 & 1.5 \end{bmatrix} = \begin{bmatrix} 2.0 & 0.5 & 1.2 \end{bmatrix}^{-1} (\begin{bmatrix} 0 & 0 & 0 \\ 2 & 2 & 4 \end{bmatrix} - 6.2642) = 9.5515E-06$$

$$P_h = \frac{P_{mh}^1(2)}{\exp\left(\frac{1}{2}nd^2 + 2Z_m\right) \cdot P_{mh}^1(1) + P_{mh}^2(2)} = \frac{0.0878}{\exp\left(\frac{1}{2}(12.2081 + 6.2642)\right) \cdot 0.8958 + 0.0878} = 9.5515E-06$$

<table>
<thead>
<tr>
<th>Sample</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$P_{mh}$</th>
<th>$\chi^2_{mh}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.9960</td>
<td>-1.2014</td>
<td>-1.2699</td>
<td>9.5515E-06</td>
<td>1.565838</td>
</tr>
<tr>
<td>2</td>
<td>2.1510</td>
<td>0.5018</td>
<td>2.1330</td>
<td>4.9970E-02</td>
<td>3.821436</td>
</tr>
<tr>
<td>3</td>
<td>-0.9846</td>
<td>-0.2391</td>
<td>-1.1371</td>
<td>1.4421E-05</td>
<td>1.09759</td>
</tr>
<tr>
<td>4</td>
<td>0.8301</td>
<td>-0.0275</td>
<td>0.7013</td>
<td>1.2638E-03</td>
<td>0.638162</td>
</tr>
<tr>
<td>5</td>
<td>0.9449</td>
<td>0.1630</td>
<td>1.1502</td>
<td>5.8367E-03</td>
<td>1.243088</td>
</tr>
<tr>
<td>6</td>
<td>3.2658</td>
<td>1.3072</td>
<td>2.2316</td>
<td>1.3439E-02</td>
<td>5.608105</td>
</tr>
<tr>
<td>7</td>
<td>1.2914</td>
<td>0.3752</td>
<td>0.0264</td>
<td>4.6338E-05</td>
<td>2.062668</td>
</tr>
</tbody>
</table>
Figure 4.1 The multivariate Bayesian control chart for the example
The Bayesian control chart and traditional Chi-square chart are shown in Fig. 4.1 and Fig. 4.2, respectively. The control limit for the Chi-square chart is 12.84 with the probability of type I error $\alpha = 0.005$ and the ARL$_1$=200. We can see that the statistic on the Bayesian control chart increases dramatically after the 17$^{th}$ sample and the 18$^{th}$ statistic value falls outside the control limit to indicate that the system is in the warning state. However, the deterioration of the system cannot be exhibited clearly by the Chi-square chart. All of the points on the Chi-square chart are below the control limit. This example illustrates that the Bayesian chart can prevent the costly system failure, while the Chi-square chart may not provide any signal when the system is in state 2.

**4.3 Model Description**

In this section, a detailed description of the model is given. First, the notation used in this chapter is defined as follows:
\( C_1 \) = the operation and maintenance cost per unit time while the system is in normal state.

\( C_2 \) = the operation and maintenance cost per unit time while the system is in warning state.

\( C_{fa} \) = the inspection cost per alarm.

\( C_p \) = the preventive maintenance cost.

\( C_f \) = the failure maintenance cost (including the system installation cost and the system down cost).

\( C_s \) = the sampling cost.

\( T_{fi} \) = expected full inspection time.

\( T_{pm} \) = expected time for preventive maintenance.

\( T_f \) = expected time to restore the system after the occurrence of system failure.

\( E(\text{Cost}) \) = expected cycle cost.

\( E(\text{CL}) \) = expected cycle length.

It is assumed that total time of \( T_{pm} \) and \( T_{fi} \) required to perform a full inspection and preventive maintenance is less than the time \( T_f \) required to perform corrective maintenance on the system to restore the failed system to a good as new condition. Similarly, it is assumed that \( C_{fa} + C_p < C_f \).

The proposed maintenance policy with the Bayesian control chart is shown in Fig. 4.3. The system starts in state 1 and runs on a continuous basis. A sample is collected from the system at time \( ih \) at the cost of \( C_s \) and the statistic on the multivariate Bayesian control chart is updated according to the formula which is derived in the previous section. Note that the lower control limit on the multivariate Bayesian control chart is zero. If the updated posterior probability falls
above the upper control limit of the control chart, it means that the chart gives a signal. If the control chart does not give any signal, the system continues without any interruption. However, if the control chart signals, the system will stop and a full inspection at the cost of $C_{fa}$ takes place to check the actual condition of the system. The system is resumed after the full inspection if no warning condition is found. In this scenario, the signal from the control chart is called false alarm. If the control chart signal is validated by the inspection, preventive maintenance at the cost of $C_p$ will be performed to bring the system to the initial normal state and the signal is treated as a true alarm. The system is resumed after the full inspection and preventive maintenance.

Figure 4.3 Condition-based maintenance policy

In Case 1 as shown in Fig. 4.3, the control chart signals when the system is in state 1, thus, the signal is treated as a false alarm after an inspection. No maintenance is triggered and the system continues to operate. In Case 2 in Fig. 4.3, the control chart signals after the system shifts to state 2 and the signal is called a true alarm. Preventive maintenance takes place after the full inspection and the system is restored back to state 1. In Case 3 in Fig. 4.3, the system shifts to state 3 from state 1 or state 2 before the control chart gives any signal. This means that the system failure occurs and costly replacement at the cost of $C_f$ is inevitable.
4.4 Economic Design and Economic Statistical Design of Multivariate Bayesian Control Chart for CBM

The objective of this research is to choose optimal Bayesian control chart parameters that minimize the expected average cost per unit time. Namely, the Bayesian control chart parameters include sampling interval $h$, sample size $n$ and upper control limit $P^*$. Since the sample size $n$ is equal to one in most maintenance applications, $n$ will not be considered as a parameter to be optimized in this study. We develop a cost model based on renewal theory to mathematically express the expected average cost per unit time. The concept of a system cycle is first introduced. The end of a cycle is triggered by an alarm from the control chart or a system failure. More specifically, if a false alarm occurs, the cycle length is equal to the time when the alarm occurs plus the time for a full inspection. In the case of a true alarm, the cycle is completed after preventive maintenance is performed. If the system fails, the cycle is ended after the system restoration. Based on the model assumption that the length of time between state transitions is exponentially distributed, the process can be divided into a series of independent successive cycles. From renewal theory, the expected average cost of the system is equal to the expected cycle cost divided by the expected cycle length. The optimization problem for the economic design of the multivariate Bayesian control chart for CBM application can therefore be formulated as

$$\begin{align*}
\min z(h, P^*) &= \frac{E(\text{Cost of Cycle})}{E(\text{CL})} \\
\text{s.t. } &0 \leq P^* \leq 1 \\
&h > 0
\end{align*}$$

(4.4)
The objective function in the above optimization problem is an expected average cost function of the chart parameters. \( P^* \) is required to be within zero and 1 since it is the upper control limit for the posterior probabilities. In the following sections, the formulas of expected cycle length and expected cycle cost will be derived to construct the objective function. The choice of a suitable statistical constraint for economic statistical design of the control chart will also be discussed.

### 4.4.1 Expected Cycle Length

Conditioning on the time when the system moves out of state 1, the expected cycle length is as follows:

\[
E(\text{CL}) = \int_0^{\infty} E(\text{CL}|T_1 = t) \nu_i e^{-\nu_i t} dt
\]

where \( T_1 \) is the time when the system moves out of state 1. By dividing the whole interval, the expected cycle length can be rewritten as:

\[
E(\text{CL}) = \sum_{m_i = 0}^{\infty} \left\{ \int_{m_{ih}}^{(m_{ih} + 1)h} E(\text{CL}|T_1 = t, 1 \rightarrow 2) p_{12} \nu_i e^{-\nu_i t} dt \right\}
\]

\[
= \sum_{m_i = 0}^{\infty} \left[ CL_{1 \rightarrow 2}(m_i) + CL_{1 \rightarrow 3}(m_i) \right]
\]

where \( CL_{1 \rightarrow 3}(m_i) = \int_{m_{ih}}^{(m_{ih} + 1)h} E(\text{CL}|T_1 = t, 1 \rightarrow 3) p_{13} \nu_i e^{-\nu_i t} dt \)

\( CL_{1 \rightarrow 2}(m_i) = \int_{m_{ih}}^{(m_{ih} + 1)h} E(\text{CL}|T_1 = t, 1 \rightarrow 2) p_{12} \nu_i e^{-\nu_i t} dt \)

First, the formula for \( CL_{1 \rightarrow 3}(m_i) \) is derived.
\[ CL_{1\rightarrow 3}(m_t) = \int_{m_h}^{(m+1)h} \left\{ E(CL|T_1=1, 1 \rightarrow 3) p_{13} V_t e^{-\nu t} dt \right\} \]

\[ = \int_{m_h}^{(m+1)h} \left\{ \sum_{i=1}^{m} E(CL|T_i=1, 1 \rightarrow 3, AL_i) \cdot P_{1\rightarrow 3}(AL_i|m_t) p_{13} V_t e^{-\nu t} dt + E(CL|T_i=1, 1 \rightarrow 3, \text{no alarm before failure}) \cdot (1 - \sum_{i=1}^{m} P_{1\rightarrow 3}(AL_i|m_t)) p_{13} V_t e^{-\nu t} dt \right\} \]  \hspace{1cm} (4.7)

where

\[ P_{1\rightarrow 3}(AL_i|m_t) = P(\text{the first alarm occurs at } ith \text{ sampling point}|T_i=1, t \in [m_h, (m+1)h], 1 \rightarrow 3) . \]

If an alarm occurs before the system transfers from state 1 to state 3, the alarm is a false alarm and the cycle length is equal to the time when the alarm occurs plus the time for a full inspection. However, if the system shifts from state 1 to state 3, the system fails before any alarm occurs and the cycle length is equal to the failure time plus the time required to restore the system. Thus,

\[ CL_{1\rightarrow 3}(m_t) = \int_{m_h}^{(m+1)h} \left\{ \sum_{i=1}^{m} (ih + T_f) \cdot P(AL_i|T_i=1, 1 \rightarrow 3) p_{13} V_t e^{-\nu t} dt \right\} \]

\[ + (t + T_f) [1 - \sum_{i=1}^{m} P(AL_i|T_i=1, 1 \rightarrow 3)] p_{13} V_t e^{-\nu t} dt \]  \hspace{1cm} (4.8)

Next, the formula for \( CL_{1\rightarrow 2}(m_t) \) is derived as follows. Let \( T_2 \) denote the time when the system moves out of state 2 to state 3. Note that the system will enter state 3 with \( p_{23} = 1 \) if it moves out of state 2.

Let \( P_{1\rightarrow 2}(AL_i|m_t, m_i) = P(AL_i|T_1=t, t \in [m_h, (m+1)h], 1 \rightarrow 2, T_2=s, t+s \in [m_h, (m+1)h]) \)
and \( P_{i\rightarrow 2}(AL_i| m_1, m_2) = \)

\[
P(AL_j| T_i = t, t \in [m_i h, (m_i + 1)h], 1 \rightarrow 2, T_2 = s, t + s \in [(m_i + 1 + m_2) h, (m_i + 1 + m_2 + 1) h])
\]

Conditioning on \( T_2 \), the time when the system moves out of state 2, we have

\[
\begin{align*}
CL_{i\rightarrow 2}(m_i) & = \int_{m_i h}^{(m_i + 1)h} \left\{ E(CL|T_i = t, 1 \rightarrow 2)p_{12}v_i e^{-\gamma t} dt \right\} \\
& = \int_{m_i h}^{(m_i + 1)h} p_{12}v_i e^{-\gamma t} \left[ \int_{0}^{(m_i + 1)h} E(CL|T_i = t, 1 \rightarrow 2, T_2 = s)p_{23}v_2 e^{-\gamma s} ds \right] dt \\
& = \int_{m_i h}^{(m_i + 1)h} \left[ \int_{0}^{(m_i + 1)h} \left\{ \sum_{i=1}^{m_i} E(CL|T_i = t, 1 \rightarrow 2, T_2 = s, AL_i) \cdot P_{1\rightarrow 2}(AL_i| m_1, m_i) \right\} + E(CL|T_i = t, 1 \rightarrow 2, T_2 = s, no\ AL\ before\ FA) \right] \cdot p_{23}v_2 e^{-\gamma s} ds \right] dt \\
& = \int_{m_i h}^{(m_i + 1)h} \left[ \int_{0}^{(m_i + 1)h} \left\{ \sum_{i=1}^{m_i} E(CL|T_i = t, 1 \rightarrow 2, T_2 = s, AL_i) \cdot P_{1\rightarrow 2}(AL_i| m_1, m_i) \right\} + E(CL|T_i = t, 1 \rightarrow 2, T_2 = s, no\ AL\ before\ FA) \right] \cdot p_{23}v_2 e^{-\gamma s} ds \right] dt \\
& = \int_{m_i h}^{(m_i + 1)h} \left[ \sum_{i=1}^{m_i} E(CL|T_i = t, 1 \rightarrow 2, T_2 = s, AL_i) \cdot P_{1\rightarrow 2}(AL_i| m_1, m_i) \right] \cdot p_{23}v_2 e^{-\gamma s} ds \right] dt
\end{align*}
\]

Given \( t \in [m_i h, (m_i + 1)h] \) and \( s \in [(m_i + 1)h - t + m_2 h, (m_i + 1)h - t + (m_2 + 1)h] \), the expected cycle lengths for three different scenarios are given as follows,

1) False alarm which is illustrated by Case 1 in Fig. 4.3

\[
E(CL|T_i = t, 1 \rightarrow 2, T_2 = s, AL_i) = ih + T_{ji} \text{ for } 1 \leq i \leq m_i
\]

2) True alarm which is illustrated by Case 2 in Fig. 4.3
For $m_1 + 1 \leq i \leq m_1 + 1 + m_2$

\[ E(\text{CL} | T_1 = t, 1 \rightarrow 2, T_2 = s, AL_i) = ih + T_{ji} + T_{pm} \]

3) No alarms before system failure which is illustrated by Case 3 in Fig. 4.3

\[ E(\text{CL} | T_1 = t, 1 \rightarrow 2, T_2 = s, \text{ no ALs before FA}) = t + T_f \]

Therefore, we have

\[ CL_{1 \rightarrow 2}(m_i) \quad (4.10) \]

\[
\begin{align*}
&= \int_{m_1 h}^{(m_1 + 1)h - t} \left[ \sum_{i=1}^{m_1} (ih + T_{ji}) \cdot P_{1 \rightarrow 2}(AL_i | m_1, m_1) \\ &+ (t + s + T_f) [1 - \sum_{i=1}^{m_1} P_{1 \rightarrow 2}(AL_i | m_1, m_1)] \right] p_{12} v_2 e^{-v_2 t} dt \\
&+ \sum_{m_2 = 0}^{+\infty} \int_{(m_1 + 1)h - t + m_2 h}^{(m_1 + 1)h - t + m_2 h} \left[ \sum_{i=1}^{m_1} (ih + T_{ji}) \cdot P_{1 \rightarrow 2}(AL_i | m_1, m_2) \\ &+ \sum_{i=m_1 + 1}^{m_1 + m_2} (ih + T_{ji} + T_{pm}) \cdot P_{1 \rightarrow 2}(AL_i | m_1, m_2) \\ &+ (t + s + T_f) [1 - \sum_{i=1}^{m_1 + m_2} P_{1 \rightarrow 2}(AL_i | m_1, m_2)] \right] p_{12} v_2 e^{-v_2 t} dt \\
&= \int_{m_1 h}^{(m_1 + 1)h - t} \left[ \sum_{i=1}^{m_1} (ih + T_{ji}) \cdot P_{1 \rightarrow 2}(AL_i | m_1, m_1) \\ &+ (t + s + T_f) [1 - \sum_{i=1}^{m_1} P_{1 \rightarrow 2}(AL_i | m_1, m_1)] \right] p_{12} v_2 e^{-v_2 t} dt \\
&+ \sum_{m_2 = 0}^{+\infty} \int_{(m_1 + 1)h - t + m_2 h}^{(m_1 + 1)h - t + m_2 h} \left[ \sum_{i=1}^{m_1 + m_2} (ih + T_{ji}) \cdot P_{1 \rightarrow 2}(AL_i | m_1, m_2) \\ &+ \sum_{i=m_1 + 1}^{m_1 + m_2} T_{pm} \cdot P_{1 \rightarrow 2}(AL_i | m_1, m_2) \\ &+ (t + s + T_f) [1 - \sum_{i=1}^{m_1 + m_2} P_{1 \rightarrow 2}(AL_i | m_1, m_2)] \right] p_{12} v_2 e^{-v_2 t} dt
\end{align*}
\]
Thus,

\[
\begin{align*}
\sum_{i=1}^{m} & \left(i\hbar P_{1\to 2} (AL_i | m_1, m_1) + \left[1 - \sum_{i=1}^{m} P_{1\to 2} (AL_i | m_1, m_1)\right] \frac{1}{\eta_i}\right) \\
& \cdot p_{12} e^{-\gamma m_h} (1 - e^{-\eta \hbar}) \\
+ (T_f - T_f) \sum_{i=1}^{m} P_{1\to 2} (AL_i | m_1, m_1) + T_f \\
+ \left[1 - \sum_{i=1}^{m} P_{1\to 2} (AL_i | m_1, m_1)\right] [(m + 1) h - h(1 - e^{-\eta \hbar})^{-1} + \frac{1}{\eta}] \cdot p_{12} e^{-\gamma m_h} (1 - e^{-\eta \hbar}) \\
- \frac{\eta}{\eta - \eta_i} P_{12} e^{-\eta \hbar} (e^{-(\eta - \eta_i) \hbar} - 1) e^{-\gamma m_h} \sum_{m=0}^{\infty} e^{-\gamma m_h} \\
& \left\{ \begin{array}{l}
\sum_{i=1}^{m_1 + 1 + m_2} \frac{1+i\hbar}{i\hbar} P_{1\to 2} (AL_i | m_1, m_2) \\
+ T_f \sum_{i=1}^{m_1 + 1 + m_2} P_{1\to 2} (AL_i | m_1, m_2) \\
+ T_f \left[1 - \sum_{i=1}^{m_1 + 1 + m_2} P_{1\to 2} (AL_i | m_1, m_2)\right] \\
+ \left[1 - \sum_{i=1}^{m_1 + 1 + m_2} P_{1\to 2} (AL_i | m_1, m_2)\right] \\
\cdot [(\frac{1}{\eta} + (m_1 + 1) h + m_2 h)(1 - e^{-\eta \hbar}) - he^{-\eta \hbar}] 
\end{array} \right\}
\end{align*}
\]
\[ \begin{align*}
CL_{1 \rightarrow 3}(m_1) + CL_{1 \rightarrow 2}(m_1) \\
= p_{13}e^{-\nu_m h} (1 - e^{-\nu_h}) \\
= p_{13}e^{-\nu_m h} \left( \frac{h}{v} \sum_{i=1}^{m_i} i P_{1 \rightarrow 3}(AL_i|m_1) + \left[ 1 - \sum_{i=1}^{m_i} P_{1 \rightarrow 3}(AL_i|m_1) \right] \left( (m_1 + 1)h - h(1 - e^{-\nu_h})^{-1} + \frac{1}{v} \right) \\
+ (T_{h_f} - T_f) \sum_{i=1}^{m_i} P_{1 \rightarrow 3}(AL_i|m_1) + T_f \right) \\
+ \left( \sum_{i=1}^{m_i} i h_i P_{1 \rightarrow 2}(AL_i|m_1,m_1) + \left[ 1 - \sum_{i=1}^{m_i} P_{1 \rightarrow 2}(AL_i|m_1,m_1) \right] \frac{1}{v_2} \right) \cdot p_{12} (1 - e^{-\nu_h}) e^{-\nu_m h} \right) \right. \\
+ (T_{h_f} - T_f) \sum_{i=1}^{m_i} P_{1 \rightarrow 2}(AL_i|m_1,m_1) + T_f \\
+ [1 - \sum_{i=1}^{m_i} P_{1 \rightarrow 2}(AL_i|m_1,m_1)] \left( (m_1 + 1)h - h(1 - e^{-\nu_h})^{-1} + \frac{1}{v} \right) p_{12} e^{-\nu_m h} (1 - e^{-\nu_h}) \\
+ [1 - \sum_{i=1}^{m_i} P_{1 \rightarrow 2}(AL_i|m_1,m_1)] \left( \sum_{i=1}^{m_i+1+m_2} i h_i P_{1 \rightarrow 2}(AL_i|m_1,m_2) \right) \\
+ T_F \sum_{i=m_1+1}^{m_1+m_2} \cdot p_{1 \rightarrow 2}(AL_i|m_1,m_2) \\
+ (T_{h_f} - T_f) \sum_{i=1}^{m_i+1+m_2} P_{1 \rightarrow 2}(AL_i|m_1,m_2) \\
+ T_f \left[ 1 - \sum_{i=1}^{m_i+1+m_2} P_{1 \rightarrow 2}(AL_i|m_1,m_2) \right] \\
+ [1 - \sum_{i=1}^{m_i+1+m_2} P_{1 \rightarrow 2}(AL_i|m_1,m_2)] \left( \left( \frac{1}{v} + (m_1 + 1)h + m_2 h(1 - e^{-\nu_h}) - he^{-\nu_h} \right) \right) \right) \\
\end{align*} \]

\[ 4.4.2 \text{ Expected Cycle Cost} \]

Similar to the previous discussion of the expected cycle length, the expected cycle cost can be obtained as follows:
\[ E(\text{Cost}) \]
\[ = \int_{0}^{\infty} E(\text{Cost}|T_1 = t)v_i e^{-\lambda t} dt \]
\[ = \int_{0}^{\infty} E(\text{Cost}|T_1 = t, 1 \rightarrow 2)p_{12}v_i e^{-\lambda t} dt + \int_{0}^{\infty} E(\text{Cost}|T_1 = t, 1 \rightarrow 3)p_{13}v_i e^{-\lambda t} dt \]
\[ = \sum_{m_i = 0}^{\infty} \left\{ \int_{m_i}^{(m_i+1)h} \left\{ E(\text{Cost}|T_1 = t, 1 \rightarrow 2)p_{12}v_i e^{-\lambda t} dt \right\} \right\} + \sum_{m_i = 0}^{\infty} \left\{ \int_{m_i}^{(m_i+1)h} \left\{ E(\text{Cost}|T_1 = t, 1 \rightarrow 3)p_{13}v_i e^{-\lambda t} dt \right\} \right\} \]

(4.12)

Let \( \text{Cost}_{t \rightarrow 3}(m_i) = \int_{m_i}^{(m_i+1)h} \left\{ E(\text{Cost}|T_1 = t, 1 \rightarrow 3)p_{13}v_i e^{-\lambda t} dt \right\} \)

and \( \text{Cost}_{t \rightarrow 2}(m_i) = \int_{m_i}^{(m_i+1)h} \left\{ E(\text{CL}|T_1 = t, 1 \rightarrow 2)p_{12}v_i e^{-\lambda t} dt \right\} \)

thus, \( E(\text{Cost}) = \sum_{m_i = 0}^{\infty} [\text{Cost}_{t \rightarrow 2}(m_i) + \text{Cost}_{t \rightarrow 3}(m_i)] \)

Next the formulas for \( \text{Cost}_{t \rightarrow 3}(m_i) \) and \( \text{Cost}_{t \rightarrow 2}(m_i) \) will be derived.

\[ \text{Cost}_{t \rightarrow 3}(m_i) = \int_{m_i}^{(m_i+1)h} \left\{ E(\text{Cost}|T_1 = t, 1 \rightarrow 3)p_{13}v_i e^{-\lambda t} dt \right\} \]

\[ = \int_{m_i}^{(m_i+1)h} \left\{ \sum_{i=1}^{\infty} E(\text{Cost}|T_1 = t, 1 \rightarrow 3, AL_i)P(AL_i|T_1 = t, 1 \rightarrow 3)p_{13}v_i e^{-\lambda t} dt \right\} \]

(4.13)

Given \( t \in [m_i h, (m_i + 1)h] \), we have

\[ E(\text{Cost}|T_1 = t, 1 \rightarrow 3, AL_i) = iC_s + ihC_i + C_{fa} = i(C_s + hC_i) + C_{fa} \]

and

\[ E(\text{Cost}|T_1 = t, 1 \rightarrow 3, \text{no ALs before FA}) = m_i C_s + tC_i + C_f \]
\[ 
Cost_{1 \rightarrow 5}(m_1) = \int_{m_h}^{(m+1)h} \{ E(Cost| T_i = t, 1 \rightarrow 3) p_{13} v_i e^{-\gamma v_i t} dt \} 
\]

\[ 
= \int_{m_h}^{(m+1)h} \left\{ \sum_{i=1}^{m} \left[ i(C_x + hC_i) + C_{fa} \right] P_{1 \rightarrow 3}(AL_i|m_1) p_{13} v_i e^{-\gamma v_i t} dt \right. 
\]

\[ + \left( m_1 C_x + tC_i + C_f \right) \left( 1 - \sum_{i=1}^{m} P_{1 \rightarrow 3}(AL_i|m_1) \right) p_{13} v_i e^{-\gamma v_i t} dt 
\]

\[ + \sum_{i=1}^{m} i(C_x + hC_i) P_{1 \rightarrow 3}(AL_i|m_1) \right\} tC_i \left( 1 - \sum_{i=1}^{m} P_{1 \rightarrow 3}(AL_i|m_1) \right) 
\]

\[ 
= p_{13} e^{-\gamma v_i m_h} \left\{ \sum_{i=1}^{m} i(C_x + hC_i) P_{1 \rightarrow 3}(AL_i|m_1) \right. 
\]

\[ + m_1 C_x + C_f + (C_{fa} - m_1 C_x - C_f) \sum_{i=1}^{m} P_{1 \rightarrow 3}(AL_i|m_1) \right\} 
\]

\[ + C_i \left[ 1 - \sum_{i=1}^{m} P_{1 \rightarrow 3}(AL_i|m_1) \right] \left[ (m_i + 1)h(1 - e^{-\gamma h}) - h + \frac{h}{n} (1 - e^{-\gamma h}) \right] \]  

(4.14)

Similarly,

\[ 
Cost_{1 \rightarrow 2}(m_1) = \int_{m_h}^{(m+1)h} \{ E(Cost| T_i = t, 1 \rightarrow 2) p_{12} v_i e^{-\gamma v_i t} dt \} 
\]

\[ = \int_{m_h}^{(m+1)h} \left\{ p_{12} v_i e^{-\gamma v_i t} \int_0^x E(Cost| T_i = t, 1 \rightarrow 2, T_2 = s) p_{23} v_2 e^{-\gamma v_2 s} ds dt \right. 
\]

\[ + \sum_{m_2=0}^{+\infty} \left\{ E(Cost| T_i = t, 1 \rightarrow 2, T_2 = s) p_{23} v_2 e^{-\gamma v_2 s} ds \right\} \] 

\[ \left. \int_{m_h}^{(m+1)h} p_{12} v_i e^{-\gamma v_i t} dt \right\} \]  

(4.15)
Next, the conditional expectations of the cycle cost when the system shifts to state 2 will be given. Given \( t, s \in [m_1h, (m_1 + 1)h] \),

\[
E(\text{Cost}|T_1 = t, 1 \rightarrow 2, T_2 = s, AL_i) = ihC_1 + iC_s + C_{fa} = i(C_s + hC_1) + C_{fa} \quad \text{for } 1 \leq i \leq m_1.
\]

and \( E(\text{Cost}|T_1 = t, 1 \rightarrow 2, T_2 = s, \text{no AL before FA}) = tC_1 + sC_2 + m_1C_s + C_f \)

Given \( t \in [m_1h, (m_1 + 1)h] \) and \( s \in [(m_1 + 1)h - t + m_2h, (m_1 + 1)h - t + (m_2 + 1)h] \), the expected cycle costs for three different scenarios are given as follows,

1) False alarm which is illustrated by Case 1 in Fig. 4.3

\[
E(\text{Cost}|T_1 = t, 1 \rightarrow 2, T_2 = s, AL_i) = ihC_1 + iC_s + C_{fa} = i(C_s + hC_1) + C_{fa} \quad \text{for } 1 \leq i \leq m_1.
\]

2) True alarm which is illustrated by Case 2 in Fig. 4.3

For \( m_1 + 1 \leq i \leq m_1 + m_2 \)
\[
E(\text{Cost}| T_1 = t, 1 \to 2, T_2 = s, AL_i) = tC_1 + (ih - t)C_2 + iC_s + C_{fa} + C_p.
\]

\[
= t(C_1 - C_2) + i(hC_2 + C_s) + C_{fa} + C_p.
\]

3) No alarms before system failure which is illustrated by Case 3 in Fig. 4.3

\[
E(\text{Cost}| T_1 = t, 1 \to 2, T_2 = s, \text{ no ALs before FA}) = tC_1 + sC_2 + (m_1 + 1 + m_2)C_s + C_f.
\]

Then,

\[
Cost_{1\to2}(m_i) = \int_{m_i}^{(m_i+1)h} [\sum_{i=1}^{m_i} [i(C_s + hC_1) + C_{fa}] \cdot P_{1\to2}(AL_i | m_i, m_i) + tC_1 + sC_2 + m_iC_s + C_f] [1 - \sum_{i=1}^{m_i} P_{1\to2}(AL_i | m_i, m_i)] \cdot p_{23} \cdot v_2 e^{-\nu_2 t} ds
\]

\[
= \int_{m_i}^{(m_i+1)h} \left[ \sum_{i=1}^{m_i} [i(C_s + hC_1) + C_{fa}] \cdot P_{1\to2}(AL_i | m_i, m_i) + tC_1 + sC_2 + m_iC_s + C_f \right] [1 - \sum_{i=1}^{m_i} P_{1\to2}(AL_i | m_i, m_i)] \cdot p_{23} \cdot v_2 e^{-\nu_2 t} ds
\]

\[
= \int_{m_i}^{(m_i+1)h} \left[ \sum_{i=1}^{m_i} [i(C_s + hC_1) + C_{fa}] \cdot P_{1\to2}(AL_i | m_i, m_i) + tC_1 + sC_2 + m_iC_s + C_f \right] [1 - \sum_{i=1}^{m_i} P_{1\to2}(AL_i | m_i, m_i)] \cdot p_{23} \cdot v_2 e^{-\nu_2 t} ds
\]

Let \( a = \int_{0}^{(m_i+1)h - t} \left[ \sum_{i=1}^{m_i} [i(C_s + hC_1) + C_{fa}] \cdot P_{1\to2}(AL_i | m_i, m_i) + tC_1 + sC_2 + m_iC_s + C_f \right] [1 - \sum_{i=1}^{m_i} P_{1\to2}(AL_i | m_i, m_i)] \cdot p_{23} \cdot v_2 e^{-\nu_2 t} ds \) and
Thus, $\text{Cost}_{t \rightarrow 2}(m_1) = \int_{m_1 h}^{(m_1 + 1)h} [a + b] p_{12} v_1 e^{-v_1 t} dt = \int_{m_1 h}^{(m_1 + 1)h} a \cdot p_{12} v_1 e^{-v_1 t} dt + \int_{m_1 h}^{(m_1 + 1)h} b \cdot p_{12} v_1 e^{-v_1 t} dt$. After simplifying $\int_{m_1 h}^{(m_1 + 1)h} a \cdot p_{12} v_1 e^{-v_1 t} dt$ and $\int_{m_1 h}^{(m_1 + 1)h} b \cdot p_{12} v_1 e^{-v_1 t} dt$, we have

\begin{align*}
\int_{m_1 h}^{(m_1 + 1)h} a \cdot p_{12} v_1 e^{-v_1 t} dt &= \left\{ \sum_{i=1}^{m_1} [i(C_s + h C_i) + C_{f_i}] \cdot P_{1 \rightarrow 2}(AL_i | m_1, m_1) \right\} \\
&= \left\{ (m_1 C_s + C_f) \cdot \left[ 1 - \sum_{i=1}^{m_1} P_{1 \rightarrow 2}(AL_i | m_1, m_1) \right] \right\} \cdot p_{12} e^{-v_1 h} (1 - e^{-v_1 h})
\end{align*}

\begin{align*}
\int_{m_1 h}^{(m_1 + 1)h} b \cdot p_{12} v_1 e^{-v_1 t} dt &= \left\{ \sum_{i=1}^{m_1} [(i + 1) C_s + (i + 1) C_f + C_{f_i}] \cdot P_{1 \rightarrow 2}(AL_i | m_1, m_1) \right\} \\
&= \left\{ [(m_1 + 1) C_s + (m_1 + 1) C_f + C_s + C_f] \cdot \left[ 1 - \sum_{i=1}^{m_1} P_{1 \rightarrow 2}(AL_i | m_1, m_1) \right] \right\} \cdot p_{12} \frac{v_1}{v_1 - v_2} e^{-v_2 h} e^{-v_1 m_1 h} (e^{-(v_1 - v_2)h} - 1)
\end{align*}

\begin{align*}
&+ C_1 \cdot \left[ 1 - \sum_{i=1}^{m_1} P_{1 \rightarrow 2}(AL_i | m_1, m_1) \right] \cdot \left\{ (m_1 + 1) h (1 - e^{-v_1 h}) - h + \frac{1}{v_1} (1 - e^{-v_1 h}) \right\} p_{12} e^{-v_1 m_1 h}
\end{align*}

\begin{align*}
&+(C_2 - C_1) \cdot \left[ 1 - \sum_{i=1}^{m_1} P_{1 \rightarrow 2}(AL_i | m_1, m_1) \right] p_{12} \frac{v_1}{v_1 - v_2} e^{-v_2 h} e^{-v_1 m_1 h} \\
&\left\{ m_1 h - \frac{1}{v_1 - v_2} (e^{-(v_1 - v_2)h} - 1) - (m_1 + 1) h e^{-(v_1 - v_2)h} \right\}
\end{align*}

(4.17)

and
\[
\int_{m_0 h}^{(m_0 +1) h} b p_{i_2} v_1 e^{-v_2 t} dt
= e^{-v_2 h} e^{-v_1 v_2 h} \frac{v_1}{v_1 - v_2} P_{i_2}^* \\
= e^{-v_2 h} e^{-v_1 v_2 h} \frac{v_1}{v_1 - v_2} P_{i_2}^*
\]
\[
\sum_{m_2=0}^{\infty} e^{-v_2 m_2 h} \\
\left[
\begin{align*}
\sum_{i=1}^{m_1} &\left[i(C_i + hC_1) + C_0 \right] P_{i-2} (AL_i | m_1, m_2) \\
+ &\sum_{i=m_1 +1}^{m_0 +1} i(hC_2 + C_1) + C_0 \\
+ &\left\{ m_1 + (m_1 + m_2) C_1 + C_f \right\} \\
\left[1 - \sum_{i=1}^{m_1 + m_2} P_{i-2} (AL_i | m_1, m_2) \right] \\
\end{align*}
\right] (1 - e^{-v_2 h}) (e^{-v_1 v_2 h} - 1)
\]
\[
\left[
\begin{align*}
\left( \frac{1}{v_2} + (m_1 + 2) h + m_2 h \right) (1 - e^{-v_2 h} - h) (e^{-v_1 v_2 h} - 1) \\
(C_i - C_0) (1 - e^{-v_2 h}) (1 - \sum_{i=1}^{m_0} P_{i-2} (AL_i | m_1, m_2)) \cdot \\
\left[ m_i h - (e^{-v_1 v_2 h} - 1) \frac{1}{v_1} - (m_1 + 1) h e^{-v_1 v_2 h} \right]
\end{align*}
\right]
\]

Then we have
\[\text{Cost}_{i ightarrow 3}(m_i) + \text{Cost}_{i ightarrow 2}(m_i)\]

\[= e^{-\gamma_i m_i h}(1 - e^{-\gamma_i h}) \left[ \sum_{i=1}^{m_i} [i(C_s + hC_f) + C_{fa}] \cdot P_{i ightarrow 3}(AL_i | m_i) \right] + (m_i + 1)h(1 - e^{-\gamma_i h}) - h + \frac{1}{\eta_i} (1 - e^{-\gamma_i h}) \right] \]

\[+ \frac{1}{\nu_i} \cdot \frac{p_{12}}{\nu_i - \nu_2} e^{-\gamma_i m_i h} e^{-\gamma_i h} (e^{-\nu_i h} - 1) \left[ \sum_{i=1}^{m_i} [i(C_s + hC_f) + C_{fa}] \cdot P_{i ightarrow 2}(AL_i | m_i, m_1) \right] \]

\[+ \frac{p_{12}}{\nu_i - \nu_2} e^{-\gamma_i m_i h} e^{-\gamma_i h} (C_2 - C_1) \left[ 1 - \sum_{i=1}^{m_i} P_{i ightarrow 2}(AL_i | m_i, m_1) \right] \left[ \frac{m_i h - \frac{1}{e^{-\nu_i h}} (e^{-\nu_i h} - 1) - (m_i + 1)h e^{-\nu_i h}}{\eta_i - \nu_2} \right] \]

\[+ \frac{p_{12}}{\nu_i - \nu_2} e^{-\gamma_i h} \sum_{m_2 = 0}^{m_2} e^{-\gamma_i m_2 h} \left[ (1 - e^{-\gamma_i h}) (e^{-\nu_i h} - 1) - C_2 \left[ 1 - \sum_{i=1}^{m_i + 1} P_{i ightarrow 2}(AL_i | m_i, m_2) \right] \right] \]

\[\cdot \left[ \frac{(1/\gamma_i + (m_i + 1)h + m_2 h) (1 - e^{-\gamma_i h}) - h}{e^{-\nu_i h} - 1} \right] + (C_2 - C_1) \left[ (1 - e^{-\gamma_i h}) (1 - \sum_{i=1}^{m_i} P_{i ightarrow 2}(AL_i | m_i, m_2)) \right] \]

\[\left[ \frac{m_i h - (e^{-\nu_i h} - 1) - \frac{1}{\eta_i - \nu_2} - (m_i + 1)h e^{-\nu_i h}}{\eta_i - \nu_2} \right] \]

### 4.4.3 Constraint Specification for Maintenance Applications

In previous sections, the formulas for the expected cycle cost and the expected cycle length have been derived to obtain the objective function of the optimization model for the pure economic design. In maintenance applications, other considerations should be taken into account in addition to the pure economic design. For example, the prevention of a system failure and the
increased rate of detection after a process shift are important issues that should be incorporated into the model. It is especially crucial for the systems which involve public service or safety, such as mass transit systems. In order to express such concerns, a statistical constraint can be added to the optimization model for the pure economic design of the control chart. First, a few important probabilities need to be considered to formulate the statistical constraint.

In order to measure the statistical performance of the control chart, the probability \( P(\text{Real Alarm}) \) that the system is ended by a real alarm from the control chart is considered. Note that the real alarm can only occur in the case where the chart signals when the system is in a warning state.

The probability \( P(\text{Real Alarm}) \) can be expressed as follows:

\[
P(\text{Real Alarm}) = \sum_{m_i=0}^{+\infty} \left\{ P(\text{Real Alarm} | T_i = t, 1 \rightarrow 2) p_{12} v_i e^{-\mathcal{V} t} dt \right\} 
\]

where

\[
\begin{align*}
&= \sum_{m_i=0}^{+\infty} \left\{ P(\text{Real Alarm} | T_i = t, 1 \rightarrow 2) p_{12} v_i e^{-\mathcal{V} t} dt \right\} \\
&= \sum_{m_i=0}^{+\infty} \left\{ P(\text{Real Alarm} | T_i = t, 1 \rightarrow 2) p_{12} v_i e^{-\mathcal{V} t} dt \right\} \\
&= \sum_{m_i=0}^{+\infty} \left\{ P(\text{Real Alarm} | T_i = t, 1 \rightarrow 2) p_{12} v_i e^{-\mathcal{V} t} dt \right\} \\
&= \sum_{m_i=0}^{+\infty} \left\{ P(\text{Real Alarm} | T_i = t, 1 \rightarrow 2) p_{12} v_i e^{-\mathcal{V} t} dt \right\}
\end{align*}
\]

\[
= \frac{v_i}{v_i - v_2} p_{12} e^{-\mathcal{V} t} \left( 1 - e^{-v_i t} \right) \left( 1 - e^{-v_2 t} \right) \sum_{m_i=0}^{+\infty} e^{-vm_i} \sum_{m_2=0}^{m_i+1} P_{1 \rightarrow 2}(AL_i | m_i, m_2) 
\]

\[
= \frac{v_i}{v_i - v_2} p_{12} e^{-\mathcal{V} t} \left( 1 - e^{-v_i t} \right) \left( 1 - e^{-v_2 t} \right) \sum_{m_i=0}^{+\infty} e^{-vm_i} \sum_{m_2=0}^{m_i+1} P_{1 \rightarrow 2}(AL_i | m_i, m_2) 
\]
Let $P(\text{Fail})$ denote the probability that a cycle is ended by a system failure. Next, the formula to calculate this probability will be derived. Conditioning on the time when the system shifts out of state 1, the probability $P(\text{Fail})$ can expressed as follows:

$$P(\text{Fail}) = \int_0^\infty P(\text{Fail}|T_1 = t)v_1e^{-v_1t}dt$$

$$= \int_0^\infty P(\text{Fail}|T_1 = t, 1 \to 2)p_{12}v_1e^{-v_1t}dt + \int_0^\infty P(\text{Fail}|T_1 = t, 1 \to 3)p_{13}v_1e^{-v_1t}dt$$

$$= \sum_{m_1=0}^{+\infty} \left\{ \int_{m_1}^{(m_1+1)h} P(\text{Fail}|T_1 = t, 1 \to 2)p_{12}v_1e^{-v_1t}dt \right\}$$

$$+ \sum_{m_1=0}^{+\infty} \left\{ \int_{m_1}^{(m_1+1)h} P(\text{Fail}|T_1 = t, 1 \to 3)p_{13}v_1e^{-v_1t}dt \right\}$$

The integrals in (4.20) can be simplified respectively as follows

$$\int_{m_1}^{(m_1+1)h} P(\text{Fail}|T_1 = t, 1 \to 3)p_{13}v_1e^{-v_1t}dt$$

$$= \int_{m_1}^{(m_1+1)h} \left[ 1 - \sum_{i=1}^{m_1} P_{1 \to 3}(AL_i|m_i) \right] p_{13}v_1e^{-v_1t}dt$$

$$= p_{13}e^{-v_1m_1h}(1 - e^{-v_1h})[1 - \sum_{i=1}^{m_1} P_{1 \to 3}(AL_i|m_i)]$$

(4.22)

and

$$\int_{m_1}^{(m_1+1)h} P(\text{Fail}|T_1 = t, 1 \to 2)p_{12}v_1e^{-v_1t}dt$$

$$= \int_{m_1}^{(m_1+1)h} \left[ 1 - \sum_{i=1}^{m_1} P_{1 \to 2}(AL_i|m_i,m_i) \right] p_{12}v_1e^{-v_1t}dt$$

$$= \int_{m_1}^{(m_1+1)h} \left[ 1 - \sum_{i=1}^{m_1} P_{1 \to 2}(AL_i|m_i,m_i) \right] p_{12}v_1e^{-v_1t}dt$$

$$= \left[ (1 - e^{-v_1h}) + \frac{v_1}{v_1-v_2} e^{-v_1h}(e^{-(v_1-v_2)h} - 1) \right] p_{12}e^{-v_1m_1h}$$

$$- \frac{v_1}{v_1-v_2} p_{12}e^{-v_1h}(e^{-(v_1-v_2)h} - 1)(1 - e^{-v_1h})e^{-v_1m_1h}$$

$$\sum_{m_1=0}^{+\infty} \sum_{i=1}^{m_1} P_{1 \to 2}(AL_i|m_i,m_i)$$

(4.23)
4.5 A Markov Chain Model for the Multivariate Bayesian Control Chart for CBM

In order to further simplify the formula for the expected cycle length, the formulas to calculate the following probabilities need to be derived,

a) \( P_{1 \rightarrow 3}(AL_i | m_i), \ 1 \leq i \leq m_i \)

b) \( P_{1 \rightarrow 2}(AL_i | m_i, m_i), \ 1 \leq i \leq m_i \)

c) \( P_{1 \rightarrow 2}(AL_i | m_i, m_2), \ 1 \leq i \leq m_i + 1 + m_2 \)

In this section, a Markov chain model is used to derive formulas to calculate the above probabilities. To approximate the continuous random variable \( P_{mh}(2) \), the interval [0, 1] is partitioned into \( l+1 \) intervals. From the definition of a Markov chain, \( \{P_{mh}(2)\} \) can be approximated by a discrete state space Markov chain model. We assume that the Markov chain \( \{Q_m\} \) is in transient state \( i \) when \( \frac{(i-1)p}{l} \leq P_{mh}(2) < \frac{ip}{l} \) for \( 1 \leq i \leq l \) and \( m=0, 1, \ldots \), and it is in the absorbing state \( l+1 \) when \( P^* \leq P_{mh}(2) < 1 \). To approximate the continuous random variable \( P_{mh}(2) \), the interval \([0, P^*] \) is partitioned into \( l+1 \) intervals. From the definition of a Markov chain, \( \{P_{mh}(2)\} \) can be approximated by a discrete state space Markov chain model. We assume that the Markov chain \( \{Q_m\} \) is in transient state \( i \) when \( \frac{(i-0.5)p}{l} \leq P_{mh}(2) < \frac{(i+0.5)p}{l} \) for \( 1 \leq i < l \), it is in state 0 when \( 0 \leq P_{mh}(2) < \frac{0.5p}{l} \) and in state \( l \) when \( \frac{(l-0.5)p}{l} \leq P_{mh}(2) < P^* \). It is in absorbing state \( l+1 \) when \( P^* \leq P_{mh}(2) < 1 \). Since state \( l+1 \) is the only absorbing state, \( P_{l(i+1)} = 0 \) for \( 0 \leq j \leq l \) and \( P_{l(i+1)} = 1 - \sum_{j=1}^{l} P_{ij} \).
4.5.1 The Transition Matrix When System is in the Normal State

Let \( P(N) = \mathbf{P}_j(N)_{i(l+1)}^{(l+1)} \) represent the transition matrix which only contains the transition probabilities for the transient states when the system is in the normal state.

From Theorem 1,

\[
\mathbf{P}_{ij}(N) = P(X_m = j|Q_0, \cdots, Q_{(m-1)} = i, \text{the system is in the normal state}) \text{ for } 1 \leq i, j \leq l
\]

\[
= P\left( \frac{(j - 0.5)P^*}{l} \leq P_{mh}(2) < \frac{(j + 0.5)P^*}{l} | Q_{(m-1)} = i \right)
\]

\[
= P\left( \frac{(j - 0.5)P^*}{l} \leq \frac{\bar{P}_{mh}(2)}{\bar{P}_{mh}(1)} < \frac{(j + 0.5)P^*}{l} | Q_{(m-1)} = i \right)
\]

\[
= P\left( \frac{U_j}{\bar{P}_{mh}(2)} - \frac{L_j}{\bar{P}_{mh}(1)} < \exp(\frac{1}{2}(nd^2 + Z_m)) \leq \frac{U_j}{\bar{P}_{mh}(2)} - \frac{L_j}{\bar{P}_{mh}(1)} | Q_{(m-1)} = i \right)
\]

\[
= P\left( 2 \ln \left( \frac{\bar{P}_{mh}(2)}{\bar{P}_{mh}(1)} - \frac{L_j}{\bar{P}_{mh}(1)} \right) - nd^2 < Z_m \leq 2 \ln \left( \frac{\bar{P}_{mh}(2)}{\bar{P}_{mh}(1)} - \frac{L_j}{\bar{P}_{mh}(1)} \right) - nd^2 | Q_{(m-1)} = i \right)
\]

where \( U_j = \frac{(j + 0.5)P^*}{l} \) and \( L_j = \frac{(j - 0.5)P^*}{l} \) for \( 1 \leq j < l \) and \( L_0 = 0, \ U_l = P^* \). Given \( Q_{(m-l)} = i \),

\( P_{(m-l),h}(2) \) can be approximated by the middle value \( \frac{a_i + b_i}{2} \) of the interval. Let

\[
a_i = \bar{P}_{mh}(1) \approx e^{-(\lambda_1 + \lambda_3)}(1 - \frac{P^*}{l}), \text{ and } b_i = \bar{P}_{mh}(2) = \frac{\lambda_1(e^{\lambda_1h} - e^{-(\lambda_1 + \lambda_3)h})}{\lambda_1 + \lambda_3 - \lambda_2} \cdot (1 - \frac{P^*}{l}) + e^{-\lambda_3h} \cdot \frac{P^*}{l}
\]

Then, for \( 0 \leq i \leq l, 0 < j \leq l \),

97
\[
P_y(N) = P \left\{ 2 \ln \left( \frac{b_j - b_i}{a_i} \right) - nd^2 < Z_m \leq 2 \ln \left( \frac{b_j - b_i}{a_i} \right) - nd^2 \mid Q_{(m-1)} = i \right\}
\]

Since \( Z_m \) follows \( N(0, 4nd^2) \) when the system is in the normal state, we have,

\[
P_y(N) = \Phi \left( \frac{b_j - b_i}{2 \ln(\frac{U_j}{a_i}) - nd^2} \right) - \Phi \left( \frac{b_j - b_i}{2 \ln(\frac{L_j}{a_i}) - nd^2} \right)
\]

(4.24)

For \( j = 0 \),

\[
P_{00}(N) = 1 - \Phi \left( \frac{b_i}{2 \ln(\frac{U_0}{a_i}) - nd^2} \right)
\]

(4.25)

4.5.2 The Transition Matrix When System is in the Warning State

Let \( P(W) = [P_y(W)]_{i(l+1)\times(l+1)} \) represent the transition matrix which only contains the transition probabilities for the transient states when the system is in the warning state. Similar to the previous development, for \( 0 \leq i \leq l, 0 < j \leq l \),

\[
P_y(W) = \Phi \left( \frac{2 \ln(\frac{L_j^i - 1}{a_i - 1}) + 2(\lambda_i + \lambda_j)h + nd^2}{2d \sqrt{n}} \right) - \Phi \left( \frac{2 \ln(\frac{U_j^i - 1}{a_i - 1}) + 2(\lambda_i + \lambda_j)h + nd^2}{2d \sqrt{n}} \right)
\]
Thus, $P_g(W) = \Phi \left( \frac{b_i - b_j}{2 \ln \left( \frac{L_j}{a_i} \right) + nd^2} \right) - \Phi \left( \frac{b_i - b_j}{2 \ln \left( \frac{U_j}{a_i} \right) + nd^2} \right)$ (4.26)

For $j = 0$, $P_{i0}(W) = 1 - \Phi \left( \frac{b_i - b_j}{2 \ln \left( \frac{U_0}{a_i} \right) + nd^2} \right)$ (4.27)

4.5.3 Several Lemmas

In the previous section, the transition matrices of the Markov chain models were obtained. Next, define $1 = (1, \ldots, 1)^T$ and let $1 \times (l + 1)$ row vector $J(i)$ have a one in the place corresponding to the starting state $i$ and zeros in the other places for $i = 0, \ldots, l$.

For $1 \leq j \leq m_i$, $0 \leq i \leq l$ and $m_i \geq 0$, let $P(AL_j \mid T_i = t, 1 \rightarrow 3, t \in [m_ih, (m_i + 1)h], Q_0 = i)$ denote the probability that an alarm first occurs on the $j^{th}$ sample given that $Q_0 = i$ and the process transfers from state 1 to state 3 at time $t$ for $t \in [m_ih, (m_i + 1)h]$.

**Lemma 1.** For $1 \leq j \leq m_i$, $0 \leq i \leq l$ and $m_i \geq 0$, $P(AL_j \mid T_i = t, 1 \rightarrow 3, t \in [m_ih, (m_i + 1)h], Q_0 = i)$ is equal to $J(i) \cdot P^{(j-1)}(N)v$, where $v = (I - P(N))1$.

Proof:
Let $\triangle$ denote the absorbing state $l+1$. Let $\Delta P_{i,a}^{(j)}(m, N)$ denote the probability that the Markov chain will be in the state $\triangle$ on the $j^{th}$ sample without having visited state $\triangle$ up to the $j^{th}$ sample given that the initial state $Q_0 = i$ and the system shifts from the normal state to the failure state between $[m_i h, (m_i + 1)h]$. Let $P(N)$ denote the transient transition matrix of the Markov chain when the system is in the normal state.

Define $P_{jk}^{(j)} = \left( P_{jj}, P_{j1}, ..., P_{jm} \right)$, $v = (v_1, ..., v_m)^T$, and $\Delta P_{i,a}^{(j)}(m, N) = (\Delta P_{i,a}^{(j)}(m, N), ..., \Delta P_{i,a}^{(j)}(m, N))^T$

For $1 \leq j \leq m$, $i = 0, ..., l$

$$P(AL_j \mid T_i = t, 1 \rightarrow 3, \ t \in [m_i h, (m_i + 1)h], P_{0h}(2) = i)$$

$$= P(P_{jh}(2) = l + 1, \text{not in state } l + 1 \text{ before the } j^{th} \text{ sample} \mid T_i = t, 1 \rightarrow 3, \ t \in [m_i h, (m_i + 1)h], P_{0h}(2) = i)$$

$$= \Delta P_{i,a}^{(j)}(m, N)$$

For $j = 1$,

$$\Delta P_{i,a}^{(1)}(m, N) = v_i$$

From which it follows that $\Delta P_{i,a}^{(j)}(m, N) = P^{(0)}(N) \cdot v$, where $P^{(0)}(N) = I$.

For $2 \leq j \leq m$,

$$\Delta P_{i,a}^{(j)}(m, N)$$

$$= P(Q(j) = l + 1, \text{not in state } l + 1 \text{ before the } j^{th} \text{ sample} \mid T_i = t, 1 \rightarrow 3, \ t \in [m_i h, (m_i + 1)h], Q(0) = i)$$

$$= \sum_{k \neq a} \left\{ P(Q(j) = l + 1, \text{not in state } l + 1 \text{ before} \mid T_i = t, 1 \rightarrow 3, \ t \in [m_i h, (m_i + 1)h], Q(0) = i, Q(1) = k) \right\}$$

$$= \sum_{k \neq a} P_{ik}(N) \cdot \Delta P_{ia}^{(j-1)}(m, N)$$

100
In the matrix form, \( \Delta P_a^{(j)}(m_1, N) = P(N) \Delta P_a^{(j-1)}(m_1, N) \).

Next, by mathematical induction, we show \( \Delta P_a^{(j)}(m_1, N) = P^{(j-1)}(N) \mathbf{v} \) for \( 2 \leq j \leq m_1 \).

From the above discussion, we know that \( \Delta P_{i,n}^{(2)}(m_1, N) = \sum_{k \in \mathcal{A}} P_k(N) \Delta P_{k,n}^{(2-1)}(m_1, N) = \sum_{k=1}^{l} P_k(N) \mathbf{v}_k \)

\( = P_r(N) \mathbf{v} \), where \( P_r(N) \) represents the \( i \)th row of the matrix \( P(N) \).

Hence, \( \Delta P_a^{(j)}(m_1, N) = P^{(j-1)}(N) \mathbf{v} \) is true for \( j=2 \).

Assume \( \Delta P_a^{(j)}(m_1, N) = P^{(j-1)}(N) \mathbf{v} \) is true for \( j=n \), then \( \Delta P_a^{(n)}(m_1, N) = P^{(n-1)}(N) \mathbf{v} \). Next, it can be shown that it is also true for \( j=n+1 \).

\( \Delta P_a^{(n+1)}(m_1, N) = P(N) \Delta P_a^{(n+1-1)}(m_1, N) = P(N) \mathbf{P}^{(n-1)}(N) \mathbf{v} = \mathbf{P}^{(n)}(N) \mathbf{v} \)

Hence, \( \Delta P_a^{(j)}(m_1, N) = P^{(j-1)}(N) \mathbf{v} \) is true for \( 2 \leq j \leq m_1 \). We have \( \Delta P_a^{(j)}(m_1, N) = P^{(j-1)}(N) \mathbf{v} \) for \( 1 \leq j \leq m_1 \). If the initial state is state \( i \), the probability is \( J(i) \mathbf{P}^{(j-1)}(N) \mathbf{v} \).

Note that \( 1 - \sum_{j=1}^{m_1} P(AL_j \mid T_1 = t, t \in [m_1 h, (m_1 + 1)h)，P_{0h}(2) = i) \) is the probability that no alarm is signaled by the Bayesian control chart before the system fails between \( m_1h \) and \( (m_1 + 1)h \). \( \square \)

Similarly, the following Lemmas 2-4 can be proved. For \( 1 \leq j \leq m_1 \), \( 0 \leq i \leq l \) and \( m_1 \geq 0 \), let

\( P(AL_j \mid T_1 = t, t \in [m_1 h, (m_1 + 1)h)，1 \rightarrow 2, T_2 = s, t + s \in [m_1 h, (m_1 + 1)h)，Q_0 = i) \) denote the probability that an alarm first occurs on the \( j \)th sample given that \( Q_0 = i \) and the process transfers from state 1 to state 2 and then shifts to state 3 in the \( m_1^{th} \) sampling interval.
Lemma 2. For $1 \leq j \leq m_1$, $0 \leq i \leq l$ and $m_i \geq 0$,

$$P(AL_j | T_1 = t, t \in [m_i h, (m_i + 1) h], 1 \rightarrow 2, T_2 = s, t + s \in [m_i h, (m_i + 1) h], Q_0 = i)$$

is equal to

$$J(i) \cdot P^{(j-1)}(N)v$$

where $v = (I - P(N))1$.

Note that $1 - \sum_{j=1}^{m_i} P(AL_j | T_1 = t, t \in [m_i h, (m_i + 1) h], 1 \rightarrow 2, T_2 = s, t + s \in [m_i h, (m_i + 1) h], Q_0 = i)$ is the probability that no alarm is signaled by the Bayesian control chart before the system fails between $m_i h$ and $(m_i + 1) h$.

For $1 \leq j \leq m_1$ and $m_i \geq 0$, let

$$P(AL_j | T_1 = t, t \in [m_i h, (m_i + 1) h], 1 \rightarrow 2, T_2 = s, t + s \in [(m_i + 1 + m_i), (m_i + 1 + m_i + 1) h], Q_0 = i)$$

denote the probability that an alarm first occurs on the $j^{th}$ sample given that $Q_0 = i$ and the process transfers from state 1 to state 2 in the $m_i^{th}$ sampling interval, then the process shifts from state 2 to state 3 between $(m_i + 1 + m_i) h$ and $(m_i + 1 + m_i + 1) h$.

Lemma 3. For $1 \leq j \leq m_1$, $0 \leq i \leq l$ and $m_i \geq 0$,

$$P(AL_j | T_1 = t, t \in [m_i h, (m_i + 1) h], 1 \rightarrow 2, T_2 = s, t + s \in [(m_i + 1 + m_i), (m_i + 1 + m_i + 1) h], Q_0 = i)$$

equals to $J(i) \cdot P^{j-1}(N)v$, where $v = (I - P(N))1$.

For $m_i + 1 \leq j \leq m_i + 1 + m_i$, $m_i, m_i \geq 0$, and $0 \leq i \leq l$, let

$$P(AL_j | T_1 = t, t \in [m_i h, (m_i + 1) h], 1 \rightarrow 2, T_2 = s, t + s \in [(m_i + 1 + m_i), (m_i + 1 + m_i + 1) h], Q_0 = i)$$
denote the probability that an alarm first occurs on the $j^{th}$ sample given that $Q_0 = i$ and the process transfers from state 1 to state 2 in the $m_i^{th}$ sampling interval, then the process shifts from state 2 to state 3 between $(m_i + 1 + m_2)h$ and $(m_i + 1 + m_2 + 1)h$.

**Lemma 4.** For $m_i + 1 \leq j \leq m_i + 1 + m_2$, $m_i, m_2 \geq 0$, and $0 \leq i \leq l$,

$$P(\text{AL}_j | T_1 = t, t \in [m_i h, (m_i + 1)h]), 1 \rightarrow 2, T_2 = s, t + s \in [(m_i + 1 + m_2)h, (m_i + 1 + m_2 + 1)h], Q_0 = i)$$

is equal to $J(i) \cdot P^{(m_i)}(N) \cdot P^{(j-m_i-1)}(W) \cdot w$, where $w = (I - P(W))1$.

Note that $1 - \sum_{j=1}^{m_i} P_{1 \rightarrow 2} (\text{AL}_j | m_1, m_2) - \sum_{j=m_i+1}^{m_i+1+m_2} P_{1 \rightarrow 2} (\text{AL}_j | m_1, m_2)$ is the probability that no alarm is signaled by the Bayesian control chart before the system fails which occurs between $(m_i + 1 + m_2)h$ and $(m_i + 1 + m_2 + 1)h$.

### 4.6. Closed-form Formulas for Expected Cycle Length and Expected Cycle Cost

With the lemmas proved in the previous section and the results for the probabilities of false alarms included in Appendix A, the next step is to develop the complete closed forms of the expected cycle length and cycle cost. Assuming that $P_0(2) = 0$, the initial state of the Markov chain $\{Q_m\}$ is state 0. Let $J$ denote $J(0)$.

#### 4.6.1 Expected Cycle Length

Substituting the results shown in Appendix A to the equation (4.19), we obtain
\[
CL_{3 \rightarrow 3}(m_i) + CL_{3 \rightarrow 3}(m_i) = p_{13} e^{-\eta m_i h} \left\{ (1 - e^{-\eta h}) h J \left[ (I - P(N))^{-2} (I - P^m(N)) - (I - P(N))^{-1} m_i P^m(N) \right] \right\} \\
+ \left\{ \left[ h J \left[ (I - P(N))^{-2} (I - P^m(N)) - (I - P(N))^{-1} m_i P^m(N) \right] \right]_{\tau_2}^{\tau_2} + [1 - J \left( I - P^{(m)}(N) \right) (I - P(N))^{-1} v] \right\} P_{12} (1 - e^{-\eta h}) e^{-\eta m_i h} \\
+ \left\{ \left[ h J \left[ (I - P(N))^{-2} (I - P^m(N)) - (I - P(N))^{-1} m_i P^m(N) \right] \right]_{\tau_2}^{\tau_2} + [1 - J \left( I - P^{(m)}(N) \right) (I - P(N))^{-1} v] \right\} P_{12} \eta \rho \left( e^{-\eta \tau_{12}} - 1 \right) e^{-\eta m_i h} \\
+ \left\{ \left[ h J \left[ (I - P(N))^{-2} (I - P^m(N)) - m_i (I - P(N))^{-1} P^m(N) \right] \right]_{\tau_2}^{\tau_2} + \left[ h J \left[ (I - P(N))^{-2} (I - P^m(N)) - (I - P(N))^{-1} m_i P^m(N) \right] \right]_{\tau_2}^{\tau_2} \right\} P_{12} e^{-\eta h} \left( e^{-\eta \tau_{12}} - 1 \right) e^{-\eta m_i h}.
\]

Thus, the final formula for \( E(CL) \) is

\[
+ \left\{ \left[ h J \left[ (I - P(N))^{-2} (I - P^m(N)) - (I - P(N))^{-1} m_i P^m(N) \right] \right]_{\tau_2}^{\tau_2} + [1 - J \left( I - P^{(m)}(N) \right) (I - P(N))^{-1} v] \right\} P_{12} \eta \rho \left( e^{-\eta \tau_{12}} - 1 \right) e^{-\eta m_i h}.
\]
\[ E(CL) = \sum_{m_i=0}^{\infty} [CL_{1\rightarrow2}(m_i) + CL_{1\rightarrow3}(m_i)] \]

\[ = hJ \bullet (I - P(N))^{-2}v - (1 - e^{-\gamma_h})hJ \bullet (I - P(N))^{-2}(I - e^{-\gamma_h}P(N))^{-1}v \]

\[ - (1 - e^{-\gamma_h})e^{-\gamma_h}hJ \bullet (I - P(N))^{-1}(I - e^{-\gamma_h}P(N))^{-2}P(N)v \]

\[ + p_{12} \frac{1}{\gamma_s}(1 - J \bullet (I - P(N))^{-1}v) + p_{12} \frac{1}{\gamma_s}(1 - e^{-\gamma_h})J \bullet (I - e^{-\gamma_h}P(N))^{-1}(I - P(N))^{-1}v \]

\[ + \frac{1}{\gamma_h}(1 - e^{-\gamma_h})e^{-\gamma_h}(1 - J(I - P(N))^{-1}v) + J(I - e^{-\gamma_h}P(N))^{-2}P(N)(I - P(N))^{-1}v \]

\[ + (h(1 - e^{-\gamma_h}) - h + \frac{1}{\gamma_i}(1 - e^{-\gamma_h}))J(I - e^{-\gamma_h}P(N))^{-1}(I - P(N))^{-1}v \]

\[ - \frac{1}{\gamma_i} \frac{1}{\gamma_i} p_{12} e^{-\gamma_i} (e^{-\gamma_i - \gamma_i})^{-1} - 1 \]

\[ \begin{align*}
& hJ(I - e^{-\gamma_h}P(N))^{-1} \left\{ (I - P(W))^{-2} \\
& - (1 - e^{-\gamma_h})(I - P(W))^{-2}(I - e^{-\gamma_h}P(W))^{-1}P(W) \right\} \bullet w \\
& - (1 - e^{-\gamma_h})(I - e^{-\gamma_h}P(W))^{-2}P(W)(I - P(W))^{-1}w \\
& - (\frac{1}{\gamma_i} + h)J \bullet (I - e^{-\gamma_h}P(N))^{-1}(I - P(W))^{-1}w \\
& \cdot \left\{ (\frac{1}{\gamma_i} + 2h)(1 - e^{-\gamma_h}) - h \right\} J \bullet (I - e^{-\gamma_h}P(N))^{-1}(I - e^{-\gamma_h}P(W))^{-1}P(W)(I - P(W))^{-1}w \\
& + e^{-\gamma_h}hJ(I - e^{-\gamma_h}P(N))^{-2}P(N)(I - e^{-\gamma_h}P(W))^{-1}P(W)(I - P(W))^{-1}w \\
& - e^{-\gamma_h}hJ \bullet (I - e^{-\gamma_h}P(N))^{-2}P(N)(I - P(W))^{-1}w \\
& + h(1 - e^{-\gamma_h})e^{-\gamma_h}J \bullet (I - e^{-\gamma_h}P(N))^{-1}(I - e^{-\gamma_h}P(W))^{-2}P^2(W)(I - P(W))^{-1}w \\
& + (1 - e^{-\gamma_h})(T_{pm} + T_{\hat{\beta} - T_{\hat{f}}})J \bullet (I - e^{-\gamma_h}P(N))^{-1} \left\{ (I - e^{-\gamma_h})^{-1}I \\
& - (I - e^{-\gamma_h}P(W))^{-1}P(W) \right\} (I - P(W))^{-1}w \\
& + (T_{\hat{\beta} - T_{\hat{f}}})J \bullet ((I - e^{-\gamma_h})^{-1}I - (I - e^{-\gamma_h}P(N))^{-1}(I - P(N))^{-1}v + (1 - e^{-\gamma_h})^{-1}T_f \right. \\
& + (1 - e^{-\gamma_h}) + p_{12} \frac{\gamma_i}{\gamma_i - \gamma_s} e^{-\gamma_i} (e^{-\gamma_i - \gamma_i})^{-1} - 1 \] 
\end{align*} \]

(4.29)

4.6.2 Expected Cycle Cost

Using the results shown in Appendix A, we obtain the closed form formula for E(Cost) shown as follows:
\[ E(\text{Cost}) = \sum_{m=0}^{\infty} [\text{Cost}_{1 \rightarrow 2}(m_1) + \text{Cost}_{1 \rightarrow 3}(m_1)] = \text{Cost1} + \text{Cost2} + \text{Cost3} + \text{Cost4} + \frac{\eta}{v_{1-v_2}} p_{12} e^{-v_h (e^{-v_{1-v_2} h}(1-(v_{1-v_2} h))) - 1} (\text{Cost51} + \text{Cost52} + \text{Cost53} + \text{Cost54} + \text{Cost55}) \]

where

\[ \text{Cost1} = (C_s + hC_i) \left[ (1-e^{-\eta h}) + p_{12} \frac{\eta}{v_{1-v_2}} e^{-v_h (e^{-v_{1-v_2} h}(1-(v_{1-v_2} h)))} \right] \cdot \left\{ (I - P(N))^{-2} (I - e^{-\eta h})^{-1} I - [I - e^{-\eta h} P(N)]^{-1} \right\} v, \]

\[ \text{Cost2} = C_{fa} [(1-e^{-\eta h}) + p_{12} \frac{\eta}{v_{1-v_2}} e^{-v_h (e^{-v_{1-v_2} h}(1-(v_{1-v_2} h))) - 1}] J_s \left( (I - e^{-\eta h})^{-1} I - [I - e^{-\eta h} P(N)]^{-1} \right) [I - P(N)]^{-1} v, \]

\[ \text{Cost3} = (1-e^{-\eta h}) C_f + C_1 \left( \frac{\eta}{v_1} (1-e^{-\eta h}) h + \frac{1}{v_2} p_{12} (1-e^{-\eta h}) C_2 \right) + \frac{\eta}{v_{1-v_2}} p_{12} e^{-v_h (e^{-v_{1-v_2} h}(1-(v_{1-v_2} h))) - 1} (C_f + C_2 \frac{1}{v_2} (1-e^{-v_{1-v_2} h}(1-(v_{1-v_2} h))) + h)(C_2 - C_1), \]

\[ \text{Cost4} = \left\{ (1-e^{-\eta h}) C_s + p_{12} \frac{\eta}{v_{1-v_2}} e^{-v_h (e^{-v_{1-v_2} h}(1-(v_{1-v_2} h))) - 1} C_s \right\} + C_i h \left\{ (1-e^{-\eta h}) + p_{12} \frac{\eta}{v_{1-v_2}} e^{-v_h (e^{-v_{1-v_2} h}(1-(v_{1-v_2} h))) - 1} \right\} \cdot e^{-\eta h} [(1-e^{-\eta h})^{-2} - J_s (1-e^{-\eta h})^{-2} I - [I - e^{-\eta h} P(N)]^{-2} P(N)] [I - P(N)]^{-1} v, \]

\[ \text{Cost51} = -(C_s + hC_i) J_s \left\{ (I - P(N))^{-2} [ (I - e^{-\eta h})^{-1} I - [I - e^{-\eta h} P(N)]^{-1} ] \right\} v - (hC_2 + C_s) J_s \left\{ (I - e^{-\eta h} P(N))^{-1} P(N) [I - P(W)]^{-2} 
\right. + e^{-\eta h} [I - e^{-\eta h} P(N)]^{-2} P(N) [I - P(W)]^{-1} \right\} w, \]

\[ - C_f J_s ((1-e^{-\eta h})^{-1} I - [I - e^{-\eta h} P(N)]^{-1}) [I - P(N)]^{-1} v \]

\[ -(C_f + C_p) J_s [I - e^{-\eta h} P(N)]^{-1} [I - P(W)]^{-1} w \]
\[ \text{Cost52} \]
\[
(1 - e^{-\gamma_B}) \left\{ (h C_s + C_f) J \cdot \left[ I - e^{-\gamma_B} P(N) \right]^{-1} \right. \\
\left. - \left[ I - P(W) \right]^{-2} \left[ I - e^{-\gamma_B} P(W) \right]^{-1} P(W) \right\} \left[ (1 - P(W))^{-1} \left[ I - e^{-\gamma_B} P(W) \right]^{-2} P(W) \right] \cdot w \\
+ h C_s e^{-\gamma_B} J \cdot \left[ I - e^{-\gamma_B} P(N) \right]^{-2} P(N) \\
\left. + (C_{fa} + C_{p}) J \cdot \left[ I - e^{-\gamma_B} P(N) \right]^{-2} \left[ I - e^{-\gamma_B} P(W) \right]^{-1} P(W) \left[ I - P(W) \right]^{-1} w \right\} \\
= (1 - e^{-\gamma_B}) \left[ I - J \cdot \left[ I - P(N) \right]^{-1} v \right] (-C_s h - C_f) \\
+ (1 - e^{-\gamma_B})^{-1} \left[ I - J \cdot \left[ I - P(N) \right]^{-1} v \right] \left\{ -C_s - C_f \right\} \\
+ e^{-\gamma_B} J [I - e^{-\gamma_B} P(N)]^{-2} P(N) [I - P(N)]^{-1} v - [I - P(W)]^{-1} w \left[ -C_s h - C_f \right] \\
+ J [I - e^{-\gamma_B} P(N)]^{-1} [I - P(N)]^{-1} v - [I - P(W)]^{-1} w \\
\left. \cdot \left\{ (C_s + C_f) - C_s \left[ \frac{1}{\gamma_B} (1 - e^{-\gamma_B})^{-1} + 2h - h(1 - e^{-\gamma_B})^{-1} \right] \right\} \right\}
\]

\[ \text{Cost53} \]
\[
= e^{-\gamma_B} (1 - e^{-\gamma_B})^{-2} \left[ I - J \cdot \left[ I - P(N) \right]^{-1} v \right] (-C_s h - C_f) \\
+ (1 - e^{-\gamma_B})^{-1} \left[ I - J \cdot \left[ I - P(N) \right]^{-1} v \right] \left\{ -C_s - C_f \right\} \\
+ e^{-\gamma_B} J [I - e^{-\gamma_B} P(N)]^{-2} P(N) [I - P(N)]^{-1} v - [I - P(W)]^{-1} w \left[ -C_s h - C_f \right] \\
+ J [I - e^{-\gamma_B} P(N)]^{-1} [I - P(N)]^{-1} v - [I - P(W)]^{-1} w \\
\left. \cdot \left\{ (C_s + C_f) - C_s \left[ \frac{1}{\gamma_B} (1 - e^{-\gamma_B})^{-1} + 2h - h(1 - e^{-\gamma_B})^{-1} \right] \right\} \right\}
\]

\[ \text{Cost54} \]
\[
= -C_s (1 - e^{-\gamma_B}) h \frac{e^{-\gamma_B}}{\gamma_B} \left[ J \cdot [I - e^{-\gamma_B} P(N)]^{-2} P(N) [I - e^{-\gamma_B} P(W)]^{-1} P(W) [I - P(W)]^{-1} w \right] \\
\left[ J \cdot [I - e^{-\gamma_B} P(N)]^{-1} [I - e^{-\gamma_B} P(W)]^{-1} P(W) [I - P(W)]^{-1} w \right] \left\{ C_s \left[ (\frac{1}{\gamma_B} + 2h)(1 - e^{-\gamma_B}) - h \right] \right\} \\
+ (1 - e^{-\gamma_B})(C_s + C_f) \\
\]

and

\[ \text{Cost55} \]
\[
= -e^{-\gamma_B} (1 - e^{-\gamma_B})^{-1} (C_s h + C_f) \left\{ (1 - e^{-\gamma_B})^{-1} - J \cdot ((1 - e^{-\gamma_B})^{-1} I - [I - e^{-\gamma_B} P(N)]^{-1} [I - P(N)]^{-1} v \right. \\
\left. -J \cdot [I - e^{-\gamma_B} P(N)]^{-1} [I - P(W)]^{-1} w \right\} \\
- (1 - e^{-\gamma_B}) e^{-\gamma_B} (C_s h + C_f) J \cdot [I - e^{-\gamma_B} P(N)]^{-1} [I - e^{-\gamma_B} P(W)]^{-2} P(W) [I - P(W)]^{-1} w \\
- (C_s - C_f) e^{-\gamma_B} [(1 - e^{-\gamma_B})^{-2} - J \cdot ((1 - e^{-\gamma_B})^{-2} I - [I - e^{-\gamma_B} P(N)]^{-2} P(N) [I - P(N)]^{-1} v \right. \\
\left. -J \cdot [I - e^{-\gamma_B} P(N)]^{-1} [I - P(W)]^{-1} w \right\} \\
\left. + e^{-\gamma_B} (C_s - C_f) \left[ \frac{1}{\gamma_B} + h \left( e^{-\gamma_B} - 1 \right) \right] \right\} \\
\left\{ (1 - e^{-\gamma_B})^{-1} - J \cdot ((1 - e^{-\gamma_B})^{-1} I - [I - e^{-\gamma_B} P(N)]^{-1} [I - P(N)]^{-1} v \right. \\
\left. -J \cdot [I - e^{-\gamma_B} P(N)]^{-1} [I - P(W)]^{-1} w \right\} \\
\]

4.6.3 Constraint for Economic Statistical Design
Applying the results in Appendix A to the equations (4.20) and (4.21), we have

\[ \int_{m_3}^{(m_3+1)h} \{ P(\text{Fail} | T_1 = t, 1 \to 2) p_{13} v_i e^{-\gamma d t} \} \]

\[ = \left[ (1 - e^{-\gamma h}) + \frac{v_i}{\eta - \gamma} e^{-\gamma h} (e^{-(\eta - \gamma)h} - 1) \right] p_{12} e^{-\gamma m_3 h} \left[ 1 - J \bullet (I - P^{(m_3)}(N))(I - P(N))^{-1} v \right] \]

\[ - \frac{v_i}{\eta - \gamma} p_{12} e^{-\gamma h} (e^{-(\eta - \gamma)h} - 1)(1 - e^{-\gamma h}) e^{-\gamma m_3 h} \]

\[ \left\{ (1 - e^{-\gamma h})^{-1} \left[ 1 - J \bullet (I - P^{(m_3)}(N))(I - P(N))^{-1} v - J \bullet P^{(m_3)}(N)[I - P(W)]^{-1} w \right] \right\} \]

\[ \left\{ J \bullet P^{(m_3)}(N)[I - (1 - B_2)P(W)]^{-1} P(W)[I - P(W)]^{-1} w \right\} \]

\[ \text{(4.31)} \]

and

\[ \int_{m_3}^{(m_3+1)h} \{ P(\text{Fail} | T_1 = t, 1 \to 3) p_{13} v_i e^{-\gamma d t} \} \]

\[ = p_{13} e^{-\gamma m_3 h} (1 - e^{-\gamma h}) \left[ 1 - J \bullet (I - P^{(m_3)}(N))(I - P(N))^{-1} v \right] \]

\[ \text{(4.32)} \]

Hence,

\[ P(\text{Fail}) \]

\[ = \int_0^{\infty} P(\text{Fail} | T_1 = t) v_i e^{-\gamma d t} dt \]

\[ = \int_0^{\infty} P(\text{Fail} | T_1 = t, 1 \to 2) p_{13} v_i e^{-\gamma d t} dt + \int_0^{\infty} P(\text{Fail} | T_1 = t, 1 \to 3) p_{13} v_i e^{-\gamma d t} dt \]

\[ = \sum_{m_3=0}^{\infty} \left\{ \left[ (1 - e^{-\gamma h}) + \frac{v_i}{\eta - \gamma} e^{-\gamma h} (e^{-(\eta - \gamma)h} - 1) \right] p_{12} e^{-\gamma m_3 h} \left[ 1 - J \bullet (I - P^{(m_3)}(N))(I - P(N))^{-1} v \right] \right\} \]

\[ - \frac{v_i}{\eta - \gamma} p_{12} e^{-\gamma h} (e^{-(\eta - \gamma)h} - 1)(1 - e^{-\gamma h}) e^{-\gamma m_3 h} \]

\[ \left\{ (1 - e^{-\gamma h})^{-1} \left[ 1 - J \bullet (I - P^{(m_3)}(N))(I - P(N))^{-1} v - J \bullet P^{(m_3)}(N)[I - P(W)]^{-1} w \right] \right\} \]

\[ \left\{ J \bullet P^{(m_3)}(N)[I - (1 - B_2)P(W)]^{-1} P(W)[I - P(W)]^{-1} w \right\} \]

\[ \text{(4.33)} \]
With further simplification,

\[
P(Fail) = 1 - J\star(I - (1 - e^{-\gamma h})[I - e^{-v y h}P(N)]^{-1})[I - P(N)]^{-1}v
\]

\[
-\frac{v}{v - \gamma} p_{12} e^{-\gamma h}(e^{-(v - \gamma)h} - 1)(1 - e^{-v y h})\left[ J\star[I - e^{-v y h}P(N)]^{-1}[I - e^{-v y h}P(W)]^{-1}[P(W)][I - P(W)]^{-1}w
\right]
\]

By using the similar derivation and the results in Appendix A, the formula for \(P(\text{real alarm})\) is shown as follows:

\[
P(\text{Real Alarm})
\]

\[
= \sum_{m_i=0}^{+\infty} \left\{ P(\text{Real Alarm}|T_i = t, 1 \rightarrow 2) p_{12} v_i e^{-v y t} dt \right\}
\]

\[
= \frac{v}{v - \gamma} p_{12} e^{-\gamma h}(1 - e^{-(v - \gamma)h}) J\star[I - e^{-v y h}P(N)]^{-1}\left\{ I - (1 - e^{-v y h})[I - e^{-v y h}P(W)]^{-1}[P(W)] [I - P(W)]^{-1}w
\right\}
\]

4.6.4 Validation of the Derived formulas using Monte Carlo Simulation

To verify the validity of the formulas of the expected cycle length and cost, a Monte Carlo simulation procedure written in MATLAB was run for four cases. The four cases used arbitrary parameters \(n, h,\) and \(P^*\) shown in Tables 4.3-4.4. In all cases, except Case 2, \(h=1\) and \(n=1\) are
selected. The other system parameters are: \(d^2 = 12.2081\), \(\lambda_1 = 0.1; \lambda_2 = 0.15; \lambda_3 = 0.01; C_1 = 1; C_2 = 2; C_{ja} = 20; C_p = 30; C_f = 150; C_s = 5; T_{pm} = 2; T_{fi} = 1\); and \(T_f = 5\) for all cases. Let \(\text{Sim}_n\) denote the total number of simulation runs. Let \(\text{Sim}_{\text{StErr}}\) denote the standard error of the simulation average and is computed as

\[
\text{Sim}_{\text{StErr}} = \sqrt{\frac{1}{\text{Sim}_n (\text{Sim}_n - 1)} \sum_{i=1}^{\text{Sim}_n} (X_i - \bar{X})^2} \quad (4.36)
\]

where \(X_i\) is the result from the \(i^{th}\) simulation run and \(\bar{X}\) is the average of all simulation results.

From Table 4.2, \(\text{Sim}_n\) is chosen to be 50,000 in this research to obtain fairly small simulation standard error which is less than 0.5% of the average of the simulation results for Case 1.

Table 4.2 Simulation standard errors for selected simulation runs for Case 1

<table>
<thead>
<tr>
<th>(n)</th>
<th>(\text{ECL}_\text{Sim})</th>
<th>(\text{Sim}_{\text{StErr}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,000</td>
<td>11.8906</td>
<td>0.1185</td>
</tr>
<tr>
<td>10,000</td>
<td>11.8470</td>
<td>0.0828</td>
</tr>
<tr>
<td>50,000</td>
<td>11.9045</td>
<td>0.0370</td>
</tr>
<tr>
<td>100,000</td>
<td>11.8948</td>
<td>0.0262</td>
</tr>
</tbody>
</table>

Table 4.3 Comparison of ECL between the cost model and simulation method

<table>
<thead>
<tr>
<th>Case</th>
<th>(p^*)</th>
<th>(h)</th>
<th>(n)</th>
<th>(\text{ECL}_\text{Sim})</th>
<th>(\text{Sim}_{\text{StErr}})</th>
<th>(\text{ECL})</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>11.9045</td>
<td>0.0370</td>
<td>11.8888</td>
<td>-0.13</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
<td>1</td>
<td>2</td>
<td>12.6777</td>
<td>0.0399</td>
<td>12.6495</td>
<td>-0.22</td>
</tr>
<tr>
<td>3</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>12.4400</td>
<td>0.0390</td>
<td>12.4098</td>
<td>-0.24</td>
</tr>
<tr>
<td>4</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>12.5953</td>
<td>0.0388</td>
<td>12.6211</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 4.4 Comparison of ECost between the cost model and simulation method

<table>
<thead>
<tr>
<th>Case</th>
<th>(p^*)</th>
<th>(h)</th>
<th>(n)</th>
<th>(\text{ECost}_\text{Sim})</th>
<th>(\text{Sim}_{\text{StErr}})</th>
<th>(\text{ECost})</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>115.7159</td>
<td>0.2742</td>
<td>115.4591</td>
<td>-0.22</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
<td>1</td>
<td>2</td>
<td>121.5719</td>
<td>0.2855</td>
<td>121.3328</td>
<td>-0.20</td>
</tr>
<tr>
<td>3</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>120.6759</td>
<td>0.2846</td>
<td>120.4553</td>
<td>-0.18</td>
</tr>
<tr>
<td>4</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>122.3101</td>
<td>0.2851</td>
<td>122.6030</td>
<td>0.24</td>
</tr>
</tbody>
</table>

In Tables 4.3-4.4, the following additional notation is used: \(\text{ECL}_\text{Sim}\) is the average of the simulated cycle lengths from all the simulation runs; \(\text{ECost}_\text{Sim}\) is the average of all simulated
cycle lengths; and Difference (%) is the ratio of the difference between the model and simulation to the model result.

From the previous tables, we can see that the difference of the results between the mathematical model and the simulation model for both the expected cycle length and cycle cost is less than 0.24%. Therefore, the formulas of the expected cycle length and the expected cycle cost are verified by simulation results.

4.6.5 Optimization Models for Economic and Economic Statistical Design

In the above sections, the formulas for the expected cycle cost and the expected cycle length are expressed in a closed form and verified by simulation. The optimization model in (4.4) for the economic design of the multivariate Bayesian control chart becomes

\[
\min z(h, P^*) = \min \frac{\text{Equation(4.30)}}{\text{Equation(4.29)}} \tag{4.37}
\]

s.t. \(0 \leq P^* \leq 1\)

\(h > 0\)

The objective function formulation of the optimization model for the economic statistical design of the Bayesian chart for CBM is the same as economic design for the Bayesian chart. To reduce the rate of false alarms and system failures, a constraint of \(P(\text{Real Alarm})\) is added to the optimization model. Thus, the model for the economic statistical design of the Bayesian chart for CBM application is shown as follows:

\[
\min z(h, P^*) = \min \frac{\text{Equation(4.30)}}{\text{Equation(4.29)}} \tag{4.38}
\]
\[ s.t. \ 0 \leq P^* \leq 1 \\
\quad P(\text{Real Alarm}) \geq b \\
\quad h > 0 \]

where \( b \) represents the lower bound of \( P(\text{Real Alarm}) \) and can be any real number between 0 and 1 determined by maintenance specialists.

### 4.7 Computational Results

In this section, a simple numerical example previously used in Montgomery and Klatt (1972) and Wu and Makis (2007) is considered. The results for the performance comparison between the Bayesian control chart and traditional Chi-square control chart are given.

Consider a system described by a two dimensional observation vector, which is \( N_2(\mu, \Sigma) \) when the system is in normal state and the mean vector shifts to \( \mu_2 \) with the same covariance matrix when the system is in the warning state. The mean vectors and the covariance matrix are given as

\[
\mu_1 = [0,0]^T, \quad \mu_2 = [5,6]^T, \quad \Sigma = \begin{bmatrix} 2.0 & 1.0 \\ 1.0 & 2.5 \end{bmatrix}.
\]

The transition rates are \( \lambda_1 = 0.1, \ \lambda_2 = 0.15 \) and \( \lambda_3 = 0.01 \) for the system deterioration process. Sample size is one. The operating and regular maintenance cost \( C_1 \) and \( C_2 \) are assumed to be zero. The other cost parameters are \( C_{fa} = 20; \ C_p = 30; \ C_f = 150 \) and \( C_s = 5 \). To be able to compare the minimum expected costs, it is assumed that \( T_{fi} = T_{pm} = T_f = 0 \).
4.7.1 Numerical Results for Economic Design

The model in (4.37) is a mixed nonlinear constrained optimization problem. MATLAB programs were written to solve this problem by using similar techniques introduced in previous chapter. The results of economic design for the Bayesian control chart control limit are shown in Table 4.5. The results for the Chi-square CBM chart are from Wu and Makis (2007).

<table>
<thead>
<tr>
<th>Chart</th>
<th>Control limit</th>
<th>$h$</th>
<th>ECost</th>
<th>P(Fail)</th>
<th>P(Real Alarm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian</td>
<td>0.4516</td>
<td>5.1868</td>
<td>8.9483</td>
<td>0.3924</td>
<td>0.5933</td>
</tr>
<tr>
<td>Chi-square</td>
<td>4.7934</td>
<td>12.5232</td>
<td>9.6125</td>
<td>-</td>
<td>0.3291</td>
</tr>
</tbody>
</table>

The optimal Bayesian chart parameters, control limit and sampling interval, obtained by the MATLAB programs are 0.4516 and 5.1867 respectively. The optimal sampling interval for the Bayesian chart is smaller than the interval for the Chi-square chart. The optimal expected average cost is 8.9483. The minimum expected average cost monitor the system by using a Chi-square chart is 9.6125. The saving from use of the Bayesian chart over the Chi-square chart is more than 6.9 percent. From Table 4.5, the probability to detect that system shifts from a normal state to a warning state is 0.5933 for the Bayesian chart and 0.3291 for the Chi-square chart. The chance to detect a shift is almost doubled when using the economically designed Bayesian chart as compared to the Chi-square chart. It indicates that Bayesian chart also has much better statistical performance than the Chi-square chart.

4.7.2 Numerical Results for Economic Statistical Design

Four experiments for different values of the lower bound $b$ were run by MATLAB programs. The comparison results of the optimal expected average cost per unit time for the Bayesian chart and the Chi-square chart are shown in Table 4.6.
Table 4.6 Results for economic statistical design of Bayesian control chart

<table>
<thead>
<tr>
<th>b</th>
<th>Bayesian Chart</th>
<th></th>
<th></th>
<th>Chi-square Chart</th>
<th></th>
<th></th>
<th>Diff(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control limit</td>
<td>$h$</td>
<td>ECost</td>
<td>P(Fail)</td>
<td>P(Real Alarm)</td>
<td>Control limit</td>
<td>$h$</td>
</tr>
<tr>
<td>Optimal</td>
<td>0.4516</td>
<td>5.1868</td>
<td>8.9483</td>
<td>0.3924</td>
<td>0.5933</td>
<td>4.7934</td>
<td>12.5232</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4516</td>
<td>5.1868</td>
<td>8.9483</td>
<td>0.3924</td>
<td>0.5933</td>
<td>3.8752</td>
<td>7.5842</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4522</td>
<td>5.0425</td>
<td>8.9489</td>
<td>0.3855</td>
<td>0.6000</td>
<td>3.3080</td>
<td>5.3464</td>
</tr>
<tr>
<td>0.7</td>
<td>0.6049</td>
<td>3.0561</td>
<td>9.1657</td>
<td>0.2863</td>
<td>0.7000</td>
<td>3.0099</td>
<td>3.4021</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8011</td>
<td>1.3410</td>
<td>10.7795</td>
<td>0.1898</td>
<td>0.8000</td>
<td>3.0900</td>
<td>1.6788</td>
</tr>
</tbody>
</table>

Figure 4.4 Results for economic and economic statistical design

Generally speaking, a penalty cost is expected to be added to the optimal cost of a pure economic design of a control chart if the statistical requirement is stricter than the level the chart with optimal parameters can reach. Table 4.6 shows that the cost is increased to 10.7795 from the optimal cost 8.948 by using the Bayesian chart to monitor the system with better statistical
performance. However, if the optimal solution of the pure economic design satisfies the constraint of the statistical requirement, no penalty cost will be added. When \( b \) equals to 0.5 and the value of \( P(\text{Real Alarm}) \) with the optimal chart setting is 0.5933, the optimal solution of the economic statistical design is unchanged.

For the Chi-square chart, in all cases the control limits and sampling intervals decrease, and expected average costs increase with the increased values of the lower bound \( b \) for the statistical constraint. This result is not surprising and matches our intuition. The control limit should be lower and samples should be collected more frequently if the chart is required to be more sensitive to the system changes. The average cost would increase due to higher sampling and alarm inspection costs. However, for the Bayesian chart, Table 4.6 shows higher control limits when the value of \( b \) increases. To reduce failures, the process is monitored more frequently but because the posterior probability is the sufficient statistic for decision-making, it tracks the process evolution very well and higher values likely indicate that the system is in the warning state. Thus, more frequent process monitoring enables to increase the control limit when the value of \( b \) increases, which reduces the false alarm cost. Because the statistic of the Chi-square chart is less efficient and cannot track the process evolution as well as the posterior probability, both sampling interval and the control limit must be decreased when the value of \( b \) increases. Table 4.6 shows that the Bayesian control chart has significantly smaller penalty cost than the Chi-square control chart. When the lower bound \( b \) increases from 0.6 to 0.8, the average maintenance cost associated with the Chi-square chart climbs rapidly from 10.2796 to 14.875, while the cost associated with the Bayesian chart increases only from 8.9489 to 10.7795. The savings from using the Bayesian chart over Chi-square chart range from 6.9 percent to 27.5
percent. This can be seen in Fig. 4.4. In summary, not only does the Bayesian chart perform better than the traditional Chi-square chart when there are no statistical requirements, but its economic performance is compromised much less than the performance of the Chi-square chart when the statistical constraint is imposed.
Chapter 5 Conclusions

The main focus of this research is to design a multivariate Bayesian control chart for both quality control and CBM applications. The class of control limit policies proved to be optimal by Makis (2006a, 2007) is considered. Also, a VARMAX model with a SPE chart has been proposed to monitor autocorrelated multivariate processes. This chapter is a discussion of the conclusions obtained from Chapters 2, 3, and 4. Suggestions for future research are included as well.

5.1 Summary

A list of conclusions and contributions in this work is summarized below.

In Chapter 2, a VARMAX model was considered for monitoring a multivariate autocorrelated process. A VARMAX (3, 3, 1) model was selected using the BIC criterion to represent the activated sludge process simulated by the ASWWTP - USP simulator. The performance of the VARMAX model for the identification of the activated sludge process was compared with the PLS model and the subspace-based models presented by Sotomayor et al. (2003). Using the two performance indicators, MRSE and MVAF, the performance of the VARMAX model is much better than that of the PLS model and also considerably better than the best choice of the subspace-based models presented by Sotomayor et al. (2003). It was also shown that the SPE
chart based on the VARMAX model has a lower false alarm rate.

In Chapter 3, a design methodology for designing a multivariate Bayesian control chart based on the economic criteria was discussed for quality control applications. The optimization model for the economically-designed multivariate Bayesian control chart to find the optimal control chart parameters was developed. A Markov chain approach was used to derive the closed form of the expected average cost based on renewal theory. Statistical constraints were added for the economic statistical design of the chart. Two sets of examples with different levels of the cost and process parameters were carried out. It was observed that for all of the examples, the expected cost per time unit for the Bayesian chart is lower than for the MEWMA chart. The cost savings range from 1.961% to 9.736%. The cost penalty for the Bayesian chart with additional statistical constraints for ARLs is less than 0.02%. It indicates that the Bayesian chart designed economically has a good statistical performance. We also studied the effect on the expected average cost of the misspecified directions of out-of-control process mean vector and misspecified shift sizes. The numerical results show that there is a small cost penalty when the ratio of $d_{12}/d_i^2$ is close to 1 and the cost penalty increases when the ratio decreases given that the shift size is correctly estimated. The sensitivity analysis of misspecified shift sizes shows that overestimated shift sizes may cause large cost penalties and underestimated shift sizes do not impact the average cost much in most cases.
In Chapter 4, a multivariate Bayesian control chart was designed for CBM applications by considering the simple control limit policy structure and including an observable failure state. A methodology for economic and economic statistical design of the control chart was presented to determine the optimal control chart parameters, namely the control limit and the sampling interval. The probability that a cycle is ended by a real alarm was calculated and used to measure the statistical performance of the Bayesian chart. A constraint on this probability was added to the pure economic design of the Bayesian chart to formulate the model for the economic statistical design of the chart. Bayesian chart performs better than the traditional Chi-square chart when there is no statistical requirement. The saving from use of the Bayesian chart over the Chi-square chart is about 6.9 percent in the numerical example. Four more experiments for different values of the lower bound for the statistical constraint were implemented. The results show that the control limits and expected average costs increase and sampling intervals decrease with the increasing value of the lower bound for all cases. The expected cost per time unit increases slightly over 20% if \( P(\text{Real Alarm}) \) is required to be greater than 0.8. Furthermore, the savings from use of the Bayesian chart over Chi-square chart are up to 27.5 percent. The computational results indicate that the Bayesian chart compromises its economic performance much less than the Chi-square chart when the statistical constraint becomes stricter.

5.2 Topics for Future Research

The following suggestions are made for future research in this area:
1. Throughout this research study, the focus of control charts is to design control charts to monitor the mean vector of the process under the assumption that the covariance matrix is unchanged. However, in real industrial applications, such assumption might not be satisfied. A thorough study of the robustness of the Bayesian chart when this assumption is not satisfied would be valuable.

2. The relaxation of exponentially distributed transition time for both quality control and maintenance applications can be considered. The performance of the Bayesian chart under different distributions, such as Weibull distribution, when the system failure mechanism is considered could be compared.

3. Although some sensitivity analyses have been studied in the present research, effects of the misspecified cost and other process parameters on the expected cost per unit time should be studied with respect to pure economic and economic statistical designs under a variety of scenarios.

4. Several multivariate CUSUM control charts have appeared in the literature and some of them have better statistical performance than the multivariate EWMA chart, but not much has been done to study the economic design of CUSUM charts. A performance comparison between the Bayesian chart and a CUSUM chart may be worthwhile.
Bibliography

Alwan, B. M. and Roberts, H. V. (1988). Time-series modeling for detecting level shifts of
Brook, D., and Evans, D. A. (1972). An approach to the probability distribution of CUSUM run
    length. Biometrika, 59, 539-549.
Box, G. E. P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of
    Variance Problems: Effect of Inequality of Variance in One-way Classification. The Annals
    of the American Statistical Association, 76, 376, 802-816.
    645.
    maintenance and statistical process control: a preliminary investigation. IIE Transactions, 32,
    471-478.


Wu, J. and Makis, V. (2007), Economic and economic-statistical design of a Chi-square chart for CBM. Accepted for publication in European Journal of Operational Research.


Appendix A

From Lemma 1-4 in Chapter 3, the following formulas can be derived:

\[ \sum_{i=1}^{m} P_{i\rightarrow 3}(AL_1|m_i) \]
\[ = \sum_{i=1}^{m} J \cdot P^{(i-1)}(N)v \]
\[ = J \left[ \sum_{i=1}^{m} P^{(i-1)}(N) \right]v \]
\[ = J \cdot P^{(-1)}(N) \cdot P(N) \cdot (I - P^{(m)}(N))(I - P(N))^{-1}v \]
\[ = J \cdot (I - P^{(m)}(N))(I - P(N))^{-1}v \]

\[ \sum_{i=1}^{m} iP_{i\rightarrow 3}(AL_1|m_i) \]
\[ = \sum_{i=1}^{m} iJ \cdot P^{(i-1)}(N)v \]
\[ = J \cdot \sum_{i=1}^{m} iP^{(i-1)}(N)v \]
\[ = J \cdot [ (I - P(N))^{-2} (I - P^{m}(N)) - (I - P(N))^{-1} m_i P^{m}(N) ]v \]

\[ \sum_{i=1}^{m} P_{i\rightarrow 2}(AL_1|m_i, m_i) = \sum_{i=1}^{m} J \cdot P^{i-1}(N)v \]
\[ = \sum_{i=1}^{m} P_{i\rightarrow 3}(AL_1|m_i) \]
\[ = J \cdot (I - P^{(m)}(N))(I - P(N))^{-1}v \]

\[ \sum_{i=1}^{m} i \cdot P_{i\rightarrow 2}(AL_1|m_i, m_i) \]
\[ = \sum_{i=1}^{m} iJ \cdot P^{(i-1)}(N)v \]
\[ = \sum_{i=1}^{m} iP_{i\rightarrow 3}(AL_1|m_i) \]
\[ = J \cdot [ (I - P(N))^{-2} (I - P^{m}(N)) - (I - P(N))^{-1} m_i P^{m}(N) ]v \]
\[
\sum_{i=1}^{m_1} P_{i\to 2}(AL_i | m_1, m_2) = \sum_{i=1}^{m_1} P_{i\to 2}(AL_i | m_1, m_1) = \sum_{i=1}^{m_1} J\cdot P^{-i}(N)\nu = \sum_{i=1}^{m_1} P_{i\to 3}(AL_i | m_1) = J\cdot (I - P^{(m_1)}(N))(I - P(N)^{-1}\nu
\]

\[
m_{i+1}^{+m_2} P_{i\to 2}(AL_i | m_1, m_2) = \sum_{i=m_1}^{m_2} J\cdot P^{(m_1)}(N)\cdot P^{(i-m_1-1)}(W)\cdot w = \sum_{j=1}^{1+m_2} J\cdot P^{(m_1)}(N)\cdot P^{(j-1)}(W)\cdot w
\]

\[
= J\cdot P^{(m_1)}(N)[\sum_{j=1}^{1+m_2} P^{(j-1)}(W)]\cdot w = J\cdot P^{(m_1)}(N)(I - P^{(m_1+1)}(W))(I - P(W))^{-1}w
\]

\[
\sum_{i=1}^{m_1} iP_{i\to 2}(AL_i | m_1, m_2) = \sum_{i=1}^{m_1} iP_{i\to 2}(AL_i | m_1, m_1) = \sum_{i=1}^{m_1} iJ\cdot P^{-i}(N)\nu = J[(I - P(N))^{-2}(I - P^{m_1}(N)) - (I - P(N))^{-1}m_1P^{m_1}(N)]\nu
\]

\[
m_{i+1}^{+m_2} iP_{i\to 2}(AL_i | m_1, m_2) = \sum_{i=m_1}^{m_2} J\cdot iP^{(m_1)}(N)\cdot P^{(i-m_1-1)}(W)\cdot w = \sum_{j=1}^{1+m_2} J\cdot (j + m_1)P^{(m_1)}(N)\cdot P^{(j-1)}(W)\cdot w
\]

\[
= J\cdot P^{(m_1)}(N)[\sum_{j=1}^{1+m_2} jP^{(j-1)}(W) + m_1P^{(j-1)}(W)]\cdot w = J\cdot P^{(m_1)}(N) \left[ (I - P(W))^{-2}(I - P^{m_1+1}(W)) - (I - P(W))^{-1}(m_2 + 1)P^{m_2+1}(W) ight] \cdot w
\]