2008

Monitoring and diagnosis of process faults and sensor faults in manufacturing processes

Shan Li
University of Iowa

MONITORING AND DIAGNOSIS OF PROCESS FAULTS AND SENSOR FAULTS IN MANUFACTURING PROCESSES

by

Shan Li

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Industrial Engineering
in the Graduate College of
The University of Iowa

December 2008

Thesis Supervisor: Assistant Professor Yong Chen

ABSTRACT

The substantial growth in the use of automated in-process sensing technologies creates great opportunities for manufacturers to detect abnormal manufacturing processes and identify the root causes quickly. It is critical to locate and distinguish two types of faults – process faults and sensor faults. The procedures to monitor and diagnose process and sensor mean shift faults are presented with the assumption that the manufacturing processes can be modeled by a linear fault-quality model.

A *W* control chart is developed to monitor the manufacturing process and quickly detect the occurrence of the sensor faults. Since the *W* chart is insensitive to process faults, when it is combined with *U* chart, both process faults and sensor faults can be detected and distinguished. A unit-free index referred to as the sensitivity ratio (*SR*) is defined to measure the sensitivity of the *W* chart. It shows that the sensitivity of the *W* chart is affected by the potential influence of the sensor measurement.

A Bayesian variable selection based fault diagnosis approach is presented to locate the root causes of the abnormal processes. A *Minimal Coupled Pattern* (MCP) and its degree are defined to denote the coupled structure of a system. When less than half of the faults within an MCP occur, which is defined as sparse faults, the proposed fault diagnosis procedure can identify the correct root causes with high probability. Guidelines are provided for the hyperparameters selection in the Bayesian hierarchical model. An alternative CML method for hyperparameters selection is also discussed. With the large number of potential process faults and sensor faults, an MCMC method, e.g. Metropolis-Hastings algorithm can be applied to approximate the posterior probabilities of candidate models.

The monitor and diagnosis procedures are demonstrated and evaluate through an autobody assembly example.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

MONITORING AND DIAGNOSIS OF PROCESS FAULTS AND SENSOR FAULTS IN MANUFACTURING PROCESSES

by

Shan Li

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Industrial Engineering
in the Graduate College of
The University of Iowa

December 2008

Thesis Supervisor: Assistant Professor Yong Chen

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Shan Li

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Industrial Engineering at the December 2008 graduation.

Thesis Committee: _____
Yong Chen, Thesis Supervisor

_____
Richard Dykstra

_____
Pavlo Krokhmal

_____
Andrew Kusiak

_____
Osnat Stramer

To My Parents

ACKNOWLEDGMENTS

I would like to express my deepest and most sincere gratitude to Professor Yong Chen, my advisor. With his enthusiasm, inspiration, and patience, he made my Ph.D study an invaluable experience which will benefit my whole life. I feel very lucky to have him as my mentor.

I gratefully appreciate the precious opportunity I got to learn from many professors, while attending the University of Iowa. I feel especially grateful for having met Professor Richard Dykstra and Professor Osnat Stramer. They always take the time to answer questions and to offer encouragement.

This thesis could not be finished without many help from my colleagues and friends. I wish to extend my warmest thanks to all of them, especially to Yulan Liang, Shan Bao and Liang Zhao for their kindness and support.

I want to give my special thanks to my parents Wenzhong Li and Juying Gao. They bore me, raised me, supported me and love me. To them, I dedicate this thesis.

The financial support of the Department of Mechanical and Industrial Engineering is gratefully acknowledged.

ABSTRACT

The substantial growth in the use of automated in-process sensing technologies creates great opportunities for manufacturers to detect abnormal manufacturing processes and identify the root causes quickly. It is critical to locate and distinguish two types of faults – process faults and sensor faults. The procedures to monitor and diagnose process and sensor mean shift faults are presented with the assumption that the manufacturing processes can be modeled by a linear fault-quality model.

A *W* control chart is developed to monitor the manufacturing process and quickly detect the occurrence of the sensor faults. Since the *W* chart is insensitive to process faults, when it is combined with *U* chart, both process faults and sensor faults can be detected and distinguished. A unit-free index referred to as the sensitivity ratio (*SR*) is defined to measure the sensitivity of the *W* chart. It shows that the sensitivity of the *W* chart is affected by the potential influence of the sensor measurement.

A Bayesian variable selection based fault diagnosis approach is presented to locate the root causes of the abnormal processes. A *Minimal Coupled Pattern* (MCP) and its degree are defined to denote the coupled structure of a system. When less than half of the faults within an MCP occur, which is defined as sparse faults, the proposed fault diagnosis procedure can identify the correct root causes with high probability. Guidelines are provided for the hyperparameters selection in the Bayesian hierarchical model. An alternative CML method for hyperparameters selection is also discussed. With the large number of potential process faults and sensor faults, an MCMC method, e.g. Metropolis-Hastings algorithm can be applied to approximate the posterior probabilities of candidate models.

The monitor and diagnosis procedures are demonstrated and evaluate through an autobody assembly example.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

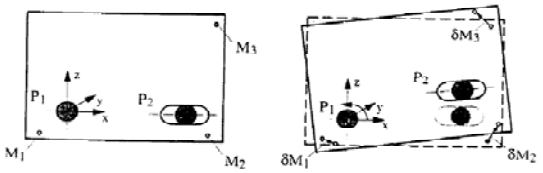
CHAPTER 1. INTRODUCTION

In recent years, there has been substantial growth in the use of automated in-process sensing and data-capture technologies such as the in-line Optical Coordinate Measurement Machine (OCMM) used in manufacturing process as automotive body assembly processes. These technologies provide 100 percent inspection of the product quality characteristics and create opportunities for manufacturers to monitor the processes and diagnose faults in real time.

Two types of root causes in manufacturing processes may cause abnormal measurements of product quality variables. One type of root cause is *process fault*, which is defined as the malfunction of tooling elements (such as fixture, cutting tool, and welding gun) in manufacturing processes. The other type of root cause is the malfunction of the sensors used to take the measurements, which are referred to as *sensor faults* occurred in the in-process sensing system. For example, Figure 1.1 demonstrates a simplified fixture layout of the flat rigid body panel in *x*-*z* plane. A four-way pin P_1 and a two way pin P_2 constrain the motion of the panel. Three points M_1, M_2 and M_3 are measured by the sensors installed in the system in both *x* and *z* directions. Figure 1.1 (a) shows the nominal position of the panel under normal condition. If a displacement occurs on pin P_2, which might be the result of tool element worn, loose or broken, there will be a motion/deviation of the panel as shown in Figure 1.1(b), resulting in the displacements of the three measurements. We call this type of root cause as process faults. Suppose there is no process fault as shown in Figure 1.1(a), we may still observe measurement of point displacements because of the malfunction of the sensing system – the occurrence of sensor faults.

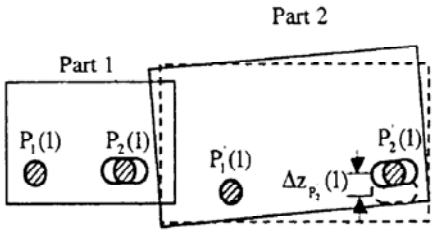
The effect of process faults and sensor faults can also be shown in machining process. Figure 1.3 demonstrates fixture locating schemes of an automotive engine head in three machining processes. The features are the cover face (M) (Figure 1.3(a)), joint

face (Figure 1.3(b)), and the slot (S) (Figure 1.3(c)). The cover face, joint face, and the slot are milled at the 1st , the 2nd , and the 3rd stages as shown in Figure 1.4.

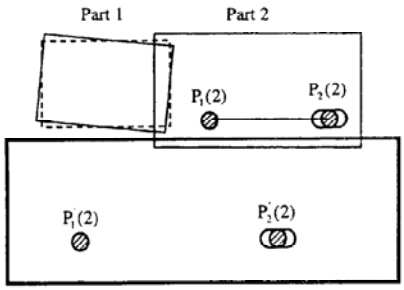


(a) Part locators and measurements layout (b) Manifestation of pin P_2 failure

Figure 1.1 Illustration of the effects of process fault and sensor fault in single-stage assemble process



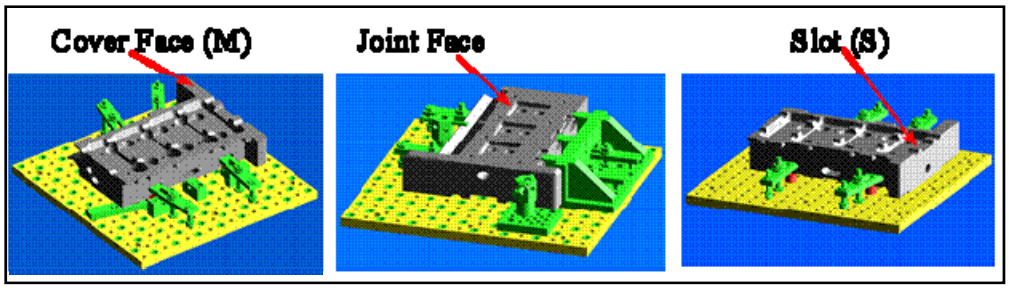
(a) Locator P2(1) malfunction in assembly station 1



(b) Manifestation of process fault propagation in assembly station 2

Figure 1.2 Illustration of process fault propagation in multistage assembly process

It can be seen that the errors of fixture and cutting tool-path at stage k (k =1, 2, 3) will cause the deviation of the quality features from their nominal positions. In addition, the sensor faults will also cause the false deviation alarm of the quality features.



(a) (b) (c)

Figure 1.3 Fixture locating schemes of an automotive engine head and its quality features

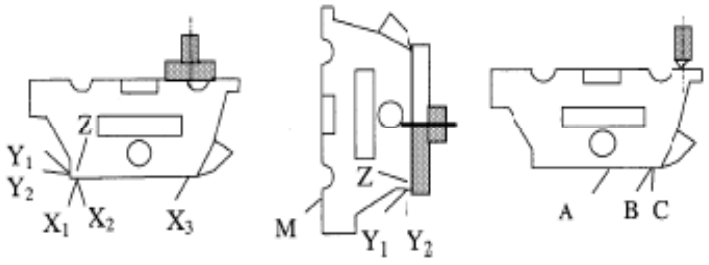


Figure 1.4 Operation sequence

1.1 Fault-Quality Model and Assumptions

We describe the relationship between processes faults, sensor faults and measurements deviation as a linear model which has been studied in the literature. We call this linear model as fault-quality model in our study. Specifically, the fault-quality

model can be written as follows: let an n×1 vector $\mathbf{y}_j = [y_{j1} \quad y_{j2} \quad \cdots \quad y_{jn}]^T$ denote the sensor measurements on n product quality characteristics of the jth product unit with sample size N :

$$\mathbf{y}_j = \mathbf{C}\mathbf{u} + \mathbf{w}_j, \quad j=1, \ldots, N \tag{1.1}$$

where the p×1 vector $\mathbf{u} = [u_1 \quad u_2 \quad \cdots \quad u_p]^T$ represents p potential process faults. The process faults $\mathbf{u}$ will affect the quality characteristics $\mathbf{y}_j$ linearly through an n×p matrix $\mathbf{C}$, which can be usually obtained through engineering analysis (for details, see Ceglarek and Shi, 1996; Jin and Shi, 1999; Zhou and Huang, 2003). The n×1 vector $\mathbf{w}_j$ denotes the aggregated effects of sensor noise and any inherent unmodeled variation in the manufacturing process for product j.

Our research is based on the *fault-quality* model. We consider the following assumptions in our studies:

(A1) The p-dimensional vector $\mathbf{u}$, which represents the deviations/errors caused by the process faults, remains constant but unknown. If no process fault exists, $\mathbf{u} = \mathbf{u}_0 \equiv \mathbf{0}$.

(A2) The noise term $\mathbf{w}_j$'s are independent random variables and identically follow a Gaussian distribution with mean $\boldsymbol{\mu}_\mathbf{w}$and covariance matrix $\mathbf{K}_\mathbf{w} = \boldsymbol{\Sigma}$, i.e., $\mathbf{w}_j i.i.d \sim N(\boldsymbol{\mu}_\mathbf{w}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive definite matrix. The mean $\boldsymbol{\mu}_\mathbf{w}$ is caused by sensor mean shift fault. If all the sensors are working properly, $\boldsymbol{\mu}_\mathbf{w} = \mathbf{0}$. σ^2 include the variances of the sensors and other process noises. We assume that the covariance matrix $\mathbf{K}_\mathbf{w}$ of noise term $\mathbf{w}_j$ is known or can be estimated from historical data.

1.2 Literature Review

1.2.1 SPC Technique

One widely applied technique to detect out-of-control conditions of a manufacturing process is Statistical Process Control (SPC) (Woodall and Montgomery,

1999). A variety of multivariate control charts in SPC can be used to monitor the collected measurements on product quality characteristics in real time to detect the presence of assignable causes and reduce process variation. Specifically, let $\mathbf{y}_j$ denote the n× 1 vector of product quality variables observed from the jth manufactured unit. $\mathbf{y}_j$ has mean vector $\boldsymbol{\mu}_{y0}$ and covariance matrix $\boldsymbol{\Sigma}$ when the process is in control. In order to monitor in-control mean vector $\boldsymbol{\mu}_{y0}$, a standard approach is to monitor the observations on $\mathbf{y}_j$ with a control chart such as Hotelling T^2 control chart, the multivariate cumulative sum of Crosier (1988) and Pignatiello and Runger (1990), or the multivariate exponentially-weighted moving average (MEWMA) of Lowry et al. (1992). Prabhu and Runger (1997) provided the recommendation of the parameter choice to improve the performance of the MEWMA control chart. Runger *et al* (1999) suggested that the MEWMA control chart can be significantly improved by transforming the original process variables to a lower-dimensional subspace through the use of a U-transformation. Alt (1984) and Montgomery (1991) presented phase I and phase II multivariate control charts to monitor the process variability. Chen, Cheng and Xie (2005) proposed a MEWMA control chart to monitor the process mean shift and variability simultaneously.

Some diagnosis approaches are studied to identify which variables are responsible for the out-of-control signal. Alt (1985) suggested applying univariate $\bar{x}$ chart on individual variables with Bonferroni-type control limits. Hayter and Tsui (1994) revised this method by producing exact individual confidence intervals such that the 'faulty' variable can be identified. Runger, Alt and Montgomery (1996) decomposed T^2 into components that reflect the contribution of each individual variable and focused on the variables highly contributing to the out-of-control signal. Mason, Tracy and Young (1995) introduced an approach to interpret the T^2 control chart based on decomposing the T^2 statistics into orthogonal components.

Two limitations exist, however, in the traditional control charts: first, the performance of the control charts will be affected by the dimension of the product quality

variables. For example, Apley and Shi (1998, 2001) noted that optical coordinate measurement machines (OCMM) installed in autobody assembly operations may obtain observations on as many as 150 product quality characteristics from each subassembly. The large number of product quality variables would create it difficulty in detecting quality change. Second, the control charts are used to monitor the manufacturing process, rather than directly identifying the root causes. Although some diagnostic methods have been studied, they are not efficient for our specific problem.

1.2.2 Linear Model and Related Monitor and Diagnosis Approaches

In recent years, the linear models, as we have described in Section 1.1, that link abnormal measurements with the potential process faults and sensor noises have been studied by several researchers. Apley and Shi's (1998) model is derived analytically from the geometry of the panel and fixture layout in assembly systems. Later on, several articles have proposed state-space models to describe the propagation of dimensional variation in discrete multi-station manufacturing processes. Such models have been developed for assembly and machining processes by Jin and Shi (1999), Ding, Ceglarek, and Shi (2002), Zhou, Huang, and Shi (2003), and others.

Based on the above fault-quality model, researchers have come up with many techniques in process fault diagnosis. Ceglarek and Shi (1996) developed a PCA (Principal Component Analysis)-based mapping procedure to diagnose a single fixture-fault in automotive body assembly processes. Ceglarek and Shi (1999) improved their previous method by considering the impact of measurement noise on the diagnostic results. They presented a diagnostic index as a function of noise, fixture geometry, and sensor location, which helped to identify a single fixture-fault in sheet metal assembly processes. Ding, Ceglarek and Shi (2002) applied PCA and pattern recognition methods to map the process faults for multistage processes. Ding, Zhou and Chen (2005) studied

the properties of a few statistical estimators of process variation sources and compared their performances with respect to computational and statistical efficiencies.

The SPC techniques are also combined with the linear fault-quality model on linear profile monitoring in recent literature. Kang and Albin (2000) and Kim, Mahmoud and Woodall (2004) studied a simple linear model that is a special case of model (1.1) in which the matrix **C** only has two columns with the first column equal to **1**, which is corresponding to the intercept term. They applied least squares regression and multivariate T^2 control chart to monitor mean shifts of the intercept and slope parameters.

The above fault monitoring and diagnosis research studies, however, only consider the malfunction of the manufacturing system and assumed that the sensing system is functioning normally. In real processes, this is not always the case. The malfunction of sensors may occur, which results in mean shifts in sensor measurements as shown in the single stage assembly and manufacturing examples. Both the process faults and the sensor faults may cause mean shifts in sensor observations. If a sensor fault is overlooked and confused with a process fault, resources will be wasted in trying to identify and eliminate process faults which may not really exist. Therefore, it is crucial to distinguish sensor faults from process faults in manufacturing quality control applications. Process fault detection based on sensor observations is valid only when it is assured that all sensors are working properly.

1.2.3 Sensor Fault Detection

Some studies in detection and identification of faulty sensor have been done in dynamic processes, especially in chemical processes. Fantoni and Mazzola (1996) applied a nonlinear principle component analysis (PCA) method to detect and identify fault sensors in nuclear power plants. Chou and Varhaegen (1997) developed a subspace algorithm to identify multivariable finite dimensional linear time-invariant systems assuming that the input and output measurements are contaminated by the white noise

and the measured output might also be affected by the propagated white input noise. Based on the subspace identification model, Qin and Li (2001) proposed a dynamic structured residual approach to identify faulty sensors in dynamic processes. These studies only focus on the occurrence of sensor faults with no process fault involved.

1.3 Objective and Outline of the Thesis

The limitations in the literature prompt us to do more research on fault monitoring and diagnosis in manufacturing processes. The objective of this thesis is to make the following contributions:

- Develop procedures to improve efficiency of traditional control charts in process fault and sensor fault monitoring by reducing the dimension of the monitored variables.
- Develop statistical monitoring methods to effectively differentiate sensor faults from process faults.
- Develop a Bayesian diagnosis approach to identify root causes responsible for the mean shift on the quality measurements.

In Chapter 2, with the assumption that there is no sensor fault, we present a new phase II multivariate control chart which can detect the process mean shift faults more efficiently by reducing the dimension of the monitored variable. The relationship between the proposed control chart procedure and some other control charts studied in the literature is also investigated in Chapter 2. In Chapter 3, we consider the occurrence of both process mean shift and sensor mean shift and develop a so called W control chart to distinguish the sensor faults from the process faults. When the W control chart is combined with the multivariate control chart in Chapter 2, the process faults and the sensor faults can be detected simultaneously. In Chapter 4, we further propose a diagnosis approach based on Bayesian variable selection methods to identify the faulty tooling elements and sensors responsible to the detected abnormal quality measurements.

CHAPTER 2.

DETECTING PROCESS FAULT FROM MULTIVARIATE DATA

In this chapter, we first assume the probability of sensor fault occurrence is negligible. Based on this condition, we modify the assumption (A2) as: the noise term $\mathbf{w}_j$ $i.i.d \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Then under model (1.1), the product quality vector $\mathbf{y}_j$ is multivariate normal with mean vector $\boldsymbol{\mu}_y = \mathbf{C}\mathbf{u}$ and covariance matrix $\boldsymbol{\Sigma}$. The process is considered to be in control if the parameter vector $\mathbf{u} = \mathbf{u}_0 \equiv \mathbf{0}$, in which case $\mathbf{y}_j$ has mean vector $\boldsymbol{\mu}_{y0} = \mathbf{C}\mathbf{u}_0 = \mathbf{0}$. The process is out of control if $\mathbf{u} \neq \mathbf{u}_0$. The objective is to detect an out-of-control process as quickly as possible. Since $\mathbf{u}$ has p elements, each of which may shift from zero, there are p possible faults. Any combination of these faults may be present simultaneously (i.e., zero, one, or multiple elements of $\mathbf{u}$ may be shifted from zero); thus, under the model (1.1), the mean vector of $\mathbf{y}_j$ may shift along any vector in the subspace spanned by columns of $\mathbf{C}$. This allows for a multitude of potential shift directions or special causes. For the time being, assume that $n \geq p$ and $\mathbf{C}$ has full column rank. In Section 2.4, we consider the case where the columns of $\mathbf{C}$ are linearly dependent.

2.1 Review of MEWMA Control Chart

As we have mentioned in Chapter 1, the observations $\mathbf{y}_j$ are usually directly monitored by a control chart such as the multivariate cumulative sums or the multivariate exponentially-weighted moving average (MEWMA). We discuss the MEWMA here because it has been shown to perform well and appears to have been studied more thoroughly than most multivariate monitoring methods. To define the MEWMA, we form the vector of MEWMA's

$$\mathbf{z}_j = \theta(\mathbf{y}_j - \boldsymbol{\mu}_{y0}) + (1-\theta)\mathbf{z}_{j-1} \quad j = 1, 2, \ldots \tag{2.1}$$

where $\boldsymbol{\mu}_{y0} = \mathbf{0}$, the smoothing constant θ is a scalar such that $0 < \theta \leq 1$ and $\mathbf{z}_0$ is a vector of all zeros. The MEWMA signals that the mean vector has shifted when (and only when) the statistic

$$\chi_j^2 \equiv \mathbf{z}_j' \boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1} \mathbf{z}_j \tag{2.2}$$

exceeds a control limit *h*, where

$$\boldsymbol{\Sigma}_{\mathbf{z}_j} = \{\theta[1-(1-\theta)^{2j}]/(2-\theta)\}\boldsymbol{\Sigma} \tag{2.3}$$

is the covariance matrix of $\mathbf{z}_j$. Instead of $\boldsymbol{\Sigma}_{\mathbf{z}_j}$ in equation (2.3), it is typically assumed in the literature that the asymptotic (as $j \to \infty$) covariance matrix

$$\{\theta/(2-\theta)\}\boldsymbol{\Sigma} \tag{2.4}$$

is used to calculate the MEWMA statistic χ_j^2 in equation (2.2). We refer to the MEWMA based on the exact covariance matrix in equation (2.3) as well as that based on equation (2.4) as *direct MEWMA* because both of these methods apply MEWMA directly to $\mathbf{y}_j$.

2.2 Proposed Procedure for Process Fault Monitoring

Instead of using direct MEWMA, we propose to detect shifts in the process vector $\mathbf{u}$ by monitoring its generalized least squares (GLS) estimator

$$\hat{\mathbf{u}}_j = (\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j \tag{2.5}$$

where it is assumed that the error covariance matrix $\boldsymbol{\Sigma}$ is known. Under the model (1.1), it is known that $\hat{\mathbf{u}}_j \sim N(\mathbf{u}, (\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1})$ since $\hat{\mathbf{u}}_j$ is a BLUE of $\mathbf{u}$. To apply a MEWMA to $\hat{\mathbf{u}}_j$, first compute the vector of EWMA's

$$\mathbf{z}_j^* \equiv \theta(\hat{\mathbf{u}}_j - \mathbf{u}_0) + (1-\theta)\mathbf{z}_{j-1}^* \quad j = 1, 2, \ldots, \tag{2.6}$$

where $\mathbf{z}_0^*$ is a vector of all zeros and, again, $0 < \theta \leq 1$. Since the in control vector $\boldsymbol{u}_0 \equiv \mathbf{0}$, the above equation can be simplified as

$$\mathbf{z}_j^* \equiv \theta\hat{\mathbf{u}}_j + (1-\theta)\mathbf{z}_{j-1}^* \quad j = 1, 2, \ldots \tag{2.7}$$

A signal of a shift in $\mathbf{u}$ is given when (and only when) the statistic

$$\chi_j^2 * \equiv \mathbf{z}_j *' \boldsymbol{\Sigma}_{\mathbf{z}_j *}^{-1} \mathbf{z}_j * \tag{2.8}$$

exceeds a control limit *h*, where

$$\boldsymbol{\Sigma}_{\mathbf{z}_j *} = \{\theta[1-(1-\theta)^{2j}]/(2-\theta)\}(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1} \tag{2.9}$$

is the covariance matrix of $\mathbf{z}_j *$. Note that the covariance matrix in equation (2.9)may not be obtained directly from $\boldsymbol{\Sigma}$ (the covariance matrix of $\mathbf{y}_j$); rather, it is obtained from $(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}$, the covariance matrix of $\hat{\mathbf{u}}_j$. We also note that, instead of using $\boldsymbol{\Sigma}_{\mathbf{z}_j *}$ in equation (2.9), the asymptotic form of this matrix, given by $\{\theta/(2-\theta)\}(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}$, may be used to compute the statistic in equation (2.8).

The MEWMA of $\hat{\mathbf{u}}_j$ reduces to a chi-squared chart for $\hat{\mathbf{u}}_j$ in the special case that $\theta = 1$. In this case, the MEWMA statistic $\chi_j^2 *$ in equation (2.8) reduces to

$$\chi_j^2 * = \hat{\mathbf{u}}_j{}'\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C}\hat{\mathbf{u}}_j$$

2.2.1 Run-Length Distribution and Design of the Procedure

It follows from Lowry et al.'s (1992) results that the run-length (RL) distribution of the MEWMA of $\hat{\mathbf{u}}_j$ depends on $\mathrm{E}(\hat{\mathbf{u}}_j)$ and $\mathrm{Cov}(\hat{\mathbf{u}}_j)$ only through the noncentrality parameter

$$\lambda = \sqrt{[\mathrm{E}(\hat{\mathbf{u}}_j)]'[\mathrm{Cov}(\hat{\mathbf{u}}_j)]^{-1}[\mathrm{E}(\hat{\mathbf{u}}_j)]} \tag{2.10}$$

where $E(\hat{\mathbf{u}}_j)$ and $Cov(\hat{\mathbf{u}}_j)$ are, respectively, the mean vector and covariance matrix of $\hat{\mathbf{u}}_j$. Using the results $E(\hat{\mathbf{u}}_j) = \mathbf{u}$ and $Cov(\hat{\mathbf{u}}_j) = (\mathbf{C}'\Sigma^{-1}\mathbf{C})^{-1}$ in (2.10) gives

$$\lambda = \sqrt{\mathbf{u}'\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C}\mathbf{u}} \tag{2.11}$$

The MEWMA of $\hat{\mathbf{u}}_j$ will have the RL distribution of a MEWMA applied to any p-variate normal vector with noncentrality λ. The above properties enable one to use existing results for the direct MEWMA to study the MEWMA of $\hat{\mathbf{u}}_j$'s RL distribution and select its two design constants θ and h. The MEWMA's RL distribution is studied, for example, by Lowry et al. (1992), Rigdon (1995a), Runger and Prabhu (1996), and

Prabhu and Runger (1997). Guidance for the selection of the MEWMA's design constants is given in Lowry et al. (1992), Rigdon (1995b), and Prabhu and Runger (1997). Prabhu and Runger (1997) find the design constants that minimize the average run length (ARL) until detection of a shift. Table 2.1 shows the minimum ARL for various values of the number of variables, the noncentrality parameter θ, and in-control ARL.

2.2.2 Comparison with Direct MEWMA

To compare the MEWMA of $\hat{\mathbf{u}}_j$'s detection performance to that of direct MEWMA, we compare the noncentralities of the two procedures. Under our general linear model (1.1), the MEWMA of $\hat{\mathbf{u}}_j$'s noncentrality parameter in equation (2.11) may be re-expressed as

$$\lambda = \sqrt{(\mathbf{Cu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Cu})} = \sqrt{\boldsymbol{\mu}_y{}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_y} \tag{2.12}$$

This result in equation (2.12) shows that the noncentrality of the MEWMA of $\hat{\mathbf{u}}_j$ is the same as that of the direct MEWMA. In the MEWMA of $\hat{\mathbf{u}}_j$, we apply a MEWMA to a p-vector, whereas in direct MEWMA a MEWMA is applied to an n-vector. Over a wide range of cases considered by Lowry et al. (1992) and Prabhu and Runger (1997), the MEWMA's detection performance always improves (in the sense of having a shorter ARL until detection of shifts) with a decrease in the number of variables that are monitored in the scheme. This finding is illustrated in Table 2.1 as well as in Table 2.2. Given the form of the MEWMA test statistic χ_j^2 in equation (2.3), this finding is to be expected from multivariate statistical theory. Based on this finding, we conclude that when $p<n$, the proposed MEWMA of $\hat{\mathbf{u}}_j$ outperforms direct MEWMA. When p (the number of columns in $\mathbf{C}$ is substantially less than n (the number of product quality variables), we see from Table 2.1 and Table 2.2 that the MEWMA of $\hat{\mathbf{u}}_j$ offers a much shorter ARL until detection of shifts.

Table 2.1 Minimal Average Run Length (ARL) of the MEWMA and associated design constants

		In-Control ARL					
		500			1000		
		Number of Variables			Number of Variables		
λ		4	10	20	4	10	20
0.5	min ARL	42.22	55.94	70.20	49.86	66.15	83.77
	θ, h	0.04, 13.37	0.03, 22.69	0.03, 37.09	0.03, 14.68	0.025,24.70	0.025,39.63
1.0	min ARL	14.60	19.29	24.51	16.52	21.74	27.65
	θ, h	0.105,15.26	0.085,25.42	0.075,40.09	0.09, 16.79	0.075,27.38	0.065,42.47
1.5	min ARL	7.65	10.01	12.70	8.50	11.07	14.01
	θ, h	0.18, 16.03	0.16, 26.58	0.14, 41.54	0.18, 17.71	0.14, 28.46	0.12, 43.80
2.0	min ARL	4.82	6.25	7.88	5.30	6.84	8.60
	θ, h	0.28, 16.49	0.24, 27.11	0.20, 42.15	0.26, 18.06	0.22, 29.02	0.18, 44.45
3.0	min ARL	2.55	3.24	4.04	2.77	3.50	4.35
	θ, h	0.52, 16.84	0.42, 27.55	0.36, 42.80	0.46, 18.37	0.40, 29.45	0.34, 45.08

Table 2.2 Average Run Length (ARL) of the MEWMA

	$\theta = 0.10$				$\theta = 0.40$			
	Number of Variables				Number of Variables			
λ	2	4	10	15	2	4	10	15
0.0	199.98	200.12	199.95	199.89	199.83	199.96	199.98	199.96
0.5	28.07	35.11	48.52	56.19	53.82	72.43	102.05	115.36
1.0	10.15	12.17	15.98	18.28	13.26	18.12	29.47	36.96
1.5	6.11	7.22	9.23	10.41	5.78	7.31	10.91	13.53
2.0	4.42	5.19	6.57	7.36	3.53	4.24	5.77	6.84
3.0	2.93	3.41	4.28	4.78	2.05	2.36	2.93	3.29

Note: All results in this table are from Prabhu and Runger (1997). The results are zero-state ARL's of the MEWMA based on the asymptotic covariance matrix in(2.4).

2.2.3 Example: Autobody Panel Assembly

Apley and Shi (1998) studied the process fault diagnosis problems for an automotive body panel assembly process as shown in Figure 2.1. In this process, a right-hand automotive bodyside is joined to the underbody and roof at the framing station. A four-way locating pin P_1 and a two-way locating pin P_2, whose positions are shown in Figure 2.1, are used to locate the right-hand bodyside in the *x*-*z* plane at the framing station. Optical coordinate sensors are used to measure deviations at points M_1 through M_4 on the bodyside in the *x*- and *z*- directions, the *x*-direction deviations of points M_5 through M_8,and the *z*-direction deviations of M_9 and M_{10}. Therefore there are totally n=14 sensor measurements in this process.

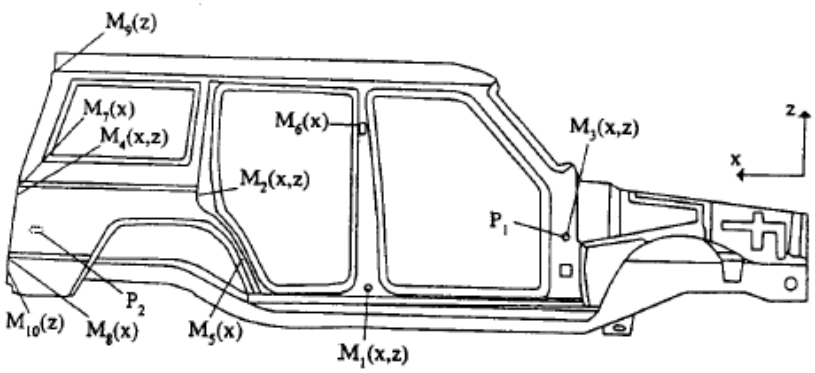


Figure 2.1 Key points on right-hand bodyside

The nominal x-z coordinates of P1, P2,and the ten measurement points are shown in Table 2.3 Let $y_j \equiv [M_1(x) \quad \cdots \quad M_8(x) \quad M_1(z) \quad \cdots \quad M_4(z) \quad M_9(z) \quad M_{10}(z)]$ represent the sensor measurement vector, where Mi(x) and Mi(z) denote the deviations from M_i's (i=1,..,10) nominal position in the x- and z- directions respectively. Assume we know pin P_2 is working under normal condition, and then there are p=2 potential process faults in this process. The mean shift of the process fault in model (1.1) is $\mathbf{u}$=[u_1, u_2],

where u_1 and u_2 are the deviations caused by errors of the 4-way locating pin P_1 in the *z*- and *x*- directions respectively.

According to the geometric relationship between locating pins and the measurement points, the $\mathbf{C}$ matrix as in model (1.1) is obtained as (readers refer to Apley and Shi (1998) for the derivation of $\mathbf{C}$ matrix):

$$\mathbf{C} = \begin{pmatrix} 0.354 & 0.057 \\ 0.354 & -0.026 \\ 0.354 & 0 \\ 0.354 & -0.004 \\ 0.354 & 0.046 \\ 0.354 & -0.087 \\ 0.354 & -0.024 \\ 0.354 & 0.043 \\ 0 & 0.187 \\ 0 & 0.361 \\ 0 & 0 \\ 0 & 0.535 \\ 0 & 0.495 \\ 0 & 0.536 \end{pmatrix}$$

with each column of $\mathbf{C}$ normalized.

In this example, we consider the potential process fault $\mathbf{u}_j$ as an independent random variable following a normal distribution rather than a constant, but the mean shift of the process faults is $E[\mathbf{u}_j] = \mathbf{u}$ and $Cov[\mathbf{u}_j] = \mathbf{K}_\mathbf{u}$. Let $\mathbf{u}_j = \mathbf{u} + \tilde{\mathbf{u}}_j$, where $\tilde{\mathbf{u}}_j$ is the random term of $\mathbf{u}_j$. We still can link the product quality vector $\mathbf{y}_j$ with the process fault in the form of model (1.1)

$$\mathbf{y}_j = \mathbf{C}\mathbf{u} + \tilde{\mathbf{w}}_j \qquad j=1, \ldots, N$$

where $\tilde{\mathbf{w}}_j = \mathbf{C}\tilde{\mathbf{u}}_j + \boldsymbol{w}_j$. The error covariance matrix $\boldsymbol{\Sigma}$ is equal to

$$\boldsymbol{\Sigma} = \mathbf{C}\mathbf{K}_\mathbf{u}\mathbf{C}' + \sigma_\mathrm{w}^2\mathbf{I}_\mathrm{w} \tag{2.13}$$

$\sigma_w^2 \mathbf{I}_w$ is the covariance matrix of the measurement errors, where $\mathbf{I}_w$ is an n×n identity matrix. From Ding, Zhou, and Chen (2005), we obtain $\sigma_w^2 = (0.1/6)^2$ and

$$\mathbf{K}_u = \begin{pmatrix} (0.2/6)^2 & 0 \\ 0 & (0.2/6)^2 \end{pmatrix}$$

Table 2.3 Nominal *x*-*z* coordinates for pins and measurement points

Point	Nominal Coordinates (mm)	
	x	*z*
P_1	2184	1489
P_2	4680	1428
M_1	3134	1200
M_2	4015	1618.5
M_3	2184	1489
M_4	4895.3	1510.5
M_5	3721	1256.5
M_6	3264	1930
M_7	4895	1608
M_8	4895.5	1273
M_9	4693.8	2228.5
M_{10}	4899	1214.5

Table 2.4 Noncentrality parameter λ corresponding to various fault mean shift

Process Mean Shift Fault	Noncentrality Parameter λ
$\mathbf{u} = (0 \quad 0)^T$	0
$\mathbf{u} = (0.039 \quad 0)^T$	1.05
$\mathbf{u} = (0 \quad 0.039)^T$	1.05
$\mathbf{u} = (0.039 \quad 0.039)^T$	1.48

Since in this example $p<<n$, the MEWMA of $\hat{\mathbf{u}}_j$ should outperform direct MEWMA. The control limit h of the MEWMA of $\hat{\mathbf{u}}_j$ may be chosen to give the desirable in-control ARL by finding the h-value for $p = 2$ in Table 1 of Rigdon (1995b). For an in-control ARL of 200 and $\theta = 0.10$, we obtain $h = 8.63$. (This is nearly equal to the 8.64 value in Prabhu and Runger (1997), upon which the first column of ARL's in our Table 2.1 is based.)

If one (or both) of the locator pins is bent or mis-located, the process fault mean shift $\mathbf{u} \neq \mathbf{0}$. For various possible shifts, we report the noncentrality parameter λ of the MEWMA of $\hat{\mathbf{u}}_j$. We see from Table 2.4 that if a single fault shifts in mean by 0.039 (a 1.17 standard-deviation shift), the noncentrality will be $\lambda = 1.05$. The ARL until detection of this shift is approximately 10.15 (from the first column of ARL's in Table 2.2). The ARL of a direct MEWMA scheme (with the same smoothing constant $\lambda = 0.10$ as the MEWMA of $\hat{\mathbf{u}}_j$) would (from Table 2.2) approximately be between 15.98 and 18.28 (the reasoning is that direct MEWMA is applied to $n = 14$ variables, and the ARL for such a scheme is between those for 10 and 15 variables). Therefore, direct MEWMA would require, on average, at least 50% longer than the MEWMA of $\hat{\mathbf{u}}_j$ to detect the shift. Now, if both faults concurrently shift by the amount 0.039, the noncentrality is $\lambda = 1.48$ (from Table 2.4). The MEWMA of $\hat{\mathbf{u}}_j$'s ARL until detection of this concurrent shift is approximately 6.11 (from Table 2.2). The ARL of direct MEWMA would approximately be between 9.23 and 10.41 (from Table 2.2), which again is more than 50% longer than that of the MEWMA $\hat{\mathbf{u}}_j$

2.3 Comparison with Other Methods

Table 2.5 shows how the proposed MEWMA of $\hat{\mathbf{u}}_j$ is related to each of five existing methods for monitoring multivariate processes.

2.3.1 Runger's (1996) U^2 Statistic

It was noted in Section 2.2 that if $\mathbf{u}$ shifts from its in-control value $\mathbf{u}_0 \neq \mathbf{0}$, the mean vector of $\mathbf{y}_j$ will shift to a vector in the subspace spanned by columns of $\mathbf{C}$. Runger (1996) proposes to monitor the squared length of the orthogonal projection of $\boldsymbol{\Sigma}^{-1/2}\mathbf{y}_j$ onto the subspace spanned by the columns of $\boldsymbol{\Sigma}^{-1/2}\mathbf{C}$, where $\boldsymbol{\Sigma}^{-1/2}$ is a symmetric square root of $\boldsymbol{\Sigma}^{-1}$. This squared length is given by

$$U^2 = \mathbf{y}_j'\boldsymbol{\Sigma}^{-1}\mathbf{C}(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j \tag{2.14}$$

Table 2.5 Relationship of proposed MEWMA of $\hat{\mathbf{u}}_j$ to existing methods

	λ	p	Columns of $\mathbf{C}$	$\boldsymbol{\Sigma}$
1. U^2 Statistic (Runger 1996)	1			
2. Healy's (1987) method	*	1		
3. Direct MEWMA (Lowry et al. 1992)		n		
4. Direct chi-squared control chart	1	n		
5. Linear profile monitoring (Kang & Albin 2000)	1	2	First column is all 1's	$\sigma^2\mathbf{I}$

Runger points out that U^2 has a chi-squared distribution with p degrees of freedom and proposes to monitor U^2 via a chi-squared chart. As indicated in Table 2.5, this proposal is equivalent to the MEWMA of $\hat{\mathbf{u}}_j$ with θ=1. We prove this statement as follows:

Proof: When θ=1, the vector of EWMA's in equation (2.7) is

$$\mathbf{z}_j^* = 1(\hat{\boldsymbol{\beta}}_j - \mathbf{0}) + (1-1)\mathbf{z}_{j-1}^*,$$

which reduces to $\mathbf{z}_j^* = \hat{\boldsymbol{\beta}}_j$. The MEWMA test statistic in equation (2.8) then reduces to

$$\chi_j^{2*} = \hat{\boldsymbol{\beta}}_j'(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})\hat{\boldsymbol{\beta}}_j$$

$$= \mathbf{y}_j'\boldsymbol{\Sigma}^{-1}\mathbf{C}(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j$$

$$= \mathbf{y}_j'\boldsymbol{\Sigma}^{-1}\mathbf{C}(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j \ ,$$

which is the same as Runger's (1996) U^2 statistic in (2.14).

The equivalence of the two procedures in this case is perhaps to be expected because of the relationship between the GLS estimator $\hat{\mathbf{u}}_j$ and the projection of $\boldsymbol{\Sigma}^{-1/2}\mathbf{y}_j$ onto the subspace spanned by columns of $\boldsymbol{\Sigma}^{-1/2}\mathbf{C}$.

Now consider the vector of EWMA's of the original observations $\mathbf{z}_j \equiv \theta\mathbf{y}_j + (1-\theta)\mathbf{z}_{j-1}$ in equation (2.1) (we omit term $\boldsymbol{\mu}_{\mathbf{y}0}$ since $\boldsymbol{\mu}_{\mathbf{y}0} = \mathbf{0}$).If the parameter vector $\mathbf{u}$ of our model (1.1) remains at its in-control value $\mathbf{u}_0 = \mathbf{0}$, the mean vector of the EWMA vector $\mathbf{z}_j$ equals zero. If a mean shift fault occurs on $\mathbf{u}$, it can be shown that the mean vector of $\mathbf{z}_j$ will shift to a vector in the subspace spanned by columns of $\mathbf{C}$. Given these results, Runger's (1996) projection method would suggest finding the orthogonal projection of the standardized MEWMA vector $\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1/2}\mathbf{z}_j$ onto the subspace spanned by the columns of $\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1/2}\mathbf{C}$, where $\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1/2}$ is a symmetric square root of $\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1}$. This projection's squared Euclidean distance from the origin

$$U^2 = \mathbf{z}_j'\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1}\mathbf{C}(\mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1}\mathbf{z}_j$$

would then be used as the control statistic. It can be shown that this statistic U^2 is equal to our MEWMA of $\hat{\mathbf{u}}_j$'s test statistic χ_j^2* in equation (2.9). The equivalence allows our χ_j^2* statistic to be interpreted as the squared length of the orthogonal projection of $\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1/2}\mathbf{z}_j$ onto the subspace spanned by the columns of $\boldsymbol{\Sigma}_{\mathbf{z}_j}^{-1/2}\mathbf{C}$.

2.3.2 Healy's (1987) Method

Healy (1987) uses the sequential-probability-ratio argument to derive the optimal cumulative sum method for detecting when the mean vector of a multivariate normal distribution shifts from $\mathbf{u}_0$ to a known vector $\mathbf{u}_1$. The case considered by Healy (1987) is the case where our $\mathbf{C}$ matrix has $p = 1$ column. In this case, $\mathbf{C}$ is an n-vector that indicates

the only possible direction of a mean shift. It can be shown as in the proof below that the statistic accumulated in the MEWMA of $\hat{\mathbf{u}}_j$ is proportional to the statistic accumulated in Healy's (1987) cumulative sum.

Proof: When $\mathbf{C}$ has p = 1 column, the component $\hat{\boldsymbol{\beta}}_j$ in equation (2.6) may be written as

$$\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_0 = \frac{\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j}{\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C}} - \boldsymbol{\beta}_0 = \frac{\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \mathbf{C}\boldsymbol{\beta}_0)}{\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C}} == c\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \boldsymbol{\mu}_{0y})$$

where c is some constant and in control $\boldsymbol{\beta}_0 \equiv \mathbf{0}$. $\mathbf{C}$ indicates the only possible direction of a mean shift, which is $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$. Substituting $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$ for $\mathbf{C}$ in the above expression gives that

$$\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_0 = c(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \boldsymbol{\mu}_0),$$

which is a scalar quantity proportional to the statistic accumulated in Healy's (1987) cumulative sum.

Therefore, the MEWMA of $\hat{\mathbf{u}}_j$ will be equivalent to Healy's (1987) method when $p = 1$, if a cumulative sum is applied in place of an EWMA.

2.3.3 Direct MEWMA and Chi-Squared Control Charts

It was shown previously that the MEWMA of $\hat{\mathbf{u}}_j$ outperforms direct MEWMA when $p<n$. In the case when $p = n$, the MEWMA of $\hat{\mathbf{u}}_j$ is equivalent to direct MEWMA. This can be proved as follows:

Proof: Assume the matrix $\mathbf{C}$ is positive definite, for the case that p = n in MEWMA, $\hat{\boldsymbol{\beta}}_j$ in (2.7) may be rewritten as

$$\hat{\boldsymbol{\beta}}_j = \mathbf{C}^{-1}\boldsymbol{\Sigma}(\mathbf{C}')^{-1}\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j = \mathbf{C}^{-1}\mathbf{y}_j$$

A MEWMA of $\hat{\boldsymbol{\beta}}_j = \mathbf{X}^{-1}\mathbf{y}_j$ must be equivalent to a MEWMA of $\mathbf{y}_j$ because MEWMA is invariant to pre-multiplying $\mathbf{y}_j$ by any (conformable) nonsingular matrix (for proof of the MEWMA's invariance, see Lowry et al. (1992)).

Direct MEWMA then reduces to the chi-squared control chart in the special case that $\theta = 1$, as noted in Lowry et al. (1992). Therefore, the chi-squared chart can be obtained from the MEWMA of $\hat{\mathbf{u}}_j$ when $\theta = 1$ and $p = n$.

2.3.4 Linear Profile Monitoring

As noted earlier, this work is also related to recent literature on monitoring linear profiles. A review and synthesis of this literature is given by Woodall et al. (2004). Many of the articles on linear profile monitoring (e.g., Kang and Albin 2000; Kim, Mahmoud and Woodall 2003; Mahmoud and Woodall 2004) assume that, in our notation, the $\mathbf{C}$ matrix has $p = 2$ columns, one of which is all ones, and the error covariance matrix $\boldsymbol{\Sigma}$ equals a constant times an identity matrix. The model considered in our work is more general in that it allows (i) $\mathbf{C}$ to have any number of columns $p \geq 1$, (ii) any entries in the $\mathbf{C}$ matrix, and (iii) the error covariance matrix $\boldsymbol{\Sigma}$ to be an arbitrary positive definite matrix (i.e., $\boldsymbol{\Sigma}$ need not be diagonal).

As indicated in Table 2.5, the Phase II multivariate method in Kang and Albin (2000) can be obtained from the MEWMA $\hat{\mathbf{u}}_j$ by setting $\theta = 1$ and $p = 2$, with all entries in one column of $\mathbf{C}$ equal to one, and the error covariance matrix $\boldsymbol{\Sigma}$ equal to a constant times an identity matrix.

2.3.5 Regression Adjustment

Several articles discuss procedures in which regression-adjusted variables are monitored. These procedures, like the MEWMA of $\hat{\mathbf{u}}_j$, also detect shifts in $\hat{\mathbf{u}}_j$. Some of these monitoring procedures are proposed for model (1.1) with only $n = 1$ product quality variable (see, e.g., Kim, Mahmoud, and Woodall 2003; Wade and Woodall 1993), whereas in the current study we consider $n > 1$. As a result of this difference, these regression-adjustment procedures are somewhat different than the MEWMA of $\hat{\mathbf{u}}_j$.

For the case $n > 1$, Hawkins (1991) develops a regression-adjustment procedure under the assumption that, at any given time, only one of the n product quality variables

in $\mathbf{y}_j$ may be shifted from its in-control mean. For the moment, suppose there is a single possible fault (i.e., $p = 1$) whose presence would shift the mean of only $\mathbf{y}_j$'s ith element. In this case, the $\mathbf{C}$ matrix in model (1.1) would just be an n-vector with the ith element equal to one and all other elements equal to zero, denoted as $\mathbf{C}_j$. This is a particular instance of the case considered in Healy (1987). It can be shown that the component of $\hat{\mathbf{u}}_j$ in equation (2.7) of the MEWMA vector $\mathbf{z}_j$* reduces to the scalar quantity $D_{ij} = c\mathbf{C}_i'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j - \boldsymbol{\mu}_0)$, where c is some constant. This statistic D_{ij} is proportional to (a) the residual from the regression of $\mathbf{y}_j$'s ith element on all other elements of $\mathbf{y}_j$, and (b) the statistic Z_i in Hawkins (1991).

Since in practice it may not be known which element of the product quality vector $\mathbf{y}_j$ will shift in mean, the statistic D_{ij} above may be computed for each product quality variable i. Hawkins's (1991) proposal is equivalent to monitoring each D_{ij}, $i = 1, 2,\ldots,n$, with a univariate CUSUM scheme. Since D_{ij} is proportional to the length of the orthogonal projection of $\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_j - \boldsymbol{\mu}_0)$ onto the vector $\boldsymbol{\Sigma}^{-1/2}\mathbf{C}_i$, Hawkins's (1991) proposal is to use simultaneous univariate CUSUM schemes to monitor n different subspaces that are not necessarily orthogonal. This monitoring procedure is designed to be sensitive to a shift in one of these subspaces (i.e., a shift in the mean of a single (arbitrary) product quality variable). It is not, however, sensitive to shifts that are (nearly) orthogonal to each of the n subspaces. Such shifts may occur under our general model (1.1) when multiple faults are jointly present. The joint presence of multiple faults is not a concern if our MEWMA of $\hat{\mathbf{u}}_j$ procedure is used because our procedure is designed to be sensitive to any shift in the subspace spanned by the columns of $\mathbf{C}$. Thus, our procedure is somewhat different from the regression-adjustment procedure of Hawkins.

The preceding enables us to categorize the various methods with respect to the number of faults they are designed to detect. Healy's (1987) method is designed for the case of $p = 1$ possible fault. Linear profile monitoring is designed for when there are $p = 2$ possible faults. Direct MEWMA and chi-squared control are obtained when $p = n$ (the

number of possible faults equals the number of product quality variables). Runger's (1996) U^2 statistic and our MEWMA of $\hat{\mathbf{u}}_j$ are more general in that they allow $1 \le p \le n$ possible faults. Finally, Hawkins's (1991) regression-adjustment procedure is designed for $p = n$ faults, but it assumes that only one of these faults may be present at any given time.

2.4 Singular Structure in **C** matrix

The GLS estimator of $\mathbf{u}$ in model (1.1) is obtained by solving the *normal equation*

$$(\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C})\hat{\mathbf{u}}_j = \mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j \tag{2.15}$$

under the assumption that $\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C}$ has an inverse. In some models, particularly for multi-station processes where sensors may not be installed at every station (see Section 2.5), the columns of the $n \times p$ matrix $\mathbf{C}$ are linearly dependent; therefore, the $p \times p$ matrix $\mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C}$ does not have an inverse. Consequently, there is no unique solution to equation (2.15) and (2.5) may not be applied to obtain the GLS estimator $\hat{\mathbf{u}}_j$.

One set of solutions for $\hat{\mathbf{u}}_j$ can be obtained as follows. Denote the column rank of $\mathbf{C}$ by r (note that $r < p$ and $r \le n$). Select r linearly independent columns of $\mathbf{C}$, denoted by $\mathbf{c}_{s_1}, \ldots, \mathbf{c}_{s_r}$, that form a basis for the column space of $\mathbf{C}$, where $1 \le s_1 < s_2 < \ldots < s_r \le p$. Define $S \equiv \{s_1 \quad s_2 \quad \cdots \quad s_r\}$ and the $n \times r$ matrix $\mathbf{C}(S) \equiv \left[\mathbf{c}_{s_1} \quad \mathbf{c}_{s_2} \quad \cdots \quad \mathbf{c}_{s_r}\right]$. Then, $\mathbf{C}(S)'\boldsymbol{\Sigma}^{-1}\mathbf{C}(S)$ is an $r \times r$ matrix of full rank. Let $\hat{\mathbf{u}}_j(S)$ denote the $r \times 1$ sub-vector consisting of the $s_1, s_2, \ldots, s_r$ elements of $\hat{\mathbf{u}}_j$ and $\hat{\mathbf{u}}_j(\bar{S})$ the $(p-r) \times 1$ sub-vector containing the remaining elements of $\hat{\mathbf{u}}_j$. A solution for $\hat{\mathbf{u}}_j$ in the normal equation (2.15) may be constructed by choosing these sub-vectors as follows:

$$\hat{\mathbf{u}}_j(\bar{S}) = \mathbf{0} \quad \text{and} \quad \hat{\mathbf{u}}_j(S) = [\mathbf{C}(S)'\boldsymbol{\Sigma}^{-1}\mathbf{C}(S)]^{-1}\mathbf{C}(S)'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j \tag{2.16}$$

It can be seen that any $\hat{\mathbf{u}}_j$ constructed from the sub-vectors in equation (2.16) will satisfy equation (2.15). Therefore, such a $\hat{\mathbf{u}}_j$ is a GLS estimator of $\mathbf{u}$ and can be used in

equation (2.7) to calculate the MEWMA vector $\mathbf{z}_j^*$. Since the sub-vector $\hat{\mathbf{u}}_j(\bar{S}) = \mathbf{0}$ does not contribute to $\mathbf{z}_j^*$ in equation (2.7), we have

$$\mathbf{z}_j^* = \theta\hat{\mathbf{u}}_j(S) + (1-\theta)\mathbf{z}_{j-1}^* \quad , \quad j = 1, 2, \ldots \tag{2.17}$$

The covariance matrix of $\hat{\mathbf{u}}_j(S)$ is $[\mathbf{C}(S)'\boldsymbol{\Sigma}^{-1}\mathbf{C}(S)]^{-1}$, and hence the covariance matrix of $\mathbf{z}_j^*$ in equation (2.17) equals

$$\boldsymbol{\Sigma}_{\mathbf{z}_j^*} = \{\theta[1-(1-\theta)^{2j}]/(2-\theta)\}[\mathbf{C}(S)'\boldsymbol{\Sigma}^{-1}\mathbf{C}(S)]^{-1} \tag{2.18}$$

The MEWMA of $\hat{\mathbf{u}}_j$'s test statistic $\chi_j^{2*} \equiv \mathbf{z}_j^{*\prime}\boldsymbol{\Sigma}_{\mathbf{z}_j^*}^{-1}\mathbf{z}_j^*$ is now computed from the $\mathbf{z}_j^*$ and $\boldsymbol{\Sigma}_{\mathbf{z}_j^*}$ given in equations (2.17) and (2.18), respectively (Alternatively, the test statistic χ_j^{2*} may be computed from the asymptotic value of $\boldsymbol{\Sigma}_{\mathbf{z}_j^*}$, which is $\{\theta/(2-\theta)\}[\mathbf{X}(S)'\boldsymbol{\Sigma}^{-1}\mathbf{X}(S)]^{-1}$). By replacing $\hat{\mathbf{u}}_j$ and $\mathbf{C}$ with $\hat{\mathbf{u}}_j(S)$ and $\mathbf{C}(S)$, respectively, all results developed in Section 2.2 can be applied for $\mathbf{C}$ matrices whose columns are linearly dependent.

When $\mathbf{C}$ has linearly dependent columns, it was noted that the MEWMA of $\hat{\mathbf{u}}_j$ is in effect applied to the *r*-vector $\hat{\mathbf{u}}_j(S)$, whereas direct MEWMA is applied to an *n*-vector. Since *r* (the rank of $\mathbf{C}$) cannot exceed the number of rows (*n*) in $\mathbf{C}$, we must have $r \le n$. Therefore, the proposed MEWMA of $\hat{\mathbf{u}}_j$ performs at least as well as direct MEWMA and outperforms direct MEWMA whenever $r<n$ (recall from Section 2.2 that a MEWMA performs better as the dimension of the vector being monitored decreases). It should be noted that the dimension of the monitored vector cannot be lower than *r*, the rank of the model matrix $\mathbf{C}$, in order to satisfy the normal equation in (2.15). Therefore, it is impossible to use a monitored vector with dimension lower than *r* to achieve an even better monitoring performance.

It is also of interest to note that, although different sets of basis vectors *S* may produce different GLS estimators $\hat{\mathbf{u}}_j$ (via equation (2.15)), the choice of *S* does not affect the results of the MEWMA of $\hat{\mathbf{u}}_j$ procedure (i.e., the MEWMA of $\hat{\mathbf{u}}_j$ is invariant to the *S* used). For convenience, we list the proof here:

Proof: Suppose $\mathbf{C}(S_1)$and $\mathbf{C}(S_2)$are two different $n \times r$ matrices, each of whose columns form a basis for the column space of $\mathbf{C}$. Then, there exists an $r \times r$ nonsingular matrix Q such that

$$\mathbf{C}(S_2) = \mathbf{C}(S_1) \tag{2.19}$$

Using the set of basis vectors in $\mathbf{C}(S_1)$, compute the vector of EWMA's $\mathbf{z}_j^*$ via (2.18) with $\hat{\boldsymbol{\beta}}_j(S)$ replaced by

$$\hat{\boldsymbol{\beta}}_j(S_1) \equiv [\mathbf{C}(S_1)'\boldsymbol{\Sigma}^{-1}\mathbf{C}(S_1)]^{-1}\mathbf{C}(S_1)'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j .$$

Now, $\mathbf{z}_j^*$ has covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{z}_j^*} = \{\theta[1-(1-\theta)^{2j}]/(2-\theta)\}[\mathbf{C}(S_1)'\boldsymbol{\Sigma}^{-1}\mathbf{C}(S_1)]^{-1},$$

and it can be verified the MEWMA of $\hat{\boldsymbol{\beta}}_j$'s test statistic $\chi_j^{2*} \equiv \mathbf{z}_j^{*\prime}\boldsymbol{\Sigma}_{\mathbf{z}_j^*}^{-1}\mathbf{z}_j^*$ equals

$$\chi_j^{2*} = \mathbf{t}_j^{*\prime}\boldsymbol{\Sigma}^{-1}\mathbf{C}(S_1)[\mathbf{C}(S_1)'\boldsymbol{\Sigma}^{-1}\mathbf{C}(S_1)]^{-1}\boldsymbol{\Sigma}_{\mathbf{z}_j^*}^{-1}[\mathbf{C}(S_1)'\boldsymbol{\Sigma}^{-1}\mathbf{C}(S_1)]^{-1}\mathbf{C}(S_1)'\boldsymbol{\Sigma}^{-1}\mathbf{t}_j^* \tag{2.20}$$

where $\mathbf{t}_j^* \equiv \theta\mathbf{y}_j + (1-\theta)\mathbf{t}_{j-1}^*$. If equation (2.19) is substituted into (2.20), the matrix $\mathbf{Q}$ vanishes. Therefore, χ_j^{2*} does not depend on the set of basis vectors chosen.

In general, it is impossible to uniquely estimate each element of $\mathbf{u}$ when the columns of $\mathbf{C}$ are linearly dependent (Ding, Shi, and Ceglarek 2002; Zhou, Ding, Chen, and Shi 2003). The subvector $\hat{\mathbf{u}}_j(\bar{S})$ is set equal to zero only for the purpose of calculating the MEWMA statistic to detect mean shifts; this practice does not imply that the elements of $\hat{\mathbf{u}}_j$ corresponding to the columns in $\bar{S}$ are estimated to be zero.

2.5 Singular Example: Multi-Station Autobody Panel Assembly

The singularity problem appears a lot in multi-station processes. Consider the multi-station, 2-D, panel-assembly process shown in Figure 2.2. At Station 1, two panel workpieces are assembled. Eight coordinate sensors, labeled M_i ,i = 1,…, 8, in Figure 2.2, are employed at Station 2 to measure the product quality variables. Each sensor observes two variables (giving a total of n = 16 measurements), corresponding to the x- and z-directions of the deviations of a selected point on the workpiece. Each workpiece is

restrained by a set of fixtures consisting of a four-way locator, which controls motion in both the x- and z-directions, and a two-way locator, which restrains motion in only the z-direction. The active locating points are labeled P_i , i = 1,..., 6, in Figure 2.2. We consider p = 4 potential process faults: a malfunction in the x-direction of four-way locators P_1 and P_3 and a malfunction in the z-direction of two-way locators P_2 and P_4. (For ease of presentation, we ignore any locator malfunctions at Station 2 as well as any malfunctions in the z-direction of Station 1's four-way locators P_1 and P_3.)

A linear model for multi-station panel-assembly processes has been established by Jin and Shi (1999). This model can be rewritten in the form of model (1.1). For our 2-D panel-assembly process, the **C** matrix that links the p = 4 process faults to the n = 16 sensor measurements is

$$\mathbf{C} = \begin{pmatrix} 0 & -0.7857 & 0 & 0.3846 \\ 0 & -0.1429 & 0 & 0.0699 \\ 0 & -0.7857 & 0 & 0.3846 \\ 0 & 0.5714 & 0 & -0.2797 \\ 0 & 0.0571 & 0 & -0.0280 \\ 0 & 0.5714 & 0 & -0.2797 \\ 0 & 0.0571 & 0 & -0.0280 \\ 0 & -0.1429 & 0 & 0.0699 \\ -1 & 0 & 1 & -1.0406 \\ 0 & 0 & 0 & -0.7804 \\ -1 & 0 & 1 & -1.0406 \\ 0 & 0 & 0 & 0.0910 \\ -1 & 0 & 1 & 0.3772 \\ 0 & 0 & 0 & 0.0910 \\ -1 & 0 & 1 & 0.3772 \\ 0 & 0 & 0 & -0.7804 \end{pmatrix}$$

This **C** has linearly-dependent columns and a rank of r = 3. Following the procedure for singularities, we choose the set S corresponding to the basis vectors $S = \{1,2,4\}$; thus, the $\mathbf{C}(S)$ matrix is obtained by deleting **C**'s third column. Similar to the

previous example, we consider the potential process fault term $\mathbf{u}_j$ as an independent random vector follows a normal distribution with mean $\mathbf{u}$ and covariance $\mathbf{K_u}$. The error covariance matrix $\boldsymbol{\Sigma}$ may be obtained by using the formula in equation (2.13) with values obtained from the design specifications in Section 2.2.3, in particular $\sigma_w^2 = (0.1/6)^2$ and

$$\mathbf{K_u} = \begin{pmatrix} (0.2/6)^2 & 0 & 0 & 0 \\ 0 & (0.2/6)^2 & 0 & 0 \\ 0 & 0 & (0.2/6)^2 & 0 \\ 0 & 0 & 0 & (0.2/6)^2 \end{pmatrix}$$

With these calculations, we now have the two constant matrices $\boldsymbol{\Sigma}$ and $\mathbf{C}(S)$ needed to implement the MEWMA of $\hat{\mathbf{u}}_j$ and study its performance.

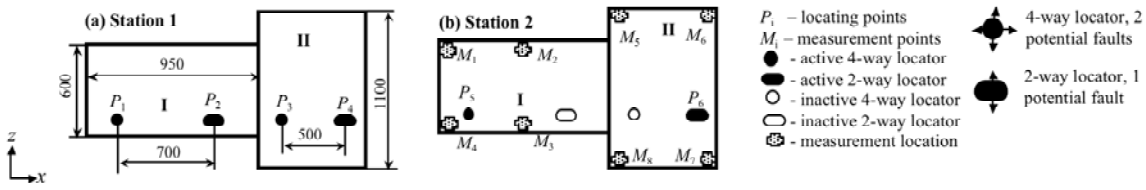


Figure 2.2 A multi-station 2-D panel-assembly process

In the MEWMA of $\hat{\mathbf{u}}_j$ procedure, a MEWMA is applied to $r = 3$ variables, whereas direct MEWMA is applied to $n = 16$ variables. ARL's for these numbers of variables are not available from previous articles, but some insight into the MEWMA of $\hat{\mathbf{u}}_j$'s performance advantage is obtained by comparing the ARL's for four and fifteen variables in Table 2.2 (note that the MEWMA of $\hat{\mathbf{u}}_j$ will have better detection performance than a MEWMA applied to four variables and direct MEWMA will have worse detection performance than a MEWMA applied to fifteen variables). As in the

previous example (Section 2.2.3), we see from Table 2.2 that the MEWMA of $\hat{\mathbf{u}}_j$ offers a noticeably shorter ARL until detection of shifts.

CHAPTER 3.
SENSOR FAULT DETECTION FOR MANUFACTURING QUALITY CONTROL

In this chapter, we consider both sensor faults in the sensing system and process faults in the manufacturing process. In assumption (A2), $\boldsymbol{\mu}_{\mathbf{w}}$, the mean of the noise term $\mathbf{w}_j$, may not equal to $\mathbf{0}$ due to existence of sensor faults. And manufacturers need to be notified as quickly as possible when $\boldsymbol{\mu}_{\mathbf{w}} \neq \mathbf{0}$. We will introduce a *W* control chart that can effectively distinguish sensor faults from process faults. We first specify the covariance of noise term $\mathbf{w}_j$ as $\mathbf{K}_{\mathbf{w}} = \sigma^2 \mathbf{I}_{\mathbf{w}}$. A more general heteroscedastic structure of $\mathbf{K}_{\mathbf{w}}$will be discussed in Section 3.3. In addition, we assume the columns of $\mathbf{C}$ matrix in model (1.1) are linear independent. In Section 3.1, we show the procedure to derive the *W* control chart from the fault-quality model. The sensitivity of the *W* chart to single, double, or multiple sensor faults is investigated in Section 3.2. In Section 3.3, the *W* chart is extended for sensors with unequal measurement precisions. Section 3.4 presents an example from automotive body assembly processes to demonstrate the performance of the *W* chart.

3.1 *W* Control Chart for Sensor Fault Monitoring

It has been noted that in SPC, people usually monitor the *n* quality characteristics $\mathbf{y}_j$ directly using a multivariate control chart, e.g. MEWMA, multivariate CUSUM and chi-squared control chart. The statistic monitored in χ^2 control chart (Montgomery, 2005) can be specified as

$$\chi_y^2 = N\bar{\mathbf{y}}^T \left(Var[\mathbf{y}_j]\right)^{-1} \bar{\mathbf{y}} = \frac{N}{\sigma^2}\bar{\mathbf{y}}^T\bar{\mathbf{y}}$$

where

$$\bar{\mathbf{y}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{y}_j$$

The upper control limit is set as $UCL = \chi^2_{\alpha,n}$ because the statistic χ^2_y follows a chi-squared distribution with *n* degrees of freedom and a noncentrality parameter

$$\lambda_{\mathbf{y}} = NE[\mathbf{y}_j]^T \left(Var[\mathbf{y}_j]\right)^{-1} E[\mathbf{y}_j] = \frac{N}{\sigma^2}(\mathbf{Cu}+\boldsymbol{\mu}_{\mathbf{w}})^T(\mathbf{Cu}+\boldsymbol{\mu}_{\mathbf{w}}).$$

In this chapter, the above χ^2 control chart is referred to as a *Y control chart* or a *Y* chart because it directly monitors the quality characteristics $\mathbf{y}_j$, j=1, …, N.

From model (1.1) we can see that both the process faults and the sensor faults will cause mean shifts in $\mathbf{y}_j$. Because the *Y* chart only detects the out-of-control conditions of the observed quality characteristics, one disadvantage of the *Y* chart is that it cannot differentiate between the sensor faults and the process faults.

To detect a mean shift in sensor measurements caused by a sensor fault, an ideal approach is to apply a χ^2 control chart to monitor the sample average of sensor measurement errors

$$\bar{\mathbf{w}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{w}_j$$

The χ^2 control chart can then generate out-of-control signals if mean shifts occur in sensor measurements. But this approach is obviously not practical because the sensor measurement errors $\mathbf{w}_j$, j=1, …, N are not directly observable for most applications. According to model (1.1), however, we can estimate $\mathbf{w}_j$ by

$$\hat{\mathbf{w}}_j = \mathbf{y}_j - \mathbf{C}\hat{\mathbf{u}}_j = [\mathbf{I}_{\mathbf{w}} - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T]\mathbf{y}_j$$

Let $\mathbf{H} \equiv \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$. The above equation can be written as

$$\hat{\mathbf{w}}_j = (\mathbf{I}_{\mathbf{w}} - \mathbf{H})\mathbf{y}_j \tag{3.1}$$

The matrix **H** is referred to as the *hat matrix* in linear regression literature. Based on the assumptions (A1) and (A2), the mean of the estimation of $\mathbf{w}_j$ is:

$$E[\hat{\mathbf{w}}_j] = [\mathbf{I}_{\mathbf{w}} - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T](\mathbf{Cu}+\boldsymbol{\mu}_{\mathbf{w}}) = (\mathbf{I}_{\mathbf{w}} - \mathbf{H})\boldsymbol{\mu}_{\mathbf{w}} \tag{3.2}$$

The variance of $\hat{\mathbf{w}}_j$

$$Var[\hat{\mathbf{w}}_j] = \sigma^2(\mathbf{I}_\mathbf{w} - \mathbf{H}) \tag{3.3}$$

Then we can apply a χ^2 control chart to monitor

$$\bar{\hat{\mathbf{w}}} = \frac{1}{N}\sum_{j=1}^{N}\hat{\mathbf{w}}_j$$

the sample average of $\hat{\mathbf{w}}_j$, j=1,2,...,N. This control chart is referred to as a *W control chart* or a *W* chart in this study. When there is no sensor fault, $\boldsymbol{\mu}_\mathbf{w} = \mathbf{0}$ and we have $E[\hat{\mathbf{w}}_j] = \mathbf{0}$. Therefore we set the in-control mean of $\bar{\hat{\mathbf{w}}}$ equal to $\mathbf{0}$. The test statistic plotted on this χ^2 control chart is

$$\chi^2 = N\bar{\hat{\mathbf{w}}}^T\left(Var[\hat{\mathbf{w}}_j]\right)^- \bar{\hat{\mathbf{w}}} \tag{3.4}$$

Substituting equations (3.2) and (3.3) into (3.4), we have

$$\chi^2 = \frac{N}{\sigma^2}\bar{\mathbf{y}}^T(\mathbf{I}_\mathbf{w} - \mathbf{H})(\mathbf{I}_\mathbf{w} - \mathbf{H})^-(\mathbf{I}_\mathbf{w} - \mathbf{H})\bar{\mathbf{y}} = \frac{N}{\sigma^2}\bar{\mathbf{y}}^T(\mathbf{I}_\mathbf{w} - \mathbf{H})\bar{\mathbf{y}} \tag{3.5}$$

where

$$\bar{\mathbf{y}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{y}_j$$

and $(\mathbf{I}_\mathbf{w} - \mathbf{H})^-$ denotes the generalized inverse matrix of $\mathbf{I}_\mathbf{w}$–$\mathbf{H}$. It can be seen that the $n\times n$ matrix $(\mathbf{I}_\mathbf{w} - \mathbf{H})$ is not invertible because its rank is equal to n–p. The χ^2 statistic plotted on the *W* control chart has the following property:

Result 3.1: Provided that there is no sensor fault, i.e. $\boldsymbol{\mu}_\mathbf{w} = \mathbf{0}$, χ^2 in equation (3.5) follows a chi-squared distribution with n–p degrees of freedom.

Proof: According to model (1.1) and assumptions (A1) and (A2), $\bar{\mathbf{y}} \sim N_n(\boldsymbol{\mu}_\mathbf{y}, \boldsymbol{\Sigma}_\mathbf{y})$, where $\boldsymbol{\mu}_\mathbf{y} = \mathbf{C}\mathbf{u} + \boldsymbol{\mu}_\mathbf{w}$ and

$$\boldsymbol{\Sigma}_\mathbf{y} = \frac{\sigma^2}{N}\mathbf{I}_\mathbf{w}$$

Under the normal working condition, we have $\mathbf{u}$=$\mathbf{0}$, $\boldsymbol{\mu}_\mathbf{w}$=$\mathbf{0}$, $\boldsymbol{\mu}_\mathbf{y}$=$\mathbf{0}$.

To find the distribution of the statistic

$$\chi^2 = \frac{N}{\sigma^2}(\bar{\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{y}})^T(\mathbf{I}_{\mathbf{w}} - \mathbf{H})(\bar{\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{y}}) = \frac{N}{\sigma^2}\bar{\mathbf{y}}^T(\mathbf{I}_{\mathbf{w}} - \mathbf{H})\bar{\mathbf{y}}$$

Let

$$\mathbf{A} \equiv \frac{N}{\sigma^2}(\mathbf{I}_{\mathbf{w}} - \mathbf{H})$$

Because $\mathbf{I}_{\mathbf{w}} - \mathbf{H}$ is an idempotent matrix with $rank(\mathbf{I}_{\mathbf{w}} - \mathbf{H}) = n - p$

$$\boldsymbol{\Sigma}_{\mathbf{y}}(\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{A} - \mathbf{A})\boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{\Sigma}_{\mathbf{y}}\left[\frac{N}{\sigma^2}(\mathbf{I}_{\mathbf{w}} - \mathbf{H}) - \frac{N}{\sigma^2}(\mathbf{I}_{\mathbf{w}} - \mathbf{H})\right]\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{0} \tag{3.6}$$

and

$$trace(\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{y}}) = trace(\mathbf{I}_{\mathbf{w}} - \mathbf{H}) = rank(\mathbf{I}_{\mathbf{w}} - \mathbf{H}) = n - p \tag{3.7}$$

According to Rao (1973, pp. 524), equations (3.6) and (3.7) are the necessary and sufficient condition for

$$\chi^2 = \bar{\mathbf{y}}^T\mathbf{A}\bar{\mathbf{y}} = \frac{N}{\sigma^2}\bar{\mathbf{y}}^T(\mathbf{I}_{\mathbf{w}} - \mathbf{H})\bar{\mathbf{y}}$$

to follow a chi-squared distribution with degrees of freedom $n–p$.

From Result 3.1, we can set $UCL = \chi^2_{\alpha,n-p}$ as the upper control limit of the W control chart, where $\chi^2_{\alpha,n-p}$ denotes the upper α-quantile of the chi-squared distribution with $n–p$ degrees of freedom. More generally, when $E[\hat{\mathbf{w}}_j] \neq \mathbf{0}$, χ^2 follows a noncentral chi-squared distribution with noncentrality parameter

$$\lambda_{\mathbf{w}} = \frac{N}{\sigma^2}E[\mathbf{y}_j]^T(\mathbf{I}_{\mathbf{w}} - \mathbf{H})E[\mathbf{y}_j] \tag{3.8}$$

It is well-known in control chart literature that the run-length distribution of the W chart, which is a χ^2 control chart, is determined by this noncentrality parameter (Montgomery, 2005).

In the rest of this section, we will simplify the MEWMA control chart we developed in the previous chapter to a χ^2 control chart which can detect the mean shift caused by the process faults. The W control chart and the following process fault control

chart should be used together to detect both sensor faults and process faults. We will see that the *W* chart should be used first in order to effectively distinguish sensor faults from process faults.

When the covariance matrix $\mathbf{\Sigma}$ of $\mathbf{w}_j$ is specified as $\sigma^2\mathbf{I}_{\mathbf{w}}$, the least squares estimator of $\mathbf{u}$ in equation (2.5) can be rewritten as

$$\hat{\mathbf{u}}_j = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{y}_j$$

According to assumptions (A1) and (A2), the mean of $\hat{\mathbf{u}}_j$ is equal to

$$E[\hat{\mathbf{u}}_j] = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{u}+\boldsymbol{\mu}_{\mathbf{w}}) = \mathbf{u}+(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\boldsymbol{\mu}_{\mathbf{w}}$$

The variance of $\hat{\mathbf{u}}_j$ is:

$$Var[\hat{\mathbf{u}}_j] = \sigma^2(\mathbf{C}^T\mathbf{C})^{-1}$$

Let

$$\bar{\hat{\mathbf{u}}} = \frac{1}{N}\sum_{j=1}^{N}\hat{\mathbf{u}}_j$$

we can plot the statistic

$$\chi^2_{\mathbf{u}} = N\bar{\hat{\mathbf{u}}}^T\left(Var[\hat{\mathbf{u}}_j]\right)^{-1}\bar{\hat{\mathbf{u}}} = \frac{N}{\sigma^2}\bar{\mathbf{y}}^T\mathbf{H}\mathbf{y}$$

on a control chart, which is called a *U control chart* or a *U* chart in this chapter, to monitor process faults. The upper control limit of the *U* chart is $UCL = \chi^2_{\alpha,p}$ because $\chi^2_{\mathbf{u}}$ follows a chi-squared distribution with *p* degrees of freedom (the proof is similar to that for the *W* chart). This *U* chart is a special case of the MEWMA control chart developed in Chapter 2 with θ=1 in equation (2.7). When process faults and/or sensor faults exist, $\chi^2_{\mathbf{u}}$ follows a noncentral chi-squared distribution with *p* degrees of freedom and the noncentrality parameter is given by:

$$\lambda_{\mathbf{u}} = \frac{N}{\sigma^2}(\mathbf{u}+(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\boldsymbol{\mu}_{\mathbf{w}})^T\mathbf{H}(\mathbf{u}+(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\boldsymbol{\mu}_{\mathbf{w}}) \qquad (3.9)$$

From equation (3.9), both process faults $\mathbf{u}$ and sensor faults $\boldsymbol{\mu}_{\mathbf{w}}$ affect $\lambda_{\mathbf{u}}$, which shows that the *U* chart is sensitive to both process faults and sensor faults. In the following section, we will see that, unlike the *U* chart, the *W* chart is only sensitive to the sensor faults, which is a desired property for distinguishing sensor faults from process faults.

As the *W* chart is only sensitive to sensor faults, while the *U* chart is sensitive to both sensor faults and process faults, regarding the usage of these two charts, we propose to first apply *W* chart to monitor sensor faults. When no sensor fault is found or each detected sensor fault has been fixed, the *U* chart can be used to detect the process faults.

Remark: The *W* chart and *U* chart are set up based on the assumption that the covariance $\mathbf{K}_{\mathbf{w}} = \sigma^2 \mathbf{I}_{\mathbf{w}}$ of $\mathbf{w}(j)$ is known so that $Var[\hat{\mathbf{w}}_j]$ and $Var[\hat{\mathbf{u}}_j]$ can be calculated. In practice, $\mathbf{K}_{\mathrm{w}}$ is not always known, in which case *W* chart and *U* chart should be changed to a Hotelling T^2 control chart (Montgomery, 2005). In the T^2 control chart, in-control means and variances of $\hat{\mathbf{w}}_j$ and $\hat{\mathbf{u}}_j$ are estimated from preliminary samples taken when the process is in control. Suppose we have *m* preliminary samples. The revised *U* chart can be obtained as follows: let $\hat{\mathbf{u}}_{pre,k}(j)$ be the least squares estimate of $\mathbf{u}_{pre,k}(j)$, where j=1, …, N and k=1, …, m. The sample mean and sample variance for the k^{th} sample are

$$\bar{\hat{\mathbf{u}}}_{pre,k} = \frac{1}{N}\sum_{j=1}^{N} \hat{\mathbf{u}}_{pre,k}(j) \text{ and } S^u_{pre,k} = \frac{1}{N-1}\sum_{j=1}^{N} (\hat{\mathbf{u}}_{pre,k}(j) - \bar{\hat{\mathbf{u}}}_{pre,k})(\hat{\mathbf{u}}_{pre,k}(j) - \bar{\hat{\mathbf{u}}}_{pre,k})^T$$

respectively.

The statistic $\bar{\hat{\mathbf{u}}}_{pre,k}$ and $S^u_{pre,k}$ is then averaged over all *m* samples to obtain

$$\bar{\bar{\hat{\mathbf{u}}}}_{pre} = \frac{1}{m}\sum_{k=1}^{m} \bar{\hat{\mathbf{u}}}_{pre,k} \text{ and } S^u_{pre} = \frac{1}{m}\sum_{k=1}^{m} S^u_{pre,k}$$

which is considered as the estimate of the in-control $E[\hat{\mathbf{u}}_j]$ and $Var[\hat{\mathbf{u}}_j]$. Finally, we plot the statistic $T_u^2 = N\left(\bar{\hat{\mathbf{u}}} - \bar{\bar{\hat{\mathbf{u}}}}_{pre}\right)^T (S^u_{pre})^{-1}\left(\bar{\hat{\mathbf{u}}} - \bar{\bar{\hat{\mathbf{u}}}}_{pre}\right)$ on the phase 2 Hotelling T^2 control chart with

$$UCL = \frac{p(m+1)(N-1)}{mN-m-p+1} F_{\alpha,p,mN-m-p+1}$$

where $F_{\alpha,p,mN-m-p+1}$ denotes the upper α-quantile of the F distribution with p and mN-m-p+1 degrees of freedom. For the revised W chart to monitor the mean shift of sensors, we can use a similar procedure to find the average sample mean $\overline{\overline{\hat{\mathbf{w}}}}_{pre}$ and sample variance S^{w}_{pre} as the estimation of $E[\hat{\mathbf{w}}_j]$ and $Var[\hat{\mathbf{w}}_j]$ and apply the T^2 control chart to detect sensor mean shift faults. It can be seen that S^{w}_{pre}, whose rank is n-p, is not a full rank matrix and does not have a regular inverse matrix. So generalized inverse of S^{w}_{pre} should be used to calculate the T^2 statistics on the revised W chart.

With a large number of preliminary samples, the Hotelling T^2 control chart is approximately equal to the χ^2 control charts (Montgomery, 2005). For simplicity, we will still consider the χ^2 control charts in the following discussion. When the number of preliminary samples is large, all the results for the χ^2 control charts will approximately apply for the T^2 charts.

3.2 Sensitivity Analysis for *W* chart

In this section we are going to discuss the sensitivity of the W control chart to sensor faults and process faults. We have seen that the χ^2 statistic in equation (2.5) generally follows a noncentral chi-squared distribution. Provided that the degrees-of-freedom is fixed, the run-length distribution of the W chart depends only on the noncentrality parameter $\lambda_{\mathbf{w}}$ given in equation (3.8). According to equation (3.2) and equation (3.8), the noncentrality parameter can be further written as

$$\lambda_{\mathbf{w}} = \frac{N}{\sigma^2} \boldsymbol{\mu}_{\mathbf{w}}^T (\mathbf{I}_{\mathbf{w}} - \mathbf{H}) \boldsymbol{\mu}_{\mathbf{w}} \tag{3.10}$$

From equation (3.10) it is obvious that $\lambda_{\mathbf{w}}$ is not related to process faults $\mathbf{u}$, which shows that the run-length of the W chart is not affected by the magnitude of the process faults. The insensitivity of the W chart to process faults is a critical property, which

assures that sensor faults can be effectively distinguished from process faults using the proposed *W* control chart.

From equation (3.10), the magnitude of the noncentrality parameter λ_w is obviously affected by the selection of the measuring units of the mean shift $\boldsymbol{\mu}_w$. We prefer a unit-free index to measure and compare the sensitivity of the *W* chart. Dividing equation (3.10) by

$$\frac{N}{\sigma^2}\boldsymbol{\mu}_w^T\boldsymbol{\mu}_w,$$

the squared length of $\boldsymbol{\mu}_w$ multiplied by a constant, the following index referred to as the *sensitivity ratio* can be obtained:

$$SR \equiv \frac{\boldsymbol{\mu}_w^T(\mathbf{I}_w - \mathbf{H})\boldsymbol{\mu}_w}{\boldsymbol{\mu}_w^T\boldsymbol{\mu}_w}, \text{ for } \boldsymbol{\mu}_w \neq \mathbf{0} \tag{3.11}$$

The sensitivity ratio defined in equation (3.11) is unit-free and always less than or equal to 1 because $\mathbf{I}_w - \mathbf{H}$,which is a projection matrix, is symmetric and semi-definite. Therefore, the sensitivity ratio always falls in the range of [0, 1]. A sensitivity ratio value closer to 1 indicates a better sensitivity of the *W* chart to the mean shift $\boldsymbol{\mu}_w \neq \mathbf{0}$. In the following subsections, the sensitivity of the *W* chart to different types of sensor faults will be studied using the sensitivity ratio defined above as the sensitivity index.

3.2.1 Sensitivity Ratio for Single-Sensor-Faults

To further study the sensitivity of the *W* chart, we consider different types of sensor mean shifts. The first case is the *single-sensor-fault*, which is a sensor fault causing a mean shift on a single sensor measurement. Under a single-sensor-fault, $\boldsymbol{\mu}_w$ can be written as

$$\boldsymbol{\mu}_w = [0 \quad 0 \quad \cdots \quad \mu_i \quad 0 \quad \cdots \quad 0]^T \tag{3.12}$$

where the i^{th} entry of $\boldsymbol{\mu}_w$ is $\mu_i \neq 0$ and the other entries are equal to zero. We denote the elements of the matrices **H** and $\mathbf{I}_w$–**H** by:

$$\mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1i} & \cdots & h_{1n} \\ \vdots & \ddots & \vdots & \iddots & \\ h_{i1} & \cdots & h_{ii} & \cdots & h_{in} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{ni} & \cdots & h_{nn} \end{bmatrix} \text{ and } \mathbf{I}_{\mathbf{w}} - \mathbf{H} = \begin{bmatrix} 1-h_{11} & \cdots & -h_{1i} & \cdots & -h_{1n} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ -h_{i1} & \cdots & 1-h_{ii} & \cdots & -h_{in} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ -h_{n1} & \cdots & -h_{ni} & \cdots & 1-h_{nn} \end{bmatrix}$$

Substituting $\boldsymbol{\mu}_{\mathbf{w}}$ of equation (3.12) into equation (3.11), we have

$$\lambda_{\mathbf{w}} = \frac{N}{\sigma^2}(1-h_{ii})\mu_i^2 \tag{3.13}$$

Further from equation (3.13), under the single-sensor-fault scenario, we have

$$SR = 1-h_{ii}. \tag{3.14}$$

So a smaller value of h_{ii} is corresponding to a larger sensitivity ratio SR.

It is interesting to note a relationship between the sensitivity ratio under the single-sensor-fault scenario and the concept of influential points in robust regression. In robust regression, an influential point is an observation having a large effect on the least squares estimation and prediction. An important indicator of a potential influential point is h_{ii}, the i^{th} diagonal element of the hat matrix $\mathbf{H}$. For example, an observation with $h_{ii}>2p/n$ is often considered as a potential influential point, which requires particular attention in regression analysis (Rousseeuw and Leroy, 2003, p. 220). In our applications, p represents the number of potential process faults and n is the number of sensor measurements. From equation (3.14), a highly influential sensor measurement (with a large h_{ii}) is corresponding to a small sensitivity ratio and in turn a low sensitivity of the W chart in detecting the sensor fault that occurs on that influential sensor. So a mean shift on a sensor corresponding to a highly influential sensor measurement will be relatively difficult to detect using the W chart. However, we can expect that a well-designed measurement system with good robustness property should have very few, if any, highly influential measurement points. For such measurement systems, the sensitivity of the W chart for most single-sensor-faults is high. Particularly, most of h_{ii} are expected to be well

below $2p/n$ in a well-designed measurement system, which corresponds to a sensitivity ratio greater than $1-2p/n$.

3.2.2 Sensitivity Ratio for Double-Sensor-Faults

The second case we study is the *double-sensor-fault,* which is a sensor fault causing mean shifts on two of the sensor measurements simultaneously. Specifically, under a double-sensor fault, the mean of the sensor noise vector is shifted from $\mathbf{0}$ to

$$\boldsymbol{\mu}_{\mathbf{w}} = \begin{bmatrix} 0 & \cdots & \mu_i & \cdots & \mu_j & \cdots & 0 \end{bmatrix}^T \tag{3.15}$$

where the i^{th} and j^{th} elements of $\boldsymbol{\mu}_{\mathbf{w}}$ are $\mu_i \neq 0$ and $\mu_j \neq 0$. Substituting equation (3.15) into equation (3.11), we have

$$SR = \frac{\begin{bmatrix} \mu_i & \mu_j \end{bmatrix} \begin{bmatrix} 1-h_{ii} & -h_{ij} \\ -h_{ji} & 1-h_{jj} \end{bmatrix} \begin{bmatrix} \mu_i & \mu_j \end{bmatrix}^T}{\begin{bmatrix} \mu_i & \mu_j \end{bmatrix} \begin{bmatrix} \mu_i & \mu_j \end{bmatrix}^T} \tag{3.16}$$

The following result is regarding the range of the values of the sensitivity ratio under a double-sensor fault:

Result 3.2*:* Let λ_S^{ij} and λ_L^{ij} represent the smallest and the largest eigenvalues of the 2×2 matrix

$$\begin{bmatrix} 1-h_{ii} & -h_{ij} \\ -h_{ji} & 1-h_{jj} \end{bmatrix} \tag{3.17}$$

Depending on different ratios of μ_i and μ_j, the sensitivity ratio in equation (3.16) ranges between λ_S^{ij} and λ_L^{ij}. That is,

$$\lambda_S^{ij} \leq SR \leq \lambda_L^{ij}.$$

Proof: Let $\mathbf{X} \equiv [\mu_i \quad \mu_j]^T$ and

$$\mathbf{A} \equiv \begin{bmatrix} 1-h_{ii} & -h_{ij} \\ -h_{ji} & 1-h_{jj} \end{bmatrix}$$

Equation (3.16) can be written as

$$SR = \frac{\mathbf{X}^T \mathbf{A} \mathbf{X}}{\mathbf{X}^T \mathbf{X}}$$

Since **A** is a symmetric matrix, the following results hold

$$\inf_{\mathbf{x}} \frac{\mathbf{X}^T \mathbf{A} \mathbf{X}}{\mathbf{X}^T \mathbf{X}} = \lambda_S^{ij} \qquad \sup_{\mathbf{x}} \frac{\mathbf{X}^T \mathbf{A} \mathbf{X}}{\mathbf{X}^T \mathbf{X}} = \lambda_L^{ij}$$

The supremum is attained when **X** is equal to the eigenvector corresponding to λ_L^{ij}, and the infimum is attained when **X** is equal to the eigenvector corresponding to λ_S^{ij}. See Rao (1973, pp. 46)for proof. Therefore,

$$\lambda_S^{ij} \le SR \le \lambda_L^{ij}.$$

From the above proof, when $[\mu_i \quad \mu_j]$ is proportional to one of the eigenvectors of the matrix in equation (3.17), the corresponding bounds, λ_S^{ij} or λ_L^{ij}, will be achieved by the sensitivity ratio.

We may define $(\lambda_S^{ij} + \lambda_L^{ij})/2$ as the "average" sensitivity ratio of a double-sensor-fault causing mean shifts of the i^{th} and the j^{th} sensor measurements. From linear algebra, we know that $\lambda_S^{ij} + \lambda_L^{ij}$, the summation of the eigenvalues of the matrix in (3.17), is equal to the trace of the matrix, which is $(1-h_{ii})+(1-h_{jj})$. As a result, we can write the "average" sensitivity ratio as:

$$\frac{(\lambda_S^{ij} + \lambda_L^{ij})}{2} = \frac{(1-h_{ii})+(1-h_{jj})}{2} = 1 - \frac{(h_{ii}+h_{jj})}{2} \tag{3.18}$$

From Section 3.2.1, we see that the diagonal elements of **H** are indicators of influential points in robust regression. We may consider

$$\frac{(h_{ii}+h_{jj})}{2}$$

as the "average" influential index for sensor measurements *i* and *j*. Recall that for single-sensor-fault on sensor measurement *i*, we have SR=1-h_{ii}. The result in equation (3.18) for double-sensor faults is similar to the result for single-sensor-faults. From the above, we

conclude that, for both single-sensor-faults and double-sensor-faults, the W chart has a higher sensitivity to sensor faults occurred on less influential sensor measurements.

The corollary below presents a lower bound of the sensitivity ratio satisfied by any double-sensor-fault for a well-designed measurement system.

Corollary 1: Let p and n represent the number of potential process faults and the measurement points, respectively. For a well-designed measurement system in which all the diagonal elements h_{ii} (i=1, …,n) in the hat matrix **H** are smaller than $2p/n$, the lower bound of the sensitivity ratio under *any* double-sensor-fault is

$$SR \geq 1 - \frac{4p}{n}.$$

Proof: Let

$$\mathbf{A} \equiv \begin{bmatrix} 1-h_{ii} & -h_{ij} \\ -h_{ji} & 1-h_{jj} \end{bmatrix}$$

where h_{ij} is the $(i,j)^{\text{th}}$ element of the matrix **H**. In the proof of Result 3.2, we showed that the sensitivity ratio can attain the values of λ_S^{ij} or λ_L^{ij} when the nonzero elements of $\boldsymbol{\mu}_{\mathbf{w}}$ are given by the eigenvector of **A** corresponding to λ_S^{ij} or λ_L^{ij}, respectively. Also we have $0 \leq \lambda_S^{ij} \leq 1$ and $0 \leq \lambda_L^{ij} \leq 1$ because the sensitivity ratio falls in [0, 1]. Since the measurement system is well-designed, the i^{th} and j^{th} diagonal elements h_{ii} and h_{jj} of the hat matrix **H** satisfy

$$h_{ii} \leq \frac{2p}{n},\ h_{jj} \leq \frac{2p}{n} \tag{3.19}$$

According to the property of eigenvalues of a matrix,

$$\lambda_S^{ij} + \lambda_L^{ij} = (1-h_{ii}) + (1-h_{jj}) = 2 - h_{ii} - h_{jj}.$$

Applying the inequalities of (3.19), it can be seen that $\lambda_L+\lambda_S$ satisfies

$$\lambda_S^{ij} + \lambda_L^{ij} \geq 2 - \frac{4p}{n} \tag{3.20}$$

Equation (3.19) can be further written as

$$\lambda_S^{ij} \geq 2 - \frac{4p}{n} - \lambda_L^{ij} \tag{3.21}$$

Substituting $\lambda_L^{ij} \leq 1$ into (3.21), we have

$$\lambda_S^{ij} \geq 1 - \frac{4p}{n}$$

Therefore $1-4p/n$ is a lower bound of the sensitivity ratio for any double sensor fault in a well-designed measurement system

The results for double-sensor-faults can be extended to the more general case of m-sensor-faults, which is a sensor fault causing mean shifts on m entries of $\boldsymbol{\mu}_{\mathbf{w}}$, $2 \leq m \leq n$. It can be seen that the sensitivity ratio of the W chart for an m-sensor fault would have a similar form to equation (3.16), with the 2×2 matrix in equation (3.17) replaced by an appropriate $m \times m$ symmetric submatrix $\mathbf{Q}$ of $\mathbf{I}_{\mathbf{w}} - \mathbf{H}$. Again, the smallest and the largest eigenvalues of $\mathbf{Q}$ are the smallest and largest values of the sensitivity ratio, respectively.

3.3 Extension to Heteroscedastic Sensor Measurements

In many real manufacturing processes, sensors of different precisions may be deployed to measure different product quality characteristics. One of the reasons is the economical concern: the high precision sensors with small measurement variance are typically expensive and therefore only used to measure the most critical quality characteristics. While for the less important quality characteristics, the less expensive sensors with relatively lower precision are installed.

To accommodate sensor measurements with unequal variances, also called heteroscedastic sensor measurements, the variance-covariance matrix $\mathbf{K}_{\mathbf{w}}$ of the sensor mean measurement error term $\mathbf{w}_j$ in assumption (A2) of model (1.1) can be specified as:

$$\mathbf{K}_{\mathbf{w}} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}$$

where $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ can be different from each other.

For the heteroscedastic model, we multiply $\mathbf{K}_{\mathbf{w}}^{-1/2}$ on both sides of model (1.1) as

$$\mathbf{K}_{\mathbf{w}}^{-1/2}\mathbf{y}_j = \mathbf{K}_{\mathbf{w}}^{-1/2}\mathbf{C}\mathbf{u} + \mathbf{K}_{\mathbf{w}}^{-1/2}\mathbf{w}_j \tag{3.22}$$

where $\mathbf{K}_{\mathbf{w}}^{-1/2}$ is the inverse of the square root of $\mathbf{K}_{\mathbf{w}}$. Let $\mathbf{y}'_j \equiv \mathbf{K}_{\mathbf{w}}^{-1/2}\mathbf{y}_j$, $\mathbf{C}' \equiv \mathbf{K}_{\mathbf{w}}^{-1/2}\mathbf{C}$ and $\mathbf{w}'_j \equiv \mathbf{K}_{\mathbf{w}}^{-1/2}\mathbf{w}_j$. Equation (3.22) can be written as

$$\mathbf{y}'_j = \mathbf{C}'\mathbf{u} + \mathbf{w}'_j \tag{3.23}$$

For the transformed sensor measurement error term $\mathbf{w}'_j$, it is easy to see that $E[\mathbf{w}'_j] = \mathbf{K}_{\mathbf{w}}^{-1/2}\boldsymbol{\mu}_{\mathbf{w}}$ and $Var[\mathbf{w}'_j] = \mathbf{K}_{\mathbf{w}}^{-1/2}\mathbf{K}_{\mathbf{w}}\mathbf{K}_{\mathbf{w}}^{-1/2} = \mathbf{I}_{\mathbf{w}}$. The transformed model in equation (3.23) is actually the same as the model in equation (1.1), with the same variance/covariance structure for the measurement error term. Therefore, the *W* chart monitoring the least squares estimation of $\mathbf{w}'_j$ based on equation (3.23) can be used to detect sensor faults. All the methods and the sensitivity results studied in the previous sections can be applied to the model in equation (3.23).

3.4 Automotive Body Assembly Example

We still use Apley and Shi (1998)'s assembly example studied in Section 2.2.3. Assuming both pin P_1 and P_2 might malfunction, there are p=3 potential process faults in this process. The process fault vector as in model (1.1) is denoted by $\mathbf{u}$=[u_1, u_2, u_3], where u_1 and u_2 are the deviations caused by errors of the 4-way locating pin P_1 in the *z*- and *x*-directions respectively, and u_3 is the deviation caused by errors of the 2-way locating pin P2 in the z- direction. In addition, sensor measurement error term $\mathbf{w}_j$ of this process is a 14×1 vector. The performance of the W chart in detecting potential sensor faults at the 14 sensor measurements will be investigated as follows.

To monitor sensor faults, as well as process faults, based on the *W* chart and *U* chart, we use a sample size N=40 and the six-sigma range of sensor noise equal to

0.2mm, which is corresponding to the variance of each sensor given by $\sigma^2 = (0.2/6\,\text{mm})^2$.

The **C** matrixis obtained as:

$$\mathbf{C} = \begin{bmatrix} 0.11579 & 1 & -0.11579 \\ -0.051883 & 1 & 0.051883 \\ 0 & 1 & 0 \\ -0.008614 & 1 & 0.008614 \\ 0.093149 & 1 & -0.093149 \\ -0.17668 & 1 & 0.17668 \\ -0.047676 & 1 & 0.047676 \\ 0.086538 & 1 & -0.086538 \\ 0.38061 & 0 & 0.61939 \\ 0.73357 & 0 & 0.26643 \\ 0 & 0 & 1 \\ 1.0863 & 0 & -0.086258 \\ 1.0055 & 0 & -0.005529 \\ 1.0877 & 0 & -0.08774 \end{bmatrix}$$

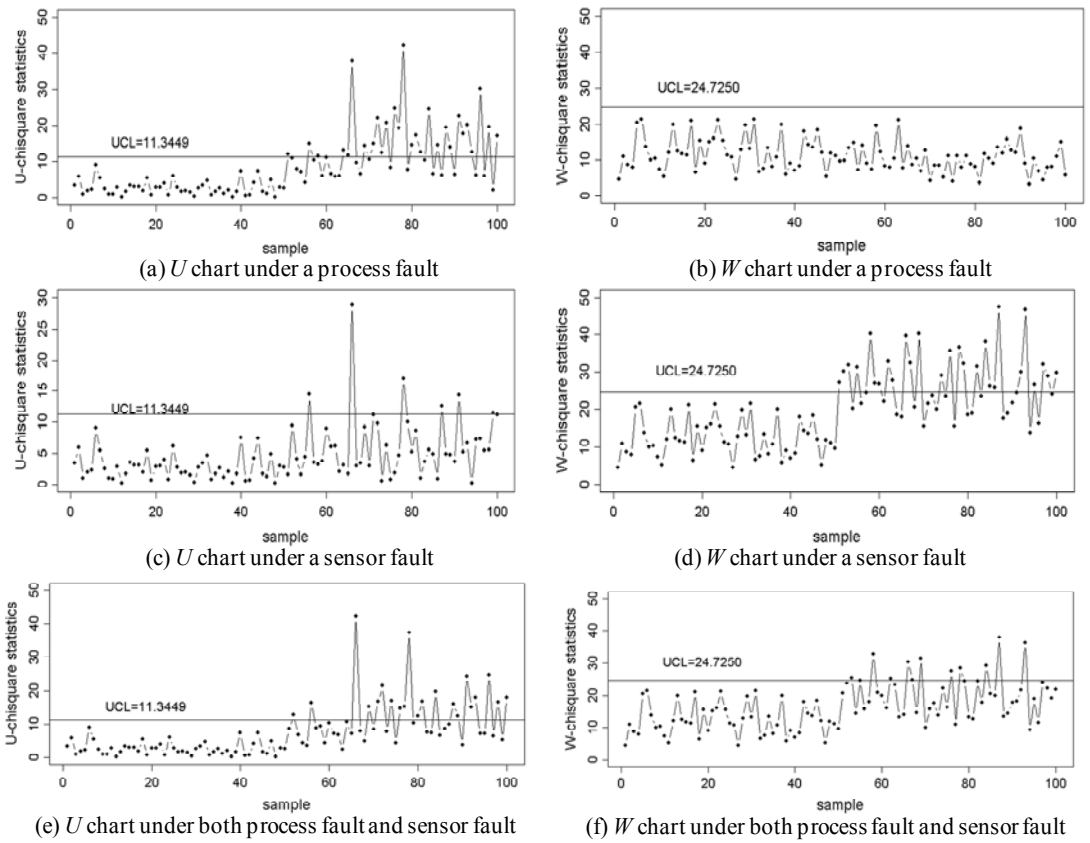
The columns of **C** matrix are not normalized in this example.

To illustrate the usage of the *W* chart and *U* chart for sensor and process fault detection, we show examples of control charts under process/sensor faults in Figure 3.3. Each control chart in Figure 3.3 is used to monitor 100 samples of data simulated from the automotive body panel assembly example described above. A mean shift caused by a sensor fault, a process fault, or both, is added for the last 50 samples. From Figure 3.3, it can be seen that the *U* chart will signal under either a process fault, or a sensor fault, or both. The *W* chart, however, will signal only if there is a sensor fault. The process fault and sensor fault can be distinguished by using the *W* chart first and then the *U* chart. For example, if an out-of-control signal is detected on the *W* chart, we know that either a sensor fault occurred or both a process fault and a sensor fault occurred, as in Figure 3.3(d) or Figure 3.3(f). Then we can focus on the sensor fault first. After the root cause of

the sensor fault is identified and all sensor faults are eliminated, we should see no out-of-control signal on the *W* chart, as in Figure 3.3(b). At this time, the *U* chart will be examined. If there is still an out-of-control signal on the *U* chart, as shown in Figure 3.3(a), then a process fault may still exist and should be identified and eliminated. Otherwise, the system is in normal working state without either process fault or sensor fault. The *Y* chart, not shown in Figure 3.3, will be sensitive to both process faults and sensor faults, similar to the *U* chart. Therefore, without the development of *W* chart, neither the *Y* chart nor the *U* chart can distinguish sensor faults from process faults.

Next, we conduct analysis of the sensitivity of the *W* chart under different sensor fault scenarios following the discussions in Section 3.3. Under a single-sensor-fault, in which only one of the sensor measurements, the i^{th} measurement, has a mean shift to μ_i. We set $\mu_i = 0.2/6$ mm to calculate the noncentrality parameter of the *W* chart. The noncentrality parameters λ_w and the sensitivity ratio *SR*, which can be calculated by equations (3.11)and (3.12) respectively, are listed in Table 3.1 for single-sensor-faults at different sensor measurements. For example, when a single-sensor-fault causes a mean shift at the sensor measuring $M_4(x)$, λ_w and *SR* are 34.996 and 0.8749, respectively.

Table 3.1 indicates that the *W* chart has a sensitivity ratio above 0.7 for most of the single-sensor-faults, with the only exception for $M_3(z)$. Further examining $M_3(z)$, we find that the diagonal element of the hat-matrix **H** corresponding to $M_3(z)$ is equal to 0.6542. It exceeds the threshold for potential influential measurement points given in Section 3.2.1, which is $2p/n = 2 \times 3/14 = 0.4286$. Consequently, $M_3(z)$ is considered as a highly influential observation, whose mean shifts are difficult to detect as discussed in Section 3.2.1. In addition, the high influence of the sensor measurement $M_3(z)$ can also be seen from its geometrical location as shown in Figure 2.1. Apparently the location of M_3 is somewhat isolated from the other sensors in the *x*- direction, which makes $M_3(z)$ a highly influential observation.



(a) *U* chart under a process fault (b) *W* chart under a process fault

(c) *U* chart under a sensor fault (d) *W* chart under a sensor fault

(e) *U* chart under both process fault and sensor fault (f) *W* chart under both process fault and sensor fault

Figure 3.3 Examples of control charts under process/sensor faults

Table 3.1 Noncentrality parameters and sensitivity ratios of single-sensor-faults

	$M_1(x)$	$M_2(x)$	$M_3(x)$	$M_4(x)$	$M_5(x)$	$M_6(x)$	$M_7(x)$
λ_w	34.501	34.892	35	34.996	34.679	33.793	34.909
SR (1-h_{ii})	0.8625	0.8723	0.8750	0.8749	0.8670	0.8448	0.8727
	$M_8(x)$	$M_1(z)$	$M_2(z)$	$M_3(z)$	$M_4(z)$	$M_9(z)$	$M_{10}(z)$
λ_w	34.724	29.059	33.321	13.831	28.102	30.133	28.06
SR (1-h_{ii})	0.8681	0.7265	0.8330	0.3458	0.7026	0.7533	0.7015

Regarding detection of a sensor fault at a highly influential sensor measurement, we suggest periodical inspections of the highly influential points using another gauging instrument. For example, in addition to the online optical sensors as shown in Figure 2.1, another gauge system called Coordinate Measuring Machine (CMM) are also used in most automotive body assembly lines to measure selected quality characteristics for a small sample of products. If the influential points are measured periodically by CMM, the measurements from CMM can be compared with those from the online optical sensors to enhance the detection sensitivity to the sensor faults at influential points. This method is feasible because the number of influential points is usually very small for a well-designed sensor system.

To compare the performance of the *W* chart with that of the *Y* chart defined in Section 3.1, we arbitrarily select three sensor measurements: $M_1(x)$, $M_4(x)$, and $M_4(z)$. Table 3.1 lists the noncentrality parameters and average run lengths (ARL) of the *W* chart and the *Y* chart in detection of various sizes of sensor mean shifts. For the *Y* chart, the noncentrality parameter λ_y is calculated by

$$\lambda_y = \frac{N}{\sigma^2}\boldsymbol{\mu}_y^T\boldsymbol{\mu}_y = \frac{N}{\sigma^2}\mu_i^2$$

So for single-sensor-faults, no matter the mean shift occurs in which sensor measurement, the values of λ_y are the same as long as the values of μ_i are the same. This is why the *Y* chart in Table 3.2 has only one column of λ_y and ARL. From Table 3.2, the sensor fault detection performance of the *W* chart is close to that of the *Y* chart. The major

Table 3.2 ARL comparison of *W* chart and *Y* chart under single-sensor-faults

Mean shift in single sensor	*W* chart						*Y* chart	
	$M_1(x)$		$M_4(x)$		$M_4(z)$			
	λ_w	ARL	λ_w	ARL	λ_w	ARL	λ_y	ARL

0	0	370.37	0	370.37	0	370.37	0	370.37
0.2σ	1.38	128.94	1.40	127.29	1.12	153.19	1.598	127.91
0.4σ	5.52	19.26	5.60	18.76	4.50	27.72	6.408	18.03
0.6σ	12.42	3.98	12.599	3.87	10.12	5.92	14.4	3.60
0.8σ	22.08	1.56	22.398	1.53	17.98	2.08	25.596	1.44
σ	34.50	1.08	34.996	1.08	28.10	1.22	39.996	1.05

advantage of the *W* chart over the *Y* chart, however, is that the *W* chart can be used to distinguish sensor faults from process faults, while the *Y* chart cannot. This can be seen clearly from Table 3.2, where the noncentrality parameters and ARL of the *W* chart, *Y* chart, and *U* chart are compared for detection of a single process fault causing a mean shift of 0.0167mm at each of the three elements of the vector **u** (corresponding to deviations $P_1(x)$, $P_1(z)$ and $P_2(z)$, respectively) .

Table 3.3 ARL of the *W* chart, *Y* chart, and *U* chart
in detection of a process fault

Location of process fault	W chart		Y chart		U chart	
	λ_w	ARL	λ_y	ARL	λ_u	ARL
$P_1(x)$	0	370.37	80	1	80	1
$P_1(z)$	0	370.37	15.36	3.18	15.36	1.51
$P_2(z)$	0	370.37	41.23	1.04	41.23	1

The results in Table 3.3 show that both the *Y* chart and the *U* chart are very sensitive to the process faults, while the *W* chart is completely insensitive to the process faults because its ARL for detection a process fault is exactly equal to the in-control ARL. Because the *Y* chart is sensitive to both the sensor faults and the process faults, it cannot be used directly to differentiate these two types of faults. The *W* chart followed by

the usage of the U chart, however, can be used to effectively distinguish sensor faults from process faults.

Next we look at the double-sensor-faults investigated in Section 3.2.3. From Figure 2.1, it can be seen that some single physical sensors (e.g., M_1) are associated with two sensor measurements (e.g., $M_1(x)$ and $M_1(z)$). Therefore, a failure at such a sensor may cause simultaneous faults on those two sensor measurements, which results in a double-sensor-fault. In this example, we focus on the double-sensor-faults at sensors M_i, i=1, 2, 3, 4, and set the mean shift at each sensor measurement as equal to 0.2/6 mm. For example, the double-sensor fault at sensor M_1 causes mean shifts of 0.2/6 mm at both $M_1(x)$ and $M_1(z)$. In Table 3.4 we show for each of the double-sensor-faults on sensors M_i, i=1, 2, 3, 4, the values of the noncentrality parameter λ_w, the sensitivity ratio SR, the largest eigenvalue (λ_L) and the smallest eigenvalue (λ_S) of the 2-by-2 submatrix given in equation (3.17), and the corresponding eigenvectors ($\mathbf{e}_1$ and $\mathbf{e}_2$).

Note from Table 3.4 that the sensitivity ratio SR for each of the four double-sensor-faults is between the largest eigenvalue λ_L and the smallest eigenvalue λ_S. In this example, the SR is calculated by setting μ_i / μ_j=1. For other ratios of μ_i / μ_j, the values of SR will be different from those in the table. But for any ratio of μ_i / μ_j, the value of the sensitivity ratios always falls in the range of [λ_S, λ_L] based on Result 2.

Table 3.4 Sensitivity results for double-sensor-faults

	M_1	M_2	M_3	M_4
λ_w	66.481	68.286	48.76	63.38
SR	0.831	0.854	0.609	0.792
λ_L	0.872	0.872	0.875	0.875
λ_S	0.717	0.833	0.346	0.702
$\mathbf{e}_1$	$[0.97\ 0.24]^T$	$[0.9997\ 0.0231]^T$	$[1\ 0]^T$	$[0.9998\ 0.0203]^T$
$\mathbf{e}_2$	$[-0.24\ 0.97]^T$	$[-0.0231\ 0.9997]^T$	$[0\ 1]^T$	$[-0.0203\ 0.9998]^T$

Table 3.5 ARL comparison of *W* chart and *Y* chart in detection of double-sensor-faults

Mean shift in two sensors	*W* chart				*Y* chart	
	M_1		M_2			
	λ_w	ARL	λ_w	ARL	λ_y	ARL
0	0	370.37	0	370.37	0	370.37
$0.2\sigma_w$	2.66	62.18	2.73	60.03	3.20	57.47
$0.4\sigma_w$	10.64	5.37	10.93	5.09	12.82	4.51
$0.6\sigma_w$	23.93	1.42	24.58	1.38	28.80	1.28
$0.8\sigma_w$	42.55	1.02	43.70	1.02	51.19	1.01
σ_w	66.48	1	68.29	1	79.99	1

In Table 3.5, we compare the noncentrality parameters and ARL of the *W* chart and the *Y* chart when a double-sensor-fault occurs at sensor M_1 or sensor M_2. The magnitudes of mean shifts in both *x*- and *z*- directions are set to be equal. The *W* chart and the *Y* chart have close performance in detection of sensor faults as for single-sensor-faults shown in Table 3.5. Again, the advantage of the *W* chart over the *Y* chart is that the *W* chart is only sensitive to sensor faults and therefore can be used to distinguish sensor faults from process faults, while the *Y* chart cannot differentiate sensor faults and process faults.

CHAPTER 4.
A BAYESIAN METHOD FOR AUTOMATIC DIAGNOSIS OF MANUFACTURING PROCESS AND SENSOR FAULTS

Although the multivariate control charts developed in Chapter 2 and Chapter 3 can help us detect process faults and sensor faults and differentiate sensor faults from process faults, further diagnosis work is needed to identify which tooling elements or which sensors are responsible for the abnormal sensor measurements. Automatic fault diagnosis can save tremendous costs and resources required to inspect root causes in a complex manufacturing process with a great number of tooling elements and sensors.

Fault diagnosis is usually conducted after faults have already been detected using a quality monitor technique, such as a multivariate control chart on measurements $\mathbf{y}_j$. Therefore, in this chapter, we assume that occurrence of process and/or sensor faults has been detected. The goal of the diagnosis procedure is to identify the faulty sensors and/or faulty tooling elements. That is, the nonzero elements in $\mathbf{u}$ and $\boldsymbol{\mu}_\mathbf{w}$ in model (1.1), which are root causes of the out-of-control signal, should be identified.

First, we combine the potential process faults and sensor faults together as a single fault vector $\boldsymbol{\beta} = [\mathbf{u}^T \quad \boldsymbol{\mu}_\mathbf{w}^T]^T$ and rewrite model (1.1) as

$$\mathbf{y}_j = \mathbf{C}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_j, \qquad j = 1, 2, \cdots, N \tag{4.1}$$

where $\mathbf{C}^* = [\mathbf{C} \quad \mathbf{I}_n]$. For the fault-quality model, the number of potential faults, which is the dimension of $\boldsymbol{\beta}$, will always exceed the number of measurements $\mathbf{y}_j$ by p – equivalent to the number of process faults, because besides the process faults, one potential sensor fault is associated with each measurement point. This causes difficulty in estimating $\boldsymbol{\beta}$ in model (4.1) by applying linear regression algorithm since $\mathbf{C}^{*T}\mathbf{C}^*$ is not a full rank matrix and the unbiased estimator of $\boldsymbol{\beta}$ doesn't exist.

Typically in moderately reliable systems, just a few process and/or sensor faults may occur simultaneously. Consequently, it is reasonable to believe that only a few

components of $\boldsymbol{\beta}$ has nonzero values. Due to this observation, we can treat the fault diagnosis as a variable selection problem, or subset selection problem as is often called. Many researchers have proposed approaches for variable selection. In this chapter, we will investigate a Bayesian variable selection method for diagnosis of both process faults and sensor faults.

4.1 Literature Review of Bayesian Variable Selection

A variety of variable selection approaches have been studied in statistics literature. Some approaches, like AIC, BIC and C_p, select the optimal subset of the explanatory variables by comparing the 2^{n+p} possible submodels (here $n+p$ is the total number of potential faults as specified in model (4.1). When $n+p$ is large, finding the best model among all candidates by exhaustive calculation is impossible. Researchers then apply heuristic stepwise selection, like forward selection, backward elimination, etc, to focus on a smaller number of potential subsets and identify a promising submodel.

George and McCulloch (1993) proposed a Bayes procedure called SSVS (Stochastic Search Variable Selection) to identify the promising subsets by specifying a Bayesian hierarchical model and applying Gibbs sampling algorithm to approximate the posterior probability of the submodels. The models showing up with more frequency in the simulation are considered as promising subsets. They mentioned that the Bayes hierarchical model was based on the work including Lempers (1971), Atkinson (1978), Perrichi (1985), Smith and Spiegelhalter (1980), Spiegelhalter and Smith (1982), Zellner (1984), Poirier (1985), Stewart (1987), and Mitchell and Beauchamp (1988). Other works in the past supporting their approach concern Gibbs sampling,such as Gelfand and Smith (1990), Gelfand et al (1990), Tanner and Wong (1987), and Verdinelli and Wasserman (1991).

Further study related to Bayesian variable selection were done by Carlin and Chib (1995), Chipman (1995), Clyde, DiSimone, and Parmigiani (1996), George and

McCulloch (1995), George, McCulloch and Tsay (1995), Geweke (1996), Hoeting, Raftery and Madigan (1997), Kuo and Mallick (1994), Meehan, Dempster and Brown (1994), Phillips and Smith (1995), Raftery, Madigan and Hoeting (1993), Raftery, Madigan and Volinsky (1995), Smith and Kohn (1995), and Wakefiel and Bennett (1996). George and McCulloch (1997) summarize and compared the performance based on various prior formulations of Bayesian variable selection. George and Foster (2000) showed the equivalence between the variable selection criterion such as AIC, BIC, C_p, and RID and specific values of hyperparamters in a Bayesian hierarchical model, and proposed an empirical Bayes variable selection procedure. Yuan and Lin (2005) applied a LASSO algorithm to compute the empirical Bayes estimate and provided a new way to select the tuning parameters in the LASSO method.

The Bayesian variable selection procedure has also been applied in the areas of experimental design. Chipman, Hamada and Wu (1997) used the Bayesian approach to analyze the designed experiments with complex aliasing pattern. Joseph (2006) also applied the Bayesian variable selection in experiment design and analysis and presented a new way to specify the hyperparameters in the priors of the Bayes model.

4.2 Bayesian Fault Diagnosis Procedure

4.2.1 Hierarchical Prior

To apply the Bayesian variable selection approach to fault diagnosis, we first need to determine the prior distributions for the parameters $\boldsymbol{\beta}$ and σ^2 in model (4.1), such that a hierarchical Bayes model can be set up. George and McCulloch (1997) presented a set of conjugate priors for Bayesian variable selection problem. We apply their hierarchical prior setup for model (4.1) as follows:

Let γ_i (i=1, 2, …, n+p) be an indicator variable of the ith component of $\boldsymbol{\beta}$. The prior distribution of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma},\sigma^2) = N_{q_{\boldsymbol{\gamma}}}\left(c\sigma^2\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\right) \text{ and } \pi(\boldsymbol{\beta}_{\bar{\boldsymbol{\gamma}}} = 0|\boldsymbol{\gamma},\sigma^2) \equiv 1$$

where $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is the subset of $\boldsymbol{\beta}$ corresponding to the components of $\boldsymbol{\gamma}$ equal to 1, while $\boldsymbol{\beta}_{\bar{\boldsymbol{\gamma}}}$ is the subset of $\boldsymbol{\beta}$ corresponding to the components of $\boldsymbol{\gamma}$ equal to 0. As we have assumed, σ^2 is the variance of noise term in model (4.1); $\mathbf{C}_{\boldsymbol{\gamma}}^{*}$ is the sub-matrix of $\mathbf{C}^{*}$ with its columns picked from $\mathbf{C}^{*}$ according to $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$. c is a hyperparameter that needs to be specified. The correlation structure $\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}$ is equivalent to the generalization of g prior of Zellner (1986). The nonzero components of variable $\boldsymbol{\gamma}$ decide which components of $\boldsymbol{\beta}$, β_i, i=1, …, n+p, should be included in the model.

The prior distribution for the indicator variable $\boldsymbol{\gamma} = [\gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_{n+p}]^T$ is assumed to have binomial distribution with parameter w_i, which can be written as

$$\pi(\gamma_i) \sim Bin(w_i) \tag{4.2}$$

Assume all the components of $\boldsymbol{\gamma}$ are independent and the joint prior of $\boldsymbol{\gamma}$ is

$$\pi(\boldsymbol{\gamma}) = \prod_{i=1}^{p+n} w_i^{\gamma_i}(1-w_i)^{1-\gamma_i} \tag{4.3}$$

From the perspective of fault diagnosis, γ_i can be understood as an indicator of the occurrence of the ith fault in the system. $\gamma_i = 0$ indicates that the ith fault doesn't occur. At the beginning of the fault diagnosis procedure, w_i, the probability a fault occurs on the ith sensor, is determined based on a subjective evaluation about the reliability of ith sensor. In Section 4.2.3, we will give more discussions about how to choose the value of w_i and other hyperparameters. In addition, George and McCulloch (1997) set the prior distribution for σ^2 as the following inverse gamma distribution:

$$\pi(\sigma^2) = IG\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right) \tag{4.4}$$

where ν and λ are two hyperparameters. The physical interpretation of ν and λ will also be given in Section 4.2.3.

A special issue in our fault-diagnosis problem is that, because $\mathbf{C}^{*}$ is not of full column rank, there will always be some fault combinations causing the same abnormal observations. For example, suppose a system can be modeled by equation (4.1) with

$$\mathbf{C}^* = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

This matrix indicates one potential process fault (β_1, the component of fault vector $\boldsymbol{\beta}$ corresponding to the first column of $\mathbf{C}^*$) and three sensor mean shift faults (β_2, β_3, β_4, the components of fault vector $\boldsymbol{\beta}$ corresponding to the last three columns of $\mathbf{C}^*$) will affect the product quality characteristics $\mathbf{y}(j)$ in this system. There would be infinitely many number of fault patterns that can cause a mean shift of the measurements $E[\mathbf{y}(j)] = [\delta \quad \delta \quad \delta]^T$. Among those fault patterns, one case is the occurrence of only process fault $\boldsymbol{\beta}(1) = [\delta \quad 0 \quad 0 \quad 0]^T$. Another one is $\boldsymbol{\beta}(2) = [0 \quad \delta \quad \delta \quad \delta]^T$, which indicates no process fault occurs and the three sensor faults occur simultaneously with all the mean shifts equal to δ. We call fault {1}, which corresponds to the first column of $\mathbf{C}^*$, and the set of faults {2, 3, 4} corresponding to the other columns of $\mathbf{C}^*$, as coupled faults. Essentially, the reason that these faults are coupled is that their corresponding columns of $\mathbf{C}^*$ are linearly dependent. It can be seen that the fault index is the same as the index of the corresponding columns of $\mathbf{C}^*$.

Before we solve the singularity problem, some definitions are first given:

A *Fault Pattern* is defined as $FP \equiv \{i: \beta_i \neq 0, i = 1, 2, \cdots, n+p\}$. A fault pattern indicates a candidate model with only the variables included in the fault pattern being selected into the model while the rest of the variables considered as zeros and excluded from the model. There are totally 2^{n+p} different fault patterns. The model corresponding to fault pattern FP_m can be denoted based on the indicator vector

$$\boldsymbol{\gamma}(FP_m) = [\gamma_1 \quad \cdots \quad \gamma_i \quad \cdots \quad \gamma_{n+p}]^T = \begin{cases} \gamma_i = 1, i \in FP_m \\ \gamma_i = 0, i \notin FP_m \end{cases}$$

A *Coupled Fault Pattern* (*CFP*) is considered as the fault pattern whose corresponding columns of $\mathbf{C}^*$ are linearly dependent. Correspondingly, we define the set of models that includes all the *CFP*s as *Coupled Model Set* (*CMS*), which can be denoted as $\boldsymbol{\Gamma}_{\text{CMS}} \equiv \{\boldsymbol{\gamma}(CFP_1), \boldsymbol{\gamma}(CFP_2), \cdots, \boldsymbol{\gamma}(CFP_K)\}$.

The singularity problem can then be solved by artificially adding a constraint that coupled fault patterns are not allowed to show up in the Bayes model. For the above example, we allow fault pattern {1} occurs and fault pattern {2, 3, 4} occurs, but set the chance of the occurrence of the coupled fault pattern {1, 2, 3, 4} as zero.

The coupling of the faults can be checked easily based on the linear dependence of the corresponding columns in $\mathbf{C}^*$. Formally, a restriction on the prior of $\pi(\boldsymbol{\gamma})$ is added by revising equation (4.3) as:

$$\pi(\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_{\text{CMS}}) = 0$$

and for $\boldsymbol{\gamma} \notin CMS$,

$$\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}}) = a \prod_{i=1}^{p+n} w_i^{\gamma_i}(1-w_i)^{1-\gamma_i} \propto \prod_{i=1}^{p+n} w_i^{\gamma_i}(1-w_i)^{1-\gamma_i} \qquad (4.5)$$

where the notation $\propto$ means $\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}})$ is proportional to the value of $\prod_{i=1}^{p+n} w_i^{\gamma_i}(1-w_i)^{1-\gamma_i}$. The constant a is obtained as

$$a = 1/\left(1 - \sum_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_{\text{CMS}}} \prod_{i=1}^{p+n} w_i^{\gamma_i}(1-w_i)^{1-\gamma_i}\right)$$

The conjugate joint prior of $\boldsymbol{\beta}$ corresponding to $\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}}$ is:

$$\pi\left(\boldsymbol{\beta}_{\boldsymbol{\gamma}} | \boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}}, \sigma^2\right) = N_{q_{\boldsymbol{\gamma}}}\left(0, c\sigma^2\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*\text{T}}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\right) \text{ and } \pi\left(\boldsymbol{\beta}_{\bar{\boldsymbol{\gamma}}} = 0 | \boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}}, \sigma^2\right) = 1 \quad (4.6)$$

The prior for σ^2 is still given by

$$\pi(\sigma^2) = IG\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right).$$

We will have a further discussion about the coupled model set CMS and the related concepts in Section 4.3.

4.2.2 Calculation of Posterior $\boldsymbol{\pi(\gamma|Y)}$

We define the vector of quality characteristics measurements from all products as $\mathbf{Y} \equiv [\mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_N]^{\text{T}}$. Given the conjugate priors in Section 4.2.1, the posterior probability for $\boldsymbol{\gamma}|\mathbf{Y}$ can be derived explicitly as

$$\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\mathrm{CMS}}|\mathbf{Y}) \propto g(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\mathrm{CMS}})$$

$$= (1+c)^{-q_\gamma/2}\left(\upsilon\lambda + \mathbf{Y}^T\mathbf{Y} - \frac{1}{1+1/c}\mathbf{W}^T\mathbf{W}\right)^{-(Nn+\upsilon)/2}\pi(\boldsymbol{\gamma} \notin CMS) \tag{4.7}$$

where $g(\cdot)$ is a function proportional to the p.m.f $\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\mathrm{CMS}}|\mathbf{Y})$ and

$$\mathbf{W} = \frac{\mathbf{1}}{\sqrt{N}}(\mathbf{Q}^{\mathrm{T}})^{-1}\mathbf{C}_{\boldsymbol{\gamma}}^{*\mathrm{T}}\sum_{j=1}^{N}\mathbf{y}(j)$$

$\mathbf{Q}$ is an upper triangular matrix satisfying $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{C}_{\boldsymbol{\gamma}}^{*\mathrm{T}}\mathbf{C}_{\boldsymbol{\gamma}}^{*}$ and can be obtained by the Cholesky decomposition. Particularly, when $\boldsymbol{\gamma} = \mathbf{0}$

$$\pi(\boldsymbol{\gamma} = \mathbf{0}|\mathbf{Y}) \propto g(\boldsymbol{\gamma} = \mathbf{0}) = (\upsilon\lambda + \mathbf{Y}^T\mathbf{Y})^{-(N(p+n)+\upsilon)/2}\pi(\boldsymbol{\gamma} = 0) \tag{4.8}$$

We also have

$$\pi(\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_{\mathrm{CMS}}|\mathbf{Y}) = g(\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_{\mathrm{CMS}}) \equiv 0 \tag{4.9}$$

The follows are the derivation of posterior $\pi(\boldsymbol{\gamma}|\mathbf{Y})$ in equations (4.7), (4.1) and (4.9).

We first assume the sample size of measurements N=1. According to Model (4.2), $\mathbf{Y}\sim\boldsymbol{N}(\mathbf{C}^*\boldsymbol{\beta}, \sigma^2\mathbf{I}_{n+p})$. This likelihood function can then be written as

$$\pi(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{C}^*\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{C}^*\boldsymbol{\beta})\right)$$

For convenience, we rewrite the conjugate priors of parameters $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and σ^2 as follows:

$$\pi(\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_{\mathrm{CMS}}) = 0 \text{ and } \pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\mathrm{CMS}}) \propto \prod_{i=1}^{p+n} w_i^{\gamma_i}(1-w_i)^{1-\gamma_i}$$

$$\pi(\sigma^2) = IG\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right)$$

$$\pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \sigma^2) = N_{q_\gamma}\left(0, c\sigma^2\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\right)$$
$$= \frac{1}{(2\pi)^{q_\gamma/2}\left|c\sigma^2\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\right|^{1/2}}exp\left\{-\frac{1}{2c\sigma^2}\boldsymbol{\beta}_{\boldsymbol{\gamma}}^T\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right\}$$

and

$$\pi\left(\boldsymbol{\beta}_{\bar{\boldsymbol{\gamma}}} = \mathbf{0}|\boldsymbol{\gamma}, \sigma^2\right) \equiv 1$$

where CMS is the *Coupled Model Set.*

Based on the artificial constraint that the coupled faults cannot occur at the same time, it is easy to see that $\pi(\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_{\text{CMS}}) = 0$. So we only focus on the probabilities for $\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}})$. By multiplying the above equations,

$$\pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \sigma^2|\mathbf{y}) \propto \sigma^{-(n+q_{\boldsymbol{\gamma}}+\nu+1)} \left|c\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\right|^{-\frac{1}{2}}$$

$$\exp[\left\{-\frac{1}{2\sigma^2}\left[\left(\mathbf{y}-\mathbf{C}_{\boldsymbol{\gamma}}^{*}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right)^T\left(\mathbf{y}-\mathbf{C}_{\boldsymbol{\gamma}}^{*}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right)+\boldsymbol{\beta}_{\boldsymbol{\gamma}}^T\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)\boldsymbol{\beta}_{\boldsymbol{\gamma}}\right]-\frac{\lambda\nu}{2\sigma^2}\right\}\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}})$$

$$= \sigma^{-(n+q_{\boldsymbol{\gamma}}+\nu+1)}\left|c\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\right|^{-\frac{1}{2}}\exp\left(-\frac{\lambda\nu}{2\sigma^2}\right)$$

$$\exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}_{\boldsymbol{\gamma}}-\frac{c}{\mathbf{1}+c}\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{y}\right)^T\frac{C+1}{C}\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)\left(\boldsymbol{\beta}_{\boldsymbol{\gamma}}-\frac{c}{\mathbf{1}+c}\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{y}\right)\right\}$$

$$\exp\left\{-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^2}\frac{c}{\mathbf{1}+c}\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{y}\right\}\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}})$$

Integrating out $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and σ^2, we can obtain

$$\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}}|\mathbf{y}) \propto g(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}})$$
$$= (1+c)^{-q_{\boldsymbol{\gamma}}}\left(\nu\lambda+\mathbf{y}^T\mathbf{y}-\frac{1}{1+1/c}\mathbf{y}^T\mathbf{C}_{\boldsymbol{\gamma}}^{*}\left(\mathbf{C}_{\boldsymbol{\gamma}}^{*\mathrm{T}}\mathbf{C}_{\boldsymbol{\gamma}}^{*}\right)^{-1}\mathbf{C}_{\boldsymbol{\gamma}}^{*T}\mathbf{y}\right)^{-(n+\nu)/2}\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}})$$

Considering sample size $N \geq 1$, we replace $\mathbf{y}$, $\mathbf{C}_{\boldsymbol{\gamma}}^{*}$ and n by $\mathbf{Y}$, $\left[\mathbf{C}_{\boldsymbol{\gamma}}^{*T} \quad \mathbf{C}_{\boldsymbol{\gamma}}^{*T} \quad \cdots \quad \mathbf{C}_{\boldsymbol{\gamma}}^{*T}\right]^T$ and Nn respectively. Then,

$$\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}}|\mathbf{y}) \propto g(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}})$$
$$= (1+c)^{-q_{\boldsymbol{\gamma}}}\left(\nu\lambda+\mathbf{Y}^T\mathbf{Y}-\frac{1}{1+1/c}\mathbf{W}^T\mathbf{W}\right)^{-(Nn+\nu)/2}\pi(\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}_{\text{CMS}})$$

where $\mathbf{W} = \frac{1}{\sqrt{N}}\mathbf{T}'^{-1}\mathbf{C}_{\boldsymbol{\gamma}}^{*\prime}\sum_{j=1}^{N}\mathbf{y}(j)$ and $\mathbf{T}'\mathbf{T} = \mathbf{C}_{\boldsymbol{\gamma}}^{*\prime}\mathbf{C}_{\boldsymbol{\gamma}}^{*}$ for $\mathbf{T}$ upper triangular. $\mathbf{T}$ can be obtained by the Cholesky decomposition. $q_{\boldsymbol{\gamma}}$ is the total number of the nonzero components of $\boldsymbol{\gamma}$.

Based on equation (4.7), exhaustive calculation of $\pi(\boldsymbol{\gamma}|\mathbf{y})$is feasible for small-sized problems (e.g., when the total number of potential faults $p+n \leq 25$). By calculating $g(\boldsymbol{\gamma})$ for each possible value of $\boldsymbol{\gamma}$, the normalization constant can be obtained as

$$G = \sum g(\boldsymbol{\gamma}) \tag{4.10}$$

and

$$\pi(\boldsymbol{\gamma}|\mathbf{Y}) = g(\boldsymbol{\gamma})/G \tag{4.11}$$

Since there are 2^{p+n} possible values of $\boldsymbol{\gamma}$, for large sized problem (p+n>25), it is time-consuming or even infeasible to calculate $g(\boldsymbol{\gamma})$ for each value of $\boldsymbol{\gamma}$ to obtain $\pi(\boldsymbol{\gamma}|\mathbf{Y})$. The major issue in fault diagnosis is to see the first few fault patterns with high posterior probability. Therefore, we can apply Markov Chain Monte Carlo (MCMC) simulation method to approximate$\pi(\boldsymbol{\gamma}|\mathbf{Y})$. By generating a Markov chain

$$\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, \ldots, \boldsymbol{\gamma}^{(k)} \tag{4.12}$$

The values of $\boldsymbol{\gamma}$ corresponding to a relatively large $\pi(\boldsymbol{\gamma}|\mathbf{Y})$ will appear with high frequency in the above simulation sequence.

We apply the Metropolis-Hastings (M-H) algorithm to generate a Markov chain. Under the simplest symmetric proposal, the M-H algorithm will move faster than the Gibbs sampler. And Liu (1995) showed that the M-H algorithm is superior to the Gibbs sampler under the asymptotic variance criterion of Peskun (1995). The details of the M-H algorithm can be referred to George and McCulloch (1997). The basic steps of the M-H algorithm are presented here for convenience.

The Metropolis-Hasting algorithm to generate a Markov chain (4.12) from $g(\boldsymbol{\gamma})$ in equations (4.1) and (4.10) is summarized as follows:

1. Generate $\boldsymbol{\gamma}^{(0)}$ by randomly generating 0 or 1 for each component of $\boldsymbol{\gamma}^{(0)}$. Set $k = 0$.

2. Generate a candidate $\boldsymbol{\gamma}^*$ by randomly changing a component of $\boldsymbol{\gamma}^{(k)}$.

3. Let $\boldsymbol{\gamma}^{(k+1)} = \boldsymbol{\gamma}^*$ with the probability

$$\alpha = min\left\{\frac{g(\boldsymbol{\gamma}^*)}{g(\boldsymbol{\gamma}^{(k)})}, 1\right\}$$

Otherwise, $\boldsymbol{\gamma}^{(k+1)} = \boldsymbol{\gamma}^{(k)}$.

4. $k = k + 1$. Return to step 2.

4.2.3 Choice of Additional Hyperparameters

In this section, we will show how to choose the additional hyperparameters ν, λ, w_i, c in the equations of priors (4.4), (4.5) and (4.6).

We first consider the prior probability w_i for $\gamma_i = 1$ (i= 1, 2, …, n+p) in equation (4.5). We recommend a way to choose w_i as: assume all the potential faults would occur with equal probability, that is $w_1 = w_2 = \cdots = w_{p+n} = w$. Since we assume that a mean shift of sensor measurements has already been detected based on a multivariate control chart and such a mean shift is caused by process faults, sensor faults, or both, it is reasonable to assume that

$$\Pr(E_1 \cup E_2 \cup \ldots \cup E_{n+p}) = 1$$

where Pr(·) is the probability of the occurrence of the events in the parenthesis. Event E_i, i=1, 2, ⋯, n+p denotes the occurrence of the ith fault (faulty tooling elements or sensors). Further applying the first-order inclusion-exclusion approximation, we have

$$1 = \Pr(E_1 \cup E_2 \cup \cdots \cup E_{n+p}) \approx \Pr(E_1) + \Pr(E_2) + \cdots + \Pr(E_{n+p})$$

$$= w_1 + w_2 + \cdots + w_{n+p} = (n+p)w$$

By solving the above approximation, we obtain $w \approx \frac{1}{n+p}$. In the rest of the study, the values of w_i (i=1, 2, …, n+p) are set as discussed above and are substituted by w.

As for the hyperparameter c used to define the prior distribution of $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ given in equation (4.6), we use a fixed value c=100 in all of our examples. We also conduct a sensitivity analysis for parameter c in Section 4.3.4. George and Foster (2000) applied both marginal and conditional maximum likelihood (CML) method to estimate the hyperparameters c and w. We recommend trying out different values of c in pilot experiments before the real diagnosis procedure such that a reasonable value can be obtained. In addition, CML method, which we will briefly introduce in Section 4.2.4, can be applied to calculate c and w. CML provides a nice quantitative guideline in selecting

the value of the hyperparameters. But as we can see in Section 4.3.4, it is not necessarily optimal in terms of the diagnosis performance.

The hyperparameter ν in equation (4.4) can be understood as the number of observations used to estimate the prior of σ^2. And it is easy to see that the mean of the prior distribution of σ^2 is $\nu\lambda/(\nu-2)$. Therefore, if historical data is used to estimate σ^2, ν and λ may be determined as follows: Denote $\hat{\sigma}^2$ as the estimation of σ^2 based on the historical data. We can set ν as the number of observations in the historical data and $\nu\lambda/(\nu-2)$ as equal to $\hat{\sigma}^2$. Therefore, λ can be obtained as $\nu\lambda/(\nu-2)$.

In summary, the whole diagnosis procedure can be demonstrated in Figure 4.1 : Firstly, a fault-quality model as of the form in Model (1.1) needs to be set up, where matrix **C** is obtained based on knowledge of the process and product design knowledge. Secondly, the hyperparameters ν, λ, w, c used in the prior distributions in equations (4.4), (4.5) and (4.6) should be selected. If expertise knowledge or experience is lacking, historical data and pilot experiments should be applied to select values of those hyperparameters according to the guidelines presented in this section. Thirdly, under the hierarchical model setting in Chapter 4, computation is conducted to calculate the posterior probabilities for different models based on equations (4.7) and (4.9). Essentially, the diagnosis problem can be considered as selecting the model (fault pattern) with the largest posterior probability. The exhaustive posterior calculation for each model can be done when the total number of potential faults $p+n<25$. Otherwise, MCMC is applied to approximate the posteriors of the models. Finally, inspection ofthe tooling elements and sensors should be conducted according to the diagnosis results to fix the faulty elements.

4.2.4 CML Method

As it can be seen, the guidelines in choosing the values of hyperparameters, such as c, given in the previous section still rely on subjective judgments. George and Foster

(2000) proposed a quantitative approach to estimate the value of c and w using the observed measurement data. Specifically, under the hierarchical model presented in Section 4.2.1, the posterior $p(\boldsymbol{\gamma}|\mathbf{Y}, c, w)$ is

$$p(\boldsymbol{\gamma}|\mathbf{Y}, c, w) \propto w^{q_\gamma}(1-w)^{p-q_\gamma}(1+c)^{-q_\gamma/2}\exp\left\{\frac{css_\gamma}{2\sigma^2(1+c)}\right\} \tag{4.13}$$

where

$$ss_\gamma = \frac{1}{N}\left(\sum_{j=1}^{N}\mathbf{y}^{\mathrm{T}}(j)\mathbf{C}^{*\mathrm{T}}\right)\left(\mathbf{C}^{*\mathrm{T}}\mathbf{C}^{*}\right)^{-1}\left(\sum_{j=1}^{N}\mathbf{C}^{*}\mathbf{y}(j)\right)$$

and N is the sample size.

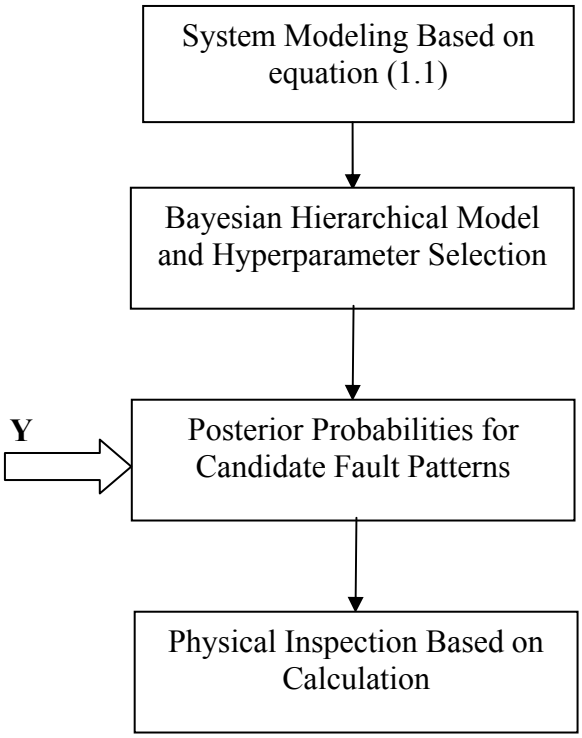


Figure 4.1 Bayesian fault diagnosis diagram

The posterior (4.13) can be considered as a function of c and w. And this function is called as the conditional 'likelihood' function $L^*(c, w, \boldsymbol{\gamma}|\mathbf{Y}) \equiv p(\boldsymbol{\gamma}|\mathbf{Y}, c, w)$. It is not a

strict likelihood function since $\boldsymbol{\gamma}$ is an unknown intermediate vector. Conditioning on $\boldsymbol{\gamma}$, the estimators of *c* and *w* are obtained as follows by maximizing the conditional likelihood function in equation (4.13).

$$\hat{c}_{\gamma} = \left(ss_{\boldsymbol{\gamma}}/\sigma^2 q_{\gamma} - 1\right)_{+} \tag{4.14}$$

and

$$\hat{w}_{\boldsymbol{\gamma}} = q_{\boldsymbol{\gamma}}/p \tag{4.15}$$

where$(\cdot)_+$ is the positive part function, that is $(\cdot)_+ = (\cdot)$ when $(\cdot)$ is a positive number, otherwise, $(\cdot)_+ = 0$. $\hat{c}_{\gamma}$ and $\hat{w}_{\boldsymbol{\gamma}}$ are the estimators of *c* and *w* for a specific model corresponding to $\boldsymbol{\gamma}$. That is, we may have different estimates of *c* and *w* for each candidate model and the hyperparameters *c* and *w* are estimated in terms of $\boldsymbol{\gamma}$. By substituting the two estimates in equations (4.14) and (4.15) into equation (4.13), the problem becomes searching for the best model such that the posterior $p\left(\boldsymbol{\gamma}|\mathbf{Y}, \hat{c}_{\boldsymbol{\gamma}}, \hat{w}_{\boldsymbol{\gamma}}\right)$ reaches its maximum. The objective function then is equivalent to the following simpler function:

$$C_{CML} = ss_{\boldsymbol{\gamma}}/\sigma^2 - B\left(ss_{\boldsymbol{\gamma}}/\sigma^2\right) - R\left(q_{\boldsymbol{\gamma}}\right)$$

where

$$B\left(ss_{\boldsymbol{\gamma}}/\sigma^2\right) = q_{\gamma}\left\{1 + \log_{+}\left(ss_{\gamma}/\sigma^2 q_{\gamma}\right)\right\}$$

$\log_+(\cdot)$ is the positive part of $\log(\cdot)$, that is, $\log_+(\cdot) = \log(\cdot)$ when $\log(\cdot)>0$; $\log_+(\cdot) = 0$ otherwise; and

$$R\left(q_{\boldsymbol{\gamma}}\right) = -2\left\{\left(p - q_{\boldsymbol{\gamma}}\right)\log\left(p - q_{\boldsymbol{\gamma}}\right) + q_{\boldsymbol{\gamma}}\log q_{\boldsymbol{\gamma}}\right\}.$$

Suppose $\boldsymbol{\gamma}^*$ represents the best model that maximizes C_{CML}, then the estimates of *c* and *w* are $\hat{c}_{\boldsymbol{\gamma}^*}$ and $\hat{w}_{\boldsymbol{\gamma}^*}$ respectively. To search for the optimal model $\boldsymbol{\gamma}^*$, the exhaustive calculation of $ss_{\boldsymbol{\gamma}}$ for every candidate model should be conducted. But with a large number of potential root causes, a heuristic strategy such as stepwise search can be applied to identify a locally optimal model. From the above discussions, the CML method can be considered as an alternative way to estimate the values of *c* and *w*.

4.3 Simulation Examples

Since the potential faults always outnumber the measurement points in our studies, even for Bayesian variable selection method, the singularity problem caused by coupled faults still exist. In Section 4.2.1, we add an artificial constraint to the prior of $\boldsymbol{\gamma}$ to force the coupled faults not to show up simultaneously. This diagnosis procedure turns out to work well in the moderately reliable systems with sparse faults. We still conduct studies to see how the coupled structure would affect the diagnosis result. Some concepts are first introduced.

Minimal Coupled Pattern (MCP): MCP is a coupled fault pattern (CFP) with the following additional property: if any element is excluded from an MCP, the columns of $\mathbf{C}^*$ corresponding to the rest of its elements are linearly independent. The degree of MCP is the number of elements in an MCP. For instance, for a system with

$$\mathbf{C}^* = \begin{bmatrix} 1 & 4 & 5 & 1 & 0 & 0 & 0 \\ 3 & 2 & 5 & 0 & 1 & 0 & 0 \\ 2 & 3 & 5 & 0 & 0 & 1 & 0 \\ 5 & 1 & 6 & 0 & 0 & 0 & 1 \end{bmatrix}$$

the first three columns of $\mathbf{C}^*$ are linearly dependent and any two of the three columns are linearly independent. Therefore, one *MCP* for the system is MCP_1={1, 2, 3} and MCP_1 has degree 3. Furthermore, another MCP in the system is MCP_2={1, 4, 5, 6, 7} and the degree of MCP_2 is 5. It can be seen from this example that the degrees may vary for different MCPs.

MCPs and the corresponding degrees describe the coupled fault structure of a system. Therefore, we study the influence of coupled fault on the diagnosis accuracy by finding the relationship between MCP and diagnosis results through simulated examples. The Bayesian variable selection approach to identify the root causes is also illustrated in the example. For choosing the values of hyperparameters including *c* and *w* in the Bayesian hierarchical model, we following the guidelines in Section 4.2.1. The diagnosis performance is compared with that of applying CML method in Section 4.3.4.

These simulated examples are designed based on a system's MCP degrees. We assume $\mathbf{y}_j$ are generated from $N_n(\mathbf{C}^*\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$ with $\sigma^2 = 1$ for all the examples without loss of generality, where n is the number of measurements obtained from one product and is equal to the number of potential sensor faults. While the matrix $\mathbf{C}^*$ and the fault vectors $\boldsymbol{\beta}$ may not be the same for different examples, the hyperparameters are set with the same values as listed in Table 4.1, where (n+p) is the total number of faults including both process faults and sensor faults.

Table 4.1 Hyperparameter settings

Sample Size N	C	λ	ν	w
20	100	1.5	10	1/(n+p)

4.3.1 Minimal Coupled Pattern of Degree7

For the first example, assume six product quality characteristics on each product need to be measured by six sensors in the system and two potential process faults would affect the product quality. So there are totally eight potential faults – two process faults and six sensor faults. Correspondingly, the $\mathbf{C}^*$ matrix in Model (4.2) is $\mathbf{C}^* = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{I}_6]$. We randomly generate the two columns $\mathbf{c}_1$ and $\mathbf{c}_2$ from normal distribution $N_6(0, \boldsymbol{I}_6)$. The mean shift is set as $\boldsymbol{\beta} = [1.5 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0]^T$, which means the mean shift faults occur on the first and the sixth components of $\boldsymbol{\beta}$ which correspond to the first tooling element and the fourth sensor, respectively. Furthermore, it can be seen that fault 1 and fault 6 belong to a minimal coupled pattern (MCP) {1, 3, 4, 5, 6, 7, 8} with degree 7.

Since this is a relatively small size problem with eight potential faults, we may exhaustively calculate $\pi(\boldsymbol{\gamma}|\mathbf{Y})$. Based on equations (4.3), (4.4), (4.5), (4.6) and (4.7), we

calculate the posterior $\pi(\boldsymbol{\gamma}|\mathbf{Y})$ 1000 times for 1000 randomly generated sets of measurements $\mathbf{Y}$ and then obtain the average $\pi(\boldsymbol{\gamma}|\mathbf{Y})$. The first four models (fault patterns) with the largest posterior probabilities in average are listed in Table 4.2. The first column represents the fault patterns or candidate models. The fault indices mark which components in $\boldsymbol{\gamma}$ are nonzero and included in the model. The column on the right shows the average posterior probabilities of the corresponding fault patterns. That is, given the product measurement information $\mathbf{Y}$, how likely a fault pattern may occur. We list the candidate models in a descending order according to the posterior probabilities.

The results in Table 4.2 show that the average posterior probability for the correct root cause $\{1, 6\}$ is 0.469, which is higher than that of the other candidate models. We set our fault diagnosis criterion as choosing the fault pattern with the largest posterior. The rate to identify the correct fault pattern $\{1, 6\}$ by applying the proposed diagnosis procedure is 59.6% among 1000 times. This example indicates that the Bayesian diagnosis procedure works well under this fault coupling situation with MCP degree relatively large.

Table 4.2 Fault patterns and the corresponding average posterior probability for MCP of degree 7

Fault Index	$\pi(\boldsymbol{\gamma}|\mathbf{Y})$
1 6	0.469
1	0.250
1 4	0.020
1 5	0.018

4.3.2 Minimal Coupled Pattern of Degree 4

In the second example, we design a $\mathbf{C}^*$ matrix as $\mathbf{C}^* = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{I}_6]$ with $\mathbf{c}_1$ generated from $N_6(0, \mathbf{I}_6)$ and $\mathbf{c}_2 = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]^T$. Same as the first example,

six quality characteristics are measured on each product. During the manufacturing process, two potential process faults and six potential sensor faults would affect the values of products' measurements. By examining the column linear independency in $\mathbf{C}^*$ matrix, one minimal coupled pattern can be identified as {2, 3, 5, 7} with MCP degree 4. The MCP degree for this example is decreased by 3 compared with that of the example in Section 4.3.1. We are interested in finding out whether the diagnosis accuracy will be sensitive to the reduction of MCP degree.

The problem is studied using two sub-cases with the mean shifts set as $\boldsymbol{\beta}(1) = [0 \quad 0 \quad 1.2 \quad 0 \quad 1.2 \quad 0 \quad 1.2 \quad 0]^T$ and $\boldsymbol{\beta}(2) = [0 \quad 0 \quad 1 \quad 0 \quad 0.7 \quad 0 \quad 1.5 \quad 0]^T$, respectively. For each case we calculate the posteriors of the candidate models 1000 times. The first three models with highest average posterior probabilities are listed in Table 4.3.

Table 4.3 Fault patterns and the corresponding average posterior probability degree 4 MCP

$\beta_3 = \beta_5 = \beta_7 = 1.2$		$\beta_3 = 1, \beta_5 = 0.7, \beta_7 = 1.5$	
Fault Index	$\pi(\boldsymbol{\gamma}\|\mathbf{Y})$	Fault Index	$\pi(\boldsymbol{\gamma}\|\mathbf{Y})$
2	0.841	2	0.650
2 3	0.030	2 5	0.184
2 5	0.030	2 3 5	0.047

From Table 4.3, we see that for both cases the fault pattern {2} stands out with the highest posterior probability 0.84 and 0.65. It can be observed that faults {3, 5, 7} belong to the minimal coupled pattern {2, 3, 5, 7}. The number of faults that occur is 3, counting for 75% within the MCP. Rather than picking up a model including multiple faults, the diagnosis approach prefers fault 2 which is also in the MCP {2, 3, 5, 7}.

The results show that when the true faults are not sparse in comparison with their corresponding MCP degree, the correct model including all those true faulty elements will not show up with high posterior probability, while the model with few variables will have high posterior probability. This can be further verified by trying another case where the mean shift is set as $\boldsymbol{\beta}_3 = [0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]^T$. That is, the fault only occurs on the third element of $\boldsymbol{\beta}$ (the first sensor) among a MCP {2, 3, 5, 7}. For 96.5% of the 1000 simulations, our fault diagnosis procedure can identify the correct fault pattern. In contrast to the cases where three faults occur in a four-degree MCP, the diagnosis procedure is much more successful in identifying the sparse fault patterns

4.3.3 Minimal Coupled Pattern of Degree 3

We further reduce the degrees of MCP to three. Consider a matrix $\mathbf{C}^* = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3 \quad \mathbf{I}_{10}]$. The first two column vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ are generated from $N_{10}(0, \mathbf{I}_{10})$ and c3=0.5c1+0.9c2. This $\mathbf{C}^*$ matrix represents a system with ten measurements on each product, three potential process faults and ten potential sensor faults. Since $\mathbf{c}_1$, $\mathbf{c}_2$ and $\mathbf{c}_3$ are linearly dependent, it can be seen that one minimal coupled pattern is {1, 2, 3} with degree 3. Three cases with mean shifts ($\beta_1 = 1.2, \beta_2 = 1.8$), ($\beta_1 = 2, \beta_2 = 1.2$), and $\beta_3 = 1.2$ are studied. Same as the examples in Section 4.3.1 and Section 4.3.2, we calculate the average posterior probabilities for each candidate models by running the simulations 1000 times. The first four fault patterns with the largest average posterior probabilities are listed in Table 4.4.

In addition, the rates that the root causes are detected for these three cases are 25%, 17%, and 98.8%, respectively. Again, the diagnosis procedure is much more successful for case 3 because the true fault pattern in case 3 is sparse compared with the MCP degree.

Table 4.4 Fault patterns and the corresponding average posterior probability for degree 3 MCP

$\beta_1 = 1.2, \beta_2 = 1.8$		$\beta_1 = 2, \beta_2 = 1.2$		$\beta_3 = 1.2$	
Fault Index	$\pi(\boldsymbol{\gamma}\|\boldsymbol{Y})$	Fault Index	$\pi(\boldsymbol{\gamma}\|\boldsymbol{Y})$	Fault Index	$\pi(\boldsymbol{\gamma}\|\boldsymbol{Y})$
3	0.3886	1	0.412	3	0.794
1 2	0.0920	2	0.412	1 3	0.016
1 3	0.0920	2 8	0.008	2 3	0.016
2 3	0.0920	1 8	0.008	1 2	0.016

All the three simulation examples show that the Bayesian diagnosis approach are effective to identify sparse fault patterns, i.e., when the number of true faults is relatively small. The sparsity of the fault pattern should be defined with consideration of the degree of the MCP containing the fault pattern. Typically, if the number of faults in a fault pattern is less than half of the MCP degree, it can be considered as sparse and the Bayesian diagnosis approach should be effective. Before identifying the faults, we don't know exactly how many faults occur and which MCP they belong to. For moderately reliable manufacturing systems, however, since only a small number of elements may have malfunctioned, the sparse fault patterns should be much more likely than the less sparse ones. Under those circumstances, the Bayesian faults diagnosis method can perform well.

In addition, it is important to consider the degrees of MCPs in the design of a manufacturing system. MCPs with large degrees will improve the diagnosability of process and sensor faults.

4.3.4 Sensitivity Analysis and Comparison with CML

In Section 4.2.3, some guidelines for the selection of the hyperparameter values are presented. For the value of c, we recommend using historical data or conducting pilot experiments to get a good estimation of c. For all the examples, we set c=100, which

works pretty well. In this section, we investigate how the diagnosis accuracy will be affected when vary the value of c, and if the magnitude of mean shift will influence the choice of c. The simulation example in Section 4.3.1 is used for the study. The example has a $\mathbf{C}^*$ matrix as $\mathbf{C}^* = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{I}_6]$ with the two columns $\mathbf{c}_1$ and $\mathbf{c}_2$ generated from normal distribution $N_6(0, \boldsymbol{I}_6)$.

Firstly, we add the same mean shift as used in Section 4.3.1 $\boldsymbol{\beta} = [1.5 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0]^T$ to the system. All the parameter values are chosen from Table 4.1 except for c. Figure 4.2 demonstrates a curve between the success rate and the corresponding values of c. The range of c is from 0.1 to 10^3. For each value of c, the diagnosis procedure is conducted 1000 times and the rate to identify the correct fault pattern $\{1, 6\}$ is then approximated. The curve shows the trend that the diagnosis accuracy will be very low when c is small. The diagnosis power will increase with the increase of c. For this case, the identification accuracy will reach its maxima when $\log_{10}(c)$ is about 1, that is c is equivalent to 10. The identification accuracy then goes down slowly but still remains above 50%. In Table 4.5, we listed the success identification rates corresponding c=1, 10, 100 and 1000. When c=100, the success identification rate is 0.596. It is slightly smaller the highest identification rate 0.622 achieved at c=10 but the performance is still very good.

In fact, c values between 10 and 1000 lead to similar diagnosis performance, which shows that the Bayesian diagnosis approach is fairly robust to the selection of the c value and there is a wide range to choose the proper c value.

In Section 4.2.4, we introduce a CML method for choosing the hyerparameters in the Bayesian hierarchical model, mainly dealing with w and c. The major advantage of CML is the success rate of fault identification is not affected by the c value.

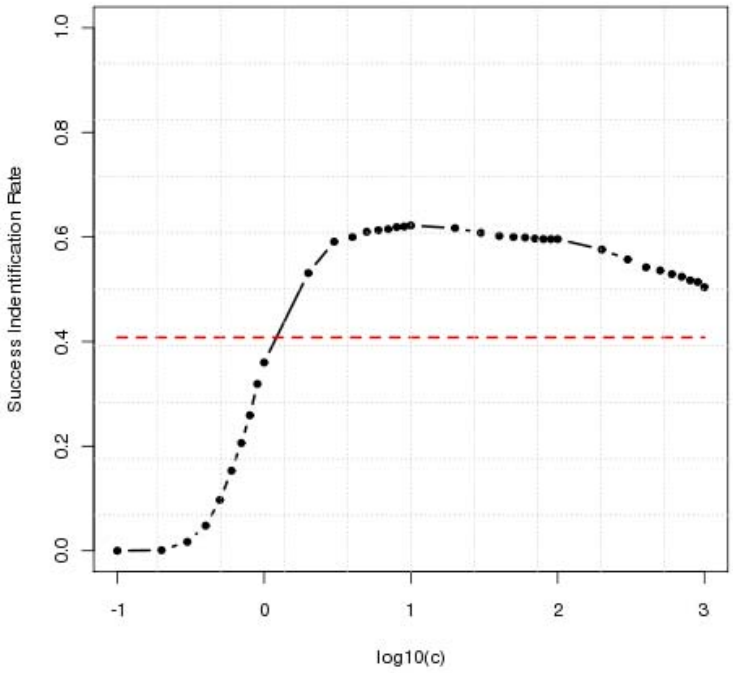


Figure 4.2 Diagnosis success rate v.s. log(*c*)

Table 4.5 Diagnosis performance for different *c* values and the CML method

Value of *c*	1	10	100	1000
Success rate of fault identification WITHOUT using CML	0.36	0.622	0.596	0.504
Success rate of fault identification using CML	0.412	0.412	0.412	0.412

The performance of the CML method is investigated using the same example. In Figure 4.2, the dashed line demonstrates the diagnosis result using CML method. The rate to identify the fault pattern {1,6} is 41.2%. The performance is not as good as choosing *c* from [10, 1000] in this example, but the diagnosis accuracy is still acceptable. In real

applications, however, the true optimal *c* value is unknown since the true fault pattern is unknown. So the CML method is an alternative method to achieve a reliable result.

Besides the parameter choice, the diagnosis performance will also depend on the process and measurement system layouts, and also the magnitude of the mean shift. For a manufacturing with multiple potential faults, since there are infinite numbers of fault combinations, we only focus on the single mean shift fault.

We still using the system $\mathbf{C}^* = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{I}_6]$ with the two columns $\mathbf{c}_1$ and $\mathbf{c}_2$ generated from normal distribution $N_6(0, \boldsymbol{I}_6)$. Figure 4.3 demonstrates four posterior probability curves by varying β_3, the third component of $\boldsymbol{\beta}$ and also the first sensor, from 0 to 1.5 while the other components of $\boldsymbol{\beta}$ remain zero. The four curves are plotted under *c*=1, 10, 100, 150. The plot shows that the posterior probability of the correct model increases with the increase of the magnitude of mean shift. That is, larger mean shift will be easier to identify. When the mean shift magnitude exceeds some threshold η, in this example $\eta = 0.6$, the posterior probability will remain stable. By comparing the four curves, it can be seen that generally, for the same magnitude of mean shift, the posterior probability of the correct fault pattern will be higher when choosing a larger *c*. This result is accordant to the result we get from the previous sensitivity analysis.

Another simulation example is conducted by adding a single process fault $\beta_3 = 0$ and keeping the rest components in $\boldsymbol{\beta}$ as zero. The posterior probability curves of the fault pattern $\{1\}$ are plotted in Figure 4.4. The results are similar to the previous example – setting a single fault $\beta_3 \neq 0$. It can be observed the threshold for the posterior probability to remain steady is about $\eta = 0.4$, which means that the diagnosis procedure tends to be more sensitive to the mean shift on process fault β_1 than the sensor fault β_3. It indicates that the proposed diagnosis procedure has different sensitivity to the root causes due to their effect on the product quality characteristics and the layout.

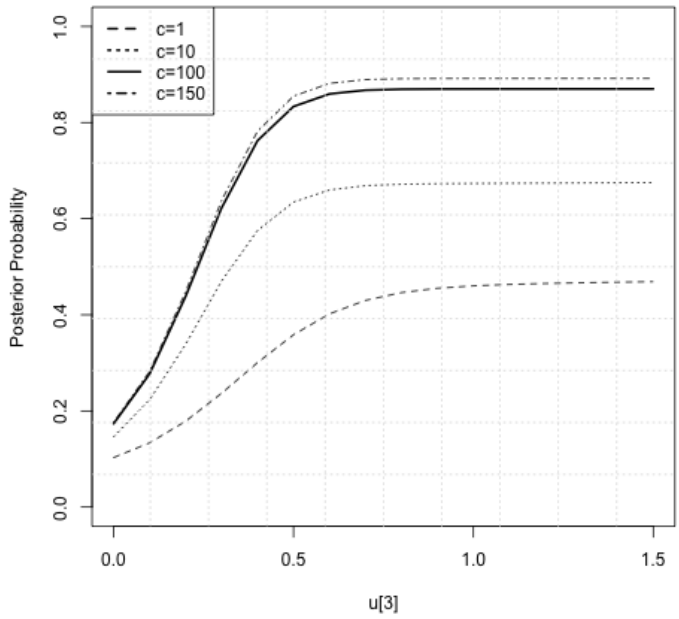


Figure 4.3 Posterior Probability v.s. Single Sensor Fault Mean Shift Magnitude

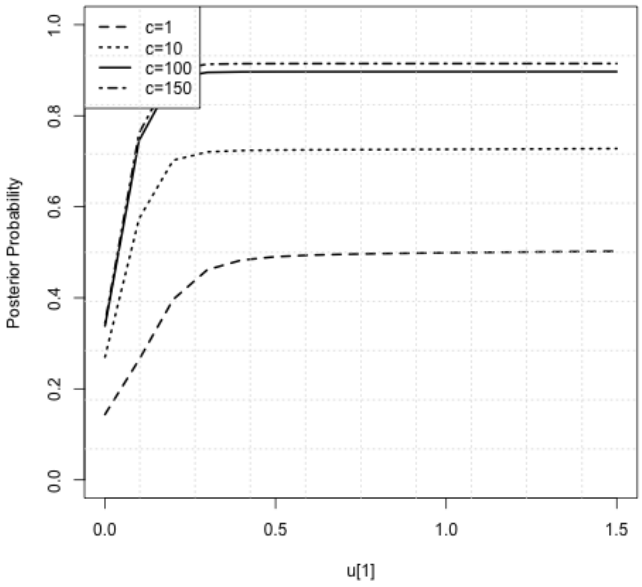


Figure 4.4 Posterior Probability v.s. Single Process Fault Mean Shift Magnitude

4.4 Case Study

In this section, we apply the Bayesian fault diagnosis approach to automotive body assembly example demonstrated in Section 2.2.3 and Section 3.4. To increase the computational complexity, we add the quality measurement points to 15. For each point, the deviations in both *x*- and *z*- directions are measured by the sensors. The revision and new added points are illustration in Figure 4.5

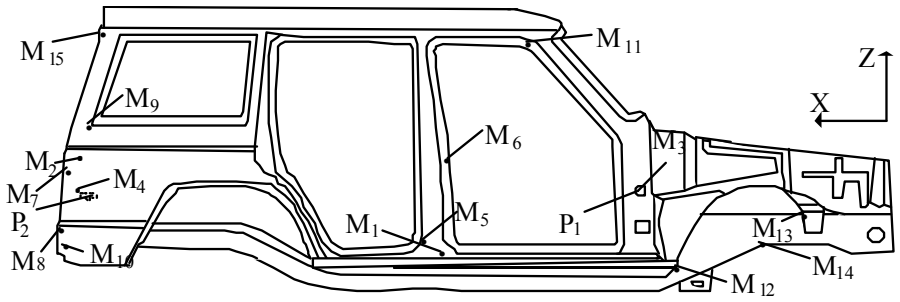


Figure 4.5 Key points on right-hand bodyside

Table 4.6 Nominal *x*-*z* coordinates for pins and measurement points

Point	Nominal Coordinates (mm)	
	x	*z*
P_1	2184	1489
P_2	4680	1428
M_1	3134	1200
M_2	4015	1618.5
M_3	2184	1489
M_4	4895.3	1510.5
M_5	3721	1256.5
M_6	3264	1930
M_7	4895	1608
M_8	4895.5	1273

M_9	4693.8	2228.5
M_{10}	4899	1214.5
M_{11}	2890	2228.5
M_{12}	2000	1031
M_{13}	1523	1640
M_{14}	1840	1300
M_{15}	3984	2228.5

The nominal *x*-*z* coordinates of P_1, P_2, and the 15 measurement points are shown in Table 4.6. The same as the notation used in Section 2.2.3, $\mathbf{y}(j) \equiv [M_1(x) \quad \cdots \quad M_{15}(x) \quad M_1(z) \quad \cdots \quad M_{15}(z)]^T$ denotes the sensor measurement vector, where $M_i(x)$ and $M_i(z)$ represents the measured deviations of M_i's (i=1,..,15) from their nominal positions in the *x*- and *z*- directions respectively.

Table 4.7 Hyperparameters settings for assembly example

c	λ	ν	w
100	0.1	10	1/33

Table 4.8 Root cause v.s. success identification rate

True Fault Patterns	Without CML	With CML
$\beta_1 = 0.4mm$	93.22%	91%
$\beta_1 = 0.15mm, \beta_6 = 0.1mm$	83.72%	73.82%
$\beta_1 = 0.1mm, \beta_8 = 0.08mm, \beta_{22} = 0.13mm$	69.88%	73.10%

Based on the geometric relationship between locating pins and the measurement points, the $\mathbf{C}$ and $\mathbf{C}^*$ matrix can be obtained as

$$\mathbf{C} = \begin{bmatrix} 0.1158 & 1 & -0.1158 \\ -0.0519 & 1 & 0.0519 \\ 0 & 1 & 0 \\ -0.0086 & 1 & 0.0086 \\ 0.0931 & 1 & -0.0931 \\ -0.1767 & 1 & 0.1767 \\ -0.0477 & 1 & 0.0477 \\ 0.0865 & 1 & -0.0865 \\ -0.2963 & 1 & 0.2963 \\ 0.1100 & 1 & -0.1100 \\ -0.2963 & 1 & 0.2963 \\ 0.1835 & 1 & -0.1835 \\ -0.0605 & 1 & 0.0605 \\ -0.2963 & 1 & 0.2963 \\ 0.3806 & 0 & 0.6194 \\ 0.7336 & 0 & 0.2664 \\ 0 & 0 & 1 \\ 1.0863 & 0 & -0.0863 \\ 0.6158 & 0 & 0.3842 \\ 0.4327 & 0 & 0.5673 \\ 1.0861 & 0 & -0.0861 \\ 1.0863 & 0 & -0.0863 \\ 1.0055 & 0 & -0.0055 \\ 1.0877 & 0 & -0.0877 \\ 0.2829 & 0 & 0.7171 \\ -0.0737 & 0 & 1.0737 \\ -0.2648 & 0 & 1.2648 \\ -0.1378 & 0 & 1.1378 \\ 0.7212 & 0 & 0.2788 \end{bmatrix}$$

And $\mathbf{C}^*$ is obtained as $\mathbf{C}^* = [\mathbf{C} \quad \mathbf{I}_{30\times 30}]$.

The variance of the sensor measurements in this assembly system is

$$\sigma^2 = \left(\frac{0.2}{6} mm\right)^2$$

We use sample size $N = 20$. To illustrate the performance of the diagnosis approach, we study three cases here. For the first case, we add only one mean shift fault on pin P_1 in the z-direction, $\beta_1 = 0.4mm$, and keep P_2 and all the sensors working properly. For the

second case, a double-mean-shift is added on P_1 in x-direction and a sensor mean shift M_3 in x direction, $\beta_1 = 0.15mm, \beta_6 = 0.1mm$, such that one process fault and one sensor fault occur simultaneously. For the third case, we introduce a triple-mean-shift-fault on P_2 in z-direction, M_5 in x direction, and M_4 in z direction by setting $\beta_1 = 0.1mm, \beta_8 = 0.08mm, \beta_{22} = 0.13mm$.

Before the diagnosis procedure, we check the linear dependency among columns in $\mathbf{C}^*$ matrix. One of the smallest MCFs is {1, 3, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30} with degree 17. Based on the discussions in Section 4.3, a fault pattern with no more than 8 faults is considered as sparse and can be identified by the Bayesian diagnosis approach.

The reason we add more measurement points is we want to increase the number of potential faults such that MCMC simulation approximation method can be applied. For this revised example, the total number of potential faults is 33, which is greater than 25, and the exhaustive calculation of all the posterior probabilities is infeasible and unnecessary. We then apply the M-H simulation algorithm to estimate $\pi(\boldsymbol{\gamma}|\mathbf{Y})$. The total number of iterations for the M-H algorithm is set to be 30000 with 4000 burn-in iterations. To find the percentage of identifying the correct patterns, we run the entire simulation for 1000 times.

The hyperparameter settings used in the diagnosis are listed in Table 4.7 and the percentages of correct root cause identification for the three cases are given in the second column of Table 4.8. It shows the success identification rates for the three cases are 93.22%, 83.72% and 69.88% respectively, which indicates that the Bayesian fault diagnosis approach is effective for all three cases. Note that the sample size used in this example is N=20. If larger sample size is used, the percentage of correct fault identification should be higher. In addition, the results using the CML method listed in the third column of Table 4.8. Again, the performance of the CML method is reliable and comparable with the result using c=100.

CHAPTER 5. SUMMARY

My research focuses on developing systematic procedures in detecting and diagnosing process and sensor mean shift faults in the manufacturing systems which can be modeled by a fault-quality model.

The studies are divided into monitor part and diagnosis part. We first propose a multivariate monitoring process in Chapter 2 with the assumption that no sensor faults present and illustrated its use in conjunction with models from the literature. Compared to standard methods for monitoring multivariate processes (e.g., chi-squared control charts and Lowry et al.'s (1992) direct MEWMA), the procedure has the advantage of monitoring the process in a potentially reduced-dimensional space. This helps to alleviate the diminished sensitivity of monitoring methods that results from increases in the number of product quality variables, a consequence of the proliferation of in-process sensors.

In Chapter 3, we develop a W control chart to detect sensor mean shifts in manufacturing processes based on a linear fault-quality model. It shows that the W chart is sensitive to sensor faults only and completely insensitive to process faults. This property enables the W chart, when combined with the U chart, to effectively distinguish sensor faults from process faults, which will greatly assist fault isolation and elimination in manufacturing processes.

Because the statistic monitored by the W control chart follows a chi-squared distribution, with the degrees of freedom fixed, the noncentrality parameter completely determines the fault detection sensitivity and the average run length of the W chart. A new unit-free index referred to as the sensitivity ratio (SR) is defined to measure the sensitivity of the W chart. Through analysis of the sensitivity ratios for single- and double- sensor faults, it is clear that the sensitivity of the W chart is affected by the potential influence, which is defined in the robust regression literature, of the sensor measurement.

In Chapter 4, a Bayesian fault diagnosis approach is developed to locate the root causes in discrete manufacturing processes. The problem is very difficult because the number of potential faults to be estimated is always larger than the number of measurements. Based on the Bayesian variable selection algorithm, our fault diagnosis approach mainly deals with the fault pattern with the largest posterior probability. In the setup of the hierarchical priors, we impose a constraint that the coupled faults cannot occur simultaneously. The diagnosis is made based on the first or first few fault patterns with the highest posterior probabilities in the system. When the size of the problem is large, the Metropolis-Hastings algorithm – an MCMC method – is applied to identify the most frequent fault patterns in the simulation. Various numerical examples are used to demonstrate the diagnosis procedure, which show that the proposed diagnosis procedure is effective in identifying sparse fault patterns. Guidelines have been provided for the hyperparameters selection in the Bayesian hierarchical model. An alternative method for selection of some hyperparameters, the CML method, is also discussed and tested through examples.

REFERENCE

Alt, F.B. (1985), " Multivariate Quality Control," Encyclopedia of Statistical Sciences, Vol.6, edited by N.L.Johnson and S. Kotz, Wiley, New York.

Alt, F. B. (2001), "Multivariate Quality Control," In Saul I. Gass and Carl M. Harris (eds.), *Encyclopedia of Operations Research and Management Science*, Kluwer Academic, Boston, pp. 544-550.

Alt, F. B., and Smith, N. D. (1988), "Multivariate Process Control," In P. R. Krishnaiah and C. R. Rao (eds.), *Handbook of Statistics*, Vol. 7, Elsevier, pp. 333-351.

Apley, D. W., and Shi, J. (1998), "Diagnosis of Multiple Fixture Faults in Panel Assembly," *ASME Transactions, Journal of Manufacturing Science and Engineering*, 120, 793-801.

Apley, D. W., and Shi, J. (2001), "A Factor-Analysis Method for Diagnosing Variability in Multivariate Manufacturing Processes," *Technometrics*, 43 (1), 84-95.

Carlin, B. P. and Chib, S. (1995)."Bayesian Model Choice via Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society*, Ser. B, 57, 473-484.

Ceglarek, D., and Shi, J. (1999), "Fixture Failure Diagnosis for Sheet Metal Assembly with Consideration of Measurement Noise," *ASMEJournal of Manufacturing Science and Engineering*, 121, 771-777.

Ceglarek, D., and Shi, J. (1996), "Fixture Failure Diagnosis for Autobody Assembly Using Pattern Recognition," *ASME Journal of Engineering for Industry*, 188(1), 55-65.

Chipman, H., Hamada, M., and Wu, C. F. J. (1997), "A Bayesian Variable-Selection Approach for Analyzing Designed Experiments with Complex Aliasing," *Technometrics*, 39(4), 372-381.

Chipman, H. (1996), "Bayesian Variable Selection with Related Predictors," *Canadian Journal of Statistics*, 24, 17-36.

Chen, G., Cheng, S.W., and Xie, H. (2005), "A New Multivariate Control Chart for Monitoring Both Location and Dispersion," *Communications in Statistics – Simulation and Computation*, 34(1), 203-217.

Chou, C. T. and Varhaegen, M. (1997), "Subspace Algorithms for the Identification of Multivariable Dynamic Errors-in-Variables Models," *Automatica*, 33(10), 1857-1869.

Clyde, M. A., DeSimone, H. and Parmigiani, G. (1996), "Prediction via Orthogonalized Model Mixing," *Journal of American Statistical Association*, 91, 1197-1208.

Crosier, R. B. (1988), "Multivariate Generalizations of Cumulative Sum Quality-Control Schemes," *Technometrics*, 30, 291-303.

Ding, Y., Ceglarek, D., and Shi, J. (2002), "Fault Diagnosis of Multistage Manufacturing Processes by Using State Space Approach," *ASME Transactions, Journal of Manufacturing Science and Engineering*, 124, 313-322.

Ding, Y., Ceglarek, D., and Shi, J. (2000), "Modeling and Diagnosis of Multistage Manufacturing Processes: Part I—State Space Model," *Japan/USA Symposium on Flexible Automation*, July 23-26, Ann Arbor, MI.

Ding, Y., Gupta, A., and Apley, D. W. (2004), "Singularity Issues in Fixture Fault Diagnosis for Multi-Station Assembly Processes," *ASME Transactions, Journal of Manufacturing Science and Engineering*, 126, 200-210.

Ding, Y., Shi, J., and Ceglarek, D. (2002), "Diagnosability Analysis of Multi-station Manufacturing Processes," *ASME Journal of Dynamic Systems, Measurement, and Control*, 124, 1-13.

Ding, Y., Zhou, S. and Chen, Y. (2005), "A Comparison of Process Variance Estimation Methods for In-Process Dimensional Control," *ASME Transactions, Journal of Dynamic Systems, Measurement, and Control*, 127, 69-79.

Fantoni, P. F., and Mazzola, A. (1996), "Multiple-failure Signal Validation in Nuclear Power Plants Using Artificial Neural Networks," *Nuclear Technology*, 113(3), 368-374.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990), "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling," *Journal of the American Statistical Association*, 85, 972-985.

Gelfand, A. E., Hills, and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398-409.

George, E. I., and Foster, D. P. (2000), "Calibration and Empirical Bayesian Variable Selection," *Biometrica*, 87(4), 731-747.

George, E. I., and McCulloch, R. E. (1997), "Approaches for Bayesian Variable Selection," *Statistica Sinica,* 7, 339-373.

George, E. I., and McCulloch, R. E. (1995), " Stochastic Search Variable Selection," in *Practical Markov Chain Monte Carlo in Practice*, eds. W.R. Gilks, S. Richardson and D.J. Spiegelhalter, Chapman & Hall, London, pp. 203-214.

George, E. I., and McCulloch, R. E. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881-889.

George, E. I., McCulloch, R. E. and Tsay, R. (1995), "Two Approaches to Bayesian Model Selection with Applications," in *Bayesian Statistics and Econometrics: Essay in Honor of Arnold Zellner*, eds. D. Berry, K. Chaloner and J. Geweke, Wiley, New York, pp. 339-348.

Geweke, J. (1996), "Variable Selection and Model Comparison in Regression," in *Bayesian Statistics 5*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Oxford Press, pp. 609-620.

Hawkins, D. M. (1991), "Multivariate Quality Control Based on Regression-Adjusted Variables," *Technometrics*, 33, 61-75.

Hayter, A. J., and Tsui, K. (1994), "Identification and Quantification IN Multivariate Quality Control Problems," *Journal of Quality Technology*, 26(3), 197-208.

Healy, J. D. (1987), "A Note on Multivariate CUSUM Procedures," *Technometrics*, 29, 409-412.

Hoeting, J., Raftery, A. E. and Madigan, D. (1997), "A Method for Simultaneous Variable Selection and Outlier Identification in Linear Regression," *Computational Statistics and Data Analysis*, 22, 251-270.

Huang, Q. and Shi, J. (2004) "Variation Transmission Analysis and Diagnosis of Multi-Operational Machining Processes," *IIE Transactions*, 36, 807-815.

Huang, Q., Zhou, S., and Shi, J. (2002), "Diagnosis of Multi-Operational Machining Processes Through Variation Propagation Analysis," *Robotics and Computer Integrated Manufacturing*, 18, 233-239.

Jin, J., and Shi, J. (1999), "State Space Modeling of Sheet Metal Assembly for Dimensional Control," *ASME Transactions, Journal of Manufacturing Science and Engineering*, 121, 756-762.

Joseph, V. R. (2006), "A Bayesian Approach to the Design and Analysis of Fractionated Experiments," *Technometrics*, 48(2), 219-229.

Kang, L., and Albin, S. L. (2000), "On-Line Monitoring When the Process Yields a Linear Profile," *Journal of Quality Technology*, 32 (4), 418-426.

Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003), "On the Monitoring of Linear Profiles," *Journal of Quality Technology*, 35 (3), 317-328.

Kuo, L. and Mallick, B. (1994), "Variable Selection for Regression Models", Department of Statistics, University of Connecticut and Department of Mathematics, Imperial College, London, England.

Lempers, F. B. (1971), *Posterior Probabilities of Alternative Linear Models,* Rotterdam: Rotterdam University Press.

Lowry, C. A., and Montgomery, D. C. (1995), "A Review of Multivariate Control Charts," *IIE Transactions*, 27, 800-810.

Lowry, C. A., Woodall, W. H., Champ, C. V., and Rigdon, S. E. (1992), "A Multivariate Exponentially Weighted Moving Average Control Chart," *Technometrics*, 34, 46-53.

Mahmoud, M. A., and Woodall, W. H. (2004), "Phase I Analysis of Linear Profiles with Calibration Applications," *Technometrics*, 46 (4), 380-391.

Mason, R.L., Tracy, N.D., and Young, J.C. (1995), "Decomposition of T^2 for Multivariate Control Chart Interpretation," *Journal of Quality Technology*, 27(2), 99-108.

Meehan, P. M., Dempster, A. P. and Brown, E. N. (1994), "A Belief Function Approach to Likelihood to Updating in A Gaussian Linear Model," Manuscript, *Frontier Science and Technology Research Foundation*, Harvard University and Massachusetts General Hospital.

Mitchell, T. J., and Beauchamp, J. J. (1988), "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, 83, 1023-1036.

Montgomery, D. C. (2005) *Introduction to Statistical Quality Control(5 th Edition)*, Wiley, New York, NY.

Pericchi, L. R. (1985), "Bayesian Hypothesis Testing in Linear Models with Continuously Induced Conjugate Priors Across Hypotheses," *Biometrika*, 71, 575-586.

Phillips, D. B. and Smith, A. F. M. (1995), "Bayesian Model Comparison via Jump Diffusions," in *Practical Markov Chain Monte Carlo in Practice*, eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter, Chapman & Hall, London, pp. 215- 239.

Pignatiello, J. J., and Runger, G. C. (1990), "Comparisons of Multivariate CUSUM Charts," *Journal of Quality Technology*, 22, 173-186.

Poirier, D. J. (1985), "Bayesian Hypothesis Testing in Linear Models With Continuously Induced Conjugate Priors Across Hypotheses," in *Bayesian Statistics 2,* eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, New York: Elsevier, pp. 711-722.

Prabhu, S. S., and Runger, G. C. (1997), "Designing a Multivariate EWMA Control Chart," *Journal of Quality Technology*, 29 (1), 8-15.

Qin, S. J and Li, W. (2001), "Detection and Identification of Faulty Sensors in Dynamic Processes," *AIChE Journal*, 47(7), 1581-1593.

Raftery, A. E., Madigan, D. M. and Hoeting, J. A. (1997), "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92, 179-191.

Raftery, A. E., Madigan, D. M. and Volinsky, C. T. (1996), "Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance," in *Bayesian Statistics 5*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Oxford Press, pp. 323-350.

Rao, C. R. (2002) *Linear Statistical Inference and Its Applications (Second Edition)*, Wiley, New York, NY.

Rigdon, S. E. (1995a), "A Double-Integral Equation for the Average Run Length of a Multivariate Exponentially Weighted Moving Average Control Chart," *Statistics and Probability Letters*, 24, 365-373.

Rigdon, S. E. (1995b), "An Integral Equation for the In-Control Average Run Length of a Multivariate Exponentially Weighted Moving Average Control Chart," *Journal of Statistical Computation and Simulation*, 52, 351-365.

Runger, G. C. (1996), "Projections and the U^2 Multivariate Control Chart," *Journal of Quality Technology*, 28 (3), 313-319.

Runger, G.C., Alt, F.B. and Montgomery, T.R. (1996) "Contributors to a Multivariate Statistical Process Control Signal", *Communication in Statistics — Theory and Methods*, 25(10). 2103-2213.

Runger, G. C., and Prabhu, S. S. (1996), "A Markov Chain Model for the Multivariate Exponentially Weighted Moving Averages Control Chart," *Journal of the American Statistical Association*, 91, 1701-1706.

Smith, M., and Kohn, R. (1996), "Nonparametric Regression Using Bayesian Variable Selection," *Journal of Econometrics*, 75, 317-343.

Smith, A. F. M., and Spiegelhalter, D. J. (1980), "Bayes Factors and Choice Criteria for Linear Models," *Journal of the Royal Statistical Society*, Ser.B, 42, 213-220.

Spiegelhalter, D. J., and Smith, A. F. M. (1982), "Bayes Factors for Linear and Log-Linear Models With Vague Prior Information," *Journal of the Royal Statistical Society*, Ser.B, 44, 377-387.

Stewart, L. (1987), " Hierarchical Bayesian Analysis Using Monte Carlo Integration: Computing Posterior Distributions When There are Many Possible Models," *The Statistician,* 36, 211-219.

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association*, 82, 528-550.

Verdinelli, I., and Wasserman, L. (1991), "Bayesian Analysis of Outlier Problems Using the Gibbs Sampler," *Statistics and Computing*, 1, 105-117.

Wade, M. R., and Woodall, W. H. (1993), "A Review and Analysis of Cause-Selecting Control Charts," *Journal of Quality Technology*, 25, 161-169.

Wakefield, J. C. and Bennett, J. E. (1996), "The Bayesian Modeling of Covariates for Population Pharmacokinetic Models," *Journal of American Statistical Association*, 91, 917-927.

Woodall, W.H., and Montgomery, D.C. (1999), "Research Issues and Ideas in Statistical Process Control," *Journal of Quality Technology*, 31(4), 376-386.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004), "Using Control Charts to Monitor Process and Product Quality Profiles," *Journal of Quality Technology*, 36 (3), 309-320.

Yuan, M., and Lin, Y. (2005), "Efficient Empirical Bayes Variable Selection and Estimation in Linear Models", *Journal of the American Statistical Association*, 100, 1215-1225.

Zellner, A. (1986), " On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distribution," in *Bayesian Inference and Decision Techniques*, eds. P.K. Goel and A. Zellner, Amsterdam, pp. 233-243.

Zellner, A. (1984), "Posterior Odds Ratios for Regression Hypotheses General Considerations and Some Specific Results," in *Basic Issues in Econometrics*, ed. A. Zellner, Chicago: University of Chicago Press, pp. 275-305.

Zhou, S., Chen, Y., and Shi, J. (2004), "Statistical Estimation and Testing for Variation Root-Cause Identification of Multistage Manufacturing Processes," *IEEE Transactions on Automation Science and Engineering*, 1 (1), 73-83.

Zhou, S., Huang, Q. and Shi, Y. (2003), "State Space Modeling of Dimensional Variation Propagation in Multistage Machining Process Using Differential Motion Vectors," *IEEE Transactions on Robotics and Automation*, 19(2), 296-309.