SENSOR FUSION AND PROCESS MONITORING FOR ULTRASONIC WELDING OF LITHIUM-ION BATTERIES

by

Weihong Guo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2015

Doctoral Committee:

Professor S. Jack Hu, Co-Chair
Professor Judy Jin, Co-Chair
Professor Elijah Kannatey-Asibu Jr.
Assistant Professor Eunshin Byon

To my family

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and respect to my co-advisors, Professor S. Jack Hu and Professor Judy Jin, for their enthusiastic inspiration, endless patience and encouragement, and insightful guidance throughout my research. Without them, this dissertation would not have become possible. Through their guidance and encouragement, I have grown both personally and professionally.

My sincere gratitude also goes to my other dissertation committee members, Professor Kannatey-Asibu Jr. and Professor Byon, who have continuously supported me over the years in many ways. Their knowledge from other points of view and valuable comments help me improve my research quality.

I am very grateful for the collaborative research environment of Hu Lab. Thanks to Prof. Mihaela Banu, Prof. Haseung Chung, Dr. Tae Hyung Kim, Dr. Hui Wang, Dr. John Wang, Dr. Jingjing Li, Dr. Xiaoning Jin, Dr. Hai Trong Nguyen, Dr. Shawn Lee, Dr. Saumuy Suriano, Dr. Robert Riggs, Dr. Kaifeng Wang, Vernnaliz Carrasquillo, Chenhui Shao, Heng Kuang, Ying Luo and Daniel Shriver for their help and friendship.

I greatly appreciate the friendship of my fellow students in Prof. Judy Jin's group: Dr. Qiang Li, Dr. Kamran Paynabar, Dr. Amir Sadrpour, Rui Chen, Hadi Masoud, Yaser Zerehsaz and Wenbo Sun. They have inspired me with their intelligence and insight. I am also grateful to the faculty and staff at the IOE department, particularly Tina, Wanda, Matt and Chris, for their good cheer and helpfulness.

I would like to thank my family for all their love and encouragement. I want to thank Shengli Wang and Sha Jiang for making Ann Arbor feel like home to me. I want to thank my parents-in-law for their love and understanding.

Most importantly, I want to thank my dear parents for always believing in me and always being there for me. They raised me with a love of knowledge and have provided a never ending source of energy for my doctoral studies and other endeavors. Last but not least, I want to thank my beloved husband, Xi, for his love, understanding, and unconditional support that have sustained me throughout this critical stage of life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Ultrasonic metal welding is used for joining lithium-ion batteries of electric vehicles. The quality of the joints is essential to the performance of the entire battery pack. Hence, the ultrasonic welding process that creates the joints must be equipped with online sensing and real-time process monitoring systems. This would help ensure the process to be operated under the normal condition and quickly address quality-related issues. For this purpose, this dissertation develops methods in process monitoring and fault diagnosis using online sensing signals for ultrasonic metal welding.

The first part of this dissertation develops a monitoring algorithm that targets near-zero misdetection by integrating univariate control charts and a multivariate control chart using the Mahalanobis distance. The proposed algorithm is capable of monitoring non-normal multivariate observations with adjustable control limits to achieve a near-zero misdetection rate while keeping a low false alarm rate. The proposed algorithm proves to be effective in achieving near-zero misdetection in process monitoring in ultrasonic welding processes.

The second part of the dissertation develops a wavelet-based profile monitoring method that is capable of making decisions within a welding cycle and guiding real-time process adjustments. The proposed within-cycle monitoring technique integrates real-time monitoring and within-cycle control opportunity for defect prevention. The optimal decision point for achieving the most benefit in defect prevention is determined through the formulation of an optimization problem. The effectiveness of the proposed method is validated and demonstrated by simulations and case studies.

The third part of this dissertation develops a method for effective monitoring and diagnosis of multi-sensor heterogeneous profile data based on multilinear discriminant analysis. The proposed method operates directly on the multi-stream profiles and then extracts uncorrelated discriminative features through tensor-to-vector projection, and thus preserving the interrelationship of different sensors. The extracted features are then fed into classifiers to detect faulty operations and recognize fault types.

The research presented in this dissertation can be applied to general discrete cyclic manufacturing processes that have online sensing and control capabilities. The results of this dissertation are also applicable or expandable to mission-critical applications when improving product quality and preventing defects are of high interests.

CHAPTER 1

INTRODUCTION

1.1 Motivation

Increasing concerns in recent years over the environmental impact of the petroleum-based transportation infrastructure and soaring gas price have led to great interest in electric vehicles. In manufacturing lithium-ion battery packs for electrical vehicles, it is critical to create reliable interconnections between battery cells, between modules, and between modules and control units. Such connections must possess reliable electrical conductivity and sufficient mechanical strength to ensure battery performance. Ultrasonic metal welding has been adopted for joining lithium-ion battery tabs due to its advantages in joining dissimilar and conductive materials (Kim *et al.*, 2011). Ultrasonic metal welding is a solid-state bonding process which uses high frequency ultrasonic energy to generate oscillating shears between two metal sheets clamped under pressure (Lee *et al.*, 2010).

The performance of an entire battery pack may not be as intended if some battery joints have low quality connections; hence, high quality interconnections between battery cells are critically needed. A low quality joint is usually caused by abnormal conditions during the welding process. Therefore, the ultrasonic welding process that creates those joints must be equipped with online sensing and real-time process monitoring systems to help ensure the process to be operated under the normal condition and quickly address quality-related issues.

With online sensing and data capturing technology, sensor signals acquired during the ultrasonic welding process provide a suitable source of information to develop process

monitoring and quality control tools. However, significant research challenges arise in developing effective methodologies for analyzing and monitoring these sensor signals to meet the increasing demand in high product quality and reliability. These critical problems are to be addressed in this dissertation as follows:

1. Process monitoring that targets a near-zero misdetection rate foremost in order to prevent any battery joints with a low quality connection going into the downstream assembly, while maintaining a relatively low Type I error rate to reduce unnecessary manual inspection.
2. Profile monitoring that is capable of making decisions within a profile cycle and guiding real-time process adjustments for the purpose of defect prevention
3. Profile monitoring and fault diagnosis that considers multi-stream signals via sensor fusion.

The case studies in this dissertation are based on ultrasonic metal welding of batteries. But the proposed methods can be applied to general discrete cyclic manufacturing processes that have the online sensing and control capabilities. The results of this dissertation are also highly applicable or expandable to mission-critical applications when improving product quality and preventing defects are of high interests.

1.2 Measurement System in Ultrasonic Welding of Lithium-ion Batteries

In ultrasonic metal welding, high-frequency (20 kHz) ultrasonic acoustic vibrations are locally applied to workpieces being held together under pressure to create a solid-state weld. As illustrated in Figure 1.1, during welding, the transducer transforms electrical energy into high frequency mechanical vibration; this mechanical vibration is transferred to a welding tip through an acoustically tuned horn. This high frequency vibration, applied under force,

disperses surface films and oxides, creating a metallurgical bond (Ultraweld® by Branson Ultrasonics Corporation).

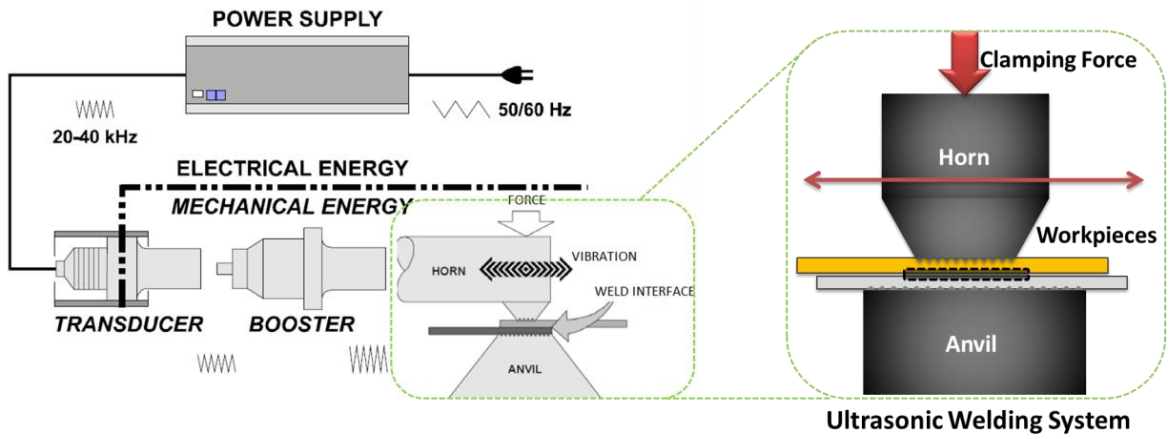


Figure 1.1. Ultrasonic welding system (Lee *et al.*, 2010)

In order to gather real-time information about the ultrasonic welding process, four sensors are applied to collect both electrical and mechanical information during welding: watt (power) meter, force sensor, linear variable differential transformer (LVDT) sensor, and microphone. Table 1.1 summarizes all the sensors, their signal types, and purposes. As shown in Figure 1.2, the watt meter and force sensor measure the transducer power and clamping force at the weld control module. In the evaluation of an ultrasonic transducer for weld quality estimation, monitoring of the output ultrasonic power and force is important to understand and identify process and tool conditions during welding.

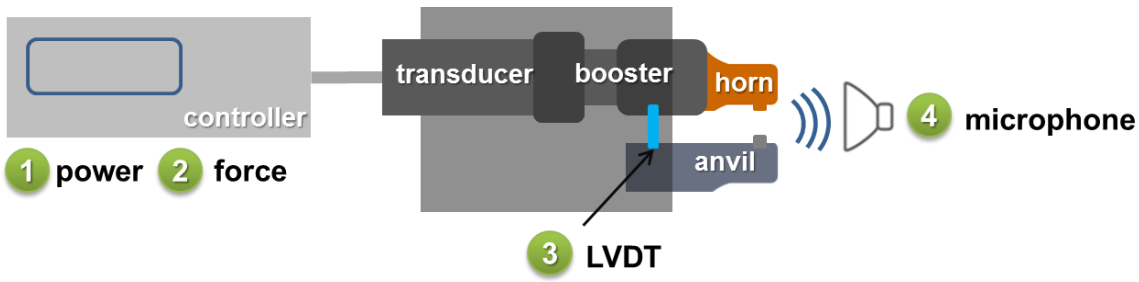


Figure 1.2. Sensors and their positions in the ultrasonic metal welding machine

Table 1.1. Applied sensors, signal types, and purposes

Sensor	Signal type	Purpose
Watt (power) meter	Ultrasonic power output at piezo-ceramic module	Monitor controller power input signal
Force sensor	Clamping force output at piezo-ceramic module	Measure clamping force at the ultrasonic transducer
LVDT sensor	Displacement between horn and anvil	Measure indentation and sheet thickness variation during welding
Microphone	Sound wave form	Detect cracking and slipping during welding

The LDVT sensor is applied to measure displacement between the horn and anvil. Material deformation from LVDT signal profile is one of the most important process variables because plastic deformation as one of main bond mechanisms of ultrasonic welding is connected closely to material deformation.

The microphone is used to collect audible sounds generated during welding. Since ultrasonic welding is a joining process based on the oscillation at the material interfaces, the input ultrasonic energy can be converted to not only material bonding energy including heat generation and plastic deformation but also energy dissipated by friction between the tool and workpiece, and between workpieces. Hence, sounds generated during welding can give a good indication about how much energy is dissipated to sounds which are not for bonding. During ultrasonic welding, the resonance frequency of the ultrasonic transducer is increasing or decreasing according to the dynamic mechanical load at the workpiece near 20 kHz.

1.3 Dissertation Overview and Organization

The objective of this research is to develop effective statistical process monitoring and fault diagnosis methods via sensor fusion and data fusion for ultrasonic metal welding. The organization of this dissertation is depicted in Figure 1.3. In this section, the research topics highlighted in the previous section are briefly discussed in the following subsections. For

each topic, an overview of research objectives, challenges, and the proposed methodology are provided. Details on these topics are presented in Chapters 2, 3, and 4, respectively, which are written as individual research papers, including the main body sections and references. Chapter 2 develops a new method for process monitoring that targets a near-zero misdetection rate in order to prevent battery joints with low quality connections going into the downstream assembly. Chapter 3 develops a new wavelet-based profile monitoring method that is capable of making decisions within an operation cycle and guiding real-time process adjustments. Chapter 4 develops a new method for multi-stream profile monitoring and fault diagnosis based on multilinear discriminant analysis. Chapter 5 draws the conclusions and summarizes the original contributions of the dissertation. Several topics of future research are suggested.

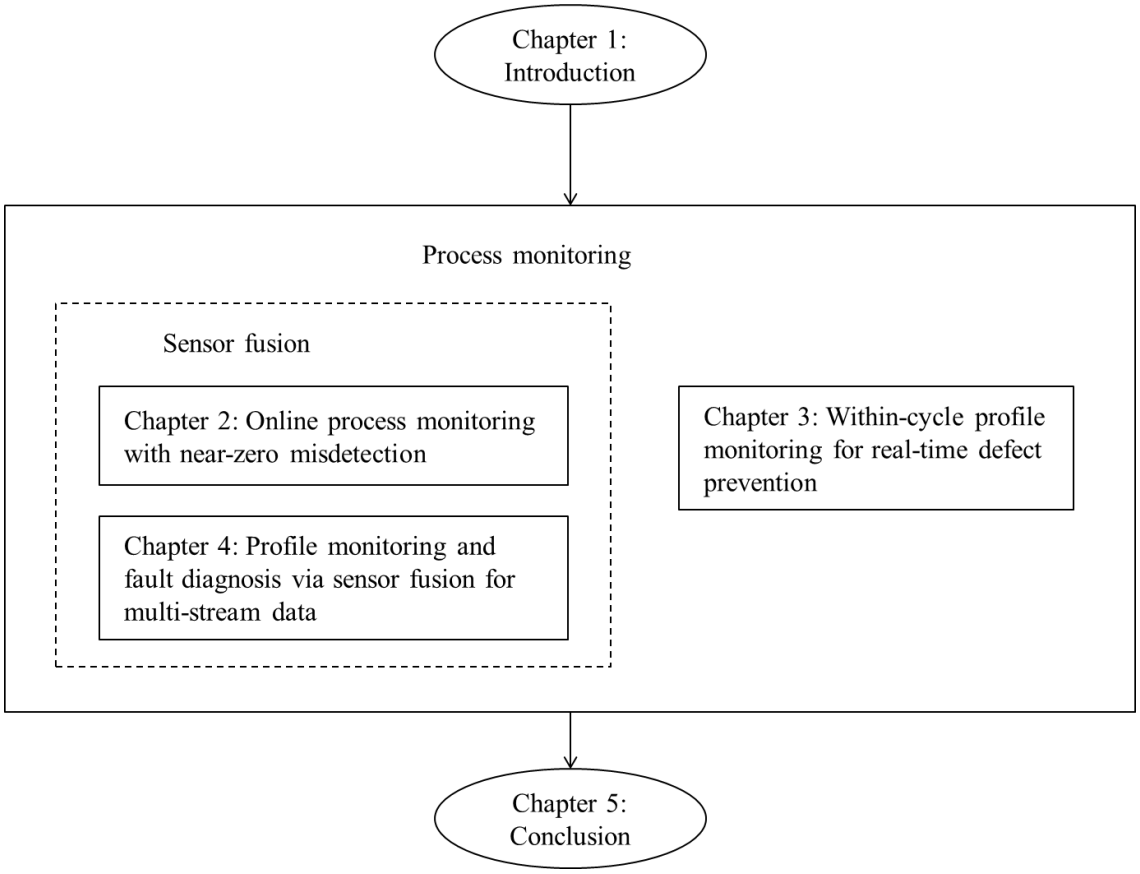


Figure 1.3. Dissertation organization

1.3.1 Online process monitoring with near-zero misdetection

In Chapter 2 of this dissertation, we develop an online monitoring algorithm that targets a near-zero Type II error rate foremost while maintaining a relatively low Type I error rate for the online quality monitoring system for ultrasonic welding of batteries. The features used in this monitoring method are extracted from online sensor signals based on the understanding of the process and domain knowledge.

In a typical battery assembly plant, the quality of the joint is inspected after the welding process through off-line manual inspection. This leads to delayed detection of low quality welds and a high manual inspection rate. According to Kim *et al.* (2011), weld quality can be classified into cold welds, good welds, and over welds. Both cold welds and over welds are considered problematic. Good welds have high peel strengths while problematic welds have low or medium peel strengths. We have observed from lab experiments and plant reports that a normal welding process, although with the presence of inherent variations, usually produces good welds. When the welding process is driven out-of-control due to assignable causes, e.g., metal surface contamination, improperly placed metal sheets, etc., bad welds are generated with a very high probability. So, the proposed monitoring system works as follows: whenever it detects an out-of-control sample, it would send an alarm to the downstream manual inspection, and the quality of that sample would then be verified by off-line inspection.

In the context of ultrasonic metal welding, Type I error occurs when the monitoring system announces a battery weld to be a suspect when it is actually in good quality, while Type II error occurs when the monitoring system fails to detect an out-of-control sample that turns out to be problematic. Thus, Type I error from the quality monitoring system results in unnecessary manual inspection efforts. On the other hand, passing a problematic weld will not only potentially impair the performance of the battery pack in electric vehicle, but also harm the performance and safety of the entire vehicle. Hence, Type II error results in passing

problematic weld to downstream processes, which is a dangerous consequence that should be avoided. Therefore, the online quality monitoring system for ultrasonic welding of batteries needs to achieve a near-zero Type II error rate foremost while maintaining a relatively low Type I error rate in order to ensure weld quality and reduce the manual inspection rate.

Developing a monitoring system for ultrasonic welding of batteries that satisfies the above requirements on Type I and Type II errors is very challenging. The smallest Type I error and the smallest Type II error cannot be achieved at the same time due to the trade-off between risks of getting these two types of errors. The conventional control chart techniques widely used in many process monitoring systems are designed to target a required Type I error rate. In operations where part quality is critical, a near-zero Type II error rate becomes the major goal for the monitoring system. It also needs a low Type I error to reduce manual inspections, but even a relatively high Type I error rate (e.g. 50%) represents a substantial reduction in manual inspection. Furthermore, the high frequency and short duration of ultrasonic welding process requires the real-time monitoring algorithm to be computationally efficient and its results to have good interpretability.

To address these challenges, Chapter 2 of this dissertation develops an online process monitoring algorithm ‘SPC-M’ by integrating univariate statistical process control method and the multivariate Mahalanobis distance approach. The acceptance region of the proposed SPC-M algorithm is the shared region of multiple univariate control limits and the Mahalanobis distance limit. In this way, SPC-M can be used to monitor multivariate processes in order to achieve a near-zero Type II error rate. The control limits for SPC-M algorithm are established and then tuned based on training data. The effectiveness of the proposed method is validated over a testing period in battery manufacturing.

1.3.2 Within-cycle profile monitoring for real-time defect prevention

Lithium-ion battery manufacturing is an example of the many applications that demand high product quality and reliability. This increasing demand calls for new methods to meet strict requirements in process monitoring and control for defect prevention. In Chapter 3 of this dissertation, a within-cycle profile monitoring method is developed for cycle-based profile data to facilitate real-time process monitoring and defect prevention. Cycle-based signals are collected from repetitive operational cycles of a discrete manufacturing process.

Traditional profile monitoring methods focus on between-cycle decisions, in the sense that the monitoring decision for each finished part is delivered after analyzing the entire cycle of signal, giving a time-delayed monitoring decision that is too late for within-cycle defect prevention. With strict quality requirement in mission-critical products like vehicle battery, process changes need to be detected before the part is finished so that corrections to the process could be made to prevent defects from occurring. In ultrasonic welding of battery tabs, weld defect prevention is critical since a low quality connection will potentially impair the performance of the battery pack in electric vehicle. It is possible, however, to adjust the clamping pressure in real time by installing external pneumatic pressure regulators to the existing ultrasonic welding machine. In the scenario of cold welds, the adjusted pressure would prevent defective joints and thus compensate the negative effects brought by the out-of-control operation. We call the monitoring strategy in this framework "within-cycle monitoring" as the monitoring decision is made within an operation cycle rather than at the end of the cycle. As opposed to the between-cycle monitoring approach, within-cycle profile monitoring shifts the detection of abnormal process conditions from post-manufacturing to real-time decisions during manufacturing. This gives real-time decision based on the analysis of an early portion of the cyclic signal; the monitoring decision can be then used to guide real-time process adjustment and enable defect prevention.

One of the biggest research challenges in within-cycle monitoring is determining the decision point. The optimal decision point should be able to balance the tradeoff between monitoring accuracy and the length of the remaining time available for control actions. On the one hand, reliable detection and accurate monitoring decision require a latter decision point so that a longer portion of profile data is included in the monitoring decision. On the other hand, early detection is desirable in order to leave sufficient time for process adjustment and control actions. This issue is not a concern in traditional control chart design and profile monitoring, and yet has not been investigated for complex profiles. Within-cycle monitoring of complex profile signals also raises challenges in modeling both between-profile and within-profile variations, detecting both profile mean shift and variance change, and effectively characterizing complex profile data.

In the proposed method, a new monitoring performance requirement is defined by considering both the monitoring accuracy and within-cycle control opportunity, and then an optimization problem is formulated and solved to find the optimal decision point. The proposed method adopts the wavelet transformation and the mixed-effect model to characterize complex profile data and capture both between-profile and within-profile variations. In order to effectively detect both profile mean shift and variance change, we further combine the wavelet-based mixed-effect model with control chart design on the monitoring of profile deviations. Simulations, sensitivity analyses, and case studies are also conducted to evaluate the performance of the proposed within-cycle profile monitoring algorithm and its effectiveness in defect prevention.

1.3.3 Profile monitoring and fault diagnosis for multi-stream data

When multiple sensors are installed to measure different variables in the welding process, we collect heterogeneous multi-stream signals. If a single signal does not provide enough information to effectively evaluate the performance of the process, multi-sensor fusion

methods are needed. In Chapter 4 of this dissertation, we develop a method based on a multilinear extension of the linear discriminant analysis to extract and analyze information from multi-stream profile data to facilitate profile monitoring and fault diagnosis.

A large portion of the existing multi-sensor data fusion methods is based on extracting a single synthetic index from the monitoring signals, e.g., a weighted summation of signals. The main limitations of this approach include the loss of information involved in the feature extraction process, the loss of sensor-to-sensor correlations, and the problem-dependent nature of the synthesizing scheme. Although profile monitoring techniques have been demonstrated to be more effective than synthetic index-based methods in monitoring processes characterized by repeating patterns (Noorossana *et al.*, 2012), only a few authors have studied profile monitoring approaches in the field of sensor fusion (Kim *et al.*, 2006; Amiri *et al.*, 2013; Chou *et al.*, 2014). Recently, with the fast development of multilinear methods for face recognition, Paynabar *et al.* (2013) proposed a multi-channel profile monitoring and fault diagnosis method based on uncorrelated multilinear principal component analysis (UMPCA) (Lu *et al.*, 2009), whereas Grasso *et al.* (2014) investigated the problem of multi-stream profile monitoring using multilinear PCA (MPCA) (Lu *et al.*, 2008). The major limitation of PCA-based methods is that they do not make use of the class information.

In Chapter 4 of this dissertation, we investigate the use of multilinear extensions of linear discriminant analysis (LDA) to deal with multi-stream signals for the purpose of process monitoring and fault diagnosis. One extension, vectorized-LDA (VLDA), involves vectorizing multi-stream profiles into a high-dimensional vector and then applies regular LDA on it; the other extension, uncorrelated multilinear LDA (UMLDA), operates directly on the multi-stream profiles and then extracts uncorrelated discriminative features through solving a tensor-to-vector projection. The disadvantages of the first extension include

creating high computational complexity, breaking the natural structure and correlation in the original data, and potentially losing more useful representations that can be obtained in the original form.

We propose a UMLDA-based approach for analyzing multi-stream profiles that considers the interrelationship of different sensors. The features extracted by the proposed UMLDA-based method can effectively discriminate different classes and provide fault diagnosis results. A simulation study is conducted to evaluate the performance of the proposed method and its performance superiority over VLDA and other competitor methods. The possibility of improving classification performance in fault diagnosis using ensemble learning with UMLDA is also explored. The effectiveness of the proposed method is further tested on a multi-layer ultrasonic metal welding process.

References

Amiri, A., Zou, C. and Doroudyan, M.H. (2013) Monitoring Correlated Profile and Multivariate Quality Characteristics. *Quality and Reliability Engineering International*, 30, 133-142.

Chou, S. H., Chang, S.I. and Tsai, T.R. (2014) On Monitoring of Multiple Non-linear Profiles. *International Journal of Production Research*, 52, 3209-3224.

Grasso, M, Colosimo, B.M. and Pacella, M. (2014) Profile Monitoring via Sensor Fusion: The Use of PCA Methods for Multi-Channel Data. *International Journal of Production Research*, 52, 6110-6135.

Kim, J., Huang, Q., Shi, J. and Chang, T.-S. (2006) Online Multichannel Forging Tonnage Monitoring and Fault Pattern Discrimination Using Principal Curve. *Journal of Manufacturing Science and Engineering*, 128, 944-950.

Kim, T.H., Yum, J., Hu, S.J., Spicer, J.P. and Abell, J.A. (2011) Process Robustness of Single Lap Ultrasonic Welding of Thin, Dissimilar Materials. *CIRP Annals - Manufacturing Technology*, 60, 17-20.

Lee, S.S., Kim, T.H., Hu, S.J., Cai, W.W. and Abell, J.A. (2010) Joining Technologies for Automotive Lithium-ion Battery Manufacturing – A Review. *Proceedings of ASME 2010 International Manufacturing Science and Engineering Conference*, 541-549.

Lu, H., Plataniotis, K.N. and Venetsanopoulos, A.N. (2008) MPCA: Multilinear Principal Component Analysis of Tensor Objects. *IEEE Transactions on Neural Networks*, 19, 18–39.

Lu, H., Plataniotis, K.N. and Venetsanopoulos, A.N. (2009) Uncorrelated Multilinear Discriminant Analysis With Regularization and Aggregation For Tensor Object Recognition. *IEEE Transactions on Neural Networks*, 20, 103-123.

Noorossana, R., Saghaei, A and Amiri, A. (2012) *Statistical Analysis of Profile Monitoring.* New York: Wiley.

Paynabar, K., Jin, J., and Pacella, M. (2013) Monitoring and Diagnosis of Multichannel Nonlinear Profile Variations Using Uncorrelated Multilinear Principal Component Analysis. *IIE Transactions*, 45, 1235-1247.

CHAPTER 2

ONLINE PROCESS MONITORING WITH NEAR-ZERO MISDETECTION: AN INTEGRATION OF UNIVARIATE AND MULTIVARIATE METHODS

2.1 Introduction

In recent years, increasing concerns over the environmental impact of the petroleum-based transportation infrastructure and soaring gas price have led to great interest in electric vehicles. Electric vehicles require high-power and high-capacity rechargeable batteries. In manufacturing such batteries, significant challenges exist in creating reliable interconnections between battery cells, between modules, and between modules and control units. Such connections must possess reliable electrical conductivity and sufficient mechanical strength to ensure battery performance.

Ultrasonic metal welding is used in joining lithium-ion batteries due to its advantages in joining dissimilar and conductive materials, as discussed by Kim *et al*. (2011). Ultrasonic metal welding is a solid-state bonding process which uses high frequency ultrasonic energy to generate oscillating shears between two metal sheets clamped under pressure (Lee *et al.*, 2010). After removing the surface films and oxides from the surface, the solid-state bond is formed through the plastic deformation of the contacting surfaces under high pressure (Kalpakjian and Schmid, 2008). As illustrated in Figure 1.1, during welding, the transducer transforms electrical energy into high frequency mechanical vibration; this mechanical vibration is transferred to a welding tip through an acoustically tuned horn. This high

frequency vibration, applied under force, disperses surface films and oxides, creating a metallurgical bond.

The performance of an entire battery pack may not be as desired if some battery joints have low quality connections. In order to ensure joint quality and not to pass any problematic welds to downstream processes, in a typical battery assembly plant, the quality of every single joint is inspected after the welding process through off-line manual inspection. This leads to delayed detection of low quality welds and a high manual inspection rate. The off-line quality inspection is a complex procedure that requires human operations and considerable time and labor in (a) visual inspection to ensure the welding spot is at the correct location, and (b) non-destructive mechanical test to ensure the bonding has sufficient strength. The cost of inspection becomes enormous when 100% manual inspection is performed on all welds. Therefore, the battery manufacturing processes used to join battery cells and modules must be equipped with online real-time quality monitoring and evaluation systems to ensure the quality of joining. This motivates our research to develop an online monitoring system for ultrasonic battery tab welding that can help reduce unnecessary manual inspection and ensure the quality of every weld. The monitoring system predicts the quality of each weld based on real-time sensor signals collected from the welding process.

Weld quality has been classified into cold welds, good welds, and over welds by Kim *et al*. (2011) through post-weld studies using the T-peel method. Both cold welds and over welds are considered problematic. Good welds have high peel strengths while problematic welds have low or medium peel strengths. We have observed from lab experiments and plant reports that a normal welding process, although with the presence of inherent variations, usually produces good welds. When the welding process is driven out-of-control due to assignable causes, e.g., metal surface contamination, improperly placed metal sheets, etc., bad welds are generated with a very high probability. For example, when the sheet metal is

contaminated with oil (or other substances), the welding power would not ramp up at the early stage as a normal weld does, thus resulting in a poor quality connection; if the metal sheets are improperly placed between the horn and anvil, the weld spot may fall on the edge of the sheets, also resulting in a poor quality connection; if one of the layers is bend when placing the sheets between the horn and anvil, the original input pressure may not be sufficient to make a strong connection on such an abnormal thickness. Therefore, it is important to detect process changes so that whenever the monitoring system detects an out-of-control sample, it would send a signal alarm to the downstream manual inspection, and the quality of that sample would then be verified by inspection.

Two types of errors may be committed by the monitoring system: false alarm, also known as the Type I error in hypothesis testing, and misdetection, also known as the Type II error. Specifically in this study, the Type I error occurs when the monitoring system announces a battery weld to be a suspect when it is actually in good quality; the Type II error occurs when the monitoring system fails to detect an out-of-control sample that turns out to be problematic. Thus, Type I error from the quality monitoring system results in unnecessary manual inspection efforts. On the other hand, passing a problematic weld will not only potentially impair the performance of the battery pack in electric vehicle, but also harm the performance and safety of the entire vehicle. Hence, Type II error results in passing problematic weld to downstream processes, which is a dangerous consequence that should be avoided. Therefore, the online quality monitoring system for ultrasonic welding of batteries needs to achieve a near-zero Type II error rate foremost while maintaining a relatively low Type I error rate in order to ensure weld quality and reduce the manual inspection rate.

Developing a monitoring system for ultrasonic welding of batteries that satisfies the above requirements on Type I and Type II errors is very challenging. The smallest Type I error and the smallest Type II error cannot be achieved at the same time due to the trade-off

between risks of getting these two types of errors. When a broader acceptance region is defined, it would reduce false alarms but increase misdetections; on the other hand, a narrower acceptance region reduces the risks of misdetection, but this increases the number of false alarms. The conventional control chart techniques widely used in many process monitoring systems are designed to target a required Type I error rate. In operations where part quality is critical, a near-zero Type II error rate becomes the major goal for the monitoring system. It also needs a low Type I error to reduce unnecessary manual inspections, but even a relatively high Type I error rate (e.g. 50%) represents a substantial reduction in manual inspection. Furthermore, the high frequency and short duration of ultrasonic welding process requires the real-time monitoring algorithm to be computationally efficient and its results to have good interpretability.

The objective of this chapter is to develop a monitoring algorithm that targets a near-zero Type II error rate foremost while maintaining a relatively low Type I error rate for the online quality monitoring system for ultrasonic welding of batteries. Specifically, the development of such a monitoring algorithm needs to effectively utilize sensor signals and integrate univariate and multivariate statistical process control methods. The developed monitoring algorithm will be used to help ensure part quality and reduce manual inspection costs in battery joining process and other mission-critical manufacturing processes as well. The remainder of this section briefly reviews existing methods on statistical process control. Section 2.2 describes the data collection procedure and the features used in this study. Detailed methodology on the proposed monitoring algorithm is presented in Section 2.3. Section 2.4 further demonstrates how the proposed monitoring algorithm works with a case study, followed by a discussion in Section 2.5. The conclusion is drawn in Section 2.6.

2.1.1 Literature review of the related work

In advanced manufacturing processes when near-zero Type II error rate is required, automated measurement technology is used, and every unit manufactured is analyzed. The univariate Shewhart control chart for individual measurements (Montgomery, 2013) is useful in such situations for separating assignable causes from chance causes. The individuals control chart detects out-of-control samples by setting 3-sigma control limits with

$$\hat{\sigma} = \frac{\overline{MR}}{d_2}\ , \tag{2.1}$$

where $\overline{MR}$ is the average of the moving ranges of two observations and d_2 is a constant representing the expected value of the relative range, and $d_2 = 1.128$ when a moving range of two observations is used. This method is based on the assumption that the observations follow a normal distribution. In many practical scenarios, however, the normality assumption cannot be met. Borror *et al.* (1999) found that the in-control average run length (ARL) is dramatically affected by non-normal data. One approach to dealing with the problem of non-normality is to determine the control limits for the individuals control chart based on the percentiles of the correct underlying distribution (Willemain and Runger, 1996). These percentiles could be obtained from a histogram when a large sample of observations is available.

Another limitation of the Shewhart individuals control chart is the restriction on the univariate perspective; that is, it is assumed that there is only one process output variable or quality characteristic of interest, or independence can be assumed among multiple output variables. In practice, however, most process monitoring and control scenarios involve several related variables, especially with the advances in sensing and data capturing technology in recent decades. Although applying univariate control charts to each individual

variable is possible, this is in fact inefficient and can lead to erroneous conclusions with a huge risk of misdetections. The Hotelling T^2 control chart for monitoring the mean vector of the process is a multivariate version of the Shewhart control chart (Hotelling, 1931). However, the Hotelling T^2 control chart is based on the assumption that the observations follow a multivariate normal distribution which may not hold in practice.

An exponentially weighted moving average (EWMA) control chart is robust to non-normal distributions (Borror *et al*., 1999). The multivariate exponentially weighted moving average (MEWMA) control chart is a multivariate version of EWMA control chart. The strength of MEWMA is in the capability of detecting small mean shifts (Stoumbos and Sullivan, 2002; Montgomery, 2013). However, MEWMA/EWMA charts are not advantageous in our problem due to the following reasons. First, MEWMA/EWMA charts are more effective than the Shewhart chart when the manufacturing process has a sustained small shift, which is not likely to exist in the ultrasonic welding process. A sustained shift usually results from tool wear, which is hardly a major concern in this study because the battery plant replaces tools based on a conservative strategy to minimize the effects of tool wear (Shao *et al*., 2014). Additionally, when compared to Shewhart control charts, the results from EWMA/MEWMA charts are not straightforward to interpret, whereas from the perspective of plant implementation, it is desirable that monitoring results should be easily interpreted and the thresholds are straightforward to adjust.

Mahalanobis distance (M-distance) measures the similarity of an unknown data set to a known one based on correlations between them (Mahalanobis, 1936). It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant. As a multivariate control chart technique, M-distance can act as the control limit to detect multivariate out-of-control situations (Bersimis *et al.*, 2007; De Maesschalck *et al.*, 2000). Moreover, M-distance method does not require the observations to follow normal

distributions. Using a robust estimator of the covariance matrix in calculating the M-distance would make the M-distance approach robust to non-normal distributed observations (Rousseeuw, 1984). More studies on M-distance include Mitchell and Krzanowski (1985), Barhen and Daudin (1995), De Maesschalck *et al*. (2000), Bedrick *et al*. (2000), and Xiang *et al*. (2008). For manufacturing processes where the observations are not highly correlated or the correlation varies during the process, however, using M-distance alone has a huge risk of misdetection. Further considerations on this issue are discussed in Section 2.5.

In summary, multiple univariate Shewhart individuals control charts can be used together for process monitoring if the features are known to be independent, while the M-distance approach is a good choice if the features are known to be highly correlated. However, in many manufacturing processes, the correlation structure among the features is unknown and may be varying over time. Now that the strict quality requirement of near-zero Type II error rate becomes paramount, there is a lack of statistical process monitoring technique to fully address this goal, but the development of a new method by integrating multiple univariate Shewhart-type charts with M-distance holds promise.

2.2 Data Collection Procedure and Data Description

This section describes the data collection procedure and the data and features used in this study. The ultrasonic welding system is controlled by several input parameters including weld time, energy, maximum power, tool displacement before vibration, and tool displacement after vibration. The welding control system automatically obtains these features in order to check the status of the welding, i.e., whether or not the welder is operating properly as the input parameters specify. When the welding system fails to achieve a targeted input parameter, the system sends an alarm for the welding process. However, these features, although easily obtained, only provide limited information about the welding process, but fail

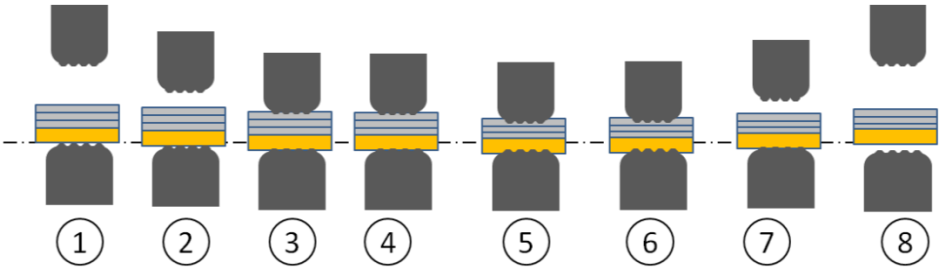
to reflect many processes changes that are caused by assignable causes such as sheet surface contamination, mislocated weld, etc., according to preliminary experiments. Therefore, the features provided by the welding system are not sensitive enough to separate problematic welds from high quality welds. Moreover, within-weld signals cannot be collected since these features do not show real-time information during the welding process. Preliminary analysis has shown that features from within-weld signals are important for process monitoring and quality prediction. For example, since the power signal rises as vibration starts and the surface films and oxides are removed from the surface, the slope of the rising of the power signal may be a good indicator of surface contamination, and some contamination may hinder the bonding of sheets and result in poor quality welds.

In order to gather physical process information for weld quality monitoring, Section 1.2 describes the additional sensors applied to collect both electrical and mechanical information during welding. Preliminary welding experiments and post-weld studies have shown that certain features from these online signals can be correlated with joint quality. Although some of the correlated features are identified in Lee *et al.* (2014), utilizing the signals systematically for weld quality monitoring remains a challenging task.

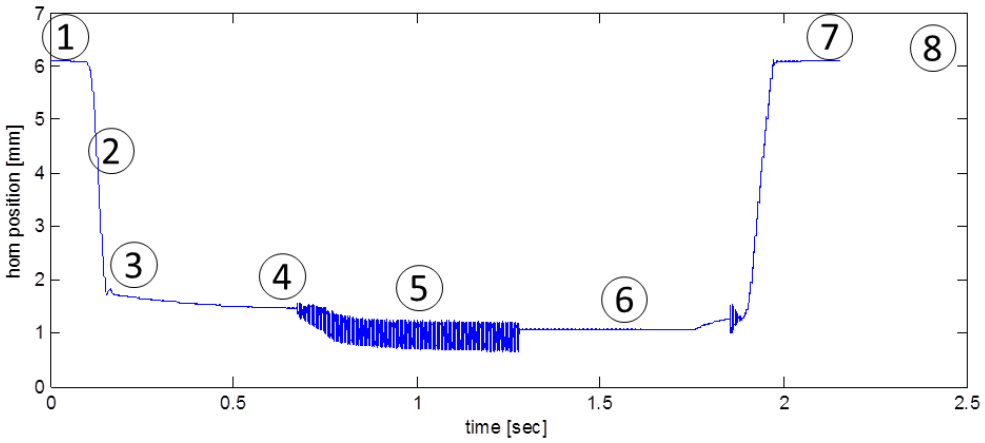
Each signal can be divided into eight segments based on the different stages of the ultrasonic metal welding operation. Figure 2.1(a) shows each stage of the operation and the positions of the horn and anvil with respect to the battery tab. A welding operation starts when the battery tab is placed on the anvil (①) and the horn starts to move down (②). The horn then touches the sheet surface (③) and starts to vibrate (④). As the vibration continues, deformation layer grows and so does the weld area (⑤). After welding, vibration stops (⑥) and the horn lifts up (⑦) to its default position (⑧). Figure 2.1(b) and (c) show different segments of the LVDT signal and the power signal corresponding to the different stages of the ultrasonic metal welding operation respectively. The relationship between signal segments

and operation stages is useful in interpreting the extracted features and relating them to the physics of the operation.

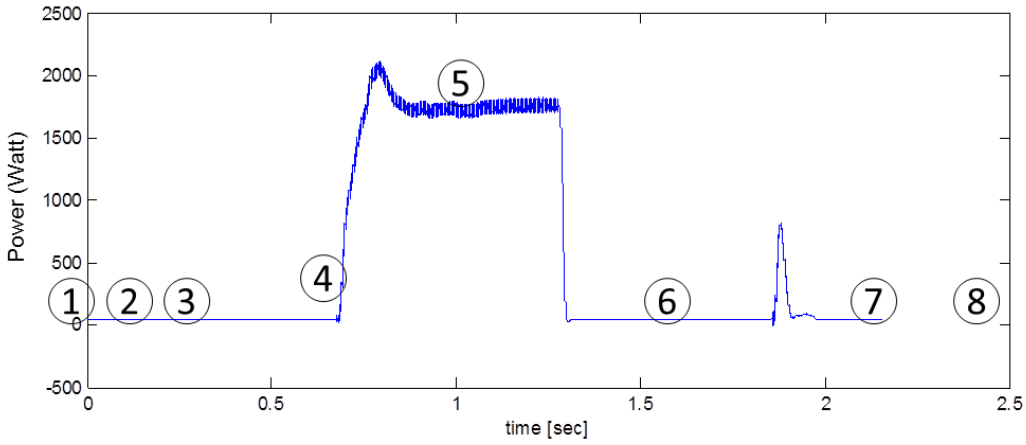
Although the welding time is short, around 0.5 sec, measured signals have a lot of data due to the high sampling rate, 100 kHz. These data are too large to be completely utilized for process monitoring and some of the data from different signals can be redundant as they contain similar process information. Hence, the raw signal data should be transformed or reduced into a set of features by feature extraction which still contains sufficient accuracy to represent the welding process information with good separability between bad welds and good welds. The feature extraction consists of transforming, mapping, simplifying, and filtering. Ten features are then selected for process monitoring, as listed in Table 2.1. Details on the selection methods of these features can be found at Shao *et al.* (2013).



(a) Different stages of ultrasonic metal welding operation



(b) Eight segments of LVDT signal



(c) Eight segments of power signal

Figure 2.1. Online signals and ultrasonic metal welding operation

Table 2.1. Features selected for process monitoring

Feature Index	Extracted From	Description
F1	LVDT signal	horn height before the main vibration
F2	LVDT signal	horn height after the main vibration
F3	power signal	maximum power value in the power signal
F4	welding control system	weld time during the main vibration
F5 ~ F9	LVDT signal	additional features
F10	power signal	

Training data was collected from an ultrasonic metal welding station over a 4-month period. Sensor signals were recorded and the features listed in Table 2.1 were extracted. Currently, most of welds produced are good and only a few cold welds are found occasionally. Thus, we focus on those cold welds as bad welds in our study. The total training data sample size is 23481 with 23437 good welds (99.81%) and 44 bad welds (0.19%). The small number of sample faults brings more challenges in establishing a threshold for the monitoring system. The proposed method was then tested on a 1-month period on the same station. The total test data sample size is 11507 with 11490 good welds (99.85%) and 17 bad welds (0.15%). The Kolmogorov-Smirnov normality test on each feature gives a very small

p-value (less than 0.01), which indicates that these features do not follow the normal distribution. As an example, Figure 2.2 shows the pre-processed values of feature F3 within a segment of the training period. Note these are pre-processed values showing the residuals of moving averages, instead of raw feature values. By taking the residuals of moving averages rather than the raw values in feature analysis, we can smooth out short-term fluctuations which are caused by material build-up, tool wear, etc. and highlight longer-term trends or cycles. The light grey dots in Figure 2.2 represent good welds, while the dark circles represent bad welds.

The above datasets were collected along with a quality report from the 100% manual inspection in plant. The plant quality report provides information (good/bad) on the quality of each weld. Although the quality report serves as a baseline for developing monitoring system, the report may not be completely accurate due to possible operator errors. Hence, special care should be taken during monitoring algorithm development stage whenever the detection result disagrees with the quality report.

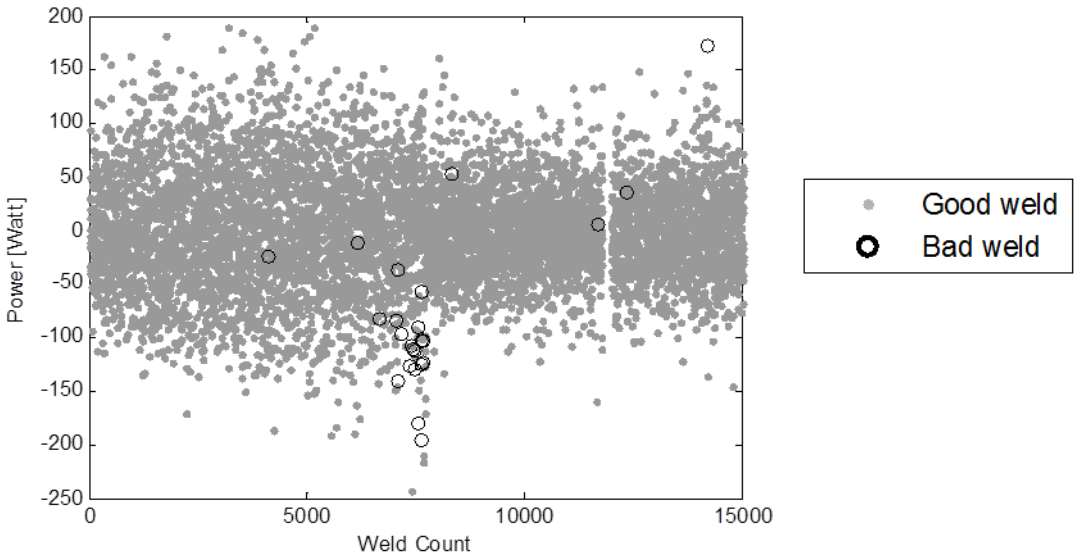


Figure 2.2. Distribution of feature F3 in residuals of moving averages

2.3 Methodology

2.3.1 Overview of proposed methodology

The proposed method is depicted in the flowchart in Figure 2.3. The methodology consists of two main phases: (1) offline control limits training to achieve near-zero Type II error and (2) online quality monitoring using the tuned control limits. In the training phase, an "SPC-M algorithm" is developed by integrating univariate Shewhart-type control charts and the M-distance approach. The training dataset is used to construct the control limits for the SPC-M algorithm, which includes multiple univariate control limits for the Shewhart individuals control charts and control limits for the M-distance approach. The SPC-M control limits are tuned in order to achieve near-zero sample Type II error.

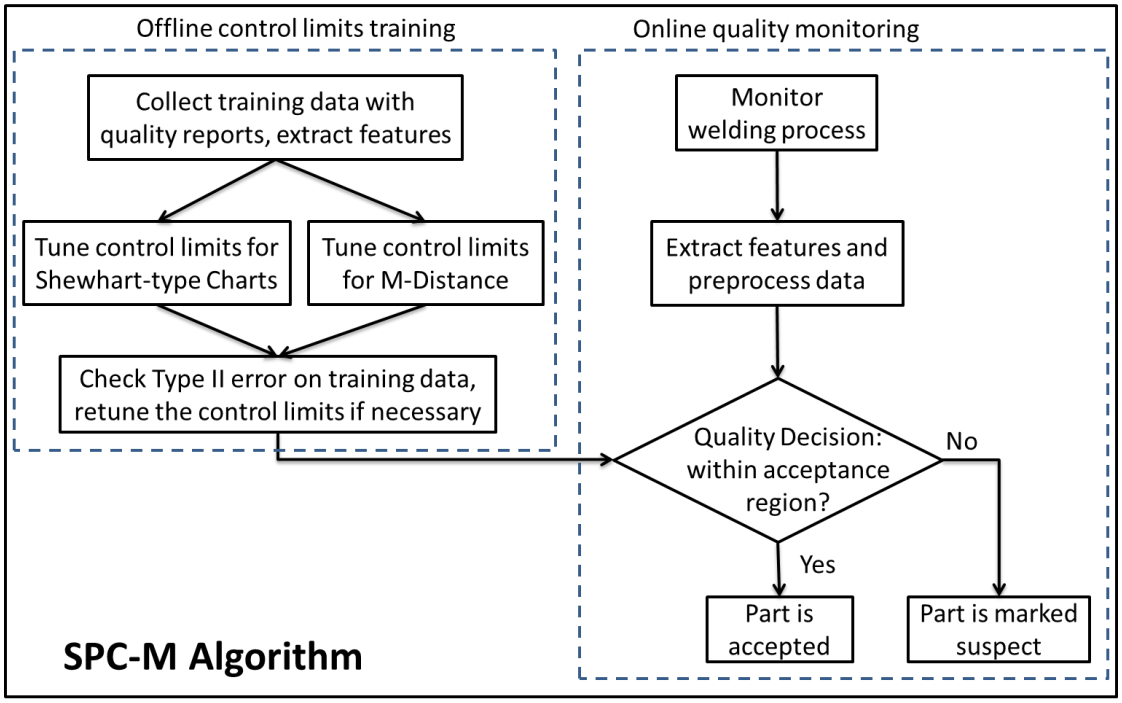


Figure 2.3. Overview of the proposed SPC-M methodology

In the second phase of online quality monitoring, welding process continues under the same operating condition as the previous period. For each test sample, features are extracted in the same way as discussed in Section 2.2. These features are then compared with the SPC-M control limits obtained from the training phase. If the test value falls into the acceptance region of SPC-M algorithm, we accept this weld as a good weld. Otherwise, the weld is marked as a suspect weld, and an alarm signal is generated. This weld will be sent to the manual inspection station for further quality check. The details of the SPC-M algorithm are explained in the following subsections.

2.3.2 SPC-M algorithm

The SPC-M algorithm is developed by integrating univariate Shewhart-type control charts and the M-distance approach to construct an integrated acceptance region. Let us look at a 2-feature case for example.

In Figure 2.4, the rectangles represent the acceptance regions given by a set of tight control limits and a set of slack control limits in the individuals control charts. For example, the tight acceptance region may be given by 1σ control limits and the slack acceptance region may be given by 3σ limits; the tight acceptance region may be given by 0.15 probability limits and the slack region may be given by 0.005 probability limits. The ellipse in Figure 2.4 represents the acceptance region given by a threshold from the multivariate M-distance approach. As illustrated in Figure 2.4, the acceptance region of the SPC-M algorithm consists of two regions, represented as A and B. Region A is the tight acceptance region from the multiple univariate control limits in individuals control charts. If the features fall into this region, it indicates that the feature values are right at the target with very small variability and that the weld should be considered as in good quality. Region B is the region excluded from region A but governed by both the multivariate M-distance's threshold and a slack acceptance region from the multiple univariate control limits in individuals control charts. If the features

fall into region B, it indicates that their values exhibit variability but are acceptable in both univariate and multivariate aspects.

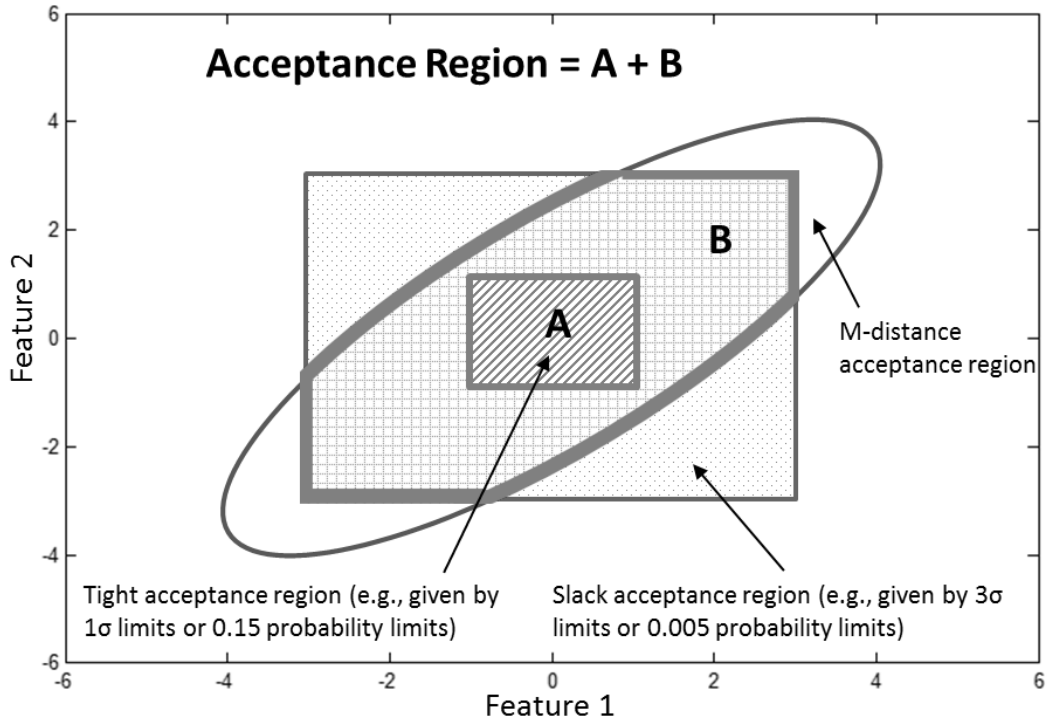


Figure 2.4. Acceptance region of SPC-M algorithm with 3σ control limits

Considering data non-normality, we specify the control limits for individuals control charts by using probability limits instead of *k*-sigma limits. For the M-distance approach, we adopt a robust estimator of covariance (Minimum Covariance Determinant (MCD) estimator, Rousseeuw, 1984) to make the method more robust to outliers and more reliable. The control limits are then tuned to achieve near-zero Type II error throughout the training period. Therefore, the integrated acceptance region, A+B, will be capable of monitoring multivariate observations with robustness to non-normal distributions and achieving near-zero Type II error rate. Details of the tuning of control limits are explained in the next subsection.

2.3.3 Tuning control limits for the SPC-M algorithm

Let f denote the feature index, $f = 1, 2, \dots, F$ and $F = 10$ is the number of features in this study. Let n_g denote the total number of good welds, n_b denote the total number of bad welds in the training dataset; the total number of welds in the training dataset is

$$n_{training} = n_g + n_b \,. \tag{2.2}$$

For the good welds, let $g_{i,f}$ denote the value of feature f of weld i, $i = 1, 2, \dots, n_g$. Then the values of feature f of all good welds form a vector $G_f = [g_{1,f} \quad g_{2,f} \quad \cdots \quad g_{n_g,f}]^T$. The feature values for all good welds can be represented as $G = [g_{i,f}]_{n_g \times F}$. Similarly, matrix $B = [b_{j,f}]_{n_b \times F}$ represents the feature values for all bad welds, where $b_{j,f}$ denotes the value of feature f of weld j, $j = 1, 2, \dots, n_b$. All training data can be represented as

$$\mathrm{Y}_{training} = \begin{bmatrix} G \\ B \end{bmatrix} = [y_{k,f}]_{n_{training} \times F} \,, \tag{2.3}$$

where $y_{k,f}$ denotes the value of feature f of weld k, $k = 1, 2, \dots, n_{training}$.

In setting up the acceptance region on univariate Shewhart-type control charts, each feature is considered individually. For G_f of feature f, we construct univariate control limits $UCL_{1,f}$, $LCL_{1,f}$, $UCL_{2,f}$, and $LCL_{2,f}$ based on percentiles p_{1u}, p_{1l}, p_{2u}, and p_{2l}. As illustrated in Figure 2.5, $UCL_{1,f}$ and $LCL_{1,f}$ are the upper and lower control limits derived from percentile $(1 - p_{1u})$ and p_{1l}, respectively, and $UCL_{2,f}$ and $LCL_{2,f}$ are the upper and lower control limits derived from percentile $(1 - p_{2u})$ and p_{2l}, respectively. These two sets of control limits represent a tight acceptance region and a wide acceptance region in the univariate control charts. With the large number of training samples available, the percentile-based control limits can be obtained from the histogram of G_f. More specifically,

$$\begin{cases} UCL_{1,f} = (1 - p_{1u}) \times 100^{th} \text{ percentile of } G_f \\ LCL_{1,f} = (p_{1l}) \times 100^{th} \text{ percentile of } G_f \end{cases},$$

$$\begin{cases} UCL_{2,f} = (1 - p_{2u}) \times 100^{th} \text{ percentile of } G_f \\ LCL_{2,f} = (p_{2l}) \times 100^{th} \text{ percentile of } G_f \end{cases}. \quad (2.4)$$

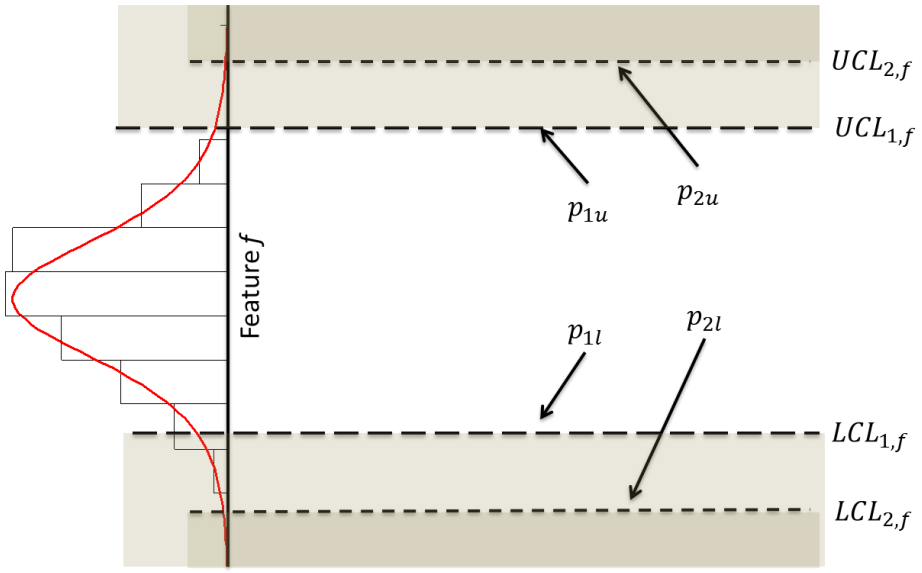


Figure 2.5. Univariate control limits used in SPC-M

For example, setting $p_{1u} = p_{1l} = 0.1587$ and $p_{2u} = p_{2l} = 0.00135$ correspond to the 1σ and 3σ control limits of Shewhart-type control limits under a normal distribution. For our training data, the univariate control limits can be determined by tuning p_{1u}, p_{1l}, p_{2u}, and p_{2l}. Note that p_{1u} and p_{1l}, p_{2u} and p_{2l} do not need to be set equal if the underlying distribution is not symmetric.

Now we consider all features together and deal with multivariate data to set up the acceptance region on M-distance part. The multivariate feature value for a good sample i in the training dataset is $x_i = [g_{i,1} \quad g_{i,2} \quad \cdots \quad g_{i,F}]^T$. Let $\bar{G}_f$ denote the average value in G_f. Then the multivariate sample mean in the training dataset is $\bar{x} = [\bar{G}_1 \quad \bar{G}_2 \quad \ldots \quad \bar{G}_F]^T$. A robust covariance S is estimated using the Minimum Covariance Determinant, as explained

by Rousseeuw (1984). The Mahalanobis distance of multivariate vector x_i from the group of good samples with mean $\bar{x}$ and covariance matrix S is given by

$$d_i = \sqrt{(x_i - \bar{x})^T \cdot S^{-1} \cdot (x_i - \bar{x})}\,. \tag{2.5}$$

The M-distance for all good welds in the training dataset is then given by $D_{good} = \begin{bmatrix} d_1 & d_2 & \cdots & d_{n_g} \end{bmatrix}^T$. The M-distance control limit CL_M can be then constructed based on a percentile of p_M with the histogram of D_{good}. More specifically,

$$CL_M = (1 - p_M) \times 100^{th} \text{ percentile of } D_{good}\,. \tag{2.6}$$

For our training data, we can tune p_M to manipulate the control limit on M-distance, which is a multivariate measure of the features.

The parameters to be tuned in the training stage form a vector

$$\boldsymbol{p} = (p_{1u}, p_{1l}, p_{2u}, p_{2l}, p_M)\,. \tag{2.7}$$

For the entire training dataset, $\mathrm{Y}_{training}$ as defined in Eq. (2.3), we tune each element in $\boldsymbol{p}$ to achieve near-zero Type II error and also keep the Type I error as small as possible. The initial values of $\boldsymbol{p}$ may be set based on sample histogram. The SPC-M algorithm with control limits tuning are elaborated in Figure 2.6. The multiple univariate control limits $UCL_{1,f}$, $LCL_{1,f}$, $UCL_{2,f}$, and $LCL_{2,f}$, and the M-distance control limit CL_M together establish the acceptance region for SPC-M. This integrated acceptance region is robust to weld quality and can be tuned to achieve zero Type II error.

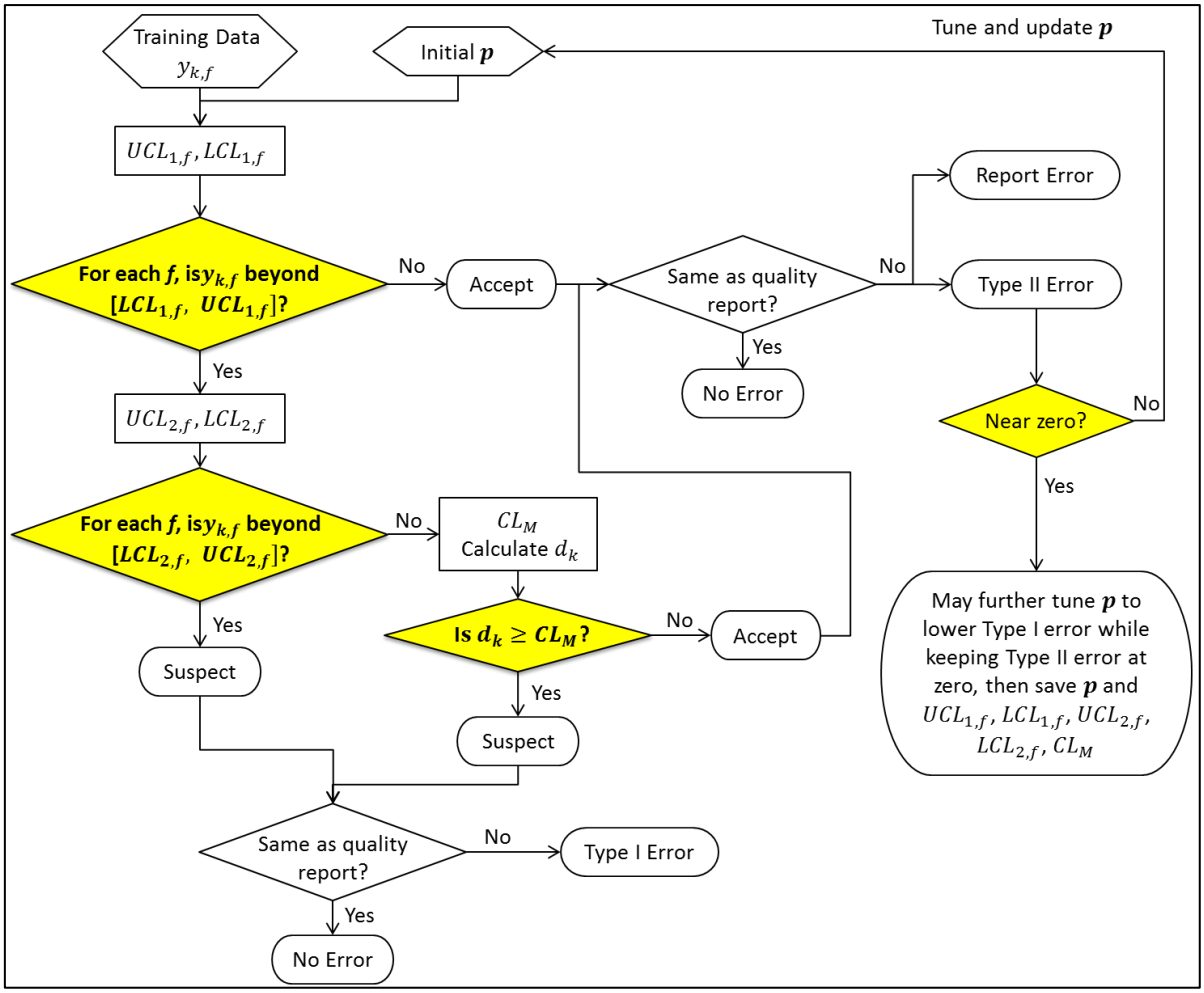


Figure 2.6. Flowchart of the SPC-M algorithm and control limits tuning

The multivariate feature value for a sample k is $y_k = [y_{k,1} \quad y_{k,2} \quad \cdots \quad y_{k,F}]^T$. For each feature f, we first compare $y_{k,f}$ with the tight control limits $[LCL_{1,f}, UCL_{1,f}]$ (similar to region A in Figure 2.4). If $y_{k,f}$ is within $[LCL_{1,f}, UCL_{1,f}]$ for all $f = 1, 2, \ldots, F$, we accept this weld k. If there is a $y_{k,f}$ beyond $[LCL_{1,f}, UCL_{1,f}]$, we then check it with the wide control limits $[LCL_{2,f}, UCL_{2,f}]$. If there is a $y_{k,f}$ beyond $[LCL_{2,f}, UCL_{2,f}]$, we then reject weld k as a suspected problematic weld. Otherwise, we estimate the M-distance from weld k to the group of good samples by

$$d_k = \sqrt{(y_k - \bar{x})^T \cdot S^{-1} \cdot (y_k - \bar{x})}\,, \tag{2.8}$$

and then check d_k with the M-distance control limit CL_M. If $d_k \geq CL_M$, we reject weld k and consider it as a suspected problematic weld; otherwise we accept it (similar to region B in Figure 2.4). After quality prediction using this set of parameters $\boldsymbol{p}$, we check the performance with quality reports to see whether Type II error is near-zero or not. Since the plant quality report may not be completely accurate, whenever a detection error is generated, detailed investigation will be performed through manual inspection and possible report errors will also be recorded. If not, we should tune $\boldsymbol{p}$ and update the acceptance region until near-zero Type II error is achieved. We may then try to further tune $\boldsymbol{p}$ to lower Type I error while keeping Type II error at zero.

2.3.4 Testing the SPC-M algorithm

Test data $\mathrm{Y}_{test} = [y_{k,f}]_{n_{test} \times F}$ is obtained as production continues, and $y_k = [y_{k,1} \quad y_{k,2} \quad \cdots \quad y_{k,F}]^T$ is the multivariate feature value for weld k in the test dataset. For each feature f, we compare $y_{k,f}$ with tight control limits $[LCL_{1,f}, UCL_{1,f}]$ and wide control limits $[LCL_{2,f}, UCL_{2,f}]$, which are obtained from previous analysis on training dataset. For each weld within $[LCL_{2,f}, UCL_{2,f}]$ but beyond $[LCL_{1,f}, UCL_{1,f}]$, we then check its M-distance with CL_M obtained from training data analysis. The M-distance from weld k to the group of good samples, d_k, is calculated according to Eq. (2.8), where $\bar{x}$ and S are also obtained from the good samples in training data. Similar to the flowchart in Figure 2.6, a weld k is rejected if there is a $y_{k,f}$ beyond $[LCL_{2,f}, UCL_{2,f}]$ or if $d_k \geq CL_M$.

2.4 Case Study

In this section, we apply the SPC-M algorithm developed in the previous section to the dataset described in Section 2.2 and show the results in (1) designing control limits in the offline training stage and (2) evaluating monitoring performance.

2.4.1 SPC-M algorithm design

Following the developed SPC-M algorithm with our training dataset, we determined $\boldsymbol{p}$ to achieve zero Type II error. The percentile values are shown in Table 2.2.

Table 2.2. SPC-M algorithm parameters

M-distance threshold	Tail probability: $p_M = 0.08$
Univariate control charts limits	Tight limits with $p_{1u} = p_{1l} = 0.15$ Wide limits with $p_{2u} = p_{2l} = 0.005$

With this acceptance region, we can achieve an overall Type I error rate of 9.8% and a Type II error rate of 0% in the training dataset. Table 2.3 shows the training data performance in a confusion matrix. Among the 23437 good welds, 2290 welds are detected as problematic from our SPC-M algorithm, resulting in an overall Type I error rate of 9.8%. Our algorithm detects all the 44 bad welds with zero Type II error. The daily Type I error rate through the training period is shown in Figure 2.7.

Table 2.3. Confusion matrix for training data

		Predicted quality			**Overall detection error**
		Good	Bad	Total	
True quality	Good	21147	2290	23437	**Type I error rate = 9.8%** **Type II error rate = 0%**
	Bad	0	44	44	
	Total	21147	2334	23481	

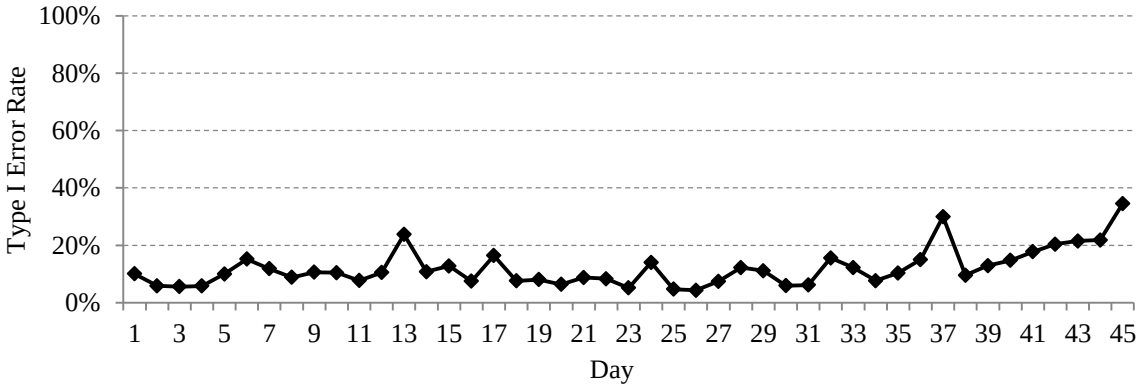


Figure 2.7. Daily Type I error rate through training period

We would like to recommend a few guidelines to help with parameter tuning. (1) For univariate control limits, the tail probability for the tight limits should be between 0.05 and 0.4 ($0.05 \leq p_{1u}, p_{1l} \leq 0.4$), whereas the tail probability for the slack limits should be between 0 and 0.05 ($0 \leq p_{2u}, p_{2l} \leq 0.05$). A good starting point for $\{p_{1u}, p_{1l}, p_{2u}, p_{2l}\}$ would be $p_{1u} = p_{1l} = 0.1587$ and $p_{2u} = p_{2l} = 0.0062$, which are equivalent to the 1σ and 3σ limits in traditional Shewhart-type control charts. (2) Among all parameter candidates that achieve zero Type II error in the method training stage, we select the set of parameters that reduces the training Type I error rate to a certain extent. Minimizing the training Type I error is not a rigid objective in parameter tuning due to three reasons: first, the computational complexity brought by the optimization problem is high; second, it is desirable to avoid over-tuned parameters; and third, even a relatively high Type I error rate (e.g. 50%) represents a substantial reduction in manual inspection.

2.4.2 Monitoring performance evaluation

As mentioned in Section 2.2, the sample size of test data is 11507 with 11490 good welds (99.85%) and 17 bad welds (0.15%). With the SPC-M algorithm parameters in Table 2.2, a zero Type II error rate was achieved through testing period, as shown in the confusion matrix in Table 2.4. Among the 11490 good welds, 2109 welds are detected as problematic from our algorithm, resulting in an overall Type I error rate of 18.4%.

Table 2.4. Confusion matrix for test data

		Predicted quality			**Overall detection error**
		Good	Bad	Total	
True quality	Good	9381	2109	11490	**Type I error rate = 18.4%** **Type II error rate = 0%**
	Bad	0	17	17	
	Total	9381	2126	11507	

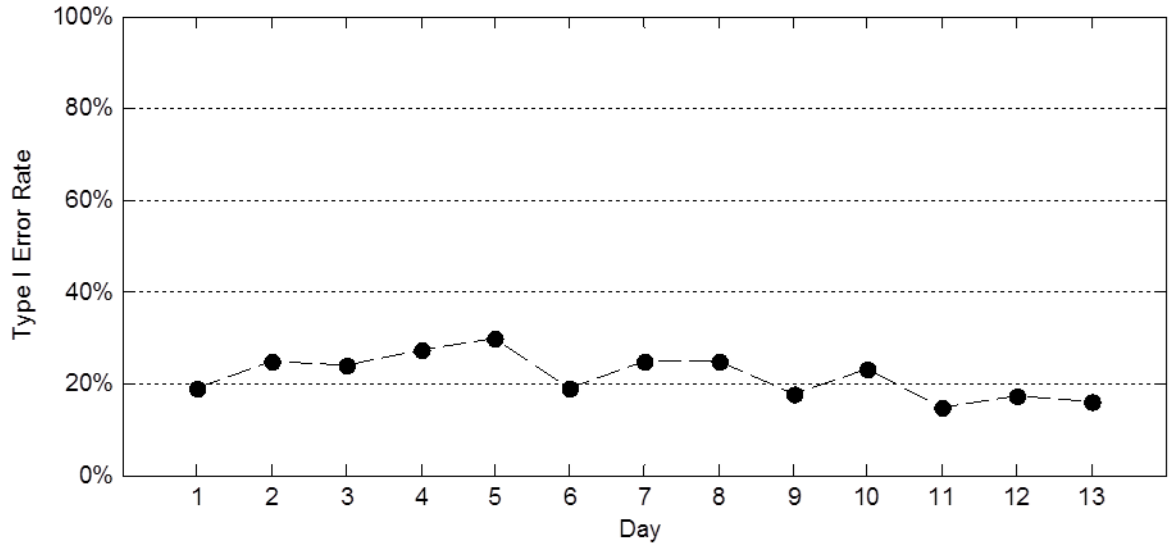


Figure 2.8. Daily Type I error rate through testing period while maintaining zero Type II error

Figure 2.8 shows the daily Type I error rate through the testing period. The maximum daily Type I error rate is less than 30% and the minimum daily Type I error rate is around 15%. We know that the ultrasonic process is influenced by various factors such as tool wear, surface condition of the workpiece, and mechanical constraint of the workpiece. These factors cause the variability in Type I error rates. Even with this variability, the manual inspection rate is reduced more than 70% without changing tuning parameters of the SPC-M algorithm. More importantly, the SPC-M algorithm did not accept any bad welds in the testing period. Although the number of bad welds varied every day, a zero Type II error rate was achieved (all 17 bad welds were detected). This indicates that the proposed method performs very well in both the training and the testing periods.

We further conducted a comparative study using either individuals control chart or M-distance approach alone. The control limits were tuned so that zero Type II error can be achieved in training period. Among all parameter candidates that meet this criterion, we selected the parameter(s) that can minimize the training Type I error rate. Table 2.5 shows the training performance and test performance in the comparative study. Using the SPC

individuals control charts alone, the Type I error rates in both training and testing periods are larger than those from SPC-M algorithm, whereas the Type I error rate from M-distance approach alone is slightly smaller than that from SPC-M. However, both these competitor methods have incurred huge costs when one bad weld in the testing period is misdetected. Comparing Table 2.5 with Table 2.3 and Table 2.4 demonstrates the superiority of using SPC-M algorithm in our case over SPC individuals control charts or M-distance approach.

Table 2.5. Performance comparison: SPC and M-distance

Method: SPC individuals control chart with $p_u = 0.02, p_l = 0$								
Training performance		Predicted quality			Test performance		Predicted quality	
		Good	Bad	Total			Good	Bad	Total
True quality	Good	20317	3120	23437	True quality	Good	9016	2474	11490
	Bad	0	44	44		Bad	1	16	17
	Total	20317	3164	23481		Total	9017	2490	11507
Overall detection error	**Type I error rate = 13.3% Type II error rate = 0%**				**Overall detection error**	**Type I error rate = 21.5% Type II error rate = 5.9%**			
Method: M-distance approach with a threshold of tail probability $p_M = 0.09$								
Training performance		Predicted quality			Test performance		Predicted quality	
		Good	Bad	Total			Good	Bad	Total
True quality	Good	21327	2110	23437	True quality	Good	9510	1980	11490
	Bad	0	44	44		Bad	1	16	17
	Total	21327	2154	23481		Total	9511	1996	11507
Overall detection error	**Type I error rate = 9.0% Type II error rate = 0%**				**Overall detection error**	**Type I error rate = 17.2% Type II error rate = 5.9%**			

2.5 Discussion

The developed SPC-M algorithm has been demonstrated to work effectively on the ultrasonic welding of batteries. In this section, we further discuss the integrated algorithm so as to provide more guidelines to practitioners.

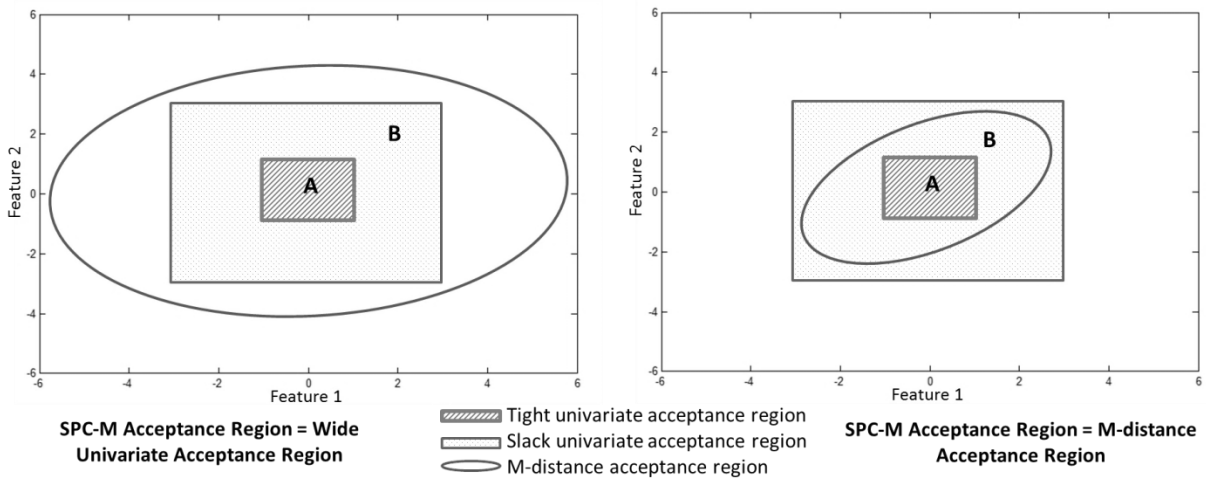
As mentioned in Section 2.3.3, the parameters to be tuned in the method training stage is a vector $\boldsymbol{p} = (p_{1u}, p_{1l}, p_{2u}, p_{2l}, p_M)$. Although the entire dataset has 10 features, it is

assumed that the univariate control chart for each feature uses the same set of parameters $\{p_{1u}, p_{1l}, p_{2u}, p_{2l}\}$ to establish probability limits. It is feasible and applicable, however, to have different percentiles for each feature, i.e., $\{p_{1u}, p_{1l}, p_{2u}, p_{2l}\}_{f=1}^{10}$, which be tuned based on the feature's sensitivity. Intuitively, this would give more flexibility in constructing the integrated acceptance region, and possibly more accurate detections. On the other hand, the disadvantages of introducing different sets of parameters for each feature are also obvious: the tuning of 40~50 parameters simultaneously brings high computational complexity, probability limits in univariate control charts would be difficult to interpret, and that the tuned parameters may be overly case-dependent and thus method generality is lost. Therefore, we would like to recommend $\boldsymbol{p} = (p_{1u}, p_{1l}, p_{2u}, p_{2l}, p_M)$ as tuning parameters in the training stage, but we also acknowledge the feasibility of having more parameters in SPC-M.

It is demonstrated in previous sections that the SPC-M algorithm is a data-driven technique targeting a near-zero sample Type II error rate and it is robust to non-normal distributed observations. The population error rate is estimated from the sample error rate based on training dataset and the control limits. In our case, the population Type I error rate is estimated as $\hat{\alpha} = 1 - (1 - \hat{\alpha}_I)^F \cdot (1 - \hat{\alpha}_M) = 1 - (1 - p_2)^{10} \cdot (1 - p_M) = 16.8\%$, where $\hat{\alpha}_I$ is the estimated sample Type I error rate from a Shewhart individuals chart and $\hat{\alpha}_I = p_2$; $\hat{\alpha}_M$ is the estimated sample Type I error rate from the M-distance approach and $\hat{\alpha}_M = p_M$. This gives a general idea of the false alarm rate when a near-zero sample Type II error rate is achieved.

The integrated acceptance region is most suitable for monitoring manufacturing processes when the correlation among multiple features is not very high and possibly varies as the process continues. Hence, the SPC-M algorithm can be replaced by multiple Shewhart individuals control charts if the features are known to be independent of each other

throughout the process; while the SPC-M algorithm can be replaced by the M-distance approach alone if the features are known to be highly correlated throughout the process. Figure 2.9 illustrates the two boundary scenarios for the 2-feature example. Given strict independence among the two features will force the elliptic acceptance region from M-distance to be completely outside the rectangular acceptance region from univariate control charts, as shown in Figure 2.9(a). Let $p_M = 0$ if the M-distance threshold is not necessary, then we have $\hat{\alpha} = 1 - (1 - \hat{\alpha}_I)^F \cdot (1 - \hat{\alpha}_M) = 1 - (1 - p_2)^{10}$ as the estimated population Type I error rate. On the other hand, a high correlation among the two features will force the elliptic acceptance region from M-distance to be completely inside the rectangular acceptance region from univariate control charts, as shown in Figure 2.9(b). Let $p_2 = 0$ if the Shewhart individuals control limits are not necessary, then we have $\hat{\alpha} = 1 - (1 - \hat{\alpha}_I)^F \cdot (1 - \hat{\alpha}_M) = 1 - (1 - p_M) = p_M$ as the estimated population Type I error rate. In our case, however, the SPC-M algorithm proves to be much better than using either individuals control chart or M-distance approach alone.



(a) SPC-M algorithm replaced by Shewhart individuals control chart

(b) SPC-M algorithm replaced by M-distance approach

Figure 2.9. Boundary scenarios of the SPC-M algorithm

2.6 Conclusion

In this study, we address a critical issue in weld quality monitoring with near-zero Type II error rate and low manual inspection rate for ultrasonic metal welding process in a battery assembly plant. The conventional control chart techniques cannot achieve the near-zero Type II error rate that is desired in monitoring the battery joining process. To deal with this problem, we developed an SPC-M monitoring algorithm by integrating univariate statistical process control method and the Mahalanobis distance approach. The SPC-M algorithm has an integrated acceptance region which is the shared region of multiple univariate control limits and the Mahalanobis distance limits. In this way, the monitoring algorithm can be used to monitor multivariate processes in order to achieve near-zero Type II error rate. The control limits for SPC-M algorithm was tuned based on training data. Then the algorithm was validated on test data from battery manufacturing. The results showed that the SPC-M algorithm achieved a 21.6% Type I error rate and 0% Type II error rate overall. Comparative studies also demonstrated the superiority of using SPC-M algorithm in our case over SPC individuals control chart or M-distance approach. With 0% Type II error rate, the SPC-M algorithm did not pass any suspected bad welds to downstream processes. The monitoring algorithm also proves to be robust against process variations such as tool wear, surface condition of the workpiece, and mechanical constraint of the workpiece, as evidenced by the 0% overall Type II error rate over the testing period.

In our study, the size of the training dataset is sufficiently large thanks to the rapid production rate in plant. Generally, 200 good samples are needed to find the control limits based on percentiles obtained from a histogram. However, a certain number of poor quality samples are also needed to help tuning the control limits to achieve near-zero Type II error rate. It is straightforward that the more poor quality samples there are in the offline training

stage, the better the control limits will be tuned for the online monitoring stage. Therefore, developing an adaptive training scheme for SPC-M has the potential to shorten the algorithm training period and reduce data collection cost, and thus is an interesting topic for future research. Furthermore, associating detection errors with cost analysis would be a valuable development for online monitoring that is worth future research efforts.

Utilizing this SPC-M algorithm, the near-zero misdetection monitoring system can be applied to the monitoring of many mission-critical processes. The development of the SPC-M algorithm should assist manufacturing and quality engineers and in their decisions to specify good welds and problematic welds in a more cost-effective manner.

References

Barhen, A. and Daudin, J.J. (1995) Generalization of the Mahalanobis Distance in the Mixed Case. *Journal of Multivariate Analysis*, 53, 332-342.

Bedrick, E.J., Lapidus, J. and Powell, J.F. (2000) Estimating the Mahalanobis Distance from Mixed Continuous and Discrete Data. *Biometrics*, 56, 394-401.

Bersimis, S., Psarakis, S. and Panaretos, J. (2007) Multivariate Statistical Process Control Charts: an Overview. *Quality and Reliability Engineering International*, 23, 517-543.

Borror, C.M., Montgomery, D.C. and Runger, G.C. (1999) Robustness of the EWMA Control Chart to Non-Normality. *Journal of Quality Technology*, 31, 309-316.

BRANSON, BRANSON Ultraweld® L20. (2007) BRANSON Ultrasonics Corporation, in: http://www.bransonultrasonics.com.

De Maesschalck, R., Jouan-Rimbaud, D. and Massart, D.L. (2000) The Mahalanobis Distance. *Chemometrics and Intelligent Laboratory Systems*, 50, 1-18.

Hotelling, H. (1931) The Generalization of Student's Ratio. *The Annals of Mathematical Statistics*, 2, 360-378.

Kalpakjian, S. and Schmid, S.R. (2008) *Manufacturing Processes for Engineering Materials*. Pearson Education, Upper Saddle River, N.J.

Kim, T.H., Yum, J., Hu, S.J., Spicer, J.P. and Abell, J.A. (2011) Process Robustness of Single Lap Ultrasonic Welding of Thin, Dissimilar Materials. *CIRP Annals - Manufacturing Technology*, 60, 17-20.

Lee, S.S., Kim, T.H., Hu, S.J., Cai, W.W. and Abell, J.A. (2010) Joining Technologies for Automotive Lithium-ion Battery Manufacturing – A Review. *Proceedings of ASME 2010 International Manufacturing Science and Engineering Conference.*

Lee, S.S., Shao, C., Kim, T.H., Hu, S.J., Kannatey-Asibu, E., Cai, W.W., Spicer, J.P. and Abell, J.A. (2014) Characterization of Ultrasonic Metal Welding by Correlating Online Sensor Signals With Weld Attributes. *Journal of Manufacturing Science and Engineering*, 136, 051019.

Mitchell, A.F.S. and Krzanowski, W.J. (1985) The Mahalanobis Distance and Elliptic Distributions. *Biometrika*, 72, 464-467.

Montgomery, D.C. (2013) Introduction to Statistical Quality Control. Wiley, Hoboken, NJ.

Shao, C., Paynabar, K., Kim, T.H., Jin, J., Hu, S.J., Spicer, J.P., Wang, H. and Abell, J.A. (2013) Feature Selection For Manufacturing Process Monitoring Using Cross-Validation. *Journal of Manufacturing Systems*, 32, 550-555.

Shao, C., Guo, W., Kim, T.H., Jin, J., Hu, S.J., Spicer, J.P., and Abell, J.A. (2014) Characterization and Monitoring of Tool Wear in Ultrasonic Metal Welding. *Proceedings of the 9th International Workshop on Microfactories (IWMF 2014)*, 161-169.

Stoumbos, Z.G. and Sullivan, J.H. (2002) Robustness to Non-Normality of the Multivariate EWMA Control Chart. *Journal of Quality Technology*, 34, 260-276.

Mahalanobis, P.C. (1936) On the Generalised Distance in Statistics. *Proceedings of National Institute of Science*, India, 49-55.

P.J. Rousseeuw, P.J. (1984) Least Median of Squares Regression. *Journal of the American Statistical Association*, 79, 871-880.

Willemain T.R. and Runger, G.C. (1996) Designing Control Charts Using an Empirical Reference Distribution. *Journal of Quality Technology*, 28, 31.

Xiang, S., Nie, F. and Zhang, C. (2008) Learning a Mahalanobis Distance Metric for Data Clustering and Classification. *Pattern Recognition*, 41, 3600-3612.

CHAPTER 3

WITHIN-CYCLE PROFILE MONITORING FOR REAL-TIME DEFECT PREVENTION

3.1 Introduction

The increasing demand in high product quality and reliability has placed strict requirements in process monitoring and control for defect prevention. For example, in lithium-ion battery manufacturing, a single defect in battery joining may lead to undesirable performance of the entire battery pack (Li *et al.*, 2010); in steel rolling, a small surface defect could cause catastrophic failure when the rolled product is in use (Jin *et al.*, 2008). The key to product quality improvement lies in reducing variability in production. Statistical Process Control (SPC) has been applied to monitoring manufacturing process and reducing variability through post-quality analysis and elimination of assignable causes. Such analyses are conducted after parts are completed and therefore not capable for defect prevention if the part is found to be defective. This motivates our research to shift the detection of abnormal process conditions from post-manufacturing to real-time decisions during manufacturing.

For discrete manufacturing, the sensor measurements provided by online sensing and data capture technology are time-dependent functional data, also called profile data or waveform signals (Woodall, 2007; Woodall *et al.*, 2004). In this chapter, we are particularly interested in cycle-based signals, which are collected from repetitive operational cycles of a manufacturing process. Figure 3.1 illustrates the cycle-based power signals collected from the ultrasonic metal welding process for joining lithium-ion batteries for electric vehicles. A profile signal is

recorded as a joint is created, with the length of the signal cycle corresponding to the pace of production. When there are assignable causes in the manufacturing process, the cycle profile significantly deviates from the "in-control" profile shape, which may indicate the occurrence of a potential defect. Three types of profile signals are shown in Figure 3.1: (i) profiles from multiple in-control operations collected at different cycles of production, (ii) one profile signal collected from an out-of-control operation when the sheet metal interface is contaminated with oil, and (iii) one profile signal collected from the same out-of-control operation as (ii) but the clamping pressure is adjusted in real time at τ^*. In (ii) and (iii), the abnormal surface conditions of the workpiece cause the profiles to change as the welding power does not ramp up at the early stage as a normal weld does, thus resulting in a poor quality connection in (ii) unless the clamping pressure is adjusted in time (iii). Hence, the conventional between-cycle monitoring strategy, which makes the detection for each finished part based on the entire cycle of signal, gives a time-delayed monitoring decision that is too late for defect prevention.

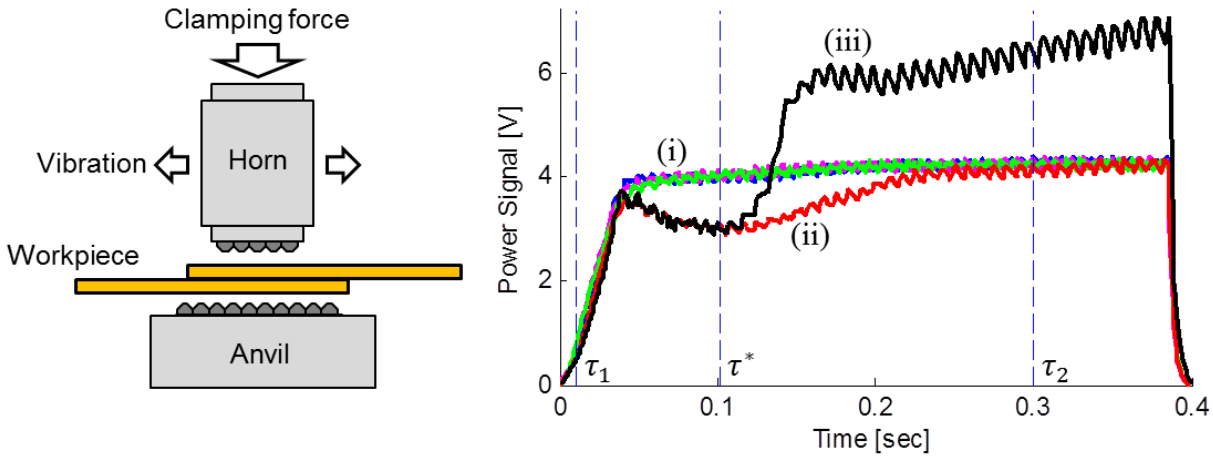


Figure 3.1. Ultrasonic welding and cycle-based power signal

With strict quality requirement in these mission-critical products, process changes need to be detected before the part is finished so that corrections to the process could be made to prevent defects from occurring. In our battery joining example, weld defect prevention is critical since a low quality connection will potentially impair the performance of the battery pack in electric vehicle. It is possible, however, to adjust the clamping pressure in real time by installing external pneumatic pressure regulators to the existing ultrasonic welding machine. In the scenario of cold welds, the adjusted pressure would prevent defective joints and thus compensating the negative effects brought by the out-of-control operation. Profile signal (iii) in Figure 3.1 is the result of such an attempt: as the abnormal shape of the profile is detected at time τ^*, clamping pressure is adjusted immediately; by the time the weld is finished, potential defects are prevented. We call the monitoring strategy in this framework "within-cycle monitoring" as the monitoring decision is made within an operation cycle rather than at the end of the cycle. The practice of engineering process control is not in the scope of this chapter. As opposed to the between-cycle monitoring approach mentioned before, within-cycle monitoring gives real-time decision which can be used to guide real-time process adjustment and enable defect prevention. Although there have been many research focusing on developing between-cycle monitoring strategies, research on within-cycle monitoring is very limited. In this chapter, we aim to develop a new monitoring method considering within-cycle control opportunity for defect prevention. Particularly, we aim to make the monitoring decision based on an early portion of the cyclic signal and to use the monitoring results to guide real-time process adjustment and defect prevention.

One of the biggest research challenges in within-cycle monitoring is determining the decision point. On the one hand, reliable detection and accurate monitoring decision require a latter decision point so that a longer portion of the profile data is included in the monitoring decision. In our ultrasonic welding example in Figure 3.1, monitoring decisions made at τ_1

(0.01sec) will not be accurate due to the limited length of signal. On the other hand, early detection is desirable in order to leave sufficient time for process adjustment and control actions. In our example, monitoring decisions made at τ_2 (0.30sec) may be too late for taking corrective actions. Therefore, the optimal decision point, τ^*, should be able to balance the tradeoff between monitoring accuracy and the length of the remaining time available for control actions. Therefore, a new monitoring strategy with optimal decision point considering both monitoring decision and control opportunity will be developed in this chapter. This issue is not a concern in traditional control chart design and profile monitoring. Although Chang *et al.* (2014) proposed an SPC framework to detect potential changes of a wave profile using partially generated profile, the issue on determining the decision point was not mentioned and their method could not be applied to complex profiles other than the wave profile.

Within-cycle monitoring of complex profile signals also raises challenges in modeling both between-profile and within-profile variations, detecting both profile mean shift and variance change, and effectively characterizing complex profile data. Most of the existing research on profile monitoring has been for the case in which the profile can be represented by a parametric model, from simple linear regression models (Zou *et al.*, 2006; Mahmoud *et al.*, 2007; Noorossana *et al.*, 2004; Chang and Gan, 2006; Gupta *et al.*, 2006) to more complicated methods like multiple and polynomial regression models (Zou *et al.*, 2007; Kazemzadeh *et al.*, 2008; Kazemzadeh *et al.*, 2009; Mahmoud, 2008), and nonlinear regression models (Ding *et al.*, 2006; Williams *et al.*, 2007; Williams *et al.*, 2007). Applying a parametric model, however, is not always achievable because it requires strong domain knowledge and major modeling efforts to identify an appropriate parametric model structure. To overcome this challenge, an alternative approach that uses nonparametric models has attracted increasing attention. The wavelet transform is a nonparametric alternative that can be effectively used for modeling nonlinear profiles with sharp jumps (Zhou *et al.*, 2006;

Jeong *et al.*, 2006; Chicken *et al.*, 2009). Using a wavelet-based method to monitor nonlinear profiles and perform statistical process control on complicated profile data has generated increasing interest in recent years. One limitation of the wavelet-based profile monitoring methods is that the between-profile variation is ignored since they assumed that the total variability of profiles can be modeled by within-profile variations as random noises, which are typically assumed to be normally independently distributed. In order to consider both within-profile and between-profile variations, Paynabar and Jin (2011) extended the wavelet-based change-point model in Chicken *et al.* (2009) by incorporating a mixed-effect model to characterize nonlinear profile variations. However, all of these wavelet-based monitoring methods are limited to detecting process mean shift while ignoring variance change. For within-cycle monitoring, detecting variance change is no less important than detecting the mean shift, especially when the abnormal profile shape is exhibited in an individual profile instead of the overall mean profile. For example, if the sample with profile signal (ii) in Figure 3.1 is the only sample in this batch with such abnormal surface conditions, then this profile change would not be characterized as a mean shift but rather a variance change. In this chapter, we adopt the wavelet transformation and the mixed-effect model to characterize complex profile data and capture both between-profile and within-profile variations. In order to effectively detect both profile mean shift and variance change, we will further combine the wavelet-based mixed-effect model with control chart design on the monitoring of profile deviations.

The remainder of this chapter is organized as follows. Section 3.2 provides an overview of the proposed methodology and formulates an integrated criterion for evaluating the performance of within-cycle monitoring considering control opportunity. Section 3.3 presents the detailed solution methodology in finding the optimal decision point, characterizing profiles, and developing control charts. Simulations and sensitivity analyses are given in

Section 3.4, followed by a case study in Section 3.5. Finally, Section 3.6 concludes the chapter.

3.2 Problem Formulation

The proposed within-cycle profile monitoring method is depicted in the flowchart in Figure 3.2. The methodology development consists of three stages: I. Collect training samples, II. Determine the optimal decision point τ^*, and III. Design control chart for online monitoring. In subsequent sections, we will elaborate the steps in Stage II in detail.

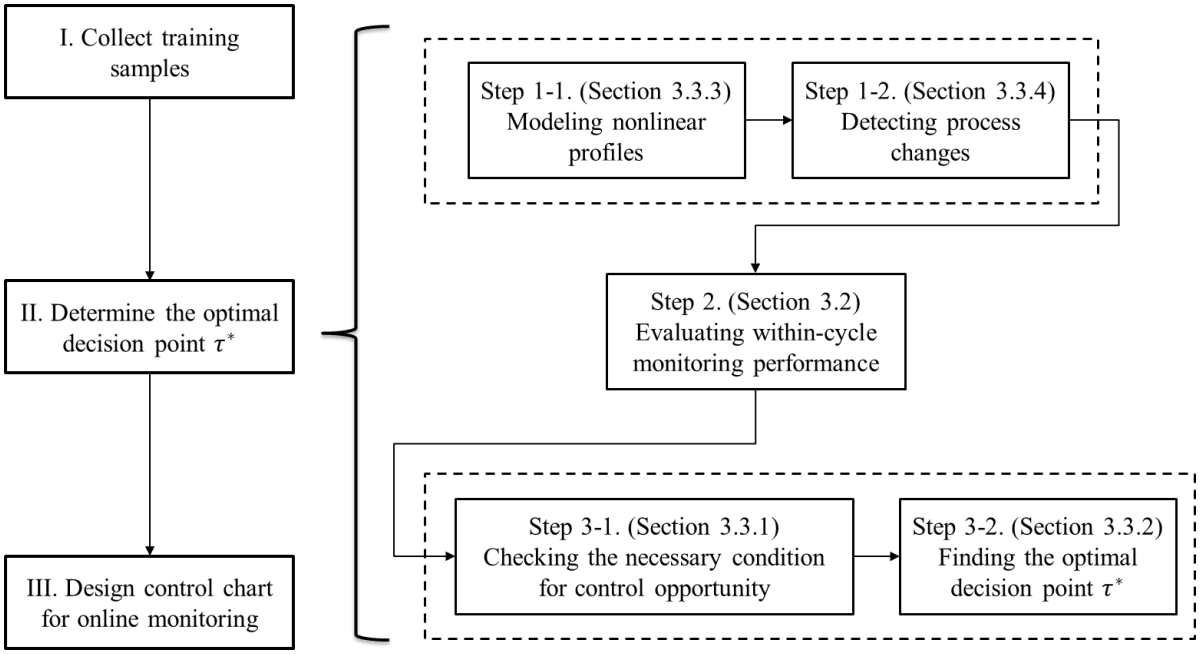


Figure 3.2. Overview of the proposed within-cycle profile monitoring methodology

The three major steps in determining the optimal decision point for online monitoring are:

Step 1-1. The measured nonlinear profile data are modeled using wavelets and mixed-effect models in which process changes in both the mean and variance can be characterized in order to characterize both within-profile and between-profile variations.

Step 1-2. Control charts are developed to monitor the amount of profile deviation from a reference profile in which the magnitude of profile deviation is modeled and monitored.

Step 2. A criterion to evaluate the within-cycle monitoring performance to consider both monitoring accuracy and control opportunity is proposed.

Step 3. Considering the trade-off between reliable detection and sufficient control opportunity, the optimal decision point τ^* is to be found through an optimization problem.

It should be noted that in practice, not all abnormal conditions can be effectively corrected with one-step control even if a deviation is detected within the profile cycle. If the profile only changes in a few data points, e.g., a spike or a local deviation, or if the profile change only occurs at the very latter portion of the signal, the monitoring technique may not be able to fully capture this scenario. This kind of shifts also poses more challenges in real-time process adjustment. In this chapter, we focus on situations where the process can be adjusted with one-step control based on within-cycle monitoring decisions. Within-cycle monitoring is most effective for situations when the abnormal condition is reflected by the entire profile change or changes in at least some early segments of the profile.

Under the situations where the process can be adjusted based on within-cycle monitoring decisions, still not all control actions will be effective due to the time constraint. We define a function $p_e(\tau)$ to describe the probability that within-cycle control actions will be effective if a correct detection is made at time τ, $0 \le p_e(\tau) \le 1$ for $\forall \tau \in (0, T]$, where T is the manufacturing process cycle. Considering the time effect of control actions, $p_e(\tau)$ is assumed to be a monotonically non-increasing function of τ. The exact expression of $p_e(\tau)$ can be obtained empirically through experiments, or, it can be theoretically derived or

reasonably assumed based on the specific machine and process. In our ultrasonic welding example, when abnormal surface conditions of the workpiece cause the welding power does not ramp up as a normal weld does, we can increase the clamping pressure with the use of external pressure regulators. But if this control action is triggered too late, there may not be sufficient time for the welding power to react. On the other hand if a control action is falsely triggered, clamping pressure will be added to the normal welding process and thus leading to a high risk of having an over weld. In view of such situation, we assume that all falsely triggered control actions will drive the process out-of-control, regardless of the decision point. Table 3.1 describes the confusion matrix of the within-cycle monitoring performance considering control opportunity.

Table 3.1. Confusion matrix of the monitoring performance considering control opportunity

		Monitoring decision and control action at τ	
		In-control	Out-of-control
True process	In-control	$1-\alpha(\tau)$ No control action triggered	$\alpha(\tau)$ Falsely triggered control
	Out-of-control	$\beta(\tau)$ No control action triggered	$1-\beta(\tau)$ Correctly triggered control; Control action is effective with a probability of $p_e(\tau)$

The performance of a process control method is evaluated by Type I and Type II errors Montgomery (2013). In order to reflect how within-cycle monitoring performance is affected by the decision point τ, denote $\alpha(\tau)$ and $\beta(\tau)$ $(0 \leq \alpha(\tau), \beta(\tau) \leq 1)$ as the Type I and Type II error rates, respectively, given by a monitoring technique when the monitoring decision is made at time τ. Now that a monitoring decision is made at time τ $(0 < \tau \leq T)$,

$\alpha(\tau)$ and $\beta(\tau)$ vary with τ as only the signal segment in $(0, \tau]$ has been generated and considered at this time.

Now we denote D_{nc} as the proportion of out-of-control profiles based on history data if no within-cycle monitoring and control actions are implemented. Denote D_{wc} to be the proportion of out-of-control profiles after the proposed within-cycle monitoring technique is adopted and real-time control actions are implemented. It can be seen from Table 3.1 that D_{wc} differs from D_{nc} in two aspects: (i) when an in-control process is falsely detected as out-of-control with a probability of $\alpha(\tau)$, unnecessary control actions will be triggered, resulting in a risk of damaging those good quality parts; (ii) when an out-of-control process is correctly detected with $1 - \beta(\tau)$, correct control actions will be taken immediately, followed by effectively correcting bad parts into good parts with a probability of $p_e(\tau)$. Therefore, D_{wc} is derived as

$$D_{wc} = D_{nc} + (1 - D_{nc})\alpha(\tau) - D_{nc}\big(1 - \beta(\tau)\big)p_e(\tau). \tag{3.1}$$

where $(1 - D_{nc})\alpha(\tau)$ is the loss due to falsely triggered control actions and $D_{nc}\big(1 - \beta(\tau)\big)p_e(\tau)$ is the gain from effective correctly triggered control actions.

3.3 Solution Methodology

3.3.1 Necessary condition for control opportunity

The proposed within-cycle process monitoring technique is effective if and only if there is an improvement from D_{nc} to D_{wc}. Thus, the necessary condition for control opportunity is given by $D_{wc} < D_{nc}$, which is simplified as

$$L(\tau) \equiv \frac{\alpha(\tau)}{1 - \beta(\tau)} < \frac{p_e(\tau)}{1/D_{nc} - 1} \equiv H(\tau). \tag{3.2}$$

where the left-hand side ratio, $L(\tau)$, is determined by the accuracy of monitoring decisions; the right-hand side ratio, $H(\tau)$, is affected by the control effectiveness function, $p_e(\tau)$, and the process status without within-cycle monitoring and control, D_{nc}. Only when Eq. (3.2) is satisfied, there are opportunities for real-time control and within-cycle monitoring to benefit.

A closer examination of Eq. (3.2) indicates that the necessary condition for control opportunity can be further expressed by two conditions that (i) $\alpha(\tau)$ should be less than $D_{nc}p_e(\tau)$ and (ii) the detection power $1-\beta(\tau)$ should be no less than $\frac{\alpha(\tau)}{D_{nc}p_e(\tau)}$. It is then noticed that D_{nc} plays an important role in this necessary condition. If the process is mostly in-control with a smaller value of D_{nc}, it poses a higher requirement on false alarm rate $\alpha(\tau)$ and detection power $1-\beta(\tau)$, and the within-cycle monitoring and control will be effective only if the monitoring decision is accurate. To have a more accurate monitoring decision, a larger value of τ will be needed and thus resulting in a smaller control opportunity. But if the process has a high probability of falling out-of-control, as indicated by a larger value of D_{nc}, within-cycle monitoring will be effective even when the monitoring decision is less accurate, giving more opportunities for real-time control.

3.3.2 Optimal decision point

In order to determine the optimal decision point for within-cycle process monitoring and control, an optimization problem can be formulated to find the decision point, τ, that:

$$\underset{\tau}{\text{minimize}}\ D_{wc}$$

$$\text{subject to}\ D_{wc} < D_{nc}$$

where D_{wc} is defined in Eq. (3.1).

We assume D_{nc} is known a priori since it is estimated from history production quality data without real-time adjustment. Type I error rate $\alpha(\tau)$ is pre-determined as a design

parameter when constructing control charts. Therefore, the optimization problem to find the decision point is further simplified as

$$\underset{\tau}{\text{maximize}}\ (1-\beta(\tau))p_e(\tau)$$

$$\text{subject to}\ (1-\beta(\tau))p_e(\tau) > \left(\frac{1}{D_{nc}} - 1\right)\alpha(\tau). \tag{3.3}$$

where $0 \le 1-\beta(\tau) \le 1$ and $0 \le p_e(\tau) \le 1$. The upper bound of the objective function is 1, which is achieved when $\beta(\tau) = 0$ and $p_e(\tau) = 1$. In most scenarios, the detection power $1-\beta(\tau)$ is a monotonically non-decreasing function of τ, while the control effectiveness function $p_e(\tau)$ is a monotonically non-increasing function of τ.

Denote the objective function as $g(\tau) = (1-\beta(\tau))p_e(\tau)$. The optimal decision point is $\tau^* = argmax_\tau\, g(\tau)$. In order to find τ^*, we need to solve for $g'(\tau) = 0$ and further prove $g''(\tau^*) < 0$. Setting $g'(\tau) = 0$ gives

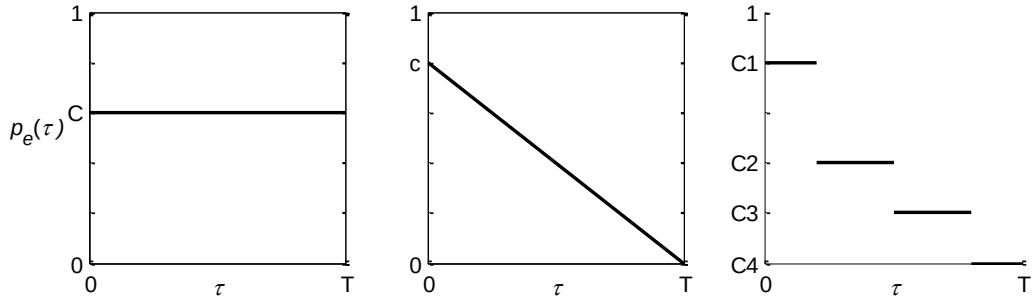
$$\frac{-\beta'(\tau)}{1-\beta(\tau)} = \frac{-p_e'(\tau)}{p_e(\tau)}. \tag{3.4}$$

The left-hand side of Eq. (3.4) is determined by the detection power $1-\beta(\tau)$ and its derivative w.r.t. τ; the right-hand side is determined by the control effectiveness and its derivative. Given $\beta(\tau)$ and $p_e(\tau)$, the optimal decision point τ^* can be found using Eq. (3.4). Table 3.2 provides several simplified expressions of Eq. (3.4) in light of some typical scenarios of $p_e(\tau)$. Figure 3.3 illustrates those $p_e(\tau)$ patterns.

It is noticed that the step function $p_e(\tau)$ is not differentiable at points $T_1, T_2, \ldots$. A feasible alternative to find τ^*would be first finding the optimal decision point among the points that are differentiable, then compare with $\tau = T_1, T_2, \ldots$. Despite non-differentiability, the step function is brought up in Table 3.2 since it is the generalization of continuous functions, linear or nonlinear.

Table 3.2. Simplified expressions of Eq. (3.4) under typical control effective functions

	$\boldsymbol{p_e(\tau)}$ **pattern over** $\boldsymbol{(0, T)}$	**Expression of Eq. (3.4)**
(a)	Constant $p_e(\tau) = C$	$\dfrac{-\beta'(\tau)}{1-\beta(\tau)} = 0$
(b)	Linear $p_e(\tau) = c(1 - \frac{\tau}{T})$	$\dfrac{-\beta'(\tau)}{1-\beta(\tau)} = \dfrac{1}{T-\tau}$
(c)	Step (piecewise constant) $p_e(\tau) = \begin{cases} C_1 & \text{if } \tau \in (0, T_1] \\ C_2 & \text{if } \tau \in (T_1+1, T_2] \\ \vdots & \vdots \end{cases}$	$\begin{cases} \dfrac{-\beta'(\tau)}{1-\beta(\tau)} = 0 & \text{if } \tau \neq T_1, T_2, \\ p_e'(\tau) \text{ does not exist} & \text{if } \tau = T_1, T_2, \end{cases}$



(a) Constant (b) Linear (c) Step

Figure 3.3. Typical control effective functions described in Table 3.2

3.3.3 Basis of mixed-effect profile modeling using wavelets

Suppose a profile f consists of n pairs of points $(t_j, y_j), j = 1,2, \dots, n$. Generally, the relationship between $\boldsymbol{y}$ and $\boldsymbol{t}$ can be described as $\boldsymbol{y} = f(\boldsymbol{t}) + \varepsilon$, where $f(.)$ is an unknown true nonlinear function of the profile f, ε is a random error with mean zero and standard deviation σ. Performing the wavelet transformation on the observe data $\boldsymbol{y} = \{y_j\}$ with a dyadic length of $n = 2^J$, the matrix form of Discrete Wavelet Transform (DWT) is represented as

$$\boldsymbol{z} = \boldsymbol{W}\boldsymbol{y} = \boldsymbol{W}(\boldsymbol{f} + \varepsilon) = \boldsymbol{\theta} + \boldsymbol{W}\varepsilon \tag{3.5}$$

where $\boldsymbol{W}$ is the $n \times n$ orthonormal wavelet transformation matrix and $\boldsymbol{\theta}$ is the transformed wavelet coefficients of the true sampled function $\boldsymbol{f}$. Since $\boldsymbol{W}$ is an orthogonal transform, $\boldsymbol{z}$ is normal with mean $\boldsymbol{\theta}$ and variance $\sigma^2 \cdot \boldsymbol{I}_{n\times n}$, where $\boldsymbol{I}_{n\times n}$ is an $n \times n$ identity matrix. The vector $\boldsymbol{z}$ represents all decomposed wavelet coefficients. More details about the wavelet transformation can be found in Daubechies (1992) and Mallat (1999).

In this chapter, an orthogonal Haar transform is used for the discretized profile data $\boldsymbol{y}_i = f_i(\boldsymbol{t}) + \boldsymbol{\varepsilon_i}$, where $\boldsymbol{y}_i$ is a vector of the discrete response measurements of profile i, $\boldsymbol{t}$ is a vector consisting of equally spaced sampling time or distance data, and $\boldsymbol{\varepsilon_i}$ is a vector of NID noises with $\boldsymbol{\varepsilon_i} \sim MVN(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ to represent the within-profile variation. The resulting wavelet coefficients of $\boldsymbol{y}_i$ are represented as $\boldsymbol{z}_i = \boldsymbol{\theta}_i + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}$, where $\boldsymbol{\theta}_i = \boldsymbol{W}\boldsymbol{y}_i$ is a vector of the true wavelet coefficients transformed from the true profile function $f_i(\boldsymbol{t})$, $\boldsymbol{z}_i = \boldsymbol{W}\boldsymbol{y}_i$ is a vector of the empirical wavelet coefficients transformed from noisy profile $\boldsymbol{y}_i$, and $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} = \boldsymbol{W}\boldsymbol{\varepsilon_i}$ is a random noise vector in the wavelet domain with $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} \sim MVN(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

To consider the between-profile variation, a mixed model, in which a few wavelet coefficients are selected to act as random effects, is utilized. Davidian and Giltinan (1995) and Demidenko (2004) have provided a comprehensive introduction to mixed models. In this chapter, we adopt the mixed-effect model presented by Paynabar and Jin (2011). To implement the mixed model based on wavelet coefficients, let $\boldsymbol{\theta}_i = \boldsymbol{\mu} + \boldsymbol{b}_i$, where $\boldsymbol{\mu}$ is the vector of fixed effects common to all profiles, $\boldsymbol{b}_i$ is the vector of random effects of profile i with $\boldsymbol{b}_i \sim MVN(\boldsymbol{0}, \boldsymbol{\Lambda})$, and $\boldsymbol{\Lambda}$ is a positive-definite matrix that represents the covariance structure of the random effects. $\boldsymbol{\Lambda}$ is assumed to be a diagonal matrix, which implies that the random effects are uncorrelated. We also assume that in the equation $\boldsymbol{z}_i = \boldsymbol{\mu} + \boldsymbol{b}_i + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}$, $\boldsymbol{b}_i$ is independent of $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}$.

Therefore, the wavelet coefficients of $\boldsymbol{y}_i$ are represented as $\boldsymbol{z}_i = \boldsymbol{\mu} + \boldsymbol{b}_i + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}$, $\boldsymbol{z}_i \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Lambda} + \sigma^2 \boldsymbol{I})$. Based on this mixed model, the parameters of $\boldsymbol{\mu}$ and $\boldsymbol{b}_i$ can be effectively used to represent the profile mean and between-profile variation, respectively.

3.3.4 Control chart design and detection performance

Let f_i denote the newly observed profile from sample i and f_0 denote a pre-defined in-control reference profile. We can then determine if f_i has changed from f_0 by comparing the corresponding wavelet vector $\boldsymbol{z}_i$ to $\boldsymbol{z}_0$, where $\boldsymbol{z}_0$ is the wavelet vector corresponding to the reference profile f_0:

$$\|f_i - f_0\|_{L^2}^2 = \int_0^T \left(f_i(t) - f_0(t)\right)^2 dt = \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0\|_{l^2}^2 \tag{3.6}$$

where $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0\|_{l^2}^2$ is estimated based on $\|\boldsymbol{z}_i - \boldsymbol{z}_0\|_{l^2}^2$. In most cases, the in-control reference profile is unknown and must be estimated from a training set of N in-control profile observations (Chicken *et al.*, 2009).

As mentioned in the previous subsection, $\boldsymbol{z}_i = \boldsymbol{\mu} + \boldsymbol{b}_i + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}$ and $\boldsymbol{z}_i \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Lambda} + \sigma^2 \boldsymbol{I})$. Denote $\boldsymbol{\mu}_0$ and $\boldsymbol{\Lambda}_0$ to be the parameters for the in-control process, while $\boldsymbol{\mu}_1$ and $\boldsymbol{\Lambda}_1$ are the parameters for the out-of-control process. Assume the standard deviation of NID noises, σ, does not change. Therefore, the wavelet vector of the reference profile has $\boldsymbol{z}_0 \sim MVN(\boldsymbol{\mu}_0, (\boldsymbol{\Lambda}_0 + \sigma^2 \boldsymbol{I})/N)$. The distribution of the wavelet vector of the incoming profile $\boldsymbol{y}_i$ has $\boldsymbol{z}_i | H_0 \sim MVN(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0 + \sigma^2 \boldsymbol{I})$ if $\boldsymbol{y}_i$ is in-control and $\boldsymbol{z}_i | H_1 \sim MVN(\boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1 + \sigma^2 \boldsymbol{I})$ if $\boldsymbol{y}_i$ is out-of-control. Therefore, the deviation of the wavelet vector of an incoming profile $\boldsymbol{y}_i$ from the reference profile vector $\boldsymbol{z}_0$ is represented as

$$\boldsymbol{z}_i - \boldsymbol{z}_0 \sim \begin{cases} MVN\left(\boldsymbol{0}, \dfrac{N+1}{N}(\boldsymbol{\Lambda}_0 + \sigma^2 \boldsymbol{I})\right), & H_0: \boldsymbol{y}_i \text{ is in-control} \\ MVN\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0, \dfrac{\boldsymbol{\Lambda}_0 + \sigma^2 \boldsymbol{I}}{N} + \boldsymbol{\Lambda}_1 + \sigma^2 \boldsymbol{I}\right), & H_1: \boldsymbol{y}_i \text{ is out-of-control} \end{cases}. \tag{3.7}$$

Define $\delta_i^2 = \|\boldsymbol{z}_i - \boldsymbol{z}_0\|_{l^2}^2$. Notice that both within-profile variation and between-profile variation are reflected in δ_i^2. Hence, δ_i^2's can be used to detect both profile mean shift and variance change.

When a within-cycle monitoring decision is made at time τ, $\tau \in (0, T]$, only the segment of the profile data in $(0, \tau]$ is observed and available for analysis. Denote $\boldsymbol{y}_{i,\tau}$ as the discrete response measurements of profile i in $(0, \tau]$ and its wavelet coefficients are represented in vector $\boldsymbol{z}_{i,\tau}$. Similar to the derivation of δ_i^2 above, denote $\delta_{i,\tau}^2 = \left\|\boldsymbol{z}_{i,\tau} - \boldsymbol{z}_{0,\tau}\right\|_{l^2}^2$, where $\boldsymbol{z}_{0,\tau}$ is the wavelet coefficients vector of the pre-known reference profile segment till time τ. At τ, a control chart is constructed for process monitoring and its detection power is evaluated. In Figure 3.4, we illustrate the development of control chart and how the results from Phase I control chart are utilized to find the optimal decision point.

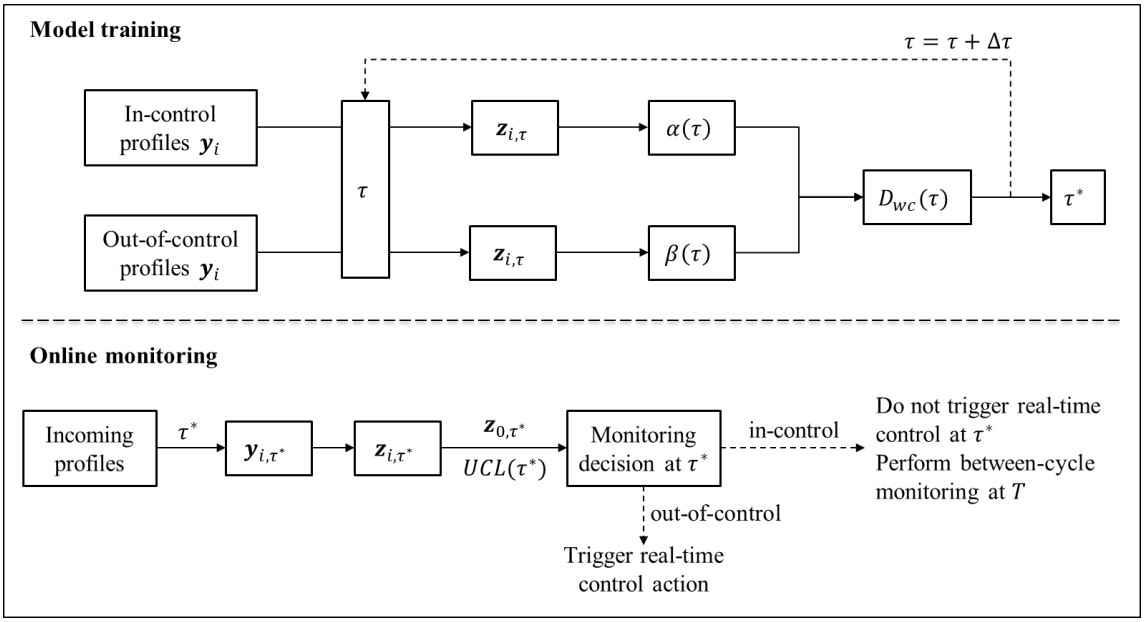


Figure 3.4. Flowchart of control chart development

3.3.4.1 General case

Without putting any other assumptions on the covariance structure of the random effects, $\boldsymbol{\Lambda}$, than the diagonality assumption described in Section 3.3.3, the monitoring statistic of the control chart at τ is

$$\delta_{i,\tau}^2 = \left\| \boldsymbol{z}_{i,\tau} - \boldsymbol{z}_{0,\tau} \right\|_{l^2}^2 = \sum_{j=1}^{n_\tau} \left(z_{i,\tau,j} - z_{0,\tau,j} \right)^2 \tag{3.8}$$

where $z_{i,\tau,j}$ is the jth entry in $\boldsymbol{z}_{i,\tau}$, $z_{0,\tau,j}$ is the jth entry in $\boldsymbol{z}_{0,\tau}$, and n_τ is the number of observed data points till τ.

Since $\boldsymbol{\Lambda}_0$ and $\boldsymbol{\Lambda}_1$ are diagonal matrices with possibly different diagonal elements, the monitoring statistic under H_0, $\delta_{i,\tau}^2 | H_0$, follows a generalized χ^2 distribution whose closed form cannot be expressed. The upper control limit of this control chart, UCL, is set based on the percentile of $\delta_{i,\tau}^2 | H_0$ values obtained from the training dataset.

The detection power at decision point τ is

$$1 - \beta(\tau) = 1 - \Pr\left(\delta_{i,\tau}^2 < UCL(\tau) | H_1 \right) \tag{3.9}$$

where the CDF of the generalized χ^2 distribution can be estimated with the help of computing software.

3.3.4.2 Special case

Now we present the design of the control chart for a special case when the distribution of the monitoring statistic, $\delta_{i,\tau}^2$, can be explicitly presented. Besides the diagonality assumption of $\boldsymbol{\Lambda}$, it is also assumed that $\boldsymbol{\Lambda}_0 = \lambda_0^2 \boldsymbol{I}$ and $\boldsymbol{\Lambda}_1 = \lambda_1^2 \boldsymbol{I}$, indicating that the diagonal elements in the covariance matrix take on the same value λ_0^2 and λ_1^2, respectively. This can be

interpreted by the special scenario when the between-profile variability is equally distributed along the profile.

With $\boldsymbol{\Lambda}_0 = \lambda_0^2 \boldsymbol{I}$ and $\boldsymbol{\Lambda}_1 = \lambda_1^2 \boldsymbol{I}$, we can derive

$$z_{i,\tau,j} - z_{0,\tau,j} \sim \begin{cases} N\left(0, \frac{N+1}{N}(\lambda_0^2 + \sigma^2)\right), & H_0: \boldsymbol{y}_i \text{ is in-control} \\ N\left(\mu_{i,\tau,j} - \mu_{0,\tau,j}, \frac{\lambda_0^2 + \sigma^2}{N} + \lambda_1^2 + \sigma^2\right), & H_1: \boldsymbol{y}_i \text{ is out-of-control for } j = 1,2,\dots,n. \end{cases} \quad (3.10)$$

Define a statistic $Q_{0,\tau}$ as

$$Q_{0,\tau} = \frac{\delta_{i,\tau}^2}{\frac{N+1}{N}(\lambda_0^2 + \sigma^2)} = \sum_{j=1}^{n_\tau} \left(\frac{z_{i,\tau,j} - z_{0,\tau,j}}{\sqrt{\frac{N+1}{N}(\lambda_0^2 + \sigma^2)}} \right)^2 \quad (3.11)$$

Since $\frac{z_{i,\tau,j} - z_{0,\tau,j}}{\sqrt{\frac{N+1}{N}(\lambda_0^2+\sigma^2)}}$ follows the standard normal distribution, $Q_{0,\tau}$ follows a χ^2 distribution with degree of freedom n_τ, denoted as $Q_{0,\tau} \sim \chi_{n_\tau}^2$ distribution. Therefore, the monitoring statistic under H_0, $\delta_{i,\tau}^2 | H_0$, can be considered as a transformation from the $\chi_{n_\tau}^2$ distributed random variable $Q_{0,\tau}$.

Set the Type I error rate to be $\alpha(\tau) = \alpha$ for $\forall \tau \in (0, T]$. Since the explicit expression of the $\chi_{n_\tau}^2$ distribution is known, the critical value for $Q_{0,\tau}$ is denoted as χ_{α,n_τ}^2. Therefore, the upper control limit for control chart at τ is

$$UCL(\tau) = \frac{N+1}{N}(\lambda_0^2 + \sigma^2)\chi_{\alpha,n_\tau}^2 \quad (3.12)$$

Since the value of χ_{α,n_τ}^2 is affected by n_τ, the number of observed data points at τ, the UCL is also a function of τ.

When $\boldsymbol{y}_i$ is out-of-control, its wavelet vector $\boldsymbol{z}_i | H_1 \sim MVN(\boldsymbol{\mu}_1, \boldsymbol{\Lambda}_1 + \sigma^2 \boldsymbol{I})$. Define a statistic $Q_{1,\tau}$ as

$$Q_{1,\tau} = \frac{\delta_{i,\tau}^2}{\frac{1}{N}(\lambda_0^2 + \sigma^2) + (\lambda_1^2 + \sigma^2)} = \sum_{j=1}^{n_\tau} \left(\frac{z_{i,\tau,j} - z_{0,\tau,j}}{\sqrt{\frac{1}{N}(\lambda_0^2 + \sigma^2) + (\lambda_1^2 + \sigma^2)}} \right)^2 \quad (3.13)$$

$Q_{1,\tau}$ follows a non-central χ^2 distribution with degree of freedom n_τ and a non-centrality parameter

$$\omega_\tau = \sum_{j=1}^{n_\tau} \left(\frac{\mu_{i,\tau,j} - \mu_{0,\tau,j}}{\sqrt{\frac{1}{N}(\lambda_0^2 + \sigma^2) + (\lambda_1^2 + \sigma^2)}} \right)^2 = \frac{||\boldsymbol{\mu}_{1,n_\tau} - \boldsymbol{\mu}_{0,n_\tau}||_{l^2}^2}{\frac{1}{N}(\lambda_0^2 + \sigma^2) + (\lambda_1^2 + \sigma^2)} \quad (3.14)$$

where $\boldsymbol{\mu}_{1,n_\tau}$ and $\boldsymbol{\mu}_{0,n_\tau}$ are the partially observed out-of-control profile mean and in-control profile mean, respectively. Denote the distribution of $Q_{1,\tau}$ as $Q_{1,\tau} \sim \chi_{n_\tau}^2(\omega_\tau)$ distribution. Based on Eq. (3.13), the detection power of the designed control chart is represented as

$$1 - \beta(\tau) = 1 - \Pr\left(\delta_{i,\tau}^2 < UCL(\tau) \middle| H_1\right) = 1 - \Pr\left(Q_{1,\tau} < \widetilde{UCL}(\tau) | H_1\right) \quad (3.15)$$

where $\widetilde{UCL}(\tau) = \frac{UCL(\tau)}{\frac{1}{N}(\lambda_0^2+\sigma^2)+(\lambda_1^2+\sigma^2)}$. Based on the CDF of the non-central χ^2 distribution, $1 - \beta(\tau)$ can be directly represented in closed form.

In this special case, part-to-part variability is reflected by the change from λ_0^2 to λ_1^2. λ_0^2, λ_1^2, and σ^2 are not affected by the monitoring decision point τ. τ can be equivalently expressed in term of n_τ, the number of observed data points till τ. Assume a total of n data points will be observed from the entire signal cycle at time T, $1 \le n_\tau \le n_T = n$.

As can be seen from Eq. (3.15), the detection power $1 - \beta$ is affected by $\widetilde{UCL}$ and the non-central χ^2 distribution. The non-centrality parameter is determined by $||\boldsymbol{\mu}_{1,n_\tau} - \boldsymbol{\mu}_{0,n_\tau}||_{l^2}^2$ and the variance components, where $||\boldsymbol{\mu}_{1,n_\tau} - \boldsymbol{\mu}_{0,n_\tau}||_{l^2}^2$ represents the overall profile changes till τ. Therefore, the detection power $1 - \beta$ is jointly affected by the designed Type I error level α, the monitoring decision point τ or n_τ, the variance

components $\lambda_0^2 + \sigma^2$ and $\lambda_1^2 + \sigma^2$, and the shape of the overall profile mean shift. Note that the mean shifts are not simple shifts and they cannot be easily described in simple terms of vertical displacement as traditional control charts. Instead, various values are used for $||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0||_{l^2}^2 = a^2$, the L2 norm for the distance between the in- and out-of-control profiles.

3.4 Simulation

This section demonstrates the design of control charts, the evaluation of detection power, and the search for optimal monitoring decision point through simulation studies. Without loss of generality, we focus simulation experiments on the special case.

3.4.1 Design of simulation experiments

We aim to obtain the objective function $g(\tau) = (1 - \beta(\tau))p_e(\tau)$ through a numerical search over the possible values of the decision point τ. In order to obtain $g(\tau)$, we need to design the functions for $p_e(\tau)$ and the mean shift. A closer examination of Eq. (3.4) indicates that the c parameter in linear $p_e(\tau)$ does not affect the search of optimal decision point. Thus, a linear trend of $p_e(\tau)$ is assumed, e.g., $p_e(\tau) = 1 - \frac{\tau}{T}$, as illustrated in Figure 3.3 (b) with $c = 1$.

In simulation, we consider both mean shift and variance change. Two patterns of the profile mean shift are considered: (i) when the mean shift is constant with $\left||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0|\right|_{l^2}^2 = a^2$, we have $\left|\left|\boldsymbol{\mu}_{1,n_\tau} - \boldsymbol{\mu}_{0,n_\tau}\right|\right|_{l^2}^2 = \frac{n_\tau}{n}a^2$, where n is the total number of data points observed from the entire signal cycle; (ii) when the mean shift is a parabolic shape with $\left||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0|\right|_{l^2}^2 = a^2$, we have $\left|\left|\boldsymbol{\mu}_{1,n_\tau} - \boldsymbol{\mu}_{0,n_\tau}\right|\right|_{l^2}^2 = \frac{3a^2n_\tau^2}{n^2} - \frac{2a^2n_\tau^3}{n^3}$. In simulating the variance change, we define the change as $\frac{\lambda_1^2}{\lambda_0^2} = m^2$. Hence, the mean shift if represented by a, while m

explains the variance change. Let $a^2 = \frac{d}{30}n$ and $m = 1 + 0.02d$, where $d = 0,1,2,3,4,5$. The following parameters are also implemented in simulation: $N = 1000, n = 2^{10}$ corresponds to $T = 0.4$sec, $\alpha = 0.01$, and $\lambda_0^2 = \sigma^2 = 1$.

3.4.2 Results and discussion

Figure 3.5 shows the objective function $g(\tau)$ under different patterns of the mean shift, different values of a^2 and m with a linear trend of the $p_e(\tau)$ function. The optimal decision point can be found at the maximum value of the objective function. Figure 3.6 further shows the left-hand and right-hand side values of Eq. (3.4). The optimal decision point can also be found in Figure 3.6 at the intersection of the right-hand side value and the left-hand side value. Figure 3.7 summarizes the optimal decision point found at each scenario.

It can be observed from Figures 3.5~3.7 that with a fixed m value, both $1 - \beta(\tau)$ and $-\beta'(\tau)$ will increase as a increases. On one hand, if $1 - \beta(\tau)$ increases faster than $-\beta'(\tau)$, $T - \tau$ needs to increase in order to achieve $g'(\tau) = 0$. With a large m, $1 - \beta(\tau)$ rises quickly, which puts $-\beta'(\tau)$ to rise slowly. As a result, τ^* decreases as a increases. On the other hand, if $1 - \beta(\tau)$ increases slower than $-\beta'(\tau)$, $T - \tau$ needs to decrease in order to satisfy $g'(\tau) = 0$. With a small m, $1 - \beta(\tau)$ values are small and rises slowly, which puts $-\beta'(\tau)$ to rise faster comparing to $1 - \beta(\tau)$. So, τ^* increases as a increases. The same trend can also be observed when m increases under a fixed a. When a value is in-between, e.g., $a^2 = 68$, τ^* increases as m increases while m is still in a relatively smaller value; when m continues to increase and becomes a large value, $1 - \beta(\tau)$ becomes more dominant due to larger shifts, then τ^* decreases when m increases.

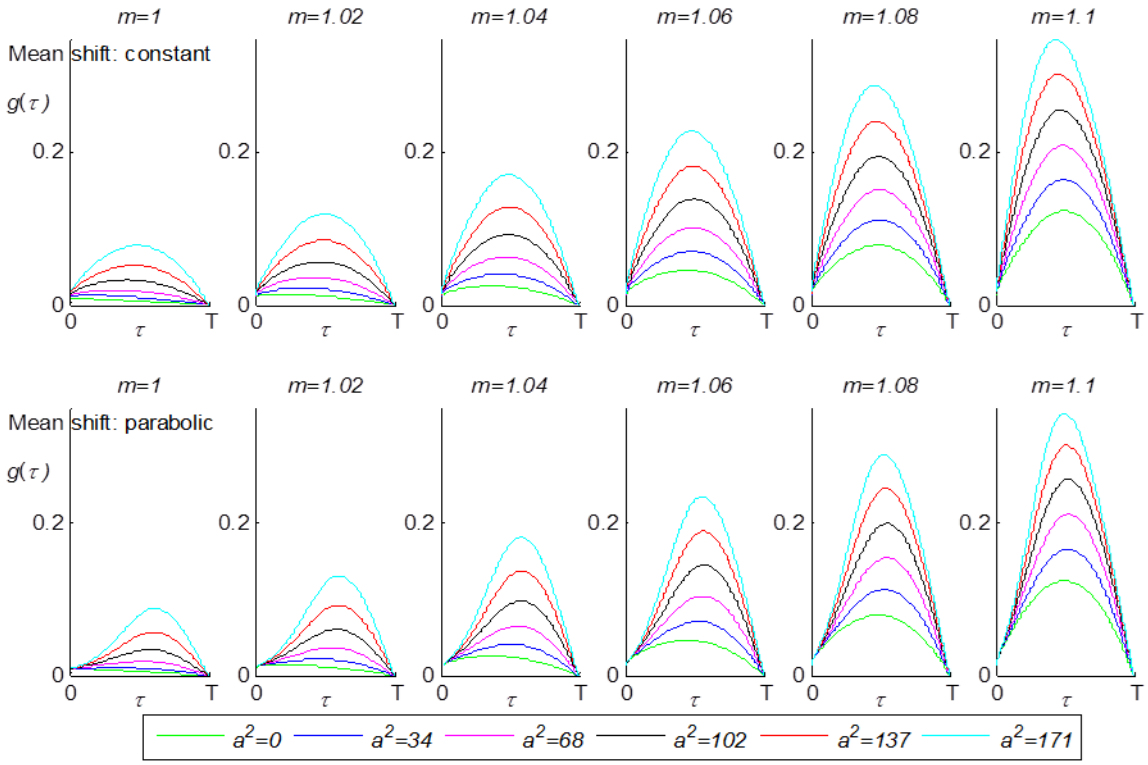


Figure 3.5. Plot of the objective function w.r.t. decision point τ

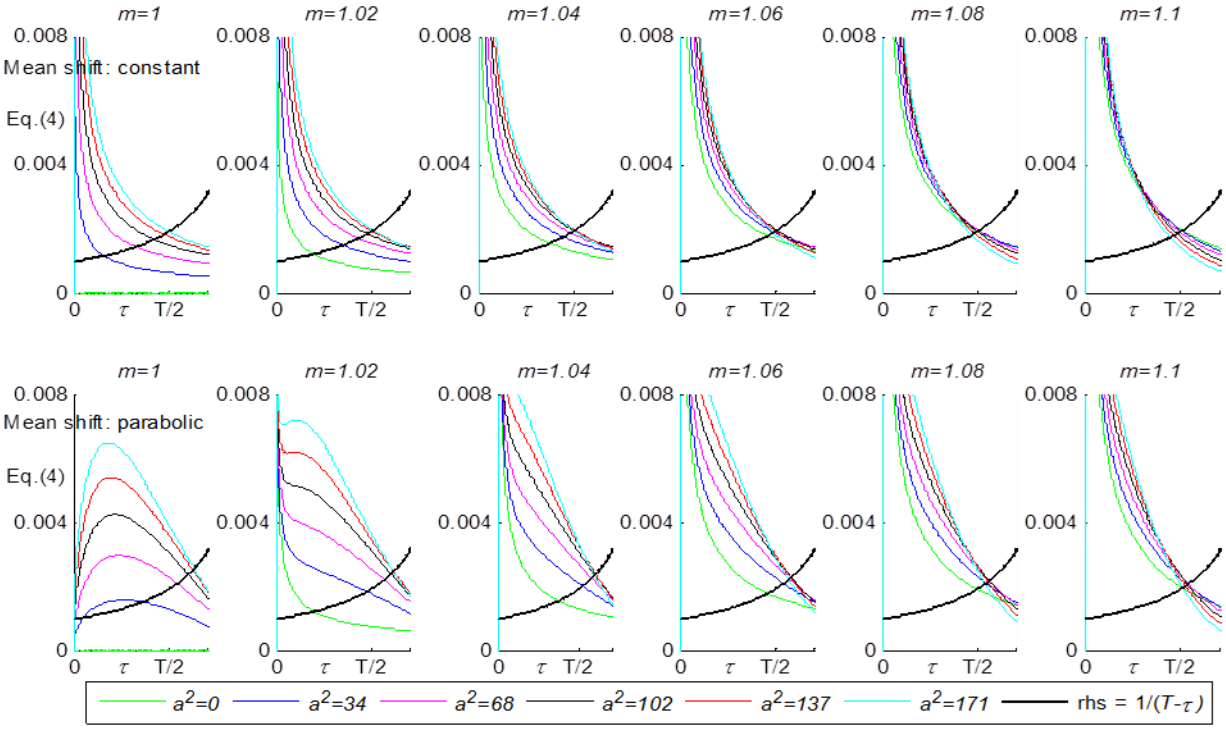


Figure 3.6. Plot of Eq. (3.4) w.r.t. decision point τ

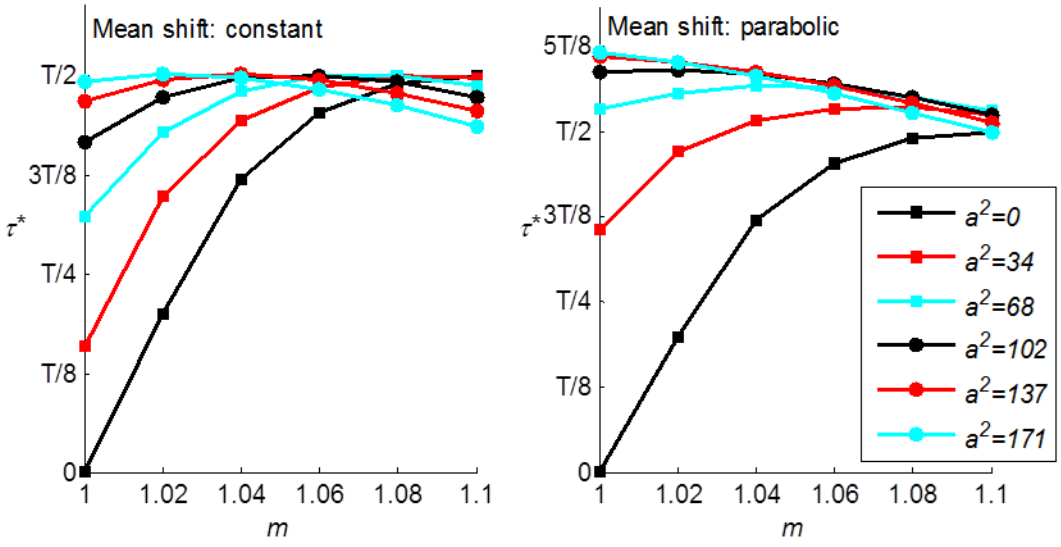


Figure 3.7. Optimal decision point τ^*

Figure 3.8 gives the output of the objective function under the optimal decision point. The objective function essentially represents the percentage of defect prevention since $g(\tau) = (1-\beta(\tau))p_e(\tau)$ describes the percentage of out-of-control control profiles that can be effectively corrected. As can be seen from Figure 3.8, the percentage of defect prevention varies from 1% to 35%, depending on the magnitude of the shift. These simulations validate that the proposed within-cycle monitoring method is effective for defect prevention.

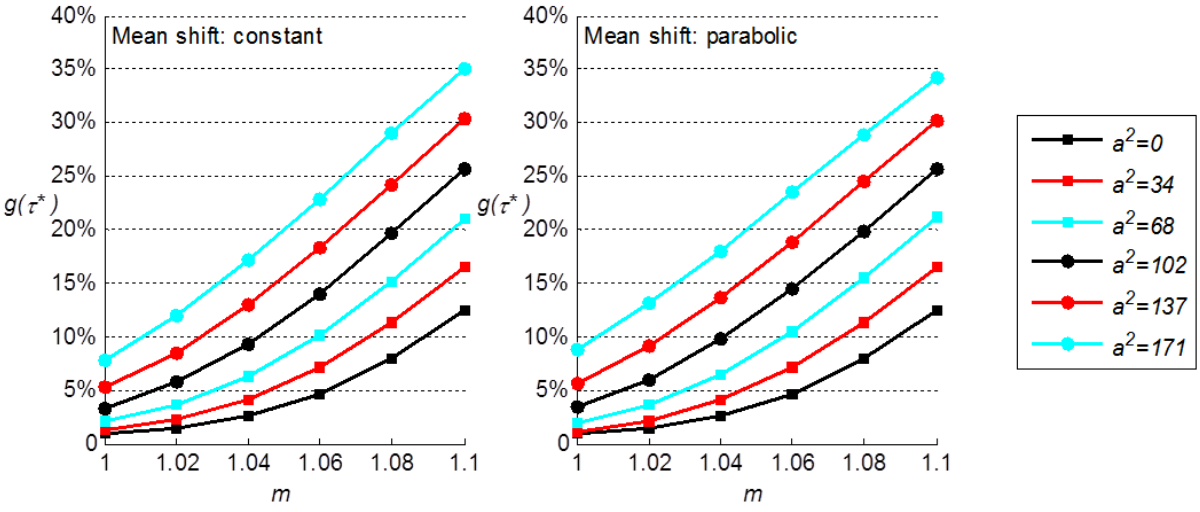


Figure 3.8. Objective function under the optimal decision point

3.4.3 Sensitivity analysis

In order to account for the possible inaccurate online estimations of the a and m values, sensitivity analysis has been performed to show how the performance would be affected by these inaccuracies. We define η to be the percentage of maximal performance that can be achieved when the estimated values deviate from the true values. If online estimation is 100% accurate, $\eta = 100\%$ can always be achieved. In this subsection, we analyze the performance of the proposed monitoring technique with respect to inaccurate estimations of the (1) mean shift and (2) variance change. Based on the findings from sensitivity analysis, we also provide some insights on the online application of the proposed methodology.

Since the mean profile, $\boldsymbol{\mu}_1$ or $\boldsymbol{\mu}_0$, is obtained based on a batch of profiles, the mean shift magnitude a cannot be estimated for each incoming profile. Instead, we would use the mean vector of the profile batch to which the incoming profile belongs. Therefore, it is important to know that if the optimal decision point τ^* is robust to small mean shift, which is not reflected by a. Denote a_0 to be the true value of a, while τ^* is found based on $a = 0$ and an estimated value of $\hat{m}$. Thus,

$$\eta_1 = \frac{g\left(\tau^*|_{a=0,\hat{m}}\right)}{g(\tau^*|_{a=a_0,\hat{m}})} \times 100\%. \tag{3.16}$$

The circles in Figure 3.9 show the η_1 values as defined in Eq. (3.16), with various combinations of a_0 and m. Each curve represents how η_1 values vary with a_0, under a certain m value, where m ranges from 1 to 1.5. It can be seen that the curve descends more significantly if m is smaller.

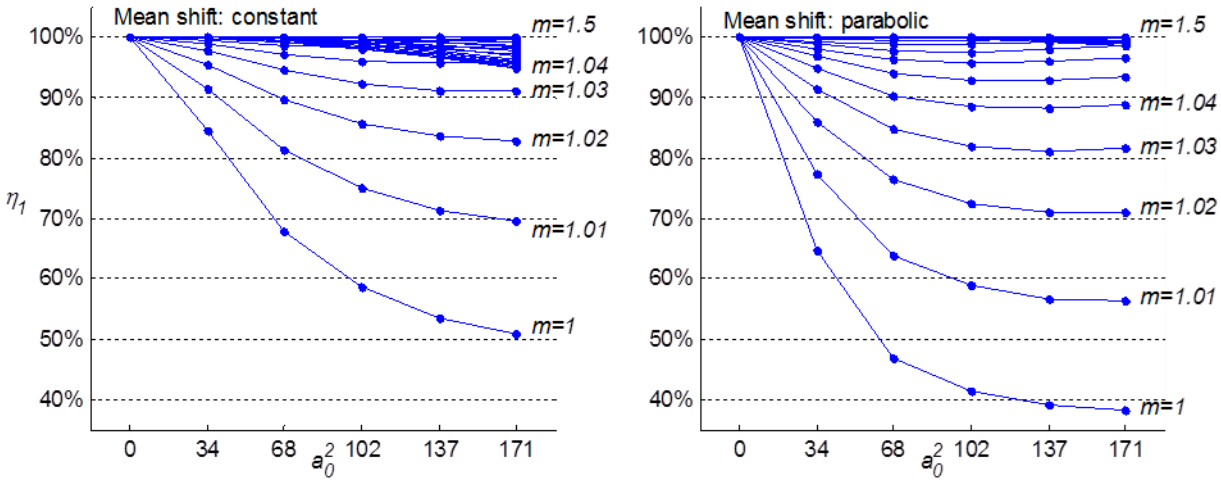


Figure 3.9. The percentage of the maximal performance that can be achieved when a mean shift of a_0 is not considered in determining τ^*

If an incoming profile deviates from the reference, although this deviation is not quantified as a mean shift, it would be considered as the variance change, leading to a certain m value. It can be seen from Figure 3.9 that $\eta_1 > 90\%$ can be guaranteed when $m \geq 1.05$, regardless of the pattern of the mean shift. This finding demonstrates that the proposed monitoring technique and the optimal decision point are robust to inaccurate a values. More importantly, it proves that the online estimation of a is not necessary; the deviations of an incoming profile from the reference should be represented by the variance change.

On the other hand, we are interested to know how the monitoring performance would be affected if the m value from online estimation, $\hat{m}$, differs from the true value, denoted as m_0. Since the optimal decision point τ^* is found based on $\hat{m}$, we have

$$\eta_2 = \frac{g(\tau^*|_{\hat{m},a=0})}{g(\tau^*|_{m_0,a=0})} \times 100\%. \tag{3.17}$$

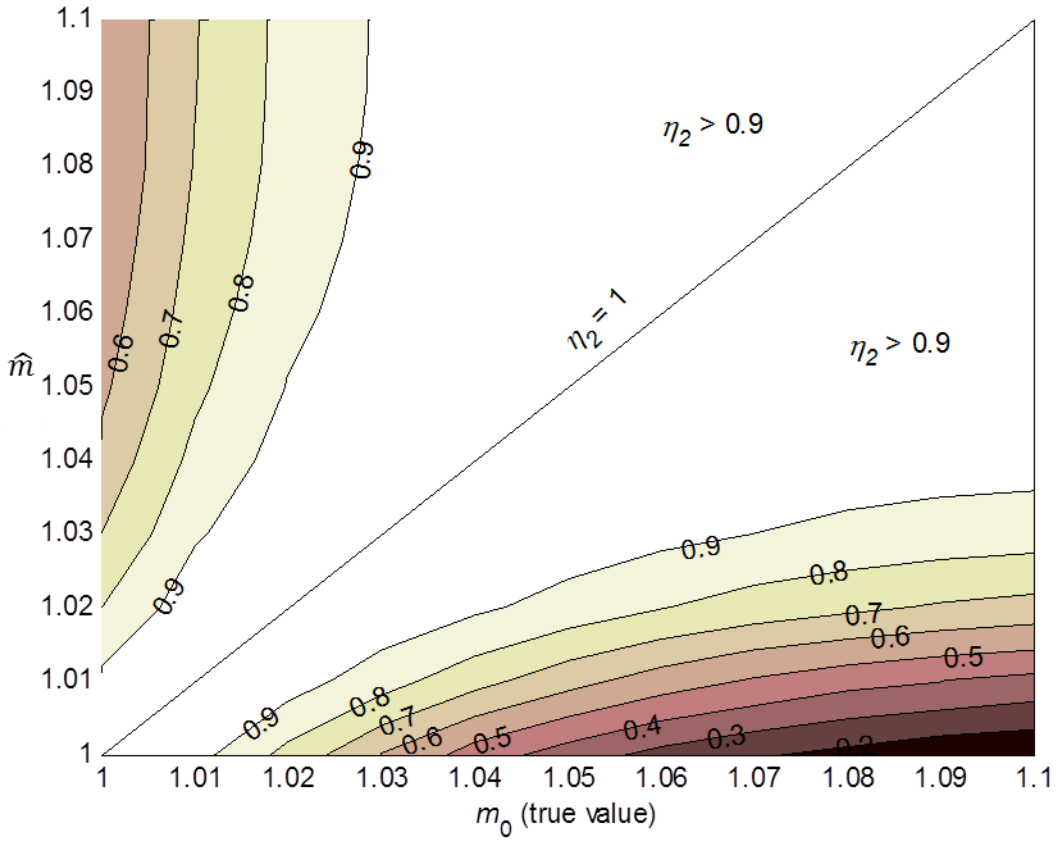


Figure 3.10. The percentage of the maximal performance that can be achieved when the estimated *m* value differs from the true value

Figure 3.10 shows the η_2 values as defined in Eq. (3.17), with various combinations of $\hat{m}$ and m_0. This contour plot shows the dividing lines at $\eta_2 = 100\%, 90\%, 80\%,$ etc. with m ranging from 1 to 1.1. It can be seen from Figure 3.10 that $\eta_2 > 80\%$ is guaranteed in a majority of the simulated scenarios. When the between-profile variation estimation is not very unreliable, $\eta_2 > 90\%$ can be achieved. Furthermore, it can be noticed that the dividing lines are not symmetric on the two sides of $\eta_2 = 100\%$, but η_2 values are much higher on the upper triangle when $\hat{m} > m_0$ than on the lower triangle. This finding demonstrates that it is generally better to round up $\hat{m}$ rather than to round down. It is also suggested to make monitoring decisions and consider control actions based on a more aggressive estimation of the between-profile variation rather than a conservative estimation.

3.5 Case Study

The ultrasonic metal welding example introduced at the beginning of this chapter is analyzed in this section to demonstrate the proposed monitoring technique. Ultrasonic welding is a critical process for joining lithium-ion batteries for electric vehicles. In this case study, welding experiments of two layers of nickel plated copper are investigated. The welding time for each part is $T = 0.4$ sec. Through online data acquisition and data preprocessing, the power signals have $n = 1024$ data points in each profile cycle, i.e., $n_T = n_{0.4} = 1024$ and $\tau = \frac{n_\tau}{n} T$. Figure 3.11 illustrates ten power signals from the in-control and out-of-control processes, respectively. The five profiles from the out-of-control process are collected from welds which have surfaces contaminated with oil.

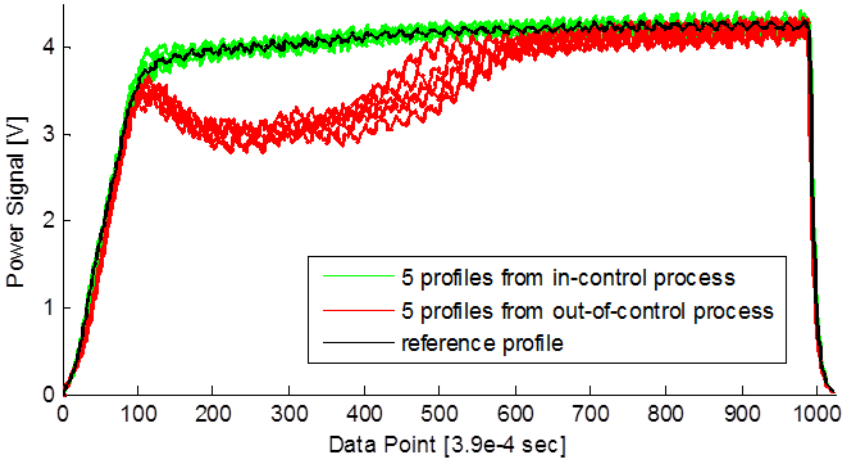


Figure 3.11. Power signals from ultrasonic welding process

Recall that the wavelet coefficients of a raw profile $\boldsymbol{y}_i$ are represented as $\boldsymbol{z}_i = \boldsymbol{\mu} + \boldsymbol{b}_i + \tilde{\boldsymbol{\varepsilon}}_i$, $\boldsymbol{z}_i \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Lambda} + \sigma^2 \boldsymbol{I})$. Based on the above five in-control profiles, we establish the baseline parameters for the in-control process: $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$, $\boldsymbol{\Lambda} = \hat{\boldsymbol{\Lambda}}$, and $\sigma = \hat{\sigma}$. The detailed derivation of these parameters can be found in Paynabar and Jin (2011). $\hat{\boldsymbol{\Lambda}}$ is a diagonal matrix with various diagonal elements.

3.5.1 Surrogate data

An in-control profile i is generated as $y_{ij}^{in} = f_j + b_{ij} + e_{ij}$, where f_j is the value of $\boldsymbol{f}$ at data point j, $\boldsymbol{f}$ is the true function in the original domain obtained by using IDWT with $\boldsymbol{\mu}$; $b_{ij} \sim N(0, s^2 f_j^2)$ represents the between-profile variation, where s is the coefficient of variation, which is the ratio of the standard deviation of $f_j + b_{ij}$ to its mean value f_j; the within-profile variation is represented by $e_{ij} \sim N(0, \sigma^2)$.

An out-of-control profile with only the mean shift is generated as $y_{ij}^{out} = \begin{cases} f_j - u_j + b_{ij} + e_{ij} & \text{if } j \in [j_0, j_1] \\ f_j + b_{ij} + e_{ij} & \text{otherwise} \end{cases}$, where the mean shift occurs in interval $j \in [j_0, j_1]$, $\boldsymbol{u}$ represents a disturbance function. To be consistent with the mean shift patterns in Section 3.4, define (i) a step function for the mean shift: $u_j = k(sf_j + \sigma)$, and (ii) a parabolic function for the mean shift: $u_j = u_0 \frac{4(j-j_0)(j_1-j)}{(j_1-j_0)^2}, u_0 = k(s\bar{f}_{[j_0,j_1]} + \sigma)$.

An out-of-control profile with only the variance change is generated as $y_{ij}^{out} = \begin{cases} f_j + m^2 \cdot b_{ij} + e_{ij} & \text{if } j \in [j_0, j_1] \\ f_j + b_{ij} + e_{ij} & \text{otherwise} \end{cases}$, where the variance change occurs in interval $j \in [j_0, j_1]$, $m^2 b_{ij} \sim N(0, m^2 s^2 f_j^2)$ is the new between-profile variation under the out-of-control scenario.

Surrogate data are generated for both in-control and out-of-control profiles. Table 3.3 lists the six out-of-control scenarios to be studied. $s = 0.02$ is selected in reference to the baseline $\boldsymbol{\Lambda}$. The process change occurs in interval $j \in [j_0, j_1], j_0 = 10, j_1 = 400$, which is an early portion of the power signal. The in-control profiles and 6 out-of-control scenarios are illustrated in Figure 3.12, where the in-control profiles are plotted in blue and the out-of-control profiles are plotted in red. An enlarged area in $j \in [200,400]$ and $y \in [3.5, 4.5]$ is also shown in each subplot.

Table 3.3. Out-of-control scenarios in case study

Case Study	Out-of-control scenario
Case A	Mean shift (step) with $k = 1$
Case B	Mean shift (step) with $k = 2$
Case C	Mean shift (parabolic) with $k = 1$
Case D	Mean shift (parabolic) with $k = 2$
Case E	Variance change with $m^2 = 2$
Case F	Variance change with $m^2 = 3$

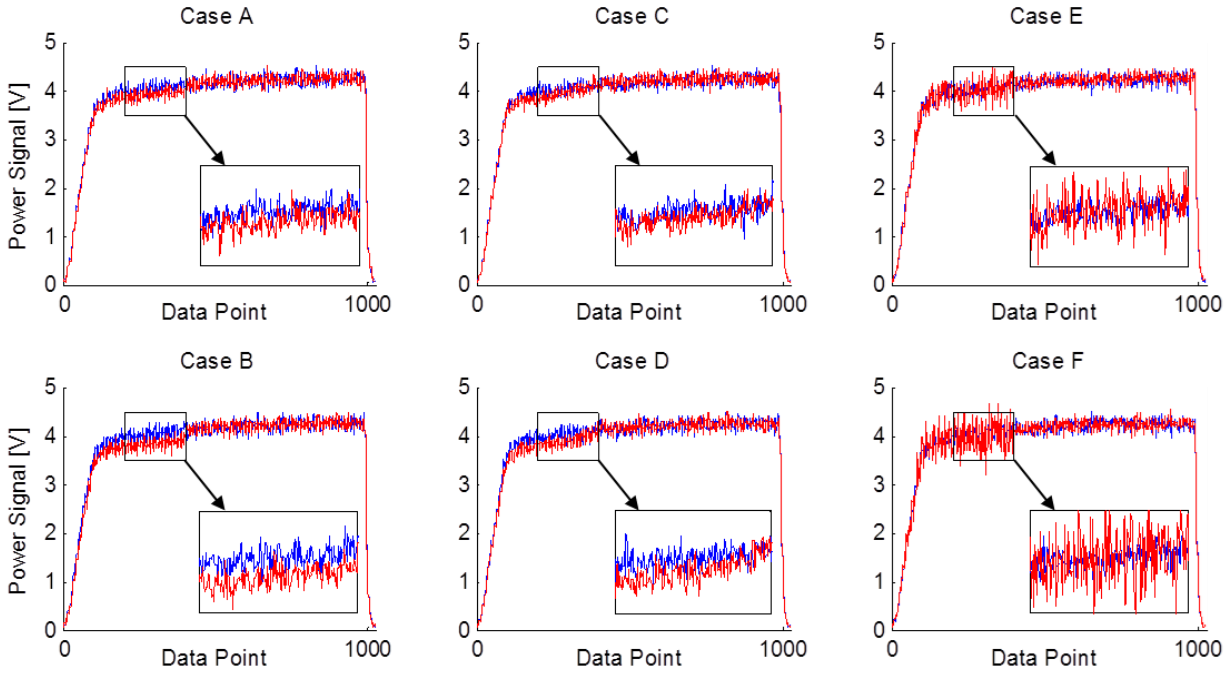


Figure 3.12. In-control profiles and out-of-control scenarios in case study

3.5.2 Results and discussion

To implement the proposed monitoring technique, let $\alpha = 0.01, D_{nc} = 0.05, p_e(\tau) = 1 - \frac{\tau}{T}$. At each possible decision point τ, control charts are established with UCL and profile deviations $\delta_{i,\tau}^2$. Control chart performance $\alpha(\tau)$ and $\beta(\tau)$ are then estimated; the overall performance of within-cycle process monitoring and control, D_{wc}, can also be derived. In

this case study, we have generated a total of 200 profiles (100 in-control, 100 out-of-control) in the training stage and a total of 2000 profiles (1000 in-control, 1000 out-of-control) in the validation stage. To further overcome the random effects in simulation, 10 replications of simulations with independent and identically distributed data have been performed in the training stage and 20 replications have been performed in the validation stage. We assume that the performance in each scenario is represented by the average performance from the multiple replications.

As mentioned in Section 3.3, the optimal decision point is found when D_{wc} is minimized and $L(\tau) \equiv \frac{\alpha(\tau)}{1-\beta(\tau)} < \frac{p_e(\tau)}{1/D_{nc}-1} \equiv H(\tau)$ is satisfied. Take Case F for example. Figure 3.13 shows these critical criteria in model training and validation for Case F. For decision point τ, the blue dot in Figure 3.13 indicates $L(\tau)$ and the red line represents $H(\tau)$. Since $D_{nc} = 0.05$ and $p_e(\tau)$ is a linear function, $H(\tau)$ is also a linear function of τ. The condition for control opportunity is given by the region where $L(\tau) < H(\tau)$ is satisfied, i.e., where the blue dots are below the red line. The green dots in Figure 3.13 represents $D_{wc}(\tau)$. Therefore, τ^* is found to be at data point $j^* = 45$, i.e., $\tau^* = 45 \cdot \frac{0.4}{1024}$(sec) in both training and validation stages. Figures on the performance measurement in other cases are given in Appendix. As a summary of the results, the optimal decision points in both training and validation stages for all cases studied are shown in Figure 3.13. Since multiple independent replications have been carried out, Figure 3.13 also shows the ranges of τ^* values obtained from each replication. Considering the length of data points is 1024, the τ^* values found in training and validation stages are very close. The following observations are also made based on the case study results:

- Comparing Case A with Case B, a smaller τ^* is suggested in Case B when the magnitude of the parameter shift is larger. Intuitively, a larger shift should be easier to

detect, but it may also require more control efforts. Hence, within-cycle process monitoring and control should be considered at an earlier time for a larger shift. The same observation can also be made when comparing Case C with Case D, or Case E with Case F.

- Comparing Case A with Case C, or Case B with Case D, a larger τ^* is suggested in Case C (or D) when the mean shift takes on a parabolic function. The mean shift under a parabolic function is not as significant as that under a step function at the beginning of the shift, thus delaying the decision point for within-cycle monitoring and control.

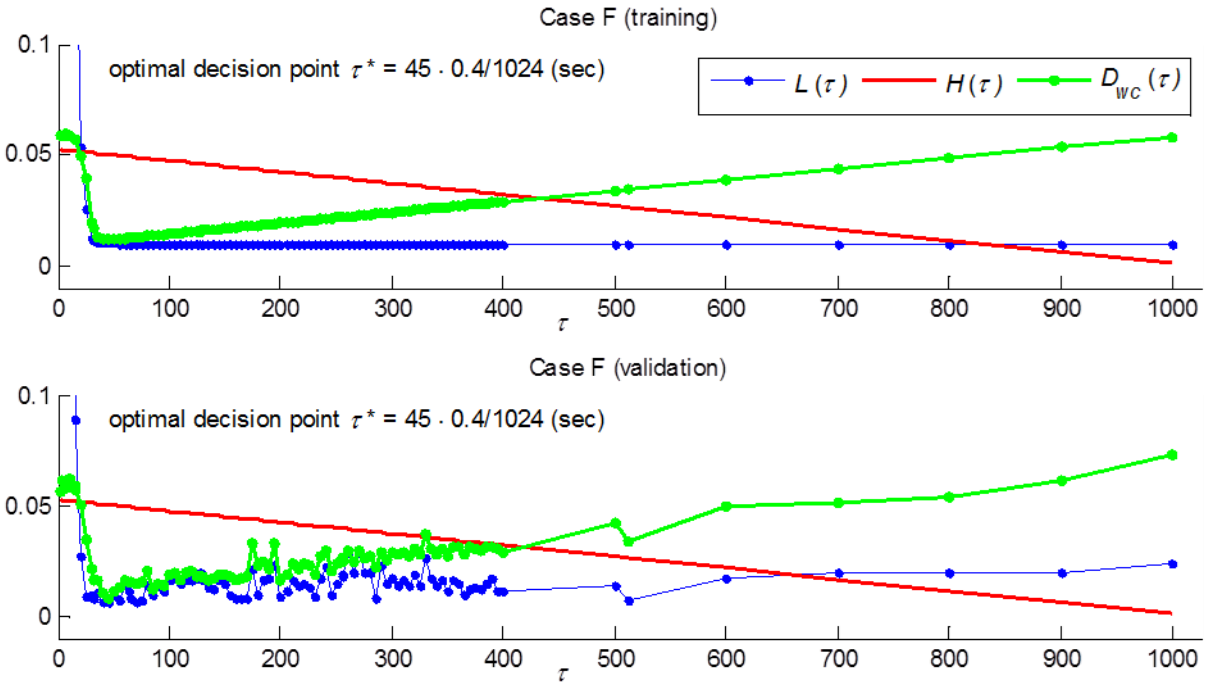


Figure 3.13. Performance measurement of within-cycle monitoring and control in Case F

Table 3.4. Optimal decision point in case study

Optimal decision point $\boldsymbol{\tau^* = \frac{j^*}{n}T}$	Training		Validation	
	average of j^*	range of j^*	average of j^*	range of j^*
Case A	45	[32, 50]	55	[64, 85]
Case B	20	[15, 20]	30	[20, 64]
Case C	155	[128, 145]	140	[165, 256]
Case D	50	[50, 55]	64	[75]
Case E	65	[60, 70]	65	[64, 95]
Case F	45	[35, 50]	45	[45]

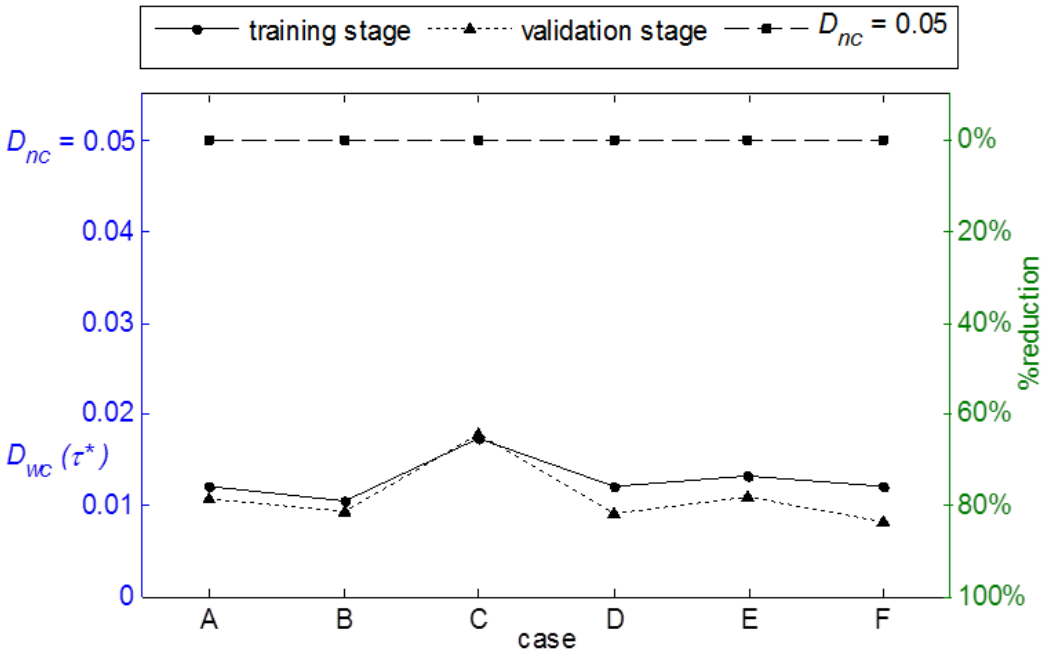


Figure 3.14. D_{wc} value at τ^* and the percentage of reduction from D_{nc} to $D_{wc}(\tau^*)$ in case study

Using the optimal decision points found in each case, the values of D_{wc} at τ^* are further summarized in Figure 3.14. Plotting $D_{wc}(\tau^*)$ against $D_{nc} = 0.05$, we notice the huge improvement from D_{nc} to D_{wc}, indicating that the proportion of out-of-control profiles is significantly reduced when the proposed within-cycle monitoring technique is adopted and

real-time control actions are implemented. We also present the percentage of this reduction on the right side vertical axis of Figure 3.14, where $\%\, reduction = (1 - D_{wc}(\tau^*)/D_{nc}) \times 100\%$. Among all six cases, the percentage of reduction from D_{nc} to $D_{wc}(\tau^*)$ ranges from 64% to as high as 84%. As a result, utilizing the proposed method is promising in preventing more than 60% of potential defects.

Based on the above results, we conclude that optimal decision point exists and that the proposed within-cycle monitoring and control strategy is effective under various scenarios in this case study. The reduction from D_{nc} to $D_{wc}(\tau^*)$ proves that the proposed method is promising in preventing defects in real time. In this case study, we have assumed that process change occurs in an early stage of the operation and that potential defects can be prevented by a one-step adjustment of process parameter(s) during the operation. These assumptions are essential in providing control opportunities and then for the proposed within-cycle monitoring and control strategy to benefit. Developing within-cycle monitoring and control strategies for multi-step adjustment of process parameter(s) will be an interesting topic for future research. Furthermore, the extension of the developed monitoring strategy for a single type of fault to within-cycle monitoring and diagnosis of multiple types of fault would be a valuable development that is worth future research efforts.

3.6 Conclusion

A new wavelet-based profile monitoring method has been developed by considering the tradeoff between real-time monitoring accuracy with within-cycle control effectiveness for defect prevention. In order to leave sufficient time for real-time process adjustment, the monitoring decision is made based on an early portion of the cyclic signal, while the optimal decision point for achieving the most benefit in defect prevention is determined by solving an optimization problem. Wavelet-based control charts are then developed to monitor profile

deviations and detect process changes. The effectiveness of the proposed method is validated and demonstrated by simulations and case studies. With the developed within-cycle profile monitoring strategy, the proportion of out-of-control profiles is significantly reduced, indicating great potential in preventing defects in real time. In addition, results from the sensitivity analysis provide insights on the online application of the proposed method and its robustness against inaccuracies in online estimation.

The proposed within-cycle monitoring technique can be applied to general discrete cyclic manufacturing processes that have the online sensing and control capabilities. The results of this research are also highly applicable or expandable to mission-critical applications when improving product quality and preventing defects are of high interests.

Appendix

The figures in Appendix show the performance measurement in Cases A~E for the case study. The blue dots in the figures indicate $L(\tau)$ values, the red lines represent $H(\tau)$, and the green dots represents $D_{wc}(\tau)$. The condition for control opportunity is given by the region where $L(\tau) < H(\tau)$ is satisfied, i.e., where the blue dots are below the red line. The optimal decision point τ^* is found when D_{wc} is minimized and the condition for control opportunity is satisfied. The results on τ^* are also indicated in the figures.

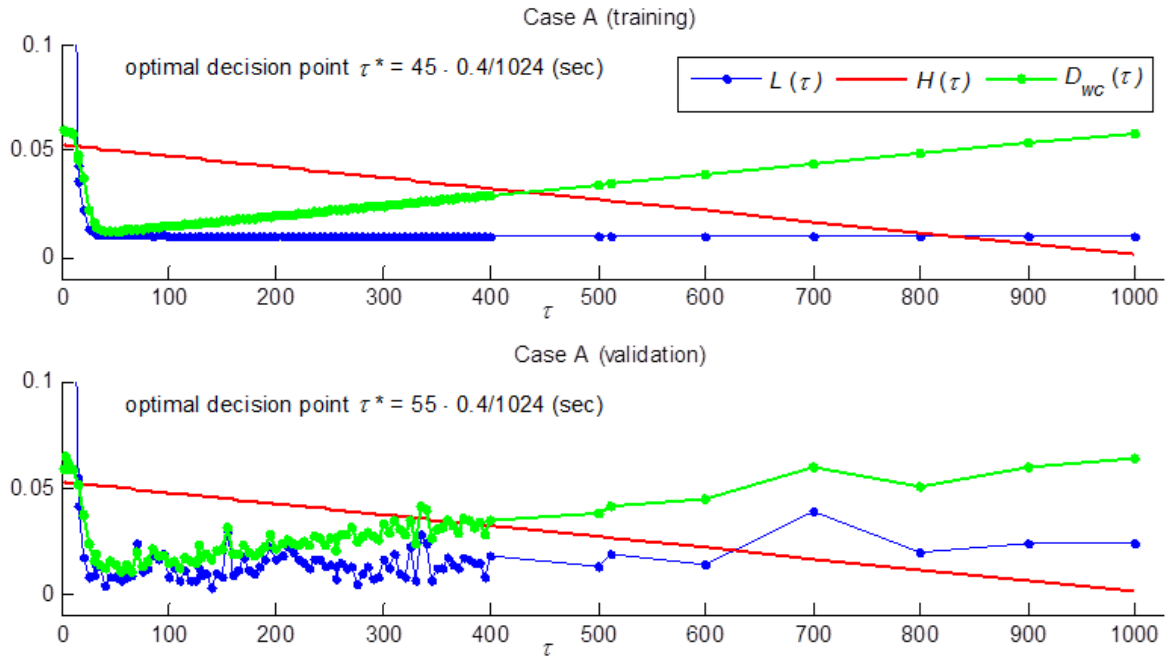


Figure 3.A.1. Performance measurement of within-cycle monitoring and control in Case A

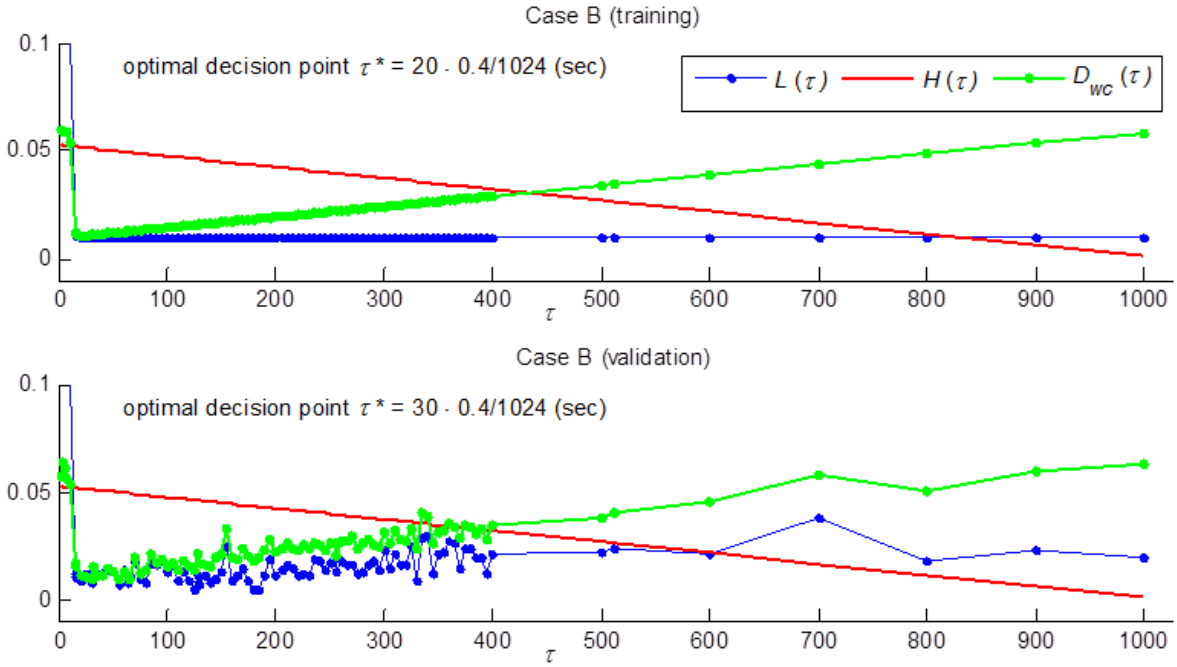


Figure 3.A.2. Performance measurement of within-cycle monitoring and control in Case B

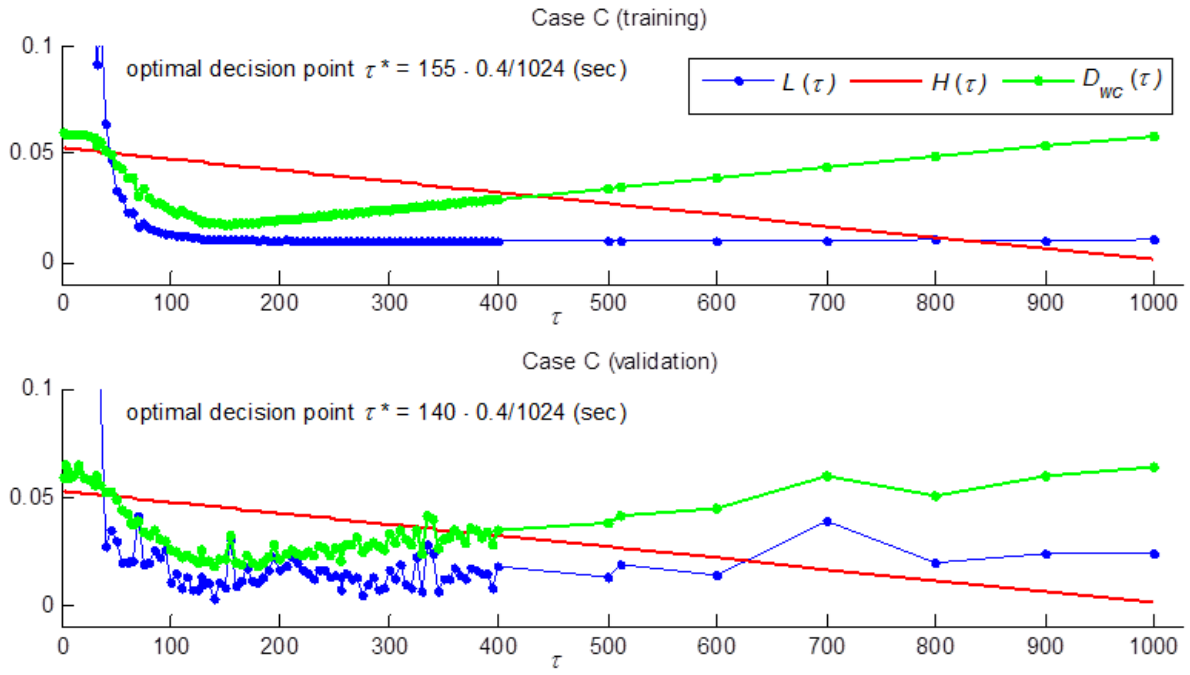


Figure 3.A.3. Performance measurement of within-cycle monitoring and control in Case C

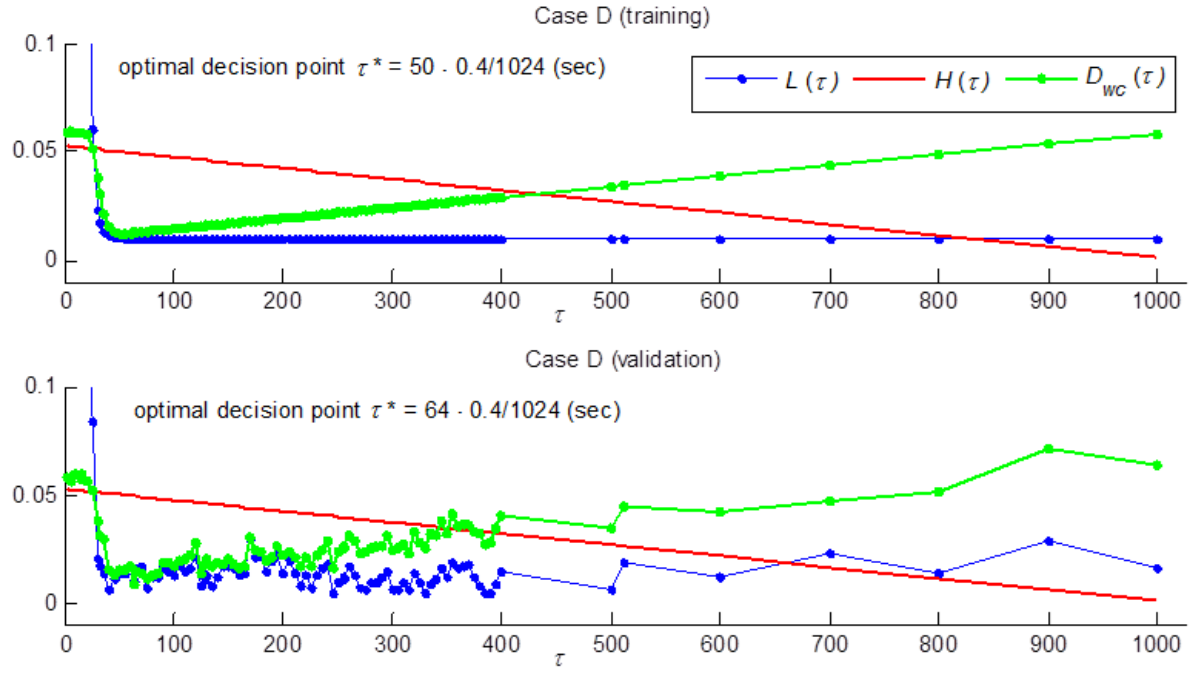


Figure 3.A.4. Performance measurement of within-cycle monitoring and control in Case D

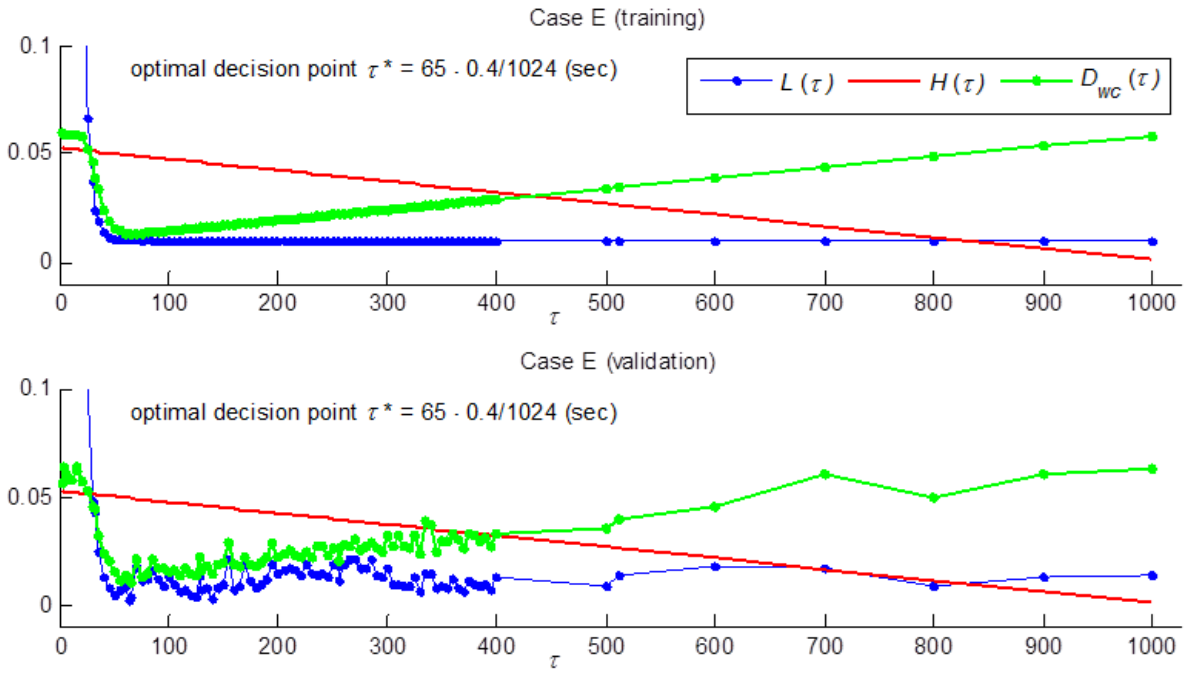


Figure 3.A.5. Performance measurement of within-cycle monitoring and control in Case E

References

Chang, S.-I., Tavakkol, B., Chou, S.-H. and Tsai, T.-R. (2014) Real-time Detection of Wave Profile Changes. *Computers & Industrial Engineering*, 75, 187-199.

Chang, T.C. and Gan, F.-F. (2006) Monitoring Linearity of Measurement Gauges. *Journal of Statistical Computation and Simulation*, 76, 889-911.

Chicken, E., Pignatiello, J., Joseph and Simpson, J.R. (2009) Statistical Process Monitoring of Nonlinear Profiles Using Wavelets. *Journal of Quality Technology*, 41, 198-212.

Daubechies, I. (1992) *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA.

Davidian, M. and Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurements Data*, Chapman and Hall, London, UK.

Demidenko, E. (2004) *Mixed Models: Theory and Applications*, Wiley, New York, NY.

Ding, Y., Zeng, L. and Zhou, S. (2006) Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes. *Journal of Quality Technology*, 38, 199-216.

Gupta, S., Montgomery, D.C. and Woodall, W.H. (2006) Performance Evaluation of Two Methods for Online Monitoring of Linear Calibration Profiles. *International Journal of Production Research*, 44, 1927-1942.

Jeong, M.K., Lu, J.C. and Wang, N. (2006) Wavelet-based SPC Procedure for Complicated Functional Data. *International Journal of Production Research*, 44, 729-744.

Jin, N., Zhou, S., Chang, T.-S. and Huang, H.-H. (2008) Identification of Influential Functional Process Variables for Surface Quality Control in Hot Rolling Processes. *IEEE Transactions on Automation Science and Engineering*, 5, 557-562.

Kazemzadeh, R.B., Noorossana, R. and Amiri, A. (2008) Phase I Monitoring of Polynomial Profiles. *Communications in Statistics - Theory and Methods*, 37, 1671-1686.

Kazemzadeh, R.B., Noorossana, R. and Amiri, A. (2009) Monitoring Polynomial Profiles in Quality Control Applications. *International Journal of Advanced Manufacturing Technology*, 42, 703-712.

Lee, S.S., Kim, T.H., Hu, S.J., Cai, W.W. and Abell, J.A. (2010) Joining Technologies for Automotive Lithium-ion Battery Manufacturing – A Review. *Proceedings of ASME 2010 International Manufacturing Science and Engineering Conference*,541-549.

Mahmoud, M.A. (2008) Phase I Analysis of Multiple Linear Regression Profiles. *Communications in Statistics - Simulation and Computation*, 37, 2106-2130.

Mahmoud, M.A., Parker, P.A., Woodall, W.H. and Hawkins, D.M. (2007) A Change Point Method for Linear Profile Data. *Quality and Reliability Engineering International*, 23, 247-268.

Mallat, S. (1999) *A Wavelet Tour of Signal Processing*, Academic Press, Burlington, MA.

Montgomery, D.C. (2013) *Introduction to Statistical Quality Control*, Wiley, Hoboken, NJ.

Noorossana, R., Amiri, A., Vaghefi, S.A. and Roghanian, E. (2004) Monitoring Process Performance Using Linear Profiles. *Proceedings of the 3rd International Industrial Engineering Conference*, Tehran, Iran.

Paynabar, K. and Jin, J. (2011) Characterization of Non-Linear Profiles Variations Using Mixed-Effect Models and Wavelets. *IIE Transactions*, 43, 275-290.

Williams, J.D., Birch, J.B., Woodall, W.H. and Ferry, N.M. (2007) Statistical Monitoring of Heteroscedastic Dose-Response Profiles from High-Throughput Screening. *Journal of Agricultural, Biological, and Environmental Statistics*, 12, 216-235.

Williams, J.D., Woodall, W.H. and Birch, J.B. (2007) Statistical Monitoring of Nonlinear Product and Process Quality Profiles. *Quality and Reliability Engineering International*, 23, 925-941.

Woodall, W.H. (2007) Current Research on Profile Monitoring. *Produção*, 17, 420-425.

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, 309-320.

Zhou, S., Sun, B. and Shi, J. (2006) An SPC Monitoring System for Cycle-based Waveform Signals using Haar Transform. *IEEE Transactions on Automation Science and Engineering*, 3, 60-72.

Zou, C., Tsung, F. and Wang, Z. (2007) Monitoring General Linear Profiles Using Multivariate Exponentially Weighted Moving Average Schemes. *Technometrics*, 49, 395-408.

Zou, C., Zhang, Y. and Wang, Z. (2006) A Control Chart based on a Change-point Model for Monitoring Profiles. *IIE Transactions*, 38, 1093-1103.

CHAPTER 4

PROFILE MONITORING AND FAULT DIAGNOSIS VIA SENSOR FUSION FOR MULTI-STREAM DATA

4.1 Introduction

The wide applications of low-cost and smart sensing devices along with fast and advanced computer systems have resulted in a rich data environment, which makes a large amount of data available in many applications. Sensor signals acquired during the process contain rich information that can be used to facilitate effective monitoring of operational quality, early detection of system anomalies, quick diagnosis of fault root causes, and intelligent system design and control. In discrete manufacturing and many other applications, the sensor measurements provided by online sensing and data capturing technology are time- or spatial-dependent functional data, also called profile data (Woodall *et al.*, 2004; Woodall, 2007). In this chapter, we are particularly interested in cycle-based profile data, which are collected from repetitive operational cycles of a manufacturing process. Examples of cycle-based profile data include the tonnage signals in stamping processes (Jin and Shi, 1999), the pressing force signals in a valve seat assembly operation (Paynabar and Jin, 2011), and the power signals and displacement signals in ultrasonic metal welding (Lee *et al.*, 2014).

There is extensive research on the modeling and monitoring of cycle-based profile data in the literature, including both linear profiles and nonlinear profiles. An overview of parametric and nonparametric approaches for profile data as well as application domains can be found in Noorossana *et al.* (2012). A large portion of profile monitoring literature focuses on single

signal analysis. This is a convenient way to extract and analyze sensory data in cases where the performance of the process can be effectively reflected by individual profiles. However, there is a strong industrial interest for multi-signal applications, especially in cases where a single signal does not provide enough information to effectively evaluate the performance of the process. This leads to an increasing demand for multi-sensor fusion methods to analyze the multiple signals captured from different sensors for process monitoring and system diagnostics purposes.

One motivating example is the ultrasonic metal welding process for joining lithium-ion batteries for electric vehicles (Lee *et al.*, 2010), as illustrated in Figure 4.1. It is important to develop an online monitoring method to facilitate in-process quality control and fault diagnosis to allow for a faster implementation of corrective actions. In order to have a better understanding of the process, four sensors are installed in the welding machine (Hu, 2011): the power meter monitors controller power signal, the force sensor measures the clamping force, the displacement sensor measures the displacement between horn and anvil, and the microphone captures the sound during vibration. Figure 4.2(a) shows the signals from these four sensors for samples from the normal welding process and three faulty processes: (1) surface contamination, (2) abnormal thickness, and (3) mislocated/edge weld. Figure 4.2(b) shows the welded tabs associated with these processes. In general, the normal welding process produces good welds with strong connections, while the faulty processes tend to create poor quality connections which may have adverse effects on the performance of the battery pack. If samples are contaminated, for example, with oil, there is less friction between the metal layers, causing insufficient vibration at the beginning of the weld. So, the power signal does not rise as fast as a normal weld does. Once oil gets removed by vibration, the power signal picks up. Abnormal welding thickness may be caused by material handling errors, or sheet metal distortion, or operation errors. The displacement signal clearly shows

how the displacement between horn and anvil is affected by thicker layers. Mislocated/edge weld may be caused by operation errors or alignment errors. With edge weld, all clamping force is applied to a smaller weld region, resulting in more displacement between horn and anvil towards the end of the weld. It can be seen from Figure 4.2 that on the one hand each signal contains richer information about product quality and process condition than any single point can provide, and on the other hand a single stream of signals is not informative enough for recognizing the type of faults.

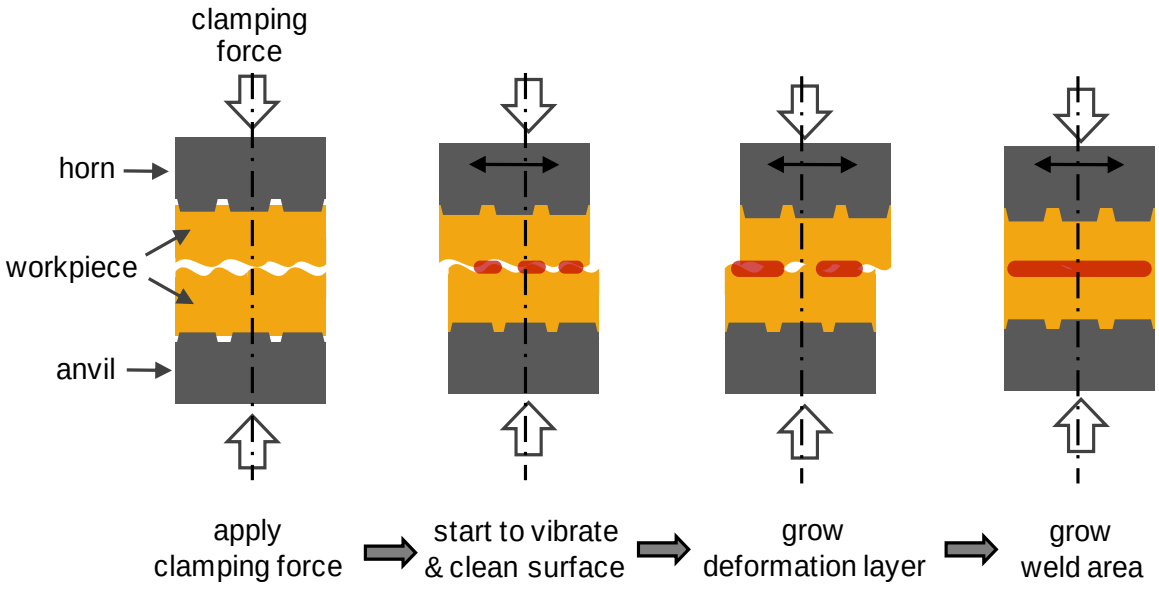
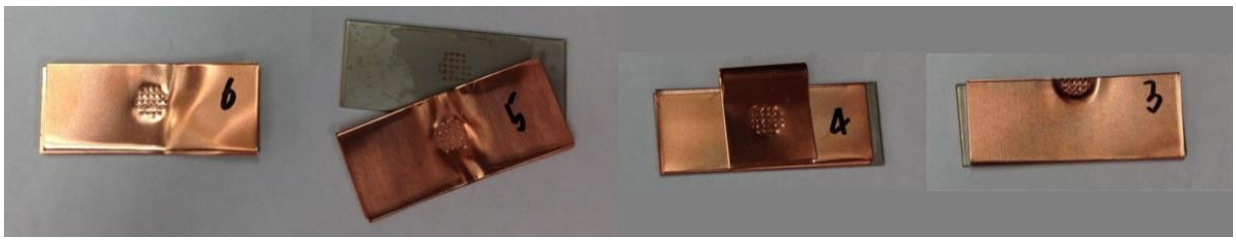
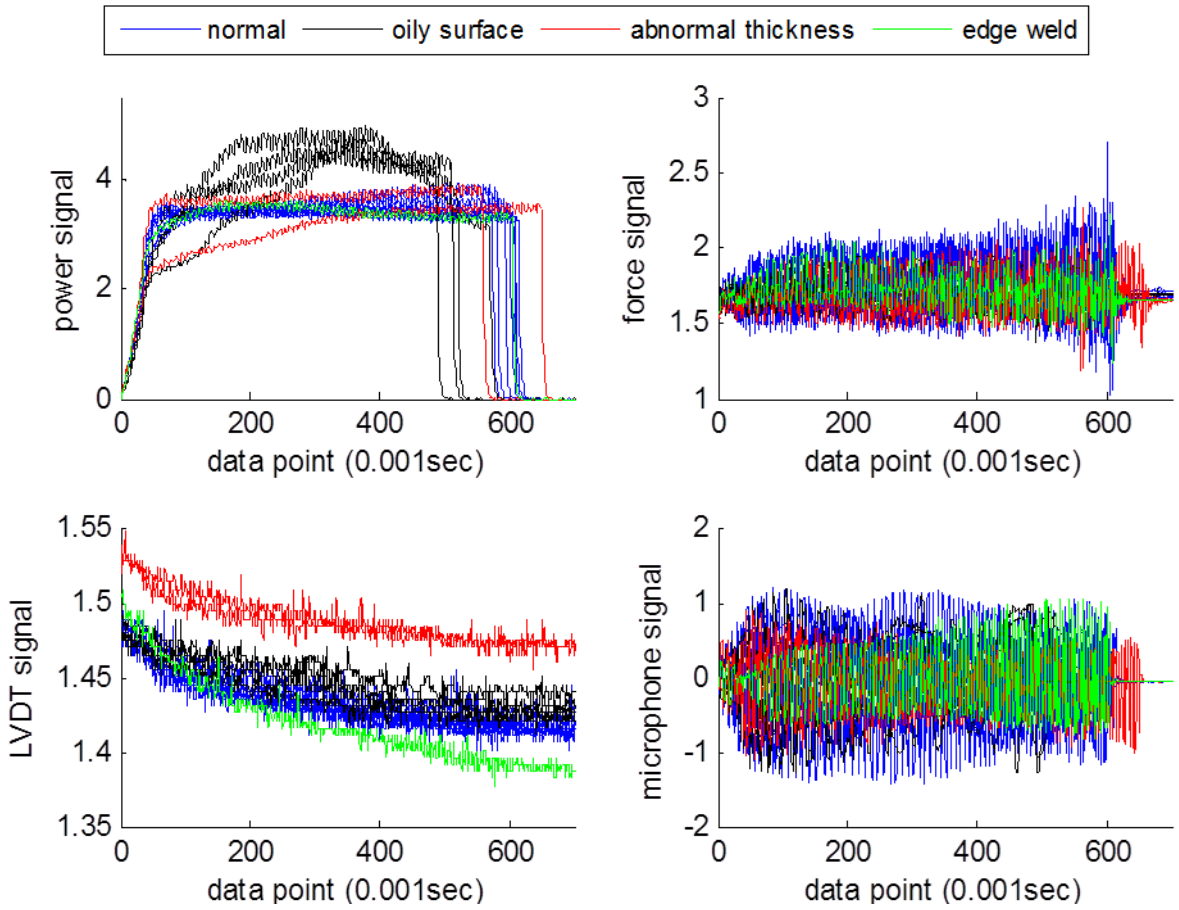


Figure 4.1. Ultrasonic metal welding process (Lee *et al.*, 2010)



(a) Welds from the normal welding process and three faulty processes: surface contamination, abnormal thickness, and mislocated/edge weld (from left to right)



(b) Sensor signals from the normal welding process and three faulty processes

Figure 4.2. Sensor signals and samples from ultrasonic metal welding processes

There have been many research efforts on multi-sensor data fusion in manufacturing operations, for example, chatter detection in milling (Kuljanic *et al.*, 2009), tool condition monitoring (Cho *et al.*, 2010; Grasso *et al.*, 2013), engine fault diagnosis (Basir and Yuan, 2007), etc. A large portion of the multi-sensor data fusion methods is based on extracting a single synthetic index from the monitoring signals, e.g., a weighted summation of signals. The main limitations of this approach include the loss of information involved in the feature extraction process, the loss of sensor-to-sensor correlations, and the problem-dependent nature of the synthesizing scheme. Although profile monitoring techniques have been demonstrated to be more effective than synthetic index-based methods in monitoring

processes characterized by repeating patterns (Noorossana *et al.*, 2012), only a few authors have studied profile monitoring approaches in the field of sensor fusion (Kim *et al.,* 2006; Amiri *et al.,* 2013; Chou *et al.,* 2014). Recently, with the fast development of multilinear methods for face recognition, Paynabar *et al.* (2013) proposed a multi-channel profile monitoring and fault diagnosis method based on uncorrelated multilinear principal component analysis (UMPCA) (Lu *et al.,* 2009), whereas Grasso *et al.* (2014) investigated the problem of multi-stream profile monitoring using multilinear PCA (MPCA) (Lu *et al.,* 2008). Multi-channel profiles are homogeneous, in which all sensors measure the same variable, whereas multi-stream signals are heterogeneous, in which various sensors measure different variables.

In this study, we investigate the use of multilinear extensions of linear discriminant analysis (LDA) to deal with multi-stream signals for the purpose of process monitoring and fault diagnosis. LDA has been widely used as an effective tool for dimension reduction and discriminant analysis of complex data. Regular LDA is a linear algorithm that can only operate on vectors, thus cannot be directly applied to multi-stream profiles. To apply LDA to multi-stream profiles, these profiles need to be combined and reshaped (vectorized) into vectors first. So, this method is referred to as Vectorized-LDA (VLDA). Applying LDA to this high-dimensional vector creates high computational complexity due to the dimension of scatter matrices. Moreover, vectorization breaks the natural structure and correlation in the original data, e.g., sensor-to-sensor correlation, and potentially loses more useful representations that can be obtained in the original form. Lu *et al.* (2009) introduced an uncorrelated multilinear LDA (UMLDA) framework as an alternative to VLDA. UMLDA is a multilinear dimensionality reduction and feature extraction method that operates directly on the multidimensional objects, known as tensor objects, rather than their vectorized versions. The UMLDA extracts uncorrelated discriminative features directly from tensorial data

through solving a tensor-to-vector projection. Although MPCA and UMPCA are also multilinear subspace feature extraction algorithms operating directly on the tensorial representations, similar to PCA, they are both unsupervised methods that do not make use of the class information. In manufacturing and many other applications, training samples from various classes can be easily collected in an efficient manner. In these applications, supervised multilinear methods like UMLDA take class information into considerations and thus may be more suitable for fault recognition. Although there is some exploratory research on the applications of UMLDA to image processing on face and gait recognition tasks (Lu *et al.*, 2009), very little research could be found in the literature on using the UMLDA technique for analyzing multi-stream nonlinear profiles for the purpose of fault detection and diagnosis. Therefore, the main objective of this chapter is to propose a UMLDA-based approach for analyzing multi-stream profiles that considers the interrelationship of different sensors. The features extracted by the proposed UMLDA-based method can effectively discriminate different classes and provide fault diagnosis results. The effectiveness of the proposed method is tested on both simulations and a real-world case study in the ultrasonic metal welding process.

The remainder of this chapter is organized as follows. Section 4.2 presents the method for analysis and dimension reduction of multi-stream profiles using UMLDA. VLDA is also reviewed in this section. Section 4.3 compares the proposed UMLDA-based method with VLDA and its variants, and other competitor methods including UMPCA-based and MPCA-based methods in the performance of extracting discriminative features and recognizing the type of faults. A case study of ultrasonic metal welding process is given in Section 4.4. Finally, Section 4.5 concludes the chapter.

4.2 Dimension Reduction of Multi-stream Signals using UMLDA and VLDA

Multi-way data analysis is the extension of two-way methods to higher-order datasets. This section first reviews the basic notations and concepts in multilinear algebra, and then introduces the implementation of UMLDA and VLDA for the purpose of dimensionality reduction in handling multi-stream signals. More details on the theoretical foundations of the mathematical development of UMLDA based on multilinear algebra can be found in De Lathauwer *et al.* (2000), Kolda (2001), and Acar and Yener (2009). The algorithm we use in this chapter for extracting uncorrelated features from tensor data is based on the theories presented in those articles.

4.2.1 Basic multilinear algebra concepts and tensor-to-vector projection

An L-way array $\mathcal{A}$ is an Lth-order tensor object $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_L}$ such that I_l represents the dimension of the l-mode, $l = 1, \ldots, L$, where the term *mode* refers to a generic set of entities (Kiers, 2000). The l-mode vectors of $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_L}$ are defined as the I_l-dimensional vectors obtained from $\mathcal{A}$ by varying the index i_l ($i_l = 1, \ldots, I_l$) while keeping all the other indices fixed. In multilinear algebra, a matrix **A** can be considered to be a second-order tensor. The column vectors and row vectors are considered as the 1-mode and 2-mode vectors of the matrix, respectively. The l-mode product of a tensor $\mathcal{A}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_l \times I_l}$, denoted by $\mathcal{A} \times_l \mathbf{U}$, is a tensor with entries $(\mathcal{A} \times_l \mathbf{U})(i_1, \ldots i_{l-1}, j_l, i_{l+1}, \ldots, i_L) = \sum_{i_l} \mathcal{A}(i_1, \ldots, i_L) \cdot \mathbf{U}(j_l, i_l)$. The scalar product of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_L}$ is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \sum_{i_2} \ldots \sum_{i_L} \mathcal{A}(i_1, i_2, \ldots, i_L) \cdot \mathcal{B}(i_1, i_2, \ldots, i_L)$.

To project tensorial data into a subspace for better discrimination, there are two general forms of multilinear projection: the tensor-to-tensor projection (TTP) and the tensor-to-vector

projection (TVP). The TVP projects a tensor to a vector and it can be viewed as multiple projections from a tensor to a scalar. A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_L}$ can be projected to a point y through L unit projection vectors $\{\mathbf{u}^{(1)^T}, \mathbf{u}^{(2)^T}, \ldots, \mathbf{u}^{(L)^T}\}$ as $y = \mathcal{A} \times_1 \mathbf{u}^{(1)^T} \times_2 \mathbf{u}^{(2)^T} \ldots \times_L \mathbf{u}^{(L)^T} = \left\langle \mathcal{A}, \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \ldots \circ \mathbf{u}^{(L)} \right\rangle \equiv \langle \mathcal{A}, \mathcal{U} \rangle$, $\mathbf{u}^{(l)} \in \mathbb{R}^{I_l \times 1}$, $\left\|\mathbf{u}^{(l)}\right\| = 1$ for $l = 1, \ldots, L$, where $\|\cdot\|$ is the Euclidean norm for vectors. This multilinear projection $\{\mathbf{u}^{(1)^T}, \mathbf{u}^{(2)^T}, \ldots, \mathbf{u}^{(L)^T}\}$ is called an elementary multilinear projection (EMP), which is the projection of a tensor on a single line (resulting a scalar) and it consists of one projection vector in each mode. The TVP of a tensor object $\mathcal{A}$ to a vector $\mathbf{y} \in \mathbb{R}^P$ in a P-dimensional vector space consists of P EMPs, which can be written as $\left\{\mathbf{u}_p^{(1)^T}, \mathbf{u}_p^{(2)^T}, \ldots, \mathbf{u}_p^{(L)^T}\right\}_{p=1,\ldots,P} = \left\{\mathbf{u}_p^{(l)^T}, l = 1, \ldots, L\right\}_{p=1}^{P}$. The TVP from $\mathcal{A}$ to $\mathbf{y}$ is then written as $\mathbf{y} = \mathcal{A} \times_{l=1}^{L} \left\{\mathbf{u}_p^{(l)^T}, l = 1, \ldots, L\right\}_{p=1}^{P}$, where the pth component of $\mathbf{y}$ is obtained from the pth EMP as $\mathbf{y}(p) = \mathcal{A} \times_1 \mathbf{u}_p^{(1)^T} \times_2 \mathbf{u}_p^{(2)^T} \ldots \times_L \mathbf{u}_p^{(L)^T}$.

In the frame of multi-stream profile data, the simplest L-way array representing the signals is a third-order tensor object $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times M}$ such that I_1 is the number of sensors, I_2 is the number of data points collected on each profile, and M is the number of multi-stream profiles or samples. Note that more articulated datasets may be generated by introducing additional modes, e.g., by adding a further mode to group together different families of sensors.

4.2.2 The UMLDA approach

Multilinear subspace feature extraction algorithms operating directly on tensor objects without changing their tensorial structure are emerging. Since LDA is a classical algorithm that has been very successful and applied widely in various applications, there have been

several variants of its multilinear extension proposed, named multilinear discriminant analysis (MLDA) in general. The projected tensors obtained from MLDA, however, are correlated contrary to classical LDA. To overcome this issue, Lu *et al.* (2009) proposed UMLDA, in which a TVP projection is used for projection. In this subsection, we review the UMLDA method proposed by Lu *et al.* (2009).

The derivation of the UMLDA algorithm follows the classic LDA derivation of minimizing the within-class distance and maximizing the between-class distance simultaneously, thus achieving maximum discrimination. A number of EMPs are solved one by one to maximize the discriminant criterion with an enforced zero-correlation constraint. To formulate the UMLDA problem, let $\left\{y_{m_p}, m = 1, \dots, M\right\}$ denote the pth projected scalar features, where M is the number of training samples and y_{m_p} is the projection of the mth sample $\mathcal{A}_m$ by the pth EMP $\left\{\mathbf{u}_p^{(1)^T}, \mathbf{u}_p^{(2)^T}\right\}$: $y_{m_p} = \mathcal{A}_m \times_1 \mathbf{u}_p^{(1)^T} \times_2 \mathbf{u}_p^{(2)^T}$. Adapting the classical Fisher Discriminant Criterion (FDC) to scalar sample, the between-class scatter $S_{B_p}^{\mathbf{y}}$ and the within-class scatter $S_{W_p}^{\mathbf{y}}$ are

$$S_{B_p}^{\mathbf{y}} = \sum_{c=1}^{C} N_c \left(\bar{y}_{c_p} - \bar{y}_p\right)^2, \; S_{W_p}^{\mathbf{y}} = \sum_{m=1}^{M} \left(y_{m_p} - \bar{y}_{c_{m_p}}\right)^2, \tag{4.1}$$

where C is the number of classes, N_c is the number of samples for class c, c_m is the class label for the mth training sample, $\bar{y}_p = (1/M)\sum_m y_{m_p} = 0$ assuming the training samples are zero-mean, and $\bar{y}_{c_p} = (1/N_c)\sum_{m,c_m=c} y_{m_p}$. Thus, the FDC for the pth scalar samples is $F_p^{\mathbf{y}} = S_{B_p}^{\mathbf{y}}/S_{W_p}^{\mathbf{y}}$. Let $\mathbf{g}_p$ denote the pth coordinate vector, with its mth component $\mathbf{g}_p(m) = y_{m_p}$. The objective of UMLDA is to determine a set of P EMPs that maximize the

scatter ratio while producing uncorrelated features. The mathematical formulation of UMLDA can be written as

$$\left\{\mathbf{u}_p^{(1)^T}, \mathbf{u}_p^{(2)^T}\right\} = argmax\, F_p^{\mathbf{y}} \tag{4.2}$$

$$\text{subject to} \qquad \left\|\mathbf{u}_p^{(1)}\right\| = 1, \ \left\|\mathbf{u}_p^{(2)}\right\| = 1,$$

$$\frac{\mathbf{g}_p^T \mathbf{g}_q}{\|\mathbf{g}_p\|\|\mathbf{g}_q\|} = \delta_{pq}, \qquad p, q = 1, \ldots, P$$

where $\delta_{pq} = 1$ for $p = q$ and $\delta_{pq} = 0$ otherwise.

The solution to this problem is provided by using the successive determination approach. The P EMPs $\left\{\mathbf{u}_p^{(1)^T}, \mathbf{u}_p^{(2)^T}\right\}_{p=1}^{P}$ are determined sequentially in P steps, with the pth step obtaining the pth EMP. Specifically, the first EMP $\left\{\mathbf{u}_1^{(1)^T}, \mathbf{u}_1^{(2)^T}\right\}$ is determined by maximizing $F_1^{\mathbf{y}}$ without any constraint; then the second EMP $\left\{\mathbf{u}_2^{(1)^T}, \mathbf{u}_2^{(2)^T}\right\}$ is determined by maximizing $F_2^{\mathbf{y}}$ subject to the constraint that $\mathbf{g}_2^T \mathbf{g}_1 = 0$; the third EMP $\left\{\mathbf{u}_3^{(1)^T}, \mathbf{u}_3^{(2)^T}\right\}$ is determined by maximizing $F_3^{\mathbf{y}}$ subject to the constraint that $\mathbf{g}_3^T \mathbf{g}_1 = 0$ and $\mathbf{g}_3^T \mathbf{g}_2 = 0$; etc. To solve for the pth EMP $\left\{\mathbf{u}_p^{(1)^T}, \mathbf{u}_p^{(2)^T}\right\}$, there are two sets of parameters corresponding to the 2 projection vectors to be determined, $\mathbf{u}_p^{(1)}$ and $\mathbf{u}_p^{(2)}$, one in each mode. Although it is most desirable to determine these 2 sets of parameters simultaneously so that $F_p^{\mathbf{y}}$ is maximized with respect to the zero-correlation constraint, this is a rather complicated nonlinear problem without an existing optimal solution. The only exception is for the case when we deal with one-stream profile, or one sensor, in which the UMLDA boils down to the classical linear case where only one projection vector is to be solved. Therefore, an approximate iterative approach that considers one mode at a time is used to determine each

EMP. Detailed information about this approach can be found in Jin *et al.* (2001) and Lu *et al.* (2009).

The implementation of UMLDA given by Lu *et al.* (2009) for the purpose of face recognition introduces a regularization parameter γ (R-UMLDA). To solve for $\mathbf{u}_p^{(l^*)}$ in the l^*-mode, assuming that $\left\{\mathbf{u}_p^{(l)}, l \neq l^*\right\}$ is given, the tensor samples are projected in these $(L-1$ modes$)$ $\{l \neq l^*\}$ to obtain vectors $\tilde{\mathbf{y}}_{m_p}^{(l^*)} = \mathcal{A}_m \times_{l=1,l\neq l^*}^{L} \left\{\mathbf{u}_p^{(l)^T}, l = 1, \ldots, l^*-1, l^*+1, \ldots, L\right\}_{p=1}^{P}$. The regularized within-class scatter matrix $\tilde{S}_{W_p}^{(l^*)}$ is defined as

$$\tilde{S}_{W_p}^{(l^*)} = \sum_{m=1}^{M} \left(\tilde{\mathbf{y}}_{m_\mathbf{p}}^{(l^*)} - \bar{\tilde{\mathbf{y}}}_{c_{m_\mathbf{p}}}^{(l^*)}\right)\left(\tilde{\mathbf{y}}_{m_\mathbf{p}}^{(l^*)} - \bar{\tilde{\mathbf{y}}}_{c_{m_\mathbf{p}}}^{(l^*)}\right)^T + \gamma \cdot \lambda_{max}\left(\check{S}_W^{(l^*)}\right) \cdot \mathbf{I}_{I_{l^*}} \qquad (4.3)$$

where $\gamma \geq 0$ is a regularization parameter, $\mathbf{I}_{I_{l^*}}$ is an identity matrix of size $I_{l^*} \times I_{l^*}$, and $\lambda_{max}\left(\check{S}_W^{(l^*)}\right)$ is the maximum eigenvalue of $\check{S}_W^{(l^*)}$, which is the within-class scatter matrix for the l-mode vectors of the training samples, defined as $\check{S}_W^{(l^*)} = \sum_{m=1}^{M}\left(\mathbf{A}_{m(l^*)} - \bar{\mathbf{A}}_{c_m(l^*)}\right)\left(\mathbf{A}_{m(l^*)} - \bar{\mathbf{A}}_{c_m(l^*)}\right)^T$, where $\bar{\mathbf{A}}_{c_m(l^*)}$ is the l^*-mode unfolded matrix of the class mean tensor $\bar{\mathcal{A}}_c = (1/N_c)\sum_{m,c_m=c}\mathcal{A}_m$.

The purpose of introducing the regularization parameter is to improve the UMLDA algorithm under small sample size scenario, where the dimensionality of the input data is high, but the number of training samples for some classes is too small to represent the true characteristics of their classes. This is a common case in small scale production like prototyping or personalized production. This scenario may also occur when a certain type of fault exists but rare, and that the data from that fault case is limited. If the number of training samples is too small, the iterations tend to minimize the within-class scatter towards zero in order to maximize the scatter ratio. Having a regularization parameter in the within-class

scatter ensures that during the iteration, less focus is put on shrinking the within-class scatter. The basic UMLDA is obtained by setting $\gamma = 0$.

Based on the observations in Lu *et al*. (2009), the sensitivity of the R-UMLDA to initialization and regularization suggests that R-UMLDA is not a very stable feature extractor and it is good for ensemble-based learning. Regularized UMLDA with aggregation (R-UMLDA-A) is hence introduced to aggregate several differently initialized and regularized UMLDA feature extractors to achieve better classification results. To focus on feature extraction performance, simple aggregation at the matching score level using the nearest-neighbor distance is implemented in R-UMLDA-A. Let A denote the number of R-UMLDA feature extractors to be aggregated. To classify a test sample $\mathcal{A}$, it is projected to A feature vectors $\{\mathbf{y}_{(a)}\}_{a=1,\ldots,A}$ using the A TVPs first. Next, for the ath R-UMLDA feature extractor, the nearest-neighbor distance of the test sample $\mathcal{A}$ to each candidate class c is

$$d(\mathcal{A}, c, a) = \min_{m, c_m = c} \left\| \mathbf{y}_{(a)} - \mathbf{y}_{m_{(a)}} \right\|. \tag{4.4}$$

$d(\mathcal{A}, c, a)$ is then scaled to the interval [0, 1] as $\tilde{d}(\mathcal{A}, c, a) = \frac{d(\mathcal{A},c,a) - \min_c d(\mathcal{A},c,a)}{\max_c d(\mathcal{A},c,a) - \min_c d(\mathcal{A},c,a)}$. The aggregated nearest-neighbor distance is obtained using the simple sum rule:

$$d(\mathcal{A}, c) = \sum_{a=1}^{A} \tilde{d}(\mathcal{A}, c, a). \tag{4.5}$$

Therefore, the test sample $\mathcal{A}$ is assigned the label $c^* = \arg\min_c d(\mathcal{A}, c)$.

4.2.3 The VLDA approach

VLDA is a generalization of LDA to tensor data, which applies the regular LDA to a tensor object reshaped into a vector. In the frame of multi-stream profile data, the third-order tensor object $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times M}$ representing the signals is unfolded slice by slice; the slices are

then rearranged into a large two-dimensional matrix $\mathbf{A} \in \mathbb{R}^{I_1 I_2 \times M}$, where I_1 is the number of sensors, I_2 is the number of data points collected on each profile, and M is the number of samples. The classical LDA is then performed on matrix $\mathbf{A}$. What we seek is a transformation matrix $\mathbf{W}$ that maximizes the ratio of the between-class scatter to the within-class scatter

$$\mathbf{W} = argmax\, J(\mathbf{W}) = argmax\, \frac{|\mathbf{W}^T \boldsymbol{S_B} \mathbf{W}|}{|\mathbf{W}^T \boldsymbol{S_W} \mathbf{W}|} \tag{4.6}$$

$$\text{subject to} \quad \|\mathbf{w}_i\| = 1, i = 1, \ldots, c-1$$

where $\boldsymbol{S_B}$ and $\boldsymbol{S_W}$ are the between-class scatter and within-class scatter, respectively, c is the number of classes. The transformed signal samples can be obtained by $\mathbf{y} = \mathbf{W}^T \mathbf{A}$. More details on the calculation of $\boldsymbol{S_B}$ and $\boldsymbol{S_W}$ using Fisher linear discriminant can be found in Duda *et al.* (2012).

4.3 Performance Comparison in Simulations

In this section, the performances of the UMLDA and VLDA methodologies are evaluated and compared by means of Monte Carlo simulations. The multi-stream signals in simulation are generated in a similar manner as in Grasso *et al.* (2014): a four-stream profile dataset is generated based on three benchmark signals proposed by Donoho and Johnstone (1994). These signals have been used by different authors to test wavelet-based algorithms, but also in the frame of statistical models and machine learning literature (e.g. see Fan *et al.*, 2012; Ko *et al.*, 2009; Koo and Kil, 2008). The complex pattern features in the benchmark signals make it difficult for profile modeling using a parametric approach. Figure 4.3 illustrates the three benchmark signals: 'blocks', 'heavysine', and 'bumps', and they are denoted as $\boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{x}_3$, respectively.

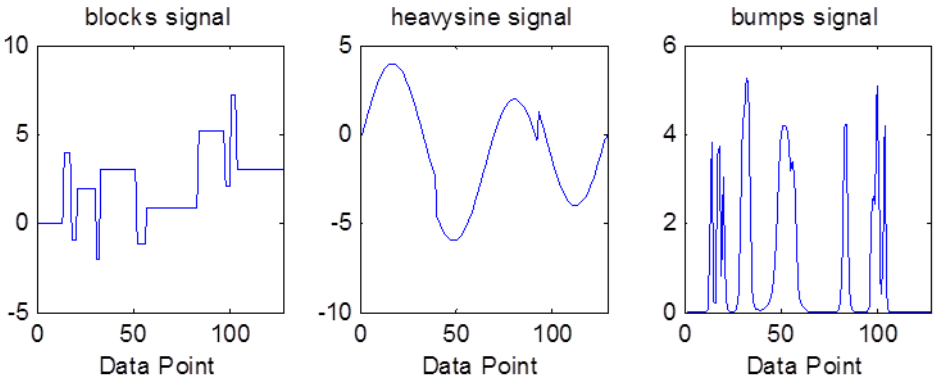


Figure 4.3. Benchmark signals 'blocks', 'heavysine', and 'bumps'

Let $\chi \in \mathbb{R}^{N \times K \times M}$ denote the third-order tensor object that represents the four-stream profile dataset, where $N = 4$ is the number of streams or sensors, $K = 128$ is the number of data points for all the signals, and M is the number of samples. χ is generated to contain different types of correlation structures: linear correlation (e.g., $\chi_{1,\cdot,m}$ and $\boldsymbol{x_1}$, $\chi_{2,\cdot,m}$ and $\boldsymbol{x_3}$, etc.), curvilinear correlation (e.g., $\chi_{2,\cdot,m}$ and $\boldsymbol{x_1}$, $\chi_{3,\cdot,m}$ and $\boldsymbol{x_2}$, etc.), and no correlation (e.g., $\chi_{3,\cdot,m}$ and $\boldsymbol{x_1}$, $\chi_{4,\cdot,m}$ and $\boldsymbol{x_3}$, etc.). χ is defined as follows:

$$\begin{aligned} \chi_{1,\cdot,m} &= b_{1,m}\boldsymbol{x_1} + b_{2,m}\boldsymbol{x_2} + \varepsilon_{1,m} \\ \chi_{2,\cdot,m} &= b_{3,m}\boldsymbol{x_1^2} + b_{4,m}\boldsymbol{x_3} + \varepsilon_{2,m} \\ \chi_{3,\cdot,m} &= b_{5,m}\boldsymbol{x_2^2} + b_{6,m}\boldsymbol{x_3^2} + \varepsilon_{3,m} \\ \chi_{4,\cdot,m} &= b_{7,m}\boldsymbol{x_1 x_2} + \varepsilon_{4,m} \end{aligned} \quad (m = 1, \dots, M) \tag{4.7}$$

where $\varepsilon_{n,m} \sim N(0, 0.5^2)$ is the random noise and $\boldsymbol{b}_m = [b_{1,m}, \dots, b_{7,m}]^T \sim MVN(\boldsymbol{\mu}_b, \Sigma_b)$ is the model parameter vector, $n = 1, \dots, 4, m = 1, \dots, M$. Similar to the dataset used in Grasso *et al.* (2014), the following settings are used to generate the dataset: $\boldsymbol{\mu}_b = [0.2, 1, 1.5, 0.5, 1, 0.7, 0.8]^T$, $\Sigma_b = diag(\sigma_{b_1}^2, \dots, \sigma_{b_7}^2) = diag(0.08, 0.015, 0.05, 0.01, 0.09, 0.03, 0.06)$. Figure 4.4 shows 100 in-control profile samples generated in this setting. As can be seen in Eq. (4.7), the four streams of signals are not independent, but the correlation structure is complex for profile modeling.

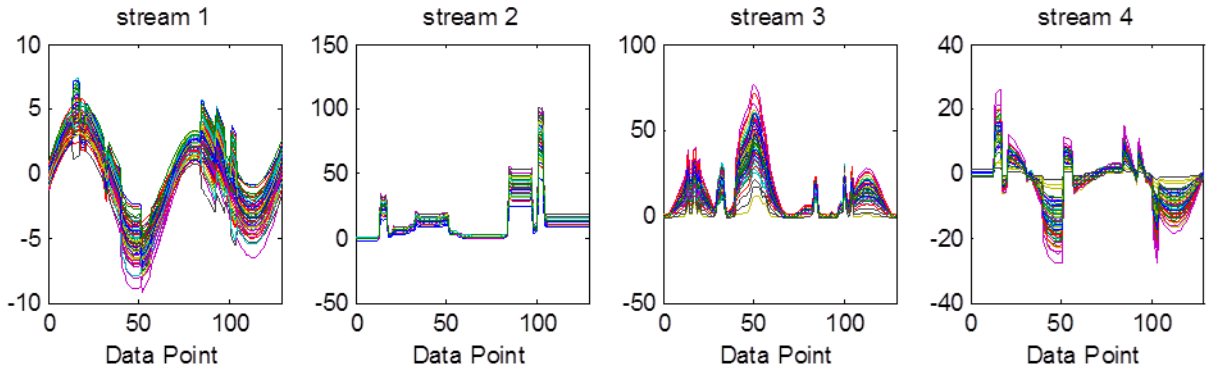


Figure 4.4. 100 in-control profile samples

Different out-of-control scenarios are generated to simulate different kinds of deviations from the natural multi-stream pattern. Each out-of-control scenario is associated with an assignable cause. In the context of ultrasonic metal welding (and many other manufacturing processes as well), these assignable causes represent different types of faults, e.g., mislocated weld, sheet metal distortion, surface contamination, etc. In this chapter, we assume multiple faults do not occur simultaneously on one part, i.e., a single part has no more than one fault. The following out-of-control scenarios are considered:

Scenario (a): Mean shift of the reference signal

$$\boldsymbol{x}_u \rightarrow \boldsymbol{x}_u + \delta_a \mathbf{1}_{K\times 1} \quad (u = 1,2,3) \tag{4.8}$$

where $\delta_a \in \{0.01, 0.025, 0.05, 0.075, 0.1\}\sigma_{x_u}$ is the magnitude of the shift, σ_{x_u} is the standard deviation of $\boldsymbol{x}_u$ reference signal, $u = 1,2,3$, and $\mathbf{1}_{K\times 1}$ is a column vector of ones.

Scenario (b): Superimposition of a sinusoid term on the reference signal

$$\boldsymbol{x}_u \rightarrow \boldsymbol{x}_u + \delta_b \boldsymbol{y}_s \quad (u = 1,2,3) \tag{4.9}$$

where $\delta_b \in \{0.025, 0.05, 0.075, 0.1, 0.125\}\sigma_{x_u}$, and $\boldsymbol{y}_s$ is the sine function over the domain $[0, K]$, with period K and peak-to-peak amplitude equal to 1, $u = 1,2,3$.

Scenario (c): Standard deviation increase of the error term

$$\sigma_{\varepsilon_{n.m}} \rightarrow \delta_c \sigma_{\varepsilon_{n.m}} \ (n = 1,2,3,4) \tag{4.10}$$

where $\delta_c \in \{1.1, 1.5, 2, 2.5, 3\}$, and $\sigma_{\varepsilon_{n.m}}$ is the standard deviation of the error term $\varepsilon_{n.m}$.

Scenario (d): Mean shift of the model parameter

$$\mu_{b_w} \rightarrow \mu_{b_w} + \delta_d \ (w = 1, \dots, 7) \tag{4.11}$$

where $\delta_d \in \{1,2,3,4,5\}\sigma_{b_w}$, μ_{b_w} and σ_{b_w} are the mean value and standard deviation of the $w$the model parameter $b_w, w = 1, \dots, 7$.

Scenario (e): Standard deviation increase of the model parameter

$$\sigma_{b_w} \rightarrow \delta_e \sigma_{b_w} \ (w = 1, \dots, 7) \tag{4.12}$$

where $\delta_e \in \{1.5, 2, 2.5, 3, 4\}$.

Scenario (f): Gradual mean shift of the reference signal

$$\boldsymbol{x}_u \rightarrow \boldsymbol{x}_u + \delta_f \mathbf{1}_{K\times 1} \ (u = 1,2,3) \tag{4.13}$$

where δ_f is the magnitude of the shift and $\mathbf{1}_{K\times 1}$ is a column vector of ones. This scenario is introduced to represent the effects of tool wear on profile data. As tool wear develops, the reference signal of the $(m + 1)$th sample would have a larger mean shift than that of the mth sample. Considering the severeness of tool wear, let $\delta_{f1} \in [0.01, 0.05]\sigma_{x_u}$ represent the deviations caused by a lightly worn tool, $\delta_{f2} \in (0.05, 0.1]\sigma_{x_u}$ represent those caused by a tool with intermediate level of worn, and $\delta_{f3} \in (0.1, 0.15]\sigma_{x_u}$ represent the deviations caused by a severely worn tool, $u = 1,2,3$.

4.3.1 Methods in comparison

The general framework of profile monitoring and fault diagnosis using multi-stream signals is illustrated in Figure 4.5. For multilinear methods like UMLDA, the multi-stream

signals can be directly represented in a tensor object, and then the tensor is normalized so that the training samples are in the same dimension and zero-mean. For linear methods like VLDA, the multi-stream signals need to be vectorized to a matrix, and then followed by normalization. Feature extraction method, e.g., UMLDA or VLDA, then produces vector features that can be fed into standard classifiers for classification. The output is a tensor class labels which represents 'normal' or some fault type.

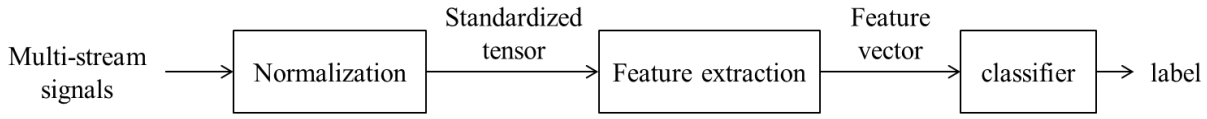


Figure 4.5. Framework of profile monitoring and fault diagnosis using multi-stream signals

Performance comparison is conducted in two levels: (1) feature extraction performance, and (2) classification performance. To compare feature extraction performance, we use the following four multilinear and three linear methods to extract features: regularized UMLDA (R-UMLDA), regularized UMLDA with aggregation (R-UMLDA-A), UMPCA, MPCA, VLDA, uncorrelated LDA (V-ULDA), and regularized LDA (V-RLDA). The feature vectors obtained are then fed into the nearest-neighbour classifier (NNC) with the Euclidean distance measure for classification.

In R-UMLDA, the regularization parameter γ is empirically set to $\gamma = 0.001$. If we let Q denote the number of training samples per class, then intuitively, stronger regularization is more desirable for a smaller Q, and weaker regularization is recommended for a larger Q. Since the tensor object $\chi \in \mathbb{R}^{4\times128\times M}$, one R-UMLDA will extract up to 4 features. In R-UMLDA-A, up to $A = 20$ differently initialized and regularized UMLDA feature extractors are combined with each producing up to 4 features, resulting in a total of 80 features. The regularization parameter ranges from 10^{-7} to 10^{-2}.

UMPCA and MPCA are unsupervised multilinear methods that seek a set of projections to maximize the variability captured by the projected tensor. UMPCA will produce up to 4 features which are uncorrelated, while MPCA will produce as many as approximately 80 features which are correlated in order to capture at least 99% of the variation in each mode. Details on the theoretical development of UMPCA and MPCA can be found in Lu *et al.* (2008, 2009).

In addition to VLDA, two more linear methods are included in comparison, V-ULDA and V-RLDA. V-ULDA and V-RLDA improve LDA on undersampled problems and small sample size problems, respectively. Each method will project to up to $C-1$ features with C being the number of classes. Details on the theoretical development of ULDA and RLDA can be found in Ye (2005) and Ye *et al.* (2006).

In order to further improve classification performance, we feed the features extracted by multiple R-UMLDA extractors into random space method, and compare its performances with the R-UMLDA-A which adopts the simple nearest-neighbour aggregation. Since classification is not the main focus of this chapter, we will not discuss the ensemble learning methods in detail. Readers interested in random space method and ensemble learning are referred to Ho (1998) and Hastie *et al.* (2008).

4.3.2 Simulation results

This subsection discusses simulation results in three main cases A, B and C.

4.3.2.1 Case A

Generate a total of 1200 profile samples with 200 samples in each class: in-control and five out-of-control scenarios (a) – (e). All 1200 samples in $C=6$ classes are plotted in Figure 4.6. Specifically, the five out-of-control scenarios are:

(a) mean shift of the 'block' reference signals: $\boldsymbol{x_1} \rightarrow \boldsymbol{x_1} + 0.1\sigma_{x_1}$, resulting in $\tilde{\chi}_{1,\cdot,m} = b_{1,m}(\boldsymbol{x_1} + 0.1\sigma_{x_1}\mathbf{1}_{K\times 1}) + b_{2,m}\boldsymbol{x_2} + \varepsilon_{1,m}$, $\tilde{\chi}_{2,\cdot,m} = b_{3,m}\left(\boldsymbol{x_1} + 0.1\sigma_{x_1}\mathbf{1}_{K\times 1}\right)^2 + b_{4,m}\boldsymbol{x_3} + \varepsilon_{2,m}$, and $\tilde{\chi}_{4,\cdot,m} = b_{7,m}(\boldsymbol{x_1} + 0.1\sigma_{x_1}\mathbf{1}_{K\times 1}) + \varepsilon_{4,m}$;

(b) superimposition of a sinusoid term on the 'block' reference signal: $\boldsymbol{x_1} \rightarrow \boldsymbol{x_1} + 0.1\sigma_{x_1}\boldsymbol{y_s}$, $\boldsymbol{y_s}$ is a sine function, resulting in $\tilde{\chi}_{1,\cdot,m} = b_{1,m}(\boldsymbol{x_1} + 0.1\sigma_{x_1}\boldsymbol{y_s}) + b_{2,m}\boldsymbol{x_2} + \varepsilon_{1,m}$, $\tilde{\chi}_{2,\cdot,m} = b_{3,m}\left(\boldsymbol{x_1} + 0.1\sigma_{x_1}\boldsymbol{y_s}\right)^2 + b_{4,m}\boldsymbol{x_3} + \varepsilon_{2,m}$, and $\tilde{\chi}_{4,\cdot,m} = b_{7,m}(\boldsymbol{x_1} + 0.1\sigma_{x_1}\boldsymbol{y_s}) + \varepsilon_{4,m}$;

(c) standard deviation increase of the error term e_1: $\sigma_{\varepsilon_{1.m}} \rightarrow 3\sigma_{\varepsilon_{1.m}}$, resulting in $\tilde{\chi}_{1,\cdot,m} = b_{1,m}\boldsymbol{x_1} + b_{2,m}\boldsymbol{x_2} + \tilde{\varepsilon}_{1,m}$, where $\tilde{\varepsilon}_{1,m} \sim N(0, (3 \times 0.5)^2)$;

(d) mean shift of the model parameter b_1: $\mu_{b_1} \rightarrow \mu_{b_1} + 5\sigma_{b_1}$, resulting in $\tilde{\chi}_{1,\cdot,m} = \tilde{b}_{1,m}\boldsymbol{x_1} + b_{2,m}\boldsymbol{x_2} + \varepsilon_{1,m}$, where $\tilde{b}_{1,m} \sim N\left(\mu_{b_1} + 5\sigma_{b_1}, \sigma_{b_1}^2\right)$;

(e) standard deviation increase of the model parameter b_1: $\sigma_{b_1} \rightarrow 4\sigma_{b_1}$, resulting in $\tilde{\chi}_{1,\cdot,m} = \tilde{b}_{1,m}\boldsymbol{x_1} + b_{2,m}\boldsymbol{x_2} + \varepsilon_{1,m}$, where $\tilde{b}_{1,m} \sim N\left(\mu_{b_1}, \left(4\sigma_{b_1}\right)^2\right)$.

Of the five scenarios above, all profiles in streams 1, 2, and 4 are affected in (a) and (b), while in (c), (d), and (e), only the profiles in stream 1 present out-of-control patterns.

Since a large amount of the $\tilde{\varepsilon}_{1,m}$'s generated by $\tilde{\varepsilon}_{1,m} \sim N\left(0, \left(3\sigma_{\varepsilon_{1.m}}\right)^2\right)$ in fault (c) would overlap with the $\varepsilon_{1,m}$'s generated by $\varepsilon_{1,m} \sim N\left(0, \sigma_{\varepsilon_{1,m}}^2\right)$ in the in-control class, and that the $\tilde{b}_{1,m}$'s generated by $\tilde{b}_{1,m} \sim N\left(\mu_{b_1}, \left(4\sigma_{b_1}\right)^2\right)$ in fault (e) would greatly overlap with the $b_{1,m}$'s generated by $b_{1,m} \sim N\left(\mu_{b_1}, \sigma_{b_1}^2\right)$ in the in-control class, faults (c) and (e) would be very difficult to separate from the in-control class.

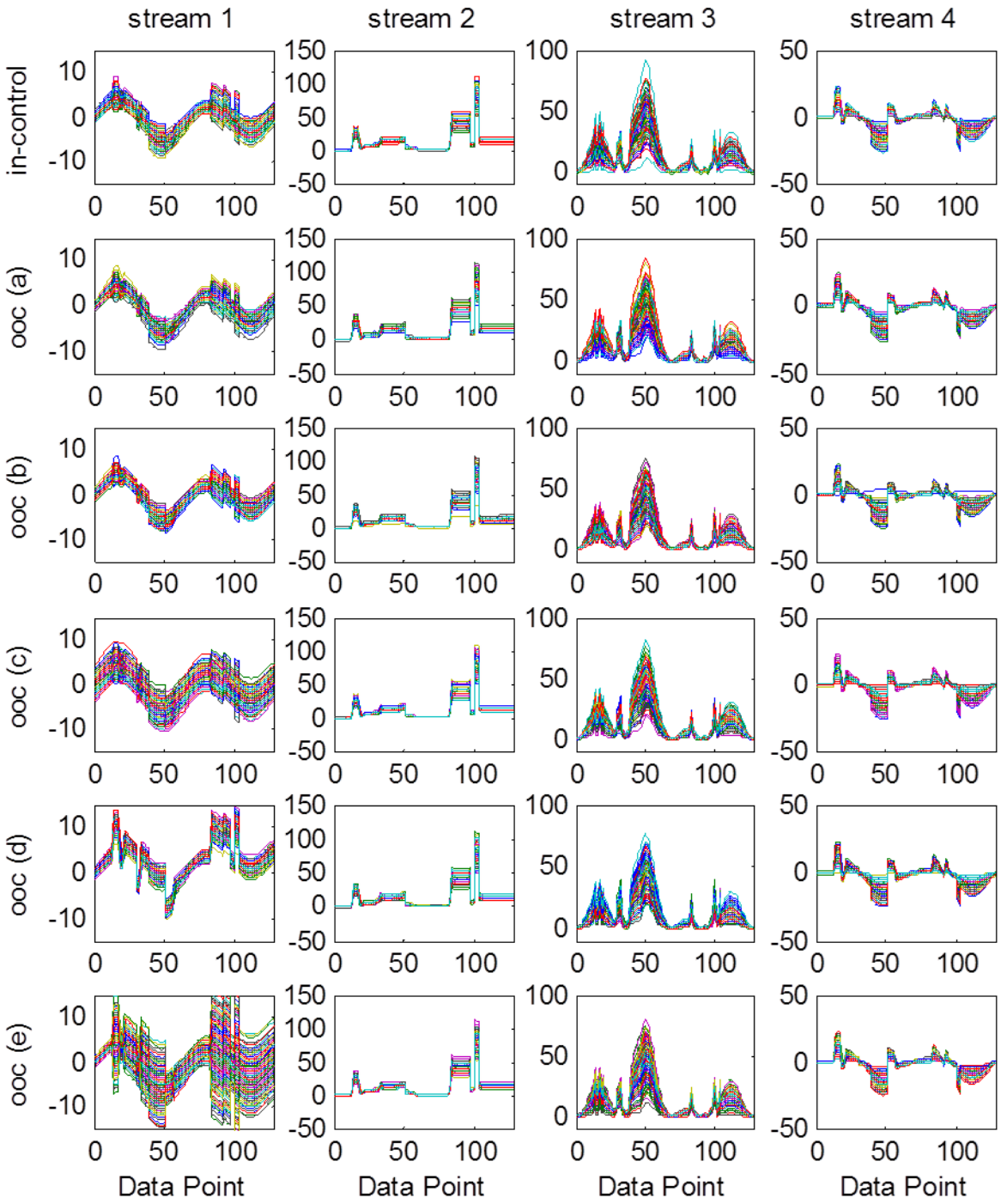


Figure 4.6. Case A dataset: 1200 samples in 6 classes

Half of these 1200 samples are considered as the training dataset. Figure 4.7 plots the normalized training data in the 6 classes in 4 streams. Normalization is performed by taking away the grand mean of all training samples from the original data.

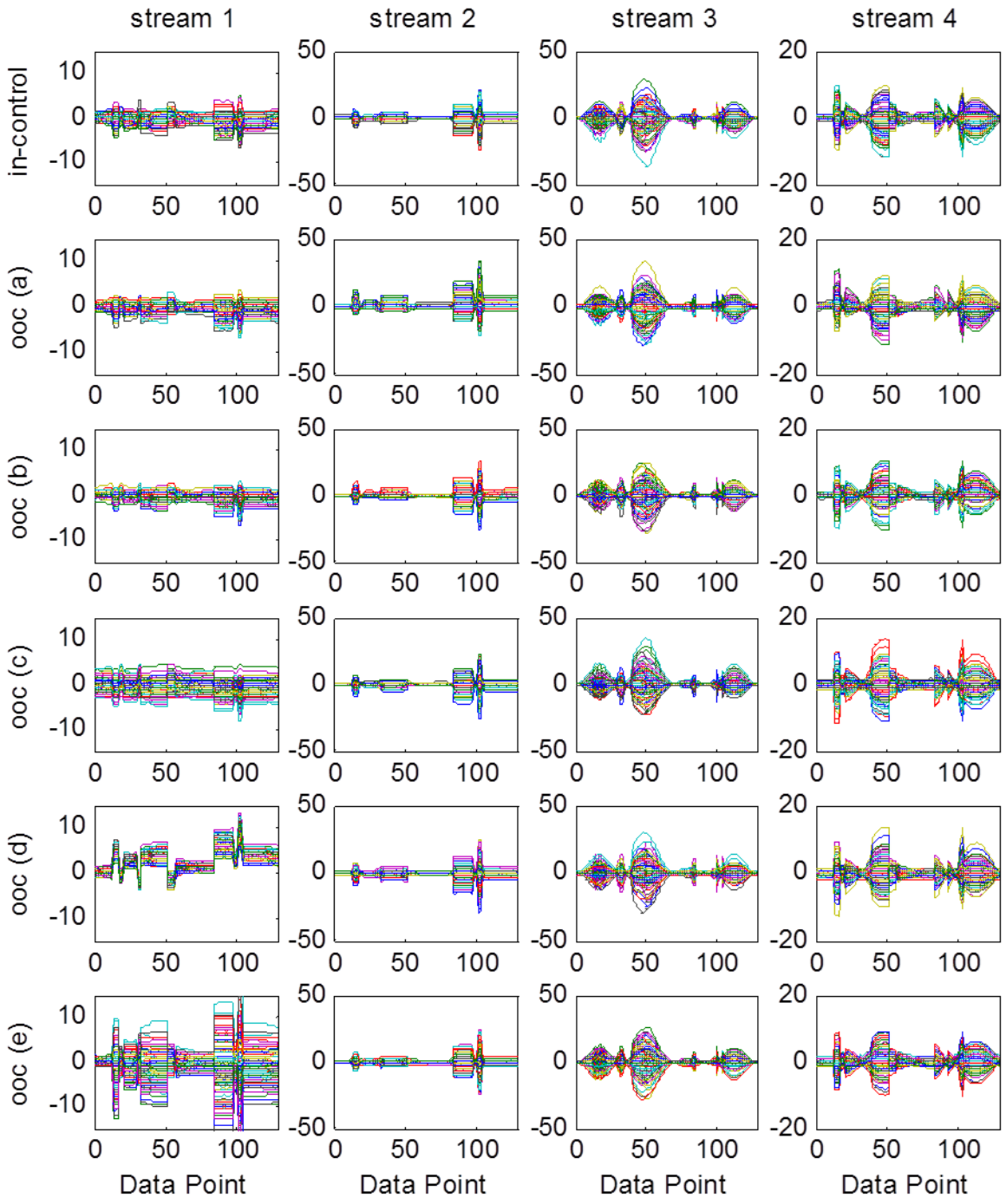


Figure 4.7. Case A dataset: normalized training data (600 samples in 6 classes)

Using the procedures described in Section 4.2 and Section 4.3.1, regularized UMLDA is applied to the generated data. In UMLDA, the eigentensors corresponding to the $p$the EMP, $\mathbf{U}_p \in \mathbb{R}^{4\times128}$, $p = 1,2,3,4$ are obtained by $\mathbf{u}_p^{(1)} \circ \mathbf{u}_p^{(2)}$, where $\mathbf{u}_p^{(1)} \in \mathbb{R}^{4\times1}$ and $\mathbf{u}_p^{(2)} \in$

$\mathbb{R}^{128\times1}$. Figure 4.8 shows $\mathbf{U}_p$, $p = 1,2,3,4$, obtained from the training dataset in a single simulation run of Case A. Each row of $\mathbf{U}_p$ corresponds to one signal stream. As can be seen from Figure 4.8, the eigenvectors corresponding to the first EMP show an efficient discrimination against streams 1 and 4, whereas those corresponding to the second EMP show a strong discrimination against stream 2. The eigenvectors corresponding to the third and fourth EMPs show weak discriminations against stream 4, whereas limited useful information is extracted from stream 3 for discriminant analysis. These results are exactly compatible with the data generation model, thus implying that R-UMLDA can effectively extract information for discriminant analysis about multi-stream profiles.

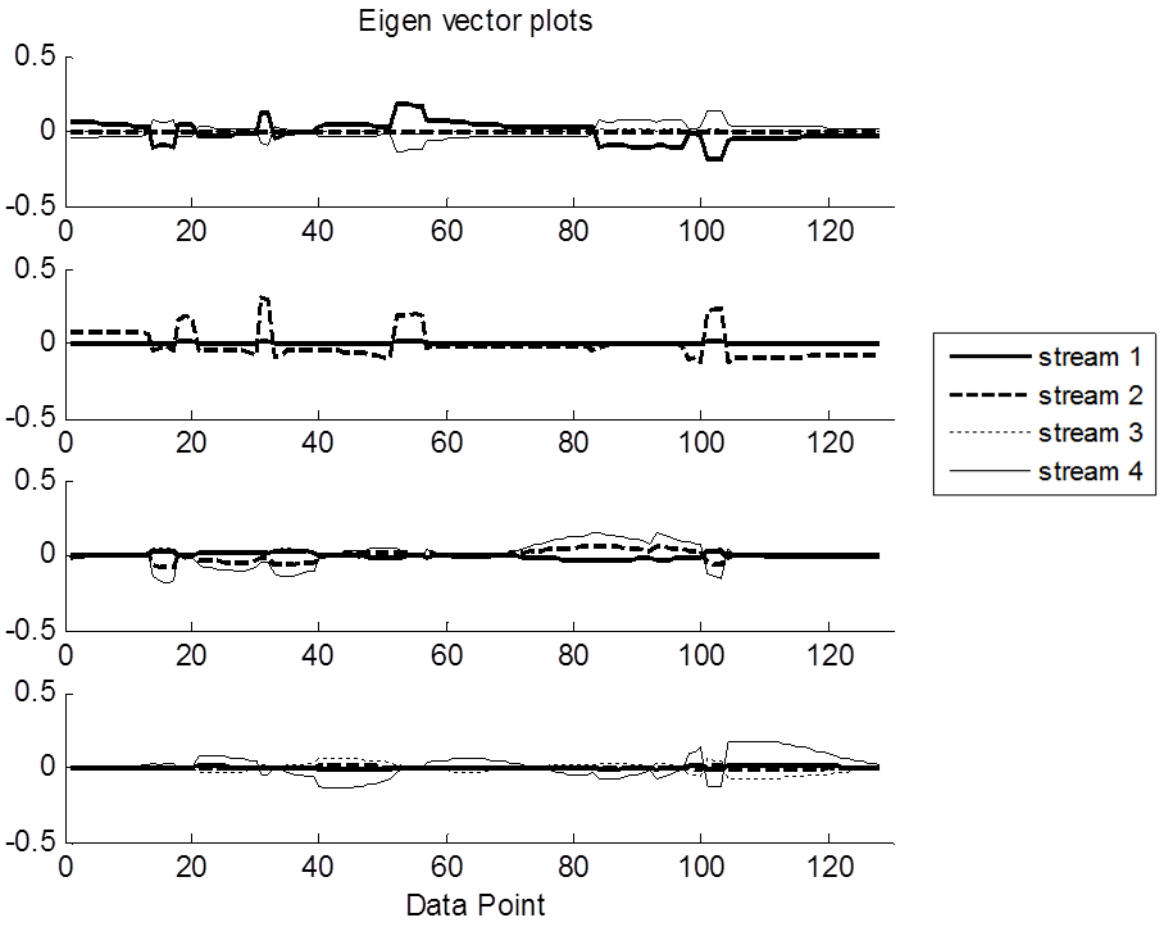


Figure 4.8. Eigentensors from R-UMLDA in simulation Case A

Using the first p EMPs ($p = 1,2,3,4$), multi-stream profiles can be projected to p uncorrelated features, which are then fed into the nearest-neighbour classifier (NNC). The classification performance in the testing dataset is shown in Figure 4.9 and Table 4.1. Figure 4.9 plots the following detailed results against the number of features used:

- correct classification rate: $\sum_{m=1}^{M_{test}} I(\hat{c}_m = c_m) / M_{test}$, where $\hat{c}_m$ is the predicted class for sample m, c_m is the true class, and M_{test} is the number of testing samples.
- correct passing rate: $\sum_{m=1}^{M_{test}} I(\hat{c}_m = 0 | c_m = 0) / M_{test}$, where '0' indicates the 'normal' class.
- correct detection rate: $\sum_{m=1}^{M_{test}} I(\hat{c}_m > 0 | c_m > 0) / M_{test}$, where $c > 0$ indicates a fault class.
- true fault classification rate: $\sum_{m=1}^{M_{test}} I(\hat{c}_m = c_m | c_m > 0) / M_{test}$.
- rate of true detection but wrong fault classification: $\sum_{m=1}^{M_{test}} I(\hat{c}_m \neq c_m | \hat{c}_m > 0, c_m > 0) / M_{test}$.

As can be seen in Figure 4.9, the first two features extracted by R-UMLDA are the most powerful features in classification. Adding the third and fourth features slightly helps improve the correct classification rate.

More detailed classification results with respect to the number of features fed into the classifier are shown in the confusion matrices in Table 4.1. From Table 4.1, we can easily observe an improvement in classification accuracy when two or more EMPs are used instead of using only the first one. We also notice that when two or more features are used, most of the classification errors come from separating the in-control class, fault (c), and fault (e) from each other. This observation is exactly compatible with the data generation model, based on which we have expected that faults (c) and (e) are the most difficult classes to separate from the in-control class.

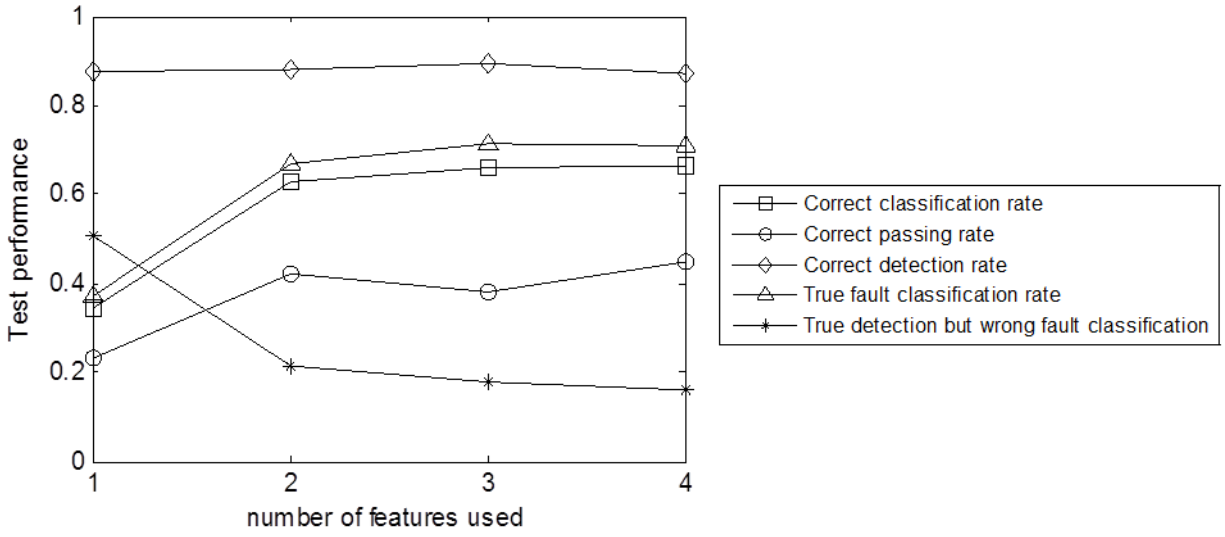


Figure 4.9. Classification performance of NNC for R-UMLDA features in Case A testing dataset

Table 4.1. Confusion matrix of NNC for R-UMLDA features in Case A testing dataset

	Classified as							**Classified as**					
	Normal	Fault (a)	Fault (b)	Fault (c)	Fault (d)	Fault (e)		Normal	Fault (a)	Fault (b)	Fault (c)	Fault (d)	Fault (e)
Actual	***1 feature***						Actual	***2 features***					
Normal	23	23	23	25	1	5	Normal	42	0	5	41	1	11
Fault (a)	18	25	25	26	0	6	Fault (a)	0	98	1	0	0	1
Fault (b)	18	27	26	24	0	5	Fault (b)	4	1	74	17	0	4
Fault (c)	17	29	21	23	1	9	Fault (c)	38	0	12	41	0	9
Fault (d)	1	0	0	1	77	21	Fault (d)	0	0	0	1	81	18
Fault (e)	8	12	12	19	15	34	Fault (e)	17	0	2	28	13	40
	Classified as							**Classified as**					
	Normal	Fault (a)	Fault (b)	Fault (c)	Fault (d)	Fault (e)		Normal	Fault (a)	Fault (b)	Fault (c)	Fault (d)	Fault (e)
Actual	***3 features***						Actual	***4 features***					
Normal	38	0	1	46	1	14	Normal	45	0	0	40	0	15
Fault (a)	0	98	1	0	0	1	Fault (a)	0	98	1	0	0	1
Fault (b)	0	1	98	0	0	1	Fault (b)	0	0	99	1	0	0
Fault (c)	38	0	2	44	1	15	Fault (c)	49	0	1	33	1	16
Fault (d)	0	0	0	2	79	19	Fault (d)	0	0	0	2	80	18
Fault (e)	16	0	0	29	17	38	Fault (e)	16	0	0	25	15	44

Applying the competitor methods described in Section 4.3.1, Figure 4.10 shows the classification performance in terms of correct classification rate of NNC for various feature extraction methods in Case A testing dataset. The plotted results are the average correct classification rates in 100 simulation runs. In Figure 4.10, the curves with triangle markers correspond to classification performance for UMPCA and MPCA features. It is obvious that these results are significantly worse than LDA-based methods, regardless of the number of features used. This agrees with our understanding of PCA-based feature extractors which do not make use of the class information and only seek projections to maximize the captured variability instead of class discrimination.

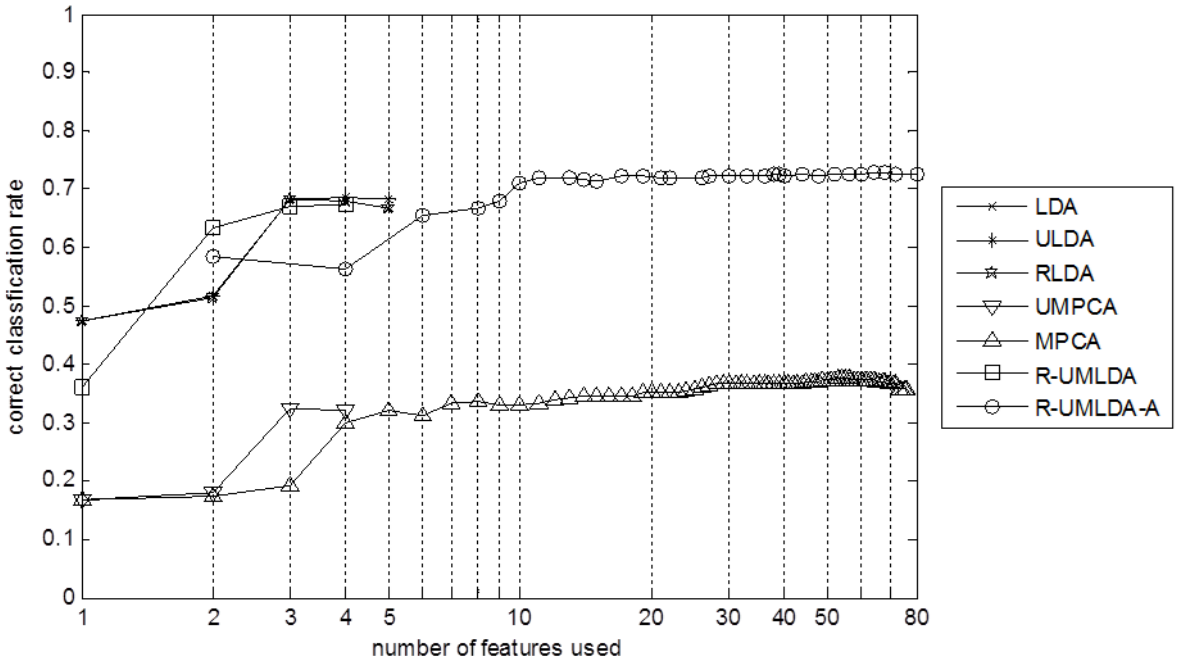


Figure 4.10. Classification performance of NNC for various feature extractors in Case A testing dataset

The curves with cross, star, and asterisk markers in Figure 4.10 correspond to vectorized LDA methods (including LDA, ULDA, and RLDA), whereas the curves with square and circle markers correspond to UMLDA methods. It can be seen from Figure 4.10 that the first

two features extracted by R-UMLDA are the most powerful features in classification. Beyond the first two features, the performance from R-UMLDA varies very slowly with an increased number of features used. The first three features extracted by vectorized LDA methods are also powerful, but the improvement from using the first two R-UMLDA features is not significant.

The best correct classification rate is achieved using R-UMLDA-A. Figure 4.10 shows that R-UMLDA-A outperforms all other algorithms. This demonstrates that aggregation is an effective procedure and there is indeed complementary discriminative information from differently regularized R-UMLDA feature extractors.

4.3.2.2 Case B

Generate a total of 800 profile samples with 200 samples in each of the following four classes: in-control and three out-of-control scenarios in (f), where three magnitudes of gradual mean shift are added to the 'block' reference signal to reflect machine tools with light worn, medium worn, and severe worn. Use half of these samples as the training dataset.

Table 4.2 presents the confusion matrix of the nearest-neighbor classifier for R-UMLDA (with $\gamma = 0.001$) features in Case B testing dataset. As more features are fed into the classifier, the classification accuracy improves significantly. We also observe that classification errors only occur in the following three scenarios: distinguishing between the normal class and fault (f-1) slight tool wear; distinguishing between fault (f-1) slight tool wear and fault (f-2) medium tool wear; and distinguishing between fault (f-2) slight tool wear and fault (f-3) severe tool wear.

Figure 4.11 shows the classification performance in terms of correct classification rate of NNC for various feature extraction methods in Case B testing dataset. The plotted results are the average correct classification rates in 100 simulation runs.

Table 4.2. Confusion matrix of NNC for R-UMLDA features in Case B testing dataset

	Classified as					Classified as			
	Normal	Fault (f-1)	Fault (f-2)	Fault (f-3)		Normal	Fault (f-1)	Fault (f-2)	Fault (f-3)
Actual		***1 feature***			Actual		***2 features***		
Normal	67	33	0	0	Normal	68	32	0	0
Fault (f-1)	1	64	35	0	Fault (f-1)	3	64	33	0
Fault (f-2)	0	0	44	56	Fault (f-2)	0	0	52	48
Fault (f-3)	0	0	0	100	Fault (f-3)	0	0	0	100
	Classified as					**Classified as**			
	Normal	Fault (f-1)	Fault (f-2)	Fault (f-3)		Normal	Fault (f-1)	Fault (f-2)	Fault (f-3)
Actual		***3 features***			Actual		***4 features***		
Normal	68	32	0	0	Normal	73	27	0	0
Fault (f-1)	3	74	23	0	Fault (f-1)	0	91	9	0
Fault (f-2)	0	0	76	24	Fault (f-2)	0	0	84	16
Fault (f-3)	0	0	0	100	Fault (f-3)	0	0	0	100

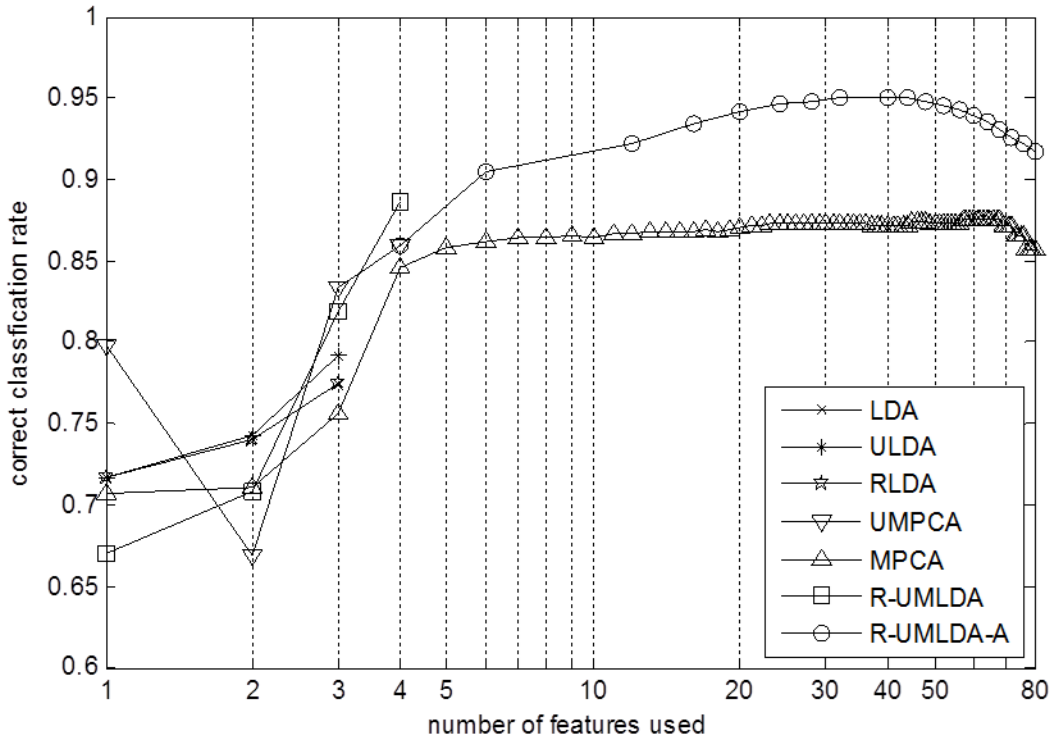


Figure 4.11. Classification performance of NNC for various feature extractors in Case B testing dataset

Similar to Case A, the features extracted by UMPCA and MPCA are the weakest features in classification. Although the first few (1~2) features extracted by VLDA, ULDA, and RLDA are the most discriminative, using three or more R-UMLDA features lead to notably enhanced results. Figure 4.11 also shows the significant improvement introduced by aggregation. In all, R-UMLDA and R-UMLDA-A outperform all other algorithms.

4.3.2.3 Case C

Generate a total of 1200 profile samples with 200 samples in each of the following six classes: in-control and five out-of-control scenarios: (d) mean shift of the model parameter b_1, (e) standard deviation increase of the model parameter b_1, and the three out-of-control scenarios in (f) as described in Case B. Use half of these samples as the training dataset.

Table 4.3. Confusion matrix of NNC for R-UMLDA features in Case C testing dataset

	Classified as							**Classified as**					
	Normal	Fault (f-1)	Fault (f-2)	Fault (f-3)	Fault (d)	Fault (e)		Normal	Fault (f-1)	Fault (f-2)	Fault (f-3)	Fault (d)	Fault (e)
Actual	***1 feature***						Actual	***2 features***					
Normal	29	30	13	5	4	19	Normal	47	28	2	0	2	21
Fault (f-1)	18	30	27	8	1	16	Fault (f-1)	9	31	46	9	0	5
Fault (f-2)	7	13	43	22	0	15	Fault (f-2)	1	5	46	45	0	3
Fault (f-3)	3	6	16	49	0	26	Fault (f-3)	0	0	12	88	0	0
Fault (d)	3	4	0	0	83	10	Fault (d)	1	1	0	0	81	17
Fault (e)	11	12	15	21	22	19	Fault (e)	21	19	3	0	17	40
	Classified as							**Classified as**					
	Normal	Fault (f-1)	Fault (f-2)	Fault (f-3)	Fault (d)	Fault (e)		Normal	Fault (f-1)	Fault (f-2)	Fault (f-3)	Fault (d)	Fault (e)
Actual	***3 features***						Actual	***4 features***					
Normal	51	21	0	0	1	27	Normal	52	20	0	0	3	25
Fault (f-1)	3	61	35	0	0	1	Fault (f-1)	0	88	12	0	0	0
Fault (f-2)	0	0	66	34	0	0	Fault (f-2)	0	0	90	10	0	0
Fault (f-3)	0	0	0	100	0	0	Fault (f-3)	0	0	0	100	0	0
Fault (d)	1	2	0	0	73	24	Fault (d)	3	0	0	0	78	19
Fault (e)	26	16	1	0	18	39	Fault (e)	29	11	0	0	19	41

Table 4.3 presents the confusion matrix of the nearest-neighbor classifier for R-UMLDA (with $\gamma = 0.001$) features in Case C testing dataset. As more features are fed into the classifier, the classification accuracy improves significantly.

We also observe that almost all classification errors occur in the following four scenarios: distinguishing between the normal class and fault (f-1), distinguishing between fault (f-1) and fault (f-2), distinguishing between fault (f-2) and fault (f-3), and separating fault (e) from the normal class. It is very difficult to separate fault (e) from the in-control class due to the fact that the $\tilde{b}_{1,m}$'s generated by $\tilde{b}_{1,m} \sim N\left(\mu_{b_1}, \left(4\sigma_{b_1}\right)^2\right)$ in fault (e) would greatly overlap with the $b_{1,m}$'s generated by $b_{1,m} \sim N\left(\mu_{b_1}, \sigma_{b_1}^2\right)$ in the in-control class.

Figure 4.12 shows the classification performance in terms of correct classification rate of NNC for various feature extraction methods in Case C testing dataset. The plotted results are the average correct classification rates in 100 simulation runs.

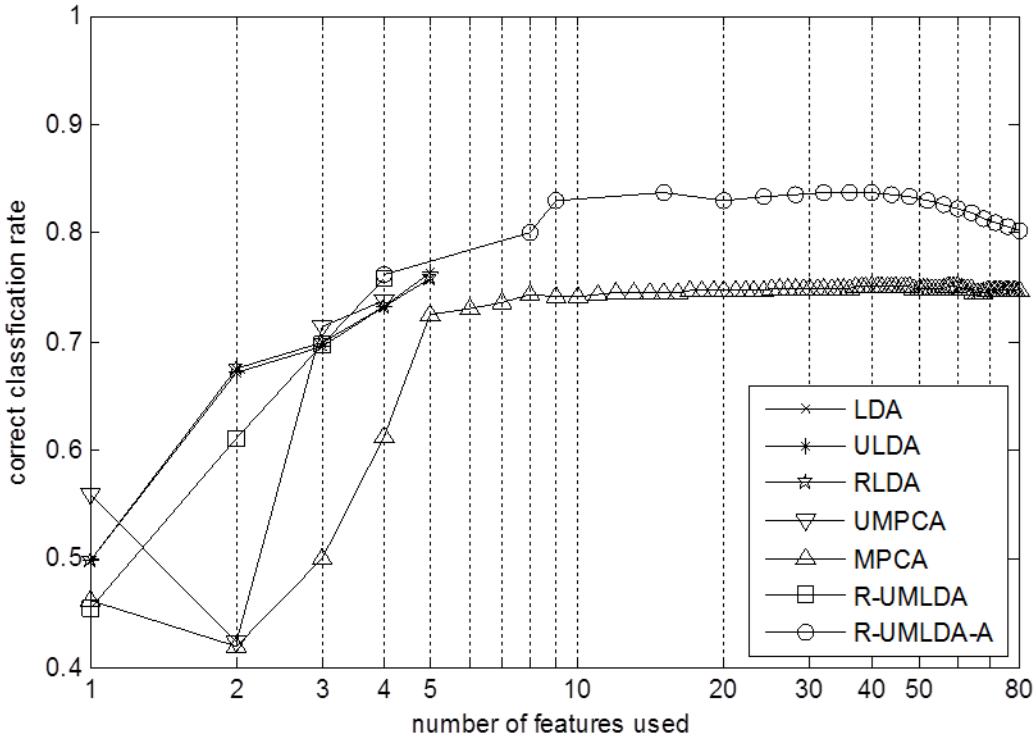
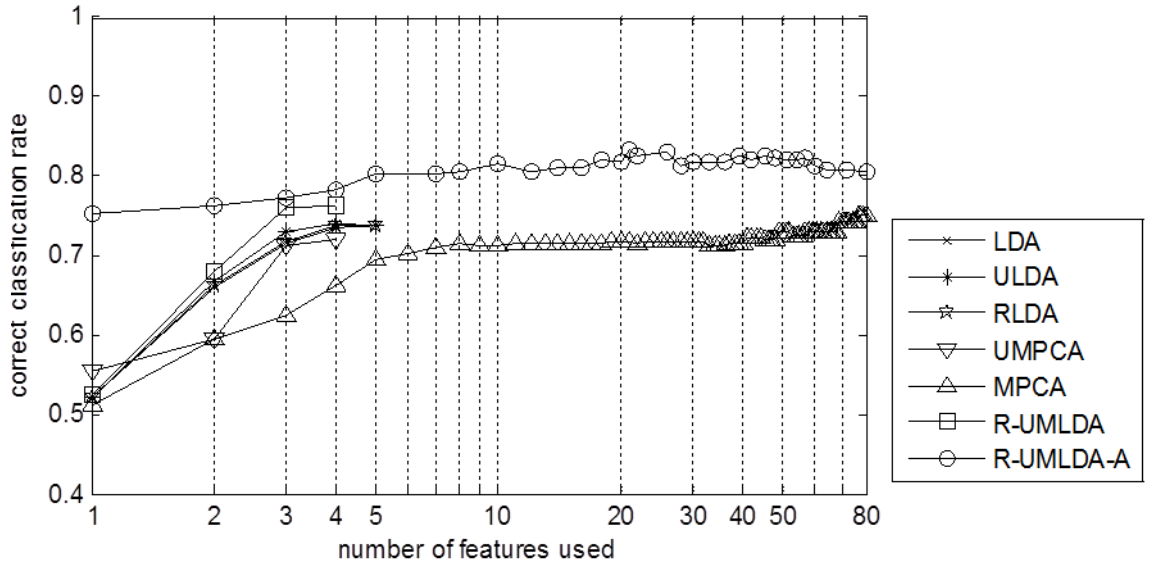


Figure 4.12. Classification performance of NNC for various feature extractors in Case C testing dataset

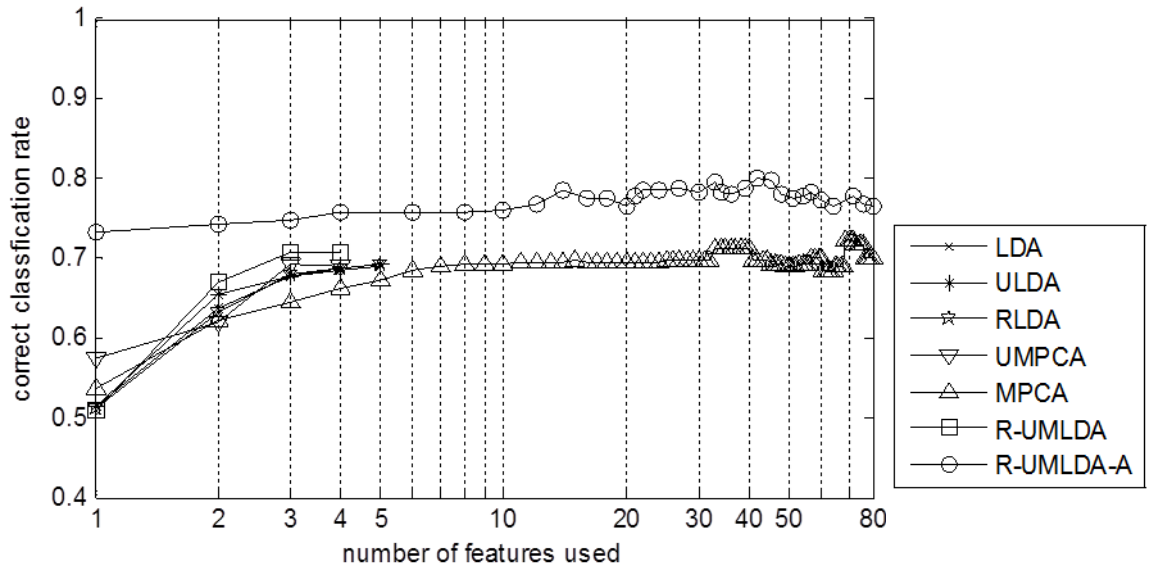
Similar to Cases A and B, the features extracted by UMPCA and MPCA are not as powerful as the other features in classification. Although the first few (1~2) features extracted by VLDA, ULDA, and RLDA are the most discriminative, using three or more R-UMLDA features lead to notably enhanced results. Figure 4.12 also shows that aggregation can effectively enhance the results, and that R-UMLDA and R-UMLDA-A outperform all other algorithms.

Under the framework of Case C, we discuss how the number of training samples in each class would affect feature extraction results. We consider two variants of Case C: C-1: generate 20 profile samples in-control and 20 samples in each of the five Case C out-of-control scenarios; C-2: generate 10 profile samples in-control and 10 samples in each of the five C out-of-control scenarios. Similarly, we use half of the samples as the training dataset in each case. Figure 4.13 shows the classification performance in terms of correct classification rate of NNC for various feature extraction methods in Cases C-1 and C-2 testing dataset. The plotted results are the average correct classification rates in 100 simulation runs.

Comparing Figure 4.12 with Figure 4.13(a) and Figure 4.13(b), we notice that although the correct classification rates in Figure 4.13 are slightly worse than those in Figure 4.12 due to the smaller sample sizes, the classification performance does not vary significantly given the different number of samples in each class. In all three cases, the best classification result is always achieved by R-UMLDA-A. If we want to limit the number of selected features to 3 or 4, then the first 3~4 features extracted by R-UMLDA are always the most powerful ones in classification. Figure 4.13, along with previous results, demonstrates that R-UMLDA-A achieves the best overall performance in all the simulation experiments, and that R-UMLDA-A is a robust and effective feature extraction and dimension reduction algorithm for multi-stream profiles.



(a)



(b)

Figure 4.13. Classification performance of NNC for various feature extractors in Case (a) C-1, and (b) C-2 testing dataset

4.3.2.4 Improving classification via ensemble learning

This subsection explores the possibility of further improving classification performance in fault diagnosis using ensemble learning. In R-UMLDA-A, 20 differently initialized and regularized UMLDA feature extractors are aggregated at the matching score level using the

nearest-neighbor distance. Although R-UMLDA-A achieves the best results in previous simulation experiments, more advanced ensemble-based learning algorithms such as boosting, bagging, and random space method are expected to achieve better results. Investigating alternative combination methods, however, is not the main focus of this chapter. So, we will only show the classification performance using the random subspace method and leave the in-depth studies in this direction to future work.

Random subspace method is an ensemble classifier that consists of several classifiers each operating in a subspace of the original feature space, and outputs the class based on the outputs of these individual classifiers. The *k*-nearest neighbor classifiers are used here as individual classifiers.

As an example, we consider the dataset from a single simulation run of Case A as described in Section 4.3.2.1 and Figure 4.6. Using the same 20 R-UMLDA feature extractors as in R-UMLDA-A, we plot the classification results of random space method and R-UMLDA-A in Figure 4.14. The curves with circle or cross markers correspond to random subspace classification with different number of nearest neighbors, i.e., different values of k. Comparing these results to R-UMLDA-A, which are plotted in square markers, we see that the random subspace ensemble significantly increases the accuracy of classification, given a proper choice of k. With $k = 20\sim25$, random subspace ensemble can achieve a relatively high correct classification rate using only 15 features, whereas R-UMLDA-A needs at least 20 features to achieve a similar performance. This also indicates more promising opportunities of using UMLDA for feature extraction and dimension reduction in handling multi-stream signals.

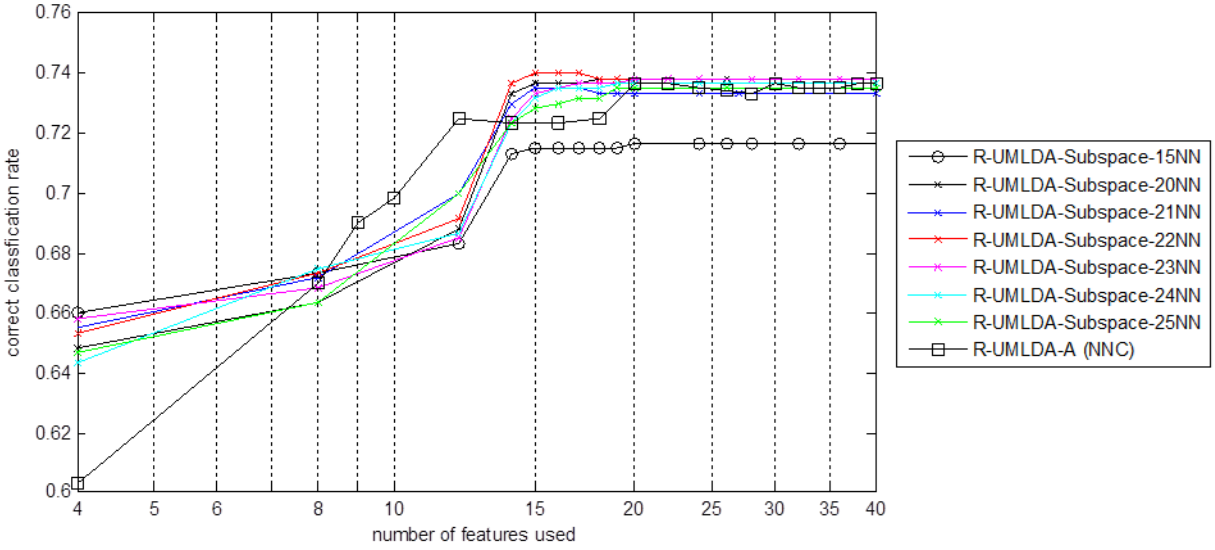


Figure 4.14. Classification performance of random space method for multiple R-UMLDA extractors in Case A testing dataset

4.4 Case Study in Multi-layer Ultrasonic Metal Welding

The ultrasonic metal welding example introduced at the beginning of this chapter is analyzed in this section to demonstrate the proposed method for multi-stream profile monitoring and fault diagnosis. Ultrasonic welding is a critical process for joining lithium-ion batteries for electric vehicles. In this case study, welding experiments of joining three layers of copper with 1 layer of nickel plated copper are investigated. The clamping pressure is 34 psi and the vibration amplitude is 40 μm.

As described in the introduction, Figure 4.2(b) shows the welded tabs from the normal welding process and three faulty processes: (1) surface contamination, (2) abnormal thickness, and (3) mislocated/edge weld. Figure 4.2(a) shows signals associated with these welds from four sensors. These sensor signals provide rich information about the product quality and process condition. Both R-UMLDA and VLDA methods are trained using 8

normal samples, 2 samples with fault 1 (oily surface), 1 sample with fault 2 (abnormal thickness), and 1 sample with fault 3 (edge weld).

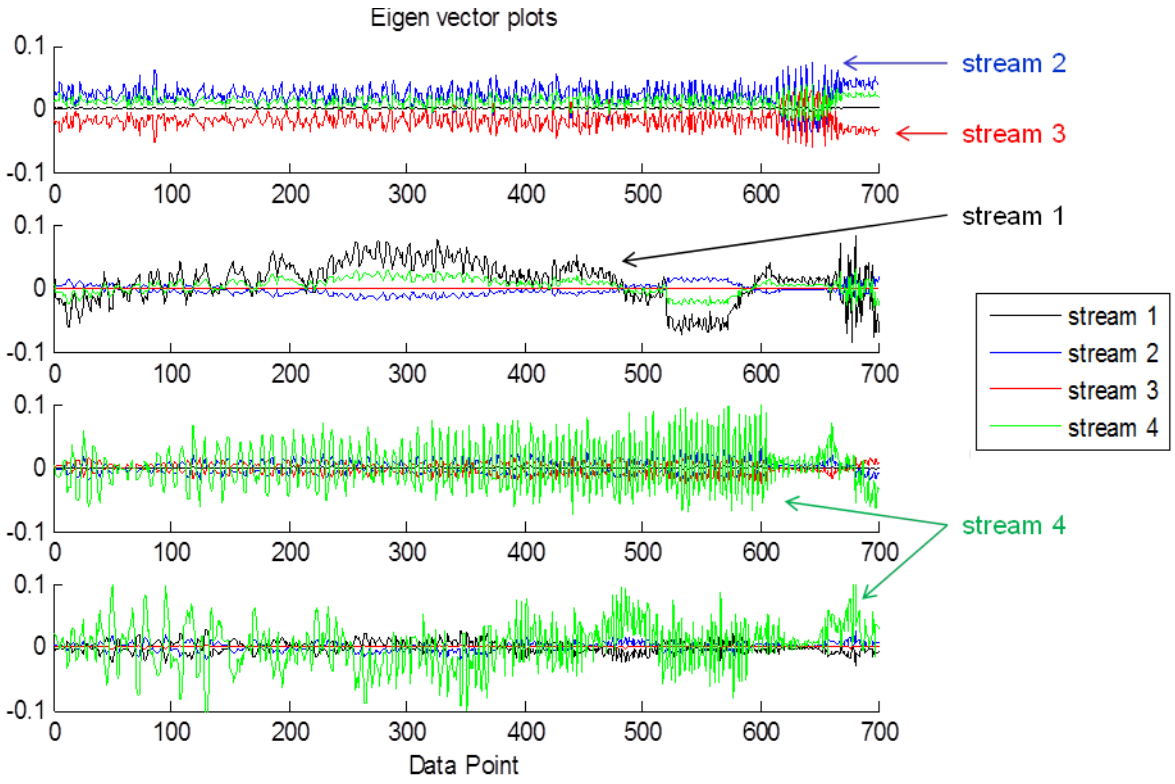


Figure 4.15. Eigentensors from R-UMLDA in ultrasonic metal welding

Using one R-UMLDA feature extractor with $\gamma = 0.001$, the eigenvectors corresponding to the four EMPs are shown in Figure 4.15. It can be seen from this figure that the eigenvectors corresponding to the first EMP show an efficient discrimination and strong negative correlation in streams 2 and 3. The eigenvectors corresponding to the second EMP show a strong discrimination in stream 1, whereas those corresponding to the third and fourth EMPs deliver similar information on discrimination in stream 4.

After training UMLDA and VLDA, the feature extractors and NNC are applied to five testing samples: 2 from the normal process, 2 from fault 1, and 1 from fault 2. Figure 4.16 plots the classification performance in terms of correct classification rate of NNC for

UMLDA and VLDA in the testing samples. For the five testing samples, it can be seen that R-UMLDA-A can easily achieve 100% correct classification using only four features, while R-UMLDA achieves 80%. The vectorized LDA methods, however, do not perform as well as UMLDA. The features extracted by RLDA achieve the same level of classification accuracy as R-UMLDA, whereas LDA and ULDA extract much weaker features. The results indicate that UMLDA-based methods, especially R-UMLDA-A, outperforms VLDA methods (including LDA, ULDA, and RLDA) in detecting abnormal processes and fault diagnosis.

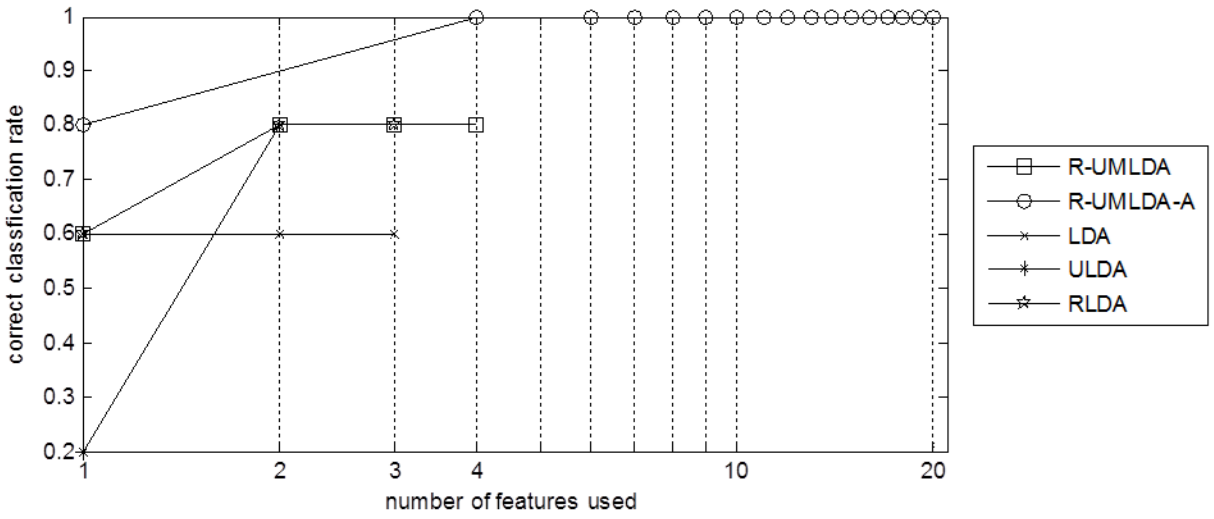


Figure 4.16. Classification performance of NNC for UMLDA and VLDA in ultrasonic welding

4.5 Conclusion

In this chapter, based on UMLDA, we proposed a method for effective analysis of multi-sensor heterogeneous profile data. With various sensors measuring different variables, information from each sensor, sensor-to-sensor correlation, and class-to-class correlation should all be considered. A simulation study was conducted to evaluate the performance of

the proposed method and its performance superiority over VLDA and other competitor methods. The results showed that the features extracted by VLDA and competitor methods are not as powerful as UMLDA in discriminating profiles and classification. The possibility of improving classification performance in fault diagnosis using ensemble learning with UMLDA was further explored. We also applied both UMLDA and VLDA to a real case study of a multi-layer ultrasonic metal welding process for the purpose of process characterization and fault diagnosis. The results indicate that UMLDA outperforms VLDA in not only detecting the faulty operations but also classifying the type of faults.

In the future, several remaining issues in this framework will be studied in more depth, such as the impact of the number of training samples on UMLDA-based feature extraction, and the combination of ensemble learning methods with R-UMLDA. More comprehensive case study will be performed in the future as we collect more samples from welding experiments. Developing tensor-based methods for monitoring manufacturing processes with vision technology will be an interesting topic for future research. Furthermore, the extension of the developed method to online process monitoring and online learning would be an interesting development.

References

Acar, E., and Yener, B. (2009) Unsupervised Multiway Data Analysis: A Literature Survey. *IEEE Transactions on Knowledge and Data Engineering*, 21, 6–20.

Amiri, A., Zou, C. and Doroudyan, M.H. (2013) Monitoring Correlated Profile and Multivariate Quality Characteristics. *Quality and Reliability Engineering International*, 30, 133-142.

Basir, O. and Yuan, X. (2007) Engine Fault Diagnosis Based on Multisensory Information Fusion Using Dempster–Shafer Evidence Theory. *Information Fusion*, 8, 379-386.

Cho, S., Binsaeid, S., and Asfour, S. (2010) Design of Multisensor Fusion-Based Tool Condition Monitoring System in End Milling. *The International Journal of Advanced Manufacturing Technology*, 46, 681–694.

Chou, S. H., Chang, S.I. and Tsai, T.R. (2014) On Monitoring of Multiple Non-linear Profiles. *International Journal of Production Researc*h, 52, 3209-3224.

De Lathauwer, L., De Moor, B. and Vandewalle, J. (2000) A Multilinear Singular Value Decomposition. SIAM Journal on Matrix Analysis and Applications, 21, 253–1278.

Donoho, D. L. and I. M. Johnstone, I.M. (1994) Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrik*a, 81, 425–455.

Duda, R.O., Hart, P.E. and Stork, D.G. (2012) Pattern Classification. John Wiley & Sons.

Grasso, M., Albertelli, P., and Colosimo, B.M. (2013) An Adaptive SPC Approach for Multi-sensor Fusion and Monitoring of Time-varying Processes. *Procedia CIRP*, 12, 61-66.

Grasso, M, Colosimo, B.M. and Pacella, M. (2014) Profile Monitoring Via Sensor Fusion: The Use of PCA Methods for Multi-Channel Data. *International Journal of Production Research*, 52, 6110-6135.

Hastie, T., Tibshirani R. and Friedman, J. (2008) *The Elements of Statistical Learning*, second edition. Springer, New York.

Ho, T. K. (1998) The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 832–844.

Hu, S.J. (2011) *Technical Report: On-line Quality Monitoring System for Ultrasonic Battery Tab Welding*, General Motors Collaborative Research Lab at the University of Michigan, Ann Arbor, MI.

Jin, J and Shi, J. 1999. Feature-preserving Data Compression of Stamping Tonnage Information Using Wavelets. *Technometric*s, 41, 327-339.

Jin , Z., Yang, J.Y., Hu, Z.S. and Lou, Z. (2001) Face recognition based on the uncorrelated discriminant transformation. *Pattern Recognition*, 34, 1405–1416.

Kiers, H. A. L. (2000). Towards a Standardized Notation and Terminology in Multiway Analysis. *Journal of Chemometrics*, 14, 105–122.

Kim, J., Huang, Q., Shi, J. and Chang, T.-S. (2006) Online Multichannel Forging Tonnage Monitoring and Fault Pattern Discrimination Using Principal Curve. J*ournal of Manufacturing Science and Engineering*, 128, 944-950.

Kuljanic, E., Totis, G., and Sortino, M. (2009) Development of an Intelligent Multisensor Chatter Detection System in Milling. *Mechanical Systems and Signal Processing* , 23, 1704–1718.

Lee, S.S., Kim, T.H., Hu, S.J., Cai, W.W. and Abell, J.A. (2010) Joining Technologies for Automotive Lithium-ion Battery Manufacturing – A Review. *Proceedings of ASME 2010 International Manufacturing Science and Engineering Conference*, 541-549.

Lee, S. S., Shao, C., Kim, T. H., Hu, S. J., Kannatey-Asibu, E., Cai, W. W., Spicer, J.P., Wang, H. and Abell, J. A. (2014). Characterization of Ultrasonic Metal Welding by Correlating Online Sensor Signals With Weld Attributes. *Journal of Manufacturing Science and Engineering*, 136, 051019.

Lu, H., Plataniotis, K.N. and Venetsanopoulos, A.N. (2008) MPCA: Multilinear Principal Component Analysis of Tensor Objects. *IEEE Transactions on Neural Networks*, 19, 18–39.

Lu, H., Plataniotis, K.N. and Venetsanopoulos, A.N. (2009) Uncorrelated Multilinear Discriminant Analysis With Regularization and Aggregation For Tensor Object Recognition. *IEEE Transactions on Neural Network*s, 20, 103-123.

Lu, H., Plataniotis, K.N. and Venetsanopoulos, A.N. (2009) Uncorrelated Multilinear Principal Component Analysis for Unsupervised Multilinear Subspace Learning. *IEEE Transactions on Neural Network*s, 20, 1820–1836.

Noorossana, R., Saghaei, A and Amiri, A. (2012) Statistical Analysis of Profile Monitoring. New York: Wiley.

Paynabar, K and Jin, J. (2011) Characterization of Non-Linear Profiles Variations Using Mixed-Effect Models and Wavelets. *IIE Transactions*, 43: 4, 275-290

Paynabar, K., Jin, J., and Pacella, M. (2013) Monitoring and Diagnosis of Multichannel Nonlinear Profile Variations Using Uncorrelated Multilinear Principal Component Analysis. *IIE Transactions*, 45, 1235-1247.

Woodall, W.H. (2007) Current Research on Profile Monitoring. Produção, 17, 420-425.

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, 309-320.

Ye, J. (2005) Characterization of A Family of Algorithms for Generalized Discriminant Analysis on Undersampled Problems. *Journal of Machine Learning Research*, 6, 483–502.

Ye, J., Xiong, T., Li, Q., Janardan, R., Bi, J., Cherkassky, V., and Kambhamettu, C. (2006) Efficient Model Selection for Regularized Linear Discriminant Analysis. *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, 532-539.

CHAPTER 5

CONCLUSION AND FUTURE RESEARCH

5.1 Conclusions and Contributions

This dissertation has focused on developing new methodologies to utilize sensor signals to monitor operational quality, detect process anomalies, diagnose fault root causes, and guide corrective actions to meet the increasing demand in high product quality and reliability in ultrasonic welding of lithium-ion batteries. First, a new process monitoring algorithm called SPC-M was proposed to achieve a near-zero misdetection error and reduce the manual inspection rate. Then, a new wavelet-based profile monitoring method was developed to consider both real-time monitoring and control opportunity to achieve maximal defect prevention. Furthermore, considering multi-sensor heterogeneous profile data, a new method for profile monitoring and fault diagnosis was developed. A detailed summary of the dissertation and its new contributions is given below:

1. *A new process monitoring algorithm for achieving a near-zero misdetection error rate and reducing the manual inspection rate.* In this study, we addressed a critical issue in weld quality monitoring that targets a near-zero Type II error rate and low manual inspection rate for ultrasonic metal welding process in a battery assembly plant. Conventional control chart techniques, which set control limits based on a given Type I error rate, cannot achieve the near-zero Type II error rate that is desired in this case. The proposed monitoring algorithm ‘SPC-M’ is developed by integrating univariate statistical process control method and the multivariate Mahalanobis distance

approach. The acceptance region of the proposed SPC-M algorithm is the shared region of multiple univariate control limits and the Mahalanobis distance limits. In this way, SPC-M can be used to monitor multivariate processes in order to achieve a near-zero Type II error rate. The control limits for SPC-M algorithm were established and then tuned based on training data. The proposed method was validated on test data from battery manufacturing. Comparative studies also demonstrated the superiority of using SPC-M in our case over SPC individuals control chart or M-distance approach. With 0% Type II error rate over the testing period, the SPC-M algorithm did not pass any suspected bad welds to downstream processes.

2. *A new wavelet-based profile monitoring method considering both real-time monitoring and control opportunity for achieving maximal defect prevention.* We developed a new wavelet-based profile monitoring method by considering the tradeoff between real-time monitoring accuracy and within-cycle control effectiveness for defect prevention. In order to leave sufficient time for real-time process adjustment, the monitoring decision is made based on an early portion of the cyclic signal, while the optimal decision point is determined by solving an optimization problem. Wavelet-based control charts are then developed to monitor profile deviations and detect process changes. The effectiveness of the proposed method was validated and demonstrated by simulations and case studies. With the developed within-cycle profile monitoring strategy, the proportion of out-of-control profiles was significantly reduced, indicating great potential in preventing defects in real time. In addition, results from the sensitivity analysis provided insights on the online application of the proposed method and its robustness against inaccuracies in online estimation.
3. *A new method based on multilinear discriminant analysis for multi-stream profile monitoring and fault diagnosis.* We proposed a method for effective analysis of multi-

sensor heterogeneous profile data using a multilinear extension of linear discriminant analysis, called the uncorrelated multilinear discriminant analysis (UMLDA). With various sensors measuring different variables, information from each sensor, sensor-to-sensor correlation, and class-to-class correlation should all be considered. A simulation study was conducted to evaluate the performance of the proposed method and its performance superiority over vectorized-LDA and other competitor methods. The results showed that the features extracted by UMLDA are most powerful in discriminating profiles and classification. The possibility of improving classification performance in fault diagnosis using ensemble learning with UMLDA was also explored. A case study on multi-layer ultrasonic metal welding demonstrated the effectiveness of the proposed UMLDA-based method in not only detecting the faulty operations but also classifying the type of faults.

5.2 Future Research

The methodologies and models developed in this dissertation could be further improved and extended in the following directions:

1. *Development of an adaptive training scheme for online process monitoring with near-zero misdetection.*

   The SPC-M algorithm developed in Chapter 2 of this dissertation makes use of a large size of training data, about 200 good samples and a certain number of poor quality samples, which is easily available from rapid mass production. If the training samples are limited, e.g., in small batch production, establishing and tuning SPC-M control limits would be very challenging. In order to shorten the algorithm training period and reduce the data collection cost, developing an adaptive training scheme with online learning capabilities for SPC-M is worth future research efforts.

2. *Development of within-cycle monitoring and control methods with multi-step adjustment of process parameter(s) for defect prevention.*

   The within-cycle profile monitoring algorithm developed in Chapter 3 of this dissertation is based on the assumption that process change occurs in an early stage of the operation and that potential defects can be prevented by a one-step adjustment of process parameter(s) during the operation. The effectiveness of the control action is limited by the remaining time for control and also the capacity of the controller. In cases where one-step adjustment is less effective or more advanced controllers are available, multiple control actions should be considered. Developing within-cycle monitoring and control methods for multi-step adjustment of process parameter(s) will be an interesting topic for future research. The consideration of multiple control actions would make it very challenging in balancing the tradeoff between real-time monitoring accuracy and within-cycle control effectiveness for defect prevention at each control step.

3. *Development of within-cycle profile monitoring algorithms using multi-stream signals and sensor fusion.*

   The profile monitoring and fault diagnosis method developed in Chapter 4 of this dissertation considers fully observed signals obtained at the end of each operation, which gives a time-delayed monitoring decision that is too late for within-cycle defect prevention. In order to enable defect prevention, the optimal decision point needs to be determined considering not only the tradeoff between real-time monitoring accuracy and within-cycle control effectiveness, but also the interrelationship of different sensors. Thus, the extension of the developed multi-stream profile monitoring and fault diagnosis method for fully observed signals to within-cycle

decisions using partially observed signals would be an interesting development that is worth future research efforts.

4. *Advancement of knowledge on the relationship between tool wear and online signals in ultrasonic metal welding.*

   It is reported that welding tool replacement is a major production cost in vehicle battery production. In the third problem studied in the dissertation, tool wear is considered as a type of faulty process conditions, and the effect of tool wear on sensor signals is represented by the gradual mean shift of the reference signal in simulation. A more accurate quantification of tool wear should consider the dynamic change of the shape of the tool and remaining tool life. Thus, more in-depth investigation on the relationship between tool wear and online signals is needed in order to extract and analyze information from online multi-stream signals to help indicate tool wear status and estimate remaining tool life.

5. *Development of real-time process monitoring methods for complex types of data.*

   The data considered in this dissertation include sensor signals and features. With the advances in online sensing and data capturing technology, high-dimensional data may be collected in broad applications, such as multi-stream signals, images, and spatiotemporal data. The problem would be more complicated if missing data and/or the mixture of continuous and categorical data are involved. It is challenging, yet crucial, to adapt and modify the current profile monitoring techniques and use them for high-dimensional data.

6. *Improvement of the performance of ultrasonic welding processes.*

   One direction to help further improve the performance of ultrasonic welding processes is to link the profile monitoring methods developed in this dissertation with the determination of process setup parameters. It would be an interesting development

for future research to establish the relationship among process parameters, profile variations, and product quality through the integration of design of experiments, response profile analysis, and engineering domain knowledge.