VARIATION MONITORING, DIAGNOSIS AND CONTROL FOR COMPLEX
SOLAR CELL MANUFACTURING PROCESSES

by

Huairui Guo

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF SYSTEMS AND INDUSTRIAL ENGINEERING
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

2004
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
The University of Arizona™
Graduate College

As members of the Final Examination Committee, we certify that we have read the
dissertation prepared by Huairui Guo
entitled Variation Monitoring, Diagnosis, and Control for Complex Solar
Cell Manufacturing Processes

and recommend that it be accepted as fulfilling the dissertation requirement for the
Degree of Doctor of Philosophy

Dr. Jionghua Jin 11/20/04
Dr. Ronald G. Askin 11/20/04
Dr. Ferenc Szidarovszky 11/20/04
Dr. Salim Hariri 11/24/04
Dr. Kevin McNeill 12/2/04

Final approval and acceptance of this dissertation is contingent upon the
candidate's submission of the final copies of the dissertation to the Graduate College.
I hereby certify that I have read this dissertation prepared under my direction and
recommend that it be accepted as fulfilling the dissertation requirement.

Dr. Jionghua Jin 11/20/04
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirement for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: [Signature]
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Jionghua Jin. Without her guidance and kindly support, I might not have the opportunity to explore the field of reliability and quality engineering. I would like to thank members of my committee, Dr. Askin, Dr. Szidarovszky, Dr. McNeill, and Dr. Hariri. I appreciate their help in both my course work and research projects.

I would like to thank my friends, former and current graduate students in the Department of Systems & Industrial Engineering: Dr. Xuejun Zhu, Mr. Zhihong Zhou and Mr. Seunghyun Baek for their advice, support and friendship.

Special thanks to my wife, Pengying Niu. It was her understanding and support that made this journey possible.
DEDICATION

This work is dedicated to my family, my father Erfu Guo, my mother Xiumei Fu, my wife Pengying Niu, and my brother Huaiming Guo. Their love, support, understanding and encouragement have been the major motivation for me in working towards the Ph. D. degree.
TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. 10

LIST OF FIGURES ............................................................................................................... 11

ABSTRACT .......................................................................................................................... 13

CHAPTER 1: INTRODUCTION .............................................................................................. 15

1.1 MOTIVATION .............................................................................................................. 15

1.2 RESEARCH FRAMEWORK IN THIS DISSERTATION .............................................. 18

1.2.1 SPC Supervisory Generalized Predictive Control for Thin Film Deposition Process .................................................................................................................. 18

1.2.2 Hierarchical ANOVA for Variance Decomposition and Diagnosis in Batch Printing Process ................................................................................................................ 25

1.2.3 Multiscale Statistical Process Monitoring Using Wavelets Analysis for Autocorrelated Data .............................................................................................................. 29

1.3 OUTLINE OF THE DISSERTATION .......................................................................... 35

CHAPTER 2: PROCESS BACKGROUND AND VARIABLE DESCRIPTION .................. 36

2.1 DESCRIPTION OF PROCESS FLOW ........................................................................ 37

2.2 CIGS THIN FILM DEPOSITION PROCESS ................................................................ 39

2.3 PANEL SILVER PRINTING PROCESS ...................................................................... 41

2.3.1 Description of Process Variables ......................................................................... 43

2.3.2 Selection of Interested Variation Components .................................................. 44
TABLE OF CONTENTS –Continued

CHAPTER 3: SPC BASED SUPERVISORY GENERALIZED PREDICTIVE
CONTROLL FOR THIN FILM DECOMPOSITION PROCESS ........... 47

3.1 PROCESS MODELING USING ARMAX MODEL ......................... 48

3.2 PROCESS CONTROL USING GENERALIZED PREDICTIVE CONTROL
ALGORITHM .................................................................................. 49

3.3 SUPERVISORY PREDICTIVE CONTROL STRATEGIES ............... 56

3.3.1 CUSUM Charts for Detection and Estimation of Shift and Drift
Disturbances ........................................................................... 56

3.3.2 X-bar Chart for the Detection and Estimation of a Spike Disturbance ...... 60

3.3.3 Control Law Revision for Supervisory Compensation of the Detected
Disturbances ........................................................................... 62

3.4 CASE STUDY ........................................................................... 64

3.4.1 Thin Film Deposition Process Modeling .................................. 64

3.4.2 Predictive Control .................................................................. 66

3.4.3 Supervisor Strategies ............................................................. 68

3.5 CONCLUSION ........................................................................... 75

CHAPTER 4: ANOVA METHOD FOR VARIANCE COMPONENT
DECOMPOSITION AND DIAGNOSIS IN BATCH
MANUFACTURING PROCESSES ..................................................... 77

4.1 GENERAL DECOMPOSITION FRAMEWORK OF TOTAL PROCESS
VARIATIONS .................................................................................. 78
TABLE OF CONTENTS –Continued

4.2 VARIANCE COMPONENT ANALYSIS USING ANOVA .......................... 80
   4.2.1 Statistical Modeling of Interested Variation Components............. 82
   4.2.2 Statistical Testing and Estimation of Variance Components.......... 83
   4.2.3 Site Variability Decomposition and Diagnosis Based on Linear Contrasts 85
4.3 CASE STUDY ............................................................................. 94
4.4 CONCLUSION ............................................................................ 98

CHAPTER 5: MULTISCALE STATISTICAL PROCESS MONITORING USING WAVELETS ANALYSIS FOR AUTOCORRELATED DATA ...... 100

5.1 REVIEW OF WAVELET TRANSFORM AND MULTISCALE ANALYSIS. 101
   5.1.1 Wavelet Transform............................................................. 101
   5.1.2 Multiscale Analysis............................................................ 103
   5.1.3 Signal Decomposition by Filter Bank Using Haar Transform........ 104
5.2 DESCRIPTION OF PROCESS MODELS AND CHARACTERISTICS...... 106
   5.2.1 ARMA(n, m) Model and Process Fault Conditions...................... 106
   5.2.2 Autospectrum Analysis of ARMA (n, m).................................. 107
   5.2.3 Response Characteristics of Variance Change .......................... 108
   5.2.4 Response Characteristics of Mean Shift .................................. 109
   5.2.5 Relationship of Wavelet Coefficients and System Frequency Responses 113
5.3 VARIANCE CHANGE MONITORING USING DETAIL COEFFICIENTS... 114
   5.3.1 Review of SD-CUSUM Chart for Process Variability Monitoring .... 114
   5.3.2 Selection of Optimal Level of Detail Coefficients for Fault 1 Detection 116
TABLE OF CONTENTS –Continued

5.3.3 Selection of Optimal Level of Detail Coefficients for Fault 2 Detection. 119
5.3.4 Simulation Study.................................................................122

5.4 MEAN SHIFT MONITORING USING SCALE COEFFICIENTS ..........131
5.4.1 Selection of Optimal Decomposition Level of Scale Coefficients for
Mean Shift Detection ....................................................................131
5.4.2 WEWMA Mean Shift Monitoring Robust to Variance Change ....133
5.4.3 Comparison of the WEWMA with Direct-EWMA and SCC Charts for
Mean Shift Detection ....................................................................135

5.5 MULTISCALE MONITORING OF BOTH MEAN AND VARIANCE
CHANGES.....................................................................................136

5.6 CONCLUSION.............................................................................141

CHAPTER 6: CONCLUSION AND FUTURE WORK .........................143
6.1 CONCLUSION.............................................................................143
6.2 FUTURE WORK.........................................................................146

REFERENCES ..................................................................................148
LIST OF TABLES

Table 2.1 Effect of process variables on potential variations ........................................ 44
Table 4.1 Expected Mean Squares for Decomposition I of Eq. (4-5) ............................. 84
Table 4.2 Analysis of Variance Table for Decomposition I of Eq. (4-5) ...................... 84
Table 4.3 Expected Mean Squares for Decomposition II of Eq. (4-7) ......................... 84
Table 4.4 Analysis of Variance Table for Decomposition II of Eq. (4-7) ................... 85
Table 4.5 Expected Mean Squares for Contrast Decomposition I of Eq. (4-29) ............ 91
Table 4.6 Analysis of Variance Table for Contrast Decomposition I of Eq. (4-29) ....... 91
Table 4.7 Expected Mean Squares for Contrast Decomposition II of Eq. (4-38) ......... 93
Table 4.8 Analysis of Variance Table for Contrast Decomposition II of Eq. (4-38) .... 94
Table 4.9 Sum of squares of the full decomposition for two factors ........................... 95
Table 4.10 $F$-test of variance components for old process ....................................... 95
Table 4.11 Sum of squares based on the full decomposition of three factors ............... 96
Table 4.12 $F$-test for the contrast variations ............................................................ 96
Table 4.13 Sum of squares based on the full decomposition of the new process ......... 97
Table 4.14 $F$-test of variance components for the new process ............................... 98
Table 5.1 Exemplary models and their characteristics .............................................. 111
Table 5.2 Comparison of ARLs: Wavelets-based EWMA, Direct EWMA and SCC
    EWMA under i.i.d. noise ......................................................................................... 135
LIST OF FIGURES

Figure 1.1 Overview of a thin film module manufacturing process ........................................ 16
Figure 1.2 Structure of SPC-supervisory GPC ................................................................... 24
Figure 1.3 Flow chart of dissertation ................................................................................... 35
Figure 2.1 Example of final solar cell product ..................................................................... 39
Figure 2.2 CIGS Thin film deposition process ................................................................... 40
Figure 2.3 Control system of the CIGS deposition process .................................................. 41
Figure 2.4 Silver Screen Printing Process .......................................................................... 42
Figure 2.5 Measurement locations on a printed panel .......................................................... 43
Figure 3.1 Initial production run data of step input and response ........................................ 64
Figure 3.2 AIC under different model structures ................................................................. 66
Figure 3.3 Comparison of the modeling results with the real process response .................... 66
Figure 3.4 Impulse and step response of the thin film deposition process model .................. 67
Figure 3.5 A simulated control performance ...................................................................... 68
Figure 3.6 Mean shift disturbance function and CUSUM monitoring control chart .......... 69
Figure 3.7 Comparison of control performance under mean shift disturbance ................. 70
Figure 3.8 Comparison of model residual errors under mean shift disturbance ................ 70
Figure 3.9 CUSUM monitoring control chart for detecting drift disturbance ..................... 72
Figure 3.10 Control performance comparison under drift disturbance .............................. 72
Figure 3.11 Comparison of model errors under drift disturbance ....................................... 72
Figure 3.12 Control performance comparison under spike disturbance ............................ 74
Figure 3.13 Control performance under a spike disturbance for a longer memory system ................................................................. 75
Figure 4.1 General decomposition of variance components .............................................. 79
Figure 5.1 Filter bank structure for wavelets transform ....................................................... 105
Figure 5.2 Filter frequency response using Haar Transform .............................................. 105
Figure 5.3 Process model structure .................................................................................... 107
Figure 5.4 Autospectrum of the models in Table 5.1 ......................................................... 111
Figure 5.5 Unit step response of the models in Table 5.1 .................................................. 112
LIST OF FIGURES –Continued

Figure 5.6 Autospectrum of the unit step responses of the models in Table 5.1 .......... 112
Figure 5.7 CUSUM chart for mean shift detection under process variance change...... 116
Figure 5.8 Relationship between $\delta_a$ and $\delta_d$ at different levels ................................. 118
Figure 5.9 Relationship between $\delta_e$ and $\delta_d$ at different levels ($\sigma_e = 1$) ..................... 120
Figure 5.10 Relationship between $\delta_e$ and $\delta_d$ at different levels ($\sigma_e = 3$)................... 121
Figure 5.11 Comparison of SD-WCUSUM and SD-CUSUM to mean shift .................. 123
Figure 5.12 Choosing an optimal detail level for variance change detection of process with measurement error (Model 3) .......................................................... 126
Figure 5.13 Choosing an optimal detail level for variance change detection of process with measurement error (Model 1) .......................................................... 127
Figure 5.14 Choosing an optimal detail level for measurement error variance change detection (Model 1) ...................................................................................... 129
Figure 5.15 Choosing an optimal detail level for measurement error variance change detection (Model 3) ...................................................................................... 131
Figure 5.16 Wavelets decomposition of the signal of Model 1 under mean shift .......... 132
Figure 5.17 The effect of variance change for mean shift detection ............................... 134
Figure 5.18 Using Multiscale method for process change detection ............................. 141
ABSTRACT

Interest in photovoltaic products has expanded dramatically, but wide-scale commercial use remains limited due to the high manufacturing cost and insufficient efficiency of solar products. Therefore, it is critical to develop effective process monitoring, diagnosing, and control methods for quality and productivity improvement. This dissertation is motivated by this timely need to develop effective process control methods for variation reduction in thin film solar cell manufacturing processes.

Three fundamental research issues related to process monitoring, diagnosis, and control have been studied accordingly. The major research activities and the corresponding contributions are summarized as follows:

(1) Online SPC is integrated with generalized predictive control (GPC) for the first time for effective process monitoring and control. This research emphasizes on the importance of developing supervisory strategies, in which the controller parameters are adaptively changed based on the detection of different process change patterns using SPC techniques. It has been shown that the integration of SPC and GPC provides great potential for the development of effective controllers especially for a complex manufacturing process with a large time varying delay and different process change patterns.

(2) A generic hierarchical ANOVA method is developed for systematic variation decomposition and diagnosis in batch manufacturing processes. Different from SPC, which focuses on variation reduction due to assignable causes, this research aims to reduce inherent normal process variation by assessing and diagnosing inherent variance components from production data. A systematic method of how to use a full factor
decomposition model to systematically determine an appropriate nested model structure is investigated for the first time in this dissertation.

(3) A multiscale statistical process monitoring method is proposed for the first time to simultaneously detect mean shift and variance change for autocorrelated data. Three wavelet-based monitoring charts are developed to separately detect process variance change, measurement error variance change, and process mean shift simultaneously.

Although the solar cell manufacturing process is used as an example in the dissertation, the developed methodologies are generic for process monitoring, diagnosis, and control in process variation reduction, which are expected to be applicable to various other semiconductor and chemical manufacturing processes.
CHAPTER 1
INTRODUCTION

1.1 MOTIVATION

Renewable energy is playing a more and more important role in the world energy supply since the gas and oil prices have increased dramatically. One important type of renewable energy is solar-electric power generated by photovoltaic (PV) products. PV products, such as solar cell battery, are made from materials that have photovoltaic effect to convert solar energy into direct current (DC) electricity. The unique advantage of using the solar-electric power is its renewable and nonpolluting merits.

Interest in photovoltaic (PV) products has expanded dramatically, but wide-scale commercial use remains limited due to the high manufacturing cost and insufficient efficiency of solar products. The industry has established an 18% to 20% conversion efficiency goal at a cost of less than 50 cents per watt for the thin film based PV products. However, current PV products based on thin film modules can only achieve 7% to 10% efficiency (U.S. Photovoltaic industry roadmap, 2003). This obstacle is mainly due to the lack of effective process control methods suitable to the large-scale automatic mass production.

In the transition of PV thin film manufacturing from small scale to mass production, it is critical to develop effective technologies which can quantitatively assess the relationship between process conditions and product properties, quickly identify the
root causes of process variations, and effectively adjust the process to compensate for the
effect of inevitable process disturbances. This dissertation is motivated by this timely
need to develop an effective process control methodology for variation reduction in thin
film solar cell module manufacturing processes.

Figure 1.1 Overview of a thin film module manufacturing process

Figure 1.1 gives a brief overview of the production flow in a thin film solar cell
manufacturing process, which consists of three major production stages: thin film
deposition, panel printing, and module assembling. Four different types of thin film
deposition processes are employed to deposit different materials on a continuous
substrate layer by layer. Each layer of materials should be uniformly deposited to meet
the specific ratio among materials and thickness uniformity requirements, which is
essential for efficiently absorbing and converting solar energy into electricity. After the
deposition process, the continuous thin film roll is cut into panels to fit into a printing
process, in which silver print gridlines are printed on the deposited thin film for
transmitting electrons to generate DC electricity. The variability of gridline thickness is a
critical concern in the panel printing process because it directly affects the circuit resistance. Finally, a panel is cut into cells, which are then sorted and assembled together based on their efficiency grades for the designed modules. More detail description of process variables at each production stage and their associated quality measurement will be given in Chapter 2.

As it is shown above, the solar cell module manufacturing is a very complex process, in which final product quality of module efficiency is affected by all above production stages. Moreover, different production stages may have different variation patterns, which are closely related to their design function, process control methods, potential process faults and disturbances. Therefore, the associated techniques for process monitoring, diagnosis, and control should be explored accordingly. Based on the most common need of the technology development, in this dissertation, three research topics will be addressed specifically. They are (a) automatic feedback control to adaptively adjust the temperature for achieving uniform thin film deposition; (b) hierarchical ANOVA to automatically identify inherent variations and root causes in a batch printing process using production observation data; and (c) multiscale SPC monitoring to automatically detect both mean shift and variance change for autocorrelated processes.

It will be shown that although this dissertation research is conducted based on a thin film solar module manufacturing process, the developed methodology holds generic research results in terms of the system modeling and automatic controller development, statistical process monitoring for autocorrelated data, and variation decomposition and
diagnosis for batch manufacturing processes. Therefore, it is expected to possibly extend those developed methodologies into other production stages or other manufacturing processes wherever those process control and available sensor data have similar characteristics.

1.2 RESEARCH FRAMEWORK IN THIS DISSERTATION

This section will provide an overview of the associated fundamental research issues for the development of the methodology highlighted in the previous section. The objective, associated literature reviews, and the proposed approach will be discussed for each topic specifically.

1.2.1 SPC Supervisory Generalized Predictive Control for Thin Film Deposition Process

Problem Statement. In thin film deposition process, because of the complexity of the source effusion, heat dissipation, source consumption, and the uncertainty of the time delay within the process, the thickness uniformity control is a critical research challenge. In order to provide a high throughput, it is expected to simultaneously deposit multiple materials on a large area of the substrate surface with a fast substrate moving speed. However, this leads to another challenge issue of how to reduce the loss due to the film non-uniformity as the production progresses across both substrate width and length.

It is known that one major process factor impacting the variation of thin film thickness is source temperature (Simpson, 2003). In order to control thickness, appropriate control of temperature is needed to compensate for unanticipated process
disturbances during production. In general, a thin film deposition control has two steps in each run: a manual process control during the initial start up of production, and an automatic feedback control during continuous production. In the beginning of each new production run, a manual control is performed to increase the temperature through several steps. After the system response (i.e. the film thickness) is close to the target point, an automatic controller is engaged to maintain the film thickness at target based on the thickness measurement data. Most existing controllers are a pre-designed time-invariant controller based on a fixed model identified off-line. Although such a controller is simple in its implementation, it cannot thoroughly consider different disturbances in a given run and the variation between run-to-run. Examples of unanticipated disturbances and variations are the properties of source materials, the operation of rebuilding and reinstallation of source cells between runs, impurity in source materials and debris in a chamber, utilization of materials over production time, characteristic change of source heat dissipation to substrate and chamber, and degradation of source heaters, etc. As a result, film thickness may vary even though the source temperatures are maintained at the same target setting points. In addition, it is extremely difficult to model the interaction of source evaporation process with other process variables such as environmental temperature and various disturbances. Thus, the compensation of disturbance and run-to-run variations cannot be achieved through neither an simple automatic feedforward control nor an off-line pre-setting of the source temperature. How to design and use suitable on-line feedback control which can consider the above process properties is a critical research problem.
In current thin film thickness control, the conventional SPC (Statistical Process Control) procedure is also widely used. But it only conducts a post-process quality analysis by using off-line measurement instruments to analyze the film layer structure and thickness. However, it is always expected that a manufacturing process have the capability of real time sensing and process changes detecting for making corresponding process adjustment to prevent or reduce defective product loss over the production. This issue is more critical for a continuous dynamic manufacturing process, like thin film deposition, whenever it has a need to accommodate unanticipated process disturbance, variations, or drifts during production time. The recent development of in-situ thickness sensors has provided a great potential of achieving in-process monitoring and online automatic process control. However, how to integrate in-process monitoring techniques with an efficient and accurate process control algorithm is still an open research problem. In the first part of this dissertation, research on how to integrate SPC monitoring for effective online feedback controller development will be addressed in Chapter 3.

State of the Art. Many research results on process feedback control have been developed and used (Astrom and Wittenmark, 1989, 1990; Butler and Stefani, 1994; Guo and Sachs, 1993; Ingolfsson and Sachs, 1993; Castillo and Hurwitz, 1997; Del Castillo, 2001; Tseng et al, 2002). When a process controller is designed for a given system, a particular system model should be identified first. A system model can either be built from historical data or built at the beginning of each process run using current run data. The later method is used when the run-to-run variation is relatively high and should be considered in the control model.
The most widely used process control model is Exponentially Weighted Moving Average (EWMA) model. Ingolfsson and Sachs (1993) used a first-order EWMA model for a process and discussed the stability conditions of the controller. In their paper, they showed that the EWMA controller would gradually tune the control value so that the expectation of the process output would meet the desired target gradually. Later Bulter and Stefani (1994) proposed a double EWMA controller to eliminate the deterministic drift within the process. However, EWMA controller only has one model parameter and double EWMA controller only has two, so it is difficult to meet the requirement of a complex dynamic system. Therefore, more general time series models were discussed by Del Castillo (1996), Del Castillo and Hurwitz (1997). Once the process model is obtained, control law can be developed based on this model and a pre-chosen control objective.

The most widely used control objective is minimum variance, which tries to minimize the variance of one-step-ahead prediction error. In the book of Box, et al (1994), the detail of minimum variance control algorithm is given. However, minimum variance control lacks robustness to the system dead-time and model order. If the model is not accurate, the prediction error cannot be minimized by using minimum variance control (Pandit and Wu, 1983).

To reduce the influence of model estimation errors in the minimum variance control, another control algorithm — generalized predictive control (GPC) was first proposed by Clark et al (1987), and the properties of GPC are further studied by Clark and Mohtadi (1989) for a set of continuous chemical process control problems. Recently,
GPC has also been used in semiconductor control applications (Lee et al, 2002). Instead of only minimizing the variance of one-step-ahead prediction error, GPC minimizes an objective function which considers several step prediction errors and also the cost of control actions. Since several step prediction errors are considered simultaneously, GPC is not as sensitive as minimum variance control to the system model, especially to the dead time of the model.

In this dissertation, GPC is used for CIGS thin film deposition process control because of its robustness to the dead time variability and estimation uncertainty, effectiveness in handling non-minimum phase systems, and potential in combining SPC through supervisory strategies. The above control procedure which involves modeling and control law design is usually called Automatic Process Control (APC). Besides of APC, as it is mentioned before, another popularly used process control method is Statistical Process Control (SPC). SPC method can trace back to the work of Shewhart in the 1930s (Shewhart, 1931). Currently, it is the most popular tool used in industry for process quality control (Montgomery, 2001).

Since both APC and SPC were widely used in process control, the potential of integrating them together was studied by many researchers (Sachs et al, 1995; Jiang and Tsui, 2002; Lin and Adams, 1996; Ruhhal, Runger and Dumitrescu, 2000; Tsung and Tsuf, 2003). In the integration, SPC is used to monitor the process while APC is used to control the process if SPC shows that the process should be taken action to control. APC modifies the behavior of the output by adjusting the input value. However, abrupt and large shifts in the output indicate potential failures in the process. So the output cannot
be simply compensated by only using APC controller to adjust the input value. The
detailed discussion about the difference and connection of SPC and APC were given by
Vander et al (1992). Some integration strategies, for example, how to use SPC to monitor
the model residuals, were studied by Box and Kramer (1992), Box, Jenkins and Reinsel
monitoring the control input instead of process output to improve the chance of failure
detection, for example, Montgomery et al (1994), Tsung et al (1999), and Jiang and Tsui
(2002). However, almost all the current APC and SPC integration methods only focus on
how to use SPC to monitor the APC process. Few papers discuss the idea of how to use
the information from SPC to adjust the control law of APC and thus control the process
more efficiently. Similar ideas can be found in the paper of Sachs et al (1995). In this
paper, a SPC and APC integration strategy which considers run to run variation in the
controller was given. The controller provides a framework for controlling a process
which is subject to disturbances such as shifts and drifts. The framework has three
components: rapid mode, gradual mode and generalized SPC. The choice between the
two models is determined by the outcome from generalized SPC. However, the
controller is designed based on a simple EWMA system model and minimum variance
control algorithm. Therefore it can only deal with a simple process with a fixed step
delay. How to integrate SPC and APC for a complex process such as thin film deposition
process with uncertain delay time and uncertain distances is still a challenging problem.

In this dissertation, a supervisory predictive control strategy is designed by
integrating SPC and APC for a complex thin film deposition process. Integrating a
supervisory strategy in a predictive controller is more critical and difficult than integrating it in a minimum variance control based on simple EWMA controller. In predictive control, more steps of disturbance prediction will be used. Therefore, in order to reduce the prediction error, a suitable on-line disturbance monitoring strategy, which not only can monitor but can also predict disturbance should be developed.

Proposed Approach. In Chapter 3, a supervisory Generalized Predictive Control (GPC) combined with Statistical Process Control (SPC) method will be used to reduce the process variability of the thin film deposition process. In the using of supervisory GPC, the deposition process is identified as an ARMAX model. Supervisory strategies, developed from SPC techniques, are used to monitor process changes and estimate the disturbance type and magnitude during production period. Based on the SPC monitoring results, different supervisory strategies are defined to revise the disturbance models and the control law in the GPC to achieve a satisfactory control performance.

The framework of the supervisory GPC is shown in Figure 1.1. Detailed discussions on each module are provided in Chapter 3.

Figure 1.2 Structure of SPC-supervisory GPC
One of the significances of this research is emphasizing the importance of developing supervisory strategies through monitoring the process changes using SPC techniques, and then revising the controller parameters accordingly. This integration of SPC with APC provides great potential for the development of effective controllers in complex manufacturing processes.

1.2.2 Hierarchical ANOVA for Variance Decomposition and Diagnosis in Batch Printing Process

Problem Statement. In a silver screen printing process, which can be considered as an example of a typical batch manufacturing process, products are produced batch-by-batch. Thus, total process variations are generally divided into two categories: one is batch-by-batch variation due to the process variability among different batches, and the other is within-batch variation due to the process variability within a batch. Since different variations are usually caused by different root causes, separation and estimation of each variation component is very critical for determining an effective inherent variation reduction strategy. SPC is popularly used for variation control in batch manufacturing processes. However, most of the SPC methods emphasized reducing variations due to assignable causes by implementing control charts for process monitoring. Therefore, control chart method only focuses on how to maintain the current process variation under the normal condition, but cannot reduce the inherent variation of a process. Different from conventional methods through removal of the assignable causes related process variations, this dissertation focuses more on how to reduce inherent natural process variations based on normal production data.
State of the Art. Process control and variation reduction for batch manufacturing process has been studied by many researchers (Yashchin, 1994; Nomikos and MacGregor, 1995; Kourtis, Nomikos and MacGregor, 1995; Woodall and Thomas, 1995; Mason, Chou and Young, 2001). Since variation of a batch manufacturing process can be generally classified into two categories: batch-to-batch variation and within batch variation, in order to control the total variation, strategies which can separate and thus monitor the different variation components should be used. Otherwise, as pointed by Yashchin (1994), in the presence of batch-to-batch variation, standard $\bar{X}$ and $S$ charts which use the control limits obtained from within batch variation will typically produce an unacceptably high number of false alarms, sometimes making the whole control system useless. In Yashchin's (1994) paper, the total variance was decomposed into several variance components. By monitoring those components individually, the corresponding assignable causes can be further identified. SPC with several components of common cause variability was also discussed by Woodall and Thomas (1995). They also pointed out that when common SPC is used, the estimate of variability used to establish control limits is almost always based on within-batch variability. However, because of the batch-to-batch variation, the control limits are often too narrow and result in an unacceptably high false alarm rate. So they present a generalized in-control model to include multiple components of common cause variability, such as batch-to-batch variability. Multivariate quality control technologies were also used for batch process variation control. Nomikos and MacGregor (1995) used Multivariate SPC charts for batch processes monitoring. In their paper, multi-way principal component analysis (PCA) is used to compress the
information contained in the data trajectories into low dimensional spaces that describe the operation of past batches. Mason et al (2001) applied Hotelling's $T^2$ statistic for batch process quality control. Discussions of necessary adaptations, such as the formulas for computing the statistic and its distribution are also included in their paper. Roes and Does (1995) developed monitoring control charts for a silicon wafer manufacturing process, in which a mix-effect model is used to include the fixed effects of the grinding wafer positions in addition to the nested random effect of process runs. In addition, Runger and Fowler (1998) further developed batch-to-batch control charts with contrasts for semiconductor wafer manufacturing processes. In their approach, the monitored contrasts are formalized based on the engineering knowledge of the site patterns of the potential assignable causes. So, the out-of-control points in the contrast control chart are able to link to the conditions of monitored sites. For multidimensional measurements, multivariate control charts were developed based on PCA (principal component analysis) and PLS (partial least squares) for online monitoring of polymerization reactor process, a stamping process, and a filament extrusion process (Nomikos and MacGregor, 1995; Jin and Shi, 2000, Wurl, et al, 2001).

All the above methods focus on how to develop effective control limits of SPC control charts for batch manufacturing process monitoring. However, it is known that SPC control charts can only maintain the process variation by analyzing "out of control points" caused by assignable causes. They cannot reduce the inherent variation, which is caused by common causes, since their control limits are built based on it (Hamada et al 1993). For further improving the quality of the final product, other methods such as DOE
should be used to reduce the inherent variance of a process. By using the method of DOE, a series of experiments were carefully designed and the corresponding experiment data were collected and analyzed to identify variation components. The difficulty of using DOE is that it need deliberately change the current process in order to obtain data with particular features. Methods of using normal process data to do variance analysis were discussed by Hamada et al (1993). In this paper, a method which uses chart techniques in conjunction with observational studies for continuous variation reduction was proposed. The basic idea of this method is to design a sampling plan that identifies the largest sources of variation, which can be used to guide the action to reduce or eliminate the variation sources. However, there is no systematic method to diagnose the root causes of inherent process variation. In this dissertation, a general hierarchical framework of how to apply ANOVA to automatically assess and diagnose the root causes of inherent process variations will be developed.

Proposed Approach. The proposed variation modeling and analysis method in Chapter 4 is based on the general methodology of ANOVA method. The critical issue of using the ANOVA method based on production data is how to determine appropriate statistical models to describe all decomposed variation components of interest. In this dissertation, a general variation decomposition framework is presented based on statistical nested effect models for testing and estimating typical variance components in batch manufacturing processes. This general framework is based on the decomposition of three typical variation components caused by three main factors in the silver gridline printing process. The three variation components are batch-by-batch variation, sample-
by-sample variation and site-by-site variation. For the significant site variation, three diagnostic contrast components are further defined for root cause identification. The three contrast components are front-back contrast, left-right contrast and diagonal contrast, respectively. When the interested variations are induced by more than three factors, it generally needs more different nested models for various variation component analyses. The use of a full factor decomposition model to determine the needed nested models is suggested in the proposed method, which can help to systematically determine the nested model structures.

Different from the previous research focus, this dissertation aims to develop a general variation decomposition and analysis methodology for assessing and diagnosing inherent variance components from production data in batch manufacturing processes. The proposed method can be used for inherent variation identification and reduction before applying SPC control charts. It is recommended to first devote the efforts to reduce or avoid systematic inherent errors rather than simply accept it as the inherent process variation in the control chart development.

1.2.3 Multiscale Statistical Process Monitoring Using Wavelets Analysis for Autocorrelated Data

Problem Statement. It is known that a process fault may occur at an unknown time with either a single mean/variance change or both. In general, there are different root causes to associate with the mean shift and variance change, which needs different process improvement and control strategies. Therefore, it is always desirable to have an effective process monitoring system that not only can detect process changes but also can
separate the mean shift and variance change. A multiscale process monitoring based on wavelets analysis is developed for this purpose to simultaneously monitor process mean shift and variance change.

In general, in-process sensing data collected from an automatic manufacturing process are autocorrelated. For example, in the CIGS thin film deposition process, Se pressure signal and thickness measurements are autocorrelated over time. It is known that the conventional control charts do not work well for autocorrelated data, specifically, those control charts will give misleading results in the form of too many false alarms. Moreover, those conventional control charts mainly focus on single fault (either mean shift or variance change). Their monitoring performance will be severely deteriorated when both faults occur at the same time.

It is known that the variance change is usually caused by a severe process failure or severe sensor failure, which requires immediate actions to check and remove the associated root causes if they exist. For the process mean shift, a feedback control or process adjustment can be used to compensate the process change. Therefore, it is desirable to develop an effective process monitoring method for autocorrelated data to simultaneously detect mean and variance change separately, which can help to make an effective decision for process improvement.

State of the Art. The commonly used process monitoring methods for autocorrelated processes are model-based methods. One popular model-based method is special cause chart (SCC) (Alwan and Roberts 1988; Wardell, Moskowitz and Plante 1994). The principal of the SCC method is to whiten an autocorrelated process based on
the process model and then use traditional control charts to monitor the whiten residuals. It is known that the performance of SCC chart is very sensitive to the model estimation accuracy (Apley and Shi, 1999). In order to avoid the model estimation problem, EWMAST chart or Direct-EWMA chart (Schmid 1997; Zhang 1998; Adams and Tseng 1998; Lu and Reynolds 1999a, b) was developed especially for small process mean shift detection. Direct-EWMA charts apply the exponentially weighted moving average (EWMA) statistic directly to the autocorrelated data without the need of identifying process models. It is shown that Direct-EWMA charts are very effective to small mean shift detection in a slow time varying processes. In fact, EWMA model is a special case of ARMA(1,1) model with $\phi_1 = 1$, and $\theta_1 = 1 - \lambda$. When a process is described by a high order ARMA $(n, m)$ $(n \geq 2)$ model having a high frequency spectrum, the effectiveness of EWMA will be significantly affected by the process variance, which will be shown in Chapter 5.

Process response is usually contaminated by noise which complicates the process change detection. An algorithm, Kalman filter, was developed by R.E. Kalman (1960) for solving this problem. The Kalman filter is a model based method. It has been used by mathematicians and physical scientists for navigation, missile guidance, satellite tracking, and other applications where short-term prediction and adjustments are required. Roush et al. (1992) used Kalman filter to detect changes in poultry production responses. In their paper, the one step prediction residual of Kalman filter was used as the variable under monitoring. However, as pointed by Basseville and Nikiforov (1993), the asymptotic time invariance and stability of the Kalman filter are important for applications. Therefore,
how to build a sufficient model which can describe the process is very critical. Similar with other model based methods, in order to obtain efficient monitoring results, the parameters in the Kalman filter should be estimated accurately. Moreover, without using the information from frequency domain, it is difficult to distinguish process changes from measurement error changes.

For process variance change detection, an $R$ chart or $S$ chart is usually developed for large variance change detection using multiple observations (Montgomery, 2001). For a single observation monitoring, individual range charts can also be developed accordingly (Montgomery, 2001). For small variance change detection, a standard deviation CUSUM (denoted as SD-CUSUM in this dissertation) chart was developed by Hawkins (1981, 1993) using individual observation. However, this SD-CUSUM chart is sensitive to both mean and variance changes, which causes the difficulty in separation of process faults whether due to mean shift or variance change. Moreover, it will be shown that the performances of these control charts are dramatically degraded in an autocorrelated process. So, there is a need to develop new control charts for autocorrelated process monitoring.

Besides the drawbacks mentioned above, conventional SPC control charts are best only for detecting certain types of changes. For example, a Shewhart chart can detect large change quickly, but is insufficient in detecting small mean shift. CUSUM, MA and EWMA charts are better at detecting a small mean shift, but may be slow in detecting a large shift. In fact, Shewhart charts monitor data at the scale of the sampling interval, which is the finest scale, MA and EWMA charts represent data at a coarser scale
determined by the parameter of $\lambda$ in EWMA and window length in MA respectively, and CUSUM charts monitor data at the scale of all the measurements, which is the coarsest scale (Bakshi 1998).

Recently, wavelet-based SPC methods were used by some researchers. (Luo, et al. 1999; Bakshi, 1998 and 1999; Jin and Shi, 2001; Lada et al. 2002; Ganesan, et al. 2003). Luo et al. (1999) applied PCA to the wavelet detail coefficients at several levels and then $T^2$ and $Q$ statistics were used to detect possible faults. Bakshi (1999) proposed a multiscale analysis method which can exploit the ability of wavelets to extract events at different scales, compress deterministic features in a small number of relatively large coefficients. The wavelet coefficients at all levels were monitored by SPC with different control limits, which may cause significant increase of the total false alarm rate. Jin and Shi (2001) proposed a feature extraction and fault detection method for waveform signal monitoring. In their paper, $T^2$ chart was used to monitor the significant coefficients. Lada et al. (2002) proposed a bootstrapping procedure based on wavelet coefficients for fault detection of functional data. Ganesan et al. (2003) proposed a moving window-based strategy for online implementation of the wavelet monitoring strategy. The energy of the detail coefficient at each level was used as an indicator for process change detection. Almost all the methods outlined above focus on how to extract features from noise corrupted data and then applied multivariate SPC, such as $T^2$ and $Q$ chart to the extract features. None of the research discuss the how to choose an optimal level for the efficient detection of mean shift and variance change simultaneously. In this dissertation, a multiscale monitoring strategy was proposed to fully utilize the frequency and time
domain characteristics of wavelets analysis. A set of control charts were developed for simultaneously monitoring mean shift and variance change in an autocorrelation process.

**Proposed Approach.** A multiscale process monitoring using wavelets analysis will be developed to fully utilize the process information in both time and frequency domains. Three typical process faults in autocorrelated processes will be investigated in this dissertation, which includes process variance change, measurement error variance change, and process mean shift. In Chapter 5, three control charts to monitor the selected wavelet coefficients will be developed for detecting those process changes specifically. Based on the different frequency characteristics of scale coefficients and detail coefficients, an EWMA chart is developed to monitor the selected scale coefficients for process mean shift detection, while two SD-CUSUM charts are developed to monitor the selected detail coefficients for detection of process variance change and measurement error variance change separately. An optimal selection of the decomposition levels is investigated by considering system response information in the frequency domain. It is shown that the proposed SD-WCUSUM chart to monitor the selected detailed coefficients is more robust to measurement errors in the process variance change detection when compared with SD-CUSUM charts on the direct signal monitoring. Moreover, the proposed WEWMA chart is more powerful in the small mean shift detection especially when a system has a high frequency spectrum. It will be seen that the proposed multiscale control chart is a model-free approach, which aims to avoid the model estimation problem in the SCC chart. The performance of the proposed monitoring charts is comparable with SCC chart that uses a true model to whiten autocorrelated data. Moreover, it will show that multiscale
monitoring on the detail coefficients will be insensitive to process mean shifts, which can help expedite process diagnosis for making an effective process improvement decision.

1.3 OUTLINE OF THE DISSERTATION

Three fundamental research issues related to process monitoring, diagnosis and control will be addressed in this dissertation. The outline of the dissertation is given in Figure 1.3.

In Chapter 1, three research issues that will be addressed in this dissertation are introduced. The corresponding state of art and the proposed research approaches are briefly reviewed. In Chapter 2, the related process background of a solar cell manufacturing process will be given. For those three research issues presented in Chapter 1, the details of each proposed methodology will be discussed in Chapter 3, 4 and 5 respectively.

Figure 1.3 Flow chart of dissertation
The Chapter 3 will discuss the integration of SPC and APC for supervisory GPC development in the CIGS thin film deposition process which has a large time delay and different process change patterns occurring at unknown times. Chapter 4 will discuss a general framework for hierarchical variance decomposition using ANOVA which emphasizes the inherent variation diagnosis and reduction based on normal production data. Chapter 5 proposes a new multiscale SPC monitoring system for detecting and separating the process mean shift and variance change. Finally, in Chapter 6, conclusions will be drawn and future research will also be discussed.
Chapter 2

Process Background and Variable Description

In this chapter, the background of solar cell module manufacturing process is briefly reviewed. Two processes, CIGS thin film deposition process and silver screen printing process will be discussed in detail. Process variables which are related to the following chapters will also be explained.

This chapter is organized as follows: First in Section 2.1, the brief review of the overall process is given. Then in Section 2.2 and 2.3, the CIGS and Silver screen printing process are discussed in detail.

2.1 Description of Process Flow

The whole solar cell manufacturing process can be divided into three major stages including thin film deposition, panel printing and module assembly. During the deposition stage, different layers of materials are deposited on thin flexible substrate. These layers, all together, make the photovoltaic film, which acts as a core part of solar cell. These layers are deposited in the sequence of 1) Molybdenum, 2) CIGS (Copper, Indium, Gallium, Selenium), 3) CDS (Cadmium Sulphate), 4) ITO (Indium Tin Oxide), which is shown in Figure 1.1. The deposited thin film is in the form of rolls.

The first deposition process has the following steps.

1) Stainless steel roll is mounted in the sputtering chamber for deposition of molybdenum.
2) After molybdenum deposition is completed, the roll is stored in a dry box, from where it is picked up for CIGS deposition. There are 3 zones with different material sources in a CIGS deposition chamber, which is used to evaporate CIGS materials to the molybdenum coated roll.

3) CIGS coated roll is then taken to CDS deposition area. CDS is deposited by chemical bath process.

4) The last deposition process is ITO. It is carried out in a similar chamber as that of molybdenum.

5) After these deposition processes are completed, samples are taken for quality checks which are called diode samples and the roll is passed to the printing process.

The second stage is panel printing process which includes the following steps.

1) A deposited thin film roll is first put into an assembly dry box after ITO is coated. Then it is cut into pieces 1 foot long (which are called panels). They are kept in a tray for silver printing.

2) Panels are printed with silver gridlines in a screening machine. Then they are stacked on a trolley and cured thermally.

3) Additional coatings are screen printed onto the panels to electrically isolate individually cell and mechanically join them.

The third stage is the module assembling process which includes the following steps.

1) Cells are cut from the panel (14 cells in each panel) and sent for sorting.
2) Cells are sorted as per category, which have been assigned through computer software on the basis of efficiency and voltage range.

3) Conductive epoxy is applied on the silver bus line for better contact between cells after assembly. After that cells are stuck together to form one sub-module. The epoxy is thermally cured in an oven.

4) These modules are tested for efficiency in a machine.

5) Sub-modules are kept together for assembly to form a module.

6) These modules are checked for efficiency and then sent for final product reliability testing.

One example of the final product is given in Figure 2.1.

![Figure 2.1 Example of final solar cell product](image)

In this dissertation, the research is mainly conducted on two critical processes which are CIGS deposition process and Silver Printing process. In the following subsections, the details of process variables and quality characteristics in these two processes are given.

### 2.2 CIGS THIN FILM DEPOSITION PROCESS

CIGS thin film deposition is conducted in a vacuum chamber. Different effusion
sources are evaporated into the stainless steel web through isolated containers. Figure 2.2 shows the distribution of effusion sources in the vacuum chamber, that is, different material sources are located in different zones respectively. With a careful design of the roll speed and an appropriate design of source locations, the web motion creates a controllable flux profile at the substrate. Sources are poured into source containers. In general, temperature is a critical factor affecting the effusion rate and evaporation deposition in thin film deposition processes.

![Figure 2.2 CIGS Thin film deposition process](image)

From Figure 2.2, we can see that on the top of the source cell, there is a cover with three nozzles for evaporation. The evaporated gas will be deposited on the substrate roll. Heater goes through the cover for heating the source materials. Thermocouples are used to measure the temperature. Substrate Roll (or web) moves from the left side through zone 1, zone 2 and zone 3 to the right side as shown in Figure 2.2. Each material is evaporated from its source cell and deposited on the passing web. Their deposition rates are adjusted by controlling source temperatures. Process temperatures are set to given points. The XRF sensor in Figure 2.2 is used to measure the thickness of the thin film, which will be used in the feedback controller.
In order to develop an automatic feedback controller in a thin film process, a process model which can represent the relationship of an in-situ measurable product quality and the controllable process variables is required. In general, after the process setup variables (such as chamber pressure, chamber environmental temperature, and web moving speed) achieve their target values, the effusion source temperature, which is the direct and most sensitive process variable affecting the physical effusion properties during the production run, will be used as the controllable input variable of the process model. The output variable of the process model is deposition film thickness, which is measured twelve times during each two minutes at the end of the decomposition chamber. Those variables will be used in Chapter 3 for system modeling and process monitoring and control. The diagram of the control system of the CIGS process is given in Figure 2.3.

![Figure 2.3 Control system of the CIGS deposition process](image)

2.3 PANEL SILVER PRINTING PROCESS

A silver gridline acts as a connector to conduct electrical current through solar cells. These gridlines are printed on the cell material by a printing machine. Liquid silver is spread over the rubber print in grid format. By the physical properties, these silver lines have resistance, which can be measured to analyze the quality and uniformity of printing.
This resistance is tested with four point probes. The Printing Machine is given in Figure 2.4.

![Silver Screen Printing Process](image)

Figure 2.4 Silver Screen Printing Process

The printing operation is performed panel by panel and each panel is a product unit called *sample*. At the beginning of printing every panel, each panel is put on the tabletop of the printing machine, and the air suction pores distributed on the machine tabletop are used to hold the panel during printing operation. The squeegee moves from front to back to press conductive silver ink through the rubber screen to print gridlines on the panel. Then, the flood bar moves from back to front to spread ink over the rubber screen ready for printing next panel. Generally, the printing tools of a rubber screen and squeegee used in a Printing Machine need to be taken out for cleaning after finishing each production run and set up again for the next product run. A production run cycle under the same tool setup is considered as one batch production.

The resistance of printed silver gridlines is a major concern of the product quality because it significantly affects the solar panel efficiency. It is known that the gridline resistance is affected by the width and height of the printed gridlines. So, the spatial uniformity of the width and the height of the printed gridlines is a critical issue for
reducing solar panel efficiency variation due to the printing operation. For inspection of the uniformity, four resistance measurements ($M_1$, $M_2$, $M_3$, $M_4$) are taken at four corners of each panel as shown in Figure 2.4. The variation among these four measurement positions called *sites* is used to represent the spatial uniformity of site-by-site variation.

![Figure 2.5 Measurement locations on a printed panel](image)

2.3.1 Description of Process Variables

The important process variables in the printing process are classified as: (1) tool conditions (screen’s tension and missing lines, squeegee’s edge shape and hardness), (2) tool setup position (the gap and relative orientation angle between the squeegee and the rubber screen, the alignment of the rubber screen and the squeegee relative to the machine tabletop, and (3) conductive ink material (ink viscosity and composition). Table 2.1 provides a summary of these process variables and whether they have an effect on each of process variation components. Since the tool condition degradation is generally very slow, it is reasonable to assume the tool condition is not changed within one production run.
<table>
<thead>
<tr>
<th></th>
<th>Batch-by-batch variation</th>
<th>Site-by-site variation</th>
<th>Sample-by-sample variation for a given site within a run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tool condition</td>
<td>Yes</td>
<td>Yes</td>
<td>Ignore tool degradation within a run</td>
</tr>
<tr>
<td>Tool setup</td>
<td>Yes</td>
<td>Yes</td>
<td>Assume tool position fixed within a run</td>
</tr>
<tr>
<td>Conductive ink material</td>
<td>Yes</td>
<td>Ignore ink variation within a run</td>
<td>Assume no new ink added within a run</td>
</tr>
</tbody>
</table>

2.3.2 Selection of Interested Variation Components

From Table 2.1, it can be seen that the batch-by-batch variation and site-by-site variation are of the major interests for process variation reduction because they are affected by many assignable process variables or working conditions. The sample-by-sample variation is considered as the inherent process variation especially sampling at a given site and within a production run. Thus, from the general variation decomposition structure, Factor batch (Factor A) and Factor site (Factor B) are considered as the main factors contributing to the process variations. The possible nested factors can be either A(B) or B(A). Based on the possible existence of potential root causes, three variation components are selected which are contributed by Factor A as well as the nested Factors A(B) and B(A). The details of each variation component and the associated root causes are discussed as follows:

1. **Batch-by-batch variation** due to the variability of the average of four sites over different runs (Variation $Q_1$ due to Factor A): It mainly represents the process variability over different batches, such as ink materials, different setup positions of the squeegee orientation angle, the distance of the rubber screen from the printing machine tabletop, the tool condition change due to squeegee wear and the loosening of rubber screen
tension. The process improvement strategy should enhance the inspection of ink property, tool setup, and tool condition at the beginning of each run.

(2) **Batch-by-batch variation** at any of four sites over different runs (Variation $Q_2$ due to the nested factor $A(B)$): It reflects the repeatability of tool setup position at each site over different runs. So, $Q_2$ is caused by different tool setup errors and/or non-uniform tool degradation at each site over different runs. Thus, the reduction of this variation should reduce operator induced variability and tool setup variability over different runs.

(3) **Site-by-site variation within a production run** (Variation $Q_3$ due to the nested factor $B(A)$): For each production run, the site-by-site variation is mainly caused by the alignment accuracy of the squeegee position relative to the screen positions, and the screen orientation relative to the printing machine tabletop. The strategy for reducing such a within-run site-by-site variation is to improve the tool setup inspection accuracy at each site and check tool performance uniformity over the printing area.

For the above listed three variation components of interest, the ANOVA method will be applied in Chapter 4 to model, test, and estimate these variation components. Then, the effective process improvement efforts can be made accordingly to remove or reduce the corresponding significant variation root causes.

From the above description, it can be seen that CIGS thin film deposition and printing processes represent two typical types of manufacturing processes. CIGS deposition process can be generally taken as a continuous manufacturing process having common characteristics in chemical industry. The printing process can be taken as a
general batch manufacturing process which has similarity with the wafer batch manufacturing processes. So, the methodologies developed in this dissertation, based on CIGS thin film deposition and printing processes, are also expected to be applied in other similar processes in which the measurement data are autocorrelated and process changes include different patterns of mean shift and variance change at an unknown time.
CHAPTER 3

SPC BASED SUPERVISORY GENERALIZED PREDICTIVE CONTROL OF THIN FILM DECOMPOSITION PROCESS

This chapter presents a Supervisory Generalized Predictive Control (GPC) by combining GPC with Statistical Process Control (SPC) for the process control of thin film deposition process. In the supervised GPC, the deposition process is identified as an Auto-Regressive, Moving Average with eXogenous input (ARMAX) model for each production run and a GPC is applied to feedback in situ thickness sensing data for thickness control. Supervisory strategies, developed from SPC techniques, are used to monitor process changes and estimate the disturbance magnitudes during production. Based on the SPC monitoring results, different supervisory strategies are defined to revise the disturbance models and the control law in the GPC to achieve a satisfactory control performance.

This chapter is organized as follows: In Section 3.1, the ARMAX model used for feedback control will be given. In Section 3.2, based the ARMAX model, the corresponding Generalized Predictive Control algorithm is discussed. Section 3.3 addresses the strategies of how to integrate SPC and GPC together under different type of disturbances. A case study is given in Section 3.4. Finally, Section 3.5 is the conclusion and future work.
3.1 PROCESS MODELING USING ARMAX MODEL

In the start up period of each production run, the temperature of each container is manually increased step-by-step with each step having more than eight sampling intervals. Those inputs, together with corresponding film thickness responses, are used to model each run of the thin film deposition process. Thus, run-to-run variations are reflected on the model parameter change. An ARMAX \((n_a, n_b, n_c, n_d, d)\) model with disturbance is used to describe the thin film deposition process,

\[ A(q)y_t = B(q)u_{t-d} + C(q)f_t + D(q)e_t \]  

(3-1)

where \(y_t\) and \(u_{t-d}\) are system response of the film thickness at time \(t\) and control input of the corresponding source temperature at time \(t-d\) respectively, \(d\) is the dead time of the process, \(A(q)\), \(B(q)\), \(C(q)\), and \(D(q)\) are polynomial functions with the orders as \(n_a\), \(n_b\), \(n_c\), and \(n_d\) respectively, and \(q^{-1}\) is the backshift operator with \(q^{-1}y_t = y_{t-1}\).

Here, \(A(q) = 1 + a_1 q^{-1} + ... + a_n q^{-n_a}\), \(B(q) = b_1 q^{-1} + ... + b_n q^{-n_b}\), \(C(q) = c_1 q^{-1} + ... + c_n q^{-n_c}\), and \(D(q) = 1 + d_1 q^{-1} + ... + d_n q^{-n_d}\). \(f_t\) represents the disturbances that are possibly existing in the process either as a spike, a mean shift, or a linear drift function, \(e_t\) is an independently identically distributed (i.i.d) zero-mean sequence of modeling residuals with \(e_t \sim N(0, \sigma_e^2)\). The model parameters in Eq. (3-1) can be estimated using least squares method if \(D(q)\) is known. In general case, if \(D(q)\) is unknown, the Two Stage Methods described in Section 10.4 of reference Ljung (1999) should be used. In our case, \(D(q) = 1\), (see the case study section), so least squares method is directly used for parameter estimation. Eq. (3-1) also can be rewritten as
Denote $\frac{B(q)}{A(q)} = \sum_{j=0}^{m_u} g^u_j q^{-j}$, $\frac{C(q)}{A(q)} = \sum_{j=0}^{m_f} g^f_j q^{-j}$, and $\frac{D(q)}{A(q)} = \sum_{j=0}^{m_e} g^e_j q^{-j}$. $m_u$, $m_f$, and $m_e$ are the maximum order of each corresponding item. They are decided based on the impulse response function $g^s_j$ from each input item to the output $y$. That is, when $j > m_s$, $|g^s_j| < \eta_0$ ($s=u,f,e$). An example of how to decide the threshold $\eta_0$ will be given in the case study part. Using impulse functions, Eq. (3-2) can be rewritten as:

$$y_t = \sum_{j=0}^{\min(i-D,m_u)} g^u_j u_{t-j} + \sum_{j=0}^{\min(i,m_f)} g^f_j f_{t-j} + \sum_{j=0}^{\min(i,m_e)} g^e_j e_{t-j}, \quad (3-3)$$

### 3.2 PROCESS CONTROL USING GENERALIZED PREDICTIVE CONTROL ALGORITHM

The main reasons for choosing the GPC (Please see Appendix A for detail) for the deposition process control in the research are:

- **Dead time estimation uncertainty:** As described in the process overview, the single thickness output $y_t$ is a compound effect of several source cells located at different places. Since there is no sensor to measure each cell effusion, the dead time cannot be determined for each cell separately. As a result, the dead time used in the controller design is a combined effect of several sources. Moreover, the web speed variation also leads to the dead time variation. So, without accurately knowing the dead time, it is difficult to design a controller using the PID or minimum variance control.
control strategy. Thus, it is desirable to design a generalized predictive controller to consider the possible range of multiple dead time values.

- **Unknown and time-varying disturbance patterns:** During production, different disturbance patterns, such as mean shift, slow drift, or spike may occur. So, an adaptive control strategy should be designed to reflect the difference of the disturbance models. In recent years, research achievements on using EWMA models to estimate the time-varying mean shift or slow linear drift have demonstrated a great success for controller design (Del Castillo and Hurwitz 1997; Moyne, et al, 2000). However, most of those methodologies assume that disturbances exist over all production time with predefined disturbance model structures. Thus, EWMA models are used to continuously estimate the unknown or time-varying disturbance model parameters. Sachs et al (1995) investigated a run-to-run controller design by characterizing run-to-run variations as two different modes, that is, a slow mode and a fast mode. Different from other methods, Sachs et al assumed that a slow mode disturbance exists over all production time. The disturbance is estimated by EWMA and continuously compensated in the controller. The compensation of the fast mode disturbance is only conducted when a large shift is detected using X-bar monitoring control chart. Therefore, the SPC monitoring technique provides a supervisory strategy to online determine the disturbance model structure. However, all those run-to-run control methodologies are developed for a minimal variance controller, which is limited to an assumption of the process having a fixed step dead time. Therefore,
they cannot be applied in the thin film deposition process in which a large and uncertain dead time exists.

The objective function (performance criterion) of the GPC control is defined as:

\[ J = \sum_{i=d}^{N+d} q(i) (\hat{y}_{t+i|t} - y^*_i)^2 + \sum_{i=0}^{N} r(i) u_{t+i}^2 \]  

(3-4)

where \( y^*_i \) is the reference (or desired target) output at the end of \( (t+i) \) th sample interval, \( q(i) \) and \( r(i) \) are the weighting coefficients, \( N \) is the sliding horizon for the output prediction and input series, and \( \hat{y}_{t+i|t} \) is the \( i \) th step ahead prediction made at the beginning of \( r \) th time interval, which is obtained based on Eq. (3-3) using the conditional expectation as:

\[
\hat{y}_{t+i|t} = E(y_{t+i|t}) = \sum_{j=0}^{i-d} g_j^y u_{t+i-d-j} + \sum_{j=i-d+1}^{\min(t+i-m_y)} g_j^y u_{t+i-d-j} + \sum_{j=0}^{\min(t+i-m_f)} g_j^f f_{t+i-j} + \sum_{j=i}^{\min(t+i-m_e)} g_j^e e_{t+i-j}
\]  

(3-5)

\( f_{t+i-j} \) \( (j=0, 1, \ldots, i-1) \) is the prediction of the deterministic disturbance function made at time \( t \) according to the discovered disturbance function \( f_t \) at time \( t \). For the random error \( e_t \), the conditional expectation under the current time \( t \) is: \( E(e_{t+i|t}) = 0 \) and \( E(e_{t-i|t}) = e_{t-i} \) \( (i \geq 0) \). Thus, it results in \( E\{ \sum_{j=0}^{i} g_j^f e_{t+i-j} \} = \sum_{j=i}^{i} g_j^f e_{t+i-j} \).

The basic concepts of the GPC strategy can be summarized as follows: At current time \( t \), an optimal control value \( u_t \) is obtained by solving the optimization problem defined in Eq. (3-4). This value will be the input for the \( (t+1) \) th time interval. By using
the sliding horizon window of every $N+1$ step predictions, the control problem is simplified from a dynamic programming problem to a static optimization problem (Astrom and Wittenmark, 1995).

In order to solve the predictive control problem, the objective function in Eq. (3-4) is equivalently represented by a general vector and matrix forms as:

$$J = \left( \hat{Y} - Y^* \right)^T Q \left( \hat{Y} - Y^* \right) + U_1^T R U_1$$

(3-6)

where $Q = \text{diag}[q(N + d), q(N + d - 1), \ldots, q(d)]$ and $R = \text{diag}[r(N), r(N - 1), \ldots, r(0)]$ are weighting coefficients, which are diagonal matrices. Vector $\hat{Y}$ represents all $N+1$ step predictions in the sliding window, which is represented based on Eq. (3-5) as:

$$\hat{Y} = G_1 U_1 + G_0 U_0 + G_r F + G_e E$$

(3-7)

where each vector of $\hat{Y}$, $U_1$, $U_0$, $F$, $E$ are defined as follows: vector $\hat{Y} = [\hat{y}_{t+d+N} | \hat{y}_{t+d+N-1} | \ldots | \hat{y}_{t+d+1} | \hat{y}_{t+d}]^T$ represents all predicted outputs based on time $t$ within the sliding window. Corresponding to $\hat{Y}$, the future optimal control input vector is denoted as $U_1 = U_1^{op} = [u_{t+N} \ u_{t+N-1} \ \ldots \ u_{t+i} \ \ldots \ u_t]^T$ based on Eq. (3-5), which will be obtained in Eq. (3-9) late. It should be noted that only single $u_i$ in $U_1^{op}$ will be used to control the process at time $t$. Similarly, from Eq. (3-5), the previously used control input vector is denoted as $U_0 = [u_{t-1} \ \ldots \ u_i \ \ldots \ u_{t-min(t, m_u)}]^T$, which is ordered backward from time $t-1$ to the maximum steps of $m_u$ defined in Eq. (3-3). The effect of those control inputs on the future system outputs is determined by the dynamic system memory on the historical control inputs. Also in Eq. (3-7), vector
\[
F = [f_{t+d+N}, f_{t+d+N-1}, \ldots, f_{t+d+N-\min(t+d+N, m_f)}]^T
\]
is used to represent the deterministic disturbances from time \( t+d+N \) backwards to the maximum steps of \( m_f \) defined in Eq. (3-3), and vector \( E = [e_t, e_{t-1}, \ldots, e_{t-\min(t,m_u-d)}]^T \) is used to denote the historical random residuals. Those residuals can be calculated up to the current time \( t \) using different Eqs. (3-32), (3-34) and (3-35) corresponding to a mean shift, a linear drift, and a spike error respectively, which will be discussed in Section 3.3.3. If \( m_u-d<0 \), we set \( E = 0 \), which means the historical random disturbances do not influence the future output.

Corresponding to each vector of \( U_1, U_0, F, \) and \( E \) used in Eq. (3-7), the matrixes of \( G_1, G_0, G_f, \) and \( G_e \) are used to reflect the different dynamic response weights respectively, which are defined based on Eq. (3-5) and Eq. (3-7):

\[
G_1 = \begin{bmatrix}
g_0^u & g_1^u & \cdots & g_N^u \\
0 & g_0^u & \cdots & g_{N-1}^u \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & g_0^u
\end{bmatrix}_{(N+1) \times (N+1)}
\]
reflects how future control input \( U_1 \) affects output \( \hat{Y} \);

\[
G_0 = \begin{bmatrix}
g_{N+1}^u & g_{N+2}^u & \cdots & g_{\min(t,m_u)+N}^u \\
g_N^u & g_{N+1}^u & \cdots & g_{\min(t,m_u)+N-1}^u \\
\vdots & \vdots & \ddots & \vdots \\
g_1^u & g_2^u & \cdots & g_{\min(t,m_u)}^u
\end{bmatrix}_{(N+1) \times (\min(t,m_u))}
\]
reflects the effect of dynamic system memory of the historical control input \( U_0 \) on the outputs \( \hat{Y} \);
$G_f = \begin{bmatrix}
g_0^f & g_1^f & \cdots & g_{\min(d+t+N, m_f)}^f \\
0 & g_0^f & \cdots & g_{\min(d+t+N, m_f)-1}^f \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & g_{\min(d+t+N, m_f)-N}^f \\
\end{bmatrix} \text{ reflects the effect of the deterministic disturbance } F \text{ on output } \hat{Y};$

$G_e = \begin{bmatrix}
g_{N+d}^e & g_{N+d+1}^e & \cdots & g_{\min(t+d, m_e)+N}^e \\
g_{N+d-1}^e & g_{N+d}^e & \cdots & g_{\min(t+d, m_e)+N-1}^e \\
\vdots & \vdots & \ddots & \vdots \\
g_d^e & g_{d+1}^e & \cdots & g_{\min(t+d, m_e)}^e \\
\end{bmatrix} \text{ reflects the effect of the dynamic system memory of the historical random error } E \text{ on the future output } \hat{Y}, \text{ which is effective only when } m_e - d \geq 0 \text{ is held.}$

It should also notice that when the production time is larger enough, the dimension of those matrixes of $G_0$, $G_f$, and $G_e$ will be constrained only by the maximum order of each corresponding impulse function as defined in Eq. (3-3). By solving the optimization problem with $\frac{dJ}{dU_1} = 0$, one has:

$$\frac{dJ}{dU_1} = 2[G_1^T Q(G_1 U_1 + G_0 U_0 + G_f F + G_e E - Y^*) + R U_1] = 0 \quad (3-8)$$

The optimal predictive control law, $U_1^{op}$, is obtained by solving Eq. (3-8) as

$$U_1^{op} = -[G_1^T Q G_1 + R]^{-1} G_1^T Q[G_0 U_0 + G_f F + G_e E - Y^*] \quad (3-9)$$

Although the vector of all future inputs from $u_t$ to $u_{t+N}$ is provided in $U_1^{op}$, only $u_t$ will be used to control the process at time $t+1$. At the next time $t+1$, another
optimization will be conducted based on the new observation of \( y_{t+1} \), which is used to obtain a new optimal control vector \( U_t^{op} = [u_{t+1}, u_{t+2}, \ldots u_{t+N+1}] \). The newly obtained \( u_{t+1} \) will be used as a control input at time \( t+2 \). This strategy is repeated until the end of the production run.

**Remark 1: Prediction horizon \( N \) selection**

The principle in the sliding horizon selection is to have its lower limit be equal to the dead time of the system. If the dead time is estimated with uncertainty, or the dead time varies during the production, the sliding horizon should be set as the lower bound of the dead time estimation or variation range. The upper limit of the sliding horizon should be larger than, or equal to, the settling time of the system responses. The settling time, \( T_s \), is defined as the time when the absolute value of the impulse response of the process is always below a small threshold value of \( \eta_0 \) as time goes to infinity. \( T_s \) can be obtained by investigating the impulse response of the model as

\[
|g_j^u| < \eta_0 \quad \forall j > T_s \tag{3-10}
\]

The determination of \( \eta_0 \) is decided by the modeling accuracy requirement, which can be set based on \( m_u \). An example of how to choose \( \eta_0 \) and the corresponding \( m_u \) will be given in the case study part. In addition, a smaller \( N \) value should be selected if the process under more significant disturbances or model uncertainty.

**Remark 2: Weighting coefficients \( R \) selection**

Parameter \( R \) has an impact on the process stability and tracking error. Correct selection of \( R \) value is an effective way to reach a satisfactory performance for
controlling non-minimum phase processes. In general, a larger $R$ value leads to a more stable system, less overshoot, slower response, and larger tracking errors. Thus, if a disturbance occurs, a smaller $R$ value is preferred under the constraints of stability and acceptable overshoot. In this way, a fast tracking performance can be achieved especially for a fast drift disturbance.

3.3 SUPERVISORY PREDICTIVE CONTROL STRATEGIES

Various disturbances exist in the thin film deposition process as described in chapter 1. Three typical disturbance patterns (mean shift, linear drift or ramp, and spike) will be studied in the development of a supervisory strategy in the GPC design.

3.3.1 CUSUM Charts for Detection and Estimation of Shift and Drift Disturbances

In order to testify the disturbance patterns and their impacts on the control performance, a statistical Cumulative-Sum (CUSUM) monitoring chart is used to detect and estimate the disturbance $f_t$ with a mean shift or linear drift function. The estimated function $f_t$ will be further used in Eq. (3-9) to revise the future control adjustment for compensation of the corresponding disturbance.

In the following section, a brief review of CUSUM chart is given to introduce the notations used in the chapter. The detailed discussion of the CUSUM chart technique can be found in Montgomery (2001).

For a CUSUM chart, it is constructed by a positive and negative CUSUM statistics as:

$$ S_H(t) = \max[0, x_t - (\mu_0^* + K) + S_H(t-1)] $$

(3-11)
\[ S^*_L(t) = \max[0, (\mu_0^* - K) - x_i + S^*_L(t-1)] \]

where \( x_i \) is the monitoring variable with the mean of \( \mu_0^* \) under the in-control condition.

The starting values of \( S^*_H(t) \) and \( S^*_L(t) \) are \( S^*_H(0) = S^*_L(0) = 0 \). \( K = |\Delta/2| \) and \( \Delta \) is the mean shift of \( x_i \) interested to be detected.

An out of control point is indicated at time \( k \) if \( S^*_H(k) > h \) or \( S^*_L(k) > h \), where \( h \) is a design parameter as the control limit of a CUSUM chart. In general, the values of \( h \) and \( K \) determine CUSUM chart sensitivity to disturbances and impact the Average Run Length (ARL) of the chart (Montgomery, 2001).

If an out of control point is observed at time \( k \), \( N^+ \) (or \( N^- \)) records the number of consecutive samples since the upper-side cusum \( S^*_H(k) \) (or lower-side cusum \( S^*_L(k) \)) have a positive value. Thus, the occurrence time of the mean shift is detected as \( s+1 \), and \( s = k - N^+ \) for an upward shift or \( s = k - N^- \) for a downward shift. The sum of mean shifts over \( N^+ \) (or \( N^- \)) steps are estimated by:

\[
\begin{align*}
\text{if } S^*_H(k) > h, & \quad \sum_{i=1}^{N^+} \mu^* = N^+ \cdot (\mu_0^* + K) + S^*_H(k) \\
\text{if } S^*_L(k) > h, & \quad \sum_{i=1}^{N^-} \mu^* = -N^- \cdot (\mu_0^* + K) - S^*_L(k)
\end{align*}
\]

In the developed control strategy, there are two different CUSUM charts used to monitor a mean shift and a linear drift respectively.

(i) CUSUM chart for mean shift disturbance monitoring:

In order to detect and estimate the mean shift, monitoring variable \( x_i \) is defined as \( \xi_i \) based on Eq. (3-1) as follows:
\[ \xi_t = \frac{A(q)}{D(q)} y_t - \frac{B(q)}{D(q)} u_{t-d} = \frac{C(q)}{D(q)} f_t + e_t \]  

(3-14)

Thus, \( \xi_t \) can be directly calculated for each newly observed output \( y_t \) and control input \( u_{t-d} \). The following hypothesis test is used in CUSUM chart to detect whether there exists a nonzero mean shift of \( f_t \):

\[
H_0 : \mu_{\xi_t} = \mathbb{E}(\xi_t) = 0; \quad \text{when } f_t = 0 \\
H_1 : \mu_{\xi_t} = \mathbb{E}(\xi_t) \neq 0; \quad \text{when } f_t = \mu_f
\]

(3-15)

So, taking the expectation on Eq. (3-14), it has:

\[
\mu_{\xi_t} = \mathbb{E}[\xi_t] = \frac{C(q)}{D(q)} f_t = \sum_{i=0}^{n_f} I_{f-i} f_{t-i}
\]

(3-16)

where \( n_f \) is the maximum order of the inverse function of \( I_f \) by expanding \( \frac{C(q)}{D(q)} \) as the backward operator of \( q^{-1} \). By substituting the condition of Eq. (3-15) into Eq. (3-16), it has

under \( H_0 \), \( \mu_{\xi_t} = 0 \)

under \( H_1 \), \( \mu_{\xi_t} = \mu_f \left( \frac{\min(N^*, n_f)}{\sum_{i=0}^{n_f} I_{i} f} \right) \)

(3-17)

where \( N^* \) is the steps of the mean shift occurring at \( t-N^* \). Thus, if either \( S_{H}^\xi(k) > h \) or \( S_{H}^\xi(k) > h \) is detected in the CUSUM chart, the mean shift can be estimated based on Eqs. (3-13) and (3-17) as:
(ii) CUSUM chart for monitoring a linear drift disturbance

When monitoring and estimating a linear drift disturbance, $x_t$ is defined as $\xi_t$ as

$$\xi_t = \xi_t - \xi_t-1$$  \hspace{1cm} (3-19)

From Eq. (3-14) and Eq. (3-16), it can be seen that:

$$\xi_t = I_0 f_t + I_1 f_t-1 + \cdots + I_{n_f} f_{t-n_f} + e_t$$  \hspace{1cm} (3-20)

$$\xi_{t-1} = I_0 f_{t-1} + I_1 f_{t-2} + \cdots + I_{n_f} f_{t-1-n_f} + e_{t-1}$$  \hspace{1cm} (3-21)

Subtracting Eq. (3-20) by Eq. (3-21), it has

$$\xi_t = I_0 (f_t - f_{t-1}) + I_1 (f_{t-1} - f_{t-2}) + \cdots + I_{n_f} (f_{t-n_f} - f_{t-1-n_f}) + e_t - e_{t-1}$$  \hspace{1cm} (3-22)

Thus, the following hypothesis test is used in the CUSUM chart to detect whether there exists a nonzero slope of the linear drift function $f_t$:

$$H_0: \mu^{\xi_t} = E(\xi_t) = 0; \quad \text{when } f_t - f_{t-1} = 0 \quad \forall t > 1$$

$$H_1: \mu^{\xi_t} = E(\xi_t) \neq 0; \quad \text{when } f_t - f_{t-1} = \beta$$  \hspace{1cm} (3-23)

By taking the expectation on Eq. (3-22) and substituting the condition of Eq. (3-23), we have

under $H_0$, \hspace{1cm} $\mu^{\xi_t} = 0$

under $H_1$, \hspace{1cm} $\mu^{\xi_t} = \beta \left( \sum_{i=0}^{\min(N^e, n_f)} I_i^f \right)$  \hspace{1cm} (3-24)
where \( N^\beta \) is the steps of the linear drift occurring at \( t^\beta \). Thus, if either \( S_H^c(k) > h \) or \( S_L^c(k) > h \) is detected in the CUSUM chart, the slope of linear drift can be estimated based on Eqs. (3-13) and (3-24) as:

\[
\text{if } S_H^c(k) > h, \quad \beta = \left( N^+ \cdot K + S_H^c(k) \right) \left( \sum_{j=1}^{N^+ \cdot \min(j,n_f)} I_f^j \right)^{-1}
\]

\[
\text{if } S_L^c(k) > h, \quad \beta = -\left( N^- \cdot K + S_L^c(k) \right) \left( \sum_{j=1}^{N^- \cdot \min(j,n_f)} I_f^j \right)^{-1}
\]

### 3.3.2 X-bar Chart for the Detection and Estimation of a Spike Disturbance

In a thin film deposition process, a spike signal may be observed from the thickness measurements. In practice, a spike signal is likely due to sensor errors. Thus, it is desirable to detect and remove the sensing errors from the system response, so the controller will not make a wrong feedback control. Since the process did not make any change, the spike, which is an outlier, should be moved from the data and should not be used for feedback control.

For the thin film deposition process model, the effect of a single spike error at time \( s \) on the true output \( y_t^0 \) without spike error can be modeled as:

\[
y_t^0 = y_t - \rho \delta(t-s) \quad \text{if a spike fault occurs at time } s
\]

(3-26)

where \( \delta(t-s) \) is the Kronecker delta function, \( y_t \) is the system output measurement including the spike error at time \( s \), and \( \rho \) is the magnitude of the spike. So, When there is no process disturbance \( f_t \) at \( t=s \), based on Eq. (3-14), we have
By taking expectation on Eq. (3-27), we have,

\[ E(\xi_i) = \begin{cases} 
\rho & t = s \\
0 & \text{others} 
\end{cases} \] (3-28)

From Eq. (3-28), it can be seen that an X-bar control chart is used to detect the larger spike error, which is defined as:

\[ \begin{cases} 
UCL = 3\sigma_e^2 \\
CL = 0 \\
LCL = -3\sigma_e^2 
\end{cases} \] (3-29)

where \( UCL \) (\( LCL \)) represents the upper (lower) control limits, \( CL \) is the central line of the control chart.

Since the spike usually occurs only at one sample interval, a decision rule is needed to identify spike errors, which is “A spike occurs at time \( s \) if \( \xi(s) \) is out of control, but all \( \xi(s+i) \) \( (i=-b, -b+1, ..., -1, 1, 2, ..., b) \) are in control”. Here \( b \) is a detection window determining the monitoring range, which is a integer value less than the minimal process delay time. If an out of control condition is detected, both the time and magnitude of the spike will be provided to the GPC for further compensation actions. The expectation of spike is \( E[\xi(t)] = \rho \) when \( t = s \), and \( E[\xi(t)] = 0 \) when \( t \neq s \). So, in a given detection window with length \( 2b \), the estimated magnitude of the spike \( \rho \) is estimated as:

\[ \hat{\rho} = \xi(s) - \frac{\sum_{i=-b}^{b} \xi(t)}{2b} \] (3-30)
The expectation \( E(\hat{\rho}) = E[\xi(t)] - E\left[ \sum_{i=-b}^{s} \frac{\xi(t)}{2b} \right] = \rho - 0 = \rho \).

3.3.3 Control Law Revision for Supervisory Compensation of the Detected Disturbances

(i) Control law revision under mean shift or linear drift disturbance:

A time delay always exists in detecting a disturbance using a CUSUM chart. Assume that a mean shift or drift occurs at time index \( s \), which is detected at time index \( k \) \((k > s)\). Thus, a detection delay of \( k-s \) is expected. In the next control law calculation for \( U_t, t \geq k \), there needs not only to revise the future prediction of the disturbance model \( (f_{k+1|k}, \ldots, f_{k+d+N|k}) \) based on the detected disturbance function at \( k \), but also to revise the disturbance model and residual errors during \( s < t \leq k \) for \( (f_s, f_{s+1}, \ldots, f_k) \) and \( (e_s, \ldots, e_k) \) to reflect the disturbance effect occurring at time \( s \).

When a mean shift actually occurs at time \( s \) but detected at time \( k \), the revised deterministic disturbance model used in the control law of Eq. (3-9) is:

\[
f_t = \begin{cases} 
0 & t < s \\
\mu_f & t \geq s 
\end{cases} \quad (3-31)
\]

\( \mu_f \) is calculated by Eq. (3-18). Substituting \( f_t \) of Eq. (3-31) into Eqs. (3-14) and (3-16), the revised residual is obtained as

\[
e_t = \begin{cases} 
\xi_t & t < s \\
\xi_t - \sum_{i=1}^{\min(t-s, n)} \xi_t \mu_f & s \leq t \leq k
\end{cases} \quad (3-32)
\]
By substituting those revised $e_i$ of Eq. (3-32) and $f_i$ of Eq. (3-31) into Eq. (3-9), the revised control law $U_1^{op} = [u_{t+N} \ u_{t+N-1} \ \ldots \ u_{t+i} \ \ldots \ u_t]^T$ can be obtained, in which the single $u_t$ is used to control the process at time $t$ to compensate the mean shift.

Similarly, when a linear drift actually occurs at time $s$ but detected at time $k$, the revised disturbance model used in the control law of Eq. (3-9) is:

$$f_t = \begin{cases} 0 & t < s \\ \frac{(t - s + 1)\beta}{(t - s + 1)} & t \geq s \end{cases} \tag{3-33}$$

and the revised residual is obtained by substituting $f_i$ of Eq. (3-33) into Eqs. (3-14) and (3-16) as:

$$e_t = \begin{cases} \xi_i & t \leq s \\ \min(-\xi_i, \xi_i) - \sum_{i=1}^{\min(t-s,n_i)} [I_f \cdot \beta \cdot (t - s + 1)] & s < t \leq k \end{cases} \tag{3-34}$$

By substituting those revised $e_i$ of Eq. (3-34) and $f_i$ of Eq. (3-33) into Eq. (3-9), the revised control law $U_1^{op} = [u_{t+N} \ u_{t+N-1} \ \ldots \ u_{t+i} \ \ldots \ u_t]^T$ can be obtained, in which the single $u_t$ is used to control the process at time $t$ to compensate the detected linear drift.

(ii) Control law revision under a spike disturbance:

If a spike due to sensor errors is detected, it is desirable to remove the spike data from the calculation of the feedback control. This effort can be achieved by recalculating the residual series $e_t$ used in the control law of Eq. (3-9) when a spike is detected.

Assume that a spike occurring at time $s$ has been detected at time $k=s+b$. In this case, the revised $e_t$ is obtained based on Eq. (3-27) as:
By substituting this revised $e_i$ of Eq. (3-35) into Eq. (3-9), the revised control law 

$$U_{i}^{op} = \begin{bmatrix} u_{t+N} & u_{t+N-1} & \cdots & u_{t+i} \end{bmatrix}^T$$

can be obtained, in which the single $u_t$ is used to control the process at time $t$.

3.4 CASE STUDY

3.4.1 Thin Film Deposition Process Modeling

Real production data were collected for the case study from a thin film deposition process. In the data set, the step input signals of the source temperature as shown in Figure 3.1(a) were applied during the process start up period and the process response is shown in Figure 3.1(b). 70 data points are collected for the process modeling. The sampling interval used in the data acquisition is 2 minutes.

![Figure 3.1 Initial production run data of step input and response](image)

Based on the process engineering knowledge, the thin film process under the normal operation condition is modeled by an ARX model with $C(q) = D(q) = 1$. The order of $A(q)$, $B(q)$, and delay step $d$ denoted as $[n_a, n_b, d]$ will be quantitatively
determined based on AIC (Akaike's Information theoretic Criterion) criterion (Ljung and Soderstrom, 1987; Apley and Shi, 1999), i.e., the smallest value of AIC indicates the best-fit model among all compared model candidates. In this case study, the possible model structures includes the following 14 combinations: \([1 1 d]; [2 1 d]; [3 1 d]; [4 1 d]; [2 2 d]; [3 2 d]; [4 2 d]; [3 3 d]; [4 3 d]; [4 4 d]; [5 5 d]; [6 6 d]; [7 7 d]; [8 8 d]\). The possible dead time are \(d=2, 3, 4, 5, 6\). Figure 3.2 shows all AIC values of these 14 models under each dead time. It is clear that the model of \([5 5 5]\) (model structure index=11) has the smallest AIC value, which is considered as the best model among these 14 models. After selecting this best-fit model, the statistical F-test is followed to further determine whether a lower order model \((n_a<5, n_b<5)\) can be used, which is conducted by statistical testing whether the sum of squares of modeling residual errors is significantly increased when a lower order is used (Box, et al., 1994). A F-test with the significance level 95% is used in the final order determination. Based on the data set, the final ARX model structure is \([2 2 5]\) with the parameters as \(A(q) = 1 - 0.1297 q^{-1} - 0.04919 q^{-2}\), \(B(q) = 0.007376 + 0.004798 q^{-1}\). The error term series is determined from \(e_t \sim N(0, 10^{-4})\).
In order to verify the model accuracy, Figure 3.3(a) shows the comparison between the real process response and its one step ahead prediction, and Figure 3.3(b) shows its corresponding residual errors. It is clear that the identified model has a good track performance to the real process output.

3.4.2 Predictive Control

The sliding window size used in the GPC are selected as $N=3$. It is determined based on the knowledge of the delay uncertainty in the thin film processes between the heater for the temperature adjustment and the XRF measurement of the thickness.
response. $Q = \text{diag}[1,1,1,1], \, R = r \times \text{diag}[1,1,1,1]$, and $r = 10^{-7}$. Generally, the weight coefficients ($Q$) of prediction error are set to 1 and the weight coefficients ($R$) of the control cost are adjusted based on the applications. (Astrom and Wittenmark, 1995). In order for the system to have the ability of quickly compensating the process change, $R$ is usually selected as a relative small value under the controller stability constraint. Here, we select a reasonable value of $r = 10^{-7}$ based on trial and error experiments. In Eq. (3-9), In order to calculate the optimal control input $U_i^{opt}$, it need to decide the dimension of matrixes $G_0, \, G_f$, and $G_e$, which are decided by $m_u, m_f$, and $m_e$ respectively. For the purpose of illustration, an example of how to choose $m_u$ based on the system step response was given in Figure 3.4. $m_f$ and $m_e$ are chosen using the same way. In Figure 3.4(b), the 99% percentage threshold, which means the step response stays within 1% of its final stable value, was also plotted. Based on this threshold of model accuracy, another threshold $\eta_0$ in Eq. (3-10) can be easy calculated and thus the value of $m_u$. The values of $m_f$ and $m_e$ are determined by similar way.

![Impulse and step response of the thin film deposition process model](image)

(a) Impulse response from $u$ to $y$
(b) Step response from $u$ to $y$

Figure 3.4 Impulse and step response of the thin film deposition process model
In Eq. (3-9), denote $\Phi = [G_1^T Q G_1 + R]^{-1} G_1^T Q$, which is shown as follows in the case study:

$$
\Phi = \begin{bmatrix}
-135.1 & 105.1 & -61.5 & 39.9 \\
-0.315 & -134.9 & 105.0 & -61.5 \\
0.17 & -0.414 & -134.9 & 105.1 \\
-0.073 & 0.17 & -0.315 & -135.1 
\end{bmatrix}
$$

![Figure 3.5 A simulated control performance](image)

A simulated tracking control performance with the target 7.5 is shown in Figure 3.5. From the Figure, it can be seen that there is a delay of 10 samples from the starting of the production run (at sample index 1), to the process output response. This 10-step delay comes from 5-step dead time after the first control input added at sample index 5. At $t \geq 10$, the output quickly achieves and maintains at the target value with a small variation. The noise standard deviation used in this simulation is 0.01, which is equal to the estimated $\hat{\sigma}_e$ of the modeling residual $e_t$.

3.4.3 Supervisory Strategies

3.4.3.1 Mean Shift Detection and Compensation

A CUSUM chart is designed to monitor a mean shift of the process. The process mean shift is simulated as $\Delta = -1.5 \sigma_e = -0.015$ with $\sigma_e = 0.01$, which is added to the
process at \( t \geq 30 \). The design parameters of the CUSUM chart is \( \mu_0^e = 0 \), \( K = |\Delta/2| = 0.0075 \), and \( h = 5\sigma_c = 0.05 \). Figure 3.6(a) shows the mean shift disturbance function of \( f_i \) and Figure 3.6(b) shows the CUSUM monitoring control chart of \( \zeta_i \) in the supervisory GPC. It can be seen that after the shift occurs at sample 30, the first out of control point is detected at sample \( k = 37 \) with \( S_L(37) = 0.0522 > h = 0.05 \). Based on Eq. (3-18) with \( n_f = 1 \) and \( I_o^f = 1 \) in the model, the shifted mean at time \( t = 37 \) is estimated as \( \mu^e(37) = -K - S_L(37)/N^- = -0.015 \) with \( N^- = 7 \). Figure 3.6(a) shows the comparison of the estimated means shift and true mean shift over different time, in which the dotted line of the estimated mean shift is very close to the dashed line of the true mean shift except for the detection delay period.

After detecting the mean shift, the supervisory GPC will compensate it by using the estimated mean shift. As shown in Figure 3.7(a) with the usage of the detection and compensation of the mean-shift in the controller, the controller can quickly compensate the mean shift and bring the output back to the setting target value. However, without the usage of compensating the means shift in the controller as shown in Figure 3.7(b), the mean shift disturbance would lead to a mean shift in the process response. It also notices
that due to the dead time \( (d=5) \), the control compensation is not fully applied to the system until time 42 that is 5-step dead time after the mean shift is detected at \( t=37 \). Thus, reducing the dead time in the process and have an early detection of the mean shift will be an effective way to reduce the impact of the mean shift disturbance.

![Graph](image1)

(a) response with mean shift compensation  (b) response without mean shift compensation

Figure 3.7 Comparison of control performance under mean shift disturbance

The comparison of model errors (difference between the one-step ahead prediction and the real measurement of the process) under mean shift was also plotted in Figure 3.8.

![Graph](image2)

(a) adding estimated mean shift to the model  (b) without adding mean shift to the model

Figure 3.8 Comparison of model residual errors under mean shift disturbance

3.4.3.2 Drift Disturbance Detection and Compensation

For detecting a linear drift disturbance using a CUSUM chart, the monitored random variable is defined as \( \xi_t = \xi_t - \xi_{t-1} \), which has a standard deviation of \( \sigma_\xi = \sqrt{2}\sigma_e = 0.0141 \) under the in-control condition. So, the simulated drift has a slope
of \( A = 1.5 \times \sigma_z = 0.0212 \), which is added to the process at \( t \geq 30 \). The parameters of the CUSUM chart is defined as \( \mu_0^z = 0 \), \( K = |A/2| = 0.0106 \), \( h = 4 \times \sigma_z = 0.0566 \). Figure 3.9 (a) shows the disturbance signal of \( f_t \) and Figure 3.9(b) shows the CUSUM monitoring chart. It can be seen that the first out of control point is indicated at time \( t = 33 \) with \( S_L(33) = 0.0716 > 0.0566 \) and \( N^- = 3 \). Similarly, based on Eq. (3-25) with \( n_f = 1 \) and \( I_o = 1 \) in the model, the drift slope at time \( t = 33 \) is estimated as \( \beta(33) = -0.0285 \). Figure 3.9(a) shows the comparison of the estimated slope and the true slope of the linear drift disturbance over different time, in which the dotted line of the estimated slope is very close to the dashed line of the true slope except for the detection delay period.

The comparison of the control performance between with and without compensation of the detected linear drift is shown in Figure 3.10. As it can be seen from Figure 3.10(a), after a drift disturbance was introduced at \( t \geq 30 \), the process output will experience a short period of large deviations from the setting point. However, after detecting the drift at \( t = 33 \), the supervisory GPC can quickly compensate the disturbances at \( t \geq 38 \), but the control compensation is not fully applied to the system within the dead time 5 after the drift is detected at \( t = 33 \). Thus, an effective way to reduce the impact of disturbance is to reduce the dead time and make an early detection of the drift. However, without the supervisory strategies (i.e. no detection and compensation of the drift), the controller output has a linear drift as shown in Figure 3.10(b).
The comparison of model errors (difference between the one-step ahead prediction and the real measurement of the process) under linear drift disturbance was also plotted in Figure 3.11.
3.4.3.3 Spike Disturbance Detection and Control Law Revision

It is assumed that a spike, which leads to 6% deviation of the output target value, is due to the thickness sensor error and introduced to the system at time 30. Due to the sensor error, the control performance with the supervision in GPC is shown in Figure 3.12(a). From the Figure, it can be seen that a significant deviation from the setting point occurs in the output response at data index 35, which is due to the misleading control reaction (or feedback) to the spike error added at $t=30$. This 5-step delay is due to the dead time of the process. Under the supervision, an $X$-bar control chart is used to detect the spike disturbance by using the decision rule defined in Section 3.3.2 ($h=2$ is used in the rule). Thus, the control chart detects the spike error at $k=37$, which is two-step delay after the spike is shown on the response at $t=35$. Once the spike is detected, a supervisory correction action is taken to remove the spike effects using Eq. (3-35). It can be seen that the output tracking errors are reduced significantly at $t=37$ after removing the spike in the control law calculation. However, due to this two-step delay in the spike detection, there is no compensation effort for the first two time indices ($t=35, 36$). Thus, a significant deviation from the target is observed at $t=35, 36$. Once a spike disturbance is confirmed, the supervisory GPC will remove the spike impacts from the process and make the output back to the normal condition quickly. By comparing Figure 3.12(a) and Figure 3.12(b), it can be seen that without supervision the output response at index 37 has a larger deviation than that with supervision.
It should be pointed out that the effectiveness and importance of having this spike error removal feature is dependent on the dynamic characteristic of the process model. Both the importance and effectiveness will be more significant if the impulse response has a long memory of the process dynamics, which is equivalent to have the poles of the characteristic equation close to 1. To illustrate this point, Figure 3.13 gives a comparison to show the effectiveness of the spike compensation technique based on a model obtained from another production run data. The model parameters are $A(q) = 1 - 0.39q^{-1} - 0.126q^{-2}$, $B(q^1) = 0.003645 + 0.004168q^{-1}$, $C(q) = D(q) = 1$, the dead time of $d=5$, and The error term series is determined from $e_t \sim N(0, 10^{-4})$. The poles of the characteristic equation are 0.6 and -0.21. In this case, a similar spike, which also leads to about 6% of the output target value, is added as the sensor error to demonstrate the effectiveness of the spike error compensation. It can be seen from Figure 3.13 that without the spike removal the output deviations from the target at indexes of 37-39 are larger than that with the spike removal.
3.5 CONCLUSION

A thin film deposition process control problem, which can be well modeled by a SISO ARMAX linear dynamic model, was investigated in this chapter. Due to the inherent characteristics of the thin film deposition process with a large and varying dead time, a GPC was adopted to the process controller design. A set of supervisory strategies is used to obtain a satisfied control performance by compensating the different types of disturbances. One significance of the chapter has been emphasizing the importance of developing supervisory strategies through monitoring the process changes using SPC techniques, and then revising the controller parameters accordingly to achieve a superior control performance. This integration of SPC with automatic process control (APC) provides great potential for the development of effective controllers in complex manufacturing processes. The case studies validate the effectiveness of the developed supervisory GPC.

There are several open issues to be further investigated for the techniques presented in this chapter. One of them is how to determine the thresholds for the SPC control charts, because the SPC is used to monitor a GPC controlled process. Thus, Type I and Type II errors discussed in the conventional SPC literatures cannot be used directly.
Although some research have been done to investigate the SPC monitoring for the PID controlled process (Tsung, et al, 1998), the SPC monitoring for the supervisory GPC control, which is more complex than the PID control strategy, deserves some further attention. Also, a cautious control strategy could be integrated to consider the estimation uncertainty of the process model and disturbance (Shi and Apley, 1998). Another open issue worth to investigate is how to systematically and simultaneously select (optimal) GPC parameters (e.g. $N$, $R$, etc.), SPC thresholds, and the alternative supervisory strategies. Currently, some trial-and-error efforts are required to select those parameters (Astrom and Wittenmark, 1995). The other topic related to the thin film thickness control is to expand the research presented in this chapter from single-input-single-output (SISO) to multiple-inputs-multiple-outputs (MIMO) cases because there are multiple material sources, as well as multiple quality indices, for a thin film deposition process. Some early works in predictive control based on the state space model can be considered as a way for the further extension.
CHAPTER 4
ANOVA METHOD FOR VARIANCE COMPONENT DECOMPOSITION AND DIAGNOSIS IN BATCH MANUFACTURING PROCESSES

In Chapter 4, ANOVA method will be used on production data to do inherent variance diagnosis for the Silver Printing Process. The research in this chapter provides a generic framework for decomposition of three typical variation components in batch manufacturing processes. For the purpose of variation root causes diagnosis, corresponding linear contrasts are defined to represent the possible site variation patterns and the statistical nested effect models are developed accordingly. It shows that the use of a full factor decomposition model can expedite the determination of the number of nested effect models and the model structure. Finally, an example is given for the inherent variation reduction in the screening conductive gridline printing process for solar battery fabrication.

This chapter is organized as follows: In Section 4.1, a general variation decomposition framework of total process variations is provided for all possible nested relationship of variance components of interest. Section 4.2 discusses the general modeling and analysis procedures using the ANOVA method. The implementation and the effectiveness of the developed methodology are illustrated in Section 4.3 with a real case study in the gridline printing process. Finally, the chapter is concluded in Section 4.4.
4.1 GENERAL DECOMPOSITION FRAMEWORK OF TOTAL PROCESS VARIATIONS

As it is shown in the literature (Woodall and Thomas, 1995; Yashchin, 1994; Roes and Does, 1995; Runger and Fowler, 1998), process variations in batch manufacturing processes can be generally classified into three types due to the change of three factors or variables, that is, (a) Factor batch leading to batch-by-batch variation due to the difference of batches; (b) Factor sample inducing sample-by-sample variation representing the difference among samples; and (c) Factor site representing the nonuniformity of site-by-site variations. In practice, depending on the potential root causes of each variation component in a particular application, the change of one factor (such as Factor A) usually occurs within another factor (such as Factor B). In the analysis of variance (ANOVA), it is called Factor A is nested by Factor B. For example, in a wafer fabrication process, the wafer-by-wafer (i.e. sample-by-sample) variation is usually analyzed within a given batch. Thus, Factor sample is nested by Factor batch. Similarly, in a screening conductive gridline printing process discussed in Section 4.3, the variance component of Factor site, which is a fixed factor, is studied within Factor run. In this case, Factor site is nested by Factor batch. Therefore, different nested models should be carefully constructed to appropriately represent those corresponding variance components.

In this Chapter, a general decomposition framework of total process variations is provided in Figure 4.1, where Layer 1 shows the process variation decomposed by each of the three factors, and Layer 2 provide a general list of how each factor's variation is
nested by other two factors. Furthermore, in order to expedite the root cause diagnosis of inherent site variation, the site-by-site variation can be further decomposed into different contrasts defined by linear combinations of site measurements, which is shown in the bottom of Layer 1 in Figure 4.1. The linear relationships among sites in a contrast should be well defined based on the particular process knowledge such that each contrast corresponds to some meaningful variation patterns of known root causes. When this contrast decomposition of the total site-by-site variation is performed, the variance components as defined in Layer 2 should be further decomposed under each site contrast rather than the total site variation as shown in layer 3 of Figure 4.1. The detailed discussion of this decomposition will be discussed in Section 4.2 in the silver gridlines printing process.

Figure 4.1 General decomposition of variance components
It should be noted that for a particular application, not all listed nest factors necessarily have a clear physical interpretation. The determination of the nest components from this general list is essentially needed for a given application, which usually relies on the process engineering knowledge of the interested variance components and the existence of the possible root causes.

In general, only one statistical model may not be sufficient to describe all interested variation components in Figure 4.1. So, there is a need to justify how many models are required and what statistical models are adequate for modeling of the selected subset of variance components. In the previous research literatures (Woodall and Thomas, 1995; Yashchin, 1994; Roes and Does, 1995; Runger and Fowler, 1998), there is no discussion on how to use a set of statistical models to fit various decompositions of variance component, because a single statistical model was sufficient to describe the selected monitoring variance components in their applications. In general, the selection of the interested factors (or nested factors) and the determination of the needed statistical models are related to the variation characteristics of a particular application process. As an example, a screening conductive gridline printing process will be used in this chapter to illustrate the details of the methodology development and implementation procedures.

4.2 VARIANCE COMPONENT ANALYSIS USING ANOVA

In the following sections, \( y_{ijk} \) is used to denote a measurement taken at site \( j \) \((j=1,...,b)\) of panel sample \( k \) \((k=1,...,n)\) within batch \( i \) \((i=1,...,a)\). In the example of the screening conductive gridline printing process, four measurement sites are fixed for each panel sample, i.e., \( b=4 \); three panel samples are taken from each batch at
each site, i.e., \( n = 3 \); and six batches of production data (i.e. \( a = 6 \)) are selected for initial evaluation of the process variation. The corresponding sum of squares of these three variations \((Q_1, Q_2, \text{and } Q_3)\) are represented by:

\[
SS(Q_1) = SS_A = bn \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_{ij} - \bar{y}_i)^2
\]

\[
SS(Q_2) = SS_{A(B)} = n \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_{ij} - \bar{y}_{.j})^2
\]

\[
SS(Q_3) = SS_{B(A)} = n \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_{ij} - \bar{y}_{i.})^2
\]

\[
\bar{y}_{ij} = \frac{\sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}}{abn}, \quad \bar{y}_{.j} = \frac{\sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk}}{nb}, \quad \bar{y}_{i.} = \frac{\sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}}{an}, \quad \bar{y}_{...} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}}{nab}
\]

Factor A is a random factor with \( a \) levels and Factor B is a fixed factor with \( b \) levels.

The critical issue of using the ANOVA method is to develop appropriate linear statistical models for the decomposed variance components of interest. For analyzing the interested sum of squares of \( SS_{A(B)} \) and \( SS_{B(A)} \) induced by those two nested factors \( A(B) \) and \( B(A) \), two different nested effect decomposition models are needed and discussed in the following subsections. Subsection 4.2.1 will present those two statistical nested models. Section 4.2.2 gives the statistical testing and estimation of each variance components. In order to perform root cause determination of the site variation, a further decomposition of the site variation by three linear contrasts is presented in details in Subsection 4.2.3.
4.2.1 Statistical Modeling of Interested Variation Components

A linear two-stage nested design model can be used to represent the interested variance components $Q_1$ and $Q_3$ induced by Factor batch (A) and the nested Factor site (B) within the batch as follows:

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{(i)jk}$$

where $\mu$ is the overall mean of all factors at all levels, $\tau_i$ is the $i$th level effect of Factor A, $\beta_{j(i)}$ is the $j$th level effect of Factor B nested under the $i$th level of Factor A, and $\epsilon_{(i)jk}$ is a random model error with $\epsilon_{(i)jk} \sim NID(0, \sigma^2)$. The variance $\sigma^2$ is assumed to be constant and independent of all levels of other factors. Since Factor A is a random factor, it is assumed $\tau_i \sim NID(0, \sigma^2)$. Although Factor B is a fixed factor, the nested factor of B-within-A is a random factor. So, it is assumed that $\beta_{j(i)} \sim NID(0, \sigma^2_{\beta(i)})$. From model (4-4), the total sum of squares can be decomposed as follows called Decomposition I here:

$$SS_T = SS(Q_1) + SS(Q_3) + SS_E = SS_A + SS_{B(A)} + SS_E$$

Similarly, a different linear two-stage nested design model is needed to represent the interested sum of squares of $Q_2$ induced by Factor batch (A) nested by Factor site (B) as follows:

$$y_{ijk} = \mu + \tau_{i(j)} + \beta_{j} + \epsilon_{(j)ik}$$

where $\tau_{i(j)}$ is the $i$th level effect of Factor A nested by Factor site (B), $\beta_{j}$ is the $j$th level effect of Factor B, and $\epsilon_{(j)ik}$ is a random model error with $\epsilon_{(j)ik} \sim NID(0, \sigma^2)$. The variance $\sigma^2$ is assumed to be constant and independent of all levels of other factors. Since Factor A is a random factor, it is assumed $\tau_{i(j)} \sim NID(0, \sigma^2_{\tau(i)})$. Although Factor B is a fixed factor, the nested factor of B-within-A is a random factor. So, it is assumed that $\beta_{j} \sim NID(0, \sigma^2_{\beta})$. From model (4-6), the total sum of squares can be decomposed as follows called Decomposition II here:

$$SS_T = SS(Q_2) + SS_E = SS_B + SS_{B(A)} + SS_E$$
Now, Factor A is random, the nested factor of A-within-B is still random, thus, 
\( \tau_{a(j)} \sim NID(0, \sigma^2) \). However, Factor B is a fixed factor satisfying the condition of 
\[ \sum_{j=1}^{b} \beta_j = 0. \] In this case, the total sum of squares is decomposed as follows called

**Decomposition II.**

\[
SS_T = SS_B + SS(Q_2) + SS_E = SS_B + SS_{A(B)} + SS_E
\]  
(4-7)

### 4.2.2 Statistical Testing and Estimation of Variance Components

In ANOVA analysis, it is necessary to construct appropriate test statistics to check the significant level of each factor. The F-test statistics are obtained as Eqs. (4-8)-(4-10) based on the relationship of the expected mean squares of the variance components (Montgomery, 1997),

\[
\frac{MS_A}{MS_E} \sim F(df_A, df_E)
\]  
(4-8)

\[
\frac{MS_{A(B)}}{MS_E} \sim F(df_{A(B)}, df_E)
\]  
(4-9)

\[
\frac{MS_{B(A)}}{MS_E} \sim F(df_{B(A)}, df_E)
\]  
(4-10)

where \( MS_A, MS_{A(B)} \) and \( MS_{B(A)} \) are the mean squares of the corresponding factors described by each subscript. \( df_A, df_{A(B)}, \) and \( df_{B(A)} \) are the corresponding degrees of the freedom. The detailed analysis of each item of **Decomposition I** is shown in the expected mean squares table of Table 4.1 and the analysis of variance table of Table 4.2.
Table 4.1 Expected Mean Squares for Decomposition I of Eq. (4-5)

<table>
<thead>
<tr>
<th>Factor</th>
<th>R</th>
<th>F</th>
<th>R</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>τ_i</td>
<td>1</td>
<td>b</td>
<td>N</td>
<td>( E(\text{MS}<em>A) = \sigma^2 + bn\sigma^2</em>{(r)} )</td>
</tr>
<tr>
<td>( \beta_{(i)} )</td>
<td>1</td>
<td>0</td>
<td>N</td>
<td>( E(\text{MS}<em>{B(A)}) = \sigma^2 + n\sigma^2</em>{(r)} )</td>
</tr>
<tr>
<td>( \epsilon_{(i)} )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( E(\text{MS}_E) = \sigma^2 )</td>
</tr>
</tbody>
</table>

Table 4.2 Analysis of Variance Table for Decomposition I of Eq. (4-5)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares (SS)</th>
<th>Degree of Freedom (DF)</th>
<th>Mean Square (MS)</th>
<th>F-test (F_0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Factor A ( A \Rightarrow Q_1 )</td>
<td>( SS_A = bn \sum_{i=1}^a (y_{i.} - \bar{y})^2 )</td>
<td>( df_A = a-1 )</td>
<td>( MS_A = \frac{SS_A}{df_A} )</td>
<td>( F_0 = \frac{MS_A}{MS_E} )</td>
</tr>
<tr>
<td>Random Factor B (A) ( B(A) \Rightarrow Q_1 )</td>
<td>( SS_{B(A)} = n \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{.j})^2 )</td>
<td>( df_{B(A)} = a(b-1) )</td>
<td>( MS_{B(A)} = \frac{SS_{B(A)}}{df_{B(A)}} )</td>
<td>( F_0 = \frac{MS_{B(A)}}{MS_E} )</td>
</tr>
<tr>
<td>Error E</td>
<td>( SS_E = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{.ij})^2 )</td>
<td>( df_E = ab(n-1) )</td>
<td>( MS_E = \frac{SS_E}{df_E} )</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y})^2 )</td>
<td>( df_T = abn-1 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Similar analysis is performed for Decomposition II of Eq. (4-7), where Factor A-within-B is a random factor but Factor B is a fixed factor. The determination of the expected mean squares table is shown in Table 4.3. A corresponding summary of the analysis of variances is given in Table 4.4.

Table 4.3 Expected Mean Squares for Decomposition II of Eq. (4-7)

<table>
<thead>
<tr>
<th>Factor</th>
<th>R</th>
<th>F</th>
<th>R</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>τ_{(i)}</td>
<td>1</td>
<td>1</td>
<td>n</td>
<td>( E(\text{MS}<em>{A(B)}) = \sigma^2 + n\sigma^2</em>{(r)} )</td>
</tr>
<tr>
<td>( \beta_j )</td>
<td>a</td>
<td>0</td>
<td>n</td>
<td>( E(\text{MS}<em>B) = \sigma^2 + n\sigma^2</em>{(r)} + (an\sum_{j=1}^b \beta_j^2)/(b-1) )</td>
</tr>
<tr>
<td>( \epsilon_{(i)} )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( E(\text{MS}_E) = \sigma^2 )</td>
</tr>
</tbody>
</table>
Table 4.4 Analysis of Variance Table for Decomposition II of Eq. (4-7)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares (SS)</th>
<th>Degree of Freedom (DF)</th>
<th>Mean Square (MS)</th>
<th>F-test ($F_0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Factor</td>
<td>$SS_{A(B)} = n \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_{ij} - \bar{\bar{y}})^2$</td>
<td>$df_{A(B)} = b(a-1)$</td>
<td>$MS_{A(B)} = \frac{SS_{A(B)}}{df_{A(B)}}$</td>
<td>$F_{A(B)} = \frac{MS_{A(B)}}{MS_{E}}$</td>
</tr>
<tr>
<td>Fixed Factor B</td>
<td>$SS_{B} = an \sum_{j=1}^{b} (\bar{y}_{j} - \bar{\bar{y}})^2$</td>
<td>$df_{B} = b-1$</td>
<td>$MS_{B} = \frac{SS_{B}}{df_{B}}$</td>
<td>$F_{B} = \frac{MS_{B}}{MS_{A(B)}}$</td>
</tr>
<tr>
<td>Error E</td>
<td>$SS_{E} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ik})^2$</td>
<td>$df_{E} = ab(n-1)$</td>
<td>$MS_{E} = \frac{SS_{E}}{df_{E}}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_{T} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}_{ik})^2$</td>
<td>$df_{T} = abn-1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The estimates of these variance components are also obtained as:

$$\hat{\sigma}_{Q_2}^2 = \hat{\sigma}_{i}^2 = \frac{(MS_{A} - MS_{E})}{bn} \quad (4-11)$$

$$\hat{\sigma}_{Q_2}^2 = \hat{\sigma}_{r(B)}^2 = \frac{(MS_{A(B)} - MS_{E})}{n} \quad (4-12)$$

$$\hat{\sigma}_{Q_2}^2 = \hat{\sigma}_{t(r)}^2 = \frac{(MS_{B(A)} - MS_{E})}{n} \quad (4-13)$$

4.2.3 Site Variability Decomposition and Diagnosis Based on Linear Contrasts

Both variation components of $Q_2$ and $Q_3$ are related to the site variability. If such variation components are significant, removal or reduction of the associated root causes are generally desired. From Table 2.1, it can be seen that the site variability is mainly related to the tool conditions and tool setups. Thus, identification of site variation patterns can expedite the diagnosis of variation root causes for process improvement.

Based on the engineering knowledge of potential fault patterns in the screening conductive gridline printing process, the total spatial variation among four sites can be further decomposed into three typical variation patterns, which are reflected by the differences of the front and back site (between M1&M4 and M2&M3), the left and right site (between M1&M2 and M3&M4), and the diagonals (between M1&M3 and M2&M4)
as shown in Figure 2.5. The statistical test and the estimates of these three variation patterns can provide a useful guideline on how to adjust the tool position to reduce tool setup errors.

For the purpose of analyzing the contribution of each variation pattern, three linear contrasts of $B^{FB}$, $B^{LR}$, and $B^{D}$ are defined to represent each contrast effect of front-back sites, left-right sites, and the diagonal respectively, that is,

$$\beta_{ik}^{FB} = \frac{(y_{i1k} + y_{i4k} - y_{i2k} - y_{i3k})}{2}$$

$$\beta_{ik}^{LR} = \frac{(y_{i1k} + y_{i2k} - y_{i3k} - y_{i4k})}{2}$$

$$\beta_{ik}^{D} = \frac{(y_{i1k} + y_{i3k} - y_{i2k} - y_{i4k})}{2}$$

The constant 2 in the denominator is used to normalize the effect of each contrast.

It is known that two contrasts with linear coefficients $\{c_i\}$ and $\{d_i\}$ are called orthogonal contrasts if the condition of $\sum_{i=1}^{b} c_i d_j = 0$ is satisfied. From Eqs. (4-14)-(4-16), it can be seen that the linear coefficients of those three contrasts $\beta_{ik}^{FB}$, $\beta_{ik}^{LR}$, and $\beta_{ik}^{D}$ are [0.5 -0.5 -0.5 0.5], [0.5 0.5 -0.5 -0.5], and [0.5 -0.5 0.5 -0.5] respectively. So, it can be proved that the defined three contrasts are orthogonal to each other. The following analysis will show how to develop appropriate statistical models to meet the need of analyzing these contrast variations for Decomposition I of Eq.(4-5) and Decomposition II of Eq. (4-7) respectively.
4.2.3.1 Analysis of Contrast Variance Components for Decomposition I

(1) Variation Decomposition I Represented by Linear Contrasts

It is known that the total sum of squares $SS_T$ can be represented by:

$$SS_T = \sum_{i=1}^{a} \sum_{k=1}^{n} [(y_{ik} - \bar{y}_{..})^2 + (y_{i2k} - \bar{y}_{..})^2 + (y_{i3k} - \bar{y}_{..})^2 + (y_{i4k} - \bar{y}_{..})^2] \quad (4-17)$$

It can be shown that:

$$y_{ik} - \bar{y}_{..} = (\bar{y}_{i..} - \bar{y}_{..}) + \frac{\bar{\beta}^{FB}_i + \bar{\beta}^{LR}_i + \bar{\beta}^{D}_i}{2} + (\bar{y}_{i..} - \bar{y}_{i.}) \quad (4-18)$$

So, the sum of squares of Eq. (4-18) is:

$$\sum_{i=1}^{a} \sum_{k=1}^{n} (y_{ik} - \bar{y}_{..})^2 = n \sum_{i=1}^{a} [(\bar{\beta}^{FB}_i)^2 + (\bar{\beta}^{LR}_i)^2 + (\bar{\beta}^{D}_i)^2] + \sum_{i=1}^{a} \sum_{k=1}^{n} (y_{i..} - \bar{y}_{i.})^2 \quad (4-19)$$

The derivation from Eq. (4-18) to Eq. (4-19) utilizes the fact that all four items of $\bar{y}_{i..}$, $\bar{\beta}^{FB}_i$, $\bar{\beta}^{LR}_i$, and $\bar{\beta}^{D}_i$ are orthogonal to each other.

Similarly, for other three variation components of Eq. (4-17), it can be obtained that:

$$y_{i2k} - \bar{y}_{..} = (\bar{y}_{i..} - \bar{y}_{..}) + \frac{-\bar{\beta}^{FB}_i + \bar{\beta}^{LR}_i - \bar{\beta}^{D}_i}{2} + y_{i2k} - \bar{y}_{i2} \quad (4-20)$$

$$y_{i3k} - \bar{y}_{..} = (\bar{y}_{i..} - \bar{y}_{..}) + \frac{-\bar{\beta}^{FB}_i + \bar{\beta}^{LR}_i + \bar{\beta}^{D}_i}{2} + y_{i3k} - \bar{y}_{i3} \quad (4-21)$$

$$y_{i4k} - \bar{y}_{..} = (\bar{y}_{i..} - \bar{y}_{..}) + \frac{\bar{\beta}^{FB}_i - \bar{\beta}^{LR}_i - \bar{\beta}^{D}_i}{2} + y_{i4k} - \bar{y}_{i4} \quad (4-22)$$

Thus, the total sum of squares of Eq. (4-17) can be represented in terms of the nested contrasts as:
88

\[ SS_T = 4n \sum_{i=1}^{d} (\bar{y}_{i.} - \bar{y}_{..})^2 + n \sum_{i=1}^{d} [(\bar{\beta}_{iF})^2 + (\bar{\beta}_{iL})^2 + (\bar{\beta}_{iN})^2] + \sum_{i=1}^{d} \sum_{j=1}^{k} (y_{ijk} - \bar{y}_{i.})^2 \]

\[ = SS_A + SS_{B^n(A)} + SS_{B^L(A)} + SS_{B^N(A)} + SS_E \] (4-23)

It concludes that the original total site variation nested by Factor run are decomposed into three orthogonal contrast variations nested by Factor run, that is,

\[ SS_{B(A)} = SS_{B^n(A)} + SS_{B^L(A)} + SS_{B^N(A)} \]. This result shows that for a factor with level \( b \), its total variation nested by another factor can still be decomposed into the variations of \( b-1 \) orthogonal contrasts under the same nest factor.

(2) Representation of Contrast Effects by Factor Effects

In fact, three contrasts defined above can also be further represented by two factors \( V \) and \( W \) with two fixed levels, which are defined as \( V=1 \) representing the front site, \( V=-1 \) representing the back site; \( W=1 \) representing the left site, and \( W=-1 \) representing the right site, that is,

\[ v_{ik} = (y_{ik} + y_{ik})/\sqrt{2} \], \quad \omega_{ik} = (y_{ik} + y_{ik})/\sqrt{2} \] (4-24)

\[ \omega_{ik} = (y_{ik} + y_{ik})/\sqrt{2} \], \quad \omega_{ik} = (y_{ik} + y_{ik})/\sqrt{2} \] (4-25)

\( \sqrt{2} \) is used to normalize each factor effect. The equivalence of the defined new factors to the contrasts can be seen by the following proof through comparing the results in Eq. (4-29) with that in Eq. (4-23).
Similarly, the sum of squares of the nested factor $W$ as well as the nested interaction of factor $V$ and factor $W$ are:

$$SS_{V(W)} = n \sum_{i=1}^{6} \left[ \frac{(\bar{y}_{A_i} + \bar{y}_{W_i}) - \bar{y}_{A_i} + \bar{y}_{W_i} + \bar{y}_{i}}{\sqrt{2}} \right]^2 + \left[ \frac{(\bar{y}_{A_i} + \bar{y}_{W_i}) - \bar{y}_{A_i} + \bar{y}_{W_i} + \bar{y}_{i}}{\sqrt{2}} \right]^2$$

$$SS_{V(W)} = n \sum_{i=1}^{6} \left[ \frac{(\bar{y}_{A_i} + \bar{y}_{W_i}) - \bar{y}_{A_i} + \bar{y}_{W_i} + \bar{y}_{i}}{\sqrt{2}} \right]^2 + \left[ \frac{(\bar{y}_{A_i} + \bar{y}_{W_i}) - \bar{y}_{A_i} + \bar{y}_{W_i} + \bar{y}_{i}}{\sqrt{2}} \right]^2$$

(4-27)

It shows that $SS_{W(A)} = SS_{LR(A)}$ and $SS_{[V,W](A)} = SS_{D(A)}$. Thus, Eq. (4-23) can be represented as:

$$SS_T = SS_A + SS_{V(A)} + SS_{W(A)} + SS_{[V,W](A)} + SS_E$$

(4-29)

Remark: An important conclusion is that the total process variation is contributed by three factors as random Factor $A$ with the level equal to the number of runs and two fixed factors of $V$ and $W$ with the level equal to two. The main effects of Factors $V$ and $W$ as well as the effect of their interaction $V \times W$ are equivalent to the contrast effects of $B^{FB}$, $B^{LR}$, and $B^{D}$ respectively. This conclusion can be directly used to expedite the
development of a statistical model for the contrast variations under *Decomposition II*,
which will be discussed in subsection 4.2.3.2.

(3) Statistical Testing and Estimation of the Decomposed Site Variance Components

Statistical modeling of the decomposition of Eq. (4-29) can be represented by an equivalent model of the two-stage nested design of three factors as:

\[
y_{pqk} = \mu + \tau_i + \nu_{p(i)} + \omega_{q(i)} + (\nu\omega)_{pq(i)} + \epsilon_{(ipq)k}\]

\[
\begin{align*}
\tau_i &\quad i = 1, \ldots, a \\
\nu_{p(i)} &\quad p, q = 1, 2 \\
\omega_{q(i)} &\quad k = 1, \ldots, n
\end{align*}
\]  

(4-30)

In the model, \(\tau_i\) is the effect of run \(i\). \(\nu_{p(i)}\) is the effect of Factor \(V\) at level \(p\) nested by Factor \(A\) at level \(i\), that is, the effect of the front-back contrast within run \(i\). Similarly, \(\omega_{q(i)}\) is the effect of the left-right contrast within run \(i\). \((\nu\omega)_{pq(i)}\) is the effect of the diagonal contrast within run \(i\). So, the main effect of \(\tau_i\) is the same as the original model of Eq. (4-4), the nested effects of \(\nu_{p(i)}\), \(\omega_{q(i)}\), and \((\nu\omega)_{pq(i)}\) represent the within-run variations induced by different site variation patterns. Now, using the decomposition model of Eq. (4-30), the \(F\)-tests of the nested contrast factors are developed in Eqs. (4-31)-(4-33) based on the expected mean squares table of Table 4.5 and the analysis of variance table of Table 4.6.

\[
\frac{MS_{V(A)}}{MS_E} \sim F(a, df_E) \quad (4-31)
\]

\[
\frac{MS_{W(A)}}{MS_E} \sim F(a, df_E) \quad (4-32)
\]

\[
\frac{MS_{V\times W}(A)}{MS_E} \sim F(a, df_E) \quad (4-33)
\]
For analysis of the nested contrast variance components of Decomposition I in Eq. (4-29), the expected mean squares table is shown in Table 4.5, where the effect of each contrast within Factor run is random. A summary of the analysis of variances is given in Table 4.6.

Table 4.5 Expected Mean Squares for Contrast Decomposition I of Eq. (4-29)

<table>
<thead>
<tr>
<th>Factor</th>
<th>R^2</th>
<th>F</th>
<th>F</th>
<th>R^2</th>
<th>n</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>R^2_i</td>
<td>1 2 2 n</td>
<td></td>
<td></td>
<td>1 2 n</td>
<td></td>
<td>$E(MS_A) = \sigma^2 + 4n\sigma^2_r$</td>
</tr>
<tr>
<td>$\beta^P_{(i)} \Rightarrow \nu_{p(i)}$</td>
<td>1 0 2 n</td>
<td></td>
<td></td>
<td>1 0 n</td>
<td></td>
<td>$E(MS_{V(A)}) = \sigma^2 + 2n\sigma^2_{\nu(\tau)}$</td>
</tr>
<tr>
<td>$\beta^R_{(i)} \Rightarrow \omega_{q(i)}$</td>
<td>1 2 0 n</td>
<td></td>
<td></td>
<td>1 2 n</td>
<td></td>
<td>$E(MS_{W(A)}) = \sigma^2 + 2n\sigma^2_{\omega(\tau)}$</td>
</tr>
<tr>
<td>$\beta^P_{(i)} \Rightarrow (\nu\omega)_{pq(i)}$</td>
<td>1 0 0 n</td>
<td></td>
<td></td>
<td>1 0 n</td>
<td></td>
<td>$E(MS_{V\times W(A)}) = \sigma^2 + n\sigma^2_{\nu\times\omega(\tau)}$</td>
</tr>
<tr>
<td>$e_{(pq)k}$</td>
<td>1 1 1 1</td>
<td></td>
<td></td>
<td>1 1 1</td>
<td></td>
<td>$E(MS_E) = \sigma^2$</td>
</tr>
</tbody>
</table>

Table 4.6 Analysis of Variance Table for Contrast Decomposition I of Eq. (4-29)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Factor A ⇒ Q1</td>
<td>$SS_A = bn \sum_{i=1}^{a} (\bar{y}<em>{i.} - \bar{y}</em>{.})^2$</td>
<td>$MS_A = \frac{SS_A}{a-1}$</td>
<td>$F_A = \frac{MS_A}{MS_E}$</td>
</tr>
<tr>
<td>Random Factor B^R(A) ⇒ V(A)</td>
<td>$SS_{V(A)} = n \sum_{i=1}^{a} (\bar{y}<em>{i1} + \bar{y}</em>{i4} - \bar{y}<em>{i2} - \bar{y}</em>{i3})^2$</td>
<td>$MS_{V(A)} = \frac{SS_{V(A)}}{a}$</td>
<td>$F_{V(A)} = \frac{MS_{V(A)}}{MS_E}$</td>
</tr>
<tr>
<td>Random Factor B^R(A) ⇒ W(A)</td>
<td>$SS_{W(A)} = n \sum_{i=1}^{a} (\bar{y}<em>{i1} + \bar{y}</em>{i2} - \bar{y}<em>{i3} - \bar{y}</em>{i4})^2$</td>
<td>$MS_{W(A)} = \frac{SS_{W(A)}}{a}$</td>
<td>$F_{W(A)} = \frac{MS_{W(A)}}{MS_E}$</td>
</tr>
<tr>
<td>Random Factor B^R(A) ⇒ V×W(A)</td>
<td>$SS_{V\times W(A)} = n \sum_{i=1}^{a} (\bar{y}<em>{i1} + \bar{y}</em>{i4} - \bar{y}<em>{i2} - \bar{y}</em>{i3})^2$</td>
<td>$MS_{V\times W(A)} = \frac{SS_{V\times W(A)}}{a}$</td>
<td>$F_{V\times W(A)} = \frac{MS_{V\times W(A)}}{MS_E}$</td>
</tr>
<tr>
<td>Error E</td>
<td>$SS_E = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (y_{ijk} - \bar{y}_{.})^2$</td>
<td>$MS_E = \frac{SS_E}{4a(n-1)}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (y_{ijk} - \bar{y}_{.})^2$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The corresponding nested variance components are estimated by.

\[
\hat{\sigma}_Q^2 = \hat{\sigma}_{v(\tau)}^2 = \frac{(MS_{v(A)} - MS_E)}{2n};
\]

\[
\hat{\sigma}_Q^2 = \hat{\sigma}_{w(\tau)}^2 = \frac{(MS_{w(A)} - MS_E)}{2n};
\]

\[
\hat{\sigma}_Q^2 = \hat{\sigma}_{v\times w(\tau)}^2 = \frac{(MS_{[v\times w]}(A) - MS_E)}{n};
\]

(4-34) \hspace{1cm} (4-35) \hspace{1cm} (4-36)

4.2.3.2 Analysis of Contrast Variance Components for Decomposition II

For Decomposition II, the original nested Factor A(B) is further decomposed into three items, that is, Factor A nested by three contrasts can be equivalent to Factor A nested by Factor V, Factor W, and their interaction V×W as A(V), A(W), and A(V×W).

From the conclusion in the remark of Section 4.2.3.1, it is known that the total process variation can be generally considered as the effect of these three factors (A, V, and W) and their interactions. Thus, a full factor decomposition of the total sum of squares of three factors can be generally expressed as:

\[
SS_T = SS_A + SS_V + SS_W + SS_{A\times V} + SS_{A\times W} + SS_{V\times W} + SS_{A\times V\times W} + SS_E
\]

(4-37)

From the nested design, it is known that \(SS_{A(V)} = SS_A + SS_{A\times V}\), \(SS_{A(W)} = SS_A + SS_{A\times W}\), and \(SS_{A(V\times W)} = SS_A + SS_{A\times V} + SS_{A\times W} + SS_{A\times V\times W}\). Thus, there need three decomposition models to represent all these nested factors of A(V), A(W), and A(V×W), that is,

\[
SS_T = SS_{A(V)} + SS_V + SS_W + SS_{V\times W} + SS_{A\times W(V)} + SS_E
\]

(4-38)

\[
SS_T = SS_{A(W)} + SS_V + SS_W + SS_{V\times W} + SS_{A\times V(W)} + SS_E
\]

(4-39)

\[
SS_T = SS_{A(V\times W)} + SS_V + SS_W + SS_{V\times W} + SS_E
\]

(4-40)
Here, both Eqs. (4-38) and (4-39) are corresponding to the standard two stage nested design models. In fact, it is known that \( SS_b = SS_p + SS_{\varphi} + SS_{\varphi(w)} \). Thus, by comparing Eq. (4-40) with Eq. (4-7), it can be obtained that \( SS_{A(V \times W)} = SS_{A(B)} \). So, there is no need to analyze Eq. (4-40) anymore. The statistical models corresponding to Eqs. (4-38) and (4-39) are:

\[
y_{ijpq} = \mu + \tau_{(p)} + \nu_{(i)} + \omega_{(q)} + (\tau \omega)_{ip(q)} + \varepsilon_{(ipq)}
\]

\[
y_{ijpq} = \mu + \tau_{(p)} + \nu_{(i)} + \omega_{(q)} + (\tau \omega)_{ip(q)} + \varepsilon_{(ipq)}
\]

From Eq. (4-41), the F-test for the nested Factor \( A(V) \) is obtained based on the expected mean squares table and the analysis of variance table shown in the Tables 4.7 and 4.8.

\[
\frac{MS_{A(V)}}{MS_E} \sim F(2(a-1), df_E)
\]

<table>
<thead>
<tr>
<th>Factor</th>
<th>R</th>
<th>F</th>
<th>F</th>
<th>R</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau(\beta^{FB}) \Rightarrow \tau_{(p)} )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>n</td>
<td>( E(MS_{A(V)}) = \sigma^2 + 2n \sigma_{\tau(v)}^2 )</td>
</tr>
<tr>
<td>( \beta^{FB} \Rightarrow \nu_{(i)} )</td>
<td>a</td>
<td>0</td>
<td>2</td>
<td>n</td>
<td>( E(MS_\nu) = \sigma^2 + 2na \sum_{j=1}^{2} \nu_j^2 + 2n \sigma_{\tau(v)}^2 )</td>
</tr>
<tr>
<td>( \beta^{LR} \Rightarrow \omega_{(q)} )</td>
<td>a</td>
<td>2</td>
<td>0</td>
<td>n</td>
<td>( E(MS_\varphi) = \sigma^2 + 2na \sum_{j=1}^{2} \omega_j^2 + n \sigma_{\tau(v)}^2 )</td>
</tr>
<tr>
<td>( \beta^{D} \Rightarrow (\tau \omega)_{ip} )</td>
<td>a</td>
<td>0</td>
<td>0</td>
<td>n</td>
<td>( E(MS_{\varphi(w)}) = \sigma^2 + 2na \sum_{j=1}^{2} \nu_j^2 + n \sigma_{\tau(v)}^2 )</td>
</tr>
<tr>
<td>( <a href="%5Cbeta%5E%7BFB%7D">\tau \times \beta^{LR}</a> \Rightarrow (\tau \omega)_{ipq} )</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>n</td>
<td>( E(MS_{A(V \times W)}) = \sigma^2 + n \sigma_{\tau(v)}^2 )</td>
</tr>
<tr>
<td>( \varepsilon_{(ipq)} )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( E(MS_E) = \sigma^2 )</td>
</tr>
</tbody>
</table>
Table 4.8 Analysis of Variance Table for Contrast Decomposition II of Eq. (4-38)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Factor A(B(2)) =&gt; A(V)</td>
<td>SS_A(V) = SS_A(V) = SS_A + SS_AxV</td>
<td>MS_A(V) = \frac{SS_A(V)}{2(a-1)}</td>
<td>F_{0}^{A(V)} = \frac{MS_A(V)}{MS_E}</td>
</tr>
<tr>
<td>Fixed Factor B^V =&gt; V</td>
<td>SS_V = SS_V = \frac{na}{4} \left( \bar{y}_1 + \bar{y}_2 - \bar{y}_3 \right)^2</td>
<td>MS_V = SS_V</td>
<td>F_{0}^{V} = \frac{MS_V}{MS_A(V)}</td>
</tr>
<tr>
<td>Fixed Factor B^W =&gt; W</td>
<td>SS_W = SS_W = \frac{na}{4} \left( \bar{y}_1 + \bar{y}_2 - \bar{y}_3 - \bar{y}_4 \right)^2</td>
<td>MS_W = SS_W</td>
<td>F_{0}^{W} = \frac{MS_W}{MS_{(AxW)(V)}}</td>
</tr>
<tr>
<td>Fixed Factor B^D =&gt; VxW</td>
<td>SS_{VxW} = SS_{VxW} = \frac{na}{4} \left( \bar{y}_1 + \bar{y}_2 + \bar{y}_3 \right)^2</td>
<td>MS_{VxW} = SS_{VxW}</td>
<td>F_{0}^{VxW} = \frac{MS_{VxW}}{MS_{(AxW)(V)}}</td>
</tr>
<tr>
<td>Random Factor [AxB^L(R) B^H(R)] =&gt; AxW(V)</td>
<td>SS_{(AxB^L)(V)} = SS_{AxB^L(V)} = SS_{A} + SS_{AxB^L}</td>
<td>MS_{(AxB^L)(V)} = \frac{SS_{(AxB^L)(V)}}{2(a-1)}</td>
<td>F_{0}^{(AxB^L)(V)} = \frac{MS_{(AxB^L)(V)}}{MS_E}</td>
</tr>
<tr>
<td>Error E</td>
<td>SS_E = \sum_{i=1}^{a} \sum_{j=1}^{k} \sum_{l=1}^{n} (y_{ijkl} - \bar{y}_{ijl})^2</td>
<td>E(\text{MS}_E) = \frac{SS_E}{4a(n-1)}</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SS_T = \sum_{i=1}^{a} \sum_{j=1}^{k} \sum_{l=1}^{n} (y_{ijkl} - \bar{y}_{ijl})^2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The corresponding estimation of this variance component is:

\[ \hat{\sigma}^2_{Q_1(V)} = \hat{\sigma}^2_{e(V)} = (MS_A(V) - MS_E) / 2n \] (4-44)

A similar F-test can be conducted for the nested Factor A(W) and estimation of \( \hat{\sigma}^2_{Q_1(W)} \).

### 4.3 CASE STUDY

A case study was conducted by initially analyzing six runs of old production data to provide some suggestions for process improvement. A follow-up validation is further made through collecting and analyzing another six runs of new production data after the process takes the corresponding improvement.

First, the variance component analysis is conducted based on two factors, i.e., a random run factor A and a fixed site factor B. If the effect of Factor B is significant, a
further diagnostic analysis is made for the contrast variation components through three factors, i.e., run factor A, front-back factor V, and left-right factor W. For the old production data, Table 4.9 shows the sum of squares of the full decomposition model based on Factor A and Factor B with the levels of $a=6$ and $b=4$ respectively.

Table 4.9 Sum of squares of the full decomposition for two factors

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>A×B</th>
<th>E</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>0.045664</td>
<td>0.017633</td>
<td>0.049859</td>
<td>0.085920</td>
<td>0.199075</td>
</tr>
<tr>
<td>Df</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>48</td>
<td>71</td>
</tr>
</tbody>
</table>

The analysis of the interested variance components $Q_1$, $Q_2$, and $Q_3$ are made and summarized in Table 4.10, where P-value indicates the significant level of F-tests.

Table 4.10 F-test of variance components for old process

<table>
<thead>
<tr>
<th>Variation Source</th>
<th>Sum of Squares (SS)</th>
<th>Degree of Freedom (df)</th>
<th>Mean Square (MS)</th>
<th>F-test ($F_0$)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A $\Rightarrow Q_1$</td>
<td>$SS_A = 0.045664$</td>
<td>$df_A = a - 1 = 5$</td>
<td>$MS_A = \frac{SS_A}{df_A} = 0.009133$</td>
<td>$\frac{MS_A}{MS_E} = 5.10$</td>
<td>0.00079</td>
</tr>
<tr>
<td>A(B) $\Rightarrow Q_2$</td>
<td>$SS_{A(B)} = SS_A + SS_{A×B}$</td>
<td>$df_{A(B)} = b(a - 1) = 20$</td>
<td>$MS_{A(B)} = \frac{SS_{A(B)}}{df_{A(B)}} = 0.004776$</td>
<td>$\frac{MS_{A(B)}}{MS_E} = 2.67$</td>
<td>0.0028</td>
</tr>
<tr>
<td>B (A) $\Rightarrow Q_3$</td>
<td>$SS_{B(A)} = SS_B + SS_{A×B}$</td>
<td>$df_{B(A)} = a(b - 1) = 18$</td>
<td>$MS_{B(A)} = \frac{SS_{B(A)}}{df_{B(A)}} = 0.003750$</td>
<td>$\frac{MS_{B(A)}}{MS_E} = 2.09$</td>
<td>0.021</td>
</tr>
<tr>
<td>Error $E$</td>
<td>$SS_E = 0.08592$</td>
<td>$df_E = ab(n-1) = 48$</td>
<td>$MS_E = \frac{SS_E}{df_E} = 0.00179$</td>
<td>$\frac{MS_E}{MS_E}$</td>
<td></td>
</tr>
</tbody>
</table>

From this analysis, it can be seen that the batch mean variation $Q_1$ is significant with a Type I error no larger than 0.00079. Also, the site effect is significant, which is indicated by the F-test for $Q_2$ with a Type I error no larger than 0.0028, and for $Q_3$ with a Type I error no larger than 0.021. Therefore, the process improvement should focus on both batch mean variation reduction and site variation reduction. The estimation of each
variance components are calculated from Eqs. (4-11)-(4-13) as:  
\[ \hat{\sigma}_{Q_1}^2 = 0.00061, \]
\[ \hat{\sigma}_{Q_2}^2 = 0.001, \hat{\sigma}_{Q_3}^2 = 0.00065. \]

In order to effectively find the root causes of site variations, the analysis of contrast variance components is further conducted. The sum of squares of the full decomposition model in Eq. (4-37) is summarized in Table 4.11. The further statistical F-tests of the contrast variation components by using the nested models of Eq. (4-38) and Eq. (4-39) are summarized in Table 4.12.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F test</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q_f(B_{FB}) ) ( \Rightarrow A(B_{FB}) \Rightarrow A(V) )</td>
<td>( SS_{A(V)} = SS_A + SS_{A\cdot V} )</td>
<td>( MS_{A(V)} = \frac{SS_{A(V)}}{2(a-1)} = 0.008376 )</td>
<td>( \frac{MS_{A(V)}}{MS_E} = 4.68 )</td>
<td>0.00016</td>
</tr>
<tr>
<td>( Q_f(B_{LR}) ) ( \Rightarrow A(B_{LR}) \Rightarrow A(W) )</td>
<td>( SS_{A(W)} = SS_A + SS_{A\cdot W} )</td>
<td>( MS_{A(W)} = \frac{SS_{A(W)}}{2(a-1)} = 0.0055327 )</td>
<td>( \frac{MS_{A(W)}}{MS_E} = 3.09 )</td>
<td>0.0041</td>
</tr>
<tr>
<td>( Q_2^2 ) ( (A) \Rightarrow B_{FB}(A) \Rightarrow V(A) )</td>
<td>( SS_{V(A)} = SS_V + SS_{V\cdot A} )</td>
<td>( MS_{V(A)} = \frac{SS_{V(A)}}{a} = 0.007625 )</td>
<td>( \frac{MS_{V(A)}}{MS_E} = 4.26 )</td>
<td>0.0016</td>
</tr>
<tr>
<td>( Q_2^2 ) ( (A) \Rightarrow B_{LR}(A) \Rightarrow W(A) )</td>
<td>( SS_{W(A)} = SS_W + SS_{W\cdot A} )</td>
<td>( MS_{W(A)} = \frac{SS_{W(A)}}{a} = 0.002262 )</td>
<td>( \frac{MS_{W(A)}}{MS_E} = 1.26 )</td>
<td>0.29</td>
</tr>
<tr>
<td>( Q_2^2 ) ( (A) \Rightarrow B_{LR}(A) \Rightarrow W(A) )</td>
<td>( SS_{V(W)(A)} = SS_{V\cdot W}(A) + SS_{A\cdot V\cdot W} = 0.008175 )</td>
<td>( MS_{V(W)(A)} = \frac{SS_{V(W)(A)}}{a} = 0.001362 )</td>
<td>( \frac{MS_{V(W)(A)}}{MS_E} = 0.7 )</td>
<td>0.60</td>
</tr>
<tr>
<td>Error E</td>
<td>( SS_E = 0.08592 )</td>
<td>( MS_E = \frac{SS_E}{4a(n-1)} = 0.001790 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is clear that the run-by-run variations nested by Factor V and Factor W \([A(V)\) and \(A(W)\)] are all significant. Therefore, the tool setup repeatability should be enhanced to reduce the run-by-run variation. However, for the within-run variation, only the effect
of the back-front contrast V(A) is significant with a Type I error no more than 0.0016.
Thus, the process improvement strategy for reducing the within-run variation should
mainly focus on reducing the back-front error in every run. It turns out that the fixed
machine tabletop position error is the major contribution of this within-run site variation.
The estimates of each significant variance component can be calculated from Eqs. (4-34)
and (4-44) yielding $\hat{\sigma}_{\nu(t)}^2 = 0.00097$, $\hat{\sigma}_{\nu(T)}^2 = 0.0011$, and $\hat{\sigma}_{\nu(o)}^2 = 0.00062$.

After a careful adjustment of the printing machine tabletop and making the
improvement on the inspection of tool setup alignment and ink variability, six runs of
new process data are collected again for validation analysis. The variance component
analysis based on the full factor decomposition model of this new process is shown in
Table 4.13. The F-test results of the nested variance components are given in Table 4.14.
Now, it shows there is no significant factor anymore. The process improvement
efficiency can be indicated by the percentage of the total variation reduction as:

$$\eta = \frac{\text{Var}(old) - \text{Var}(new)}{\text{Var}(old)} \times 100\% = \frac{SS_T(old) - SS_T(new)}{SS_T(old)} \times 100\% = 41.13\%$$

| Table 4.13 Sum of squares based on the full decomposition of the new process. |
|---|---|---|---|---|---|
| A | B | AxB | E | Total |
| SS | 0.008621 | 0.002841 | 0.021428 | 0.08425 | 0.11714 |
| Df | 5 | 3 | 15 | 48 | 71 |
### Table 4.14 F-test of variance components for the new process

<table>
<thead>
<tr>
<th>Variation Source</th>
<th>Sum of Squares (SS)</th>
<th>Degree of Freedom (df)</th>
<th>Mean Square (MS)</th>
<th>F-test (F₀)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ⇒ Q₁</td>
<td>SSₐ = 0.008621</td>
<td>dfₐ = a-1 = 5</td>
<td>MSₐ = 0.001724</td>
<td>MSₐ / MSₑ  = 0.98</td>
<td>0.44</td>
</tr>
<tr>
<td>A(B) ⇒ Q₂</td>
<td>SSₐ(B) = SSₐ + SSₐB = 0.030049</td>
<td>dfₐ(B) = b(a-1) = 20</td>
<td>MSₐ(B) = 0.001502</td>
<td>MSₐ(B) / MSₑ = 0.86</td>
<td>0.64</td>
</tr>
<tr>
<td>B (A) ⇒ Q₃</td>
<td>SSₐ(B) = SSₐB + SSₐₜA = 0.024268</td>
<td>dfₐ(B) = a(b-1) = 18</td>
<td>MSₐ(B) = 0.001348</td>
<td>MSₐ(B) / MSₑ = 0.77</td>
<td>0.72</td>
</tr>
<tr>
<td>Error E</td>
<td>SSₑ = 0.08425</td>
<td>dfₑ = ab(n-1) = 48</td>
<td>MSₑ = 0.00175</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 4.4 CONCLUSION

In this Chapter, the variation decomposition and analysis for batch manufacturing processes has been conducted by using the ANOVA method. A generic framework is provided for decomposition of three typical variation components in batch manufacturing processes. For the significant site variation, three diagnostic contrast components are defined for root cause identification of the potential site variation patterns. When the interested variations are induced by more than three factors, it generally needs many different nested models for various variation component analyses. The use of a full factor decomposition model to determine the needed nested models is suggested in this chapter, which can expedite the statistical model development. Finally, this chapter illustrates the effectiveness of the proposed method in the screening conductive gridline printing process.

Although the proposed variation analysis method is illustrated in a specific application of the printing process, the proposed analysis method can be applied to other
batch manufacturing processes where separation of the batch-by-batch variation and the within-batch variation is needed. Also, when the potential variation patterns can be well defined by linear contrasts, the use of the diagnostic contrasts can expedite the variation root cause determination. If the contrast effects cannot be transferred into factors' effect, a direct decomposition of the sum of squares of the contrast components is needed to the development of the nested effect models.
CHAPTER 5

MULTISCALE STATISTICAL PROCESS MONITORING USING WAVELETS
ANALYSIS FOR AUTOCORRELATED DATA

This chapter proposes a wavelets-based multiscale monitoring method for process change detection with autocorrelated data, especially for a commonly used dynamic process of ARMA \((n, m)\) with \(n \geq 2\). Both types of process changes of mean shift and variance change will be addressed specifically. Based on the relationship between the signal variance in the time domain and signal spectrum in the frequency domain, a set of multiscale monitoring charts are systematically developed. For variance change detection, wavelet-based SD-CUSUM (standard deviation CUSUM) charts denoted as SD-WCUSUM charts are constructed. Two SD-WCUSUM charts are developed to monitor two different levels of wavelet detail coefficients for detecting process variance change and measurement error variance change respectively. It has been shown that the proposed SD-WCUSUM chart is only sensitive to variance change but not to mean shift. This property is in contrast with the traditional SD-CUSUM monitoring chart which is sensitive to both mean and variance changes. For the purpose of mean shift detection, a new wavelet-based EWMA chart called WEWMA chart is proposed by monitoring wavelet scale coefficients. It shows that WEWMA chart will achieve a better performance on detecting small mean shifts than Direct-EWMA charts especially for a dynamic system with high frequency responses. One critical issue in developing the proposed SD-WCUSUM charts or WEWMA charts is how to determine an appropriate decomposition level of detail coefficients or scale coefficients to monitor. A systematic
method for automatic determination of the best decomposition level is investigated correspondingly by fully utilizing the system frequency response information.

This chapter is organized as follows: Section 5.1 gives a brief review of wavelets transform and multiscale analysis used in this chapter. In Section 5.2, the general model of dynamic processes and associated process faults will be given. Section 5.3 will discuss multiscale SD-WCUSUM chart development for variance change detection. A systematic procedure for selecting optimal level of detail coefficients will be discussed. For efficient detection of different variance increasing due to process condition changes or sensor performance changes, detail coefficients at different levels should be selected. Section 5.4 presents a WEWMA chart for small mean shift change detection using scale coefficients at a selected level. The impact of the system frequency response on the monitoring performance is also discussed. The ARL is used to compare the performance of the proposed WEWMA chart with Model Based EWMA and Direct- EWMA charts. Finally, Section 5.5 discusses the general procedures of using the proposed multiscale monitoring charts for simultaneously detecting both process mean shift and variance change separately.

5.1 REVIEW OF WAVELET TRANSFORM AND MULTISCALE ANALYSIS

5.1.1 Wavelet Transform

If \( y(t) \) is an integrable function on \( L^2(R) \), i.e., \( \int_{-\infty}^{\infty} y(t)^2 dt < \infty \), it can be expressed by wavelets decomposition as (Daubechies, 1992):
where \( c_{j_0, k} \) is called a scale or approximation coefficient at level \( j_0 \) and location \( k \), and \( d_{j, k} \) is called a detail coefficient at level \( j \) and location \( k \), which can be calculated respectively as:

\[
c_{j_0, k} = \int y_t \cdot \phi_{j_0, k}(t) \, dt \quad ; \quad d_{j, k} = \int y_t \cdot \psi_{j, k}(t) \, dt
\]  

(5-2)

The function \( \phi(t) \) and \( \psi(t) \) are two basic functions known as the scaling function and mother wavelet respectively (Cohen, 1992), which satisfy the following multiresolution property defined as:

\[
\phi_{j, k}(t) = 2^{-j/2} \cdot \phi(2^{-j} t - k) \quad ; \quad \psi_{j, k}(t) = 2^{-j/2} \cdot \psi(2^{-j} t - k)
\]

(5-3a)

For example, the scaling function and mother wavelet of Haar wavelets are:

\[
\phi(t) = \begin{cases} 
1 & \text{for } 0 < t < 1 \\
0 & \text{otherwise}
\end{cases} \quad ; \quad \psi(t) = \begin{cases} 
1 & \text{for } 0 < t < 1/2 \\
-1 & \text{for } 1/2 < t < 1 \\
0 & \text{otherwise}
\end{cases}
\]

(5-3b)

In general, a normalized scaling function is used with \( \int \phi(t) \, dt = 1 \). Therefore, \( d_{j, k} \) quantifies the basis function \( \psi_{j, k}(t) \) yielding information of \( y_t \) near position \( k \), or time \( k \) in this dissertation, on the scale of \( 2^{-j} \) (i.e., at the frequency near \( 2^{-j} \omega_0 \), where \( \omega_0 \) is the center frequency of the mother wavelet \( \psi(t) \)). Similarly, \( c_{j, k} \) quantifies the basis function \( \phi_{j, k}(t) \).

When a discrete wavelet transform \( W \) is used for a data set \( Y = [y_1, y_2, ..., y_N]^T \),
with \( N = 2^n \), all transformed wavelet coefficients of vector \( C \) can be obtained by:

\[
C = WY \quad (5-4)
\]

where the rows of \( W \) correspond to a discretized version of the mother wavelet at various different scaling and translation. If \( W \) is an orthogonal wavelet, the original data \( Y \) can be recovered by the inverse discrete wavelet transform (without considering the boundary filter effect) as:

\[
Y = W^T C \quad (5-5)
\]

5.1.2 Multiscale Analysis

The wavelet transform has a unique merit of multiscale resolution analysis, that is, \( V_L \) (\( L = n-j, j \in Z \), \( L \) is the resolution level and \( j \) is the decomposition level) can be decomposed by a nested sequence of closed subspaces as (Chui 1992):

\[
V_L = V_{L-1} \oplus W_{L-1} = \cdots \oplus W_{L-2} \oplus W_{L-1} \quad (5-6)
\]

where \( W_{L-1} \) is the orthogonal complement of \( V_{L-1} \) with respect to \( V_L \),

\[
W_L = \text{closure}_{L^2(R)} \left\{ \psi_{j,k} : k \in Z \right\} \quad (5-7)
\]

The defined subspaces \( V_L \) satisfy:

\[
\begin{align*}
(1^0) & \quad V_{-1} \subset V_0 \subset V_1 \subset \cdots \\
(2^0) & \quad \text{closure}_{L^2(R)} \left( \bigcup_{L \in Z} V_L \right) = L^2(R) \\
(3^0) & \quad \bigcap_{L \in Z} V_L = \{0\} \\
(4^0) & \quad V_L = V_{L-1} \oplus W_{L-1}, \quad L \in Z \\
(5^0) & \quad y(t) \in V_L \Leftrightarrow y(2t) \in V_{L+1}, \quad L \in Z
\end{align*} \quad (5-8)
\]

The intuitive meaning of \((5^0)\) in Eq. \((5-8)\) is that when passing from \( V_L \) to \( V_{L+1} \), the function resolution is doubled. For an original data set \( Y = [y_1, y_2, \cdots, y_n] \in V_n \), the
The maximum decomposition level is \( n \). The data set \( Y \) can be decomposed in different resolution levels as:

\[
V_n = V_{n-1} \oplus W_{n-1} \\
= \cdots \\
= V_0 \oplus W_0 \oplus W_1 \oplus \cdots \oplus W_{n-2} \oplus W_{n-1}
\] (5-9)

Normally, the decomposition is stopped at a satisfied resolution level \( L>0 \) (or a decomposition level \( j_0 < n \)) instead of the coarsest level \( (L=0, j_0 = n) \). The stop level \( j_0 \) is determined by the resolution requirement based on applications. In Section 5.2 and Section 5.3, the detail discussion will be given on how to select an appropriate decomposition level \( j_0 \) for process change detection. Different optimal decomposition levels will be determined according to the different purposes of process variance change detection and mean shift detection.

5.1.3 Signal Decomposition by Filter Bank Using Haar Transform

The Haar transform is the first order of Daubechies wavelet transform denoted as DB1. For a time series signal \( y_t \), the Haar transform of \( y_t \) can be easily implemented by using filter banks as shown in Figure 5.1. \( h_i \) and \( l_i \) are the corresponding high pass and low pass filter at level \( i \) which are defined as:

\[
h_i = \frac{1}{\sqrt{2}} [-1 \ 1]; \quad l_i = \frac{1}{\sqrt{2}} [1 \ 1] \quad i = 1, 2, \ldots
\] (5-10)

The wavelets coefficients in Figure 5.1 are redundant when used to reconstruct the original signal [Burrus et al., 1998]. Therefore, coefficients after down sampling are used in Eq. (5-1). In this dissertation, for the purpose of online detection, coefficients in
Figure 5.1 are directly used.

\[ y_i \xrightarrow{h_i} y \xrightarrow{l_i} \]

Figure 5.1 Filter bank structure for wavelets transform

The frequency responses of filter \( h_i \) and \( l_i \) are given in Figure 5.2, in which \( x \)-axis called the normalized frequency is the ratio of the actual frequency to the Nyquist Frequency (half of the sampling frequency).

![Haar Transform Filters](image)

Figure 5.2 Filter frequency response using Haar Transform

Based on the iterative multiscale decomposition using the Haar basis, the scale coefficients \( c_{i,j}^v \) and detail coefficients scale signal \( d_{i,j}^v \) can be obtained by:
If process changes occur at time $k$ with $k \geq 2^j$ ($j_0$ is the maximum decomposition level), those coefficients calculated by Eq. (5-11) and Eq. (5-12) can be used for the online process monitoring. Therefore, the proposed approach of constructing monitoring control charts on the wavelet coefficients can be taken as a model free approach since it does not rely on a process model.

5.2 DESCRIPTION OF PROCESS MODELS AND CHARACTERISTICS

5.2.1 ARMA($n$, $m$) Model and Process Fault Conditions

An autocorrelated stationary process under the normal operation condition can be generally modeled by an ARMA ($n$, $m$) model represented as:

$$A(q)x_t = D(q)a_t$$ (5-13)

where $a_t$ is an i.i.d. Gaussian process with $a_t \sim NID(0, \sigma_a^2)$. $A(q)$ and $D(q)$ are polynomial functions of backshift operator defined by $q^{-1}$ ($q^{-1}x_t = x_{t-1}$), which are refereed to as AR and MA respectively:

$$A(q) = 1 - \phi_1 q^{-1} - \phi_2 q^{-2} - \ldots - \phi_n q^{-n}$$ (5-14)

$$D(q) = 1 - \theta_1 q^{-1} - \theta_2 q^{-2} - \ldots - \theta_n q^{-m}$$ (5-15)

To consider the inevitable measurement error $e_t \sim NID(0, \sigma_e^2)$ as shown in Figure 5.3, an autocorrelated measurement data $y_t$ under a normal operation condition can be
generally modeled by

\[ y_t = x_t + e_t \]
\[ x_t = \phi_1 x_{t-1} + \cdots + \phi_n x_{t-n} - \theta_1 a_{t-1} - \cdots - \theta_m a_{t-m} + a_t \]  

(5-16)

Figure 5.3 Process model structure

Therefore, we have \( H_0 : y_t \sim N(0, \sigma_x^2 + \sigma_e^2) \) under the normal process condition. In this chapter, process monitoring control charts will be developed to detect the occurrence of following possible faulty conditions:

Fault 1: \( H_1^1 : a_t \sim N(0, \delta_a^2 \sigma_a^2) \);

Fault 2: \( H_1^2 : e_t \sim N(0, \delta_e^2 \sigma_e^2) \);

Fault 3: \( H_1^3 : a_t \sim N(\mu_a, \sigma_a^2) \).

(5-17)

where Fault 1 and Fault 2 correspond to the variance change ratio with \( \delta_a > 1 \) and \( \delta_e > 1 \), and Fault 3 corresponds to the mean shift with the relative shift of \( \delta_\mu = \frac{\mu_a}{\sigma_a} \). The procedure for monitoring those faults will be discussed in subsections 5.3-5.5.

5.2.2 Autospectrum Analysis of ARMA \((n, m)\)

A deterministic output from a system can be expressed as a mixture of sine and cosine waves at different frequencies analyzed by using FFT. Similarly, for a time series
or a stationary stochastic system response, it is shown that the Fourier transform of the 
autocovariance function is a well-behaved deterministic function for which the usual FFT 
can be used (Pandit and Wu, 1983). This Fourier transform of the autocovariance 
function, which is known as autospectrum or power spectrum, can effectively show how 
the variance of the process output $x_i$ is distributed over the frequency bands. If a 
process model of Eq. (5-13) is known, a direct autospectrum representation of $x_i$ under 
the normal process condition can be obtained as (Pandit and Wu, 1983):

$$f_0^x(\omega) = \frac{\Delta \sigma_a^2}{2\pi} |e^{j\omega \Delta_t} - \theta_{1}e^{j(m-1)\omega \Delta_t} - \cdots - \theta_{m}e^{jn\omega \Delta_t}|^2$$  
(5-18)

where $\Delta_t$ is the sampling time interval and $\omega$ is the angular frequency in radians per unit time with the region of $-\frac{\pi}{\Delta_t} \leq \omega \leq \frac{\pi}{\Delta_t}$. When the process model is unknown, the autospectrum can be estimated from samples by taking the Fourier transform of sample autocovariance (Kay, 1988). From Eq. (5-18), it can be seen that the autospectrum of $e_t$, which is white noise under the normal condition, is constant over all frequency bands:

$$f_0^e(\omega) = \frac{\Delta \sigma_e^2}{2\pi}$$  
(5-19)

5.2.3 Response Characteristics of Variance Change

It can be seen that the autospectrum is proportional to the process variance $\sigma_a^2$ at a given frequency band based on Eqs. (5-18) and Eq. (5-19). Therefore, for the variance
change defined by the fault conditions of $H_1^1$ and $H_1^2$, the autospectrum ratio under a fault condition to that under the normal condition will have the following relationship:

$$\frac{F^x_{H_1^1}(\omega)}{F^x_{H_0}(\omega)} = \delta_a^2,$$

$$\frac{F^y_{H_1^1}(\omega)}{F^y_{H_0}(\omega)} = \delta_a^2 \cdot \frac{f^x_0(\omega) + f^e_0(\omega)}{f^x_0(\omega) + f^e_0(\omega)}$$

(5-20)

$$\frac{F^e_{H_1^1}(\omega)}{F^e_{H_0}(\omega)} = \delta_e^2,$$

$$\frac{F^y_{H_1^1}(\omega)}{F^y_{H_0}(\omega)} = \frac{f^x_0(\omega) + \delta_e^2 f^e_0(\omega)}{f^x_0(\omega) + f^e_0(\omega)}$$

(5-21)

From Eq. (5-20), it can be seen that if there are no measurement errors, the spectrum changes will have a constant ratio over all frequency bands, and the ratio does not depend on the spectrum magnitude $f^x_0(\omega)$ under the normal condition. However, when measurement errors are not ignorable, the variance change ratio will be affected by both the original spectrum $f^x_0(\omega)$ and the measurement error spectrum $f^e_0(\omega)$.

5.2.4 Response Characteristics of Mean Shift

This subsection is used to understand the characteristics of process responses in both time and frequency domains when a mean shift occurs. If the process mean shift $\mu_a = \delta \mu a$ (Fault 3) occurs at time $\tau$, it will act like a step input added into the process.

The resultant step response will be added to the output $x_t$. Based on the impulse response representation of the model in Eq. (5-13), it can be obtained as

$$x_t = \sum_{j=0}^{t-1} g_j a_{t-j} + \sum_{j=\tau}^{t} g_j (a_{t-j} + \mu_a); \text{ for } t \geq \tau$$

(5-22)

where $g_j$ is the impulse response function or Green function (Pandit and Wu, 1983). It
is known that if a process is stable, $g_j$ will converge to zero when $j$ is large enough (e.g. $j>M$). Based on Eq. (5-22), the mean of $x_i$ is obtained by:

$$
\mu_{x_i} = \sum_{j=\tau}^{t} g_j \mu_a = \mu_a \sum_{j=\tau}^{\min(t,M)} g_j; \quad \text{for } t \geq \tau
$$

(5-23)

Therefore, the mean of $x_i$ will converge to a constant value after the dynamic transition time period of $M$.

Different from the existing research that mainly focuses on autocorrelated processes with a slow dynamic change behavior described by AR(1) or ARMA(1,1) model, this research will focus more on high order autocorrelated processes represented by ARMA($n$, $m$) with $n \geq 2$, in which the normal system response spectrum falls in the frequency region larger than $0.1 \cdot \omega_q$ ($\omega_q = \pi / \Delta_q = 2\pi f_q$, $f_q$ called Nyquist frequency) under the normal process condition. As examples, three following models given in Table 5.1 will be used in this chapter for the control chart development. It can be seen that all models are stable since all their characteristic roots are in the unit circle. The corresponding autospectrum of those models are given in Figure 5.4 where the $x$-axis is the unified frequency represented by the ratio of the true frequency to the Nyquist frequency, $y$-axis gives the unified spectrum magnitude which equals to the ratio of the true spectrum magnitude to the maximum spectrum magnitude. Figure 5.5 plots the corresponding unit step response of each model. Based on Eq. (5-22), it shows that the step response will finally converge to a constant value. Thus, a significant low frequency component can be observed from the corresponding spectrum of the step response as shown in Figure 5.6. Therefore, for variance monitoring, we should select such wavelet
coefficients that are insensitive to those low frequency components of the step response.

As a result, the constructed variance monitoring charts will robust to the process mean shift.

Table 5.1 Exemplary models and their characteristics

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Characteristic roots</th>
<th>Characteristic Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 (M1)</td>
<td>$A(q) = 1 - 0.99 q^{-1} + 0.49 q^{-2}$</td>
<td>$B(q) = 1 - 0.7 q^{-1}$</td>
<td>$0.4950 \pm 0.4949i$</td>
</tr>
<tr>
<td>Model 2 (M2)</td>
<td>$1 - 0.6 q^{-1} + 0.4 q^{-2}$</td>
<td>$1 - 0.7 q^{-1}$</td>
<td>$0.3000 \pm 0.5568i$</td>
</tr>
<tr>
<td>Model 3 (M3)</td>
<td>$1 - 0.1 q^{-1} + 0.8 q^{-2}$</td>
<td>$1 - 0.7 q^{-1}$</td>
<td>$0.0500 \pm 0.8930i$</td>
</tr>
</tbody>
</table>

Figure 5.4 Autospectra of the models in Table 5.1
Figure 5.5 Unit step response of the models in Table 5.1

Figure 5.6 Autospectrum of the unit step responses of the models in Table 5.1
5.2.5 Relationship of Wavelet Coefficients and System Frequency Responses

Based on the frequency response of the Haar transform as shown in Figure 5.2, it can be seen that the high pass filter used to generate detail coefficients will suppress the low frequency components very well, and the low pass filter to obtain the scale coefficients will suppress the high frequency components well. From the properties of filter banks used for wavelets transform as shown in Figures 5.1 and 5.2, if the cut off frequency is selected as 50% reduction, the normalized frequency range of the scale coefficients at decomposition level $j$ will be equal to $[0, 2^{-j})$, while the normalized frequency of the detail coefficients at level $j$ will be $[2^{-j}, 2^{-j+1})$. So, the low frequency response of a mean shift (in general, the normalized frequency $\gamma \leq 0.1$ as shown in Figure 5.6) will be mainly reflected on the scale coefficients and suppressed on the detailed coefficients. So, the detail coefficients obtained through the high pass filter will contain little mean shift information in terms of the frequency domain. As a result, variance monitoring using the detail coefficients will be robust to the process mean change.

Moreover, if the system frequency response mainly falls in the relative high frequency $\gamma > 0.1$, process monitoring of mean shift using scale coefficients can suppress the impact of a variance change. The reason is the frequency response of process variance change will cause the same proportional magnitude change of the spectrum over all frequency bands as given in Eq. (5-18). The scale coefficients through low pass filters will suppress the high frequency components. When the significant system frequency
responses fall in the high frequency range, monitoring scale coefficients for mean shift
detection will be robust to the process variance change.

5.3 VARIANCE CHANGE MONITORING USING DETAIL COEFFICIENTS

5.3.1 Review of SD-CUSUM Chart for Process Variability Monitoring

The standard deviation CUSUM (also called scale CUSUM) chart for process
variability monitoring can be conducted for an individual observation, which is denoted
as SD-CUSUM chart in this dissertation. It is assumed that each observation \( y_i \) be the
normally distributed process measurement with \( y_i \sim N(\mu_0, \sigma^2) \). The standardized
monitored variable \( z_i \) is defined as (Hawkins, 1981):

\[
z_i = \frac{y_i - \mu_0}{\sigma}
\]

(5-24)

\[
v_i = \left( \sqrt{z_i} - 0.822 \right) / 0.349
\]

(5-25)

Since the in-control distribution of \( v_i \) approximately follows \( N(0,1) \), two one-side
standardized stand deviation cusums \( S_i^+ \) and \( S_i^- \) can be established as follows:

\[
S_i^+ = \max[0, v_i - k + S_{i-1}^+]; \quad S_i^- = \max[0, -k - v_i + S_{i-1}^-]
\]

(5-26)

where \( S_0^+ = S_0^- = 0 \), and the value of \( K \) and \( h \) are selected as in the regular CUSUM chart
(denoted as M-CUSUM) for process mean monitoring. The interpretation of SD-
CUSUM is similar to the interpretation of the M-CUSUM for the mean. If the process
standard deviation increases, the value of \( S_i^+ \) will increase and eventually exceed \( h \),
whereas if the process standard deviation decreases, the value of \( S_i^- \) will increase and
eventually exceed $h$. In fact, monitored variable $v_i$ is sensitive to both mean and variance changes. Hawkins (1993) suggests plotting two CUSUM charts for the mean and standard deviation monitoring together. If only SD-CUSUM shows out-of-control, one would suspect a change in variance, but if both SD-CUSUM and M-CUSUM charts signal out-of-control, one would suspect a shift in the mean. However, when a process has both standard deviation change and mean shift, it may cause the miss detection of mean shift, as shown in Figure 5.7. In Figure 5.7, model 1 is used with the distribution of $a_i$ changing from $N(0,1)$ to $N(1, 2.5^2)$ at time 64. Because of the variance change, the M-CUSUM chart for mean shift detection generates a lot of miss detection points. Therefore, those two CUSUM charts on the original data are not sufficient in identifying process change whether due to the mean shift, standard deviation, or both changes. The following subsections will discuss how to construct monitoring charts on wavelet coefficients, which can simultaneously monitor mean shift and variance change separately.
5.3.2 Selection of Optimal Level of Detail Coefficients for Fault 1 Detection

It has been shown that detail coefficients will be less sensitive to the process mean shift. Based on Eq. (5-20) it is also seen that if there would be no measurement errors, a variance change will cause the same ratio of the spectrum change over all frequency bands. In this case, variance monitoring using detail coefficients could be conducted at any decomposition level unless it does not overlap with the frequency bands of mean shift response. However, the measurement errors are general inevitable in practice, which causes the variance change ratio decrease, i.e., \( \frac{F_{H_1}^{Y}(\omega)}{F_{H_0}^{Y}(\omega)} < \frac{F_{H_1}^{X}(\omega)}{F_{H_0}^{X}(\omega)} \). In this case, the selection of the optimal decomposition level is needed to obtain a maximal variance...
change ratio. In this subsection, we will discuss how to select an optimal decomposition level for Fault 1 detection.

As given in Eq. (5-12), the detail coefficients $d_{j,k}$ is the linear combination of measurement data ($y_t = x_t + e_t$), and the covariance of $x_t$ and $x_{t-s}$ can be obtained as (Pandit and Wu, 1983):

$$\text{cov}(x_t, x_{t-s}) = \sigma_a^2 \sum_{j=s}^{M} g_{j-s}g_j$$  \hspace{1cm} (5-27)

Therefore, in addition to a constant $\sigma_e^2$ added to all levels, the variance of $d_{j,k}$ is proportional to $\sigma_a^2$ with the coefficient depending on level $j$. In this case, the distribution of the detail coefficients can be represented by:

$$H_0: d_{j,k} \sim N(0, \rho_j \sigma_a^2 + \sigma_e^2)$$
$$H_1: d_{j,k} \sim N(0, \delta_{a,j}^2 \rho_j \sigma_a^2 + \sigma_e^2)$$  \hspace{1cm} (5-28)

The variance change ratio on the detail coefficients can be obtained as:

$$\delta_{a,j}^2 = \frac{\sigma_{d,j}^2}{\sigma_{d,j}^2} = \frac{\delta_{a,j}^2 \rho_j \sigma_a^2 + \sigma_e^2}{\rho_j \sigma_a^2 + \sigma_e^2} = \frac{\delta_{a,j}^2 \rho_j \sigma_a^2 + \sigma_e^2}{\rho_j \sigma_a^2 + \sigma_e^2} + 1 = \delta_{a,j}^2 - \frac{\delta_{a,j}^2 - 1}{1 + \xi_j}$$  \hspace{1cm} (5-29)

where $\xi_j = \rho_j \sigma_a^2 / \sigma_e^2$. Based on Eq. (5.29), since $\delta_{a,j}^2$ is greater than 1, it can be seen that $\delta_{a,j}^2 \leq \delta_a^2$, and the maximum ratio $\delta_{a,j}^2$ is achieved with the maximum $\xi_j$ at an optimal level $j^*$.

Now Model 1 and 3 will be used to illustrate how to select an optimal decomposition level of detail coefficients for process variance change detection. For both models, it is assumed that $a_i$ follows $N(0,1)$ and $e_i$ follows $N(0,\sigma_e^2)$. Under this setting,
the $\xi_j$ values of Model 1 are 0.0997, 0.2200, 0.2405 and 0.1115 for the first four detail levels respectively; and $\xi_j$ of model 3 are 0.4574, 0.7916, 0.0539 and 0.0503. $\xi_j$ is affected by the ratio of $\sigma_a^2 / \sigma_e^2$ and the value of $\rho_j$. $\rho_j$ value is a constant for a given model or system. A set of true variance change ratio $\delta_a$ and the corresponding change ratio $\delta_d$ on detail coefficients are plotted in Figure 5.8 for each level. From this plot, it is can be seen that level 3 and level 2 are the optimal monitoring levels for Model 1 and Model 3, respectively.

Figure 5.8 Relationship between $\delta_a$ and $\delta_d$ at different levels
5.3.3 Selection of Optimal Level of Detail Coefficients for Fault 2 Detection

Based on Eq. (5-16), the distribution of the detailed coefficients under the normal condition and Fault 2 condition can be represented by:

\[ H_0: d_{j,k} \sim N(0, \rho_j \sigma_a^2 + \sigma_e^2) \]
\[ H_1^1: d_{j,k} \sim N(0, \rho_j \sigma_a^2 + \delta_e^2 \sigma_e^2) \]  

In this case, the variance change ratio on the detail coefficients can be obtained as:

\[ \frac{\delta^2_{d_{j,k}}}{\sigma^2_{d_{j,k}|H_s}} = \frac{\rho_j \sigma_a^2 + \delta_e^2 \sigma_e^2}{\rho_j \sigma_a^2 + \sigma_e^2} = \frac{\delta^2_e + \xi_j}{1 + \xi_j} = 1 + \frac{\delta^2_e - 1}{1 + \xi_j} \]  

(5-31)

Based on Eq. (5-31), it can be seen that the maximum ratio \( \delta^2_{d_{j,k}} \) is achieved with the minimum \( \xi_j \) at an optimal level \( j^* \). In a high order dynamic system, a maximum \( \xi_j^* \) is corresponding to the maximal energy of detail coefficients close to the characteristic frequency of the normal system response. Therefore, a maximum \( \xi_j^* \) used for Fault 1 detection is generally found after a few steps of decomposition. However, \( \xi_j \) could be monotonically decreasing after a certain steps of decomposition when the decomposed detail coefficients’ frequency band is below all characteristic frequency components. In another aspect, with the increase of the decomposition level, the detection delay in the time domain will increase. In order to compromise these two aspects, an optimal level which is able to detect the process change and also has a minimum detection time delay should be used.

Model 1 and 3 will be used to illustrate how to select an optimal decomposition level for measurement error variance change detection. For both model, it is assumed
that both $a_i$ and $e_i$ follow $N(0,1)$. Under this setting, the $\xi_j$ values of Model 1 are 0.8976, 1.9800, 2.1648 and 1.0032 for the first four detail levels respectively; and $\xi_j$ of model 3 are 4.1162, 7.1242, 0.4852 and 0.4528. A set of true variance change ratio $\delta_e$ and the corresponding change ratio $\delta_d$ are plotted in Figure 5.9 for each level. From the plot, it can be seen that level 1 and level 4 are the best levels among the first four detail levels for Model 1 and Model 3 respectively.

![Figure 5.9 Relationship between $\delta_e$ and $\delta_d$ at different levels ($\sigma_e=1$)](image)

Figure 5.9 Relationship between $\delta_e$ and $\delta_d$ at different levels ($\sigma_e=1$)

To further illustrate the effect of the normal measurement error $\sigma_e$ on $\delta_d$ to
detect measurement error variance change, Figure 5.10 shows the relationship between $\delta_e$ and $\delta_d$ with a large measurement error of $e \sim N(0,\sigma^2_e)$, in which $\sigma_e$ does not change with $e \sim N(0,1)$. For Model 1, the plots of level 1 and level 4, level 2 and level 3 are almost overlap with each other. For Model 3, level 3 and 4 almost cannot be distinguished. These because their $\xi$ values, which are given in Case 2 study, are so close to each other.

![Graph showing the relationship between $\delta_e$ and $\delta_d$ at different levels for Models 1 and 3.](image)

Figure 5.10 Relationship between $\delta_e$ and $\delta_d$ at different levels ($\sigma_e = 3$)
Comparing Figure 5.10 with Figure 5.9, it can be found that the difference of detection ability between different detail levels becomes smaller when $\sigma_e$ is increased. Because $e_i$ is white noise which has the same effect at all the levels. However, level 1 and level 4 are still the most efficient levels among the first four levels for Model 1 and Model 3 respectively.

5.3.4 Simulation Study

In this subsection, the following case studies will be used to illustrate the effectiveness of the proposed methods for process variation monitoring. The comparison with the SD-CUSUM chart will also be given.

Case 1: Detail coefficient monitoring insensitive to process mean change

In this case, a simulation data generated from Model 1 was used to show that the proposed variance monitoring strategy is robust to a process mean shift. It is assumed that $\alpha_i$ follows $N(0,1)$ under the normal process condition and then changes to $N(3,1)$ with $\delta_a = 3$ at time $\tau = 64$. Measurement error $e_i$ follows $N(0,1)$. The control limits $h$ for M-CUSUM charts for mean shift detection and SD-CUSUM for variance change detection are set as 7 and 5.8 respectively. Control limits $h$ for wavelets based M-WCUSUM and SD-WCUSUM are set as 9.8 and 7.5 respectively. The $K$ values in all CUSUM charts are set to 0.5. All these parameters of the CUSUM charts are set to make the $\text{ARL}_0$ equal to 370. The results are shown in Figure 5.11.
In Figure 5.11, the original signal is plotted in (a) and (b). (c) and (d) are the regular M-CUSUM chart and SD-CUSUM chart, which are used to monitor process mean shift and variance change based on the original signal data. Figure 5.11 (e) and (f) are the M-WCUSUM and SD-WCUSUM chart based on the selected wavelet coefficients, in which the scale coefficients at level 3 and the detail coefficients at level 2 are used. From Figure 5.11 (d) and (f), it can be seen that SD-CUSUM is sensitive to the process mean shift, although there is no process variance change. In contrast to this, the SD-WCUSUM chart does not show any out-of-control signal. In this case study, the process
mean shift (3σ) is large which can be detected by both direct M-CUSUM and M-WCUSUM chart. The advantage of using wavelet based method to monitor scale coefficients for small mean shift detection will be discussed in Section 5.4.

Case 2: Optimal level selection for process variance change detection

As discussed before, when measurement errors exist, the detection ability for process variance change detection is different at different detail levels. From Figure 5.8, it can be seen that the detection ability difference between each detail level of Model 3 is more significant than that of Model 1. Model 3 is first chosen to illustrate how to select an optimal level for better process variance change detection. Assume \( a_j \) follows \( N(0,1) \) under normal process condition and changes to \( N(0,2.8^2) \) with \( \delta^2 = 2.8 \) from time \( t = 64 \) under the fault condition. Process measurement error \( e_i \) follows the \( N(0,3^2) \).

From the spectral analysis of Model 3 in Figure 5.4, it can be seen that the second detail level \( (j^* = 2) \) contains the maximum frequency responses, which should generate the maximum \( \rho^*_2 \) and \( \xi^*_2 \) value. From the simulation, the \( \rho \) values of the original signal and the first four detail levels are also estimated, which are 3.95, 4.1162, 7.1242, 0.4852 and 0.4528 respectively. The simulation results are consistent with the theoretical analysis results and Level 2 is chosen to monitor the process variance change. In implementation of the proposed method, we will rely on the simulation to select the optimal level which has a maximum sample \( \rho_j \). Therefore, it is a model free approach. So, the maximum \( \xi \) value can be calculated as:
\[
\xi_2 = \frac{\rho f \sigma_a^2}{\sigma_e^2} = \frac{7.1242 \times 1}{9} = 0.7916
\]  
(5-32)

and the corresponding \( \sigma_{d_2}^2 \) is:

\[
\sigma_{d_2}^2 = \sigma_a^2 - \frac{\sigma_a^2 - 1}{1 + \xi_3} = \frac{7.84 - 1}{1 + 0.7916} = 4.022
\]  
(5-33)

Therefore \( \delta_d \) is \( \sqrt{4.022} = 2.0055 \). This means that when the standard deviation of \( a \) changes with \( \delta_a = 2.8 \), the standard deviation of detail coefficients at level two changes with \( \delta_{d_2} = 2.0055 \). For comparison, we construct four CUSUM charts for each level of detail coefficients. For the parameters of these SD-WCUSUM charts, the \( K \) is set to 0.5 and the corresponding control limits of \( h \) values are adjusted to make the ARL_0 equal to about 370. The resultant \( h \) values are 9.5, 10, 8 and 8.2 to each level respectively. The results are plotted in Figure 5.12.

From Figure 5.12, it can be seen that the process variance change was detected by monitoring both level 1 (\( \delta_{d_1} = 1.77 \)) and level 2 (\( \delta_{d_2} = 2.0055 \)) detail coefficients. This is because the variance change amount \( \delta_a (2.8) \) is relatively large, level 1 can also detect such a change. Considering the robustness of detection power to the measurement errors, level 2 is the most efficient level. In the practical application, it suggests to apply SD-CUSUM chart at the optimal detail level rather than at all the levels.
Figure 5.12 Choosing an optimal detail level for process variance change detection of process with measurement error (Model 3)

Similar results can be obtained when Model 1 is used. Assume $a$, follows $N(0,1)$ under the normal process condition and thus changes to $N(0, 2.8^2)$ from time $\tau = 64$. 
Process measurement error \( e \), follows the \( N(0, 3^2) \). The results are plotted in Figure 5.13.

(a) Original signal and detail coefficients

(b) SD-WCUSUM Chart

Figure 5.13 Choosing an optimal detail level for process variance change detection of process with measurement error (Model 1)
Case 3: Optimal level selection for measurement error variance change detection

Model 1 is first investigated, in which \( e \) follows \( N(0,3^2) \) under the normal process condition and then changes to \( N(0,4.5^2) \) from time \( \tau = 64 \) under the fault condition. \( a \), follows \( N(0,1) \). \( \rho \) values of the original signal and the first four detail levels can be obtained from the simulation, which are 1.32, 0.8976, 1.9800, 2.1648 and 1.0032 respectively. Therefore, among the first four levels, level 1 with the smallest \( \rho_1 = 0.8976 \) and the minimal detection delay is selected. Although level 4 has a small value of \( \rho_4 = 1.0032 \), it has a longer detection delay than level 1. Based on the selected level 1, the \( \xi_1 \) value is calculated as:

\[
\xi_1 = \frac{\rho_1 \sigma_a^2}{\sigma_e^2} = \frac{0.8976 \times 1}{9} = 0.1
\]  

(5-34)

and the corresponding \( \delta_{a_i}^2 \) is:

\[
\delta_{a_i}^2 = 1 + \frac{\delta_e^2 - 1}{1 + \xi_1} = 1 + \frac{2.25 - 1}{1 + 0.1} = 2.14
\]  

(5-35)

Therefore \( \delta_{a_i} = \sqrt{2.14} = 1.46 \), which means that when the standard deviation of \( e \) changes with \( \delta_e = 1.5 \), the standard deviation of detail coefficients at level 1 changes with \( \delta_{a_1} = 1.46 \). The \( K \) value of the CUSUM charts is set to 0.5 and the \( h \) value of each level is adjusted to make \( ARL_0 \) equal to 370. The resultant \( h \) values are 7.4, 8, 8.5 and 11.6 for the first four detail levels respectively. The results are plotted in Figure 5.14.
Figure 5.14 Choosing an optimal detail level for measurement error variance change detection (Model 1)
From Figure 5.14, it can be seen that the variance change of $e$, was detected by both level 1 and level 4. In the practice, it suggests to only monitor the optimal level which is chosen based on the interested $\delta_e^2$ amount and the property of the system spectrum.

Similar results can be obtained when Model 3 is used, in which the same simulation condition is used as Model 1. The monitoring charts were plotted in Figure 5.15, in which the optimal level is level 4. From Figure 5.15, it can be seen that the measurement error variance change can be detected at level 3 and level 4 which have the relative small $\rho$ value of $\rho_3 = 0.4852$ and $\rho_4 = 0.4528$ respectively.

(a) Original signal and detail coefficients
Figure 5.15 Choosing an optimal detail level for measurement error variance change detection (Model 3)

5.4 MEAN SHIFT MONITORING USING SCALE COEFFICIENTS

5.4.1 Selection of Optimal Decomposition Level of Scale Coefficients for Mean Shift Detection

As discussed in Section 5.2, the scale coefficients will be used for the process mean shift monitoring. The selection of an optimal decomposition level will be determined by process frequency response range. In general, the increase of the decomposition level will suppress more autocorrelated noise at the high frequency range; however, it also causes more steps of time delay in the time domain. Therefore, an optimal level is selected with the minimal decomposition level, at which the upper boundary of the low pass filter is equal to the low boundary of the significant frequency response. For example, in Figure 5.4 and Figure 5.6, the normal process frequency
response mainly falls in the high frequency range $\gamma > 0.1$, while the mean shift falls in low frequency range of $\gamma \leq 0.1$. It is also known if the filter cut off magnitude is selected as 50%, the low pass filter frequency range of the scale coefficients is about $[0, 0.125]$ at the decomposition level 3, and about $[0, 0.0626]$ at the decomposition level 4. So, for the mean shift detection, the maximum decomposition level is generally selected as 3 or 4.

The effectiveness of using scale coefficients to monitor a mean shift can be seen from Figure 5.16, in which the process shift of $\delta_{\mu} = 1$ is hardly seen from the original data generated from Model 1 with $a_i \sim N(0,1)$ and $e_i = 0$. With the increase of the decomposition level of scale coefficients, the mean shift trend becomes more obvious by removal of the high frequency oscillation pattern.

Figure 5.16 Wavelets decomposition of the signal of Model 1 under mean shift
5.4.2 WEWMA Mean Shift Monitoring Robust to Variance Change

The following simulation results are given to illustrate the effectiveness of using scale coefficients to monitor process mean shift. It will show that when the process has both mean and variance changes, using scale coefficients as the monitored variable is more robust to the process variance change than using the original signal directly. In this simulation, Model 1 is used and the distribution of \( a_i \) changes from \( N(0,1) \) to \( N(0,2^2) \) at time \( \tau = 64 \), In Figure 5.17 (a), the original signal data and scale coefficients at first four levels are plotted. The monitoring charts of using model based SCC EWMA, wavelet-based EWMA (WEWMA) and Direct EWMA are compared in Figure 5.17 (b). The monitoring variables used in each control chart are selected as follows. The SCC EWMA is applied to the model residual obtained by the true model. WEWMA is used to monitor the level 4 scale coefficients while Direct EWMA applies directly on the original signal. The \( \lambda \) of all EWMA chart is set to 0.2. The control limits are set to make \( ARL_0 \) equal to 370. The resultant limits of each control chart are ±0.9381, ±1.5 and ±1 corresponding to SCC EWMA, WEWMA and Direct EWMA, respectively.
From Figure 5.17, it can be seen that although Direct EWMA is popularly used as a model free approach, the detection performance is severely affected by the process.
variance change, especially when a system has a significant high frequency response. For the SCC EWMA, even the true model is used, the mean shift detection performance is also degraded when variance change occurs.

A detail ARL comparison for mean shift detection of the proposed wavelet-based method and these two existing methods are given in the following section.

5.4.3 Comparison of the WEWMA with Direct-EWMA and SCC Charts for Mean Shift Detection

In order to quantitatively demonstrate the effectiveness of the mean shift detection using the scale coefficients, the ARL performance of WEWMA chart will be compared with that of Direct-EWMA charts. Moreover, the ARL performance of SCC chart using the true process model to whiten the auto-correlated data will also be compared. The simulation results were summarized in Table 5.2, in which the value in the bracket is the standard deviation of run length and the number without bracket is the average of run length. For all the three methods, the control limits are adjusted to make the in control ARL0 = 370 with $\lambda = 0.2$.

Table 5.2 Comparison of ARLs: Wavelets-based EWMA, Direct EWMA and SCC EWMA under i.i.d. noise

<table>
<thead>
<tr>
<th>Shift ($\sigma_a$)</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wavelets-based EWMA</td>
<td>SCC EWMA</td>
<td>Direct EWMA</td>
</tr>
<tr>
<td>0.5</td>
<td>30.44 (21.04)</td>
<td>39.64 (36.25)</td>
<td>94.6 (97.7)</td>
</tr>
<tr>
<td>0.75</td>
<td>16.47 (8.7)</td>
<td>18.87 (15.86)</td>
<td>35.8 (33.7)</td>
</tr>
<tr>
<td>1</td>
<td>11.89 (4.3)</td>
<td>11.23 (6.47)</td>
<td>18.1 (15.3)</td>
</tr>
<tr>
<td>1.25</td>
<td>10.5 (2.27)</td>
<td>7.36 (3.84)</td>
<td>10.6 (8.7)</td>
</tr>
<tr>
<td>1.5</td>
<td>8.7 (2.3)</td>
<td>5.26 (2.9)</td>
<td>7.37 (5.28)</td>
</tr>
</tbody>
</table>
As shown in Table 5.2, for the mean shift less than $1 \sigma_i$, WEWMA has general better detection performance than SCC EWMA and Direct-EWMA although SCC uses the perfect true model. When the mean shift is above $1 \sigma_i$, WEWMA can also achieve a comparable performance as SCC EWMA especially when a system contains significant high frequency response as in Model 3. It is clear that WEWMA always shows a better performance than Direct-EWMA chart.

5.5 MULTISCALE MONITORING OF BOTH MEAN AND VARIANCE CHANGES

It is known that a process fault may occur at an unknown time with either a single mean shift/variance change or both. In general, there are different root causes to associate with the mean shift and variance change, which requires different process improvement and control strategies. Therefore, it is always desirable to have an effective process monitoring system that not only can detect process changes but also can separate the mean shift and variance change. A multiscale process monitoring using wavelets analysis is developed for this purpose. Based on the discussion in Section 5.3 and Section 5.4, one WEWMA chart on the selected scale coefficients is used for process mean shift detection, and two SD-WCUSUM charts on the selected detail coefficients are used for process variance change detection. A general implementation procedure for the multiscale monitoring chart system development can be summarized as follows:

(a) WEWMA chart development for process mean shift detection. This EWMA chart is developed to monitor the selected scale coefficients for process mean shift detection ($\delta_u \neq 0$). The selection of an optimal decomposition level will be determined
based on the system spectrum estimated from signals under the normal working condition. A lower optimal decomposition level is selected, which is sufficient to remove the irrelevant high frequency components and also has a short time delay. The selection of the optimal level will stop at this level when the increase of the decomposition level cannot remove any more irrelevant frequency components. It should also be clarified that the maximum decomposition level is constrained by the upper bound of the significant frequency range of a step mean shift response, which usually limits the decomposition level of scale coefficients to level 3 or level 4. Finally, the EWMA chart is developed with a presetting \( \lambda \) (such as \( \lambda = 0.2 \)) and the control limits adjusted by the in control ARL\(_0\) (such as ARL\(_0\)=370).

(b) SD-WCUSUM chart development for process variance change detection. This SD-WCUSUM chart is developed to monitor the selected detail coefficients for process variance change detection (\( \delta_a > 1 \)). The optimal level \( j^* \) is selected with a maximum \( \rho_j^* \) value defined in Eq. (5-29). It is proved that the monitoring of detail coefficients at this optimal level \( j^* \) will provide the maximal detection power to the process variance change. Moreover it is also be most robust to the process measurement errors than other levels. After selecting an optimal decomposition level, the SD-WCUSUM chart on the detail coefficients is developed as regular SD-CUSUM charts.

(c) SD-WCUSUM chart development for measurement error variance change detection. This SD-CUSUM chart is developed to monitor the selected detail coefficients for measurement error variance change detection (\( \delta_e > 1 \)). The optimal level \( j^* \) is
selected to have a relatively small $\rho_f^*$ value defined in Eq. (5-31) and a shorter time delay. The monitoring of detail coefficients at this level will provide a better detection power for the measurement error variance change. Moreover, it is also robust to the process variance $\sigma_a$. Similarly, after selecting an optimal decomposition level, the SD-WCUSUM chart based on the selected detail coefficients is developed as the regular SD-CUSUM chart.

These three control charts will be used simultaneously for detecting mean shift and variance change separately. It should be pointed out that when a single chart shows out-of-control, it can directly conclude what type of process change would be possibly occur. Moreover, if only two SD-WCUSUM charts signal, it would also be possible to conclude only variance change rather than mean shift. In general, when variance change is detected, it usually reflects a severe process failure or sensor failure, which requires immediate attention to check and remove the associated root causes. Therefore, the proposed method to separately monitor mean shift and variance change can help make a quick decision for the process improvement.

In general, measurement errors can be calibrated offline and tested before each production run. If we assume there is no dramatic change on measurement errors during each production run, process monitoring can be simplified by using two monitoring control charts, that is, WEWMA for mean shift detection and SD-WCUSUM for process variance change detection. As a result, three out of control monitoring results of two control charts can be directly mapped with three possible process change condition, i.e., mean shift only, variance change only, or both mean and variance changes at the same
time. A simulation based on Model 1 in Table 5.1 is conducted for such an illustration. The measurement error $e_i$ is assumed small and unchanged. The distribution of $a_i$ changes from the normal condition of $N(0, 1)$ into each of the three following fault condition at the time $\tau = 60$: (a) mean shift with $N(1, 1)$, (b) variance change with $N(0, 2^2)$ and (c) both mean shift and variance change with $N(1, 2^2)$. For mean shift detection, the WEWMA chart is used with $\lambda$ equal to 0.2 and control limits equal to $\pm 1.5$. For process variance change detection, the SD-WCUSUM chart is developed with $K$ equal to 0.5 and control limit $h$ equal to 4.68. All the control limits are adjusted to make in control ARL equal to 370. The corresponding monitoring results for these three different process faults are shown in Figure 5.18. It can be seen that for a single mean shift or variance change, only a single chart signal a out-of-control as shown in Figure 18 (a) and (b) respectively, while both charts signal out-of-control when mean shift and variance change occur simultaneously as shown in Figure 5.18(c).
EWMA for mean
CUSUM for Variance

EWMA for mean
CUSUM for Variance

(b) Only Variance Change Occurs

(a) Only Mean Shift Occurs
5.6 CONCLUSION

In this chapter, a multiscale SPC monitoring method using wavelets analysis has been developed to detect both process mean shift and variance changes. Three monitoring charts are developed to separately detect process variance change, measurement error variance change, and process mean shift simultaneously. Based on the different frequency characteristics of scale coefficients and detail coefficients, one WEWMA chart is developed to monitor the selected scale coefficients for process mean shift detection, meanwhile two SD-WCUSUM charts are developed to monitor the selected detail coefficients for variance change detection. An optimal selection of the decomposition levels is investigated by considering system response information in the frequency
domain. Compared with the traditional method to monitor original signals directly, the proposed method to monitor the selected wavelet coefficients is more robust to measurement errors in the variance change detection and more powerful in the small mean shift detection especially when a system has a high frequency spectrum.

It should also be pointed out that the main contribution of this research is to select effective features from wavelet coefficients for process monitoring rather than simply using the original signals. If the process has no significant high frequency spectrum, such as ARMA(1,1) model, the selected optimal decomposition level will be the original signal level. In this case, there is no need to do wavelet transform for the monitoring chart development. So, it can be taken as the existing method for monitoring chart development using the original data is a special case of the proposed method. Although EWMA and SD-CUSUM charts are chosen for the selected wavelet coefficients monitoring, it can be expected that other control charts can also be applied to the selected wavelet coefficients based on the fault characteristics (for example, X-bar chart can be used for large mean shift detection). Therefore, after selecting the optimal decomposition level of wavelet coefficients, the development of the control charts on the wavelet coefficients will mainly rely on the existing methods for the regular control chart development.
6.1 CONCLUSION

The solar cell manufacturing process is a very complex process, which involves three important production stages: thin film deposition, screen printing and cell sorting, and module assembling. This dissertation mainly focuses on the variation reduction and process control at the first two production stages. The key process control problems of how to monitor, diagnose, and control the critical process variations for quality improvement are especially addressed. Several fundamental research issues have been identified and the respective general contributions are summarized as follows:

(1) **Online SPC is integrated with generalized predictive control (GPC) at the first time for effective process monitoring and control.** The thin film deposition process is a very complex process having various inevitable process disturbances. Current in-situ sensor technology development makes it feasible to implement online monitoring and feedback control for the process variation reduction. However, the traditional minimum variance control method cannot be directly applied due to its complex dynamic characteristics of the process, such as dead time estimation uncertainty, unknown and time varying disturbance patterns. Generalized predictive control that is robust to the dead time has the potential to be applied in the thin film deposition process.

One significant aspect of this dissertation research is emphasis on the importance of developing supervisory strategies through monitoring the process changes using SPC
techniques, and then revising the controller parameters accordingly. The integration strategy of how to monitor three different disturbances and how to modify the control law based on the online monitoring results are investigated in detail. It has been shown that the integration of SPC with GPC provides great potential for the development of effective controllers especially for a complex manufacturing process with a large time varying delay and different process disturbance patterns.

(2) A hierarchical ANOVA method is developed for variance component decomposition and root cause diagnosis in a screen-printing process for inherent process variation reduction based on the normal production data. Different from conventional methods through removal of the assignable causes related process variations, this dissertation focuses more on reducing inherent natural process variations based on normal production data. A generic hierarchical framework for how to apply ANOVA method to assess and diagnose the root cause of inherent process variations is developed. Three typical variation components in a batch manufacturing process are investigated which includes batch-by-batch variation, sample-by-sample variation, and site-by-site variation. For the root cause diagnosis of the site variations, three diagnostic contrast components are further defined and their significance levels are systematically checked based on the developed nested model using ANOVA. It is shown that when there are more than three factors, the variance component decomposition is very complex which generally needs multiple nested models for ANOVA. A systematic method of how to use a full factor decomposition model to determine an appropriate nested model structure is investigated for the first time in this dissertation. By using the proposed method in the
case study of the screen printing process, it was found that the significant inherent process variations are the batch-by-batch variation, and the front-back contrast variation within a batch. After identifying the root causes of those two variation components, the suggestions for the process improvement are made for the future production improvement.

(3) A Multiscale statistical process monitoring method for autocorrelated data is proposed at the first time to simultaneously monitor the process mean shift and variance change separately. Based on the spectrum properties of autocorrelated data, especially for a higher order dynamic system with ARMA \((n, m) (n \geq 2)\), three control charts have been developed to simultaneously monitor mean shift and variance change. It has been shown that the proposed SD-WCUSUM chart is only sensitive to variance change and robust to process mean shift. This property is in contrast with a traditional SD-CUSUM monitoring chart which is sensitive to both mean and variance changes. For the purpose of mean shift detection, a new wavelet-based EWMA chart called WEWMA chart is proposed by monitoring wavelet scale coefficients. It shows that WEWMA chart will achieve a better performance on detecting small mean shifts than Direct-EWMA charts especially for a dynamic system with high frequency responses. One critical issue in developing the proposed SD-WCUSUM charts or WEWMA charts is how to determine an appropriate decomposition level of detail coefficients or scale coefficients to monitor. A systematic method for automatic determination of the best decomposition level is investigated, corresponding to different process variance change and measurement error variance change respectively. It should also be clarified that the main contribution of this research is to select effective features from wavelet coefficients for process monitoring.
rather than simply using the original signals. If the process has no significant high frequency spectrum, such as the ARMA(1,1) model, the selected optimal decomposition level will be the original signal level. In this case, there is no need to do wavelet transform for the monitoring chart development. So, development of monitoring charts on the original data can be considered as a special case of the proposed method for a lower order dynamic system.

It should be pointed out that CIGS thin film deposition and printing processes represent two typical types of manufacturing processes. CIGS deposition process can be generally taken as a continuous manufacturing process having common characteristics in chemical industry. The printing process can be taken as a general batch manufacturing process which has the similarity with the wafer batch manufacturing processes. So, the methodologies developed in this dissertation based on CIGS thin film deposition and printing processes are also expected to be applied in other similar processes, in which the measurement data are autocorrelated and process changes include different patterns of mean shift and variance change at an unknown time.

6.2 FUTURE WORK

Process monitoring, diagnosing, and control for solar cell manufacturing process improvement is a new and challenging research problem. There are many remaining research issues, both fundamental research and real implementations, that need to be further investigated in the future. A few examples include:

(1) Multivariate process monitoring to consider the dependency among different process variables;
(2) MIMO controller development in thin film deposition process to ensure the thickness ratio specification among different sources;

(3) Multivariate process capability analysis to optimize the design tolerance and quality specifications at each production stage.

(4) Multistage variation propagation modeling and analysis to consider the interactions among different production stages.
REFERENCES


