View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Scholar Commons - University of South Florida

University of South Florida
Scholar Commons

Graduate Theses and Dissertations

Graduate School

5-11-2004

A Wavelet Based Multiscale Run-by-Run Controller for Multiple Input Multiple Output (MIMO) Processes

Santosh Kothamasu
University of South Florida

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the American Studies Commons

Scholar Commons Citation

Kothamasu, Santosh, "A Wavelet Based Multiscale Run-by-Run Controller for Multiple Input Multiple Output (MIMO) Processes" (2004). *Graduate Theses and Dissertations.*
https://scholarcommons.usf.edu/etd/1117

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

A Wavelet Based Multiscale Run-by-Run Controller For Multiple Input Multiple Output (MIMO) Processes

by

Santosh Kothamasu

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Industrial Engineering
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: Tapas K. Das, Ph.D.
Ashok Kumar, Ph.D.
Jose L. Zayas-Castro, Ph.D.
Qiang Huang, Ph.D.

Date of Approval:
May 11, 2004

Keywords: RbR, CMP, multiresolution, denoising, dEWMA

© Copyright 2004, Santosh Kothamasu

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES v

ABSTRACT vii

CHAPTER 1 INTRODUCTION 1
- 1.1 Run-by-Run (RbR) Control 2
- 1.2 The Basic Structure for a Run-By Run Controller 2
- 1.3 Need for Run-by-Run Control 3
- 1.4 Control Procedure 4
- 1.5 Filters or Model Tuning Strategies 4
- 1.6 A Basic SISO Run-by-Run (RbR) Control Methodology 5
- 1.7 Multiple Input Multiple Output (MIMO) Systems 6
- 1.8 Need for Multiscale Analysis 7
- 1.9 Multiscale Analysis Using Wavelets 9
- 1.10 A Brief Description of the Problem 9
- 1.11 Thesis Organization 10

CHAPTER 2 LITERATURE REVIEW 11
- 2.1 Introduction 11
- 2.2 Statistical Process Control (SPC) 11
- 2.3 Engineering Process Control (EPC) or Automatic Process Control (APC) 12
- 2.4 Integrating SPC and EPC 13
- 2.5 Evolution of Run-by-Run Process Control 13
- 2.6 Run-by-Run Control Strategies 14
- 2.7 Applications of Run-by-Run Control 15
- 2.8 Filters 15
  - 2.8.1 Linear Filtering 16
  - 2.8.2 Shewhart Filter 17
  - 2.8.3 Moving Average (MA) or Mean Filter 17
  - 2.8.4 CUSUM Filter 18
  - 2.8.5 EWMA Filter 18
- 2.9 Disadvantages of Linear Filters 20
- 2.10 Literature Review on the Multiscale Methods and their Applications 21

2.11 Fourier Analysis 21
2.11.1 Fourier Series 21
2.11.2 Fourier Transform 22
2.12 Definition of Scale (Resolution) 22
2.13 Dilation and Translation 23
2.14 A Multiresolution Analysis Example 23
2.15 Wavelets 26
2.16 Wavelets as a Multiresolution Analysis (MRA) Tool 26
2.16.1 Wavelet Expansion 27
2.16.1.1 Characteristics of a Wavelet System 28
2.16.2 Scaling Function 28
2.16.3 Multiresolution Analysis 30
2.16.4 Wavelet Function 31
2.16.5 Discrete Wavelet Transform 32
2.16.6 Mallat's Decomposition and Reconstruction Algorithms 33
2.17 Review on the Multiscale Methodology as Applied to Process Monitoring and Control 34
2.18 Examples of Multiple Input Multiple Output (MIMO) Systems 34
2.18.1 Chemical Mechanical Planarization (CMP) 35
2.18.2 Silicon Epitaxy Process 35

CHAPTER 3 PROBLEM DEFINITION 36
3.1 Introduction 36
3.2 Need for a Double EWMA Controller 36
3.3 Control Procedure for a MIMO System 37
3.3.1 Process Model 37
3.3.2 Online Model Tuning and Process Control 38
3.4 Control Methodology 38
3.4.1 Ridge-Solution Controller 41
3.4.2 Right-Inverse Controller 41
3.5 Problem Definition 41

CHAPTER 4 PROPOSED CONTROLLER 43
4.1 Introduction 43
4.2 Wavelet Based Run-by-Run Controller 43
4.2.1 Choice of Wavelet 45
4.2.2 Determining the Number of Levels for Decomposition 45
4.2.3 Wavelet Decomposition 45
4.2.4 Denoising by Thresholding 46
4.2.5 Wavelet Reconstruction 47
4.2.6 Calculation of the Recipe/Control Action 47

CHAPTER 5 EXPERIMENTAL SETUP AND RESULTS 48
5.1 Test Model 48
5.2 Initial Values 49
5.3 Cases Considered 49
5.4 Results 50
5.4.1 Results for Case1 50
5.4.2 Results for Case2 57
5.5 Other Research Efforts 60

CHAPTER 6 CONCLUSIONS AND FUTURE RESEARCH 61
6.1 Conclusions 61
6.2 Future Research 61

REFERENCES 63

LIST OF TABLES

Table 5.1 Average Absolute Input Deviation ($\times 10^{-2}$) 54

Table 5.2 Absolute Square Deviation Form Target ($\times 10^{3}$) 57

Table 5.3 Standard Deviation of the Process Outputs 57

Table 5.4 Average Absolute Input Deviation ($\times 10^{-2}$) 58

Table 5.5 Absolute Square Deviation Form Target (ASD) ($\times 10^{4}$) 59

Table 5.6 Standard Deviation of the Process Outputs 60

LIST OF FIGURES

Figure 1.1	Basic Structure of a Run-by-Run Controller (Del Castillo and Hurwitz, 1997)	3
Figure 1.2	A Process Signal and it's Time Frequency Representation (Bakshi (1999))	7
Figure 1.3	Tiling of the Time Frequency Space (Bakshi, 1999)	8
Figure 2.1	Shewhart Filter	17
Figure 2.2	Moving Average Filter	18
Figure 2.3	CUSUM Filter	19
Figure 2.4	EWMA Filter, with Smoothing Constant -0.4	20
Figure 2.5	Dilation of Sin(x) and Cos(x)	24
Figure 2.6	Haar Translation Example	24
Figure 2.7	Multiresolution Analysis Example	25
Figure 3.1	A Run-by-Run Controller for MIMO Systems	39
Figure 4.1	A Wavelet Based Multiscale Run-by-Run Controller for MIMO Systems	44
Figure 5.1	Process Outputs	51
Figure 5.2	Prediction Error for Offset for Process Output 1	52
Figure 5.3	Prediction Error for Offset for Process Output 2	52
Figure 5.4	Prediction Error for Drift for Process Output 1	53
Figure 5.5	Prediction Error for Drift for Process Output 2	53
Figure 5.6	Input 1	54
Figure 5.7	Input 2	55
Figure 5.8	Input 3	55

Figure 5.9 Input 4 56

Figure 5.10 Average Absolute Input Deviation ($\times 10^{-2}$) 56

Figure 5.11 Process Outputs Under Mean Shift 58

Figure 5.12 Average Absolute Input Deviation ($\times 10^{-2}$) 59

A WAVELET BASED MULTISCALE RUN-BY-RUN CONTROLLER FOR MULTIPLE INPUT MULTIPLE OUTPUT (MIMO) PROCESSES

Santosh Kothamasu

ABSTRACT

Run-by-Run (RbR) control is an online supervisory control strategy designed for the batch manufacturing industry. The objective of RbR control is to minimize process drift, shift and variability between machine runs, thereby reducing costs. The most widely used RbR controllers use the Exponentially Weighted Moving Average (EWMA) filter. However, the linear nature of the EWMA filter makes these RbR controllers inefficient for processes with features at multiple frequencies (also known as multiscale processes). Recent developments in wavelet theory have enhanced the ability to analyze events in multiscale processes. New RbR control strategies have started to emerge that incorporate wavelet analysis. These controllers, developed at the University of South Florida, seem to be robust in dealing with multiscale processes. The objective of this research is to integrate the wavelet based, multiscale analysis approach with the existing double EWMA RbR control strategy for controlling a multiple input multiple output (MIMO) process. The new controller (WRbR controller) is applied on a Chemical Mechanical Planerization process having four inputs and two outputs. A continuous drift and mean shift are introduced in the process, which is then controlled using both the existing double EWMA and the new wavelet based RbR controllers. The results indicate that the wavelet based controller is better in terms of the average square deviation and the standard deviation in the

process outputs. Moreover, the observed decrease in the magnitude of the average absolute input deviation indicates a smoother process operation.

CHAPTER 1

INTRODUCTION

Run-by-Run control is a combination of statistical process control (SPC), Design of Experiments (DOE) and Engineering process control (EPC) techniques that is suitable for controlling batch processes. A run-by-run process control system assumes that automatic controllers like PID controllers, which control the process *in situ* or during a run, cannot be kept at the same set-points from one run to the other because of the various disturbances affecting the process, like deterministic trend and shifts in the mean. A run-by-run process control system primarily consists of two steps - process modeling followed by online model tuning and control. Process modeling is done offline using response surface methods, design of experiments, or ordinary least square estimation. The model, which is continuously updated or tuned based on the observed process data, is also used to determine the control action for the next run.

This thesis concentrates on the online model tuning and control for processes with shift and drift disturbances. One of the most commonly used model-tuning strategy for a process with the above characteristics is the Double Exponentially Weighted Moving Average (EWMA) filter. However, the double EWMA filter is inefficient as it operates at a fixed scale (or frequency). Proposed in this research, is a multiscale adaptation of the double EWMA filter for multiple input, multiple output (MIMO) processes.

1.1 Run-by-Run (RbR) Control

The semiconductor manufacturing industry is among the fastest growing industries that require state of the art tools and techniques for process monitoring and control. One such process control strategy is Run-by-Run control, which is popular as an online, efficient, automatic and low cost control strategy in the semiconductor manufacturing industry.

Many of the semiconductor manufacturing processes operate in batch mode, e.g., wafer-fabrication. A "run" is described as an event of processing a batch (or lot) of silicon wafers. The batch size can be as small as one. Run-by-Run control refers to a situation where the control action is taken at the end of each run. For example, consider a metal sputter deposition process. Metal is deposited on a wafer by the sputtering process [1], which is carried out for a batch of wafers. The goal of the RbR control procedure for such a process is to maintain a desired deposition thickness for a batch of wafers.

Run-by-Run control techniques are well suited for processes where the cost of a run operating far from the target is very high, but the cost of the control action is relatively low. This is typical for most of the semiconductor manufacturing processes.

1.2 The Basic Structure for a Run-By Run Controller

The basic structure of the run-by-run controller is shown in Figure 1.1. There are two feedback or control loops. The outer control loop is the Run-by-Run controller, and it acts as the supervisor for the inner control loop where automatic PID controllers are used. The objective of the PID controller is to maintain constant input settings (or "setpoints") during a run. The run-by-run controller action (or recipe) consists of varying these setpoints of the automatic (PID) controllers after every run in a supervisory manner. The setpoints, in turn, regulate the physical pro-

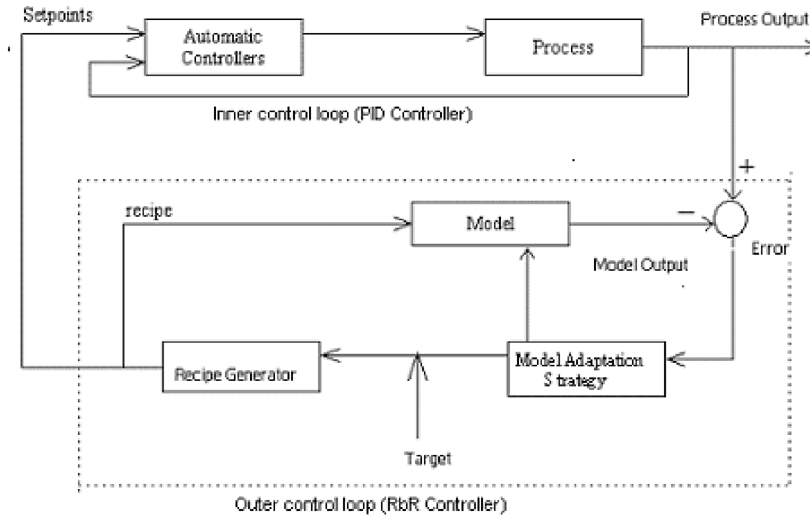

Figure 1.1. Basic Structure of a Run-by-Run Controller (Del Castillo and Hurwitz, 1997)

cess variables (or inputs) within each run, which affect the quality characteristics (or outputs). The difference between the functions of the automatic controller and the Run-by-Run controller is that the automatic controller controls the process *within* a run, while the Run-by-Run controller controls the process *between* successive runs. The control action for a run is based on the results of the most recent runs.

1.3 Need for Run-by-Run Control

For any process, the value of the controllable variables, or inputs, are preset by the machine manufacturers. This recipe remains fixed during a run. Different recipe settings can be desired to fabricate different products, or similar types of products with different quality characteristics, leading to change in specifications from run to run. Moreover, the equipment may suffer from ageing and tool wearing effects that may induce trend and autocorrelation disturbances. Likewise, a maintenance operation may cause process shifts. Hence the setpoints of the automatic controllers cannot be kept at the same value from one run to the next. This calls for a supervisory control that alters the values of the process parameters on a run-to-run basis.

1.4 Control Procedure

A run-by-run control procedure consists of two major steps. Firstly, a regression model is developed that relates the input variables with the output variables. This model is usually fitted offline with some experimental data using response surface of least square estimation methods.

The second step involves online estimation and process control. The model developed offline is continuously tuned based on the observed process data to compensate for the disturbances and also used to compute the control action for the subsequent run. This control action is known as a recipe in the microelectronics literature [2].

1.5 Filters or Model Tuning Strategies

Measured data is generally contaminated with errors due to factors such as sensor noise, disturbances, instrument degradation and human errors [3]. Since the decision about the process performance is based on the quality of information extracted from the measured data, the collected data must be cleaned or filtered for efficient process operation. Filters are generally used for data rectification. Some of the popular filters in the industry are the Shewhart, Exponentially Weighted Moving Average (EWMA), CUSUM (Cumulative Sum), and MA (Moving Average) filters. They are called linear filters as they rectify the data by taking the linear combination of the measured data.

Run-by-Run control, in most of the cases, is a model based online control strategy. At the end of each run, the difference between the actual measured output and its target (or predicted) value is computed. This difference is called the prediction error. This error is processed by filters and consequently the model parameters are updated from run to run. Hence, the filters are also referred to as *model tuning strategies.*

1.6 A Basic SISO Run-by-Run (RbR) Control Methodology

The preliminary RbR control scheme can be traced back to the work by Sachs *et al.* [2] and Ingolfsson and Sachs [4]. They presented a simple Exponentially weighted moving average (EWMA) statistic based Run-by-Run controller that recommends recipe adjustments for a silicon expitaxial growth process in a barrel reactor. Del Castillo and Hurwitz[5] discuss a similar controller for a Single Input Single Output (SISO) process. The true process model is assumed to be a first order linear model of the form,

$$y_t = \alpha + \beta u_{t-1} + \epsilon_t \tag{1.1}$$

where, t is the time index denoting the run number, ϵ_t is a white noise sequence with, $\epsilon_t \sim (0,\sigma^2)$ representing measurement error and other unpredictable noise sources, the parameters α and β are the intercept (or offset) and the process sensitivity (or gain), respectively, y_t is the process output at the end of the run t and u_{t-1} is the recipe or the process input at run $t-1$, i.e., recipe at the start of run t or the recipe at the end of run $t-1$. It is assumed that there is no time lag between the control action and its effect on the process output. Also, the process exhibits no dynamics, i.e., the output y_t, is dependent only on the input value at the start of run t (or at the end of the run $(t-1)$), u_{t-1}. The optimal control action for a *minimal variance* around target value T is given by :

$$u_t = \frac{T - \alpha}{\beta} \tag{1.2}$$

However, most industrial processes are prone to errors, which induce dynamics into the process. To compensate for these dynamics, the intercept α, is assumed to be time varying. The predicted output response for run t, is of the form,

$$\hat{y}_t = a_{t-1} + b u_{t-1} \tag{1.3}$$

where, a_{t-1} and b are estimates of the parameters, α and β respectively, obtained prior to run t with the available information. The recipe u_t, is chosen such that $a_{t-1} + bu_{t-1}$ =T, where T is the target for the response, y. Ingolfsson and Sachs [4] also assume that the estimate b of the process gain ,β, is available from some off-line designed experiment, and that this value remains constant over time. The intercept estimate a_t is updated after each run using the current output measurement, y_t. This is done by the EWMA equation,

$$a_t = \lambda(y_t - bu_t) + (1 - \lambda)a_{t-1} \tag{1.4}$$

where, λ is the smoothening constant of the EWMA filter. The optimal control action with a *minimum variance* around a desired target T is given by :

$$u_t = \frac{T - a_t}{b} \tag{1.5}$$

This is the control action for the next run, $t + 1$.

1.7 Multiple Input Multiple Output (MIMO) Systems

Most of the manufacturing processes are multivariate in nature. This is particularly true for the semiconductor manufacturing industry where the processes are of the Multi input Multi Output (MIMO) type. Consider, for example, the wafer fabrication processes like etching, sputtering and polishing (CMP). In a CMP (Chemical Planerization) process, the inputs are the table speed, back pressure, polishing downforce and the profile of the conditioning system. The characteristics of interest (or outputs) are the material removal rate and within wafer non-uniformity. The application of Run-by-Run control for a MIMO process is relatively new. The control structure and the methodology are discussed in detail in the Sections 3.2 and 3.3.

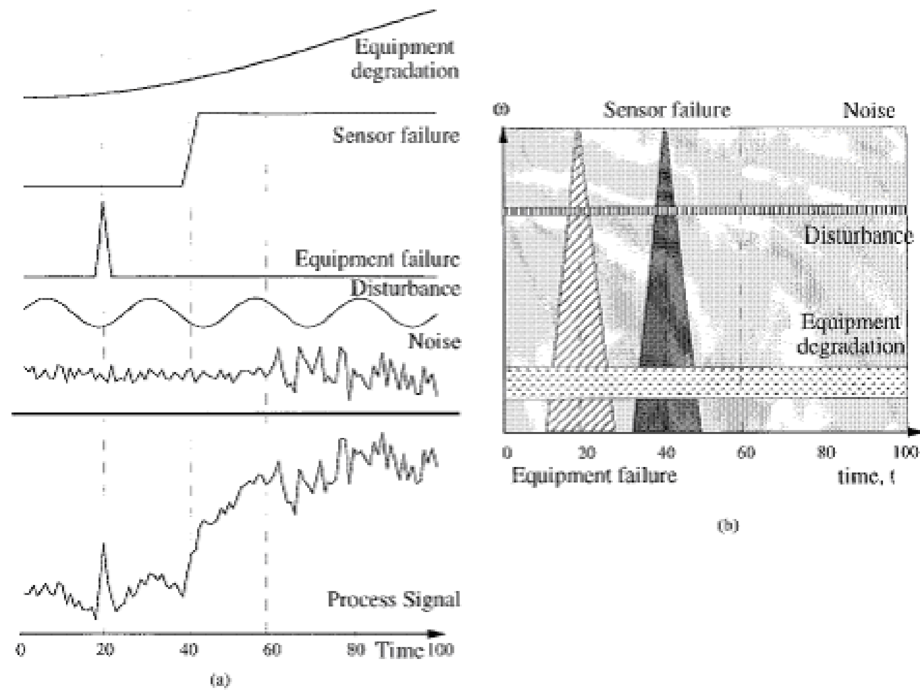

Figure 1.2. A Process Signal and it's Time Frequency Representation (Bakshi (1999))

1.8 Need for Multiscale Analysis

EWMA filter is a single scale linear filter. The linear filters, though easy to work with and suitable for online analysis, have some limitations. They analyze the data at a single scale, i.e. at a fixed resolution in time and frequency. For example, Shewhart filters analyze the measurements at the sampling interval or the finest scale, similar to the delta functions in Figure 1.3(a), and are best for detecting large, localized changes. In contrast, the MA, EWMA and CUSUM charts inherently filter the measurements at a coarser scale and hence can detect small shifts more efficiently.

In reality, measured data in most processes is multiscale in nature due to contributions from events occurring at different locations in time and frequency. As can be seen from Figure 1.2, a process signal contains contributions from a variety of sources such as, sensor failure, equipment failure, equipment degradation, disturbance and inherent noise in the process. These features have different time-frequency localiza-

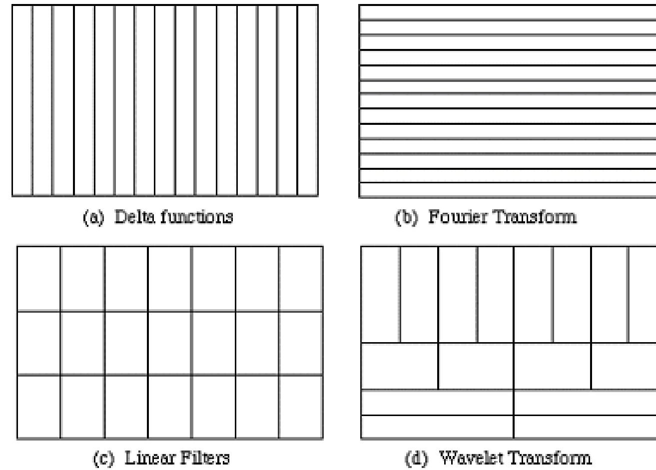

Figure 1.3. Tiling of the Time Frequency Space (Bakshi, 1999)

tion as shown in the adjoining figure. To control such a process, it is imperative to examine these features over the entire time frequency domain. For example, a step change within a process is localized in time domain while it is distributed in the frequency domain. Similarly, a variance change in the process is localized in the frequency domain but distributed in the time domain.

If the signal contains such data at multiple scales, the EWMA filter has the tendency to trade-off the extent of error removal with the quality of the retained local features. Also, the EWMA filter is not robust against a higher autocorrelation within the data. Tiling of the time frequency space as shown in Figure 1.3(c), indicates that the analysis of the signal at single scale will not permit efficient feature extraction or noise removal from typical process signal.

As can be seen in Figure 1.3(b), the Fourier transform would provide only the frequency related information within the signal; in other words, the data is localized in frequency. Noise removal by eliminating the high frequency contribution by Fourier transform will distort the localized features by over smoothing, as their high frequency components will also be removed.

1.9 Multiscale Analysis Using Wavelets

Recently, wavelets have been popular as a multiscale analysis tool. Wavelets are a family of basis functions with different time-frequency localization (or scale). As can be seen from Figure 1.3(d), the wavelet functions span the entire time frequency space. Most of the signals can be represented at multiple scales by decomposition using a set of wavelet and scaling functions. This property is called as Multiresolution Analysis (MRA). Wavelets are capable of adjusting their scale to the nature of the signal and thus provide complete information about a signal at different localizations in time and frequency by adjusting its scale to the nature of the signal. Wavelet transform is computationally efficient than the traditional Fourier transform. Wavelet analysis can often compress or *de-noise* a signal without appreciable degradation. Moreover, compared to the Fourier transform, wavelets have the additional property of compressing multiscale features and approximately decorrelating many stochastic processes. All the above properties of wavelets are suitable for online, multiscale analysis.

1.10 A Brief Description of the Problem

Most of the run-by-run control strategies are model based. Castillo and Rajagopal[6] discussed a run-by-run strategy based process control for a Multiple Input Multiple Output (MIMO) system. The model tuning strategy for such a system is based on a double EWMA filter. This research is motivated by the developments in the multiscale wavelet based analysis and the limitations posed by the single scale filters. The aim of this thesis is to integrate the wavelet based methodology with the existing double EWMA controller for online analysis. The integrated controller (called the WRBR controller) is tested for its performance in the presence of a continuous drift and mean shift.

1.11 Thesis Organization

Chapter 2 provides a literature review about RbR control and wavelet based multiscale methodology. Chapter 3 gives a detailed problem description. Chapter 4 discusses the proposed controller. Chapter 5 presents the experimental setup and results.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Any industrial process, irrespective of how perfectly it is designed, is prone to disturbances resulting in performance variability. Such inconsistency may be due to factors such as variability in raw material properties, tool wear, equipment ageing or change in the ambient conditions. These factors induce disturbances in the form of mean shifts, variance shifts, autocorrelation and trend in the process output. SPC and EPC are two strategies for process control. These control techniques evolved independently in two different types of industries. While SPC originated in the discrete parts industry, EPC is rooted in the continuous process industry. Box and Kramer [7], provided a detailed comparison of SPC and EPC. They referred to SPC as statistical process monitoring and to EPC as automatic process control (APC).

The control objectives in these industries were often different. In the parts industry, the objective was to consistently reproduce parts with the *smallest possible variation about the fixed target value.* The process industries were attempting to obtain the *highest possible mean value for the process* such as a purification process, with the smallest variation.[7]

2.2 Statistical Process Control (SPC)

SPC comprises of monitoring a process through *statistical analysis.* It may also be defined as a collection of statistical tools for process control. SPC looks for signals representing assignable causes, which may be thought of, as external disturbances

that increase variability. At the same time, it ignores the chance causes of variation that are inherent to the process. Thus, SPC does not control a process, but rather performs a monitoring action that signals when control, in the from of identification and removal of root causes, is needed [8]. SPC assumes that adjusting a process is very expensive activity and hence should be utilized only when there is strong evidence that the process is affected by an external source of variation. SPC is "top-down" tool, driven by upper management as part of a company-wide quality improvement drive. Some of the tools for achieving SPC are histogram, Pareto chart and control chart. The control charts are the most powerful SPC tools. Shewhart and EWMA control charts are popularly used in industry.

2.3 Engineering Process Control (EPC) or Automatic Process Control (APC)

EPC is also referred to as Automatic Process Control (APC) in the literature. It is an adaptive control policy commonly followed in process industries like chemical industries. EPC actively reverses the effect of process disturbances by making regular adjustments to the manipulatable process variables. In other words, EPC achieves target output by transferring variability in the output variable to an input control variable. [8]

Thus, EPC does not remove the root or assignable causes: it uses continuous adjustments to keep the process variables on target. EPC is a "bottom-up" procedure, driven by process control or manufacturing engineers. EPC assumes that the cost of controlling a process is far less as compared to the cost of being off-target.Feed forward and feed back controls are some commonly used EPC strategies.

2.4 Integrating SPC and EPC

In the last decade, there were prominent efforts to integrate process monitoring and engineering control strategies. Such integration is useful because an integrated system can use EPC to reduce the effect of predictable quality variations and SPC to monitor the process for detection of assignable causes of variation. The elimination of these assignable causes results in additional reduction of the overall variability. Improved quality would be the most likely outcome. The problem of integrating SPC and EPC for process control has been studied by Box and Kramer [7], Van der Weil *et al.* [9], Truker *et al* [10], Montgomery *et al.* [8] and Janakiram and Keats [11].

The semiconductor manufacturing industry is a hybrid industry in the sense that, it uses some of the attributes of that parts and process industries. For example, in the manufacturing of computer chips, certain aspects of the parts industry and others of the chemical industry are needed. Consequently, a control strategy, which could integrate SPC and EPC strategies, was needed. Run-by-Run control technique is a result of such efforts to integrate SPC and EPC.

2.5 Evolution of Run-by-Run Process Control

Sachs *et al.* [2], and Ingolfsson and Sachs [4] proposed a model based process control technique, which they termed as Run-to-Run process control. They describe a controller, which uses feedback control to regulate a process, while at the same time the diagnostic capability of a generalized version of SPC is used to detect major disturbances to the process. They studied the application of the EWMA statistic for process adjustment. Butler and Stefani [12] discussed a Double EWMA controller to take care of a deterministic trend within the process. Del Castillo [5] presented a literature review of the run-by-run process and also suggested some extensions. Pioneering work has been undertaken by Del Castillo [13] and Del Castillo *et al.*

[14] wherein conditions for stability of the run-by-run controllers were discussed. SEMATECH International, a semiconductor research corporation, was continually involved in the development and implementation of Run-by-Run control. The above studies were restricted to Single Input Single Output (SISO) systems. Del Castillo [6], Rajagopal [15] developed a controller for Multiple Input Multiple Output systems (MIMO) as applied to a CMP device. Tseng it et al. [16] have also presented a Double EWMA based controller for MIMO systems.

2.6 Run-by-Run Control Strategies

The EWMA based strategy has been widely used as a model tuning strategy in Run-by-Run control because of its simplicity and ability to compensate for drifts and other small disturbances.

Some other strategies for Run-by-Run control have been discussed in literature. The strategies can be classified on the basis of the assumption of the process models being linear or nonlinear. Recursive Least Square Estimation Method (RLSE) has been used to update the process model. A description of the recursive estimation methods can be found in [17], [18] and [19]. Del Castillo and Yeh [20] have discussed the application of this strategy for the Optimizing Adaptive Quality Controller (OAQC) and self-tuning controllers [5]. The application of the OAQC strategy for non-linear processes has been studied by Del Castillo and Yeh [20].

The Kalman filtering approach [21] uses a linear model to describe a process. It is different from EWMA filter in the sense that it can update both the gain and the offset parameters. Zhang *et al.* [22] have discussed about a Set valued R2R strategy with ellipsoidal approximation.

Some methods for Run-by-Run control for non-linear process models are the machine learning approach [23] and the artificial neural network based approach [24].

2.7 Applications of Run-by-Run Control

Run-by-Run control has been successfully implemented in many of the semiconductor manufacturing processes. Ingolfsson and Sachs [4] have discussed the application of the run-by-run controller to the control of a silicon epitaxy process in a barrel reactor. Butler and Stefani [12] discuss a supervisory run-to-run control of polysilicon gate etch process. The Double EWMA model tuning strategy was utilized for tuning the process model. Mozumder and Barna [25] have applied the run-by-run strategy for the control of the PECVD (Plasma Enhanced Chemical Vapor Deposition) process.

The application of Run-by-run control to a Chemical Mechanical Planarization (CMP) process has been widely discussed in literature. Boning *et al.* [26] have presented the application of an integrated hardware software control system utilizing run-by-run methods to overcome common CMP equipment difficulties. Del Castillo and Rajagopal [6] also have successfully tested the EWMA and PCC based run-by-run controllers for the CMP process. [27] have applied the run-by-run control methodology for the CMP process. They have designed experiments to determine the transition period necessary to reacquire equilibrium after a recipe change. Work has been done in applying run-by-run controllers to the process of Metal Sputter Deposition by Smith et al. [1]. They have applied the EWMA and the Predictor Corrector controller (PCC) for process control. Palmer *et al.* [21] have applied the Kalman filtering approach in the photolithography step of the semiconductor manufacturing process.

2.8 Filters

Control charts are popular in the industry to monitor changes in the process. Some of the commonly used control charts are Shewhart, CUSUM, EWMA and MA

charts. Shewhart (or $\overline{X}$) charts are suited to detect large shifts in the process mean whereas EWMA, CUSUM and MA are better in detecting small shifts in the process parameters. Hence the selection of the type of the control chart in an important issue for a process engineer. The control charts, though effective, have a limitation. They assume that the observations are statistically independent. In reality, in process and hybrid industries like semiconductor manufacturing industry, this assumption can lead to wrong conclusion.

The term 'filter' is analogous to the digital filters referred to in signal processing. Filters are a set of weight functions, which on convolution with the data extract the information contained in a specific frequency range. The process of filtering is also referred to by the term rectification. The filters are broadly classified as Linear and Non-linear filters based on the method of transforming the data.

2.8.1 Linear Filtering

The linear filters take a linear combination of the weights and the measured data as:

$$X_i = \sum_{i=1}^{N} w_i x_i \tag{2.1}$$

Where w_i are the filter weights, which satisfy

$$\sum_i w_i = 1 \tag{2.2}$$

x_i is the measured data and X_i is the filtered data.

The filter weights are called impulse response of the filter or filter coefficients. If the filter window is of finite length, then these filters are called as finite impulse response (FIR) filters. If the filters combine all measurements, they are called infinite impulse response (IIR) filters. The linear filtering methods have found wide use in statistical process control, since after proper tuning, they permit easier identification

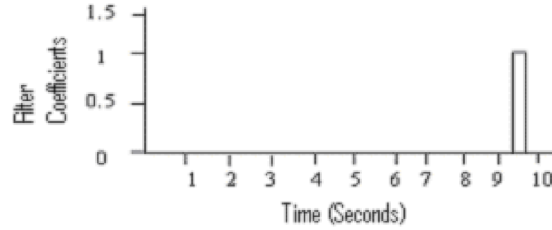

Figure 2.1. Shewhart Filter

of a change in the process mean. Some of the commonly used linear filters namely the Shewhart, CUSUM, EWMA, and MA filter.

2.8.2 Shewhart Filter

A Shewhart chart, is a FIR filter with N=1 and w=1, which can be given by y_i=1 $* x_i$. Thus, a Shewhart chart uses only the last observed data point, ignoring any information given by the entire sequence of points, i.e., the filter represents the data at the finest scale (smallest window) or highest resolution of measurement. This feature makes the Shewhart chart relatively insensitive to small shifts in the process. Thus the quality characteristic obtained by Shewhart filter is sensitive to large shifts. As can be seen form the Figure 2.1, the only filter coefficient for has a fixed value of 1.

2.8.3 Moving Average (MA) or Mean Filter

a process is monitored using the Moving Average control chart, the filtered data or the quality characteristic of interest is the average of the measurements in the window. The procedure consists of calculating the new moving average Y_i as each observation x_i becomes available. The filtered data is given by:

$$Y_i = \frac{X_i + X_{i-1} + X_{i-2} + \ldots + X_{i-w+1}}{w} \tag{2.3}$$

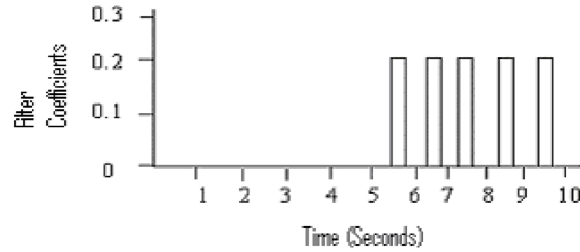

Figure 2.2. Moving Average Filter

The MA filter is a FIR filter with equal weights (as shown in Figure 2.2), and is more effective in detecting small process shifts. However, it is generally not as effective against small shifts as the EWMA or the CUSUM filters, which are described next.

2.8.4 CUSUM Filter

A CUSUM filter (Figure 2.3) is given by the following recursive equation :

$$Y_i = w_i X_i + Y_{i-1} \tag{2.4}$$

where $w_1 = w_{i-1} = \ldots = w_1$. The length of the filter window increases as the window of the measured data increases and hence it is an infinite impulse response (IIR) filter. The CUSUM filter represents data at the coarsest scale (lowest resolution). The CUSUM chart is very effective in detecting small shifts, but not as effective as the Shewhart filters in detecting large shifts. An approach to improve the ability of the CUSUM filters to detect large process shifts is to use a combined CUSUM-Shewhart procedure for online control.

2.8.5 EWMA Filter

The Exponentially Weighted Moving Average filter is a good alternative to the Shewhart filters when we are interested in detecting small shifts. The performance of the EWMA is approximately equivalent to that of the CUSUM filters but it is

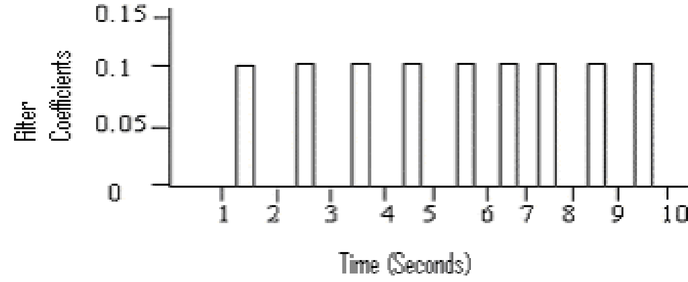

Figure 2.3. CUSUM Filter

easier to set up and operate. Hence, the EWMA filter is one of the most widely used model adaptation strategies in run-by-run control. An EWMA filter is given by the following recursive equation:

$$Y_i = \lambda X_i + (1-\lambda)Y_{i-1} \tag{2.5}$$

where λ is the filter parameter. The above equation can be expanded as :

$$Y_i = (1-\lambda)^i Y_0 + \lambda(1-\lambda)^{i-1}X_1 + \ldots + \lambda(1-\lambda)^{i-2}X_2 + \ldots + \lambda X_i \tag{2.6}$$

and $Y_0 = \mu_0$.

The length of the filter window increases with the length of the data; hence it is an infinite impulse response filter. The filtered data is exponentially averaging the current measurement with all the previous measurements. The EWMA thus gives more weight to the current observations and exponentially less to the previous measurements. The EWMA weights decrease geometrically. Since an EWMA filter considers an infinite data window (though with diminishing weights), it represents data at a coarser scale than the MA filter, but finer scale than the CUSUM filter.Figure 2.4 below represents an EWMA filter.

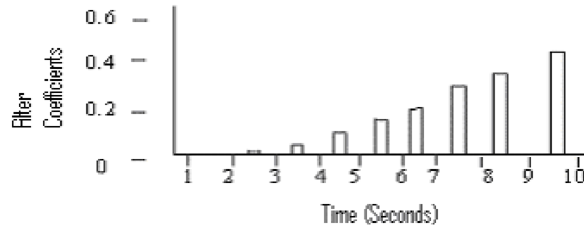

Figure 2.4. EWMA Filter, with Smoothing Constant -0.4

2.9 Disadvantages of Linear Filters

The linear filters discussed above operate at a fixed scale and hence are best for detecting changes at a single scale. For example, Shewhart charts analyze the measurements at the scale of the sampling interval or the finest scale, and hence are best suited for detecting large changes. In contrast, MA, CUSUM and EWMA charts, process measurements at a coarser scale. Hence, they are best suited for detecting small shifts or features at coarse scales. Thus, the use of any one particular filter is likely to fail in extracting an unknown feature of a data set. It means that if a signal contains data at multiple scales, the linear filter will be forced to trade off the extent of error removal with the quality of the retained local features [3]. With the removal of more errors, the retained local features in the filtered signal become smoother and more distorted.

Many SPC methods assume uncorrelated measurements, but in practice, auto-correlated measurements are very common. On using the single sale filters, auto-correlation is still present among the measurements. In other words, the linear filters are unable to de-correlate the data in the process of filtering. Also, the linear filters are also not robust in the presence of non-Gaussian errors such as outliers in the measured data.

Wavelet based analysis is useful for understanding the features at different scales. The multiscale literature is discussed henceforth.

2.10 Literature Review on the Multiscale Methods and their Applications

Wavelets were first defined by Morlet [28]. Daubechies [29], one of the brightest stars in the field of wavelets then developed her family of wavelets with compact support. After their conceptualization, wavelets were applied in a wide variety of applications. Some of the most significant applications have been in the areas of signal processing, image and data compression. Grossman and Morlet [30] used wavelet in seismic analysis. Wavelets have been widely used in the noise removal for a variety of processes. Donoho [31], and Donoho and Johnstone [32] have studied the mathematical background for wavelets.

To understand the concepts behind wavelet analysis, the stepping-stones of frequency analysis, namely the Fourier series and the Fourier transform should be understood. They are explained below.

2.11 Fourier Analysis

The basic concept of frequency-domain analysis is that a signal can be considered as the sum of sinusoidal basis functions. A continuous sine function $\sin(\omega t)$ is a single frequency wave of frequency ω radians/second, and the frequency domain description consists of a single value at a particular frequency. The Fourier analysis is done for periodic and non-periodic signals. For periodic signals, analysis is done by Fourier series and for non-periodic signals analysis is done by Fourier transform.

2.11.1 Fourier Series

A periodic signal $f(t)$ can be decomposed on a set of sine and cosine basis function as follows.

$$f(t) = \sum_{n=-\infty}^{\infty} a_n e^{jn\omega_1 t} \tag{2.7}$$

where $a_n = \frac{1}{T}\int_{t_1}^{t_1+T} f(t)e^{-jn\omega_1 t}dt$.

It can be seen that the Fourier coefficient a_n represents a contribution of the signal on a basis function with frequency $n\omega_1$. Thus the information obtained is based on the complete contribution over a particular basis function, hence this information is localized in frequency.

2.11.2 Fourier Transform

The Fourier transform is applied for non-repetitive signal. It does the same operation as what the Fourier series do. For a signal $f(t)$, the Fourier transform is given as follows

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{2.8}$$

where

$$f(t) = \int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega \tag{2.9}$$

The Fourier transform of a signal indicates the frequency content of the signal and the magnitude of the Fourier transform indicates the energy content of the signal. A filter is designed accordingly with respect to its Fourier transform. For a low-pass filter (which filter out high frequency component) the Fourier transform of the filter should be bounded within a low frequency range. For a high pass filter, the Fourier transform would be bounded within a higher range.

2.12 Definition of Scale (Resolution)

To understand the Multi-resolution or Multi-scale analysis it is necessary to understand the concept of scale. Scale can be described based on the filters used. There are various ways by which we can describe a scale. Whenever a data set is filtered out or the measurements are approximated to obtain a particular output, the description obtained from the output depends on the scale of measurement or the size of the

operator used [33]. Given a data set the scale (resolution) of the output obtained can be described by the amount of filtering done to data set to obtain the output. The original dataset represents the highest resolution of measurement. As the frequency components are filtered out from the measurements, it results in a smoother version of the input. This representation of the data is coarser, i.e., at lower scale. One can obtain various levels of scale by applying different type of filters to the data set.

2.13 Dilation and Translation

This section is introduced to have a better understanding of the concept of wavelets. The dilation of a function can be described by the Fourier representation of the signal. Consider the Fourier approximation of a function $f(x)$ as

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{j}(a_j cos(jx) + b_j sin(jx)) \tag{2.10}$$

The function is approximated by a set of sine and cosine basis functions for various J. It can be seen from Figure 2.5 that sine and cosine at various j or levels represent the dilated versions of the sine and cosine basis functions. It can be seen that as the dilation level increases, the function frequency increases. Translation of basis functions is explained by Figure 2.6, which shows the same basis functions translated at various integer values.

2.14 A Multiresolution Analysis Example

Consider the signal shown in Figure 2.7. It consists of signals at various resolutions (scales). The complete set of the Figures 2.7 (a) through (g) represents a multiresolution representation of the signal. The original signal consists of oscillations at various time intervals. The multiresolution description of the signal can be obtained at various scales by filtering out the highest frequencies present at every

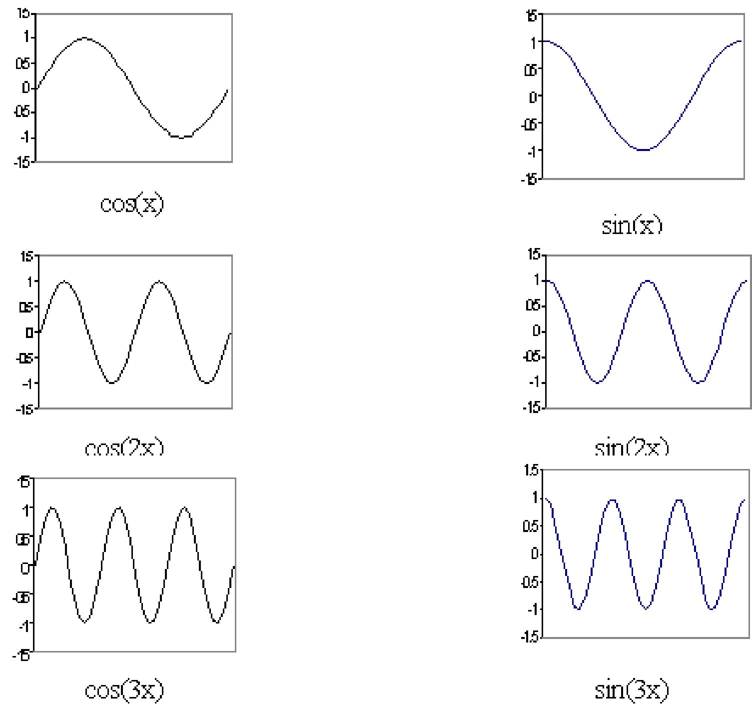

Figure 2.5. Dilation of Sin(x) and Cos(x)

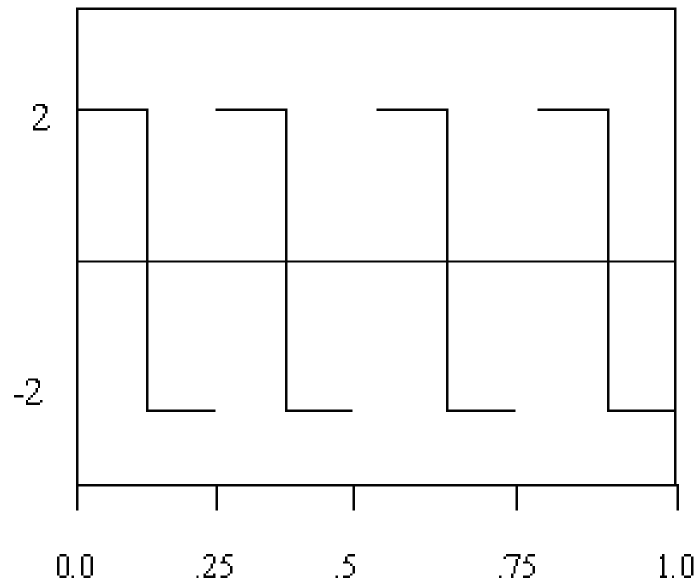

Figure 2.6. Haar Translation Example

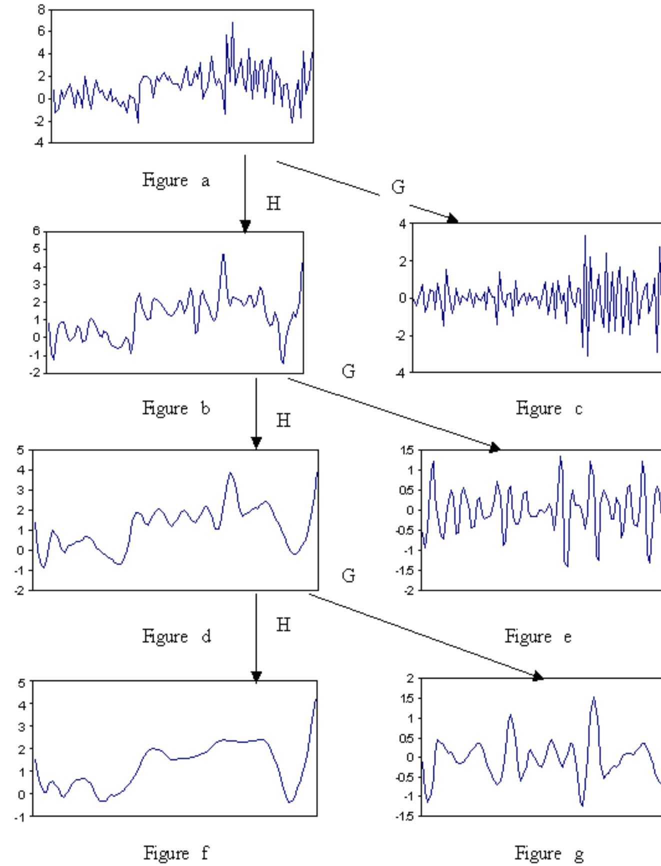

Figure 2.7. Multiresolution Analysis Example

scale. The high frequency components are removed by approximating the signal on a set of basis functions (piece wise constant function) for that resolution. In a similar way, the difference (the high frequency components) between the first approximation and the original signal (the information filtered out) is shown in Figure 2.7(c). To get the signal in Figure 2.7(d) the first approximation is again approximated by a set of basis function (piece wise constant functions). The above function is repeated until the desired level j. Thus we obtain the approximation at every level and also information about the frequency components filtered out. This description is called as the multiresolution description of the signal. The signal from Figure 2.7 (b,d,f) are called as scaled signals and that from Figure 2.7 (c,e,g) are called detail signals. Wavelet basis functions are introduced in the next section.

2.15 Wavelets

Wavelets are localized waves. Instead of oscillating forever, they drop to zero. A basic requirement for a function to be wavelet is that it should integrate to zero. In the Fourier analysis, functions are represented a linear combinations of sine and cosine waves, and hence the representations are localized in frequency, not in time. On the contrary, the wavelet analysis uses linear combinations of wavelet basis functions, localized in both time and frequency, to represent any function in the square integral space i.e. $L_2(R)$ space. Some of the commonly used wavelets are the Harr wavelet, Daubachies and Morlet wavelets.

2.16 Wavelets as a Multiresolution Analysis (MRA) Tool

Figure 2.7 shows an example of the mulitresolution analysis. In the figure, each approximation represents a signal at a particular resolution. To obtain multiresolution approximations of the signal and satisfy the properties for multi-resolution analysis, good basis functions are needed. The basis function should have support over finite interval and also satisfy multi-resolution conditions. Wavelets are basis functions, which have compact support, and they satisfy multi-resolution conditions. The wavelets have various useful properties (given below), which make MRA a useful tool in analyzing signals.

1. The wavelet decomposition approach provides a sound mathematical background for multi-scale analysis.

2. The wavelet basis function have compact support, thus any signal can analyzed on a piece wise basis.

3. The basis functions represent finer details of the signals more quickly due to its oscillation property. The coefficients obtained (wavelet coefficient) determines

the significant events in a signal, since they are measure of the amount of fluctuation in the function about a particular point $x = 2^{-j}k$, where the frequency is determined based on the dilation index j.

2.16.1 Wavelet Expansion

Any signal can be analyzed, if expressed as linear decomposition by

$$f(t) = \sum_{l} a_l \psi_l(t) \tag{2.11}$$

where l is an integer index for the finite or infinite sum, a_l is the expansion coefficient, and ψ_l represents the basis functions of t called the expansion set. If the expansion is unique, the set is called basis functions [34] For orthogonal basis function the inner product is given by

$$\langle \psi_l(t), \psi_k(t) \rangle = \int \psi_l(t)\psi_k(t)dt = 0 \qquad k \neq l \tag{2.12}$$

thus the coefficients can be calculated by the inner product

$$a_k = \langle f(t), \psi_k(t) \rangle = \int f(t)\psi_k(t)dt \tag{2.13}$$

For the wavelet expansion, a two-parameter system is constructed by using (3.13), it is given by the following equation

$$f(t) = \sum_{k} \sum_{j} a_{j,k} \psi_{j,k}(t) \tag{2.14}$$

where both j and k are integers : j is called the dilation parameter, k is the translation parameter, and $\psi_{j,k}$ are the wavelet expansion functions. The coefficients $a_{j,k)}$ are

called the discrete wavelet transform (DWT). For an understanding about the basis function, it is necessary to understand the wavelet system.

2.16.1.1 Characteristics of a Wavelet System

A wavelet system is a set of two-dimensional expansion set (basis functions) for a class of functions or signals. From 3.16, we can say that if a wavelet set is given by $\psi_{j,k}$ for $j, k = 1, 2, 3 \ldots$, a linear expansion would be $f(t) = \sum_k \sum_j a_{j,k} \psi_{j,k}(t)$ for some set of coefficients $a_{j,k}$. There are three important characteristics for a wavelet system (refer [35]

The wavelet system is generated from dilation and translation of scaling functions or wavelets. The two dimensional parameterization is achieved from the function $\psi(t)$ by

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \qquad j, k \in Z, \tag{2.15}$$

where Z is the set of integers. The factor $2^{j/2}$ maintains the norm for the wavelet function. All wavelet system satisfies multiresolution condition. That is if a signal can be expressed as a weighted sum of larger set of basis function then the same signal can be expressed as a weighted sum of smaller set of basis functions. The coefficients representing the larger set of basis function are the high-resolution coefficients and smaller set of basis functions is made up of the lower resolution coefficients.

These lower resolution coefficients can be calculated from higher resolution coefficients by a tree-structured algorithm (pyramid tree) called filter bank.

2.16.2 Scaling Function

The scaling function forms the basic step for derivation of wavelet bases. A set of scaling functions in terms of its translates is given by

$$\phi(t) = \phi(t - k), \qquad k \in Z, \qquad \phi \in L^2(R). \tag{2.16}$$

The subspace of L2(R) spanned by the scaling functions is defined as

$$V_0 = span\{\phi_k(t),\ k\ \in\ Z\}, \tag{2.17}$$

which implies that,

$$f(t) = \sum_l a_l \phi_l(t), \qquad \text{for} \qquad f(t)\ \in\ V_0. \tag{2.18}$$

A two dimensional family of scaling functions are generated from the basic scaling functions and translations are given by

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \tag{2.19}$$

where span over k is

$$V_j = span\{\phi_{j,k}(t),\ k\ \in\ Z\}, \tag{2.20}$$

It implies that,

$$f(t) = \sum_l a_l \phi_l(2^j t + k), \qquad \text{for} \qquad f(t)\ \in\ V_j. \tag{2.21}$$

For larger values of j the scaling function has a larger span hence it can represent the signal in more number of steps (dilations and translation), thus the signal which represents the finest scale of measurement would be represented by scaling function which have translation equal to the number of measurements and the scaling function will be dirac function hence the coefficients would be the measurements (samples of the function). For small values of j the scaling function are translated in larger steps thus can represent only coarse information. The importance of how the scaling function can be known after knowing about MRA and the necessary condition for MRA.

2.16.3 Multiresolution Analysis

Multiresolution provides a formal approach for constructing orthonormal basis function. It provides a particular framework for understanding of wavelet bases, and for construction of newer examples [29]. The idea behind multiresolution analysis is to express a function or signal as a limit of successive approximations, each of the approximations indicating a smoother version of the function or signal. These successive approximations correspond to different resolution thus the name multiresolution. The multiresolution analysis of $L^2(R)$ is defined as a sequence of closed subspaces V_j of $L^2(R)$, $j \in Z$, with the following properties. *Completeness in* $L^2(R)$.

$$\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots , \tag{2.22}$$

$$\cap_j V_j = \{0\}, \ \overline{\cup_j V_j} = L^2(R). \tag{2.23}$$

This implies that the space that contains the higher resolution signals would also contain the lower resolution signals.

Scale Invariance.

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}, \tag{2.24}$$

This implies that elements in a space are scaled versions of elements in the next space.

Shift invariance.

$$f(x) \in V_j \Leftrightarrow f(x+k) \in V_j, \qquad k \in Z. \tag{2.25}$$

V_0 *has an orthonormal basis* $\phi(t-k)$, and there exists a scaling function $\phi(t) \in V_0$ such that

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \qquad j, k \in Z, \tag{2.26}$$

which satisfy the property of *Riesz basis* and are linearly independent.

From this it is known that $\phi(t) \in V_0 \in V_1$, and $\phi(2t)$ is a basis for the subspace V_1 or in other words, V_1 is the space spanned by $\phi(2t)$. This implies that $\phi(t)$ can be expressed as weighted sum of shifted $\phi(2t)$ as

$$\phi(t) = \sum_k h(k)\sqrt{2}\phi(2^j t - k) \qquad n \in Z, \tag{2.27}$$

where the coefficients $h(k)$ are a sequence of numbers which are called the scaling function coefficients (scaling filter or scaling vector) and $\sqrt{2}$ maintains the norm. This recursive equation is of primary importance in derivation of the scaling function, the complete basis set. The recursive equation is called as Dilation equation or refinement equation or the two-scale equation. The scaling function has the following property

$$\int_{-\infty}^{\infty} \phi(t)dt = 1 \quad \text{and} \quad \sum_k h(k) = 1. \tag{2.28}$$

2.16.4 Wavelet Function

The resolution of a signal can be described by scaling functions and they describe smoother versions of the signals. But to identify differences between the scaling spaces (V_n) a different set of functions have to be defined. These functions are called wavelet functions (ψ) and they span the differences between the spaces spanned by the scales of the scaling function. The difference spaces are useful in identifying significant changes in the scaling space. Here the space representations are orthogonal complements. The orthogonal complement of V_n in V_{n+1} is given by W_n. If the wavelet spanned space is defined such that

$$V_1 = V_0 \oplus W_0, \tag{2.29}$$

then

$$V_2 = V_0 \oplus W_0 \oplus W_1. \tag{2.30}$$

Thus it can be written that

$$L^2(R) = \bigoplus_{j \in Z} W_j = V_{j_0} \bigoplus_{j \geq j_0} W_j. \tag{2.31}$$

The wavelet space$W_0 \subset V_1$, and thus can be represented as weighted sum of translated $\phi(2t)$ as follows

$$\psi(t) = \sum_k h_l(k)\sqrt{2}\phi(2t-k), \qquad n \in Z, \tag{2.32}$$

where $h_l(k) = (-1)^k h(1-k)$. This function obtained from (wavelet) gives or generates other equations for set of expansions of the form

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \qquad j,k \in Z. \tag{2.33}$$

2.16.5 Discrete Wavelet Transform

According to Multiresolution analysis given by Equation 2.31, any function $f(t) \in L^2(R)$ can be written using Equation 2.19 (scaling expansion) and Equation 2.33 (wavelet expansion) as

$$f(t) = \sum_k c_{j_0}\phi_{j_0,k}(t) + \sum_k \sum_{j=j_0}^{\infty} d_{j,k}\psi_{j,k}(k). \tag{2.34}$$

The coefficients in the wavelet expansion are called the discrete wavelet transform of the function $f(x)$. If the wavelet system is orthogonal then coefficients can be calculated by

$$c_{j_0} = \langle f(t), \phi_{j_0,k}(t)\rangle = \int f(t)\phi_{j_0,k}(t)dt, \tag{2.35}$$

$$d_{j,k} = \langle f(t), \psi_{j,k}(t)\rangle = \int f(t)\psi_{j,k}(t)dt. \tag{2.36}$$

At high resolution, the scaling function is similar to the Dirac function and the inner product just samples the function, in other words the coefficients are just the measurements. This is a very important property about when to apply wavelet transformation to the data set.

2.16.6 Mallat's Decomposition and Reconstruction Algorithms

In most of the application we cannot deal with the scaling functions and wavelet function, only the coefficients of dilation equation and the coefficients in the expansions are needed. Thus given a multiresolution analysis, a pyramidal algorithm can be employed to calculate the DWT and the various approximations at every space.

Mallat [36] developed a very efficient algorithm (Pyramidal Algorithm) for multiresolution analysis using wavelets. This algorithm depends on two sequences, called filters. Using this algorithm the DWT can be performed. For any wavelet family the two equations needed to derive the entire family are

$$\phi(t) = \sum_k h_k \sqrt{2}\phi(2t-k), \qquad n \in Z, \tag{2.37}$$

$$\psi(t) = \sum_k g_k \sqrt{2}\phi(2t-k), \qquad n \in Z, \tag{2.38}$$

The sequences h_k,g_k, are used to get the DWT, they are called as low pass and high pass filters. Let

$$X_{n,k} = [X_{n,0}, X_{n,1}, \ldots, X_{n,m-1}] \tag{2.39}$$

be a signal of length 2^n. We know that this signal defines a function $f \in v_n$ given by

$$f = \sum_{k=0}^{m-1} X_n(k)\phi_{n,k} \tag{2.40}$$

The filters process the signal X by two operators H and H_1. When these two operators are applied to the signal X two new signals (subband signals) are obtained

and are denoted by HX and H_1X. The entries in the two new signals are denoted by

$$(HX) = X_{j-1}(k) = \sum_m h(m-2k)X_j(m) \tag{2.41}$$

$$(H_1X) = Y_{j-1}(k) = \sum_m h_1(m-2k)X_j(m) \tag{2.42}$$

Equation2.41 contains the set of scaling coefficients at a particular scale and Equation2.42 contains the set of detail coefficients at a particular scale.

2.17 Review on the Multiscale Methodology as Applied to Process Monitoring and Control

Rajesh *et al.* [37] have discussed a wavelet based multiscle approach for statistical process monitoring. The application of the multiscale methodology for process control is relatively new. Rao [38] presented a wavelet based Run-by-Run controller for a Single Input Single Output System (SISO). The process offset is recursively updated, after each run, using a single Exponentially Weighted Moving Average (EWMA) filter. They compared the performance of the wavelet-based controller against that of the simple EWMA filter. They tested the integrated controller for a variety of disturbances like the mean shift, spike and variance shift under different levels of autocorrelations. The conclusions state that the wavelet based controller and the EWMA controller were similar in performance for lower levels of autocorrelations. Contrastingly, for higher levels of autocorrelation, the wavelet based controller is able to control the process faster.

2.18 Examples of Multiple Input Multiple Output (MIMO) Systems

The examples of a MIMO process are the CMP process and the silicon Epitaxy process. They are briefly discussed in the next section.

2.18.1 Chemical Mechanical Planarization (CMP)

CMP is a process of polishing the surface of a wafer smooth by abrasive action and chemical slurries. The CMP machine has carriers to hold the wafer and to press it on the table pad. During the polishing process, the table pad and the carriers are rotated. The mechanical force and the chemical action of the slurry thus remove the material on the wafer. Two important variables of a CMP process are the removal rate and the within-wafer non-uniformity. The inputs (or recipe) to be controlled are the table speed, back pressure, polishing downforce and the profile of the conditioning system.

The goal of the Run-by-Run control is to adjust these recipes so that the outputs meet the specific targets as close as possible from run to run.

2.18.2 Silicon Epitaxy Process

Sachs *et al.* [4] illustrated a feedback controller using a silicon epitaxy process in a barrel reactor. One important characteristic of interest is the uniformity of the thickness of the wafers within a batch. The uniformity is greatly influenced by two input variables, the balance between the metering valves of the left and the right bellows and the horizontal angle of the injectors.

CHAPTER 3

PROBLEM DEFINITION

3.1 Introduction

Run-by-Run control is a low cost and profitable supervisory control strategy devised for the batch processing industry. The objective of the RbR control strategy is to minimize process drift, shift, and variability between machine runs, thereby minimizing costs. Most of the RbR controllers are model based and single scale in nature. This strategy can be coupled with the recent developments in multiscale analysis, which forms the core of this research. This chapter first discusses the existing double EWMA controller, followed by the problem definition.

3.2 Need for a Double EWMA Controller

The model described in Equation 1.1 does not adequately describe those real time-manufacturing processes, which are subject to various disturbances, some of which are unpredictable. It is observed that many of the processes are subject to the drift disturbance. For example, the polysilicon gate etch process is known to drift considerably as the reactor ages. In a machining process, a drift in the response represents the tool wear. In the epitaxial growth process, deposition of material on the walls of the reactor chamber is a potential cause of process drift. Hence, a more reasonable model should include an additional term representing a 'drift' in the process. However, the drift is not always constant and not always known before. A single EWMA based run-by-run controller tends to be off target for a drifting process [13]. Butler and Stefani [12] suggested a double EWMA controller

(also referred to as the "Predictor Corrector Controller (PCC)") that compensates against different types of stochastic drifts. This is achieved by an additional term in the model, δt, which models a time dependent unidirectional drift disturbance (δ is the drift parameter and t is the time index or run number). Thus the double EWMA based run-by-run controller updates two parameters, namely the process offsets (or shift) and the process drift, from one run to the next.

3.3 Control Procedure for a MIMO System

Some of the MIMO systems are described in section 2.18. The steps involved in an RbR control strategy are: developing a process model followed by online model tuning and process control. This strategy, as presented below, is discussed in detail by Del Castillo and Rajagopal [6].

3.3.1 Process Model

The process model for a multiple input multiple output (MIMO) system, with p outputs and m inputs, is of the form:

$$\mathbf{Y}_t = \alpha + \beta \mathbf{U}_{t-1} + \delta \mathbf{t} + \epsilon_t \qquad (3.1)$$

Here,

$\mathbf{Y}_t = (p \times 1)$ vector of the quality characteristics (outputs or response) measured at the end of run t,

$\alpha = (p \times 1)$ vector of process offsets (intercept) for p responses,

$\beta = (p \times m)$ matrix of process gains,

$\mathbf{U}_{t-1} = (m \times 1)$ vector giving the levels of the controllable variables (or inputs) set at the end of run $t-1$,

$\delta = (p \times p)$ diagonal matrix containing the average drift per run for each of the

responses,

$\mathbf{t}$ = $(p \times 1)$ vector with entries equal to t, the time index (or run number) and,

ϵ_t = multivariate white noise.(i.e. E(ϵ_t)= 0, Var(ϵ_t) = σ^2_ϵ, Cov(ϵ_k,ϵ_{t+k})=0 for k$\neq$0) .

The model 3.1 assumes a Deterministic Trend (DT) disturbance, given by the terms $\delta t + \epsilon_t$. Del Castillo and Rajagopal [6] assume that the estimate $\mathbf{B}$, of the process gain β, is obtained offline, using methods such as Design of Experiments and linear regression. The other parameters to be estimated are the offsets, α, and the drift, δ. This procedure is explained in the subsequent section.

3.3.2 Online Model Tuning and Process Control

Once the process model is optimized, the second step is tuning the model online and hence maintaining the process in control. The model developed offline is continuously tuned based on the observed process data and used to determine an optimal control action.

3.4 Control Methodology

In the process Equation 3.1, it is assumed that the process exhibits no dynamics. However, to be realistic, the intercept α is assumed to by time varying to account for these dynamic process disturbances. Similarly, to compensate for the machine or process drift, the drift parameter δ is also assumed to be time varying. The offsets, α, and the drift, δ, will be estimated online and updated after each run, given by the estimates $\mathbf{A}_t$ and $\mathbf{D}_t$ respectively.

Figure 3.1 shows the schematic representation of a double EWMA RbR Controller. At the end of each run t, the model parameters $\mathbf{A}_t$ and $\mathbf{D}_t$ are updated by the double EWMA strategy. These updated parameters, along with the desired process targets, are then used to compute the recipe (or input settings for the Au-

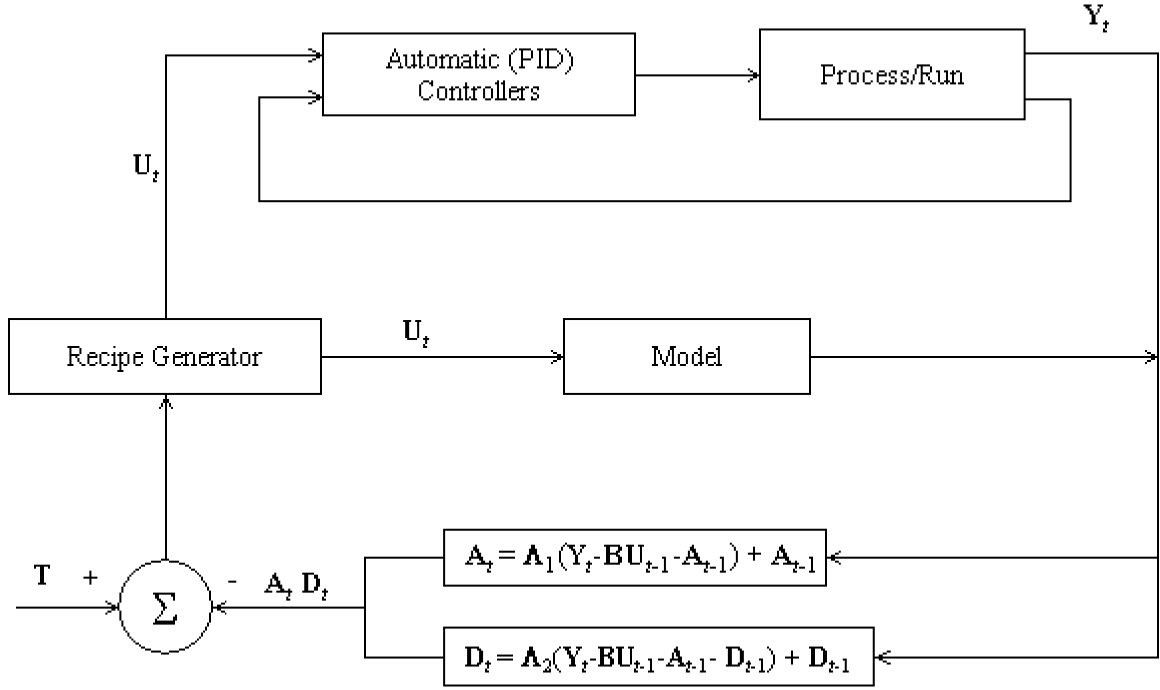

Figure 3.1. A Run-by-Run Controller for MIMO Systems

tomatic PID controllers) for the next run. The model is also updated with the new recipe.

Thus, a corrective action, $\mathbf{U}_t$, is computed at the end of run t, which gives a prescribed set of inputs, to the process engineer for the next run. This is done by choosing the control action that corrects for the one-step-ahead predicted deviation from the target [6]. That is, $\|\hat{\mathbf{Y}}_{t+1|t} - \mathbf{T}\|$ is minimized, where $\mathbf{T} = (p \times 1)$ vector of targets for the responses in $\mathbf{Y}$. Hence the objective function is:

$$\text{Min Z} = \|\hat{\mathbf{Y}}_{t+1|t} - \mathbf{T}\| \tag{3.2}$$

The predicted process output for the run (t+1) is given by :

$$\hat{\mathbf{Y}}_{t+1|t} = \hat{\alpha} + \hat{\delta}(t+1) + \mathbf{B}\mathbf{U}_t \tag{3.3}$$

Replacing α and $\delta(t+1)$ by their estimates, the above equation may be written as:

$$\hat{\mathbf{Y}}_{t+1|t} = \mathbf{A}_t + \mathbf{D}_t + \mathbf{B}\mathbf{U}_t \tag{3.4}$$

For a square system, i.e., with the same number of input and output elements (m = p), the controller equation, as a result of the minimization Equation 3.2, is given by:

$$\mathbf{U}_t = \mathbf{B}^{-1}(\mathbf{T} - \mathbf{A}_t - \mathbf{D}_t) \tag{3.5}$$

The online estimates $\mathbf{A}_t$ and $\mathbf{D}_t$ are obtained using the multivariate double EWMA equations at the end of each run t, for the next run, given by :

$$\mathbf{A}_t = \Lambda_1(\mathbf{Y}_t - \mathbf{B}\mathbf{U}_{t-1}) + (\mathbf{I} - \Lambda_1)\mathbf{A}_{t-1} \tag{3.6}$$

$$\mathbf{D}_t = \Lambda_2(\mathbf{Y}_t - \mathbf{B}\mathbf{U}_{t-1} - \mathbf{A}_{t-1}) + (\mathbf{I} - \Lambda_2)\mathbf{D}_{t-1} \tag{3.7}$$

where, $\mathbf{I}$ is a (p×p) identity matrix, and Λ_1 & Λ_2 are (p × p) diagonal EWMA weight matrices with values (0, 1]. $\Lambda_1 = \lambda_1$I and $\Lambda_2 = \lambda_2$I. λ_1 and λ_2 are the smoothing constants for the above EWMA equations. The conditions for optimal selection of the values λ_1 and λ_2 are discussed by Del Castillo [13]. When Λ_1 and Λ_2 are diagonal, $(\mathbf{A}_t + \mathbf{D}_t)$ provides the unbiased estimate of $(\alpha + \delta(t+1))$, i.e., the estimate provides an asymptotically one step ahead prediction of where the quality characteristics would have drifted in the absence of any control action.

Most industrial processes are not squared. The common case is when the number of controllable inputs is greater than the number of output responses, i.e. $m > p$. Rajgopal and Del Castillo [15] extend the above controller for non-square systems. Two controllers discussed are the Ridge-Solution Controller and the Right-Inverse Controller.

3.4.1 Ridge-Solution Controller

The Ridge Solution Controller equation is similar to Equation 3.5. The only change is that the term $\mu\mathbf{I}$ is added to the term $(\mathbf{B}'\mathbf{B})$, which makes it invertible as long as $\mu \neq 0$. The controller equation is given by:

$$\mathbf{U}_t = (\mathbf{B}'\mathbf{B} + \mu\mathbf{I})^{-1}\mathbf{B}'(\mathbf{T} - \mathbf{A}_t - \mathbf{D}_t) \tag{3.8}$$

3.4.2 Right-Inverse Controller

Tseng *et al.* [16] extended the multiple outputs and single input controller proposed by Sachs *et al.* [4] to the multivariate case. For a double EWMA based model tuning strategy, the controller equation is given by :

$$\mathbf{U}_t = (\mathbf{I} - \mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1} - \mathbf{B})\mathbf{U}_{t-1} + \mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1}(\mathbf{T} - \mathbf{A}_t - \mathbf{D}_t) \tag{3.9}$$

Thus, the controller equation provides the prescribed input settings to the process for run $(t+1)$. The recipe is also sent to update the process model.

3.5 Problem Definition

As discussed in the literature, the EWMA filter is not suitable for multiscale analysis because it's parameters do not adapt to the nature of the features within the signal. Measured data, is most industrial cases, is found to be multiscale in nature due to contributions from the error features located at different scales. There has been a significant development in the area of wavelet based multiscale analysis. Researchers in various fields like image processing and data compression have exploited this ability. Wavelet based strategy, as applied to process control operations, is relatively new. Rao [38] integrated the multiscale wavelet based analysis strategy

with the EWMA controller or a single input single output (SISO) system subjected to Gaussian errors.

Most manufacturing processes are of the multiple input multiple output (MIMO) type. The focus of this research is to develop a wavelet based multiscale double EWMA RbR controller for multiple input multiple output processes.

CHAPTER 4

PROPOSED CONTROLLER

4.1 Introduction

This chapter presents wavelet based model adaptation strategy for the double EWMA RbR controller.

4.2 Wavelet Based Run-by-Run Controller

To overcome the drawbacks of the single scale filters, Bakshi *et al.* [3] suggested an online, multiscale (OLMS), wavelet based filtering approach. It consists of the following steps:

1. Decompose the measured data within a window of dyadic length (in power of 2) using a wavelet filter.

2. Threshold the wavelet coefficients and reconstruct the filtered signal.

3. Retain only the last data point of the reconstructed signal for online-use.

4. When the measured data are available, move the window in time to include the most recent measurement while maintaining the maximum dyadic window length.

The objective of this research is to integrate the above wavelet based online multiscale filtering approach with the double EWMA RbR control strategy.

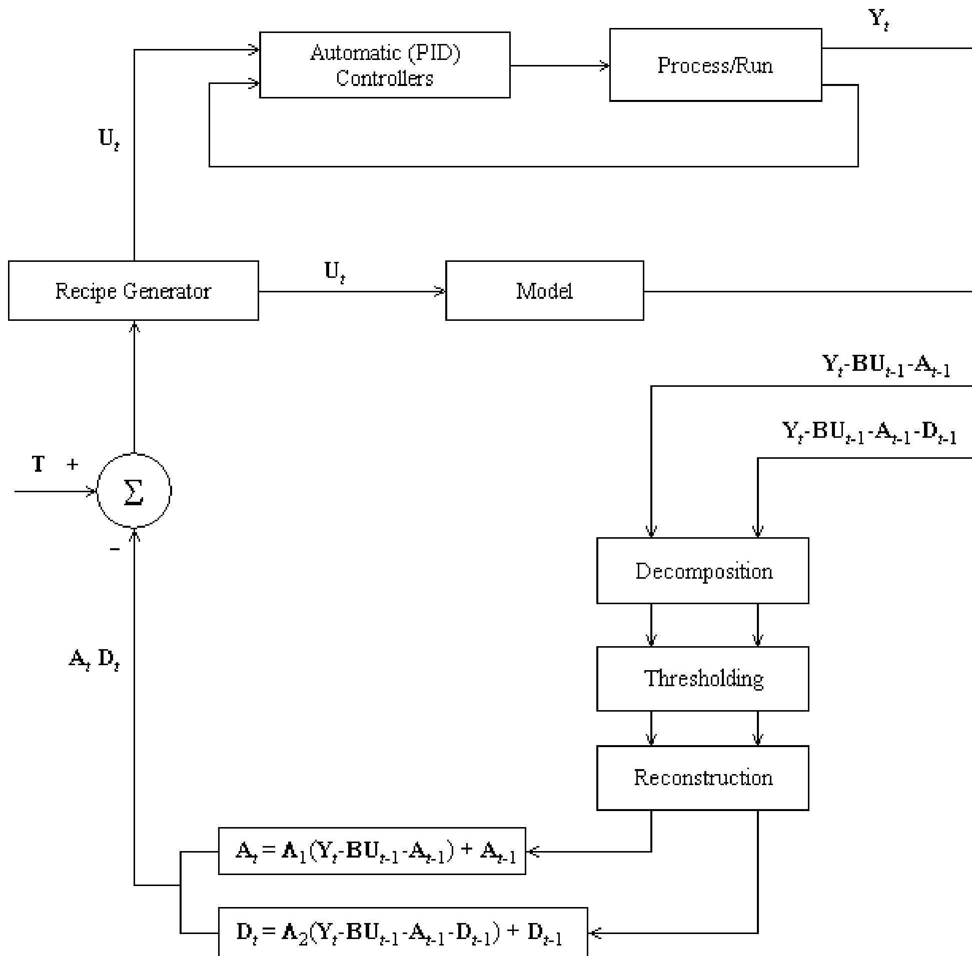

Figure 4.1. A Wavelet Based Multiscale Run-by-Run Controller for MIMO Systems

As can be seen from Figure 4.1, the wavelet based multiscale strategy differs from the double EWMA controller in that it preprocesses the prediction errors before sending them to the EWMA filters.

The proposed strategy consists of the following steps- multiscale decomposition of the prediction errors, thresholding the resulting wavelet coefficients (i.e. denoising) and finally reconstructing the denoised error signal. The steps involved in wavelet based analysis used are discussed henceforth.

4.2.1 Choice of Wavelet

The Daubechies 4th order (Db4) wavelet has been chosen in our analysis because it is compactly supported with external phase and highest number of vanishing moments for a given support width. Also, it has the property of exponentially smoothing the signal.

4.2.2 Determining the Number of Levels for Decomposition

The next step, after choosing the type of the wavelet, is to determine the number of levels up to which the signal should be decomposed. The number of levels should be 4-5 for one-dimensional analysis [39].

4.2.3 Wavelet Decomposition

In order to be able to use the wavelet based methodology online, a moving window approach is used. A dyadic length (of the order 2) of the prediction error for offset ($\mathbf{Y}_t - \mathbf{B}\mathbf{U}_{t-1} - \mathbf{A}_{t-1}$) and drift ($\mathbf{Y}_t - \mathbf{B}\mathbf{U}_{t-1} - \mathbf{A}_{t-1} - \mathbf{D}_{t-1}$) is initially chosen as 2 runs, and is then sequentially increased up to decided number of levels. The data values, within the moving window, are then decomposed using the Db4 wavelet. At each level of decomposition, approximate and detail (wavelet) coefficients are obtained. The

detail coefficients at each scale are the difference between the approximate coefficients at the earlier scale and at that scale.

4.2.4 Denoising by Thresholding

The resulting wavelet coefficients are thresholded using the Visushrink method. The concept of denoising and the Visushrink method are discussed in this section. Any process signal consists of deterministic features and stochastic noise component. The wavelet based filtering has the effect of capturing the deterministic features in a relatively small number of large wavelet coefficients, while distributing the stochastic component among all the coefficients according to its energy at each scale. This noise component does not provide any useful information and is dispensable. When the wavelet coefficients are small, they can be omitted without substantially affecting the general picture. Thus, the rationale behind thresholding wavelet coefficients, or "denoising" as per signal processing terms, is a way of cleaning out 'unimportant' details considered to be noise.

One of the methods of for determining the threshold at each scale is the Visushrink method [40] (or Donoho's universal threshold rule):

$$t_j = \sigma_j \sqrt{2log(n)}, \tag{4.1}$$

where n is the signal length and σ_j is the standard deviation of the noise at scale j. The value of σ_j is estimated from the median of absolute deviation of the wavelet coefficients at scale j as

$$\sigma_j = \frac{1}{0.6745} median(|d_{j,k}|), \tag{4.2}$$

where $|d_{j,k}|$ are the wavelet coefficients. The significant wavelet coefficients are then extracted using the soft thresholding approach as follows.

$$\text{Soft Thresholding: } \tilde{d}_{j,k} = \begin{cases} sign(d_{j,k})(|d_{j,k}| - t) & |d_{j,k}| \geq t \\ 0 & |d_{j,k}| < t \end{cases} \quad (4.3)$$

where, $sign(d_{j,k})$ is the sign of the wavelet coefficient $|d_{j,k}|$. Calculations of the threshold limits were done in *real time.*

4.2.5 Wavelet Reconstruction

The denoised signals, i.e., the prediction errors, are reconstructed from the approximate and detail coefficients. The last data point in the reconstructed signal is then considered and used to recursively update the $\mathbf{A}_t$ and $\mathbf{D}_t$ terms using the two EWMA filters, respectively. The window is then moved to include the latest (data from next run) data point, and as more data is generated.

4.2.6 Calculation of the Recipe/Control Action

The updated offset parameter $\mathbf{A}_t$, and the drift parameter $\mathbf{D}_t$, are now used to compute the input settings (or recipe) for the next run. The control action for the next run is calculated from Equations (3.5), (3.8) or (3.9) depending on the type of the system. This is the setting for the automatic controllers (e.g., PID) for the next run or batch.

CHAPTER 5

EXPERIMENTAL SETUP AND RESULTS

In this chapter the test process on which the controllers are tested is presented first. Thereafter various conditions of the process to which the controllers were applied are outlined. The results for the above study are presented.

5.1 Test Model

Rajagopal and Del Castillo [6] presented a double EWMA based RbR controller for multiple input multiple output processes. They tested their controller's performance on a, four-input and two-output, CMP process [15]. The same CMP process model is used to examine the performance of the wavelet based double EWMA RbR controller (WRBR controller). The process output equations are:

$$Y_1 = 1563.5 + 159.3(U_1) - 38.2(U_2) + 178.9(U_3) + 24.9(U_4) - 0.9t + \epsilon_{1,t}, \qquad (5.1)$$

$$Y_2 = 254 + 32.6(U_1) + 113.2(U_2) + 32.6(U_3) + 37.1(U_4) + 0.05t + \epsilon_{2,t}, \qquad (5.2)$$

where,

t is the time index or run number,

Y_1 = material removal rate,

Y_2 = within-wafer non-uniformity,

U_1 = table speed,

U_2 = back pressure,

U_3 = polishing downforce,

U_4 = the profile of the conditioning system,

$\epsilon_1 \sim N(0, 60^2)$,and

$\epsilon_2 \sim N(0, 30^2)$

The values for other parameters are:

$\Lambda_1 = (0.15)\mathbf{I}$,

$\Lambda_2 = (0.35)\mathbf{I}$,

$\mathbf{I} = (p \times p)$ identity matrix,

$\mu = 0.001$,

T= Target values for the responses in Y = $\begin{pmatrix} 2000 \\ 100 \end{pmatrix}$,

Estimated Gain B= $\begin{pmatrix} 150 & -40 & 180 & 25 \\ 30 & 100 & 30 & 35 \end{pmatrix}$, and

Process Gain $\beta = \begin{pmatrix} 159.3 & -38.2 & 178.9 & 24.9 \\ 32.6 & 113.2 & 32.6 & 37.1 \end{pmatrix}$

5.2 Initial Values

The initial values for the process model were provided as follows:

Offset Estimate $A_t = \begin{pmatrix} 1600 \\ 250 \end{pmatrix}$, and

Drift Estimate $D_t = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

The input parameters not available in [15] are suitably assumed. Initially the process is assumed to be at target.

5.3 Cases Considered

The performance of the wavelet based double EWMA RbR (WRbR) controller is examined for the following cases.

Case 1: Process outputs given by (5.1) and (5.2).

Case 2: Process outputs given by (5.1) and (5.2) with a mean shift.

The controller equation is given by Equation (3.8). The program is coded in MATLAB 6.5, Release 13, using the Wavelet Toolbox functions. The performance measures used are the following.

1. Average Squared Deviation (ASD) from the target

$$\text{ASD}_i = \sum_{t=1}^{m} [\mathbf{Y}_{i,t} - \mathbf{T}]^2/m \quad \text{for all outputs } i = 1, 2. \tag{5.3}$$

2. Standard Deviation of the process outputs.

3. Average Absolute Input Deviation (AAD)

$$\text{AAD}_j = \sum_{t=1}^{m} |\mathbf{U}_{i,t} - \mathbf{U}_{i,t-1}|/m \quad \text{for all inputs } \ j = 1, .., 4. \tag{5.4}$$

where m is the total number of runs.

5.4 Results

The simulation was carried out for 100 runs. The performance of the WRbR and the RbR controller was tested under identical conditions.

5.4.1 Results for Case1

The process outputs are shown in Figure 5.1. Though the plots appear to be similar it is shown later that the WRbR controller has a better performance. The prediction error plots for process offset are shown in Figures 5.2 and 5.3 for outputs 1 and 2 respectively. Similarly, the prediction error plots for process drift are shown in Figures 5.4 and 5.5 for processes 1 and 2 respectively. The black lines represent the prediction errors from the RbR controller, whereas the colored lines represent the

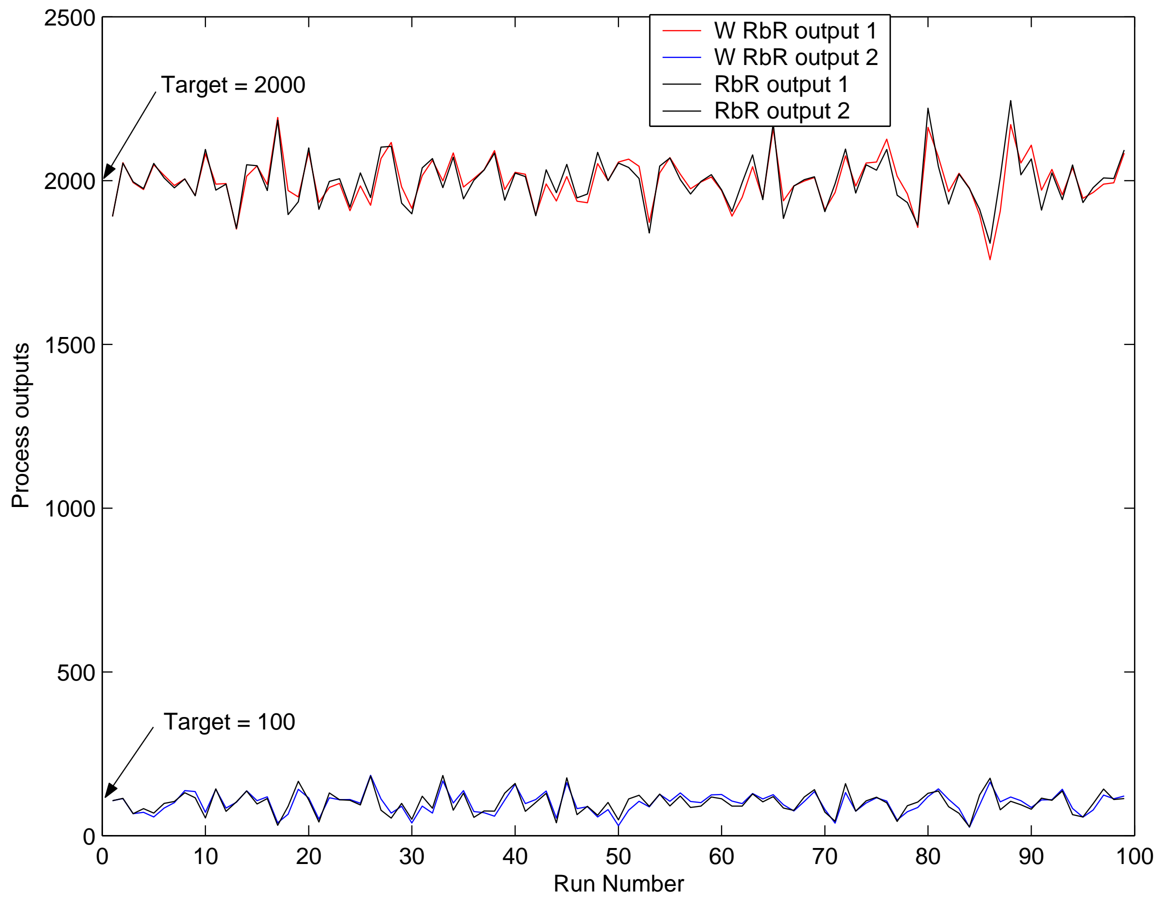

Figure 5.1. Process Outputs

prediction errors from the WRbR controller. As a result of the wavelet based multi-scale analysis, the prediction errors (for offsets and drift) from the WRbR controller are observed to have lesser variation for both the outputs.

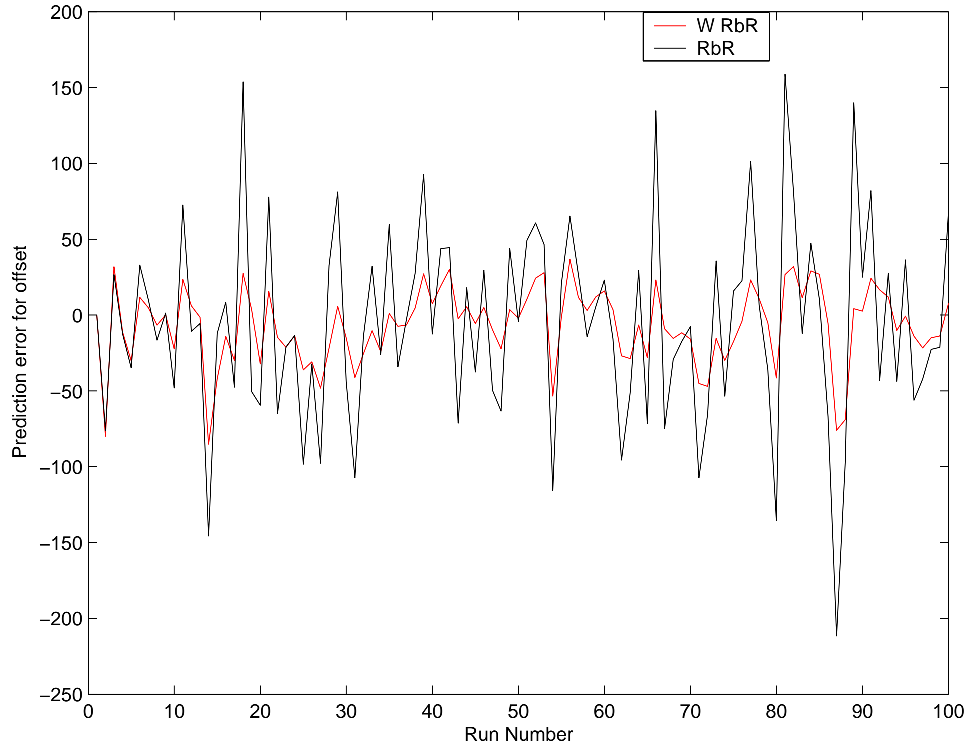

Figure 5.2. Prediction Error for Offset for Process Output 1

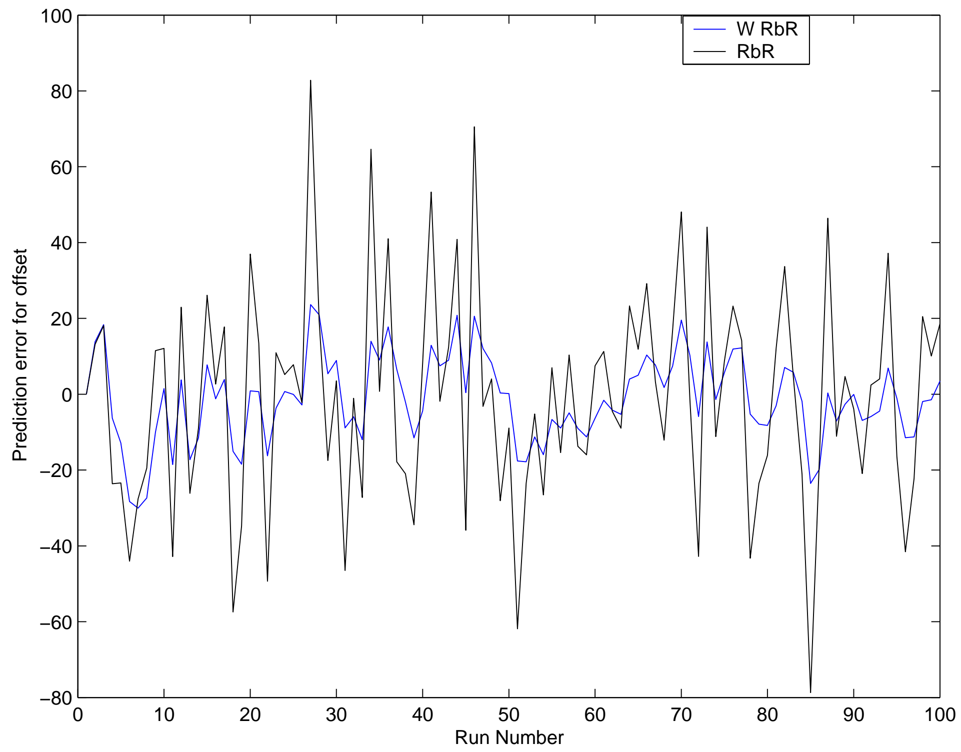

Figure 5.3. Prediction Error for Offset for Process Output 2

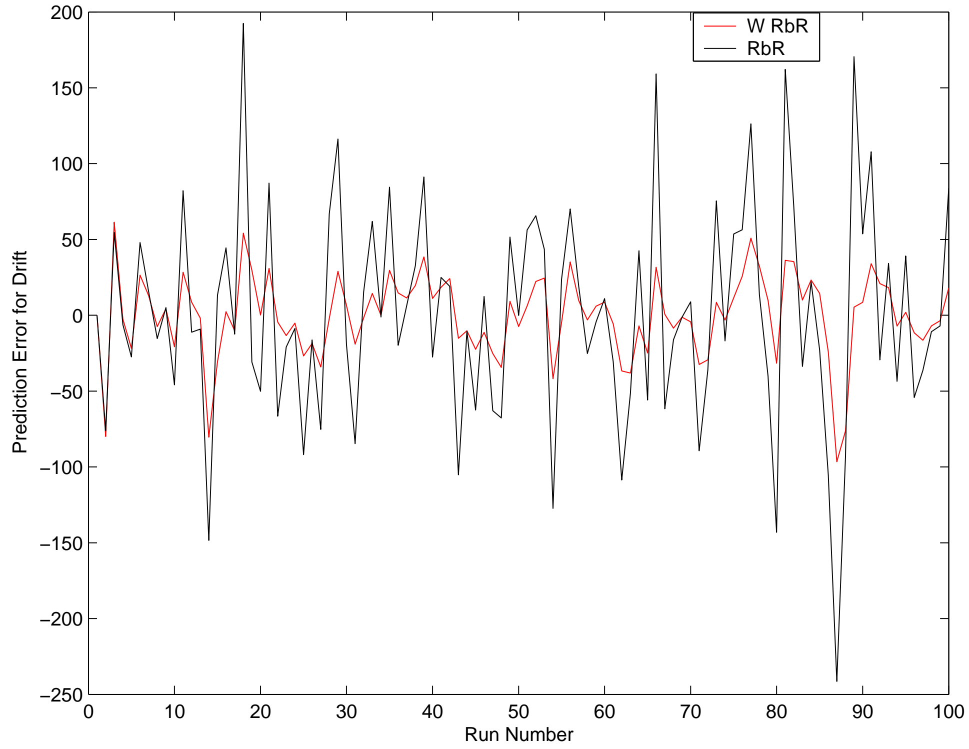

Figure 5.4. Prediction Error for Drift for Process Output 1

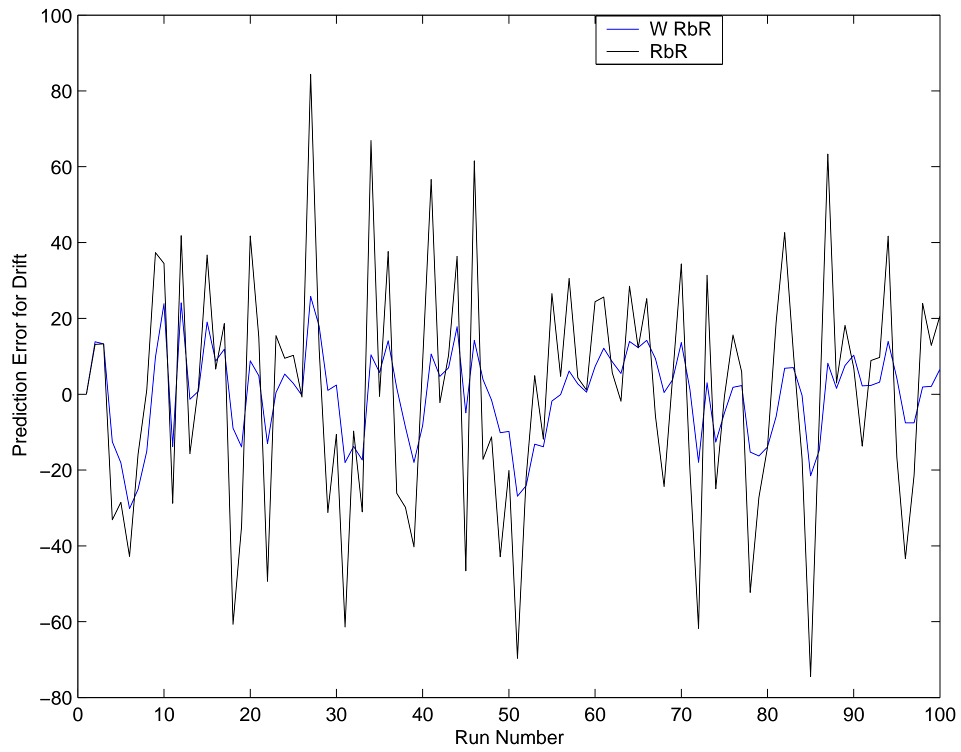

Figure 5.5. Prediction Error for Drift for Process Output 2

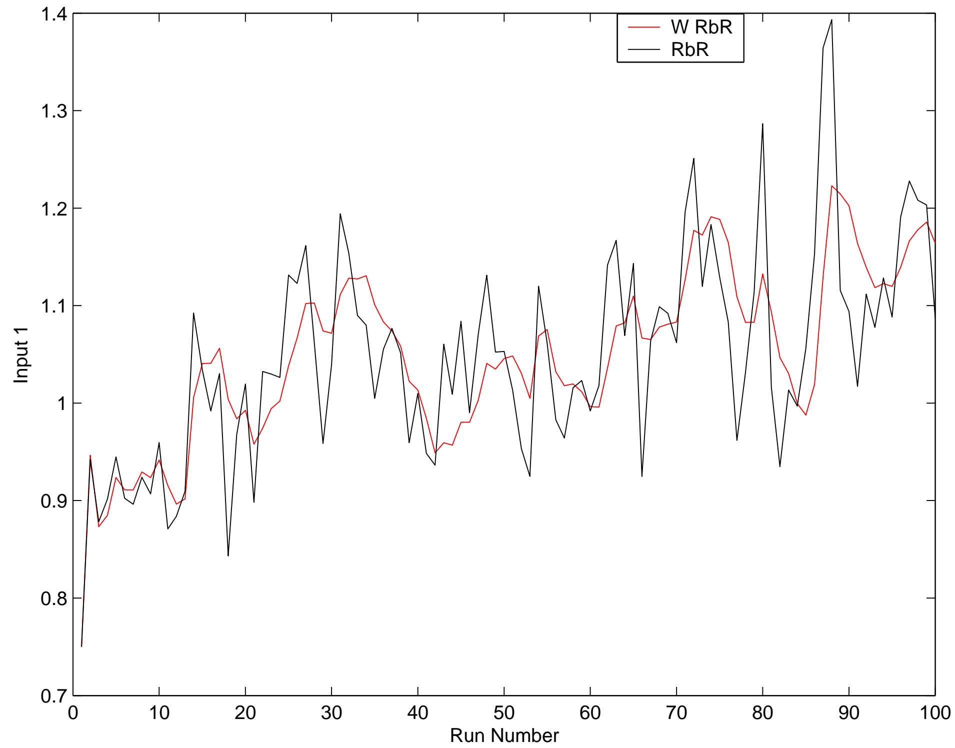

Figure 5.6. Input 1

Table 5.1. Average Absolute Input Deviation ($\times 10^{-2}$)

Inputs	RbR	W RbR	%Decrease
U1	7.19	2.46	65.82
U2	11.62	4.35	62.58
U3	8.67	3.03	65.03
U4	3.38	1.26	62.78

As the result of the smoother prediction errors, the generated recipe would also have lesser fluctuations. This is verified in the input plots shown in Figures 5.6, 5.7, 5.8 and 5.9. The average sum of absolute deviation (AAD) in all the four inputs is summarized in Figure 5.10. Table 5.1 presents the AAD for the process inputs. A decrease of more than 60% is observed for all the four inputs to the WRbR controller. A smoother process operation is the likely result.

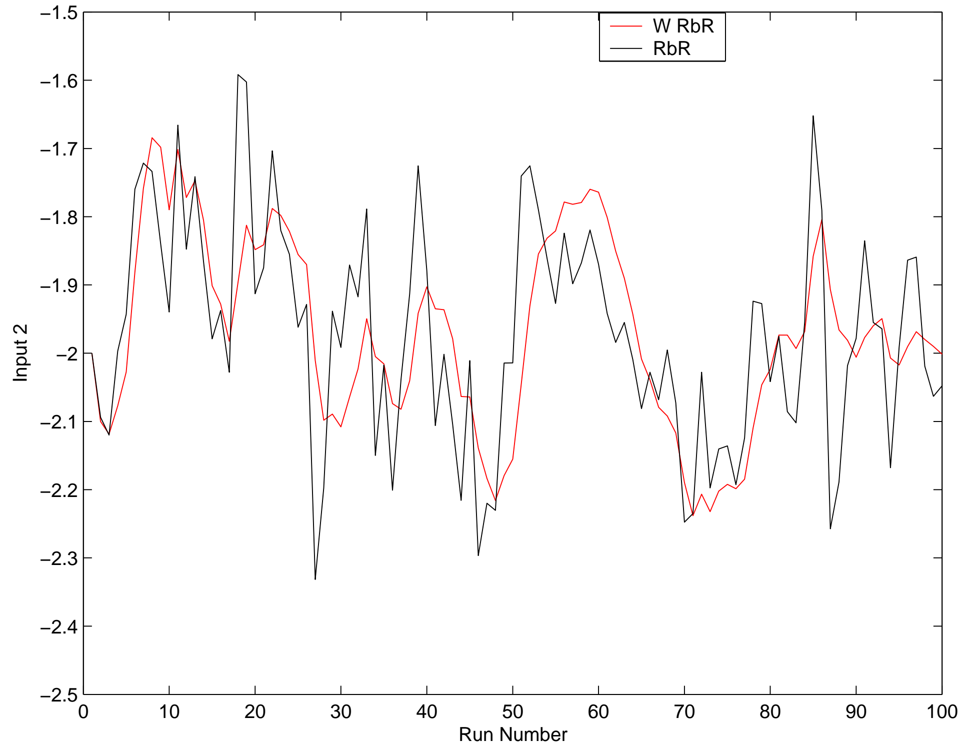

Figure 5.7. Input 2

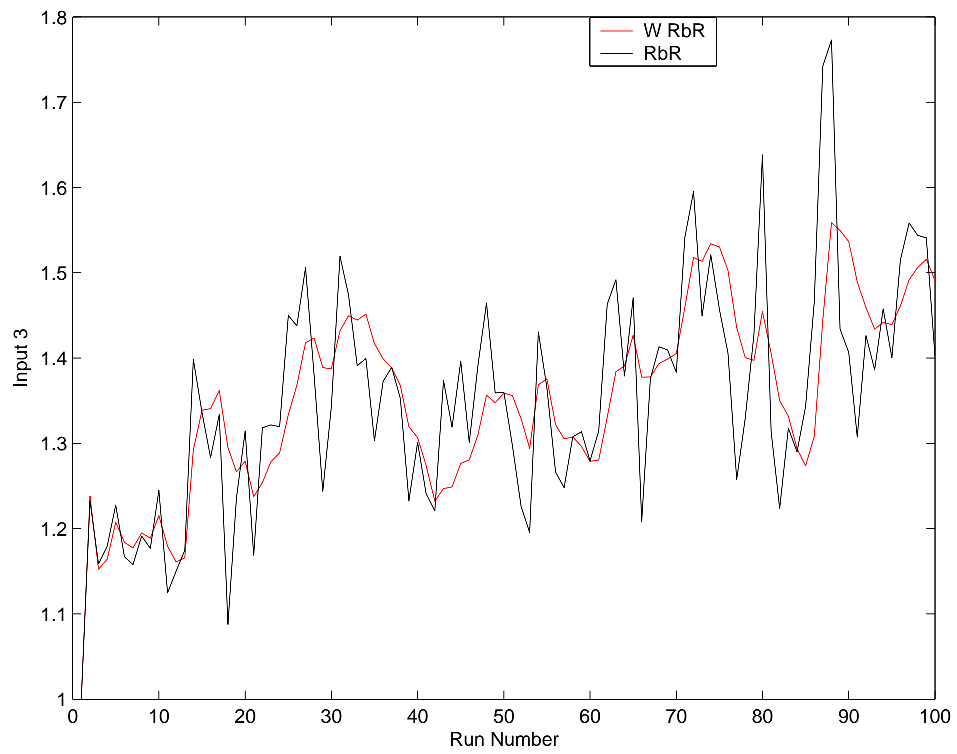

Figure 5.8. Input 3

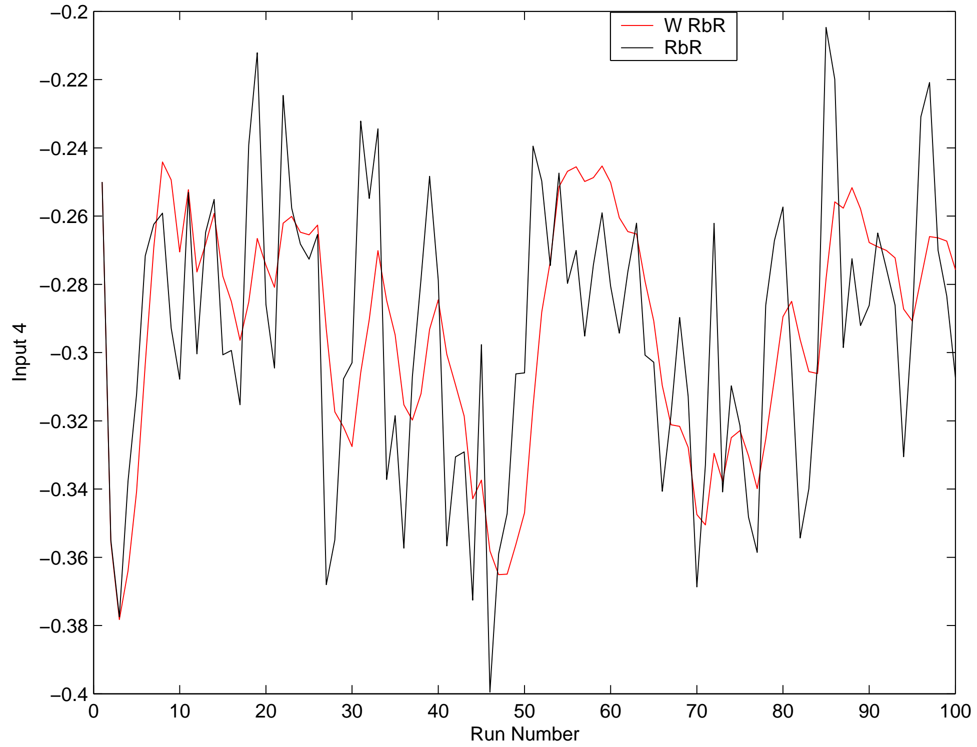

Figure 5.9. Input 4

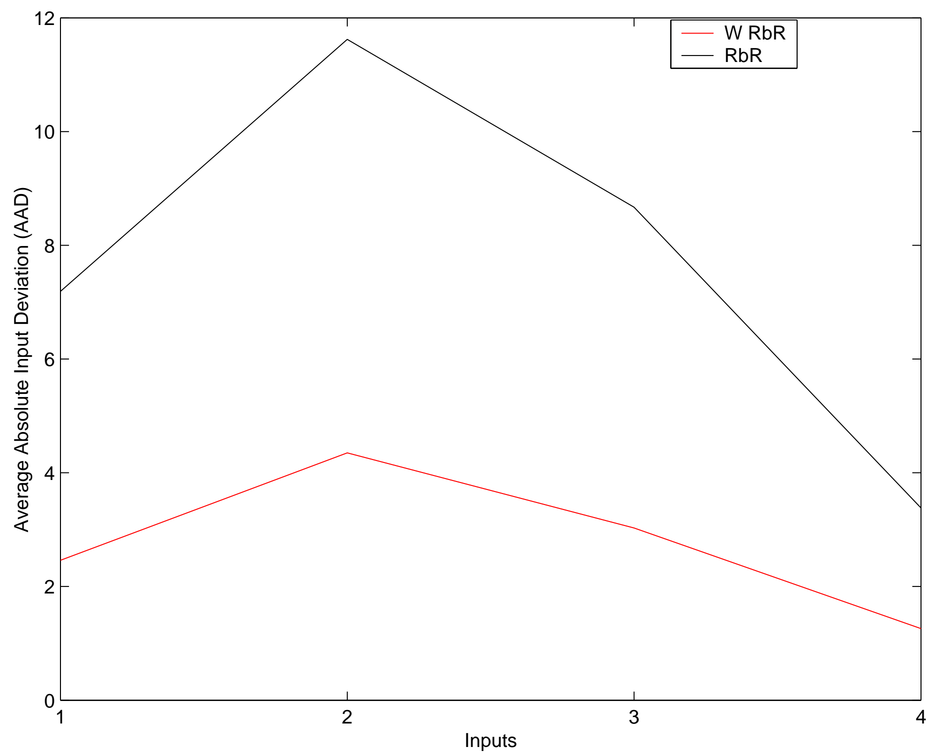

Figure 5.10. Average Absolute Input Deviation ($\times 10^{-2}$)

Table 5.2. Absolute Square Deviation Form Target ($\times 10^3$)

Process	RbR	W RbR	%Decrease
Y1	5.793	5.080	12.31
Y2	1.098	0.968	11.83

Table 5.3. Standard Deviation of the Process Outputs

Process	RbR	W RbR	%Decrease
Y1	76.50	71.63	6.36
Y2	33.31	31.27	6.10

This directly results in a more stable process performance. Tables 5.2 and 5.3 show the numerical values of ASD and Standard Deviation for both the controllers. The WRbR controller is observed to be better than the RBR controller with respect to the standard deviation in the process outputs and ASD. As an illustration, the ASD from the target is observed to be lesser by 12.31% for process output 1. Similarly, the standard deviation for output 1 is observed to be lower by 6.36%.

5.4.2 Results for Case2

A mean shift of magnitude 1000 for Process 1 and 500 for Process 2 are generated at run number 50. The process outputs are plotted in Figure 5.11. As can be observed from the figure, the WRbR controller is faster in bringing the process back to control. Similar to Case1, the wavelet based multiscale treatment of the prediction errors has resulted in lesser variation in all the four process inputs to the WRbR controller, summarized in Table 5.4. The AAD decreased by more than 50% in all the inputs. This can also be confirmed from Figure 5.12, showing a lower magnitude of AAD for all the four inputs.

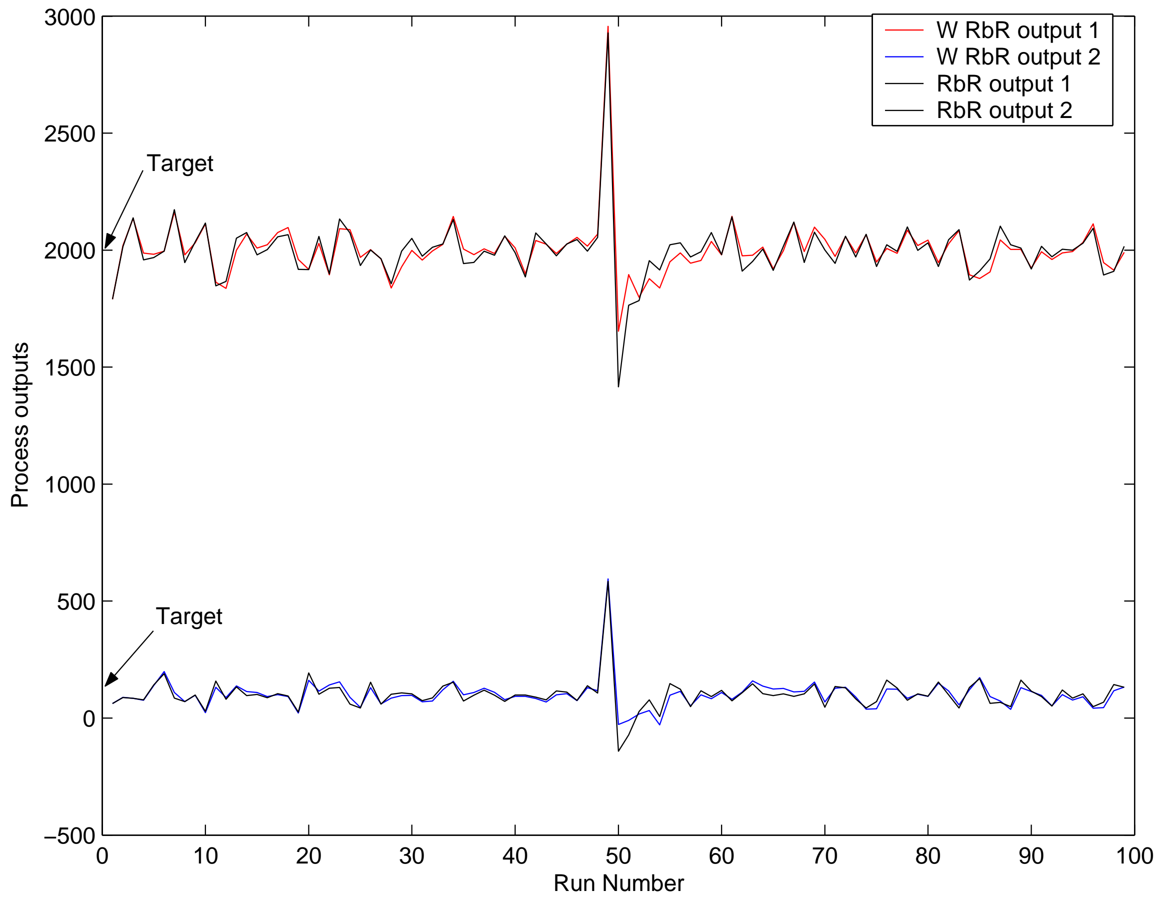

Figure 5.11. Process Outputs Under Mean Shift

Table 5.4. Average Absolute Input Deviation ($\times 10^{-2}$)

Inputs	RbR	W RbR	%Decrease
U1	9.38	4.43	52.78
U2	14.57	5.77	60.42
U3	11.16	5.20	53.42
U4	4.56	2.11	53.78

Figure 5.12. Average Absolute Input Deviation ($\times 10^{-2}$)

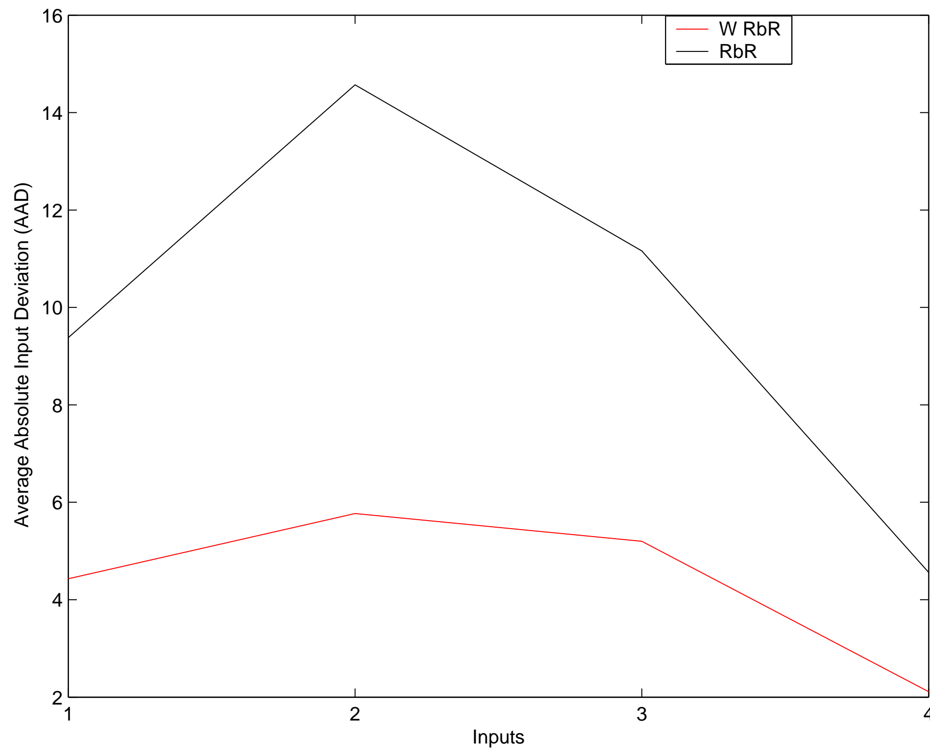

Table 5.5. Absolute Square Deviation Form Target (ASD) ($\times 10^{4}$)

Process	RbR	W RbR	%Decrease
Y1	1.804	1.579	12.45
Y2	0.452	0.407	9.93

The lower AAD directly influences the process performance resulting in a lesser ASD and standard deviation in the process outputs from the WRbR controller. These results are summarized in tables 5.5 and 5.6. For example, in process output 2, the ASD decreased by 9.93%. Similarly, the standard deviation decreased by 5.03% indicating a smoother process performance.

Table 5.6. Standard Deviation of the Process Outputs

Process	RbR	W RbR	%Decrease
Y1	134.98	126.30	6.43
Y2	67.58	64.14	5.09

5.5 Other Research Efforts

In the course of this research, some other avenues were also tried.

1. The wavelet based double EWMA RbR controller was tested for autocorrelation it the process outputs. It is observed that, in the presence of autocorrelation, the WRbR controller had similar performance to that of the RbR controller. Further, it is suggested that for an autocorrelated process, a Triple EWMA RbR controller (refer Tseng *et al.* [41]) would provide an efficient control action.

2. The wavelet based controller was tested for non-linear CMP processes. It is observed that the linear nature of the EWMA equations restricts the ability of the RbR strategy in controlling non-linear processes.

3. The EWMA Equations were considered with some modifications. An additional term $\mathbf{D}_{t-1}$ was introduced in Equation 3.6 and the $\mathbf{A}_{t-1}$ term was replaced by $\mathbf{A}_t$ in Equation 3.7. The results indicate a mean shift in the response outputs.

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH

6.1 Conclusions

This thesis presents a wavelet based multiscale approach for process control. The multiscale technique is integrated with the existing double EWMA RbR control strategy. This research presents a first attempt to introduce the multiscale methodology for run-by-run control of a multiple input multiple output (MIMO) process, which is a norm in a manufacturing environment. The performances of both the WRbR and the RbR controllers are compared for a multiple input multiple output CMP process. The wavelet based controller is shown to provide more effective control action, resulting in a smoother process performance, evaluated through three different control performance measures in comparison to the double EWMA RbR controller in a variety of circumstances. The process outputs from the WRbR controller are observed to have lower ASD and standard deviation. Moreover, it is interesting to note that the AAD decreased by more than 60% in all the four inputs. Further, the WRbR controller is observed to have a better control performance in dealing with mean shifts in a process.

6.2 Future Research

The wavelet based controller is tested for generic processes. A logical extension would be to validate the results using real time data. Development of software packages for practical implementation would follow. The multiscale methodology

can further be tested for other single scale filters. The control strategy can be modified to account for process dynamics like autocorrelation.

REFERENCES

[1] T. H. Smith, D. S. Boning, J. Stefani, and S.W. Butler. Run by run advanced process control of metal sputter deposition. *IEEE Transactions on Semiconductor Manufacturing*, 11(2):276–284, 1998.

[2] E. Sachs, A. Hu, and A. Ingolfsson. Stability and sensitivity of an ewma controller. *Journal of Quality Technology*, 25(8):271–287, 1993.

[3] M. N. Nounou and B. R. Bakshi. On-line multiscale filtering of random and gross errors without process models. *Journal of Chemometrics*, 12:415–434, 1999.

[4] A. Ingolfsson and E. Sachs. Run-by-run process control: Combining spc and feedback control. *IEEE Transactions on Semiconductor Manufacturing*, 8(8):26–43, 1995.

[5] E. Del Castillo and A. M. Hurwitz. Run-to-run process control: literature review and extensions. *Journal of Quality Technology*, 29(1):184–196, 1997.

[6] E. Del Castillo and R. Rajagopal. A multivariate double ewma process adjustment scheme for drifting processes. *IIE Transactions*, 34, 2002.

[7] G. E. P. Box and T. M. Kramer. Statistical process monitoring and feedback adjustment - a discussion. *Technometrics*, 34, 1992.

[8] D. C. Montgomery, J. B. Keats, G. C. Runger, and W. S. Messina. Integrating statistical process control and engineering process control. *Journal of Quality Technology*, 26(2):76–87, 1994.

[9] S. A. Van der Weil, W. T. Tucker, F. W. Faltin, and N. Doganaksoy. Algorithmic statistical process control: Concepts and an application. *Technometrics*, 34:286–297, 1992.

[10] W.T. Truker, F. W. Faltin, and S. A. Van der Weil. Algorithmic statistical process control: an elaboration. *Technometrics*, 35:363–375, 1993.

[11] M. Janakiram and J. B. Keats. Combining spc and epc in a hybrid industry. *Journal of Quality Technology*, 30:189–200, 1998.

[12] S. W. Butler and J. A. Stefani. Supervisory run-to-run control of a polysilicon gate etch using in situ ellipsometry. *IEEE Transactions on Semiconductor Manufacturing*, 7:193–201, 1998.

[13] E. Del Castillo. Long run and transient analysis of a double ewma feedback controller. *IIE Transactions*, 31:1–13, 1999.

[14] E. Del Castillo. Some properties of the ewma feedback quality adjustment schemes for drifting disturbances. *Journal of Quality Technology*, 33(2):153–166, 2001.

[15] R. Rajagopal and E. Del Castillo. An analysis and mimo extension of a double ewma run-to-run controller for non-squared systems. *International Journal of Reliability, Quality and Safety Engineering*, 10(4):417–428, 2003.

[16] S. T. Tseng, R. J. Chou, and S. P. Lee. A study of a multivariate ewma controller. *IIE Transactions*, 36(4):541–550, 2002.

[17] L. Ljung. *System Identification.* Prentice Hall, Englewood Cliffs, NJ, 1987.

[18] S. L. Shah and W. R. Cluett. Recursive least squares based estimation schemes for self-tunging control. *Canadian Journal of Chemical Engineering*, 69:89–100, 1991.

[19] P. Young. *Recursive estimation and time-series analysis: an introduction.* Springer Verlag, New York, 1984.

[20] E. Del Castillo and J. Yi-Yeh. An adaptive run-to-run optimizing controller for linear and nonlinear semiconductor processes. *IEEE Transactions on Semiconductor Manufacturing*, 11(2):285–295, 1998.

[21] E. Palmer, W. Ren, and C. J. Spanos. Control of photoresist properties: A kalman filter based approach. *IEEE Transactions on Semiconductor Manufacturing*, 9(2):208–214, 1996.

[22] C. Zhang, H. Deng, and J. S. Baras. "the set valued run-to-run controller with ellipsoid approximation". In *Proceedings of the AEC/APC 2000*, Lake Tahoe, NJ, 2000.

[23] Z. Ning, J. R. Moyne, T. Smith, D. Boning, E. DelCastillo, J. Yi-Yeh, and A. Hurwitz. Comparitive analysis of run-to-run control algorithms in the semiconductor manufacturing industry. In *Proceedings of 32nd Hawai international conference on system sciences*, pages 375–381. IEEE/SEMI, 1996.

[24] Y. L. Huang, T. L. Edgar, Himmelblau, and I. Trachtenberg. Constructing a reliable neural network model for a plasma etching process using limited experimental data. *IEEE Transactions on Semiconductor Manufacturing*, 7(3):333–345, 1994.

[25] P. K. Mozumder and G. G. Barna. Statistical feedback control of a plasma etch process. *IEEE Transactions on Semiconductor Manufacturing*, 7:1–11, 1994.

[26] D. Boning, W. Moyne, T. Smith, J. Moyne, R. Telfeyan, A. Hurwitz, S. Shellman, and J. Taylor. Run-by-run control of chemical-mechanical polishing. *IEEE Transactions CPMT (C)*, 19(4), 1996.

[27] A. Hu, X. Zhang, E. Sachs, and P. Renteln. Application of run by run controller to the chemical mechanical planarization process part i. In *IEEE Proc. Of the15th Int. Elect. Manuf .Tech. Symp.* IEEE, 1993.

[28] D. Morlet. Time scale analysis of high resolution signal averaged surface ecg using wavelet transformation. In *Proceedings. Computers and Cardiology.*, pages 393–396, 1991.

[29] I. Daubechies. *Ten Lectures in Wavelets.* John Wiley, 1992.

[30] A. Grossman and D. Morlet. Decomposition of hardy functions in to square integrable wavelet of constant shape. *SIAM J. Math. Anal.*, 15:723–730, 1984.

[31] D. L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrica*, 81:425–455, 1994.

[32] D. L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrica*, 57(2), 1995.

[33] A. Witkin. Scale space filtering: A new approach to multiscale description. In *Proceedings - ICASSP, IEEE International Conference on Acousitcs, Speech and Signal Processing.*, volume 3, pages 1–39A, 1984.

[34] G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley MA 02181, 1996.

[35] C. Burrus, R. Gopinath, and H. Guo. *Introduction to wavelets and wavelet transforms: a primer.* Prentice Hall, Upper Saddle River, N.J., 1998.

[36] S. G. Mallat. Multifrequency channel decompositionsof images and wavelet models. *IEEE Transactions on Acoustic Speech and Signal Processing*, 37:2091–2110, 1989.

[37] R. Ganesan, T.K. Das, and V. Venkataraman. Wavelet based multiscale statistical process monitoring - literature review and extensions. *Submitted to IEEE Transactions*, 37, 2002.

[38] B. Rao. A wavelet modulated run-by-run controller. Master's thesis, University of South Florida, 2003.

[39] M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi. *Wavelet Toolbox: For Use with MATLAB.* The Mathworks, Inc., 3, Apple Hill Drive, Natick, MA, 1996.

[40] D. L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage : Asymptotia? (with discussion). *Journal of the Royal Statistical Society*, 57(2), 1995.

[41] Shu-Kai S. Fan, Bernard C. Jiang, Chih-Hung Jen, and Chien-Chih Wang. Siso run-to-run feedback controller using triple ewma smoothing for semiconductor manufacturing processes. *International Journal of Production Research*, 40(13):3093–3120, 2002.