Sheet Profile Estimation and Machine Direction Adaptive Control

by

Lee Rippon

B.A.Sc., The University of British Columbia, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Chemical and Biological Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2017

Abstract

Sheet and film process control is often structured such that separate controllers and actuators are dedicated to either the temporal (i.e, machine direction) variations or the spatial (i.e., cross direction) variations. The dedicated machine direction (MD) and cross direction (CD) controllers require separate measurements of the MD and CD sheet property profiles, respectively. The current industrial standard involves a traversing sensor that acquires a signal containing both MD and CD property variations. The challenge then becomes how does one extract separate MD and CD profiles from the mixed signal. Numerous techniques have been proposed, but ultimately the traditional exponential filtering method continues to be the industrial standard. A more recent technique, compressive sensing, appears promising but previous developments do not address the industrial constraints. In the first part of this thesis the compressive sensing technique is developed further, specifically with regards to feasibility of implementation. A comparative analysis is performed to determine the benefits and drawbacks of the proposed method.

Model-based control has gained widespread acceptance in a variety of industrial processes. To ensure adequate performance, these model-based controllers require a model that accurately represents the true process. However, the true process is changing over time as a result of the various operating conditions and physical characteristics of the process. In part two of this thesis an integrated adaptive control strategy is introduced for the multi-input multi-output MD process of a paper machine. This integrated framework consists of process monitoring, input design and system identification techniques developed in collaboration with multiple colleagues. The goal of this work is to unify these efforts and exhibit the integrated functionality on an industrial paper machine simulator.

Preface

This thesis consists of the following two distinct parts:

- **Part I:** Sheet Profile Estimation,
- **Part II:** Machine Direction Adaptive Control.

In Part I the discussion is relevant to various sheet and film processes whereas in Part II the proposed techniques have widespread applicability to a larger variety of industrial processes that implement model-based control systems. For the purpose of this thesis, the techniques in both Part I and Part II are developed and tested specifically in the context of application on a paper machine. Focusing on a specific industrial process allows for the important and necessary consideration of the realistic industrial constraints. Furthermore, it allows for enhanced testing and result verification using the Honeywell paper machine simulator. Incorporating the expertise and resources of Honeywell, our industrial collaborator, is crucial to the practical significance of the resulting analyses.

Part I begins with introductory material into the paper machine, scanning sensor, signal processing and the problem of MD aliasing. Various MD-CD separation techniques that have been proposed in literature are then reviewed and some of the more promising methods are selected for further development. Specifically, the power spectral analysis and compressive sensing techniques are developed further in this work to enable implementation with the sampling structure of the scanning sensor. A simulated comparative analysis is performed that incorporates a realistic CD profile and the sampling geometry of the scanning sensor. Analyzing the experimental results helps to reveal the benefits and drawback of the various MD-CD separation methods. Concluding statements and recommendations for future work are provided in the final chapter of Part I.

The second part of this thesis involves unifying a series of developments made by multiple colleagues throughout our collaborative research initiative with Honeywell. Specifically model-plant mismatch detection techniques developed by Qiugang Lu and closed-loop system identification

techniques developed in collaboration with Qiugang Lu are implemented in this work. Combining these techniques with input design methods developed in collaboration with Mahdi Yousefi facilitate the development of a unified adaptive control framework. Contributions made in this work include integrating these developments into a functional adaptive control architecture and demonstrating the effectiveness on a Honeywell paper machine simulator of the machine direction process.

This work was made possible thanks to the valuable advice of my academic supervisors Bhushan Gopaluni and Philip Loewen at the University of British Columbia as well as my industrial supervisors Michael Forbes and Johan Backstrom at Honeywell Process Solutions. Furthermore, my research colleagues Qiugang Lu and Mahdi Yousefi were instrumental in the development of this research, both in terms of technical developments and as mentors. Finally I would like to acknowledge my family and friends for helping to support me throughout this work.

Table of Contents

Appendices

List of Tables

List of Figures

Part I

Sheet Profile Estimation

Chapter 1

Introduction

Improved sheet, web and film property estimation is motivated by the ongoing desire for higher quality, more consistent products and the constant desire for maximum production efficiency. In a competitive market customers are able to set strict constraints and producers are often at a strategic advantage if they can offer a premium product. In this chapter the property estimation problem will be described for a broad set of relevant industries as well as the specific industry of study.

1.1 Sheet and film processes

Collectively, sheet and film processes such as paper making, metal rolling, coating and polymer film extrusion constitute a vast global market. As such, improving the operational efficiency, and thus the control of sheet and film processes, such as those shown in Figure 1.1, is of significant industrial importance. Broadly speaking, optimal operational efficiency involves maximizing the rate of on-grade product (i.e, conforms to customer specifications) while minimizing operational expenses (e.g., energy, raw materials, human intervention). In reality, the objective is much more complex, requiring adherence to a variety of safety, environmental, and economic constraints. A common control objective is to minimize the variation of a measured output around a specified target value (i.e., set-point). This allows for the production of a high quality sheet that meets customer specifications [57].

What makes sheet and film processes similar is that property variations are minimized along both the length and the width of the sheet, also known as the machine direction (MD) and the cross direction (CD), respectively [22]. In applications where the sheet is moving, the MD profiles represent temporal variations and the CD profiles represent spatial variations. Specific MD actuators are controlled by dedicated MD controllers that require an accurate profile of the MD variations. Similarly, the CD variations are controlled by a dedicated CD controller that uses a CD profile of

Figure 1.1: Various sheet and film processes [34][39][53]

the sheet to control spatially distributed actuators. In order to accurately control the MD and CD variations it is necessary to reconstruct accurate MD and CD output profiles from the measured outputs. Outputs are often measured by a scanning sensor that traverses the moving sheet and samples a mixture of MD and CD variations [15][57]. The problem of reconstructing accurate MD and CD profiles from this scanner signal is commonly referred to as MD-CD separation and it constitutes the primary research problem addressed in Part I of this thesis.

Improved estimation of sheet and film profiles is motivated from the following two perspectives:

1. **From a control perspective:** in order to minimize output variations around a set-point, it is important to observe where the output is relative to the set-point. From a control perspective, the focus is on producing MD and CD output profiles that help increase the performance of the respective controllers.

2. **From a quality reporting perspective:** improved estimation of sheet and film properties can help provide a more accurate account of product quality to the customer. From a reporting

perspective, the focus is on quantifying an accurate representation of the sheet/film properties for the batch of product sold to the customer.

Although these two perspectives share a large over-lap in their motivations, the constraints and requirements involved in each problem are unique. The majority of this thesis is motivated from the control perspective, although quality reporting is also addressed. For thorough consideration of the practical industrial complexities, the remainder of this thesis narrows in scope and focuses specifically on application with a paper machine. However, it is important to note that the MD-CD separation analysis is relevant to all sheet and film processes where the sheet is moving in the MD while being measured by a sensor traversing in the CD.

1.2 The paper machine

As a staple of the Canadian economy, the multibillion dollar pulp and paper industry employs thousands of people and is a major source of Canada's manufactured goods and exports. Increased global competition, reduced growth in newsprint demand and volatile commodity prices have created challenges for the industry [45][61]. Moreover, decreasing profitability margins and increasing customer constraints further intensify the need for consistent product quality and improved control of operations [50]. Perhaps most importantly, the pulp and paper industry is an integral component of the larger forest products industry as the profitability of saw mills depends on revenue generated through selling chips to pulp mills [49]. Thus, it is imperative for the Canadian economy to ensure the pulp and paper industry is a global leader in efficiency and technological innovation.

A Fourdrinier paper machine, as depicted in Figure 1.2, consists of a series of dewatering and pressing operations that transform a slurry of pulp fiber into a uniform sheet of paper. These complex and expensive machines can be over 100 m long in the MD and produce a sheet of paper over 10 m wide at rates exceeding 30 m/s [5]. In the forming section the pulp fiber flows from the thick stock pump to the headbox where it is diluted and sprayed through a slice lip onto a moving wire or felt. In the pressing and drying sections the wet sheet of fiber is compressed and dried by passing through a series of metal rollers, some of which are heated with high pressure steam. Finally, the sheet is smoothed into a final product before collecting on a reel at the end of the machine [59].

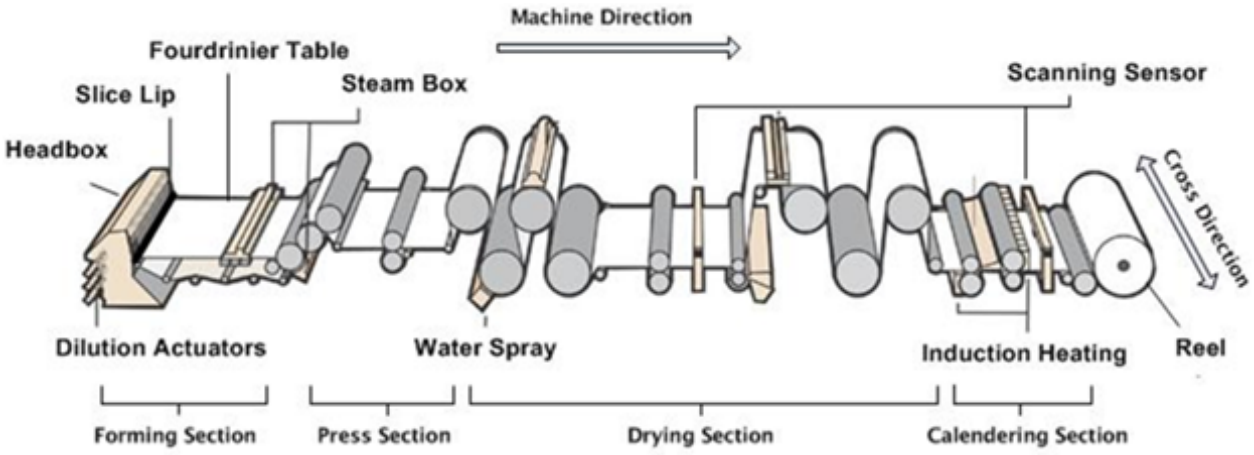

Figure 1.2: Schematic of a Fourdrinier paper machine [16]

The objective in paper-making, from a quality control perspective, is to efficiently produce a uniform sheet of paper with consistent properties that match customer specifications. Measured outputs such as caliper, moisture content and basis weight are used to determine actuator movements that correct deviations from the set-point [29]. Control of sheet properties is separated into orthogonal temporal and spatial components. Dedicated MD and CD control systems are used to address temporal and spatial variations, respectively. Set-point tracking changes are primarily addressed with the MD control system and the spatial variations across the sheet are tuned with the CD control system. The MD control system uses dedicated MD actuators such as thick stock valve positioning and dryer steam pressure to minimize the temporal variations. These actuators have a relatively uniform effect on the spatial properties of the sheet [60]. Alternatively, the spatial variations across the sheet are minimized by manipulating spatially distributed CD actuators such as headbox dilution valves and induction profilers [21]. To implement feedback control the MD and CD controllers require measured output profiles which are typically acquired using a scanning sensor.

1.3 The scanning sensor

Although various sampling configurations exist, a prevalent method in paper-making is to use a scanning sensor. Scanning sensors were first introduced around 1980 when they began to replace the single-point sensors which had a fixed CD position [58]. The scanning sensor framework consists of a

scanner head mounted on a frame that spans the width of the sheet. The scanner head is filled with a variety of sensors that can measure sheet properties such as basis weight, moisture content, and caliper (i.e., thickness) [4]. Sometimes paper machines use multiple scanners at different locations along the MD to obtain additional samples, as shown in Figure 1.2. However, the cost associated with adding multiple scanners is generally prohibitive due to the expensive sensor technology [47].

As the sheet moves at $v_{\text{sheet}} \approx 30$ m/s in the MD, the scanner traverses the sheet at $v_{\text{scan}} \approx 1$ m/s in the CD. The resulting sampling trajectory forms a diagonal zig-zag pattern across the sheet, as shown in Figure 1.3 below. The scanner frame and scanner head are depicted in grey while

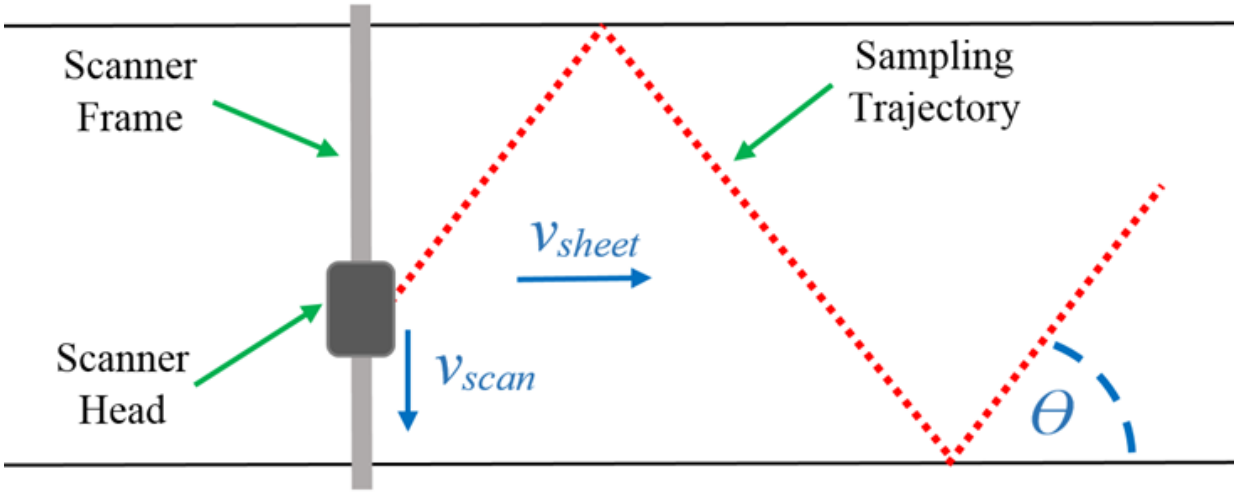

Figure 1.3: Schematic of a scanning sensor

the blue arrows indicate the direction in which the scanner and sheet are moving. The resulting sampling trajectory is described by the dotted red line although it is important to note that this depiction is not to scale.

In reality, the angle θ from the sampling trajectory to the MD axis is much smaller, typically $\theta \approx 1°$. Specifically, θ is a function of the scan time (t_{scan}), the sheet speed (v_{sheet}), and the sheet width (w_{sheet}). Although varying scan speeds can be used, it is typical in industry for the scan speed (v_{scan}) to be fixed for long periods of operation. Ignoring occasional stationary periods for scanner calibration, the scan time can be determined as simply $t_{\text{scan}} = \frac{w_{\text{sheet}}}{v_{\text{scan}}}$. Assuming realistic industrial values of $w_{\text{sheet}} = 10$ m and $v_{\text{scan}} = 0.5$ m/s, the time required for one traverse of the sheet is $t_{\text{scan}} = 20$s. Furthermore, using a realistic sheet speed of $v_{\text{sheet}} = 30$ m/s, the scanning

angle, θ, from the MD can be determined using [2]

$$\theta = 90 - \arctan(v_{\text{sheet}} \frac{t_{\text{scan}}}{w_{\text{sheet}}}), \tag{1.1}$$

to be $\theta = 0.95°$. However, this value is dependent on application and operation, e.g., an 8 m wide sheet moving at 20 m/s with a 1 m/s scanner results in $\theta = 2.86°$. Assuming a constant scan speed, the expression on the right hand side (RHS) of (1.1) simplifies to $90 - \arctan(\frac{v_{\text{sheet}}}{v_{\text{scan}}})$. This detailed analysis of the sampling trajectory provides important background information for the simulated experimental setup described in Chapter 3.

Due to the diagonal sampling trajectory the resulting signal acquired by the scanner contains a mixture of MD and CD property variations. From this signal separate MD trends and CD profiles need to be determined and provided to the MD and CD control systems, respectively. One of the more straightforward proposals to address the need for MD-CD separation is to use an array of stationary sensors that span the width of the sheet. A sensor array would likely result in a straightforward acquisition of enhanced MD trends and CD profiles as well as a vastly improved determination of the overall product quality. However, as mentioned before the sensor technology can be prohibitively expensive and the economic justification based on improved control and product quality reporting is not straightforward. Until sheet spanning sensing mechanisms become commonplace in industry, the problem of MD-CD separation with a scanning sensor configuration remains relevant [37].

1.4 Characterizing paper sheet variations

Understanding the nature of paper machine disturbances and sheet variations is critical when deciding which scanner signal processing techniques are appropriate. This is particularly true when distinguishing between techniques that are suitable for either control or product quality reporting. As alluded to earlier, the requirements for each of these perspectives are not equivalent. For example, while a higher resolution of the estimated profile might be beneficial for quality reporting, it might not improve the control of sheet properties. To further appreciate the MD-CD separation study, some background information on paper sheet variations is provided here.

Paper sheet variability is quantified primarily along the MD and CD axes, with the remaining

variability referred to as residual variability. A more detailed discussion of paper machine variance partitioning is provided in Section 1.6. Variability within a paper sheet ranges over a large spectrum from the level of individual fibers to the length of an entire reel of paper. A study of MD basis weight variation on a paper machine with a sheet speed of 11.7 m/s using a single-point sensor led to the data shown in Table 1.1 below [19]. These MD basis weight variations are categorized based on their relationship to the process, as shown in Table 1.2.

Table 1.1: MD basis weight variance distribution ($v_{\text{sheet}} = 11.7$ m/s) [19]

Frequency (Hz)	Range (s)	Percent of total variance[1]
25-250	0.004-0.04	15
2.5-25	0.04-0.4	40
0.25-2.5	0.4-4.0	10
0.006-0.25	4.0-167	20
0.0016-0.006	167-600	15

Table 1.2: Causes of MD basis weight variations ($v_{\text{sheet}} = 11.7$ m/s) [19]

Category	Frequency (Hz)	Potential sources
Formation	> 117	fiber properties, turbulence in headbox
Short term	1-117	hydraulic pulsations, equipment vibration
Medium term	0.005-1	blending, flow stability, fast control loops
Long term	< 0.005	stock prep, pulp handling, slow control loops

Unfortunately this work does not provide the same analysis for CD basis weight variations. Instead, the common assumption that CD profiles are essentially constant over time was studied and found to be technically incorrect. However, the maximum variation in consecutive CD basis weight profiles was found to be small (i.e., $\approx 6\%$) when 10 plies were tested. It was necessary to use multiple plies because otherwise the analysis would be incorrectly dominated by residual variations [19]. So while the CD profile is technically not independent of time, for control purposes assuming it is constant for short periods of time (e.g., a few scans) is reasonable.

Of particular interest for MD-CD separation is estimating the MD and CD variations that are within the range of controllable frequencies. For a sheet moving at 16.7 m/s the spectrum of variations related to process control starts at approximately 1 Hz and extends to slower variations. The highest cutoff frequency for MD control loops is for the headbox total head at around 0.05 Hz. Other control loops, such as basis weight, have a much lower cutoff frequency [8]. The spatially

[1] Average of eight separate records

distributed CD actuators may have up to 300 individual actuators and the actuator spacing (for a slice lip system) is anywhere from 7.0 cm to 20 cm with a bandwidth of around 15 cm [16] [54]. The MD and CD controllable frequencies provide a lower bound on the required resolution for MD-CD separation methods to provide adequate MD and CD profiles. For the purpose of this thesis, the target resolution in the developed MD-CD separation methods is conservatively higher than the aforementioned lower bounds. This allows for both improvements in the MD and CD control system bandwidths as well as for improvements in the quality reporting statistics.

1.5 Preliminary scanner signal processing

From both a control perspective and a quality reporting perspective the acquired scanner signal, $\boldsymbol{y} \in \mathbb{R}^n$, needs to be processed such that it represents the two dimensional reality of the process, i.e., $\mathbf{Y} \in \mathbb{R}^{n_{\text{CD}} \times n_{\text{MD}}}$. Here n is equal to the total number of measurements taken by the scanning sensor. In an ideal setting, such as with a high density sensor array, n_{CD} would represent the number of sensors distributed along the CD and n_{MD} would represent each sample time. Unfortunately, with a scanning sensor some simplifying assumptions are often relied upon instead.

Each time the scanner traverses the moving sheet of paper is considered one scan, during which up to 10,000 locations along the CD profile can be sampled [57]. An initial data verification step is performed to remove any illogical sample values. Next, these measurements are down-sampled into evenly spaced CD bins at a lower common resolution (e.g., 500 bins) that is more computationally tractable for the model-based control systems [33]. Therefore, if $\boldsymbol{y} \in \mathbb{R}^n$ represents the scanner signal after this initial processing then the total number of measurements, n, is equal to the number of CD bins multiplied by the number of scans being analyzed. For both control purposes and quality reporting the standard method for achieving $\mathbb{R}^n \mapsto \mathbb{R}^{n_{\text{CD}} \times n_{\text{MD}}}$ is to simply collapse the measurements of each scan into a column vector where the rows represent the common resolution CD bins. The resulting matrix of scanner values, $\mathbf{Y}$, is depicted in Table 1.3.

In Table 1.3, n_{MD} is equal to the number of successive scans being analyzed and n_{CD} is equal to the number of common resolution CD bins. While the rows correctly represent the approximate spatial positioning of each measurement, the columns instead treat the temporal positioning of each measurement during a scan as if the sheet was stationary and they all occurred at the same time.

Table 1.3: Two-dimensional representation of scanner signal

		Scans		←	MD	→
		Scan 1	$\cdots$	*Scan j*	$\cdots$	*Scan* n_{MD}
Bins	*Bin 1*	$Y_{1,1}$	$\cdots$	$Y_{1,j}$	$\cdots$	$Y_{1,n_{MD}}$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
↑	*Bin i*	$Y_{i,1}$	$\cdots$	$Y_{i,j}$	$\cdots$	$Y_{i,n_{MD}}$
CD	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
↓	*Bin* n_{CD}	$Y_{n_{CD},1}$	$\cdots$	$Y_{n_{CD},j}$	$\cdots$	$Y_{n_{CD},n_{MD}}$

In fact, for control purposes the MD trend is often determined by simply taking the average of each scan and ignoring the MD variations observed within the scan [21]. Sometimes subsequent scans are averaged or temporal filtering is applied to determine the CD control profile (discussed further in Chapter 2). For product quality reporting the matrix $\mathbf{Y}$ is analyzed further in a procedure referred to as variance partitioning.

1.6 Variance partitioning analysis

The sheet property matrix $\mathbf{Y}$ can be broken down into the following form:

$$\mathbf{Y} = \bar{\mathbf{Y}}^{\mathbf{MD}} + \bar{\mathbf{Y}}^{\mathbf{CD}} + \mathbf{Y}^{\mathbf{R}}, \tag{1.2}$$

where $\bar{\mathbf{Y}}^{\mathbf{MD}}$, $\bar{\mathbf{Y}}^{\mathbf{CD}}$ and $\mathbf{Y}^{\mathrm{R}}$ represent data associated with MD variations, CD variations and residual variations, respectively [44]. More specifically, if the average of all data points across scan j is

$$\bar{Y}_j^{\mathrm{MD}} = \frac{1}{n_{\mathrm{CD}}} \sum_{i=1}^{n_{\mathrm{CD}}} Y_{i,j} \tag{1.3}$$

then

$$\bar{\mathbf{Y}}^{\mathbf{MD}} = \boldsymbol{q} \begin{bmatrix} \bar{Y}_1^{\mathrm{MD}} & \cdots & \bar{Y}_{n_{\mathrm{MD}}}^{\mathrm{MD}} \end{bmatrix}, \tag{1.4}$$

where $\boldsymbol{q}$ is a column vector of length n_{CD} filled with all ones. In other words, $\bar{\mathbf{Y}}^{\mathbf{MD}}$ represents a matrix where the data points in each column are the average of all the data points for the scan of that particular column. Values along the rows of a specific column are fixed as the matrix has a flat CD profile. Therefore, when $\mathbf{Y} - \bar{\mathbf{Y}}^{\mathbf{MD}}$ is computed it is equivalent to removing the scan average from every data point along that particular scan, i.e., removing the MD variations. Similarly, if the

average of all data points across bin i is

$$\bar{Y}_i^{\mathrm{CD}} = \frac{1}{n_{\mathrm{MD}}} \sum_{j=1}^{n_{\mathrm{MD}}} Y_{i,j} \tag{1.5}$$

then

$$\bar{\mathbf{Y}}^{\mathbf{CD}} = \begin{bmatrix} \bar{Y}_1^{\mathrm{CD}} \\ \vdots \\ \bar{Y}_{n_{\mathrm{CD}}}^{\mathrm{CD}} \end{bmatrix} \boldsymbol{p}, \tag{1.6}$$

where $\boldsymbol{p}$ is a row vector of length n_{MD} filled with all ones. Here, $\bar{\mathbf{Y}}^{\mathbf{CD}}$ represents a matrix with each data point in each row as the average of data points from all scans for the CD bin corresponding to that particular row. In other words the values along a particular row are fixed as the average over all scans and the matrix has a flat MD profile. Therefore, when $\mathbf{Y} - \bar{\mathbf{Y}}^{\mathbf{CD}}$ is computed it is equivalent to removing the CD bin average from every data point along that particular CD bin, i.e., removing the CD variations. Understanding $\bar{\mathbf{Y}}^{\mathbf{MD}}$ and $\bar{\mathbf{Y}}^{\mathbf{CD}}$ provides important context for interpreting the MD and CD profiles estimated in the simulated experiments. It is also worth noting that the matrices $\bar{\mathbf{Y}}^{\mathbf{MD}}$, $\bar{\mathbf{Y}}^{\mathbf{CD}}$ and $\mathbf{Y}^{\mathrm{R}}$ are useful for controller performance monitoring [44][57].

Variance partitioning is concerned with obtaining scalar values for describing the amount of total MD, CD and residual variation in a reel of paper. These statistics can be used to assess the quality of the paper and the performance of the paper machine relative to other machines. The average across a scan is defined in (1.3), the average across a bin is defined in (1.5) and the overall average is defined as

$$\bar{Y} = \frac{1}{n_{CD} \cdot n_{MD}} \sum_{i=1}^{n_{CD}} \sum_{j=1}^{n_{MD}} Y_{i,j}. \tag{1.7}$$

The total variance of the paper sheet data is then expressed as

$$\sigma_T^2 = \frac{1}{(n_{CD} \cdot n_{MD}) - 1} \sum_{i=1}^{n_{CD}} \sum_{j=1}^{n_{MD}} (Y_{i,j} - \bar{Y})^2. \tag{1.8}$$

Similarly the MD and CD variances are calculated as

$$\sigma_{MD}^2 = \frac{1}{n_{MD} - 1} \sum_{j=1}^{n_{MD}} (\bar{Y}_j^{MD} - \bar{Y})^2, \tag{1.9}$$

and

$$\sigma_{CD}^2 = \frac{1}{n_{CD}-1}\sum_{i=1}^{n_{CD}}(\bar{Y}_i^{CD} - \bar{Y})^2, \tag{1.10}$$

respectively. The residual variance is determined by taking the difference of the total variance and the combined MD and CD variances, i.e.,

$$\sigma_R^2 = \sigma_T^2 - \sigma_{MD}^2 - \sigma_{CD}^2. \tag{1.11}$$

Instead of reporting the actual variance, it is common in industry for these values to be reported as '2-Sigma' values (e.g., $2\sigma_{MD}$) [1].

As mentioned before, this treatment of scanner data ignores the geometry of the sampling trajectory and instead treats each scan as if it were taken on a stationary sheet. However, due to the shallow sampling angle from the MD axis, θ, over 600 m of sheet can pass in the MD during a single 10 m traverse in the CD. This MD variation leakage into the CD profile can significantly distort the scalar values determined by variance partitioning. The CD variance is likely to be over estimated due to the inclusion of intra-scan MD variations. Arguably more problematic however, is the distorting effect of these MD variations when they alias into the CD profile used by the CD control system. This distortion of the estimated CD profile due to the MD variations is referred to here as MD aliasing, and it represents one of the important challenges to overcome in MD-CD separation.

1.7 MD aliasing

While variance partitioning is primarily concerned with determining scalar values of sheet variations, MD/CD separation involves obtaining profiles of the separate variations that exist either along the MD or the CD. As with variance partitioning, the CD profiles can become distorted by MD variations leaking into the CD profile, i.e., MD aliasing. Here the effects of MD aliasing are displayed with a scanner signal that has been processed to use $\bar{Y}^{\mathrm{CD}} - \bar{Y}$ as the CD profile.

To show the effects of MD aliasing some simplifying assumptions are made. Moreover, the following assumptions are common in literature and some of them are applied in the experimental analysis as well.

Assumption 1. *The scanner traverses the moving sheet at a constant and equivalent speed in both directions.*

Assumption 2. *The scanner reverses exactly at the edges of the sheet.*

Assumption 3. *The CD profile is independent of time.*

As mentioned before, Assumption 1 is essentially true for long periods of time in many paper machine operations. However, more recent work has sparked interest in variable rate scanning which will be discussed in Chapter 2. In reality, the scanner does not follow such an ideal trajectory as described by Assumption 2. Instead, the scanner comes off the sheet briefly between scans and occasionally it will even stop for a short period while off the sheet in order to calibrate. As discussed earlier, while Assumption 3 is not technically correct, it is true for the vast majority of CD variability over short periods of time.

According to the assumptions described above, the time series of the scanner data forms an even, periodic function with respect to the CD variations (i.e., MD and residual variations notwithstanding). Each time the scanner completes two scans back and forth across the sheet represents one period ($2t_{\text{scan}}$) in this even periodic time series function. Therefore, if one were to take the Fourier transform of the scanner signal time series, the CD variations would contribute to coefficients at integer multiples of $(2t_{\text{scan}})^{-1}$. Consequently, the misinterpretation of MD variations in the CD profile (i.e., MD aliasing) is most apparent when the frequency of the MD variations occur at integer multiples of $(2t_{\text{scan}})^{-1}$ [27]. To show the effect of MD aliasing a simulated paper sheet with a flat CD profile and a single MD variation of frequency $(2t_{\text{scan}})^{-1}$ is sampled by a simulated scanning sensor and the CD profile $\bar{Y}^{\text{CD}} - \bar{Y}$ is shown in red in Figure 1.4. The reason the observed CD profile extends into negative values is because the average of the scanned samples is removed according to standard industrial processing. If the frequency of the MD variation is doubled such that it equals $(t_{\text{scan}})^{-1}$ the resulting observed CD profile would take the form shown in Figure 1.5.

Observing these examples, one might conclude that simply averaging over two consecutive scans would alleviate the bias in the CD profile. While this may be true for the current phase of the MD variations, consider if the MD frequencies were phase shifted by $\pi/2$. In Figure 1.6, four consecutive scans are simulated on a sheet with a flat CD profile and an MD variation at a frequency of $(2t_{\text{scan}})^{-1}$. By phase shifting the MD variations by $\pi/2$ the peaks and troughs of the MD variation consistently

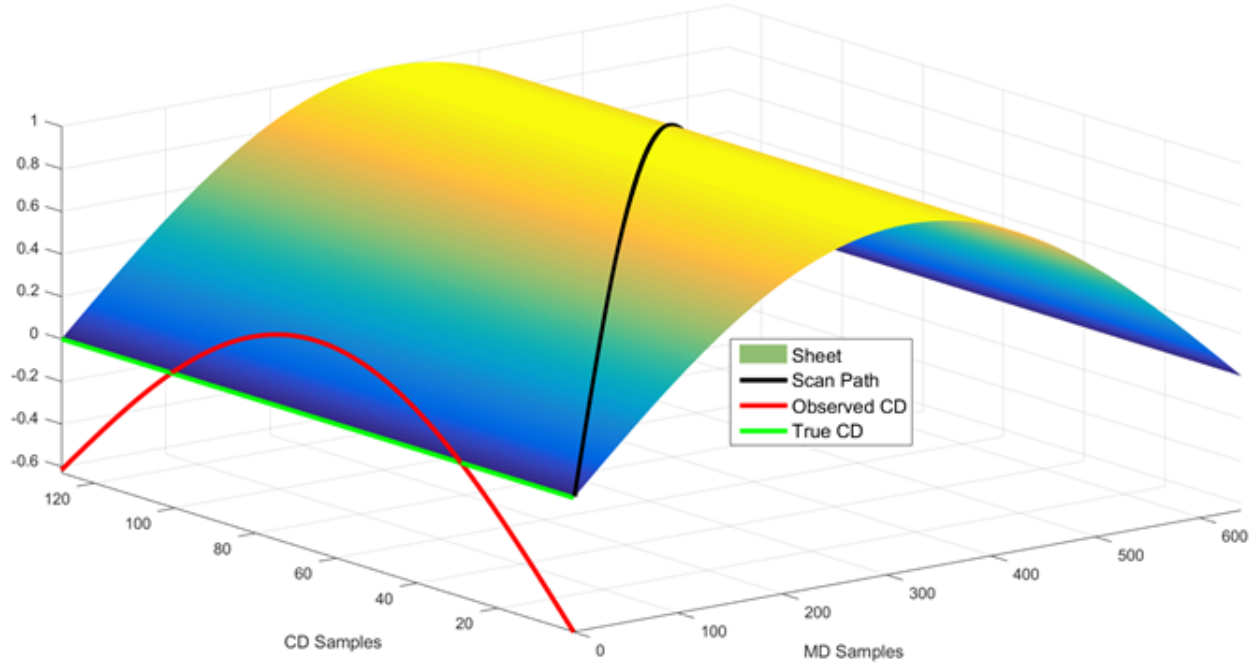

Figure 1.4: MD variation at frequency $1/(2t_{\text{scan}})$

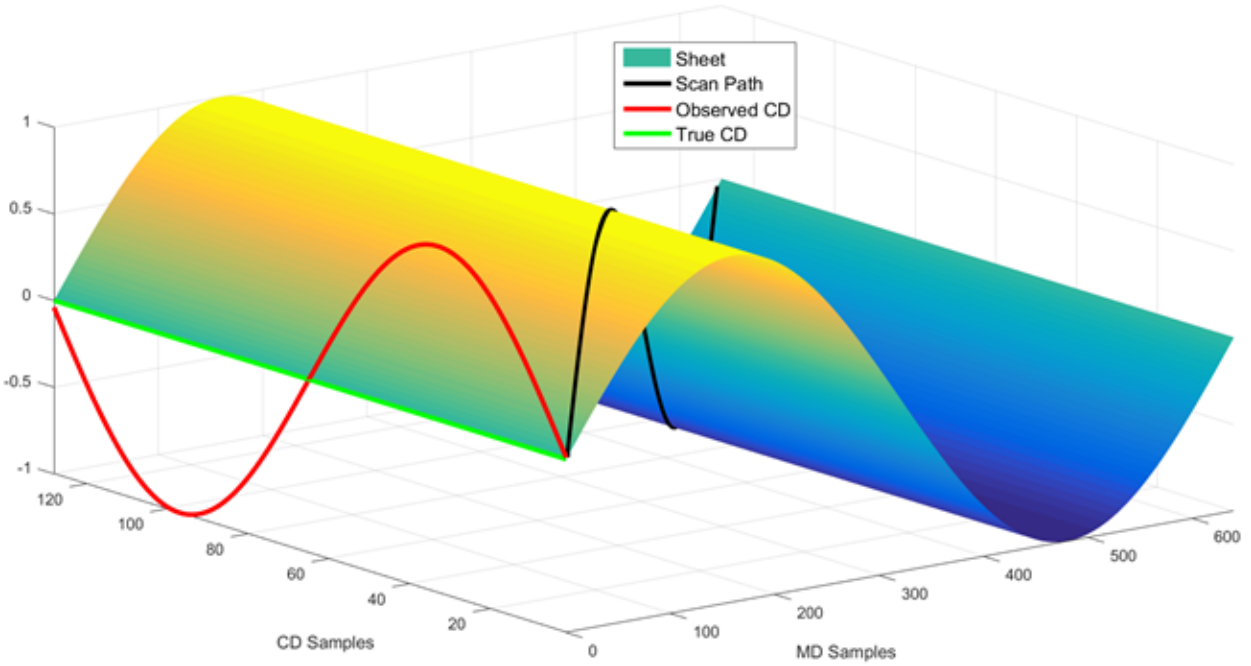

Figure 1.5: MD variation at frequency $1/(t_{\text{scan}})$

occur at the close and far side of the sheet, respectively. In this situation, no amount of averaging consecutive scans will eliminate or even alleviate the bias in the CD profile.

The aliasing of MD variations into the CD profile can have serious implications for CD control.

These perceived CD variations do not actually exist spatially but nevertheless the CD control system may attempt to correct for them which can further exacerbate the problem. By attempting to mitigate fictitious CD variations the CD control system could introduce new disturbances to the detriment of product quality. In some cases, averaging or temporal filtering can successfully reduce the effect of MD aliasing, however this usually comes at the detriment of profile resolution and controller response [37]. Until recently, the MD aliasing problem was scarcely addressed in the proposed MD-CD separation methods. In this thesis MD aliasing will be a primary focus of evaluating the effectiveness of various MD-CD separation methods.

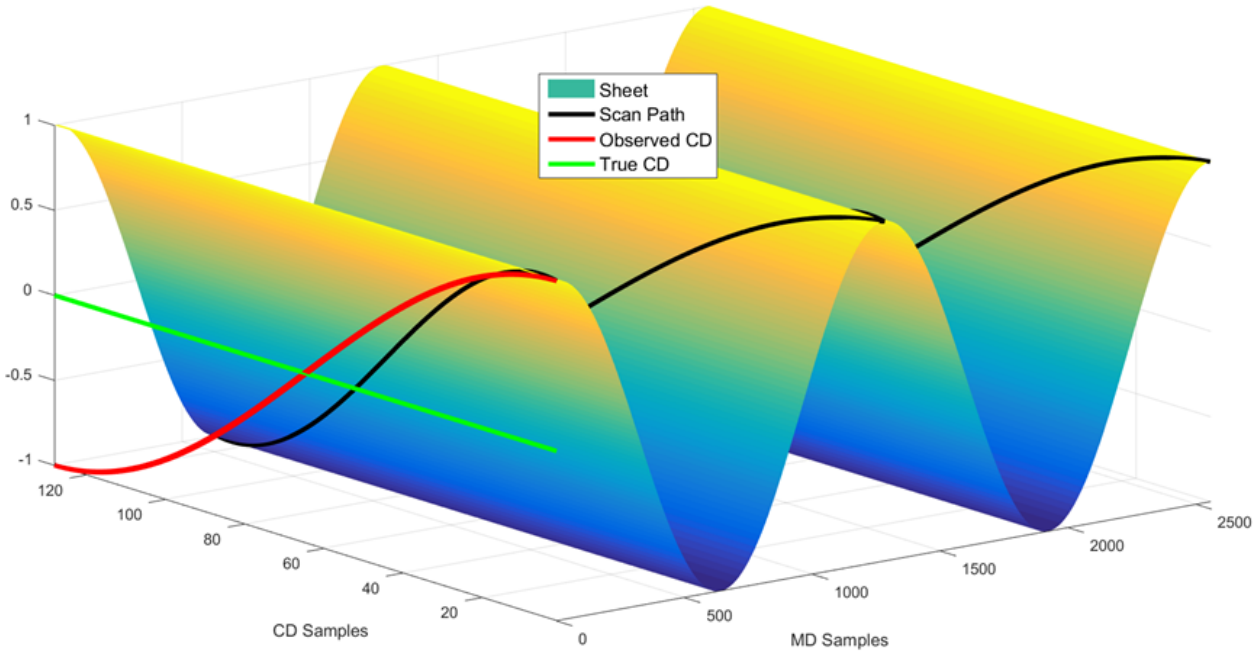

Figure 1.6: Four scans with MD variation at frequency $1/(2t_{\text{scan}})$

Chapter 2

MD-CD Separation

Over the previous thirty years a variety of different techniques have been proposed to separate MD and CD profiles from the scanning sensor signal. Unfortunately, to date none of the proposed methods have succeeded in gaining widespread industrial implementation. In this chapter, the previous developments in the MD-CD separation literature are reviewed and a few of the more promising methods are selected for further development and comparison. Next, the typical industrial method known as exponential filtering (EF) is presented with a detailed review of the various benefits and limitations. One of the more recently proposed methods, power spectral analysis (PSA), is considered with simulated examples provided. Finally, the chapter concludes with the introduction and development of the compressive sensing (CS) method for MD-CD separation.

2.1 Background

The majority of MD-CD separation literature involves the industrially supported assumptions that MD variations are faster and more dynamic than their CD counterparts, which are considered relatively time invariant. Furthermore, MD variations are assumed to have a uniform effect across the width of the sheet. These assumptions allow the control of MD and CD variations to be treated as independent problems which motivates the need to estimate separate MD and CD profiles.

The well established industrial method for MD-CD separation is referred to as exponential filtering. Essentially the signal processing described previously to obtain $\boldsymbol{Y}$ is carried out and then a temporal averaging is applied along each CD bin to determine the CD profile. Each scan is averaged into a single point to determine the MD trend. Among other problems, this method suffers from very limited MD resolution and slow upset detection in the CD profile because each new CD profile update includes a weighted average of previous CD profiles. Traditional exponential filtering ignores the CD position of the scanned data and therefore the sampling geometry itself. Up until

recently, disregarding the sampling geometry has been fairly common in the MD-CD separation literature.

In the 1980s and early 1990s developments in MD-CD separation focused primarily on modeling the process outputs such as basis weight and moisture. In 1986 Lindeborg proposes a nonlinear moisture model that uses scanning sensor measurements to estimate MD moisture variations at all times [40]. A dual Kalman filter approach is introduced by Chen in 1988 that combines both temporal and spatial models estimated from scanning sensor data [15]. Lindeborg's moisture model is leveraged by Dumont et al. in 1993 to develop an on-line estimation and identification of moisture content (EIMC) algorithm. The EIMC algorithm uses a recursive least squares exponential forgetting and resetting algorithm (EFRA) to estimate the CD profile, while the MD profile is estimated with a Kalman filter [21]. Following along this line of work, Wang et al. apply a similar approach for estimation of basis weight to develop the identification of basis weight profile and auto-regressive moving average (ARMA) MD model (IBPAM) algorithm. The IBPAM algorithm uses recursive least squares EFRA to estimate the CD profile while an extended Kalman filter (EKF) is used to estimate the MD basis weight profile [60]. The EIMC and IBPAM techniques are consolidated by Wang et al. and tested with both simulated and industrial data [59].

Although the industrial and simulated tests presented are encouraging, these model based MD-CD separation methods have not gained wide-spread industrial implementation. This might be due to the fact that model based methods rely on specification of parameters that are dependent on the particular machine and vary with operating conditions. Furthermore, to ensure convergence of the individual EKF and EFRA algorithms the tuning parameters must meet certain constraints. Even if said constraints are met, the convergence of the composite EKF and EFRA algorithm cannot be guaranteed [59]. Finally, the ability of the EIMC algorithm to provide unbiased estimates has not been proven [21].

In the mid-to-late 1990s wavelet compression and filtering techniques are introduced to approach the MD-CD separation problem. Wavelet basis functions such as Daubechies, Haar and Symmlet are tested and shown to have potentially useful compression ratios on industrial data [46]. Early efforts failed to recognize the geometry of the scanner but in 2004 Duncan and Wellstead develop a technique that considers the sampling geometry and applies the generalized sampling theorem to reconstruct MD and CD profiles. However, a non-causal filter is required that depends on the use

of both past and future data [22]. Aslani et al. propose an algorithm in 2009 that recognizes the sampling geometry with a causal approximation of the non-causal filter based on periodic sampling theory. Their algorithm combines this causal filter with recursive wavelet denoising to estimate MD and CD profiles [5]. Unfortunately this technique does not sufficiently address the issue of MD aliasing.

Explicit attention to the issue of MD aliasing has drawn increased attention in the MD-CD separation literature. The MD aliasing problem is discussed in detail by Gendron in 2005 [27]. Gendron's work is similar in many aspects to a 2008 thesis by Karimzadeh that also exploits the fact that the CD variations form an even periodic function in the time domain. By exploiting the fact that the discrete cosine transform (DCT) coefficients of the CD variations appear only at integer multiples of twice the scanning frequency, the CD profile can be estimated from the raw scanner data. Specifically, taking the DCT of the scanner signal and setting all of the coefficients that are not at integer multiples of $(2t_{\text{scan}})^{-1}$ to zero, the CD profile is estimated by taking the inverse DCT of the remaining frequency content. The MD profile is then estimated by subtracting the CD profile from the raw scanner data. This technique is compared to the wavelet and exponential filtering methods and is shown through simulations to have better performance [36]. One disconcerting fact about this DCT technique is that it does not distinguish between the CD variations and the MD variations that occur at integer multiples of $(2t_{\text{scan}})^{-1}$. Therefore, it does not offer a solution for handling the MD variations that have the most detrimental aliasing effect on the CD profile.

Recent attempts to address MD aliasing have explored the use of variable speed scanning or even random scanning speeds to minimize the impact of MD variations leaking into the CD profile. In 2009, Aslani et al. show that varying the scan speed can drastically alter the perception of MD and CD variations. Assuming the CD profile does not have any fast variations it is shown that the MD frequency components will shift if the scan speed shifts. Unfortunately in this work a systematic way to exploit the benefits of variable speed scanning is not proposed [4]. In 2011 Chen (with ABB) introduces the use of power spectral analysis (PSA) with a fast Fourier transform (FFT) to improve on traditional variance partitioning analysis (VPA) by extracting dominant variation patterns from MD, CD and residual profiles [14]. A formal procedure for using the PSA with two different scan speeds to separate MD and CD variations is patented by Chen and ABB in 2012 [13]. A variation of this technique will be one of the primary subjects of research in this thesis and thus will be covered

in more detail in Section 2.3. Work by Lang and Fu with Metso in 2013 also confirms the benefits of variable speed scanning and the potential for separating MD and CD components by observing a spectral frequency shift in the MD components with two different scan speeds [37].

In a 2011 thesis by Towfighi compressive sensing (CS) for MD-CD separation is introduced. Although initial results are promising, the practical considerations of implementation on a scanning sensor are not sufficiently considered [55]. In Section 2.4 the CS technique is developed further while taking into consideration the practical constraints of the scanning sensor framework. This new CS method is compared to the industrial standard (exponential filtering) and a variation of the PSA technique. A comparative analysis is conducted through simulations with these techniques to determine potential benefits and limitations associated with each.

2.2 Exponential filtering

Recall, exponential filtering (EF) is the longstanding industrial technique used to help estimate separate MD and CD profiles. The average of each scan is stored in a vector, $\bar{Y}^{\mathrm{MD}}$, and this is used as the input profile for the MD controller. The CD profile is calculated as a weighted average of current and past values across each CD bin by passing through an exponential filter, i.e.,

$$Y_{i,j}^{\mathrm{CD,f}} = \alpha(Y_{i,j} - \bar{Y}_j^{\mathrm{MD}}) + (1-\alpha)Y_{i,j-1}^{\mathrm{CD,f}}. \tag{2.1}$$

Here, $Y_{i,j}^{\mathrm{CD,f}}$ is the exponentially filtered data point for CD position i and MD position j. The CD position, $i = 1, \cdots, n_{\mathrm{CD}}$ where n_{CD} is the number of CD bins and the MD position, $j = 1, \cdots, n_{\mathrm{MD}}$ where n_{MD} is the number of scans in this particular context. After initial processing (as described in Section 1.5) the data matrix $\mathbf{Y}$ is created from individual output measurements $Y_{i,j}$. When data from scan j is collected, the average $\bar{Y}_j^{\mathrm{MD}}$ is determined. For EF the scan average is subtracted from the measured value at each CD bin position and multiplied by a filter factor, α. The exponential filter factor, $\alpha \in [0, 1]$, typically has a nominal value of 0.33 although in some arrangements it is dynamic according to the number of consecutive scans considered. The second term on the RHS of (2.1) shows $(1-\alpha)$ being multiplied by the filtered value from the same CD bin but the previous scan [2][5].

To summarize, EF removes the scan average from the values in the most recent scan, multiplies

these zero average bin values by 0.33 and then adds 0.67 multiplied by the filtered value from the same CD position but the previous scan. Essentially, EF works along each CD bin making the updated value a function of all of the previous values. This creates a form of temporal averaging across each CD bin where the most recent measurement is given a weight of α and the previous filtered measurement is given a weight of $(1-\alpha)$. The hope is that averaging over consecutive scans will remove high frequency MD variations from the CD profile and thus reduce the effect of MD aliasing. Employing the exponential filter requires the assumptions that MD variations are generally much faster than CD variations and that the CD profile is relatively stable.

To demonstrate the effects of EF, a sheet of paper is simulated with MD variations at 0.05 Hz and 0.10 Hz, while a scanner traverses at a speed such that $(2t_{\text{scan}})^{-1} = 0.0568$ Hz. This difference might seem trivial but it is important because the slight difference between the MD variations and integer multiples of $(2t_{\text{scan}})^{-1}$ allows the EF technique to reduce some of the aliasing over time. On the other hand, the more unique situation shown in Figure 1.6 where the MD variation is exactly $(2t_{\text{scan}})^{-1}$ renders EF entirely ineffective. Again, the sheet is simulated with no noise but this time two additive MD variations exist along with a realistic CD variation profile provided by Honeywell, our industrial sponsors. The CD profile is static, as the first four scans show in Figure 2.1. It is worth noting that the observed CD profile shown in red is actually the average of the four separate zero-mean scanned CD profiles. From Figure 2.1 it is apparent that averaging over the four scans has itself helped to dampen the MD variations in the perceived CD profile.

With fairly severe MD aliasing, as in Figure 2.1, EF may require many scans before it can effectively diminish the bias in the CD profile. To demonstrate this the exact same sheet and scanner setup in Figure 2.1 is simulated for a total of eighty scans. The effect of EF with $\alpha = 0.33$ is compared at the second, fortieth and eightieth scans as shown in Figure 2.2 with scans increasing from the top to bottom plots, respectively. The green line shows the true CD profile with no MD aliasing, the red line is the observed CD profile with the scan average removed and the blue line shows the CD profile after EF is conducted. As the number of scans increase the EF is able to reduce more of the MD aliasing because the CD profile is static and the MD variations are not exact integer multiples of twice the scan time. Figure 2.2 shows that as the scans increase the difference in aliasing between the observed CD profile (red) and the exponentially filtered CD profile (blue) decreases.

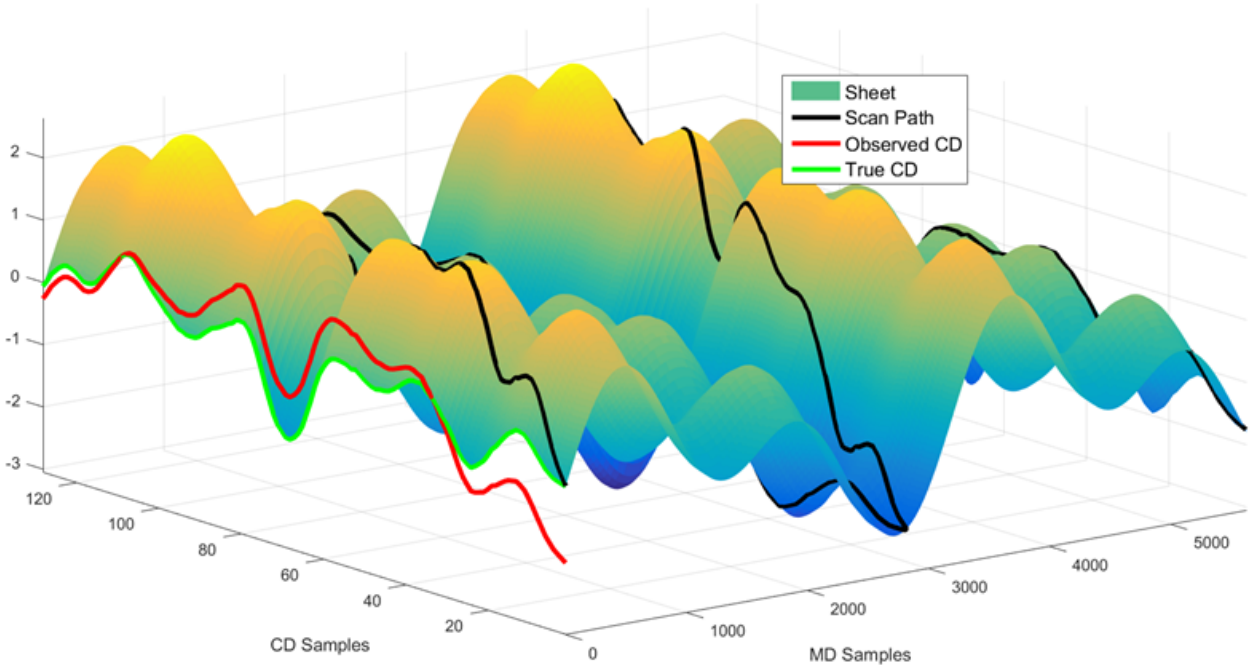

Figure 2.1: Four scans with MD frequencies near $1/(2t_{\text{scan}})$ and $1/(t_{\text{scan}})$

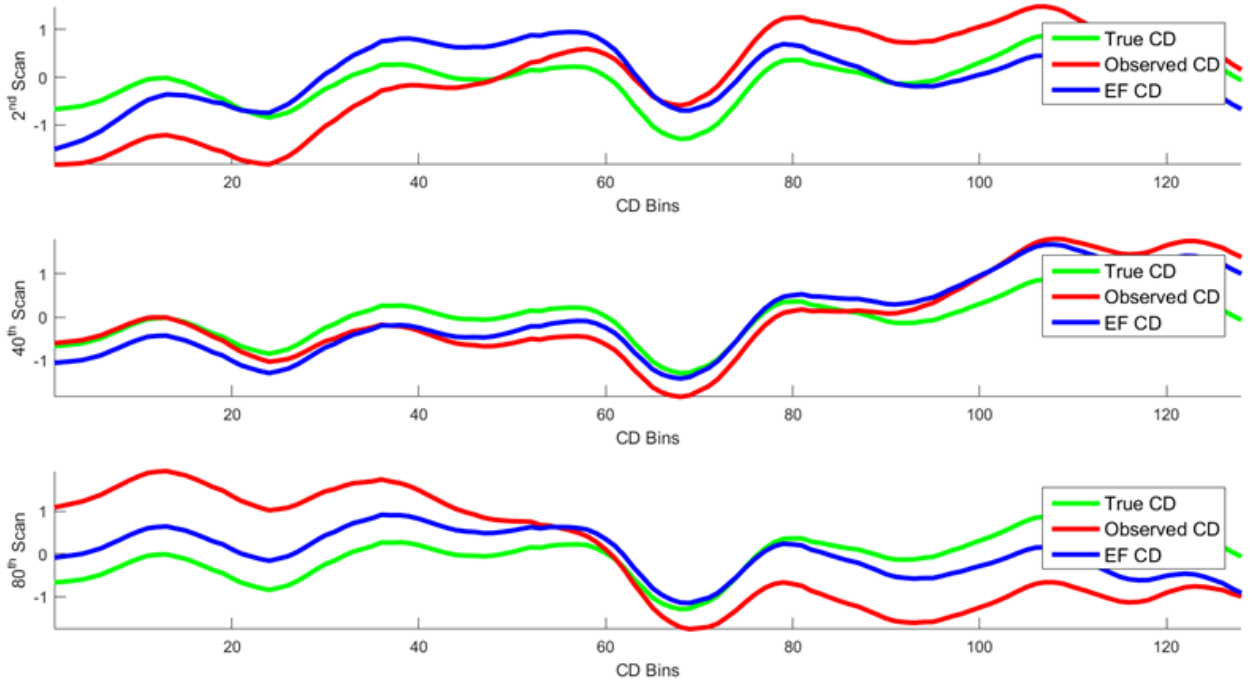

Figure 2.2: Exponential filtering after 2, 40 and 80 scans

The number of scans required to average out the high frequency MD variations can be quite significant, especially considering the necessary assumption of a static CD profile during this time. Since the most recent scan still receives α proportion of the weight, EF never fully dampens the

aliased MD variations. Another significant drawback of EF is that if the CD profile is not static, the averaging effect can mask the change for long periods of time, ultimately preventing detection of the upset and the necessary control response. Furthermore, the use of the scan average for the MD profile is not ideal as it provides unnecessarily low resolution which results in failure to acknowledge faster (potentially controllable) variations and ultimately slow upset detection by the controllers. Finally, exponential filtering does not address limitations set by the Nyquist sampling theorem which states that for accurate signal reconstruction the sampling frequency must be at least twice the highest frequency in the original band-limited signal (i.e., the Nyquist rate) [27] [51].

2.3 Power spectral analysis

The power spectral analysis (PSA) technique patented by Chen and ABB in 2012 involves comparing the Fourier transform of two consecutive scans at alternating scan speeds. By varying the scan speed and observing the shift in the support of the spectral contents, dominant MD frequencies can be distinguished from dominant CD frequencies [13]. Prior to discussing specifically how this technique is implemented for MD-CD separation, some background material on Fourier series expansion is provided.

2.3.1 Fourier analysis and orthogonal basis functions

In short, Fourier analysis involves representing a signal, $f(x)$, as a combination of sine and cosine waves at various frequencies and magnitudes. Here, some background information on this representation process is provided. Please note that attempts are made to use standard notations which may result in conflicts with other notation used throughout this thesis. Efforts are made to help clarify the use of notation, such as for the complex number, $i = \sqrt{-1}$, here as opposed to the CD bin index, i, from before. Furthermore, representation basis functions are traditionally denoted by ϕ (as seen in the references provided in this section) but to be consistent with the compressive sensing literature discussed later they will be denoted here as ψ instead. Notation is presented such that reader consideration of the current context should be sufficient to avoid confusion.

The concept of orthogonality is often described as two perpendicular lines that form a 90° angle with one-another. Extending this to higher dimensions, orthogonality can be understood as two

vectors having an inner product that is equal to zero. A classic set of orthogonal functions is $\sin(x)$ and $\cos(x)$ on the interval $[-\pi, \pi]$, i.e.,

$$\langle \sin, \cos \rangle = \int_{-\pi}^{\pi} \sin(x)\cos(x)dx = 0. \tag{2.2}$$

This orthogonal property is critical to the application of Fourier analysis and more generally to signal representation through basis functions. In Fourier analysis the orthogonal basis functions are $\sin(nx)$ and $\cos(nx)$, however other types of basis functions exist, such as wavelets, to represent signals. For generality, in what follows the set of basis functions will be denoted $\{\psi_i(x)\} = \Psi$.

A set of basis functions, Ψ, are orthogonal over the interval $[a, b]$ with respect to the weighting function $w(x)$ if for $i \neq j$

$$\int_a^b \psi_i(x)\psi_j(x)w(x)dx = 0. \tag{2.3}$$

However, in the case where $i = j$, the scalar product of the two vectors does not vanish, i.e.,

$$\int_a^b \psi_i^2(x)w(x)dx = N_i^2. \tag{2.4}$$

Now if the set of basis functions, Ψ, is normalized such that each $\psi_i(x)$ is multiplied by N_i^{-1} then the resulting set of basis functions will satisfy

$$\int_a^b \psi_i(x)\psi_j(x)w(x)dx = \delta_{ij} \equiv \begin{cases} 0, & \text{for } i \neq j \\ 1, & \text{for } i = j \end{cases} \tag{2.5}$$

where δ_{ij} is the Kronecker delta. A set of basis functions satisfying (2.5) are considered orthonormal basis functions, i.e., both orthogonal and unit normalized [3] [11].

Another important property to consider when using orthogonal functions to estimate a signal is completeness. A set of orthogonal functions is complete if a piece-wise continuous function, $f(x)$ for $x \in [a, b]$, can be represented by the series

$$f(x) = \sum_{n=0}^{\infty} a_n \psi_n(x) \tag{2.6}$$

such that the limit of the integrated squared error vanishes [3], i.e.,

$$\lim_{m\to\infty}\int_a^b \Big[f(x)-\sum_{n=0}^{m} a_n\psi_n(x)\Big]^2 w(x)dx = 0. \tag{2.7}$$

From (2.7) it is apparent that orthogonal completeness is an important property for ensuring the accuracy of the approximation of $f(x)$.

To demonstrate the importance of the orthonormal property for basis functions, suppose the function $f(x)$ is to be estimated as in (2.6). If the vector space is complex, as it is with Fourier expansion, some review of complex inner products is required. First, if z is a complex variable with both real and imaginary parts, i.e., $z = x + iy$, then the complex conjugate of z, denoted z^* is $z^* = x - iy$. For a complex matrix, $\boldsymbol{A}$, the complex conjugate, $\boldsymbol{A}^*$, is formed by taking the complex conjugate of each element [3]. For two vectors u and v in a complex vector space $\mathbb{C}^n$ their Euclidean inner product is given by [38]

$$\langle u, v\rangle = u_1 v_1^* + u_2 v_2^* + \cdots + u_n v_n^*. \tag{2.8}$$

Now, given (2.8) and the inner product defined in (2.5), consider taking the inner product of the signal we wish to estimate, $f(x)$, with the orthonormal basis function ψ_m over the interval $[a, b]$, i.e.,

$$\langle f, \psi_m\rangle = \int_a^b f(x)\psi_m^*(x)w(x)dx. \tag{2.9}$$

Plugging in the series expansion for $f(x)$ from (2.6), i.e.,

$$\langle f, \psi_m\rangle = \int_a^b \sum_{n=0}^{\infty} a_n\psi_n(x)\psi_m^*(x)w(x)dx, \tag{2.10}$$

and then extracting the summation of coefficients outside of the integral yields

$$\langle f, \psi_m\rangle = \sum_{n=0}^{\infty} a_n \int_a^b \psi_n(x)\psi_m^*(x)w(x)dx. \tag{2.11}$$

Since the set of basis functions Ψ are orthonormal as defined in (2.5), expression (2.11) simplifies to

$$\langle f, \psi_m\rangle = \sum_{n=0}^{\infty} a_n\delta_{nm} \tag{2.12}$$

and by the definition of the Kronecker delta

$$\langle f, \psi_m \rangle = a_m. \tag{2.13}$$

Therefore, by relating back to (2.6) the function to be estimated can be written directly in the orthogonal basis, i.e.,

$$f(x) = \sum_{n=0}^{\infty} \langle f, \psi_n \rangle \psi_n(x). \tag{2.14}$$

This property of orthogonal basis functions is critical for both spectral analysis and compressive sensing.

The expansion coefficients, a_m, given by

$$a_m = \int_a^b f(x)\psi_m^*(x)w(x)dx, \tag{2.15}$$

are commonly referred to as the generalized Fourier coefficients [3] [11]. As mentioned before the usual Fourier series uses the orthogonal functions $\sin(x)$ and $\cos(x)$ on the interval $[-\pi, \pi]$. The Fourier series expansion for a function $f(x)$ is given by

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx), \tag{2.16}$$

where, for $n = 0,\ 1,\ 2, \cdots, \infty$, we have

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x)\cos(nx)dx, \tag{2.17}$$

$$b_n = \frac{1}{\pi} \int_0^{2\pi} f(x)\sin(nx)dx. \tag{2.18}$$

Note that a_0 is covered by (2.17) which explains the coefficient of $1/2$ in (2.16).

The Fourier series expansion can also be extended to complex coefficients [3]. Recall, from Euler's formula, trigonometric functions can be related to the complex exponential function for any real number x as [18]

$$e^{ix} = \cos(x) + i\sin(x). \tag{2.19}$$

Therefore, the Fourier expansion from (2.16) can be expressed as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx} \tag{2.20}$$

where

$$c_n = \begin{cases} \frac{1}{2}(a_n + ib_n), & \text{for } n < 0 \\ \frac{1}{2}a_0, & \text{for } n = 0 \\ \frac{1}{2}(a_n - ib_n), & \text{for } n > 0. \end{cases} \tag{2.21}$$

If $f(x)$ is periodic in $[-L/2, L/2]$ the Fourier expansion can be expressed as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i(2\pi nx/L)}, \tag{2.22}$$

which forms the basis for the Fourier transform, $F(\omega)$.

The Fourier transform of a continuous band-limited signal $f(x)$ is given by

$$F(\omega) = \int_{-\infty}^{+\infty} f(x) e^{(-i\omega x)} dx. \tag{2.23}$$

A data set, $f(nT)$, that is sampled at uniform intervals of T, can be used to estimate $f(x)$ with the periodically extended spectrum $F^T(\omega)$, i.e.,

$$F^T(\omega) = \sum_{n=-\infty}^{+\infty} f(nT) e^{(-i\omega nT)}, \tag{2.24}$$

where

$$|F(\omega)| = 0 \quad \text{for} \quad |\omega| \geqslant \frac{1}{2}[2\pi/T], \tag{2.25}$$

$$F^T(\omega) = F(\omega) \quad \text{for} \quad |\omega| \leqslant \frac{1}{2}[2\pi/T]. \tag{2.26}$$

With $F^T(\omega)$ the estimate of $f(x)$ becomes

$$f(x) = \frac{1}{2\pi} \int_{-\pi/T}^{+\pi/T} F^T(\omega) e^{(+i\omega x)} d\omega, \tag{2.27}$$

but for practical purposes a finite data set of length N is used and the finite approximation is

performed with the discrete Fourier transform (DFT), i.e.,

$$F_D(\omega_k) = \sum_{n=0}^{N-1} f(nT)e^{(-i\omega_k nT)}, \tag{2.28}$$

where

$$\omega_k = \frac{2\pi}{NT}k, \quad \text{and } k = 0,\ 1, \cdots, \text{N-1}. \tag{2.29}$$

Other finite variations of the Fourier transform exist as well such as the finite Fourier transform and the discrete cosine transform [31].

In this work the fast Fourier transform (FFT), a more computationally efficient version of the DFT, is used for PSA. The FFT uses symmetries between the sine and cosine transforms to reduce the number of operations for a signal of N samples from $O(N^2)$ operations to $O(N\log_2 N)$ operations [17][18]. Here the notation $O(N)$ describes how the number of operations grows as a function of the signal length N. In the following section the use of the FFT is described with an example on a simulated paper sheet.

2.3.2 Power spectral analysis for MD-CD separation

The PSA technique requires the comparison of spectral components of two scans with varying scan speeds. Here the simulated scanner is set up such that the scan speed alternates every scan, resulting in different forward and reverse scanning trajectories. This allows each new scan to be compared to the most recent previous scan with the PSA method to perform MD-CD separation. To demonstrate the use of PSA for MD-CD separation, consider a simulated sheet of paper that is 8 m wide and moving at 10 m/s in the MD while a scanning sensor traverses the sheet at alternating speeds of 0.4762 m/s and 0.9091 m/s. The simulated sheet has one sinusoidal MD variation at a frequency of 0.10 Hz and one sinusoidal CD variation at a frequency of 0.5 m^{-1}. Furthermore, zero-mean normally distributed random noise is added to the sheet with a standard deviation of $\sigma_e = 0.1$. The simulated sheet has a total of n_{CD} =128 CD bins and is scanned twice, as shown in Figure 2.3. It is clear from Figure 2.3 that the slower scan speed of the first scan results in a longer MD trajectory and a smaller scanning angle θ. For each scan, the scanning angle is $\theta_1 = 2.7263°$ and $\theta_2 = 5.1944°$.

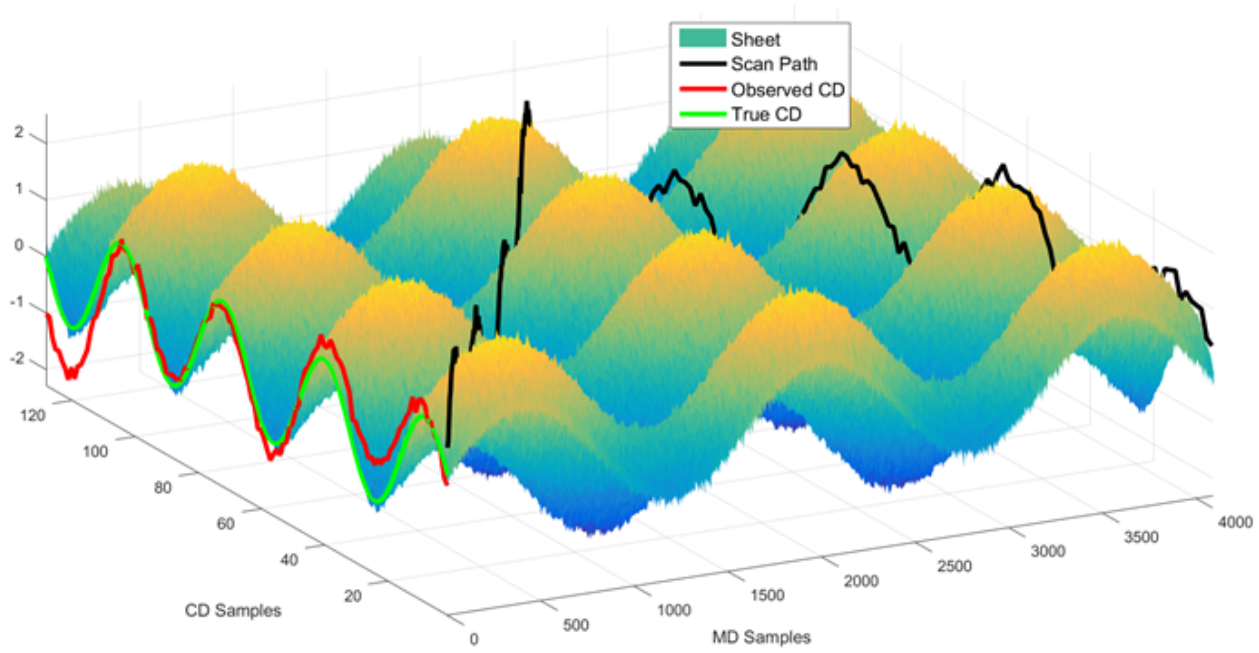

Figure 2.3: Two scan speeds with noisy sinusoidal MD and CD profiles

According to the Shannon-Nyquist theorem, in order to perfectly reconstruct a signal with maximum bandwidth ω_B one must take equispaced samples of that signal with a rate of at least $2\omega_B$, i.e., the Nyquist rate [20]. Likewise, the FFT of a scan of length n_{CD} collected with a constant sampling rate of $1/n_{CD}$ contains $n_{CD}/2$ frequency bins that range from frequencies as low as $1/n_{CD}$ ($\frac{\text{cycles}}{\text{sample}}$) to frequencies as high as 0.5 ($\frac{\text{cycles}}{\text{sample}}$). In other words, the FFT applies sinusoids with a period as large as the scanning length and as low as twice the sample spacing.

The output of the FFT contains both real and imaginary components, mirrored (for real data) at the center, where the second half of the output is the complex conjugate of the first half. For a real-valued signal, as we have here, the second half of the FFT output offers no additional information. The first value of the FFT is commonly referred to as the DC component and it represents the average value of the scanned signal multiplied by the number of samples (for the MATLAB implementation which does not normalize by $1/n_{CD}$ as others do). Furthermore, the bin at $n_{CD}/2$ where the FFT becomes mirrored represents the signal energy at the Nyquist frequency. To identify the dominant temporal and spatial frequency components the magnitudes of the complex FFT outputs are calculated.

The magnitude of the spectral content for both scans is shown in Figure 2.4. The top profile

shows the FFT magnitude plotted with respect to the spatial frequency whereas the bottom profile shows the FFT output magnitude plotted with respect to the temporal frequency. Notice in the top plot both the first and second scans have frequency content over the same spatial frequency range because both scans span the 8 m wide paper sheet. However, on the bottom plot the frequency content of the first scan does not cover the same temporal frequency range as the second scan because the first scan takes the same number of samples over a longer period of time. The second scan is effectively sampling at a faster rate so it is able to measure higher MD frequency content.

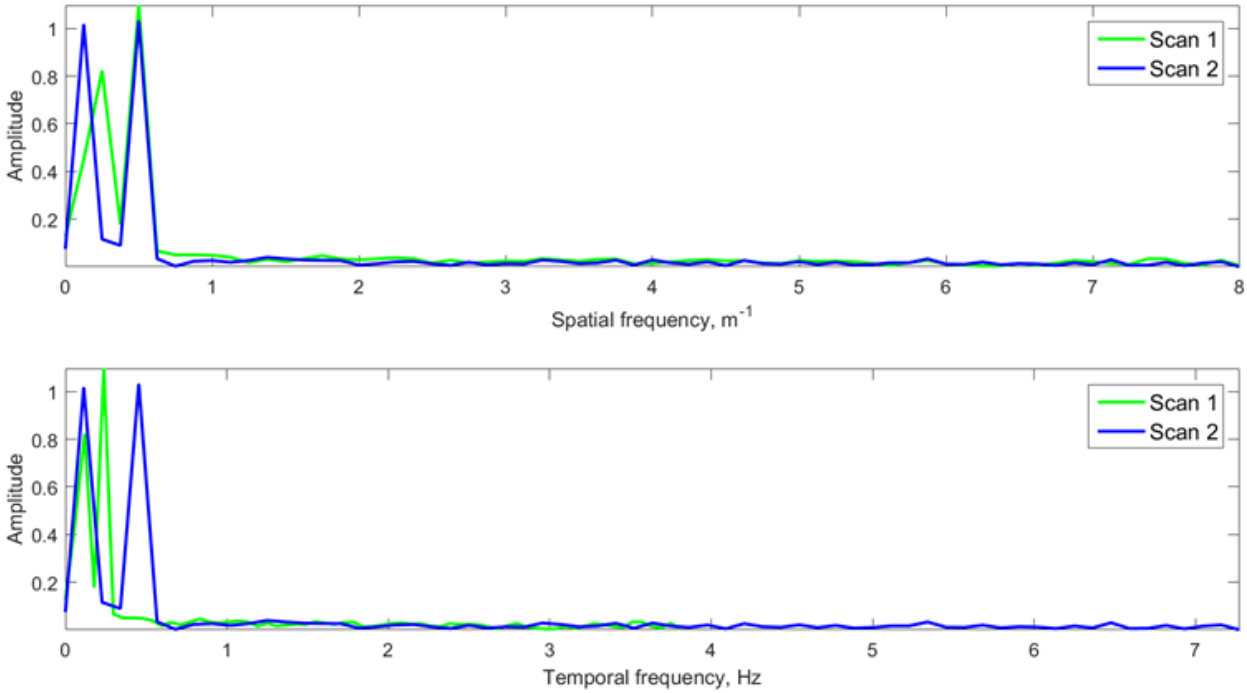

Figure 2.4: Spectral content of scans at different scan speeds

The benefit of alternating scan speeds is apparent from Figure 2.4 by observing the difference in alignment of the spectral content between the top and bottom plots. As expected, each scan has two dominant frequencies. In the top plot the higher dominant frequency of each scan aligns at the same spatial frequency, 0.5 m^{-1}, which is equivalent to the frequency of the CD variation from Figure 2.3. By accounting for the alternating scan speed in order to represent the temporal frequency content the peaks of the spectral content shift. In the bottom plot the lesser dominant frequency of each scan aligns at a temporal frequency of 0.1 Hz, which coincides with the frequency of the MD variation from Figure 2.3. This shift in the spectral content of each scan when presented with respect to both spatial and temporal frequency provides the basis for which MD-CD separation

can be conducted with PSA.

To determine the dominant frequency components a threshold is applied such that only the magnitudes that exceed the threshold are retained and the rest are set to zero. Of the remaining values, the ones that have intersecting spatial frequencies are stored as dominant CD frequencies and the ones that have intersecting temporal frequencies are stored as dominant MD frequencies. Next, the original FFT output from each scan is processed such that the values with a magnitude below another threshold are set to zero (to remove the noise). For the MD reconstruction the identified CD dominant frequencies are set to zero and for the CD reconstruction the identified MD dominant frequencies are set to zero. Finally, to perform the reconstruction the inverse of the FFT is determined for both the MD and CD frequency content. The resulting CD profiles are shown in Figure 2.5 and the reconstructed MD profiles are shown in Figure 2.6.

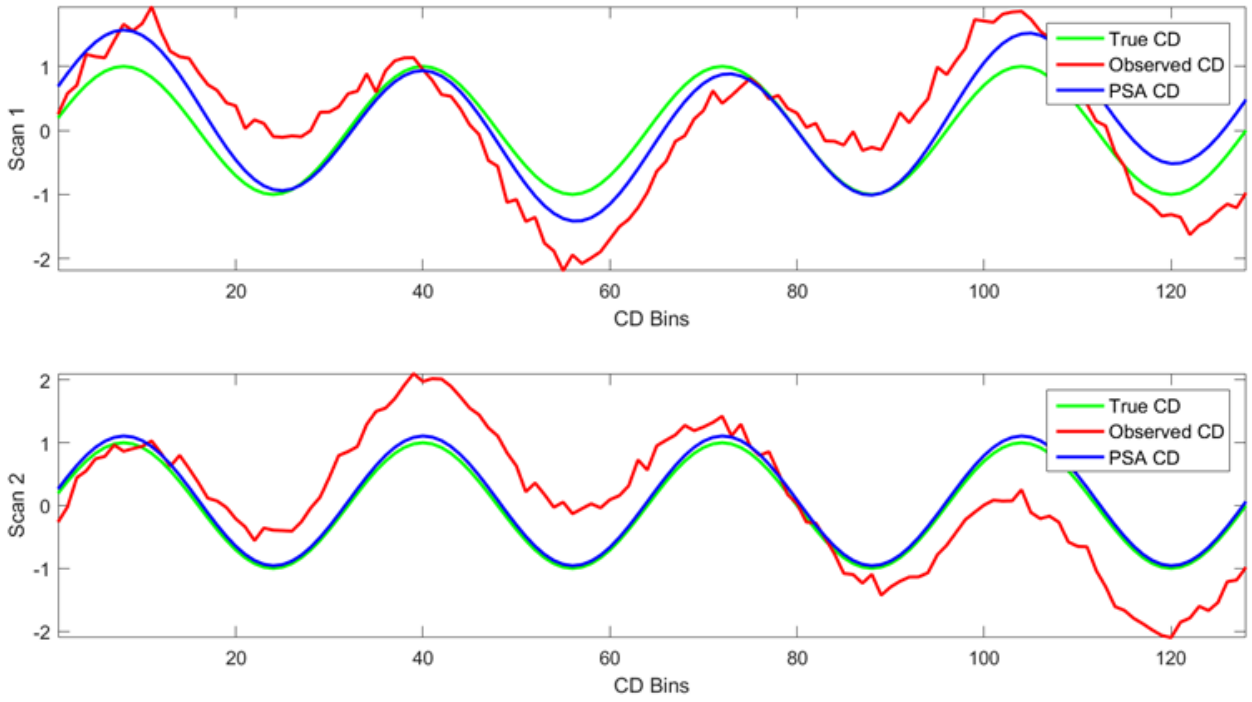

Figure 2.5: CD profiles after MD-CD separation with PSA

The estimated CD profiles of the first and second scans are shown in the top and bottom plots of Figure 2.5, respectively. The green signal represents the true CD variation without noise, the red profile represents the unprocessed observed scanner signal and the blue profile represents the PSA estimate. It is clear from Figure 2.5 that the PSA method helped to reduce the noise and more importantly the MD aliasing from the CD profile, particularly for the second scan which traverses

the CD at a faster rate. In Figure 2.6 the reconstructed MD profiles of the first and second scans are shown in the top and bottom plots, respectively. Here the red line shows the scan average which would likely be used for both control and quality reporting purposes in industrial practice. The PSA estimate (blue) of the first scan is very close to the true MD variation (green) but for the second scan (bottom plot) the PSA MD estimate is out of phase from the true MD variation. Since the dominant MD and CD spectral contents have a similar frequency it is possible that some MD and/or CD frequency content leaked into neighboring frequency bins and was lost during the processing. The reason the two plots look like they have different true MD variation frequencies is because the bottom plot is taken over a shorter time frame due to the faster scan rate.

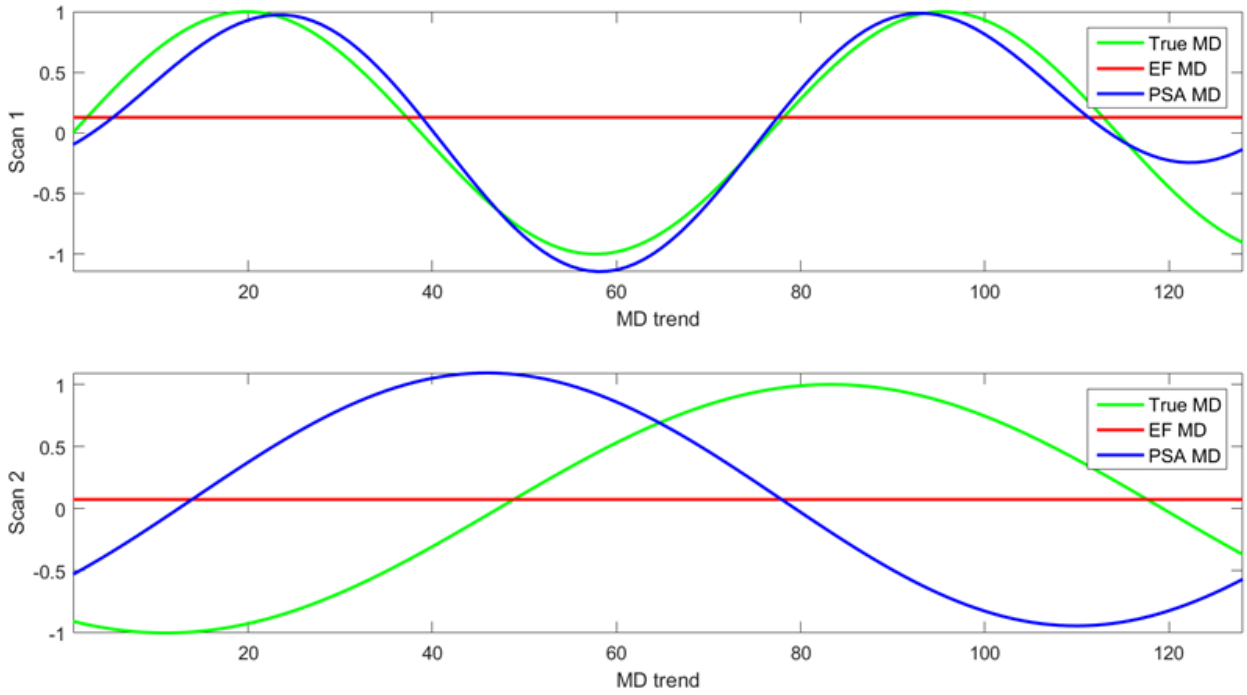

Figure 2.6: MD profiles after MD-CD separation with PSA

The PSA method for MD-CD separation is summarized in Algorithm 1. In this example the constant for the first threshold is set as $c_1 = 10$ and the constant for the second threshold is set as $c_2 = 3$. Unfortunately one of the drawbacks of the PSA technique is the selection of the thresholding constants to optimally remove noise and unwanted variations while maintaining all of the valuable information. It is also worth noting that the inverse FFT (IFFT) contains both real and imaginary components so for the purpose of this work only the real components are used to reconstruct the MD and CD profiles.

Algorithm 1 Power spectral analysis for MD-CD separation

Input: Two consecutive scans y_1, y_2 at alternating scans speeds.

Output: MD trends y_1^{MD}, y_2^{MD} and CD profiles y_1^{CD}, y_2^{CD}

for each scan **do**

Take $x = \mathrm{FFT}(y)$ and $|x|$

Set a threshold as $T_1 = c_1 \cdot \mathrm{mean}(|x|)$ and set values of $|x| \leq T$ to zero

Find support of remaining nonzero coefficients

Create separate spatial and temporal frequency supports

end for

for spatial and temporal frequency supports **do**

Find and store supports on consecutive scans that intersect

Set values of x below noise threshold ($T_2 = c_2 \cdot \mathrm{mean}(|x|)$) to zero

Set values of x at intersecting supports to zero

Reconstruct $y^{\mathrm{MD}} = \mathrm{IFFT}(x)$ if intersecting spatial supports were removed

Reconstruct $y^{\mathrm{CD}} = \mathrm{IFFT}(x)$ if intersecting temporal supports were removed

end for

2.4 Compressive sensing

The goal of compressive sensing (CS) is to exploit signal structure (specifically sparsity) in order to accurately reconstruct a signal with a much lower sampling rate than is dictated by the Shannon-Nyquist theorem. Compressive sensing has already been shown to successfully reconstruct paper profiles although implementing the structured sampling of the scanning sensor was not sufficiently addressed [55]. Random sampling is often necessary when applying compressive sensing but due to the nature of the sheet forming process it is not practically achievable. Instead, random sampling needs to be approximated. In this section the CS theory is first presented, followed by a summary of the previous work to apply CS for MD-CD separation. Finally, the proposed technique for applying CS to data from a scanning sensor for both MD and CD profile estimation is introduced.

2.4.1 Theory

Shannon's theorem states that for one to reconstruct a signal without error, the sampling rate must be at least twice the maximum frequency in the signal (i.e., the Nyquist rate). Although signals such as paper sheet variations may not have natural bandlimits, low-pass filters are often used to bandlimit these signals so Shannon's theorem still plays a role. Compressive sensing claims that as long as certain conditions pertaining to sparsity, incoherence and the restricted isometry property (RIP) are met, the signal can be perfectly reconstructed with high probability using less samples than is required by the Nyquist rate [10]. A detailed analysis of these conditions is beyond the scope of this thesis but they are presented briefly in what follows. The CS theory is developed with respect to the notation introduced in Section 1.5 for the scanner signal processing.

Recall from Section 1.5 the matrix $\mathbf{Y} \in \mathbb{R}^{n_B \times n_S}$ of scanner signal data with n_B rows where n_B is equal to the number of CD bins and n_S columns where n_S is equal to the number of MD scans. The vector form $\boldsymbol{y} = \text{vec}(\mathbf{Y})$ with $\boldsymbol{y} \in \mathbb{R}^m$ where $m = n_B \times n_S$ represents the values from each scan stacked upon one another. Please note the change in matrix and vector size notation from the previous sections which is due to the fact that the estimated paper sheet has larger matrix dimensions than the measured data matrix. Specifically, with CS the geometry of the scanning trajectory is retained and instead of stacking the data from a single scan at the same MD position the actual (or approximate) MD position of the scan data is represented. In other words, with CS the matrix of the estimated paper sheet is $\mathbf{F} \in \mathbb{R}^{n_{CD} \times n_{MD}}$ with n_{CD} rows where n_{CD} is the desired CD resolution of the estimated sheet which in this work is the number of CD data points measured across the sheet (i.e., here $n_{CD} = n_B$). The number of columns n_{MD} represents the desired MD resolution of the reconstruction which is defined here such that the scanning geometry is considered in the reconstruction ($n_{MD} \gg n_S$). The estimated sheet profile is originally reconstructed in vector form $\boldsymbol{f} = \text{vec}(\mathbf{F})$ with $\boldsymbol{f} \in \mathbb{R}^n$ where $n = n_{CD} \times n_{MD}$ and $n \gg m$.

Sparsity refers to the number of non-zero coefficients a signal has in a given representation (or sparsifying) basis, Ψ. Sparse signals have relatively few non-zero coefficients which allow them to be compressed and reconstructed accurately without storing the large amount of small coefficients. A signal, $\boldsymbol{f}$, is considered S-sparse in the representation basis Ψ if the coefficient sequence $\boldsymbol{x} \in \mathbb{R}^n$ has $S \ll n$ non-zero values and $\boldsymbol{f} = \Psi\boldsymbol{x}$ [20]. For example if the Fourier transform of a sinusoidal

signal is taken, the resulting spectrum will have large coefficients at the frequencies apparent in the signal while the rest of the spectrum will be noise around zero magnitude. If there are relatively few large coefficients then the Fourier transform is referred to as a sparsifying representation basis of the signal.

Consider the representation basis $\Psi \in \mathbb{R}^{n\times n}$ as an orthonormal basis (e.g., Fourier basis or wavelet basis) that provides a sparse representation of $\boldsymbol{f}$. Recall from equation (2.13) for an orthonormal basis the coefficients, x_i, of the original signal can be determined by the inner product, i.e.,

$$x_i = \langle \boldsymbol{f}, \psi_i \rangle \tag{2.30}$$

where

$$\boldsymbol{f} = \Psi \boldsymbol{x} = \sum_{i=1}^{n} x_i \psi_i. \tag{2.31}$$

Here $\boldsymbol{x}$ is the sparse coefficient sequence of $\boldsymbol{f}$ in the orthonormal representation basis Ψ [10].

In addition to sparsity, another condition of CS referred to as coherence is related to the projection matrix, $\Phi \in \mathbb{R}^{n\times n}$. The projection matrix describes how samples are taken from the domain of the original signal, i.e.,

$$\boldsymbol{y} = \mathbf{R}\Phi\boldsymbol{f} + \nu, \tag{2.32}$$

where $\mathbf{R} \in \mathbb{R}^{m\times n}$ is the measurement matrix that contains the sampling coordinates and ν represents additive noise that corrupts the measured data $\boldsymbol{y}$. The measurement matrix contains information regarding the geometry of the scanner trajectory. The measured signal can be expressed in terms of the representation basis coefficient sequence by inserting (2.30) into (2.32) to obtain

$$\boldsymbol{y} = \mathbf{R}\Phi\Psi\boldsymbol{x} + \nu = \mathbf{A}\boldsymbol{x} + \nu, \tag{2.33}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is referred to as the compressive sensing matrix.

Compressive sensing is concerned with minimizing the coherence, μ, of the sampling which is determined by the maximum correlation between Φ and Ψ as

$$\mu(\Phi, \Psi) = \sqrt{n} \max_{1\leq i,\, j\leq n} |\langle \phi_i, \psi_j \rangle|. \tag{2.34}$$

If Φ and Ψ are orthonormal bases then $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$. For CS it is desirable to select incoherent basis pairs for Φ and Ψ as this reduces the number of measurements required for accurate reconstruction. An example of a low coherence pair is the identity or spike basis as the projection matrix (Φ) and the Fourier basis as a representation matrix (Ψ) since spikes and sinusoids are maximally incoherent (i.e., $\mu = 1$) [9][10]. The importance of sparsity and low coherence is shown in the following formula where the probability of exact reconstruction exceeds $1 - \delta$ if

$$m \geq C \cdot \mu^2(\boldsymbol{\Phi}, \boldsymbol{\Psi}) \cdot S \cdot \log(n/\delta). \tag{2.35}$$

For maximally incoherent projection and representation bases the number of samples required for exact reconstruction is on the order of $S \cdot \log(n)$.

Finally, in order to guarantee efficient and robust reconstruction with CS it is desirable to satisfy the restricted isometry property (RIP). The CS matrix $\mathbf{A}$ is loosely said to obey the RIP if δ_S, the smallest number such that

$$(1 - \delta_S)||\boldsymbol{x}||_2^2 \leq ||\boldsymbol{A}\boldsymbol{x}||_2^2 \leq (1 + \delta_S)||\boldsymbol{x}||_2^2 \tag{2.36}$$

is not too close to one [10]. Although the RIP is NP-hard[2] to verify it is generally satisfied by ensuring a sufficient number of samples are taken for a given sparsity level, S, coherence, μ and randomness of sampling specified by $\mathbf{R}$ [26]. Selecting completely random coordinate sampling locations is usually used to satisfy the RIP and improve reconstruction for a given number of samples. However, in the paper machine application complete random sampling is not practically achievable due to the structured sampling scheme of the traversing scanning sensor.

The ℓ_2-norm applied in (2.36) is one example of the more general ℓ_p-norm. Since other types of ℓ_p-norms are used in what follows it is useful to define the ℓ_p-norm as [52]

$$||\boldsymbol{x}||_p = \Big(\sum_{i=1}^{n} |x_i|^p \Big)^{\frac{1}{p}}. \tag{2.37}$$

The objective in CS is to find the sparsest vector $\boldsymbol{x}$ that yields the measurement vector $\boldsymbol{y}$, which

[2] non-deterministic polynomial-time hard

involves solving the ℓ_0-norm, i.e.,

$$\begin{aligned} &\min ||\boldsymbol{x}||_0 \\ &\text{s.t.} \quad \mathbf{A}\boldsymbol{x} = \boldsymbol{y} \end{aligned} \tag{2.38}$$

where $\mathbf{A}$ is the CS matrix described earlier. Unfortunately (2.38) is NP-hard, therefore an equivalent ℓ_1 relaxation of the problem is solved instead. Furthermore, to account for the measurement noise, ν, the equality constraint in (2.38) is changed into an inequality constraint resulting in the following optimization

$$\begin{aligned} &\min ||\boldsymbol{x}||_1 \\ &\text{s.t.} \quad ||\mathbf{A}\boldsymbol{x} - \boldsymbol{y}||_2 \leq \sigma \end{aligned} \tag{2.39}$$

where σ is related to the noise level. In statistics this optimization is known as the Least Absolute Selection and Shrinkage Operator (LASSO) [26]. It is also know as the basis pursuit with denoising (BPDN) problem [30]. By duality, an equivalent version of (2.39) is given by

$$\begin{aligned} &\min ||\mathbf{A}\boldsymbol{x} - \boldsymbol{y}||_2 \\ &\text{s.t.} \quad ||\boldsymbol{x}||_1 \leq \tau \end{aligned} \tag{2.40}$$

where τ is a sparsity parameter. For future reference (2.39) is referred to as BPDN and (2.40) is referred to as LASSO. Here LASSO is a convex program that can be recast as a quadratic program and solved via interior point methods or proximal gradient methods for non-smooth convex optimization [26]. In this work the spectral projected gradient for ℓ_1 (SPGL1) minimization algorithm is used which obtains the solution to the BPDN problem by solving a sequence of LASSO problems. The SPGL1 method is chosen because it is reported as a fast method for large scale sparse reconstruction [30].

2.4.2 Previous work

As mentioned before, previous work by Towfighi in 2011 tested the concept of CS for sheet profile reconstruction on both simulated and industrial data. The study tested the Fourier representation basis in addition to the Haar, Symlet and Daubechies wavelet representation bases. For both the simulated and industrial trials the Fourier bases provided the lowest reconstruction error. This result was expected for the simulated trial because the simulated sheet was based on sinusoidal

variations. For industrial data the Fourier basis also makes intuitive sense if one considers the physical characteristics of the operations such as pressing and drying rollers that may result in roughly sinusoidal variations [55].

One problem with the simulated testing was that the dimensions of the simulated sheet fall woefully short of representing the true industrial problem. Furthermore, the relied upon sampling ratio used is unrealistically high and is not close to what is practically achievable given the scanning geometry. Some tests are conducted with complete random sampling which is not possible with a scanning sensor. Another test involved two random speed scanners but the traversing angle of these scanners is significantly greater than what is attainable in reality. Testing the CS technique on industrial data provided encouraging results although once again realistic structured sampling was not considered. For a high speed paper machine to satisfy the RIP condition it is proposed that either a sensor array or a large number of scanning sensors be used [55].

2.4.3 Compressive sensing for MD-CD separation

Based on Towfighi's results the Fourier basis is used as the representation basis for CS in this work. While a sensor array spanning the width of the sheet would be ideal, here the focus is on performing CS with a scanning sensor. Achieving the necessary number of samples to meet the RIP is not achievable in reality given the high speed of the paper machine and relatively slow speed of the scanning sensor. Furthermore, in industrial settings it is likely impractical to assume the sparsity of the moving paper sheet can be known in order to verify the RIP and determine the minimum number of samples required. Therefore, instead of attempting to guarantee perfect reconstruction according to the CS conditions, this work focuses on maximizing the obtained information to reconstruct the sheet profile through CS. The absence of CS theoretical guarantees should not be a deterrent especially considering CS is successfully used in many applications (e.g., tomography and spectroscopy) without such guarantees [7].

For industrial model predictive control (MPC) applications it is common for the scanner signal to be averaged into as many as 500 CD bins. Given a somewhat realistic sheet speed and scan speed, the shallow traversing angle $\theta \approx 2^\circ$ results in a very large amount of unmeasured data. For example, consider a scanner traversing the paper sheet at 2° from the MD averaging measurements over 500 CD bins per scan. Now consider the 'true' MD resolution of the paper sheet is limited to that which

is determined by the CD bins and the geometry of the scanning trajectory. That is to say, the 'true' sheet has a width of 500 equally spaced CD bins and the 2° traversing angle determines that approximately 14,500 MD data positions are covered by a single scan. This massive sheet size can be problematic with regards to simulating CS tests due to storage limitations in MATLAB. Recall, the CS matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ where m is the number of measured samples and n is the total size of the simulated sheet, i.e., $m = n_B \times n_S$ and $n = n_{CD} \times n_{MD}$. Simulating two scans with 500 CD bins and a 2° traversing angle yields $n = 14.5$ million and $m = 1000$ which results in the CS matrix $\mathbf{A}$ requiring roughly 116 GB of storage in MATLAB. Even with sparse matrix representation the necessary storage requirements far exceed the memory limitations of a standard personal computer.

To build on Towfighi's work it is desirable to provide a more realistic representation of the sampling ratio and scanning structure determined by the geometry of the scanning sensor. Furthermore, it is desirable to determine whether or not CS can provide effective MD-CD separation especially with respect to the effects of MD aliasing. Unfortunately due to the memory limitations available for this work, some compromise is necessary with regards to representing the true dimensions of the industrial problem. To balance these desires and limitations a smaller number of CD bins are used while maintaining a fairly accurate reflection of the sampling geometry and traversing angle (i.e., the 'true' simulated sheet still contains a large MD dimension). This is demonstrated by Figure 2.7 where four scans are performed on a simulated sheet with 64 CD bins that is 8 m wide and moving at 6 m/s. The simulated sheet contains single MD and CD sinusoidal variations with $\omega_{MD} \approx 1/(2t_{\text{scan}})$ and additive zero mean normally distributed random noise with standard deviation $\sigma_e = 0.1$. Although Figure 2.7 is not to scale, the scanner moves at 0.86 m/s creating a traversing angle of 8.1° that results in a total of $n_{MD} = 1792$ MD data points. The effect of MD aliasing is evident by comparing the observed CD profile averaged over four scans (in red) to the true CD profile (in green).

Even with the slow moving sheet and relatively simplistic MD and CD profiles the sampling ratio achieved is still only 0.2% of the total sheet. To improve this sampling ratio for CS a simple assumption is proposed here that allows for an increased sampling ratio.

Assumption 4. *The CD profile variations are slow enough to be assumed static over a short period (e.g., ≈10%) of a single scan.*

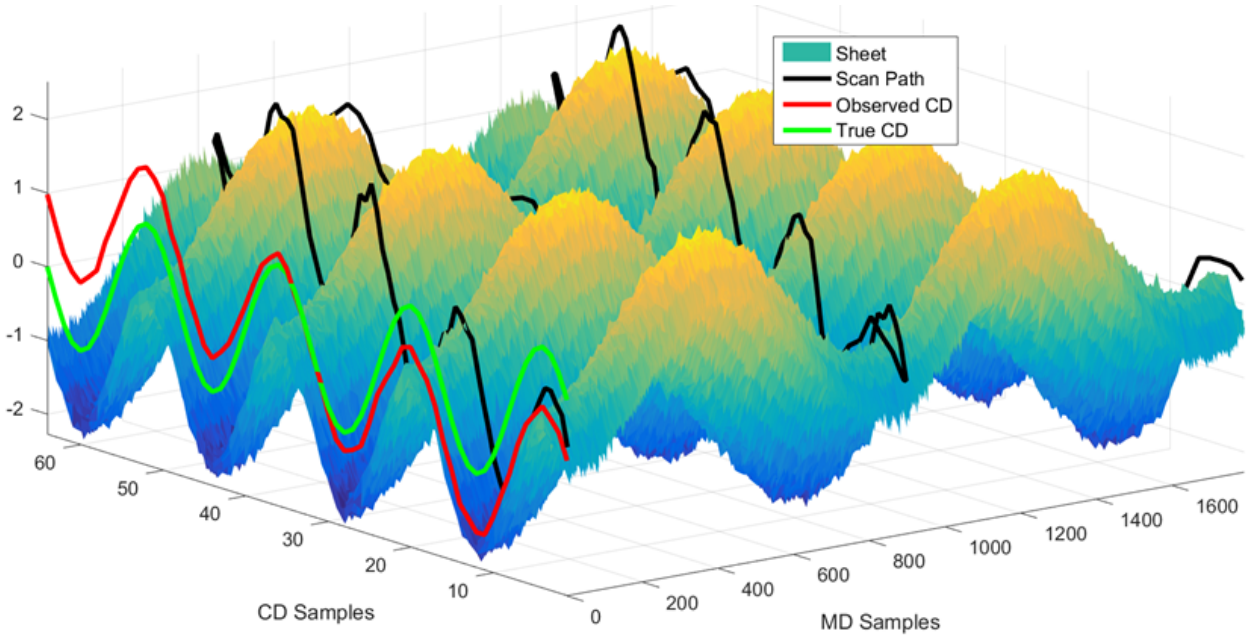

Figure 2.7: Simulated sheet with 64 CD bins and 4 scans

Note that this assumption is reasonably justifiable based on the characterization of paper sheet variations presented in Section 1.4. Furthermore, it is much less assertive than the assumption used by most MD-CD separation methods (including EF) which assume the CD profile is essentially independent of time. In practice this allows the scanner signal to represent a band of measurements across the paper sheet as opposed to just a line composed of single points. The band is created by using the measurement at a given CD bin and applying it to neighboring MD columns (along the same bin) that were missed by the scanner. No new information is taken from the true paper sheet but instead the near static behavior of the CD profile is used to better leverage the obtained measurements and increase the sampling ratio.

For example, from the first scan in Figure 2.7 the value measured at the second bin is $\mathbf{Y}(2,1) = \mathbf{F_0}(2,8) =$ -0.2668 where $\mathbf{Y}$ is the matrix of scanned data and $\mathbf{F_0}$ is the matrix of the full sheet. Assumption 4 is applied in this example to four neighboring values, i.e., it is assumed that $\mathbf{F_0}(2,6) = \mathbf{F_0}(2,7) = \mathbf{F_0}(2,8) = \mathbf{F_0}(2,9) = \mathbf{F_0}(2,10) =$ -0.2668. This is performed for each measurement from each scan providing a band of measurements across the sheet where the measurements from a particular CD bin of a particular scan all share the actual measured value. This assumption is applied to all of the scans and then the measurement vector $\boldsymbol{y}$ and CS matrix $\mathbf{A}$ are updated to reflect these new values. Next, CS is performed by solving the BPDN problem shown in (2.39) using

the SPGL1 technique described before. The sheet is estimated from the coefficients, i.e., $\boldsymbol{f} = \Psi \boldsymbol{x}$, and reshaped into the estimated sheet $\mathbf{F}$ shown in Figure 2.8 below. This simple demonstration precedes the more thorough comparative analysis of CS, PSA and EF which is described in the following chapter.

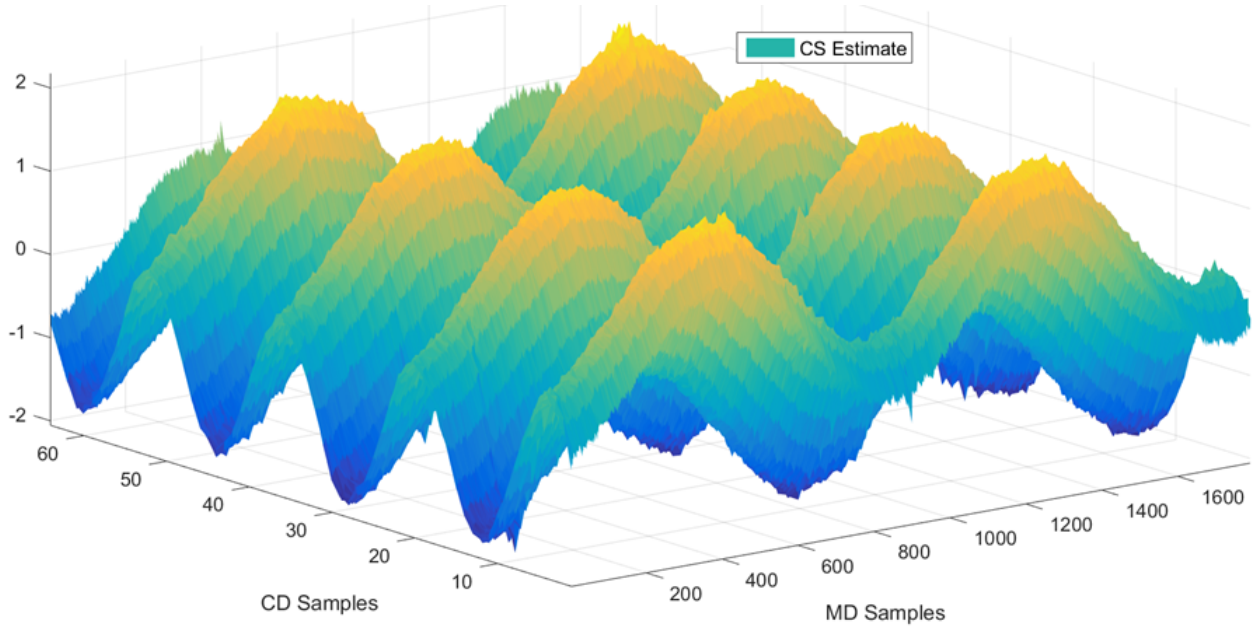

Figure 2.8: Simulated sheet reconstruction with compressive sensing

Chapter 3

Experimental Setup

In this chapter the experimental setup for comparing the exponential filtering (EF), power spectral analysis (PSA) and compressive sensing (CS) techniques is described. During industrial operation only the measurements obtained from the scanning sensor are typically available. Unfortunately since the true sheet profile is not measured there is no way to measure the reconstruction accuracy. To determine the true sheet, post-production laboratory tests can be performed but these tests are limited in scope and expensive to perform [5]. Consequentially, the analysis presented in this work is limited to simulated experiments.

To compensate for the lack of industrial data Honeywell has provided a CD profile that replicates a realistic industrial profile. This profile is shown in Figure 3.1 and a compressed version of it (i.e., less CD bins) is used in the experiments to represent the CD variation on the simulated paper sheet. Since the primary focus of this work is to study the ability of MD-CD separation methods to reduce

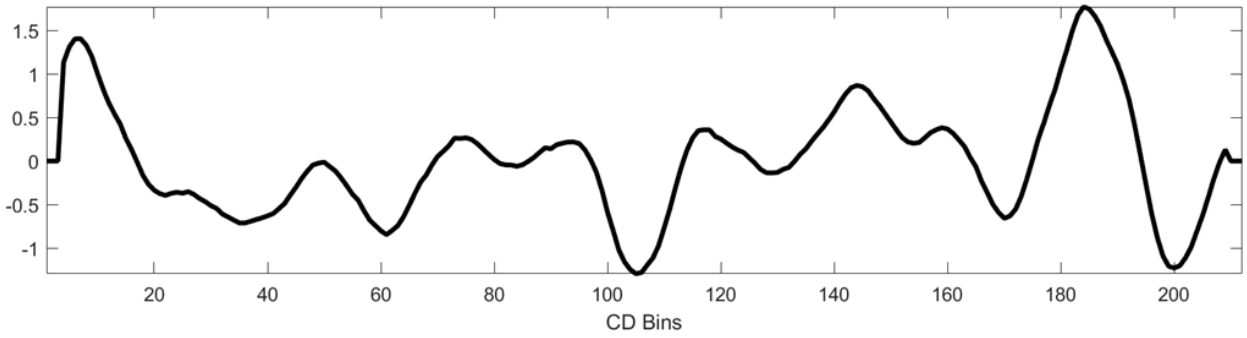

Figure 3.1: Realistic CD profile provided by Honeywell

MD aliasing, the sinusoidal MD variations are specifically chosen to have frequencies near integer multiples of $(2t_{\text{scan}})^{-1}$.

To generate the simulated sheet the number of CD bins is specified along with the number of scans, the scan speed, sheet speed and sheet width. From this information the traversing angle θ

is calculated as shown in equation (1.1). The MD length of the simulated sheet is based on the number of discrete MD data points necessary to approximate θ for the specified number of CD bins. With the necessary dimensions of the simulated sheet determined, the CD profile is repeated at each MD position and the MD profile is repeated at each CD bin to construct $\bar{\mathbf{F}}_{\mathbf{0}}^{\mathbf{CD}}$ and $\bar{\mathbf{F}}_{\mathbf{0}}^{\mathbf{MD}}$, respectively. As mentioned before, zero mean normally distributed random noise with standard deviation σ_e is used to generate the residual variation matrix $\mathbf{F}_0^{\mathrm{R}}$. Finally, the simulated sheet is constructed similar to equation (1.2), i.e,

$$\mathbf{F_0} = \bar{\mathbf{F}}_{\mathbf{0}}^{\mathbf{MD}} + \bar{\mathbf{F}}_{\mathbf{0}}^{\mathbf{CD}} + \mathbf{F}_0^{\mathrm{R}}. \tag{3.1}$$

The simulated scanner extracts measurements of $\mathbf{F_0}$ at each CD bin and MD positions determined by the traversing angle. When the scanner reaches the end of the sheet it stops and reverses ultimately taking two measurements at that CD bin such that each scan has a total n_B measurements.

Comparing the EF, PSA and CS methods for MD-CD separation is not straightforward since the favorable implementation of each technique is unique. For example, the EF method is typically more effective once a large number of scans have been processed whereas the PSA method compares only two consecutive scans. Furthermore, the PSA method requires alternating scan speeds whereas the CS method achieves a higher sampling ratio if the fastest scan speed is maintained. Instead of trying to compare the methods with the exact same implementation, each method will be implemented to allow for a favorable reconstruction given the same simulated sheet and scanning setup. This prevents improper implementation of an MD-CD separation technique to be a reason why one reconstruction is better than another. The advantages and disadvantages of the implementation of each technique are discussed in Chapter 5.

The sheet profile estimates obtained from each MD-CD separation method are compared based on how well they match the true MD and CD variation profiles. Specifically, the root mean square error (RMSE) is calculated for the estimated profiles relative to the true MD and CD profiles as follows:

$$\epsilon = \sqrt{\frac{1}{\mathrm{k}}\sum_{i=1}^{k}(p_i - \hat{p}_i)^2}, \tag{3.2}$$

where k is the number of samples in the profile, p_i is the true property value and $\hat{p}_i$ represents the

estimated value. The smaller the RMSE the better the estimated profile represents the true profile with an RMSE of 0 indicating a perfect fit. For EF the scan average is used as the MD profile so a zero order hold will be placed on the average value in order to extend to the true length of the sheet. Similarly for PSA the reconstructed MD profile is only the length of the scanned signal so interpolation is performed to generate a signal of desired length. In situations where a different number of scans are used for each method the specific reconstructed scans are typically from the end of the sheet but for clarity this is indicated along with the results.

Finally, with regards to the specific set of tests conducted to compare each method, the simulation parameters are provided with the results but the general experimental outline is as follows:

Trial 1: MD aliasing with slow moving sheet and 100 CD bins,

Trial 2: MD aliasing with fast moving sheet and 60 CD bins,

Trial 3: compressive sensing with two scanning sensors.

For each of the trials multiple tests are performed where the simulation configurations are altered slightly to study additional factors such as the effect of noise. Generally the major simulation parameters within a trial remain constant and the ones that change are stated explicitly before new results are presented. This allows for a more condensed representation of the experimental results without having to repeat similar setup parameters for each new test.

Chapter 4

Experimental Results

The results presented in this chapter consist of a series of tests that have been categorized into three major trials. The first trial compares the MD-CD separation methods with a fairly large CD profile, $n_{CD} = n_B = 100$, but a very slow sheet speed, $v_{\text{sheet}} = 1$ m/s. This unrealistic sheet speed is required due to memory limitations posed by constructing the CS matrix, $\mathbf{A}$. This trial is performed to showcase that the CS method is capable of handling a larger CD dimension if a large enough memory is provided. The second trial is conducted with a smaller CD dimension but a faster sheet speed. Increasing the sheet speed allows for the MD-CD separation problem to be studied with a fairly realistic scanning geometry. The focus of the first two trials is comparing the various MD-CD separation methods particularly with regards to alleviating the MD aliasing problem. Finally, the third trial provides some insight into the potential benefits of using compressive sensing with multiple scanning sensors.

4.1 Trial 1: Slow moving sheet with 100 CD bins

The simulation parameters for Trial 1 are summarized in Table 4.1. To build memory into the

Table 4.1: Simulation parameters for Trial 1

Parameter	Description	Value
w_{sheet}	sheet width (m)	8
v_{scan}	scan speed(s) (m/s)	0.33, 1
t_{scan}	scan time (s)	24, 8
v_{sheet}	sheet speed (m/s)	1
θ	traversing angle from MD (°)	18.4, 45
$n_{CD} = n_B$	number of CD bins	100
n_S^{EF}, n_S^{CS}	number of scans for EF and CS	20, 8
ω_{MD}	frequency of MD sinusoid(s) (Hz)	0.062, 0.12

exponential filter the sheet is simulated for 20 scans but the CS method is only performed on the

final 8 scans and the PSA technique only compares the final two scans. Since the PSA technique requires alternating scan speeds a separate scanning trajectory is performed over the final eight scans as the blue line in Figure 4.1 demonstrates. In order to compare the sheet profile estimates of the various MD-CD separation methods the reverse scan speed of the PSA method is equivalent to the regular scan speed from the CS and EF methods. This results in the final scan from all 3 methods sharing the same trajectory which allows the profile reconstructions to be compared directly. The sheet is constructed with the realistic industrial CD profile and two MD sinusoidal variations with frequencies near $(2t_{\text{scan}})^{-1}$ to create MD aliasing.

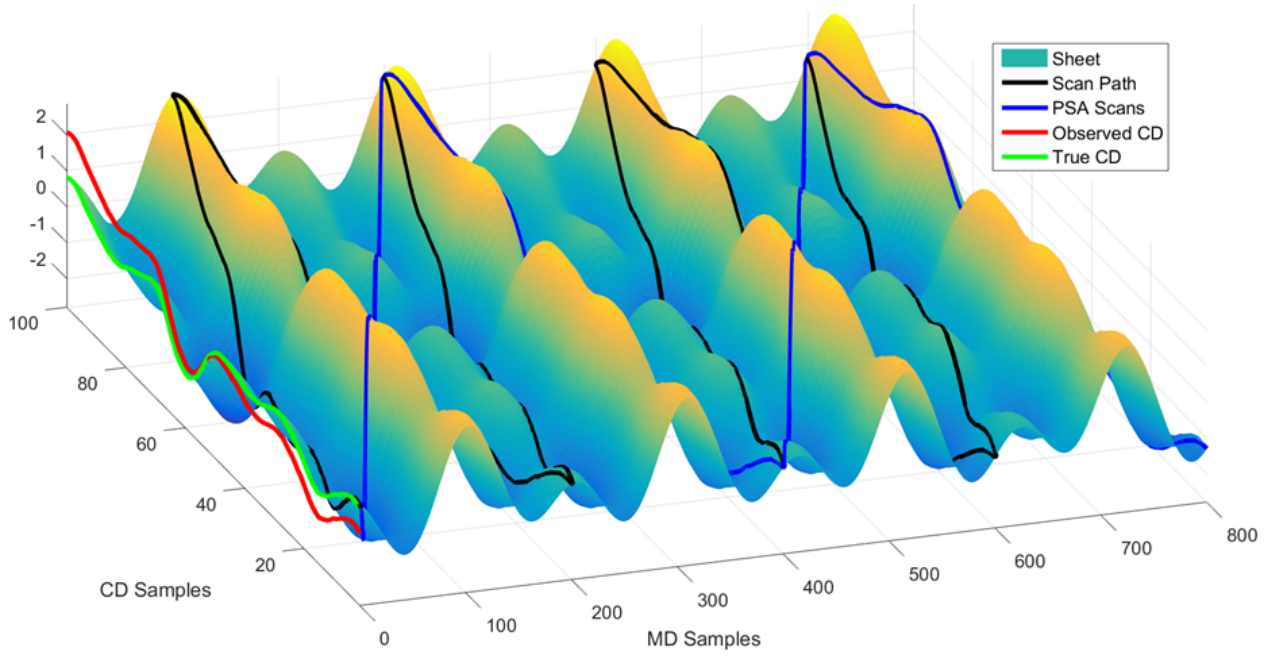

Figure 4.1: Simulated sheet for Trial 1

All 20 scans of the original sheet are passed through an exponential filter with a filter factor equal to $\alpha = 0.33$ to generate an EF estimate of the final scan. The PSA method is performed as described by Algorithm 1 with $c_1 = 4$ and $c_2 = 3$ to reconstruct the PSA estimate for the final scan. The 8 scans shown in Figure 4.1 are used to reconstruct the entire sheet via compressive sensing with the help of Assumption 4. After each of these MD-CD separation methods are performed the resulting CD profile and MD trends are presented together along with the true profile in Figure 4.2. The RMSE of each profile with respect to the true profile is presented in the legends which are located in different positions on the plot to provide the reader with the clearest possible view of the estimated profiles.

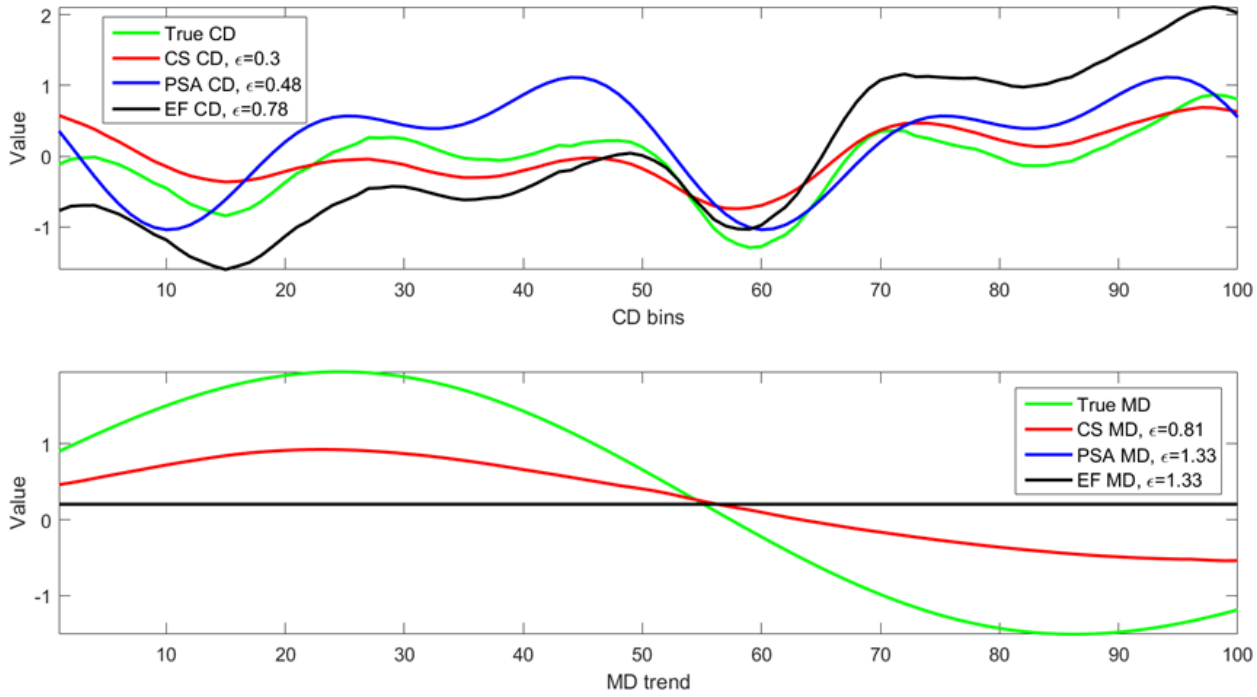

Figure 4.2: MD-CD separation results for Trial 1

Of the three methods the CS method provides the best reconstruction followed by the PSA technique and lastly the EF method. Although the PSA method seems to have removed part of the MD aliasing, the profile still over-estimates some of the CD variations which is potentially due to MD contributions that could not be distinguished from the CD frequencies. On the other hand, the CS estimate captures the essential shape of the true CD profile but the variation amplitude appears underestimated. The soft thresholding technique used by the SPGL1 method sparsifies the solution by removing small coefficients at the expense of reducing the amplitude of the remaining coefficients. This is also apparent in the MD profile where the CS method once again gives the best reconstruction and the PSA method is equivalent to the EF method meaning that only the scan average was recovered. The low frequency MD variation was likely either removed by the PSA threshold or leaked into the CD frequency bins.

To address the bias with the CS method caused by the soft thresholding the following scaling constant is proposed:

$$\beta = \frac{1}{2}\big((\max(\mathbf{Y}) - \max(\mathbf{F})) + |(\min(\mathbf{Y}) - \min(\mathbf{F})|\big). \tag{4.1}$$

Here the scalar β is multiplied by the CS reconstructed profile to increase the amplitude of the variations as shown in Figure 4.3. In terms of the RMSE the CS MD trend (bottom) is improved

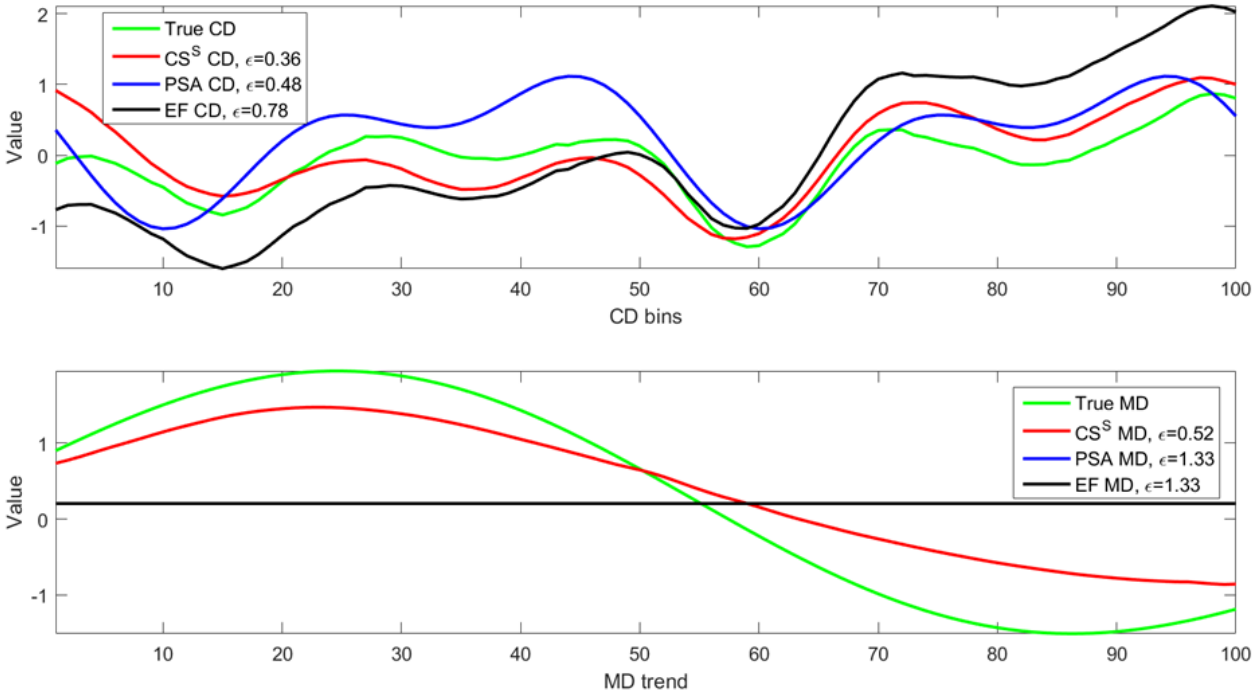

Figure 4.3: MD-CD separation after scaling for Trial 1

after scaling but the CD profile (top) has a slightly larger error. The scaling of the CS CD profile does however help visualize that the original CS reconstruction captures the general shape of the true CD profile, albeit with lower magnitude variations.

The second part of this trial involves adding zero mean normally distributed random noise with standard deviation $\sigma_e = 0.15$ to the simulated sheet. For reference the noisy simulated sheet is shown in Figure 4.4. Again, the EF uses all 20 scans, CS uses the final 8 scans and the PSA method compares only the final 2 scans. The results of the MD-CD separation for all three methods with noise are shown in Figure 4.5 where the CS method provides the most accurate CD profile and MD trend reconstructions. In this trial the inclusion of noise appears to only slightly degrade CD profile and MD trend estimates from each MD-CD separation method. Both the PSA and CS methods are effective in removing the noise with only a small number of scans and the EF method is able to reduce the effect of the noise by averaging over 20 separate scans.

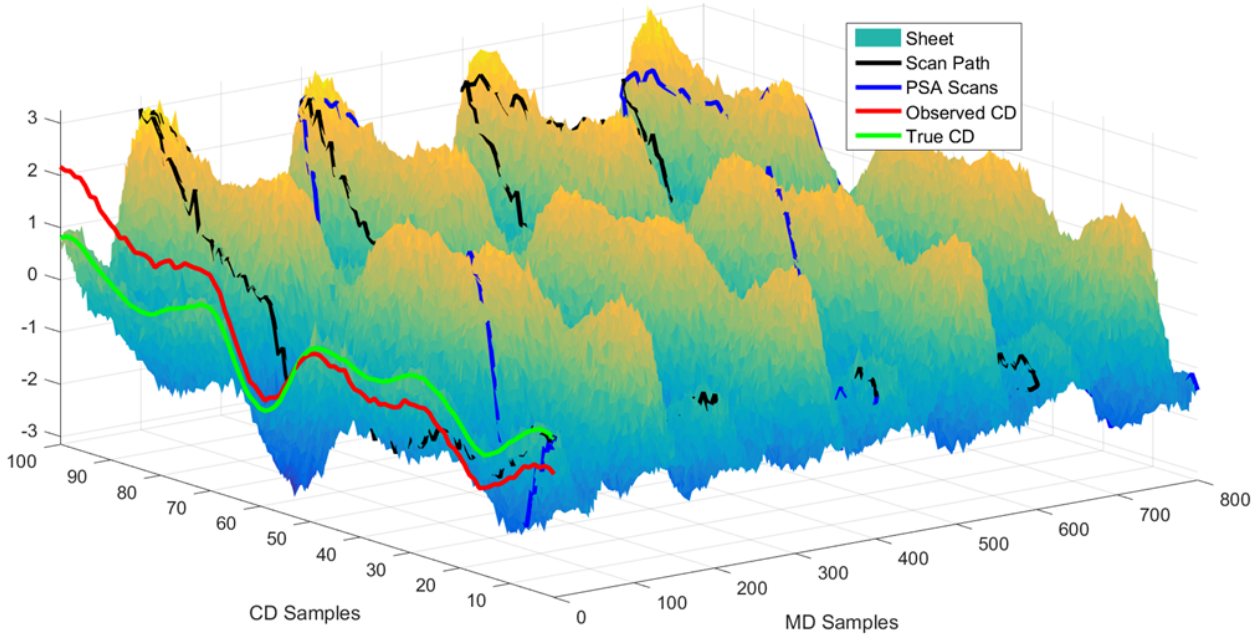

Figure 4.4: Simulated sheet with noise for Trial 1

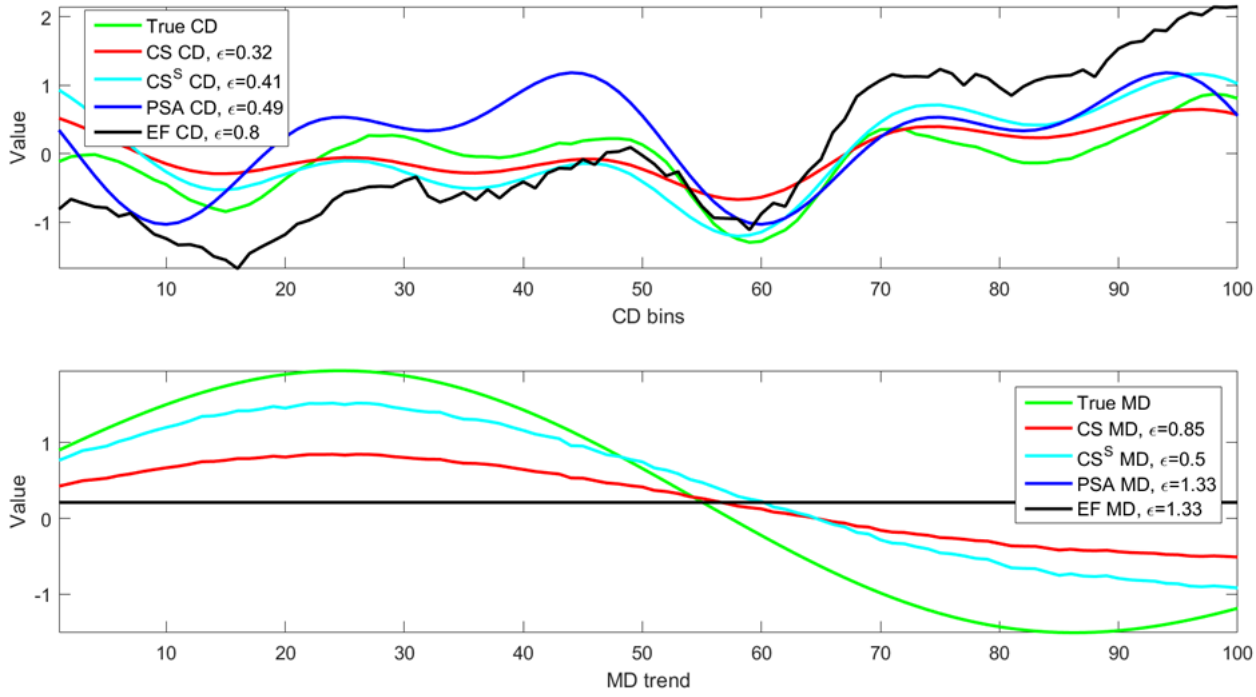

Figure 4.5: MD-CD separation results with noise for Trial 1

4.2 Trial 2: Fast moving sheet with 60 CD bins

The second trial involves reducing the number of CD bins to 60 in order to extend the sheet speed up to 8 m/s and study a more realistic traversing angle. The simulation parameters presented in

Table 4.2 show that the regular traversing angle for the CS and EF scans is 7.1° from the MD which is significantly closer to a realistic 2° value than the 45° trajectory presented in the first trial.

Table 4.2: Simulation parameters for Trial 2

Parameter	**Description**	**Value**
w_{sheet}	sheet width (m)	8
v_{scan}	scan speed(s) (m/s)	0.5, 1
t_{scan}	scan time (s)	16, 8
v_{sheet}	sheet speed (m/s)	8
θ	traversing angle from MD (°)	3.6, 7.1
$n_{CD} = n_B$	number of CD bins	60
n_S^{EF}, n_S^{CS}	number of scans for EF and CS	20, 6
ω_{MD}	frequency of MD sinusoid(s) (Hz)	0.062, 0.12

The simulated sheet shown in Figure 4.6 is very similar to the one from Trial 1 except with respect to the reduced CD dimension and the increase in the MD dimension from $n_{MD} = 800$ to $n_{MD} = 2880$. Furthermore, only 6 scans are used for the CS reconstruction in this trial in order to reduce the required size of the CS matrix. Once again the PSA scan speeds are designed such that the final scan follows the same trajectory as the CS scans in order to allow for a direct comparison of the profile estimates from the final scan. The MD-CD separation methods are performed as before with the only difference being the thresholding coefficients of the PSA method in Algorithm 1 which are decreased to $c_1 = 2$ and $c_2 = 1.5$. The reconstructed CD profiles are compared in the top part of Figure 4.7 and the estimated MD trends are compared in the lower plot with CS^{S} indicating the scaled CS reconstruction.

Comparing the RMSE of each reconstruction, displayed in the legends of Figure 4.7, shows that the scaled CS method provides the most accurate reconstruction for both the CD profile (top) and MD trend (bottom). For the CD profile the next best estimate is the regular CS method followed by the PSA technique and finally the EF method which suffers from severe MD aliasing. The MD trend estimates follow the same general trend except now the PSA MD estimate is worse than the EF estimate which just uses the scan average. It is clear that the PSA MD trend contains a significant amount of CD variation, likely due to frequency bin overlap caused by the relatively small number of CD bins in combination with the lower threshold values. In fact, the PSA MD trend estimate closely represents the CD profile estimate which strongly suggests that the MD variations were not sufficiently distinguished in the spectrum of the final scan.

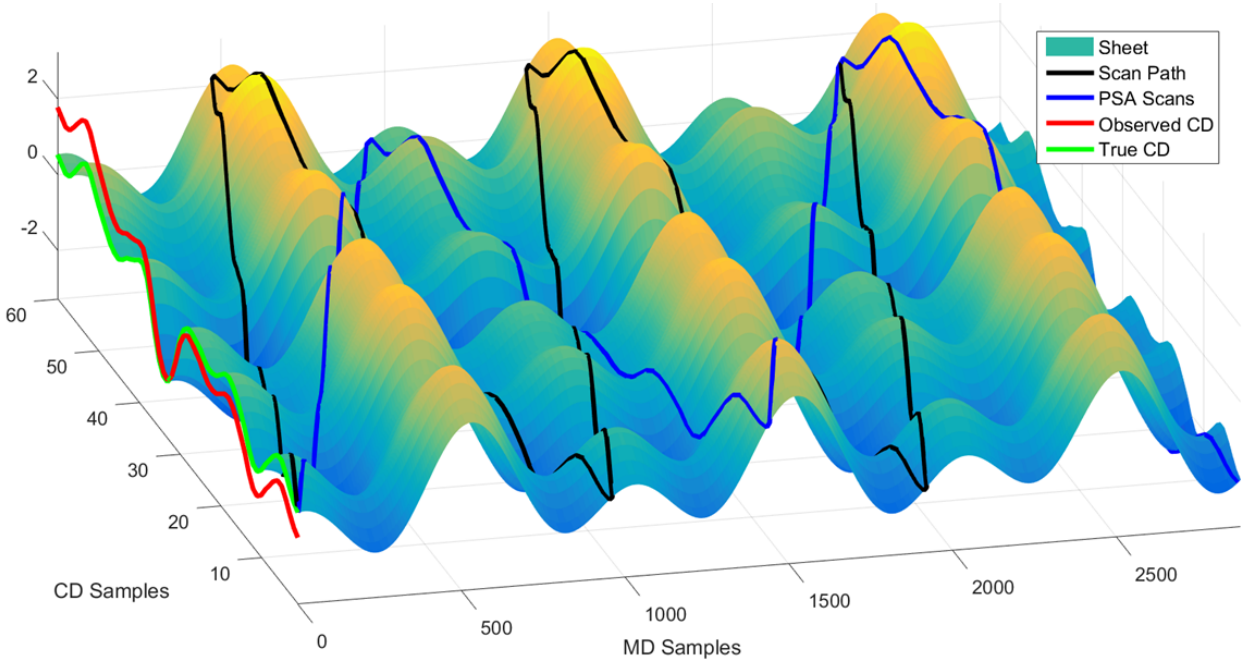

Figure 4.6: Simulated sheet for Trial 2

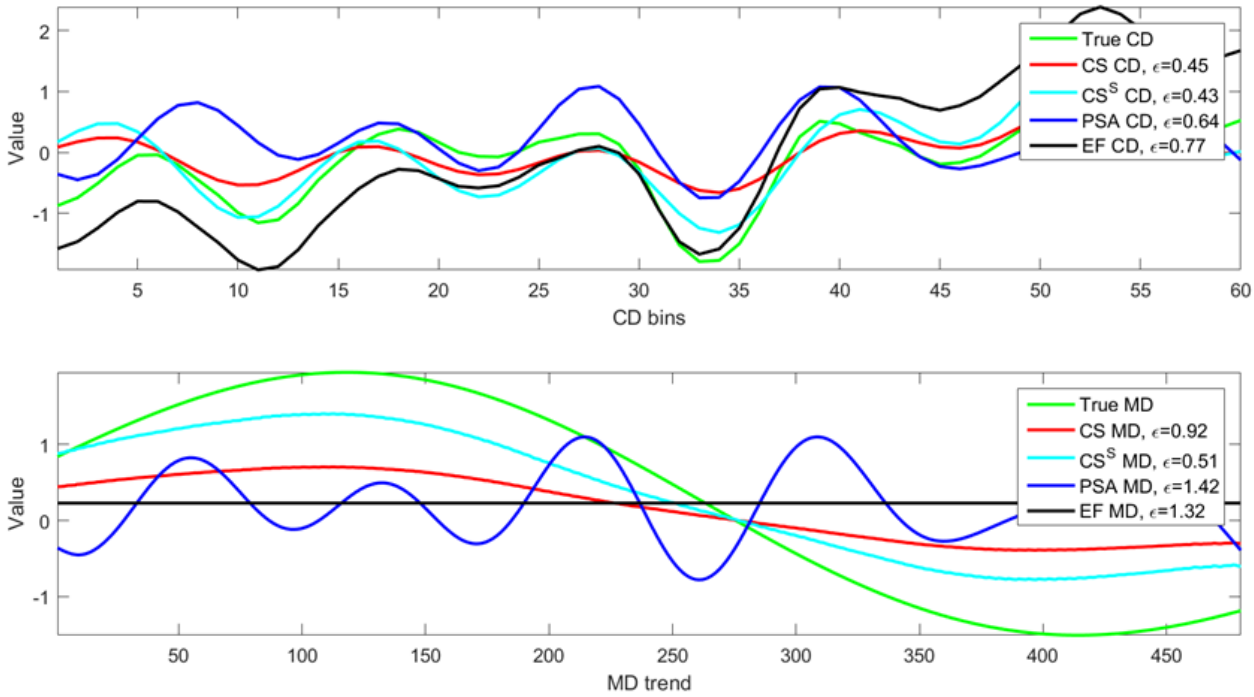

Figure 4.7: MD-CD separation results for Trial 2

Once again zero mean normally distributed random noise with standard deviation $\sigma_e = 0.15$ is added to the sheet. For concise presentation and less redundancy the noisy simulated sheet is not presented here but Figure 4.4 can be referred to because the same level of noise was added.

The effect of noise is less apparent in the EF CD profile when there are less CD bins, as is clear by comparing Figure 4.5 and Figure 4.8. The MD-CD separation results with additive noise are very similar to those without noise, i.e., the scaled CS estimate still provides the most accurate reconstruction for both the CD profile and MD trend. The only significant difference is that a high frequency MD variation occurs in the CS estimates of the MD trend as a result of the noise.

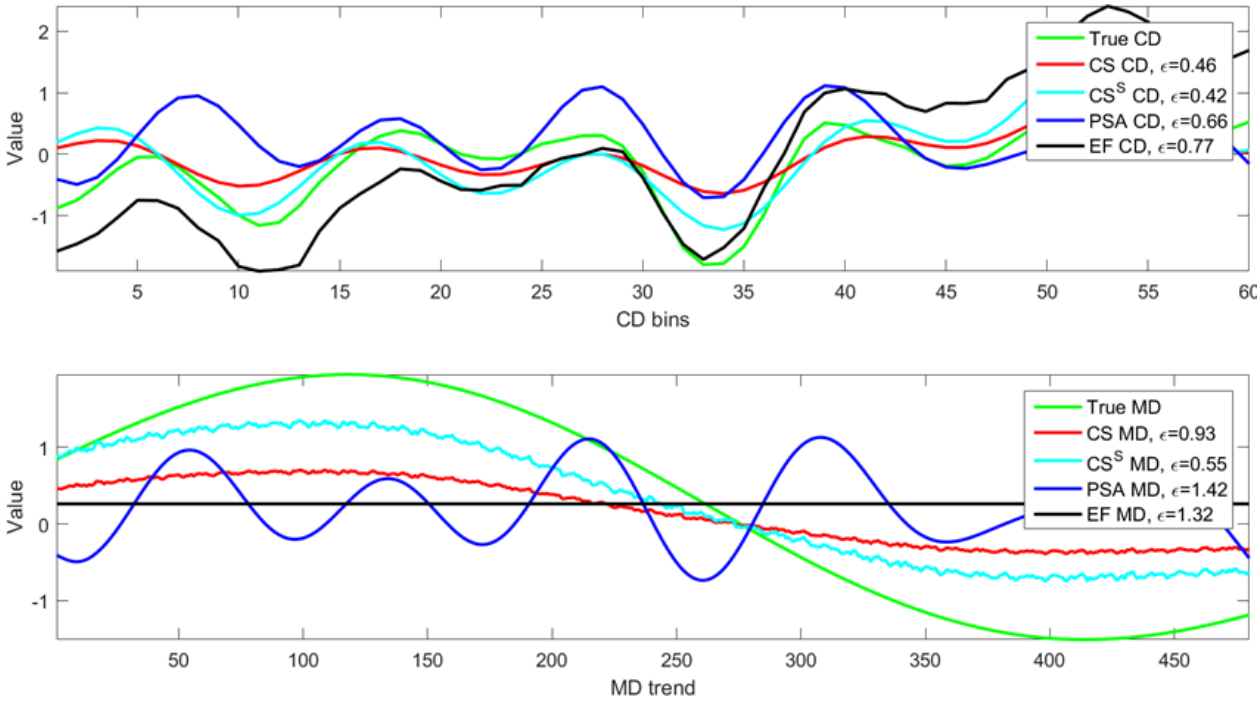

Figure 4.8: MD-CD separation results with noise for Trial 2

4.3 Trial 3: Compressive sensing with two scanning sensors

One of the major differences between CS relative to the EF and PSA MD-CD separation methods is that CS actually reconstructs the entire sheet of paper for the specified resolution. In simulations the desired resolution can be set as the resolution of the true simulated sheet but in reality the true paper sheet does not have a specific resolution. The desired resolution of the reconstruction must be specified based on the control requirements and the available measurements. This is often accomplished by compressing the large amount of single scan data into a common resolution of CD bins and a single average for the MD value.

With the CS method the long MD dimension of the scan can be split into n_B equal size sections

and the paper sheet can be reconstructed with an equal number of MD and CD values per scan. This can be interpreted in simulations as having a slow moving sheet with $\theta = 45°$ but in reality the longer MD dimension has been compressed and the data for each MD position represents a larger spatial dimension than for each CD position. Even though the MD resolution is artificially low, this type of MD binning can allow for better consideration of MD variations than is currently provided by averaging all of the scanned values into a single point.

To better observe the potential of CS this trial focuses solely on the CS method with two scanning sensors traversing in opposing directions. Two scanners represent a realistically achievable sensing framework and provides insight into the benefit of additional samples with CS. The following tests begin with a large CD dimension of 200 bins and proceed to reduced CD dimensions of 100 and 60 bins, respectively. The CS reconstructed sheet is compared visually to the true simulated sheet and the average CD and MD profiles are compared with both the true profiles and the observed scanner profiles.

4.3.1 200 CD bins

Due to memory limitations posed by constructing the CS matrix, $\mathbf{A}$, the simulated sheet speed is restricted to 1 m/s and only two scans (per scanner) are considered, as shown in Table 4.3. As with the other trials, MD profile frequencies are specifically chosen such that significant MD aliasing results. The same zero mean normally distributed random noise with standard deviation $\sigma_e = 0.15$ is added to the sheet. Furthermore, the same industrial CD profile provided by Honeywell is used to create the simulated sheet. In Figure 4.9 the simulated sheet and the trajectory of the two scanners is shown on the left and the CS reconstruction of the sheet is presented on the right.

Table 4.3: Simulation parameters for Trial 3, 200 CD bins

Parameter	**Description**	**Value**
w_{sheet}	sheet width (m)	8
v_{scan}	scan speed(s) (m/s)	1
t_{scan}	scan time (s)	8
v_{sheet}	sheet speed (m/s)	1
θ	traversing angle from MD (°)	45
$n_{CD} = n_B$	number of CD bins	200
n_S	number of scans (per scanner)	2
ω_{MD}	frequency of MD sinusoid(s) (Hz)	0.062, 0.12

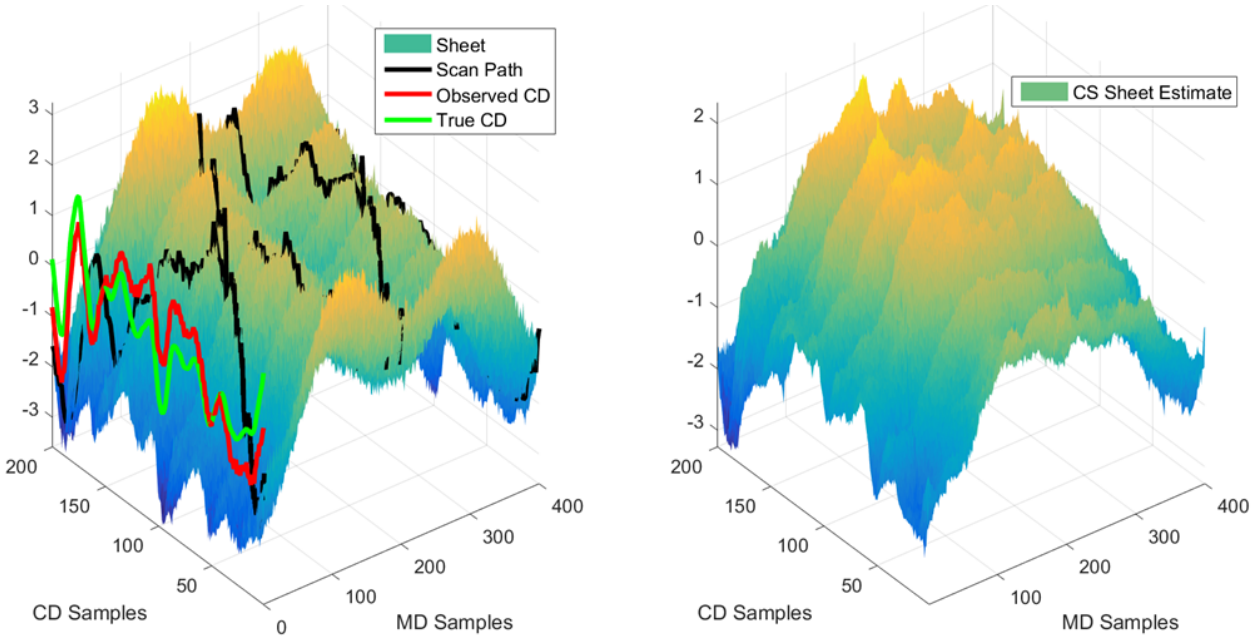

Figure 4.9: Simulated sheet and CS estimate with 200 CD bins

A larger CD dimension results in a lower overall sampling ratio because the number of samples per MD position is limited to the number of scanners. Although the CS reconstruction on the right side of Figure 4.9 has obvious imperfections it still provides valuable additional information relative to traditional scanner signal processing and exponential filtering. By leveraging the geometry of the scanning sensors the CS method makes efficient use of the available scanner information to provide an informative estimate of the sheet.

To determine the results of MD-CD separation, averages are taken along each MD position of the CS reconstruction for the MD profile and along each CD bin for the CD profile. The same averaging is conducted for the raw scanner data to generate the profile comparison presented in Figure 4.10. From the average CD profiles (top plot) the CS estimate (red) reduces some of the aliasing exhibited by the CD profile of the scan average (blue) which is evident numerically with the RMSE values presented in the legend. While the CS estimate provides a clear improvement, some aliasing is still present and the magnitudes of the estimated CD variations are smaller than the true profile (green). It is worth noting that the MD aliasing observed in the raw scan data is reduced due to the consideration of both scanners. The bottom plot compares the average MD trends where the scanned values (blue) are the average value over each scan. Again, the CS estimate contains obvious imperfections but still provides a clear improvement over simply taking the scan average.

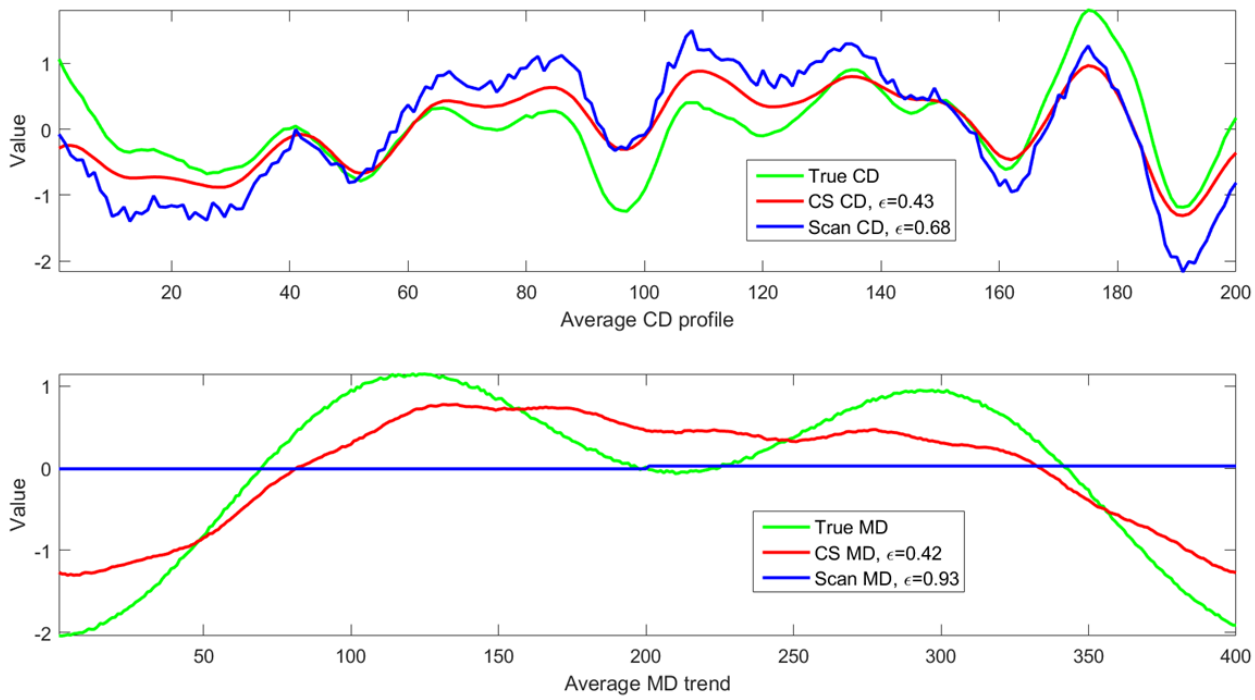

Figure 4.10: MD-CD separation results for Trial 3, 200 CD bins

4.3.2 100 CD bins

Reducing the CD dimension to 100 CD bins allows an increase in the sheet speed to 3 m/s with $\theta = 18.4°$ for two scans per scanner. Otherwise the sheet is simulated with essentially the same parameters as before, including the additive noise of standard deviation $\sigma_e = 0.15$. Again the MD variation frequencies are chosen to intentionally create severe MD aliasing in the observed measurements. The simulated true sheet is presented with the scanning trajectory on the left side of Figure 4.11 and the CS reconstruction is displayed on the right. In this test the CS reconstruction provides a fairly accurate representation of the true sheet. Note that the change in increment on the z-axis is indicative of reduced magnitude for the variations in the CS reconstruction.

To observe the results of the MD-CD separation the same averaging as before is performed to generate the CD profile and MD trend for both the CS reconstruction and the raw scan data. Observing the average CD profiles displayed in the top part of Figure 4.12 reveals a CS estimate that has removed the majority of the MD aliasing present in the raw scan data. Furthermore, the RMSE of the CS estimate for the CD profile is much smaller than that of the scanned data and also smaller than the RMSE of the previous CS reconstruction with 200 CD bins. Similarly, the

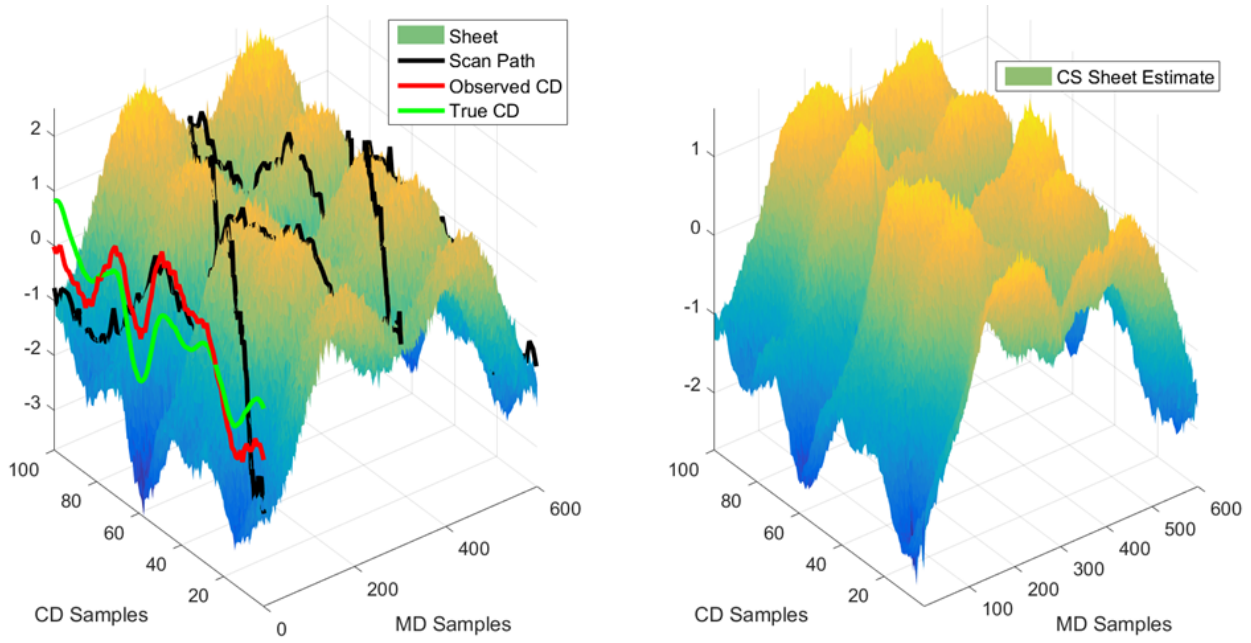

Figure 4.11: Simulated sheet and CS estimate with 100 CD bins

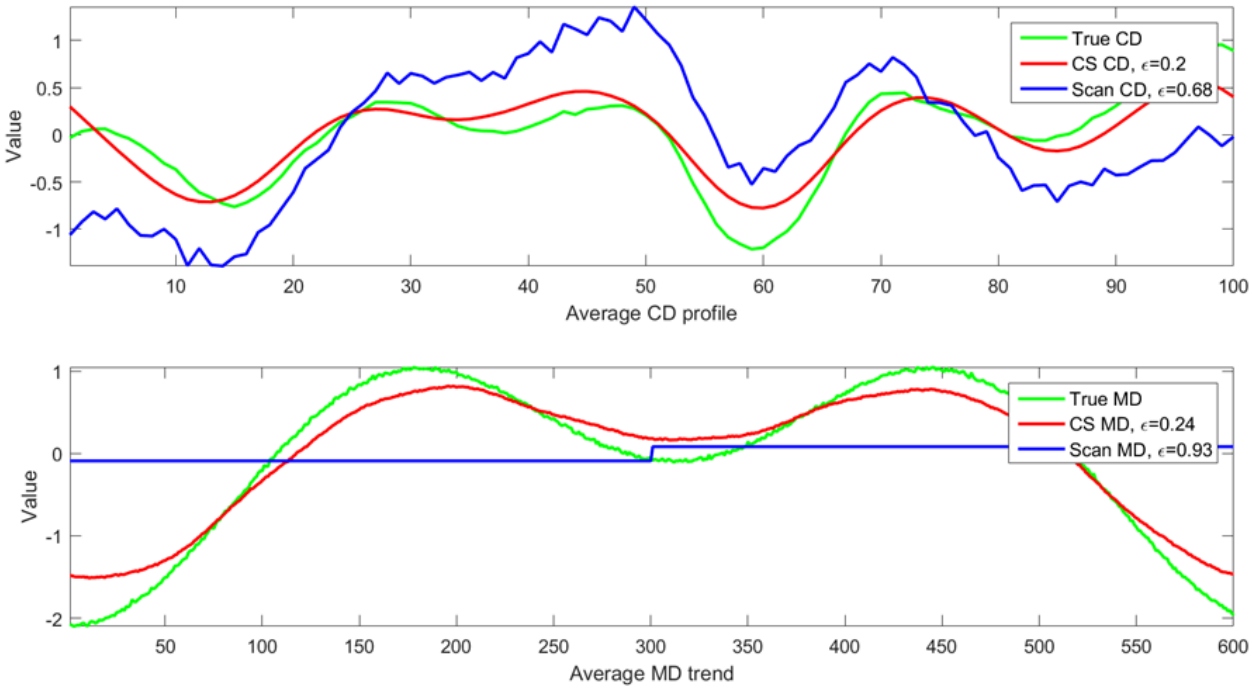

Figure 4.12: MD-CD separation results for Trial 3, 100 CD bins

CS estimate of the MD trend in the bottom part of Figure 4.12 provides a significant improvement over the scan average values and the estimate from the previous test with 200 CD bins. As before, the CS reconstruction of the sheet is not perfect but it clearly provides potentially important

additional information about the paper sheet. It is also clear from these results that given the right circumstances the CS method is able to successfully reduce the detrimental effects of MD aliasing. This test demonstrates the strong potential of CS for sheet profile estimation with a slightly improved sensing framework.

4.3.3 60 CD bins

The final test involves two scans per scanner over 60 CD bins with a sheet speed of only 1 m/s. The reduced CD dimension allows for an increased MD dimension (i.e., sheet speed) with regards to the memory limitations posed by constructing the CS matrix. However, the slow sheet speed of 1 m/s is chosen here to study the computation requirements with the smallest CS reconstruction. In Section 4.4 the results are summarized and the troublesome computation requirements of the CS method are considered. Other than the reduced CD dimension and sheet speed the simulated sheet is constructed with essentially the same parameters as the previous test. This is evident by observing the simulated sheet presented on the left side of Figure 4.13. Interestingly, even with the increased sampling ratio due to less CD bins, the CS reconstruction on the right side of Figure 4.13 is actually less accurate than the previous CS estimate with 100 CD bins. To speculate as to why, it is potentially possible that a trade-off exists with regards to the complexity (i.e., sparsity) of the CD profile and the resolution of measurements available to estimate it.

The same averaging as before is performed to evaluate the resulting MD-CD separation by comparing the average CD profile and MD trend. The CD profile estimates are compared in the top part of Figure 4.14 and the MD trends are compared in the bottom plot. Although the CS estimates for both the CD profile and MD trend outperform the raw scan estimates, they both fall short of the accuracy presented by the previous CS estimates with 100 CD bins. Even though the MD aliasing is only partially removed, the CS estimate still provides a valuable improvement over the traditional scanner signal processing.

The full tables of simulation parameters for the 100 CD bin and 60 CD bin tests were omitted here for concise presentation and limited redundancy. These parameters are presented along with additional simulation results in Appendix A. The additional simulation results include reconstructions using 4 scans for both 100 CD bins and 60 CD bins.

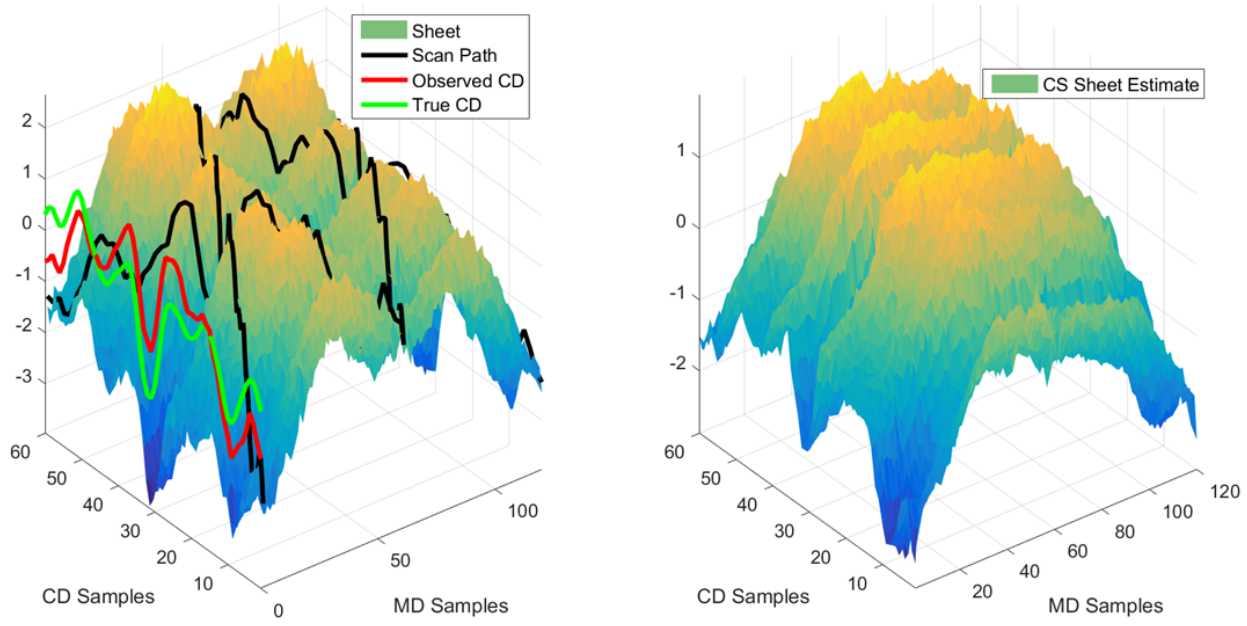

Figure 4.13: Simulated sheet and CS estimate with 60 CD bins

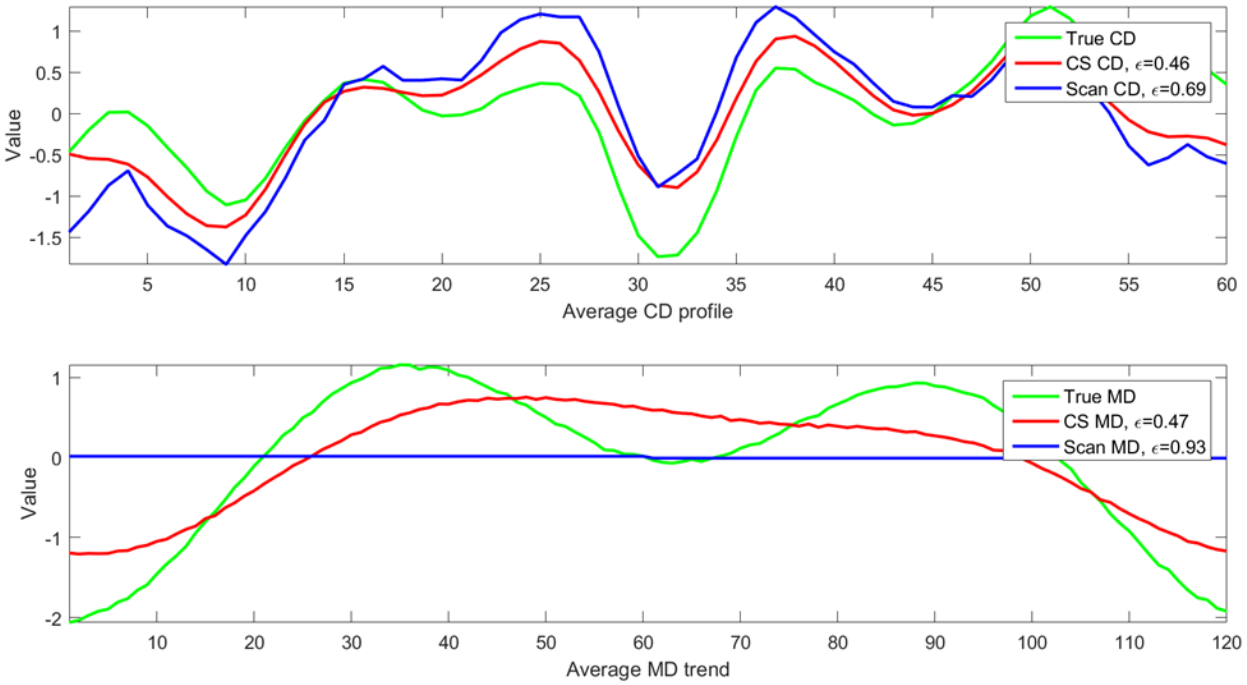

Figure 4.14: MD-CD separation results for Trial 3, 60 CD bins

4.4 Results summary and computational considerations

The results of the various simulations are summarized here and consideration is given to the computational requirements of the CS method. In Table 4.4 the RMSE results for Trial 1 with 100 CD

bins are summarized for each method. Table 4.5 portrays the summarized RMSE results for Trial 2 with 60 CD bins and a sheet speed of 8 m/s. The results from Trial 3 with two scanning sensors are summarized in Table 4.6. In the first two trials the CS and scaled CS method provide the lowest RMSE results followed by the PSA technique with the industrial standard EF technique generally providing the worst MD and CD profile estimates. Additive noise appears to cause a relatively minor increase in the RMSE of each MD-CD separation method. The CS method also fares well in the third trial relative to the raw scan averages which serve as the benchmark because not enough scans are considered for proper implementation of the EF method. Comparing all three trials reveals that the best CS reconstruction is achieved with 100 CD bins and two scanning sensors.

Table 4.4: Summary of experimental results for Trial 1

Description	**Estimation Method**	**CD RMSE**	**MD RMSE**
$\sigma_e = 0$	CS	0.30	0.81
$\sigma_e = 0$	PSA	0.48	1.33
$\sigma_e = 0$	EF	0.78	1.33
$\sigma_e = 0$	Scaled CS	0.36	0.52
$\sigma_e = 0.15$	CS	0.32	0.85
$\sigma_e = 0.15$	PSA	0.49	1.33
$\sigma_e = 0.15$	EF	0.80	1.33
$\sigma_e = 0.15$	Scaled CS	0.41	0.50

Table 4.5: Summary of experimental results for Trial 2

Description	**Estimation Method**	**CD RMSE**	**MD RMSE**
$\sigma_e = 0$	CS	0.45	0.92
$\sigma_e = 0$	PSA	0.64	1.42
$\sigma_e = 0$	EF	0.77	1.32
$\sigma_e = 0$	Scaled CS	0.43	0.51
$\sigma_e = 0.15$	CS	0.46	0.93
$\sigma_e = 0.15$	PSA	0.66	1.42
$\sigma_e = 0.15$	EF	0.77	1.32
$\sigma_e = 0.15$	Scaled CS	0.42	0.55

The results presented thus far have demonstrated the compressive sensing method performs favorably relative to the exponential filtering and power spectral analysis methods. However, one significant disadvantage of the CS method is the increased computation time that could render online implementation impractical. The required computation time varies significantly depending on the sampling ratio and size of the desired reconstruction, as is demonstrated in Table 4.7.

Table 4.6: Summary of experimental results for Trial 3

Description	Estimation Method	CD RMSE	MD RMSE
200 CD bins	CS	0.43	0.42
200 CD bins	Scan average	0.68	0.93
100 CD bins	CS	0.20	0.24
100 CD bins	Scan average	0.68	0.93
60 CD bins	CS	0.46	0.47
60 CD bins	Scan average	0.69	0.93

Table 4.7: Compressive sensing computation requirements

Test Description	Approximate computation time (s)
Trial 1, $\sigma_e = 0.15$	1358
Trial 2, $\sigma_e = 0$	2986
Trial 3, 200 CD bins	2477
Trial 3, 100 CD bins	320
Trial 3, 60 CD bins	21

It is only with the smallest CS reconstruction, i.e., 60 CD bins with two scans, that the CS method provides a practical computation time for online implementation. This is an important consideration for the MD-CD separation control problem which significantly favors rapid computation, i.e., in less than one minute. Another important consideration is the fact that an increased sampling ratio drastically improves the computation time. Unless a much faster CS method or processor is used, the implementation of CS for online control purposes is not practical with only a single scanning sensor. In Chapter 5 the benefits and limitations of each MD-CD separation method are summarized and recommendations for application and future studies are provided.

Chapter 5

Conclusion

The industrial MD-CD separation problem has proven surprisingly difficult to overcome. Of all the proposed MD-CD separation methods presented in Chapter 2 the long-standing EF technique is still the generally preferred industrial method. However, as the MD and CD control systems become capable of removing higher frequency variations the limitations of the EF method may become restrictive to improved control. Furthermore, finding means to better address the problem of MD aliasing is necessary for further progress in control of sheet properties. The standard use of a single scanning sensor creates significant challenges for proposing an improved MD-CD separation technique. While new sensing strategies such as sensor arrays and multiple scanners have been developed, the prohibitive cost of this sensor technology has limited widespread adoption.

In this work a comparative analysis is conducted between the EF method and two recent techniques, i.e., CS and PSA. These methods are compared with a single scanning sensor and the CS method is studied further using two scanning sensors. The qualitative results of the simulation studies (summarized in Section 4.4) indicate the strength of the compressive sensing method under the given simulation conditions. However, the CS method has limitations that are not adequately represented by the results of the experimental analysis. A more extensive consideration of the benefits and limitations associated with the studied MD-CD separation methods is summarized in Table 5.1.

The comparison of features in Table 5.1 stems primarily from research during the literature review and experience implementing each method in simulations. Both the PSA and EF methods have computation speeds that are easily suitable for online implementation but the computation requirements of the CS method are problematic. As seen in Section 4.4 the CS method only becomes potentially suitable for online implementation if a large sampling ratio and a small sheet dimension is applied. With the large common CD resolution of modern industrial control systems the CS

Table 5.1: Comparison of MD-CD separation methods

Feature	**CS**	**PSA**	**EF**
Computation speed	Slow	Fast	Fast
User complexity	Medium	High	Low
Required memory	High	Low	Low
Tuning importance	Minor	Crucial	Minor
MD aliasing attenuation	Good	Fair	Poor
CD profile estimation	Good	Good	Poor
MD profile estimation	Good	Fair	Poor
Reconstruction resolution	Good	Fair	Poor
Required number of scans	Few	Few	Many
Assumption reliance	Low	Medium	High

method is likely only suitable for online implementation if a sampling strategy that is significantly better than a single scanning sensor is used.

With respect to user complexity the PSA method is the most demanding particularly because of the importance of proper selection of the threshold tuning parameters. If these parameters are selected too high some of the dominant frequencies might be removed from the estimate and if they are chosen too low MD aliasing and unwanted noise might not be removed. After the desired MD resolution is selected and the CS matrix is constructed the CS method is not very demanding on the user. Only the parameter σ from (2.39) requires specification and for all of the tests conducted in this report it was simply set as $\sigma = 0.001$. Since σ is related to the noise level and fit of the data it can be increased to reduce the number of necessary computations or it can be reduced to increase the number of computations. Here $\sigma = 0.001$ is chosen because it was determined that a lower value did not provide substantive increases in the estimation accuracy. In this sense the CS method was tuned for accuracy as opposed to speed which likely contributed to some of the slow computation speeds. The EF method also has a single tuning parameter, i.e., the filter factor, which balances the weighting of the most recent measurement against the accumulated average. Only the CS method has difficulty with required memory due to the construction of the large CS matrix.

The attenuation of MD aliasing is most successful with the CS method although optimal tuning of the PSA method could potentially provide improved results. This reflects directly on the CD profile estimation which naturally favored the CS and PSA methods. The PSA method struggles to provide consistent estimates of the MD profile to improve upon the scan average benchmark whereas the CS method was more consistent in providing valuable MD information that is other-

wise not available. Another feature that strongly favors the CS method is the ability of the CS method to increase the resolution of the sheet reconstruction beyond what is generally possible under the Shannon-Nyquist sampling theorem. This ability to accurately estimate a high resolution reconstruction from relatively limited samples is arguably the largest strength of CS.

The final features covered in Table 5.1 relate to the necessary number of scans that need to be considered for each technique as well as the reliance of each method on assumptions. For accurate reconstruction the CS method was found to require consideration of at least two consecutive scans. More scans often resulted in an improved estimate but once again memory limitations become restrictive. The PSA method requires at least two scans at alternating speeds to compare spectra. Depending on the value of the filter factor and the degree of MD aliasing, the EF method requires many scans in order to build up an average of previous scans that can help eliminate the effect of MD aliasing. Reliance on assumption is primarily evaluated based on the common assumption of a static CD profile. The EF method requires the CD profile to be static for the entirety of the filter memory, i.e., potentially many scans. If the CD profile changes and the EF is not reset the changes in the CD profile can be masked by the filter and may take a long time to be recognized. The PSA method only requires the CD profile to be static over the two scans that are being compared and the CS method only requires the CD profile to be static over a small portion of a single scan. In the following two sections these features and the experience acquired are used to provide recommendations for future researchers as well as recommendations for industrial practitioners.

5.1 Recommendations for future work

The comparative analysis provided in this work provides valuable insight into the use of PSA and CS for MD-CD separation. The preliminary implementation of both methods allows for further development and potential for improvement that is beyond the scope of this research. For example, one area for recommended development is in the selection of the threshold for PSA. A more ideal threshold would better adapt to the noise level and number of primary frequency components in the scanned signal instead of relying on the specification of the tuning constants. Another potentially useful threshold could simply allow the user to select the maximum number of dominant MD frequencies to be identified and removed from the CD profile. Since the PSA method has already

been patented it is recommended that a similar technique be developed that leverages alternating scan speeds with a different basis function representation. Finally, instead of simply setting a particular frequency component to zero it is recommended that the MD contribution and the CD contribution at that particular frequency bin be quantified and reduced accordingly.

With respect to the CS method there are many areas available for suggested improvement. It is recommended to study the CS method on a more powerful computing system with a larger memory to evaluate performance with a larger CD dimension. Similarly, it would be ideal to develop a CS methodology that did not require such a large memory. With regards to biasing of CS estimates caused by soft thresholding it is recommended that a more robust debiasing technique be explored as opposed to the multiplicative scaling used in this work. A better debiasing technique could identify the supports of the large coefficients and scale the magnitude of these coefficients to the value of these supports in the original signal. Another important area for further development is improving the speed of the CS technique. A simple suggestion to improve the computation speed would be to use a more powerful computing system. Additionally, the parameter σ could be increased to relax the required accuracy of the CS reconstruction and reduce the number of required computations. Ultimately, it would be ideal to develop a recursive CS method for online implementation. Finally, it is recommended that a robust examination of both the PSA and CS techniques with industrial data be conducted. Unfortunately such a study is difficult to acquire as comparison to the true sheet properties requires expensive offline testing.

5.2 Recommendations for industrial progress

Many MD-CD separation techniques have been explored in literature and proposed as potential improvements to the EF technique regularly employed by industry. Furthermore, alternate sampling schemes such as sensor arrays, multiple scanners and variable speed scanners have been proposed to improve upon the typical constant speed single scanning sensor. Unfortunately, these proposed changes have not been successful in gaining widespread industrial implementation. For the alternative sensing frameworks the challenge involves providing clear justification that the improved property estimation and control justifies the high cost of the additional sensor technology. Understanding the continued dominance of the EF method for MD-CD separation in light of the literature

developments requires consideration of the user complexity and reliability of each method.

Relative to alternative techniques proposed in literature, EF offers very simple implementation in addition to high reliability up to a certain accuracy. It is speculated that the complex specification of tuning parameters that may vary across different operating conditions and different paper machines provides a strong deterrent for industrial adoption. To compete with EF an MD-CD separation technique needs to maintain straightforward implementation with reliability of estimation while also providing improved resolution and accuracy of the paper sheet properties. To date the literature has focused predominantly on the latter aspects of improved performance and has fallen short of considering the prior aspects of user-friendliness and reliability. This likely helps to explain a large amount of the lack of industrial acceptance for proposed MD-CD separation methods.

The PSA and CS techniques described and implemented in this work are insufficiently developed to successfully provide a robust estimation for industrial control purposes. Although neither are ready for standalone implementation they both are capable of offering valuable insight not provided by the traditional EF. The PSA method is recommended for online implementation to supplement the EF method by identifying MD variations and removing these from the CD profile. To be conservative only the largest identified MD variations can be removed. Similarly, CS could be used to help retroactively adjust the CD profile and improve the resolution of the MD trend. Furthermore, CS could help identify if the CD profile has changed which would indicate that the exponential filter should be reset.

As a final consideration, the offline implementation of CS could offer valuable information for quality reporting purposes. Accounting for the geometry of the scanning sensor allows the CS method to generate valuable information about sheet variations not available with the standard variance partitioning analysis. Even if the true MD dimension needs to be condensed due to memory limitations the CS method could identify high frequency MD variations that are ignored by simply taking the scan average. This increased resolution could provide important information for performing fault detection and determining better operating practices to reduce high frequency MD variations. Ultimately, while further development is required for the CS and PSA techniques to replace EF, both methods still offer potential sources of insight into achieving a more accurate representation of the true sheet property variations.

Part II

Machine Direction Adaptive Control

Chapter 6

Introduction

In the second part of this work the focus shifts to the development of an adaptive control strategy for the MD process of a paper machine. Model-based control methods such as model predictive control (MPC) have gained widespread implementation in many industries including sheet and film processes such as paper-making. The use of MPC for MD control has become increasingly popular since it was introduced by Honeywell in 1994 [16]. One of the difficulties associated with model-based control methods is maintaining an accurate model of the process when the true plant is changing according to the varying operating conditions and physical characteristics. A unified adaptive control framework is introduced and implemented to automatically update the controller model while maintaining closed-loop control when the true plant deviates from the previous model.

This work unifies collaborative efforts made by various members of our research group into an integrated MD adaptive control framework. The adaptive control strategy consists of techniques developed for MD process monitoring (PM), input design (ID) and system identification (SI). By integrating these separate procedures an adaptive tuning strategy for the MD process is developed. In this work the integrated framework is tested using an industrial simulator of an MD control system provided by our industrial collaborators at Honeywell. Specifically the presented case studies involve a simulated multi-input, multi-output (MIMO) MD process with an MPC control system that undergoes model-plant mismatch (MPM). The adaptive framework algorithms automatically identify the MPM, trigger an identification experiment, design the excitation signal, identify a new process model and update the controller. Using this experimental setup a series of tests are presented to examine alternative configurations of the adaptive control framework.

The objective of this work is to briefly familiarize the reader with each of the separate tasks in the framework before exhibiting the integration and functionality of the unified adaptive control technique. First, the simulated plant and control system for this case study are described and the

outline of the adaptive framework is introduced. Next, the specific techniques currently implemented for mismatch detection (also referred to here as process monitoring), input design and system identification are described. Afterwards, the experimental setup for the various tests is reviewed followed by a discussion of the experimental results. Concluding remarks and recommendations for future work are provided in the final chapter.

Chapter 7

System Overview

The industrial MD system simulated in this work involves both the true process dynamics and the MPC control system. The adaptive framework integrates to the closed-loop MD system and performs various tasks such as monitoring the input-output (IO) data and exciting the process input. In this chapter the simulated MD process is first presented before discussing how the adaptive control sub-functions interface with the closed-loop system.

7.1 Simulating the MD process

The MD paper machine process is modeled as a MIMO lower triangular system composed of six first order plus deadtime (FOPDT) transfer functions with three manipulated variables (MVs) and three controlled variables (CVs). The process output, including an additive disturbance, is fed back to a model predictive controller which determines the new process input values. A simple schematic of this closed-loop control system is shown below in Figure 7.1.

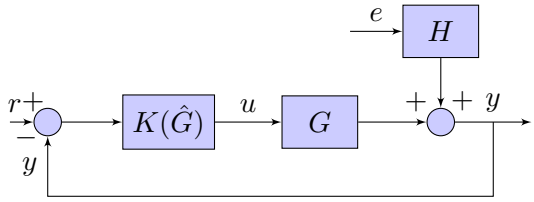

Figure 7.1: Closed-loop MD control system

From Figure 7.1 we obtain the following expression for our process outputs:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = G(q) \begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{bmatrix} + H(q) \begin{bmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{bmatrix}, \tag{7.1}$$

where $r(t)$ is the reference signal (or setpoint) and each $e_i(t)$ is an independent and identically distributed (IID) Gaussian white noise sequence with zero mean and variance $\sigma^2_{e_i}$ [41]. Process outputs $y_1(t)$, $y_2(t)$ and $y_3(t)$ represent the measured basis weight (lbs/3000ft^2), press moisture (%) and reel moisture (%), respectively. Furthermore, process inputs $u_1(t)$, $u_2(t)$ and $u_3(t)$ represent the thick stock flow (gpm), press section steam pressure (psi) and reel section steam pressure (psi), respectively. Apart from some user-specified tuning parameters, the details of the MPC controller, $K(\hat{G})$, presented in this chapter will be limited to the update of $\hat{G}$.

Both the nominal model used by the controller, $\hat{G}(q)$, and the true plant transfer function, $G(q)$, have the following form:

$$G(q) = \begin{bmatrix} G_{11}(q) & 0 & 0 \\ G_{21}(q) & G_{22}(q) & 0 \\ G_{31}(q) & G_{32}(q) & G_{33}(q) \end{bmatrix}. \tag{7.2}$$

Each G_{ij} in (7.2) represents an FOPDT transfer function which is defined in continuous time according to

$$G_{ij}(s) = \frac{b_{ij}}{a_{ij}s + 1} e^{-d_{ij}s}, \tag{7.3}$$

where b_{ij}, a_{ij} and d_{ij} represent the continuous gain, time constant and time delay parameters, respectively [23]. The nominal model used by the controller, $\hat{G}(q)$, takes the exact same form except the sub-transfer functions are denoted $\hat{G}_{ij}$ with parameters $\hat{b}_{ij}$, $\hat{a}_{ij}$ and $\hat{d}_{ij}$. Initially there is a 10% negative MPM in each of the process parameters, i.e., each of the parameters in the nominal model are 10% lower than the true process parameters presented in Table 7.1.

Table 7.1: Continuous MD Process Parameters

G_{ij}	b_{ij}	a_{ij}	d_{ij}
G_{11}	1.1600	73.26	77
G_{21}	0.3256	50.622	99
G_{22}	-0.1540	297.0	121
G_{31}	0.8283	170.28	43
G_{32}	-0.2618	232.32	33
G_{33}	-0.0611	27.324	99

The noise model, $H(q)$, is a diagonal matrix of stable and inversely stable filters, i.e.,

$$H(q) = \begin{bmatrix} H_1(q) & 0 & 0 \\ 0 & H_2(q) & 0 \\ 0 & 0 & H_3(q) \end{bmatrix}, \tag{7.4}$$

where each $H_i(q)$ can be either high pass or low pass filters of varying order. The specific noise models used in this case study vary so these will be defined along with the noise variance in the experimental setup. Next, a discussion on how the various tasks in the adaptive framework interface with the main control loop is provided.

7.2 Adaptive control framework

The adaptive control framework involves a large number of sub-functions that integrate into the main control loop. The specific details of this integration is described along with a review of these methods in the following chapters. First, however, a general overview of how the adaptive control framework integrates with the closed-loop system is provided. The adaptive control framework implemented in this work consists of three major elements: process monitoring, input design and system identification. It is worth mentioning that although the term process monitoring is used here to preserve generality, the current implementation of this PM element is equivalent to MPM detection. An illustrative schematic outlining the adaptive framework on a closed-loop system is provided in Figure 7.2. For simplicity, the specific IO data sent to the PM, ID and SI blocks is clarified in the detailed explanations of each of these blocks.

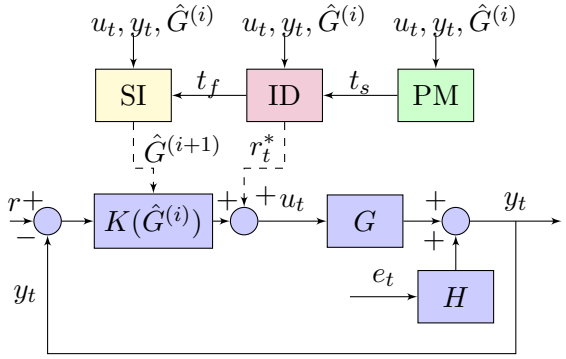

Figure 7.2: Adaptive framework for closed-loop MPC

During normal operation the model predictive controller, $K(\hat{G}^{(i)})$, controls the process, G, and the process monitoring algorithm, PM, collects process input data, u_t, and output data, y_t. The current plant model estimate, $\hat{G}^{(i+1)}$, is also provided to the PM algorithm for reference. When G deviates from the estimated process model, $\hat{G}^{(i)}$, the PM algorithm detects the MPM and triggers an excitation experiment by sending the start time, t_s, to the ID algorithm. An optimally designed excitation signal, r_t^*, is then calculated by the ID algorithm and added to the actuator movements, u_t, to provide additional process excitation for a specified experiment duration. A pseudo-random binary sequence (PRBS) can also be used to excite the system as is discussed in the experimental setup. Once the end of the experiment, t_f, is reached, the SI algorithm collects the informative IO data from the experiment duration (u_t and y_t between t_s and t_f) to determine a new process model estimate. The new estimate of the process, $\hat{G}^{(i+1)}$, is then implemented in the controller and normal operation resumes. Successful implementation of this framework allows the controller to automatically respond to changes in the process while maintaining feedback control, without requiring operator intervention. In what follows, each of the major elements of the adaptive control framework are reviewed.

Chapter 8

System Identification

In addition to re-tuning the process model, system identification is of critical importance to the MPM detection aspect of the adaptive control framework. A detailed description of the system identification methods is presented here in order to provide necessary context for the following sections. First some general background and motivation is provided regarding the particular selection of identification methods. This is followed by a theoretical review of how these methods are implemented on the closed-loop system. The chapter concludes with a user-focused description of the use of these methods in the simulated closed-loop MD system.

8.1 Background

System identification methods can be broadly categorized as follows:

(a) The prediction error family.

(b) Subspace approaches.

(c) Nonparametric correlation and spectral analysis methods.

An important difference between open-loop and closed-loop data is the correlation between the noise and the input in closed-loop data as a result of feedback control. This correlation between the noise and the input can be problematic for extending open-loop identification methods such as subspace and nonparametric methods to closed-loop data [24]. For this reason, the analysis presented here focuses on category (a), i.e., identification procedures that fit under the framework of the prediction error method (PEM). Closed-loop identification methods can be classified based on their respective feedback assumptions as follows [28]:

(i) **Direct identification**: neglects the effect of feedback and the correlation between

input and noise. Directly identifies the open-loop system from input and output measurements.

(ii) **Indirect identification**: uses knowledge of the controller to determine the open-loop system from an identified closed-loop transfer function (e.g., the sensitivity function).

(iii) **Joint input-output identification**: identifies closed-loop transfer functions from an external excitation signal (e.g., dither signal or set-point change) to both the process input and output. Determines the open-loop system from these closed-loop transfer functions. Includes the projection method and the two-stage approach.

For process monitoring, closed-loop identification must be performed on routine operating data (i.e., without external excitation). Joint input-output identification methods are thus ruled out as they depend on some form of external excitation. Furthermore, both indirect and joint input-output identification techniques can become overly complex in the presence of nonlinear feedback [24] [56]. As it is not uncommon for industrial MPC to display nonlinear dynamics, the indirect and joint input-output identification techniques both fall outside the scope of this report. Instead the focus of this report is constrained to comparing the direct identification (DI) method to a newly proposed method that utilizes both an autoregressive exogenous (ARX) model and an output-error (OE) model. This newly proposed method is referred to as the closed-loop ARX-OE identification method. Implementation of both of these techniques is discussed in the following sections.

8.2 Prediction error method

The closed-loop identification methods introduced in (i)-(iii) as well as the ARX-OE method can all be seen as varying parameterizations of the PEM. The PEM minimizes a cost function of the squared 2-norm of the prediction error, i.e.,

$$V_N(G,H) = \frac{1}{N}\sum_{t=1}^{N}\hat{\epsilon}^2(t,\theta), \tag{8.1}$$

where $\hat{\epsilon}(t,\theta)$ represents the difference between the measured and predicted outputs, i.e.,

$$\hat{\epsilon}(t,\theta) = y(t) - \hat{y}(t|\theta). \tag{8.2}$$

Please note the distinction made between the actual process white noise sequence, $e_i(t)$, and the prediction error $\hat{\epsilon}(t,\theta)$ [25]. For plant and noise models defined by the parameter vector θ, the predicted output, $\hat{y}(t|\theta)$, is given by [41]

$$\hat{y}(t|\theta) = \hat{H}^{-1}(q,\theta)\hat{G}(q,\theta)u(t) + \big[1 - \hat{H}^{-1}(q,\theta)\big]y(t), \tag{8.3}$$

and after some algebraic manipulation the following expression for the prediction error is obtained

$$\hat{\epsilon}(t,\theta) = \hat{H}^{-1}(q,\theta)\big[y(t) - \hat{G}(q,\theta)u(t)\big]. \tag{8.4}$$

Ultimately the goal is to obtain the parameter vector θ_o such that the parameterized models are equal to the true system, i.e,

$$\hat{G}(q,\theta_o) = G(q) \qquad \hat{H}(q,\theta_o) = H(q). \tag{8.5}$$

where $G(q)$ and $H(q)$ represent the true plant and noise systems, respectively. Minimizing (8.1) with respect to θ results in the following PEM estimate of the parameter vector

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta, Z^N), \tag{8.6}$$

where Z^N represents the provided input and output data, i.e.,

$$Z^N = \{u(1), y(1), \ldots, u(N), y(N)\}, \tag{8.7}$$

for a given time span $t = 1, \ldots, N$ [24][25].

The PEM implemented here uses a polynomial representation of the transfer functions $\hat{G}(q,\theta)$ and $\hat{H}(q,\theta)$ for the MIMO MD system. As an example, the expression for y_3 in (7.1) becomes

$$\begin{aligned} A_3(q)y_3(t) = \frac{B_{31}(q)}{F_{31}(q)}u_1(t-n_{k_1}) + \frac{B_{32}(q)}{F_{32}(q)}u_2(t-n_{k_2}) \\ + \frac{B_{33}(q)}{F_{33}(q)}u_3(t-n_{k_3}) + \frac{C_3(q)}{D_3(q)}e(t), \end{aligned} \tag{8.8}$$

where n_k represents the various delays of each input. Expressions for y_1 and y_2 follow in the same

fashion. The polynomials presented in (8.8) are defined as follows:

$$A(q,\theta) = 1 + \hat{a}_1 q^{-1} + \cdots + \hat{a}_{n_a} q^{-n_a}, \tag{8.9}$$

$$B(q,\theta) = \hat{b}_1 + \hat{b}_2 q^{-1} + \cdots + \hat{b}_{n_b} q^{-n_b}, \tag{8.10}$$

$$C(q,\theta) = 1 + \hat{c}_1 q^{-1} + \cdots + \hat{c}_{n_c} q^{-n_c}, \tag{8.11}$$

$$D(q,\theta) = 1 + \hat{d}_1 q^{-1} + \cdots + \hat{d}_{n_d} q^{-n_d}, \tag{8.12}$$

$$F(q,\theta) = 1 + \hat{f}_1 q^{-1} + \cdots + \hat{f}_{n_f} q^{-n_f}. \tag{8.13}$$

Where model orders n_a, n_b, n_c, n_d, n_f and n_k must be specified. In this work the time delays, n_{k_i}, are estimated by comparing a series of ARX models with different delays and determining which delay value corresponds to the ARX estimate that provides the best fit to the IO data.

8.3 Direct identification

The closed-loop DI method begins by defining the model orders specified above and proceeds by solving (8.6) directly with process input and output data. It is assumed that the plant model is known *a priori* to have a first order numerator and denominator, i.e., $n_b = n_f = 1$. Although the noise model is unknown, an estimate of the noise model order is necessary for DI. In this report the noise model is estimated to have both a first order numerator and denominator, i.e., $n_c = n_d = 1$. For SISO simulations, it has been shown that improper specification of the noise model order can result in increased bias and variance of the parameter estimates. Although early MIMO simulations failed to consistently exhibit this behavior, it is considered as part of our experimental investigation. Finally, $A(q)$ is assumed to be zeroth order which leads to the following output expressions for the direct identification method:

$$y_1(t) = \frac{\hat{b}_{11}}{1 + \hat{f}_{11} q^{-1}} u_1(t - n_{k_{11}}) + \frac{1 + \hat{c}_1 q^{-1}}{1 + \hat{d}_1 q^{-1}} e_1(t), \tag{8.14}$$

$$y_2(t) = \frac{\hat{b}_{21}}{1+\hat{f}_{21}q^{-1}}u_1(t-n_{k_{21}}) + \frac{\hat{b}_{22}}{1+\hat{f}_{22}q^{-1}}u_2(t-n_{k_{22}}) + \frac{1+\hat{c}_2q^{-1}}{1+\hat{d}_2q^{-1}}e_2(t), \tag{8.15}$$

and

$$y_3(t) = \frac{\hat{b}_{31}}{1+\hat{f}_{31}q^{-1}}u_1(t-n_{k_{31}}) + \frac{\hat{b}_{32}}{1+\hat{f}_{32}q^{-1}}u_2(t-n_{k_{32}}) + \frac{\hat{b}_{33}}{1+\hat{f}_{33}q^{-1}}u_3(t-n_{k_{33}}) + \frac{1+\hat{c}_3q^{-1}}{1+\hat{d}_3q^{-1}}e_3(t). \tag{8.16}$$

Comparing the above output expressions to the original Box-Jenkins system it is easy to see that for the DI method

$$\hat{G}_{ij}(q,\theta) = \frac{B_{ij}(q,\theta)}{F_{ij}(q,\theta)} \tag{8.17}$$

and

$$\hat{H}_i(q,\theta) = \frac{C_i(q,\theta)}{D_i(q,\theta)}. \tag{8.18}$$

As is apparent from (8.14) - (8.16), each transfer function is modeled with the same FOPDT structure as the true process, G_{ij}, shown in (7.3). Values of $\hat{b}_{ij}$ and $\hat{f}_{ij}$ in (8.14) - (8.16) represent the estimated gain and time constant parameters, respectively. Although the presentation here is focused on the more complex MIMO system it is worth noting that the output expression for the SISO implementation of DI is equivalent to (8.14).

An additional variation of the DI method specified above has been investigated in previous simulations. This variation is referred to as direct output error (OE) identification, as the noise model order is specified such that (8.14) - (8.16) take the form of output error models, i.e., $n_c = n_d = 0$ and thus $\hat{H}_i(q,\theta) = 1$. Although a detailed analysis of this technique is beyond the scope of this report, it is worth noting that the results obtained using the MIMO industrial MD simulator were interestingly comparable in quality to applying the true noise model order. Further investigation may be warranted.

8.4 ARX-OE identification

The ARX-OE method is a recent technique developed by Q. Lu et al., in 2016 for performing closed-loop identification using routine operating data without *a priori* knowledge of the noise model order.

A thorough documentation of this technique is currently pending finalization and submission but a brief introduction is provided in what follows. The ARX-OE method is carried out through the following two step procedure [43]:

1. A high order ARX estimate of the noise model is generated and IO data is filtered with the inverse of the noise model estimate.

2. An OE model with the filtered IO data is solved via the PEM.

After introducing the ARX-OE identification method for the SISO system a straightforward extension to the MIMO case is presented. First, the Box-Jenkins system from (7.1) is represented with the following ARX model structure:

$$A(q,\theta)y(t) = B(q,\theta)u(t-n_k) + e(t), \tag{8.19}$$

where $A(q,\theta)$ and $B(q,\theta)$ are defined as before. Dividing both sides of (8.19) by $A(q,\theta)$ yields

$$y(t) = \frac{B(q,\theta)}{A(q,\theta)}u(t-n_k) + \frac{1}{A(q,\theta)}e(t) \tag{8.20}$$

and by relating back to (7.1), it is apparent that

$$\hat{G}(q,\theta) = \frac{B(q,\theta)}{A(q,\theta)} \quad \text{and} \quad \hat{H}(q,\theta) = \frac{1}{A(q,\theta)}. \tag{8.21}$$

The order of $A(q,\theta)$ is selected sufficiently high enough to capture the dynamics of the true noise model (e.g., $n_a \approx 15$). This has been made evident by conducting tests that compare the impulse response (IR) coefficients of both the true and estimated inverse noise models such as the example shown in Figure 8.1.

Upon estimating the inverse noise model, $A(q,\theta)$, the next step is to filter the IO data by $A(q,\theta)$. Filtered input and output data are defined respectively as

$$u^f(t) = A(q,\theta)u(t) \quad \text{and} \quad y^f(t) = A(q,\theta)y(t). \tag{8.22}$$

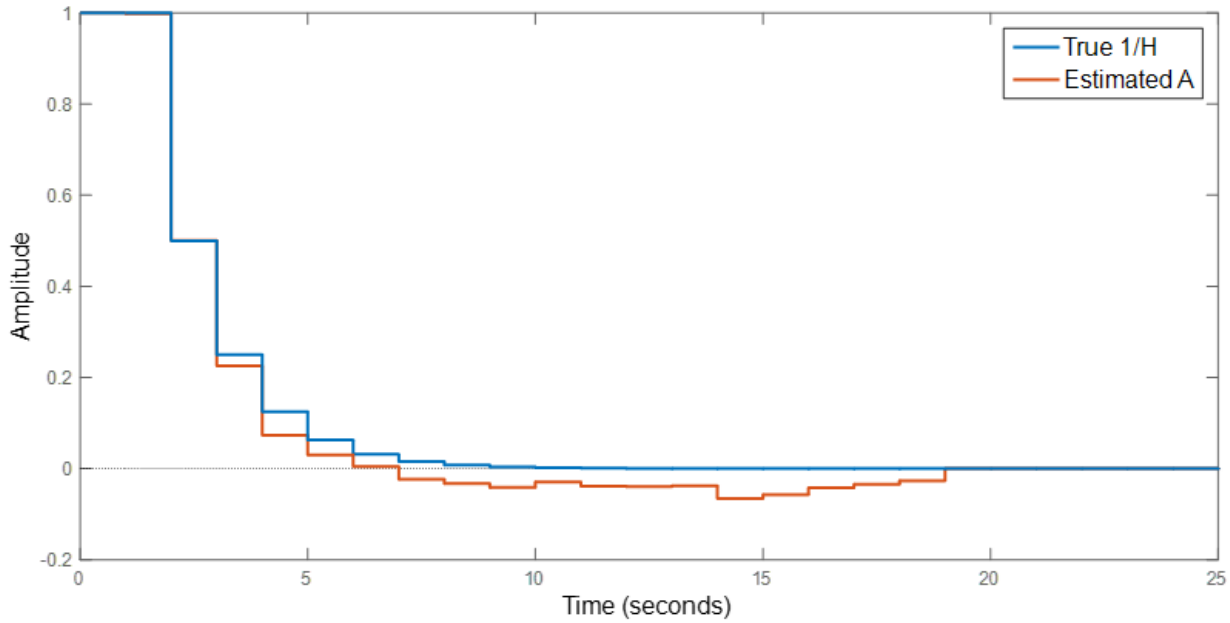

Figure 8.1: Impulse response of true and estimated inverse noise models

Inserting IO expressions from (8.22) into (8.20) yields

$$\frac{y^f(t)}{A(q,\theta)} = \frac{B(q,\theta)}{A(q,\theta)}\frac{u^f(t-n_k)}{A(q,\theta)} + \frac{1}{A(q,\theta)}e(t) \tag{8.23}$$

and multiplying both sides of (8.23) by $A(q,\theta)$ results in the following output error model

$$y^f(t) = \frac{B(q,\theta)}{A(q,\theta)}u^f(t-n_k) + e(t). \tag{8.24}$$

Finally, this output error model can be solved via the PEM and equation (8.1) as described before.

For the MIMO system, (8.24) extends to the following system of equations:

$$y_1^f(t) = \frac{B_{11}(q,\theta)}{A_{11}(q,\theta)}u_1^f(t-n_{k_{11}}) + e_1(t), \tag{8.25}$$

$$\begin{aligned} y_2^f(t) = &\frac{B_{21}(q,\theta)}{A_{21}(q,\theta)}u_2^f(t-n_{k_{21}}) \\ &+\frac{B_{22}(q,\theta)}{A_{22}(q,\theta)}u_2^f(t-n_{k_{22}}) + e_2(t), \end{aligned} \tag{8.26}$$

and

$$y_3^f(t) = \frac{B_{31}(q,\theta)}{A_{31}(q,\theta)} u_1^f(t - n_{k_{31}}) + \frac{B_{32}(q,\theta)}{A_{32}(q,\theta)} u_2^f(t - n_{k_{32}}) + \frac{B_{33}(q,\theta)}{A_{33}(q,\theta)} u_3^f(t - n_{k_{33}}) + e_3(t). \tag{8.27}$$

Model orders of $A(q,\theta)$ and $B(q,\theta)$ for the second part of the ARX-OE method are specified as $n_a = n_b = 1$. The final filtered output-error expressions are obtained by simply substituting

$$\frac{B_{ij}(q,\theta)}{A_{ij}(q,\theta)} = \frac{\hat{b}_{ij}}{1 + \hat{a}_{ij} q^{-1}} \tag{8.28}$$

into (8.25) - (8.27). The values of $\hat{b}_{ij}$ and $\hat{a}_{ij}$ obtained from this identification procedure provide estimates of the process gain and time constant, respectively. Furthermore, these estimates are used to evaluate the accuracy of the identification as is described in the experimental setup. Ultimately, the purpose of identifying the process model is to be able to re-tune the controller and obtain improved controller performance. The implementation of these closed-loop identification techniques in the MD industrial simulator is described in the following section.

8.5 Implementation

Closed-loop system identification is fundamental to both the PM and SI aspects of the adaptive control framework. The PM algorithm is setup such that only the ARX-OE method is used whereas the SI algorithm applies both the DI method and the ARX-OE method. When the SI stage is complete the user can select whether to update the controller model with either the DI estimate or the ARX-OE estimate. In this work both are selected to provide a comparison of the two methods on control performance. Here the specific implementation of closed-loop identification is described for both the PM and SI stages.

For PM a series of ARX-OE identifications are performed using over-lapping moving windows of IO data. The moving window has a length of 1440 samples and takes steps of length 60 samples with a sample time of 5s. In other words the identifications are performed every 5 minutes on the previous 2 hours worth of IO data. In the first stage of the ARX-OE method the inverse noise model estimate $A(q,\theta)$ is estimated with an order of $n_a = 18$ and the polynomial $B(q,\theta)$ is estimated with an order of $n_b = 10$. The time delay estimates used by the controller are also provided for the

ARX-OE identification.

The large orders of the first stage ARX model require estimation of many parameters and since there is only a limited amount of data the resulting estimates can exhibit high variance. Errors in an estimated model generally take the form of either bias or variance. Bias can occur as a result of having a model structure that does not capture the complexities of the true process whereas variance is typically due to disturbances in the measured data. Minimizing the cost function, $V_N(\theta, Z^N)$, involves a trade-off between bias and variance that is determined by the selection of model order. Increasing the order of a model generally decreases the bias while simultaneously increasing the variance. To address the large model order and reduce the variance of the estimates regularization constants Λ and R are determined and the regularized ARX model is estimated, i.e.

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \hat{\epsilon}^2(t,\theta) + \Lambda(\theta - \theta^*)^T R(\theta - \theta^*), \tag{8.29}$$

where θ^* is the mean of the assumed Gaussian prior distribution of θ [12].

Since the PM algorithm uses small amounts of routine operating data the signal to noise ratio is improved by filtering out the high frequency noise. This denoising step is performed after the IO data is filtered with the inverse noise model estimate. The final stage of the ARX-OE method involves solving the OE model. To improve the OE estimation the range of possible estimates are bounded using *a priori* knowledge of acceptable values. Finally the results of the ARX-OE identification are sent to the next stage of the PM algorithm, as will be discussed in detail in the following chapter.

The primary closed-loop identification used to re-tune the controller model is performed after the excitation experiment conducted by the ID algorithm. In this work the excitation experiment has a duration of 3000 samples (i.e., ≈ 4.2 hours). The IO data from the duration of the experiment is collected by the SI algorithm to perform both DI and ARX-OE identification. First the time delays are estimated by comparing a series of ARX models with varying time delays and determining which model provides the best fit. The ARX-OE method proceeds by solving the regularized ARX model as described before. The IO data is then filtered with the high order inverse noise model approximation and the constrained OE model is estimated to determine the gain and time constant parameters. The DI method is performed with similar constraints on the acceptable parameter

values. Both the OE model in the ARX-OE method and the DI model are estimated with first order numerator and denominator polynomials for the process model estimate. Furthermore, the DI method uses a first order numerator and denominator structure to estimate the noise model. After the results of both identifications are determined the user specifies which method to use for controller tuning. Finally, the controller is re-initialized with new process model estimates and regular operation resumes until another MPM is detected. This concludes the discussion of the use of closed-loop ARX-OE and DI in the adaptive control framework. Next, a brief introduction to the MPM detection technique is provided.

Chapter 9

Model Plant Mismatch Detection

The process monitoring (PM) aspect of the adaptive control framework, shown in Figure 7.2, determines whether or not the controller model estimate is adequately representative of the true process. In order to determine whether this mismatch exists, a novel model-plant mismatch (MPM) detection algorithm is employed. In this chapter some general background into the MPM detection algorithm is provided followed by a discussion of the implementation of this technique in the adaptive control framework. A thorough treatment of the theory involved with the MPM detection technique is beyond the scope of this work so the interested reader is referred to the development of the technique by Q. Lu et al., in 2017 [42].

9.1 Background

For model-based control methods such as MPC the existence of MPM can be detrimental to controller performance. Failure to acknowledge and correct significant MPM can result in sub-optimal control and production losses. It is estimated that up to 60% of industrial controllers exhibit controller performance problems from issues such as poor tuning, lack of maintenance and inappropriate control structure, among others [35]. Controller performance monitoring is a well established field that has drawn increasing interest since the development of the minimum variance benchmark by Harris in 1989 [32]. Since monitoring is performed almost constantly it is beneficial to employ a technique that passively monitors the process. Furthermore, since both the true noise model and the true process model can change during operation it is necessary to distinguish process model changes from noise model changes. If the noise model changes and the process model does not change it is unlikely that conducting an identification experiment will improve controller performance [6]. The MPM detection implemented here uses the ARX-OE identification method on routine operating data in combination with a support vector machine (SVM) in order to distinguish between changes

in the process or noise models and ultimately determine whether MPM has occurred.

To determine whether or not MPM has occurred a series of closed-loop ARX-OE identifications are performed as was described in the previous chapter. Due to the presence of noise and the lack of significant excitation in routine operating data the variance of the ARX-OE estimates can be considerable even if the true process behavior has not changed. In other words, directly comparing the routine ARX-OE estimates to the current process model does not provide a reliable MPM detection procedure since the ARX-OE estimates have a natural range of variation. Instead, a range of acceptable variation of process model estimates is established as a benchmark and new IO data is compared to this benchmark. Specifically, an SVM is used to compare new process model estimates to the benchmark and if these estimates fall outside of the acceptable range they are regarded as MPM indications.

For the purpose of this work the SVM can be understood as a binary classifier that indicates whether or not ARX-OE estimates belong to the model implemented in the controller or whether MPM has occurred. In order to classify new data the SVM model must first be trained to establish the acceptable range of variation associated with the current model. Immediately after a new controller estimate is determined a training stage begins during which a series of overlapping moving-window ARX-OE identifications are performed. During the training stage it is assumed that no MPM exists. For PM it is desirable to be able to detect mismatch in the gain, time constant and time delay parameters of the process and controller. Therefore the SVM model is trained using the IR coefficients of the ARX-OE estimated process models since the IR coefficients capture changes in all of the aforementioned parameters. The SVM is used to map the IR response coefficients into a high dimensional feature space for comparison and ultimately classification. An illustration of the high dimensional SVM classification is shown in Figure 9.1. This mismatch classification is performed for both process and noise model estimates in order to confirm whether identifying a new process model will improve controller performance. Further details on the implementation of the MPM algorithm are described in the following section but for a more theoretical description of the SVM please refer to the work by Q. Lu et al. [42].

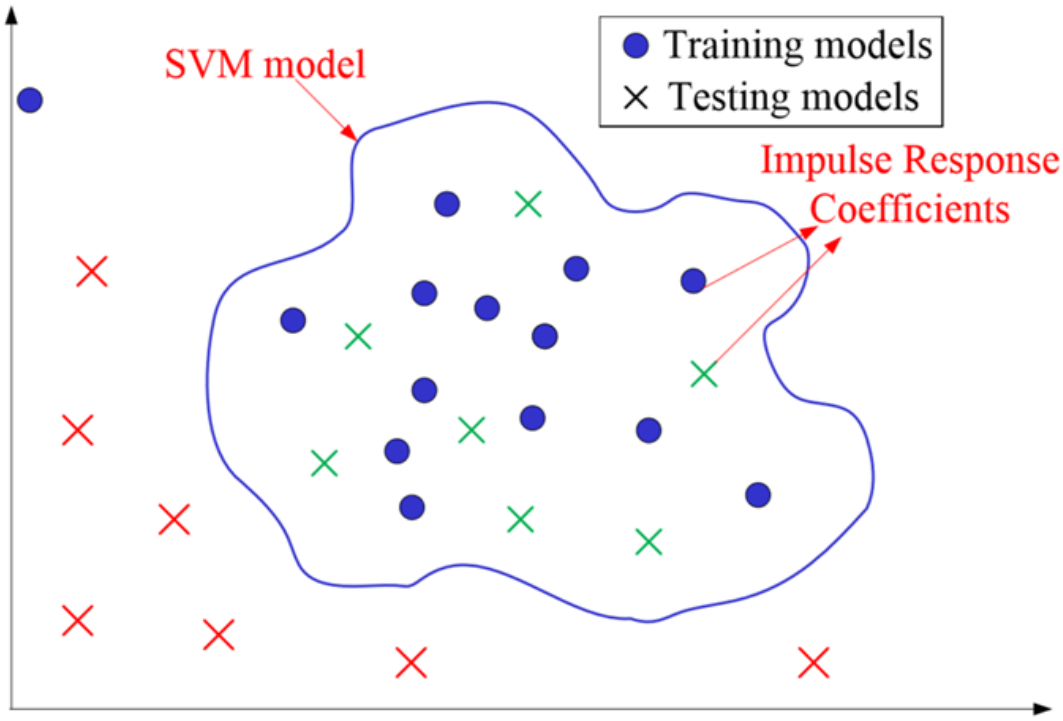

Figure 9.1: Illustration of high dimensional SVM classification [42]

9.2 Implementation

To implement SVM classification the SVM model must first be trained with data that does not contain MPM in order to establish a benchmark for comparison. When the controller is initialized, or a new model is estimated, the PM algorithm collects data over a period of operation that is assumed to have a relatively static true process, $G \approx \hat{G}$. Recall, the PM algorithm uses routine overlapping moving-window closed-loop ARX-OE identification estimates to train the SVM with IR coefficients. For this particular study the duration of the training period is 3500 samples with a sample time of 5s (i.e., ≈ 4.9 hours of training). During the training period the closed-loop ARX-OE identifications are performed with a window length of 1440 samples and step size of 60 samples (i.e., 2 hour window length and 5 minute step size). The IR coefficients of the resulting process and noise model estimates are provided to the SVM algorithm to establish an acceptable range of estimates that are considered free of MPM. Now that the SVM is trained, the MPM detection algorithm is considered active and begins comparing new IO data to the benchmark established by the SVM model.

The period of operation where the MPM detection algorithm is active is referred to as the testing period. As with the training period, the testing period involves routine ARX-OE identifications that are performed with an equivalent moving-window procedure. The SVM model constructed from the training data is illustrated by the blue boundary line in Figure 9.1. When new process and noise model estimates from the testing data are mapped into the high dimensional feature space they are compared to the SVM model. The testing models that fit with the SVM model (marked in green in Figure 9.1) are considered free of MPM and those that violate the SVM model boundary (marked in red in Figure 9.1) are considered to exhibit MPM. The SVM algorithm returns plant and noise model indicators that provide positive values if no MPM is suspected and negative values if the estimate is classified as exhibiting MPM.

To avoid triggering unnecessary identification experiments the PM algorithm requires a large number of consecutive negative indications. In this work the plant and noise model indicators are observed over a duration of an hour (i.e., 720 samples) and if over 95% of the values are negative it is determined that MPM has occurred. The noise model indicators are analyzed first and if the alarm threshold of negative values is reached the PM algorithm resets. Mismatch detection is stopped so that the SVM can be retrained. If the noise model does not exhibit mismatch then the process model indicators are analyzed and if the alarm threshold is reached an identification experiment is triggered. Now that MPM has been identified the current sample time marks the beginning of the identification experiment during which a series of steps are performed to calculate an optimal excitation signal, r_t^*, as is described in the following chapter.

Chapter 10

Input Design

Input design (ID) is used to calculate an optimal excitation signal, r_t^*, that is added to the process input, u_t, to create more informative process IO data. Informative IO data has a relatively high signal to noise ratio which enables improved system identification and more accurate parameter estimation. Industrial identification experiments are often performed in the open-loop mode where feedback control is removed and the set-point values can be manipulated. An example of an open-loop identification experiment is referred to as the bump test where the set-point is altered and the output response to the set-point change is observed. A common closed-loop industrial technique for exciting the process is to simply inject a pseudo-random binary sequence (PRBS) as a perturbation signal. Disrupting the process with perturbation sequences or set-point changes and removing feedback control can lead to lost production time by failing to meet stringent product specifications [63]. Therefore it is imperative to determine an optimal excitation signal that can be implemented in the closed-loop and minimizes the cost of performing an identification experiment. This chapter proceeds by presenting some background material on ID before describing the specific implementation of ID in the simulated case studies.

10.1 Background

Typically, for identification purposes a higher quality estimate involves a lower covariance of estimated parameters. Thus, traditional approaches to ID often involve minimizing some function of the covariance matrix such as the determinant, trace or largest eigenvalues [48]. In this work, instead of minimizing the trace of the covariance matrix, an adaptation of the trace of it's inverse, i.e., the partial Fisher information matrix is maximized. Specifically, a convex quadratic function is maximized while enforcing explicit constraints on the inputs and outputs. A moving horizon predictive (MHP) framework is used to compute a sequence of future input perturbations at each

sample time. The length of this excitation signal is referred to as the prediction horizon.

Outputs are predicted using the most recent process model estimate and predicted outputs are used to construct the partial Fisher information matrix. As will be seen in the experimental results, it is difficult (perhaps even impossible) to guarantee that the inputs and outputs always satisfy their respective constraints. This is due to the presence of output disturbances in addition to inevitable discrepancies between the true process model and the current estimate. As the prediction horizon increases, the potential for constraint violation also increases, which explains why this calculation is repeated at each sample time. The ID technique implemented in this thesis is adapted from the developments made by M. Yousefi et al. in 2015 which can be referred to for a more robust background and theoretical review [62].

The Fisher information matrix is defined as the inverse of the parameter covariance matrix as follows:

$$F = \mathrm{cov}(\theta_N)^{-1} = \Psi\Psi'. \tag{10.1}$$

The matrix Ψ in this part of the thesis is composed of a series of vectors ψ, i.e.,

$$\Psi = \begin{bmatrix} \psi_t & \dots & \psi_{t-N+1} \end{bmatrix}. \tag{10.2}$$

Each vector ψ contains the relevant IO data determined by the orders of the ARX polynomials $A(q,\theta)$ and $B(q,\theta)$, i.e.,

$$\psi_t = \begin{bmatrix} y'_{t-1} & \dots & y'_{t-n_a} & u'_{t-d} & \dots & u'_{t-d-n_b} \end{bmatrix}', \tag{10.3}$$

such that the vector ψ_t and the parameter vector θ model the process output as follows:

$$y_t = \theta\psi_t. \tag{10.4}$$

For a process described by an ARX model the parameter vector θ is given by the following expression:

$$\theta = \begin{bmatrix} a_1 & \dots & a_{n_a} & b_1 & \dots & b_{n_b} \end{bmatrix}. \tag{10.5}$$

The trace of the Fisher information matrix $R_u = \text{trace}(F)$ is defined by

$$R_u = \sum_{p=1}^{n} \sum_{j=1}^{n_a} \sum_{i=j}^{N} (y_p(t-i))^2 + \sum_{q=1}^{m} \sum_{j=1}^{n_b} \sum_{i=j}^{N} (u_q(t-n_k+1-i))^2 \tag{10.6}$$

where n is the number output variables, m is the number input variables, n_a is the order of $A(q,\theta)$, n_b is the order of $B(q,\theta)$, n_k is the input delay and N is the number of IO samples collected [62].

The ID technique applied in this work can be summarized as follows:

$$\begin{aligned} \max_{u_t} \quad & \tilde{R}_u \\ \text{subject to} \quad & y^L \leq y(t) \leq y^H, \\ & u^L \leq u(t) \leq u^H, \\ & |u(t) - u(t-1)| \leq \Delta u^H, \end{aligned} \tag{10.7}$$

where the objective function $\tilde{R}_u$ is defined as

$$\tilde{R}_u = \sum_{q=1}^{m} \left(\frac{1}{\Delta u_q^H} \text{var}(u_q^*) \right). \tag{10.8}$$

The input sequence u_q^* contains all possible combinations of input movements over the specified prediction horizon where each movement is limited to a binary choice of $\pm \Delta u_q^H$. Originally, the trace of the Fisher information matrix defined in (10.6) was maximized subject to the convex input and output constraints. Since this optimization involves maximizing a high dimensional convex quadratic function it can be difficult for quadratic solvers to handle and there is no guarantee that a global optimum is determined.

Since R_u is convex, the maximum over a compact convex domain defined by the convex constraints in (10.7) occurs at one of the vertices, i.e., when the constraints are active. The number of vertices grows exponentially with the dimension of the problem so it was originally thought that a solution based on visiting each vertice individually would be computationally infeasible [62]. However, in this work the number of input variables is $m = 3$, so if the specified prediction horizon is not too large the dimensionality of the problem is small enough to solve via determining the maximum at all of the vertices. The input sequence u_q^* is selected such that only combinations of

inputs that obey the IO boundaries while taking the maximum allowed step size are considered. In other words, of all of the possible combinations of performing binary movements of $\pm\Delta u_q^H$ over the prediction horizon, only those that meet the IO boundaries are evaluated in (10.8). The output boundaries are evaluated based on the predicted outputs resulting from taking the proposed set of input movements. Predicted outputs are generated based on the current process model estimates. The implementation of this new ID method in the experimental case studies is described further in the following section.

10.2 Implementation

After the PM algorithm triggers the start of an identification experiment the ID algorithm begins determining a perturbation sequence to excite the process. Excitation signals are determined during each of the 3000 samples of the identification experiment. Previous input movements extending for a length greater than the sum of the largest time delay and the largest time constant are provided to the ID algorithm. Combinations of all possible future input sequences are determined by computing binary permutations equal to $\pm\Delta u_q^H$ over the prediction horizon. To ensure computational tractability the prediction horizon is limited to a length of 4 steps after which the input value is held constant. This limited prediction horizon is sustainable because the ID calculation is repeated at each new sample. The previous model estimates are supplied to the ID algorithm which uses the impulse response models in combination with past and future input values to predict future outputs.

In the experimental simulations presented in this thesis the maximum input movements are specified as

$$\Delta u_q^H = [\Delta u_1^H, \Delta u_2^H, \Delta u_3^H] = [3.0,\ 1.4,\ 0.6]. \tag{10.9}$$

The input upper and lower bound deviations from the set-point are restricted to

$$u_q^H = -u_q^L = [u_1^H, u_2^H, u_3^H] = [30,\ 15,\ 8], \tag{10.10}$$

and similarly the output upper and lower bound deviations from set-point are defined as

$$y_p^H = -y_p^L = [y_1^H, y_2^H, y_3^H] = [15,\ 1.0,\ 2.0]. \tag{10.11}$$

To conservatively account for both the time delay and process time constant a total of the $n_{u_p} = 111$ most recent input values are provided to the ID algorithm at each iteration. This value is determined by taking the sum of the maximum time constant estimate, the maximum time delay estimate and a contingency factor of 35.

Combinations of future input movements are determined over a 4 step prediction horizon after which the input values are held constant for the remainder of the horizon. The total prediction horizon has a length of 27 which is defined by the maximum time delay estimate plus a contingency factor of 5. Previous plant model estimates are used to generate an impulse response model of the process which is multiplied by both the past and future input values to determine the predicted future outputs. Input sequences that correspond to predicted values of future inputs and outputs that meet the input and output upper and lower bounds are stored and values of $\tilde{R}_u$ in (10.8) are determined. After all of the possible combinations of input sequences have been evaluated the sequence that provides a maximum value of $\tilde{R}_u$ is selected as the perturbation sequence.

This concludes the description of the various elements of the adaptive control framework and their integration in the closed-loop system. System identification, performance monitoring (i.e., MPM detection) and input design algorithms have all been presented. Hopefully the reader now has a grasp of how the various stages of the adaptive framework are implemented and unified. In the following section, the focus shifts to describing the setup of the experimental simulations.

Chapter 11

Experimental Setup

To provide the necessary context for understanding the experimental results, the details of the experimental setup are described here. First, an illustration is provided to describe the general simulation time-line and common setup parameters are defined. This is followed by a review of the distinguishing configuration features for the various experiments. Finally, this chapter concludes by describing the techniques used to analyze the resulting simulation data.

11.1 Experimental time-line and general setup

The experiments presented in this work represent only a small subset of the large number of experiments conducted to integrate the various aspects of the adaptive control framework. The MPM detection algorithm is capable of detecting gain, time constant and time delay mismatch. However, the simulations presented here focus on detecting and resolving a mismatch in the gain of the process. Limiting the focus to a single type of mismatch provides more stringent control of experimental variables to better compare various experimental configurations such as tuning the process with DI or ARX-OE identification estimates. Since the MIMO process is highly coupled any identification experiment in a single IO pairing will result in excitation of all of the process inputs and outputs. For this reason when MPM is determined the identification algorithm simply re-identifies the gain, time constant and time delay values for all of the process models.

A generalized illustration of the experimental simulation time-line is provided in Figure 11.1. This figure is used to guide the discussion of the general simulation setup. A small inaccuracy in Figure 11.1 is that the controller model gain is initially 10% less than the plant gain. One important common feature of all of the experimental simulations is that a $2\times$ gain MPM in b_{11} exclusively is introduced at sample number $n_{MPM} = 4000$. The remaining gain and time constant values of the true process are held constant along with the controller model. Another common element is that

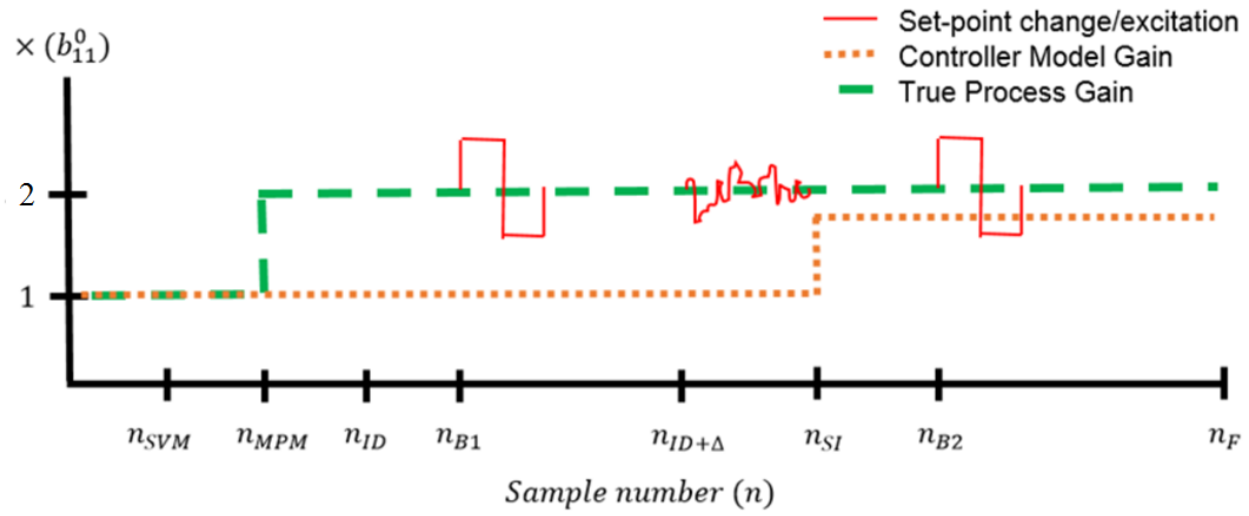

Figure 11.1: Illustrative time-line of the general setup

the SVM is trained for a period of $n_{SVM} = 3500$ samples at which point MPM detection becomes active. The sample at which MPM is detected depends on the IO data and SVM models but it is described here as n_{ID}. A user specified delay $n_{ID+\Delta} - n_{ID} = 1600$ separates the MPM detection point and the actual ID excitation experiment. Typically the ID experiment would start at n_{ID}, however the delay provides the time necessary to perform set-point changes and collect response data.

After a pre-specified wait period $n_{B_1} - n_{MPM} = 500$, a set-point change at sample n_{B_1} increases the set-point of y_1 by 10% of y_1^o. This set-point is held for 300 samples before it is decreased by twice as much, held for another 300 samples and then returned to the original set-point y_1^o. This series of set-point changes provides information on the dynamic response of the closed-loop system under MPM which is used to evaluate controller performance. An identical procedure is used at sample n_{SI} after the controller has been re-tuned. This allows us to compare the dynamic closed-loop responses both with and without adaptive tuning.

A summary of the general setup parameters is provided in Table 11.1 below. Controller parameters α^{sp} and α^d are used to tune the MPC for reference tracking performance and disturbance rejection performance, respectively. The ID algorithm is compared to using PRBS excitation signals of magnitude $u_q^{PRBS} = \pm$ [2.0, 1.0, 0.5]. Since the PRBS does not constrain the input and output values the magnitudes of the PRBS are conservatively less than Δu_q^H. The remaining experimental setup parameters depend on the various test configurations and are outlined in the following section.

Table 11.1: General experimental setup parameters

Symbol	Value	Description
n_F	14000	Simulation duration in samples
n_{MPM}	4000	Sample to trigger MPM
α^{sp}	1.5	Controller reference tracking
α^{d}	1.5	Controller disturbance rejection
y_p^o	[174, 2.3, 6.2]	Nominal output set-points
u_q^o	[165, 115, 35]	Nominal input set-points
$\Delta y_p^o\%$	[10, 0, 0]	Magnitude of set-point change
$n_{ID+\Delta} - n_{ID}$	1600	Delay after MPM detection
$n_{B1} - n_{MPM}$	500	Wait period before set-point change
$n_{\Delta y^o}$	300	Samples to hold each set-point change
H_1, H_2, H_3	1^{st} or 6^{th} order	High pass noise models
u_q^{PRBS}	$\pm$[2.0, 1.0, 0.5]	Magnitude of PRBS excitation
σ_p^e	[0.08, 0.02, 0.01]	Standard deviation of noise

11.2 Experimental configurations

As mentioned before, the primary interest of conducting the experiments is to determine how the controller performs before and after adaptive tuning. Analyzing the dynamic response of the closed-loop system helps to evaluate the effectiveness of the adaptive control framework. For a particular configuration, this effectiveness can be observed with a single simulation. However, it is also desirable to observe which configurations maximize the effectiveness of the adaptive framework. Specifically, it is important to analyze the effectiveness of the ID algorithm relative to a PRBS sequence and determine whether the DI method or the ARX-OE identification method provides the best controller performance after re-tuning.

In this report six separate experiments are performed. The first four experiments are designed to investigate whether the ID sequence is more or less effective than a simple PRBS. The last two tests are conducted to evaluate the effect of incorrect noise model order specification for the DI method. Another way to understand the series of tests is as a set of 3 trials each of which tests both the DI method and the ARX-OE method. The first trial is conducted with PRBS excitation, the second trial uses the ID algorithm and the third trial uses the ID algorithm but with a different type of noise model. A sixth order noise model is used in the third trial while the DI method still assumes that the noise model is first order. The noise models in all three trials are high pass (HP) filters. A general description of each test is provided in Table 11.2 below.

Table 11.2: Various experimental configurations

Trial/test No.	Description
1.1	PRBS excitation, ARX-OE controller tuning, 1st order HP noise
1.2	PRBS excitation, DI tuning, 1st order HP noise
2.1	ID excitation, ARX-OE tuning, 1st order HP noise
2.2	ID excitation, DI tuning, 1st order HP noise
3.1	ID excitation, ARX-OE tuning, 6th order HP noise
3.2	ID excitation, DI tuning, 6th order HP noise

The first order noise models used in the first four tests are all HP filters defined as follows:

$$\begin{aligned} H_1(q) &= \frac{1-0.5q^{-1}}{1+0.9q^{-1}}, \\ H_2(q) &= \frac{1+0.3q^{-1}}{1+0.6q^{-1}}, \\ H_3(q) &= \frac{1+0.2q^{-1}}{1+0.7q^{-1}}. \end{aligned} \tag{11.1}$$

Similarly, the sixth order noise models in the final two tests are all HP filters defined as

$$\begin{aligned} H_1(q) &= \frac{1-0.9q^{-1}+0.4q^{-2}-0.3q^{-3}-0.2q^{-4}+0.1q^{-5}+0.1q^{-6}}{1+0.9q^{-1}+0.6q^{-2}+0.3q^{-3}+0.2q^{-4}+0.1q^{-5}-0.1q^{-6}}, \\ H_2(q) &= \frac{1-0.9q^{-1}+0.4q^{-2}-0.2q^{-3}-0.1q^{-4}+0.1q^{-5}+0.1q^{-6}}{1+0.9q^{-1}+0.5q^{-2}+0.2q^{-3}+0.2q^{-4}+0.1q^{-5}-0.1q^{-6}}, \\ H_3(q) &= \frac{1-0.7q^{-1}+0.3q^{-2}-0.2q^{-3}-0.1q^{-4}+0.1q^{-5}+0.1q^{-6}}{1+0.6q^{-1}+0.4q^{-2}+0.3q^{-3}+0.2q^{-4}+0.1q^{-5}-0.1q^{-6}}. \end{aligned} \tag{11.2}$$

These noise models were selected based on their frequency response characteristics and the fact that they are all stable and inversely stable. The techniques used to analyze the experimental data are described in what follows.

11.3 Experimental analysis

As mentioned before, the dynamic output response of the closed-loop system is compared for each test, before and after adaptive tuning. Furthermore, for each test the DI and ARX-OE identification estimates are compared to the true plant values. Apart from within each test, it is also desirable to compare the responses and parameter estimates across the various testing configurations. The dynamic response data is evaluated both qualitatively and quantitatively. The qualitative assessment involves observing the output response shape to the set-point changes and the quantitative

assessment involves comparing the observed response to the ideal response. The ideal response shape is generated based on the true process model. To quantify the fit of the response to the ideal response the mean square error (MSE) is calculated as follows:

$$\epsilon = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2, \tag{11.3}$$

where n is the number of data points collected during the dynamic response test and $\tilde{y}$ represents the values of the ideal response shape. Finally, the parameter estimates are reported in terms of their percent error (PE), i.e.,

$$PE_{ij}(\%) = 100 \times \frac{|\hat{b}_{ij} - b_{ij}|}{|b_{ij}|}, \tag{11.4}$$

for some parameter estimate, $\hat{b}$, and true value, b.

Chapter 12

Experimental Results

This chapter contains three trials each of which has two tests. The first trial explores the use of a PRBS excitation signal with both DI and ARX-OE identification for controller tuning. The second trial applies the same two tests but now the ID algorithm is used to replace the PRBS. Finally, the third trial involves changing the noise model order to test whether an incorrectly specified noise model order will be detrimental to the DI method.

12.1 Trial 1: PRBS excitation

The first trial applies PRBS excitation signals for the identification experiment. Using a PRBS for excitation provides a reference for evaluating the ID algorithm. A first order high pass filter is implemented as the noise model. Two tests are conducted to evaluate the difference between tuning the controller with the ARX-OE method or the DI method.

12.1.1 Test 1.1: Tuning with ARX-OE identification

The 2× gain MPM is detected by the PM algorithm after only 860 samples, i.e., $n_{ID} = 4860$. Since a delay of 1600 samples is used the PRBS excitation experiment begins at sample number $n_{ID+\Delta} = 6460$. After the PRBS experiment an ARX-OE identification is performed and the estimated parameters are used to tune the controller. The raw input profiles for this test are shown below in Figure 12.1. It is apparent from these input profiles that the set-point change in y_1 is causing undesirable behavior in u_2 and u_3. This input behavior ultimately degrades the output responses of y_2 and y_3, as seen in Figure 12.2. From Figure 12.2 it is clear that a generous amount of PRBS excitation is used to excite the system. For y_1 the response improves after the controller is re-tuned but the responses of y_2 and y_3 appear to degrade. Once the controller is re-initialized y_2 and y_3 are unable to stabilize before the start of the second set-point change.

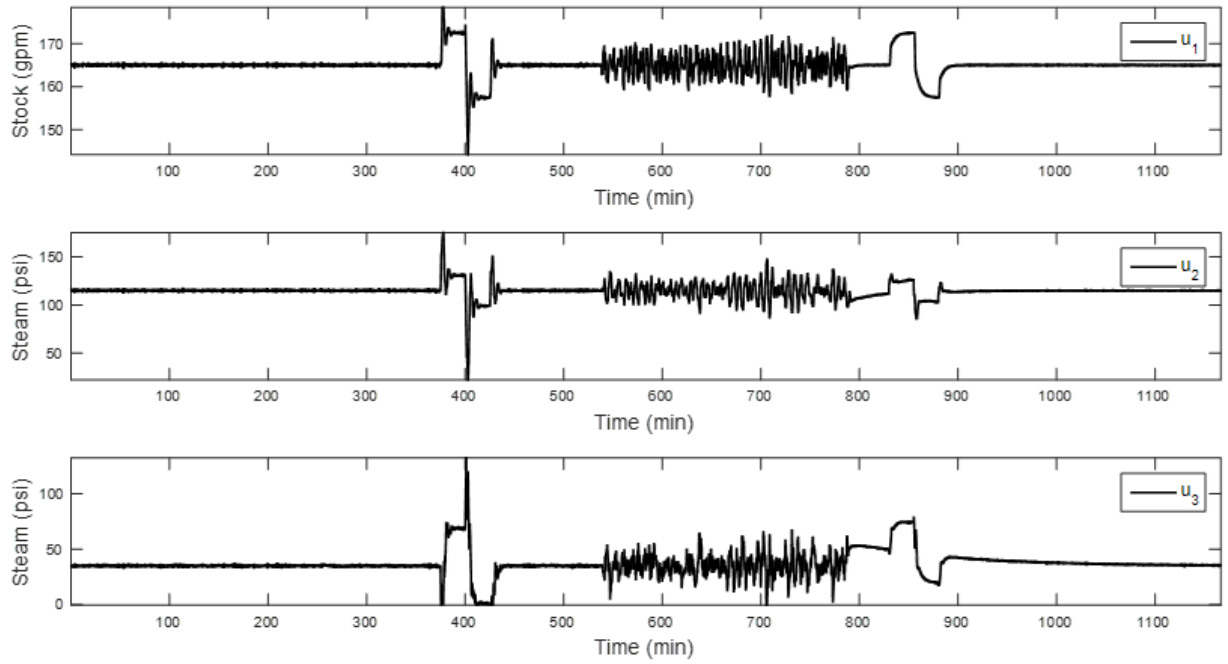

Figure 12.1: Raw input profiles for Test 1.1

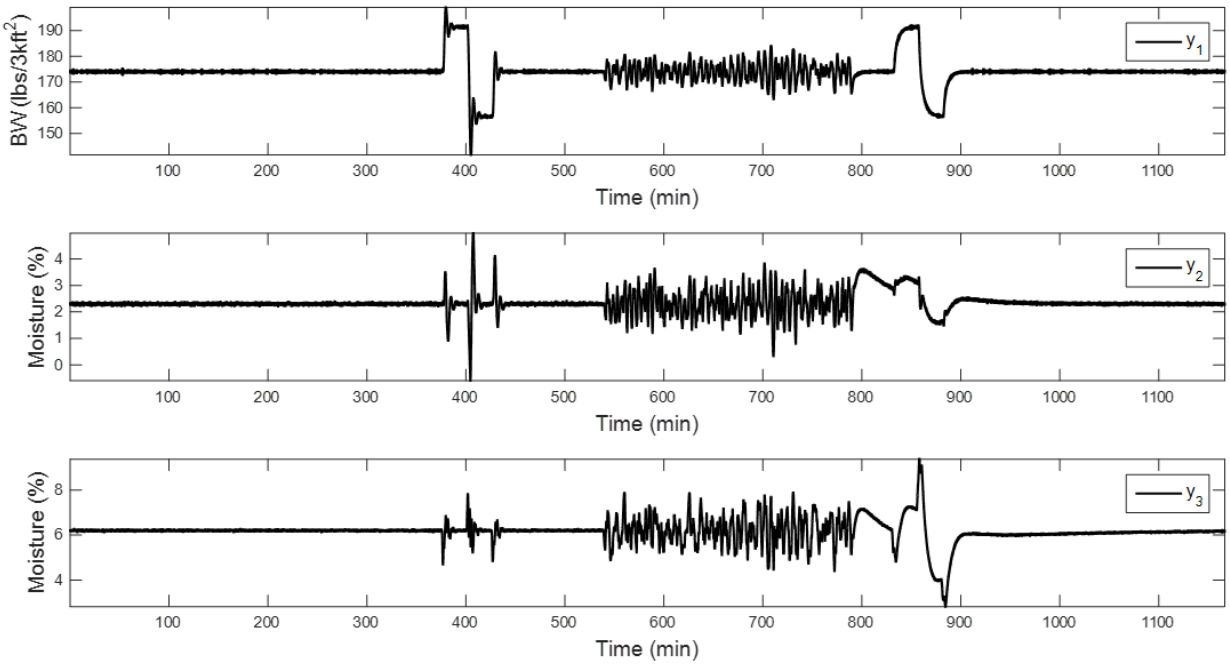

Figure 12.2: Raw output profiles for Test 1.1

Zooming in on the process outputs during the first set-point change yields Figure 12.3. It is immediately clear that the closed-loop dynamic response under MPM exhibits undesirable overshoot and oscillatory behavior. In fact, the responses of both y_2 and y_3 also exhibit oscillations

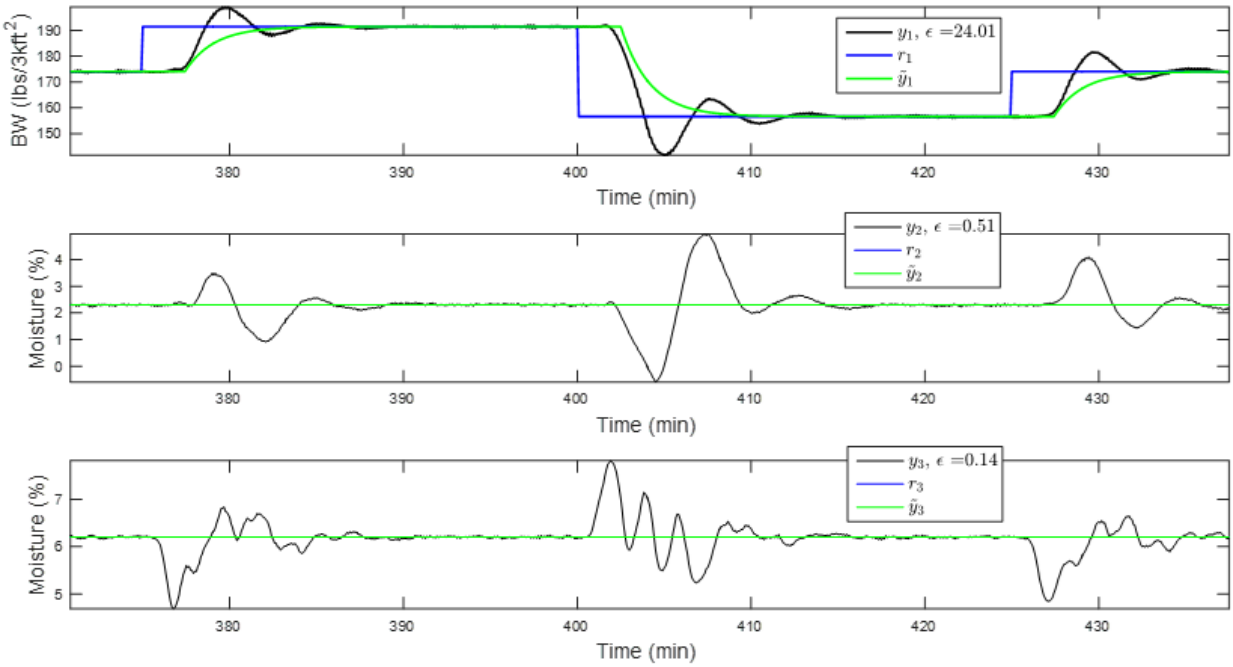

Figure 12.3: Dynamic response with MPM for Test 1.1

even though they do not undergo set-point changes. After the controller is tuned using the ARX-OE model estimates, another set-point change is performed. The output response to this set-point change is shown in Figure 12.4. Clear differences exist between the dynamic responses before

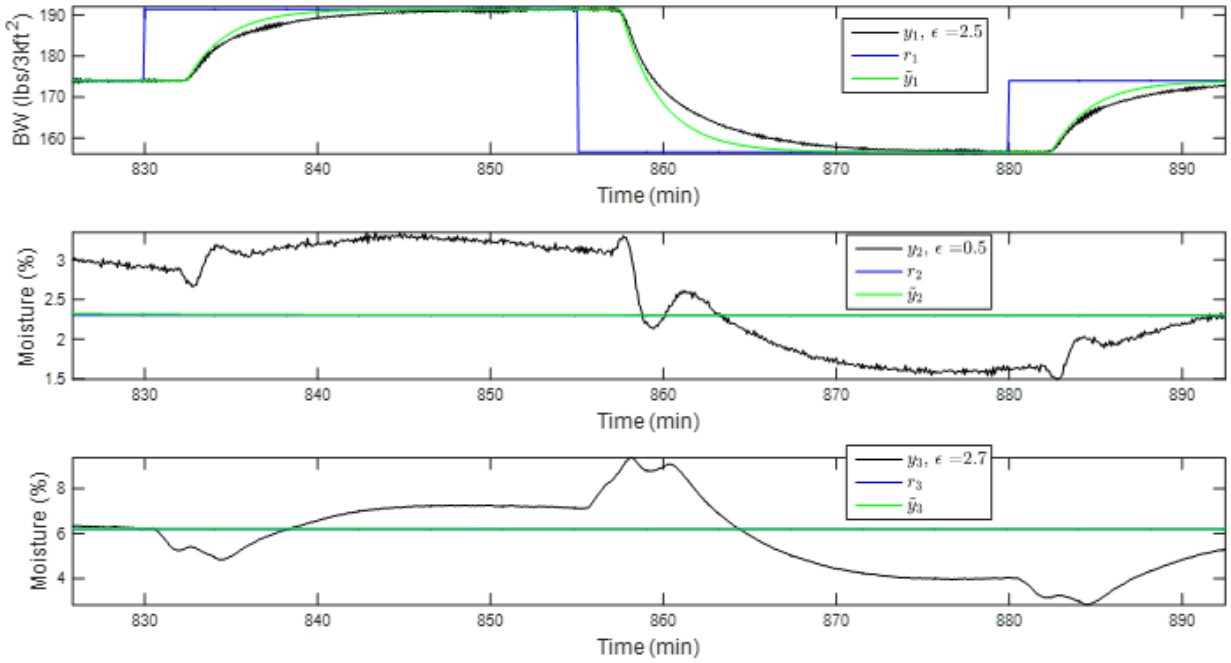

Figure 12.4: Dynamic response after tuning for Test 1.1

and after adaptive tuning. For output y_1 the response after tuning exhibits essentially no overshoot or oscillation which is in stark contrast to the response with MPM. One potential downside to the adaptive tuning response is that it appears more sluggish than the response with MPM. Furthermore, it appears that re-identifying the process models that did not contain MPM has had a detrimental effect on the response of outputs y_2 and y_3. However, quantitatively as shown by the MSE values in the legend, generally the adaptive tuning improves control. Only for the response of y_3 does the MSE increase after re-tuning. The MSE for y_1 which is the variable of particular interest is reduced dramatically.

To shed some light on the dynamic response behavior the percent error (PE) of the DI and ARX-OE identification parameter estimates are presented in Table 12.1 and listed as (G_{11}, G_{21}, G_{22}, G_{31}, G_{32}, G_{33}). It is immediately clear that the ARX-OE time constant estimates for the G_3

Table 12.1: Percent error of DI and ARX-OE estimates for Test 1.1

Method	Parameter	PE (%)
Both	Time Delay	(0, 0, 4.3, 11.1, 14.3, 21.1)
Direct	Gain	(13.9, 25.5, 100.5, 100.4, 3.8, 33.3)
Direct	Time Constant	(11.6, 26.2, 67.5, 134.3, 94.3, 480.7)
ARX-OE	Gain	(28.7, 36.8, 117.5, 113.2, 34.7, 537.5)
ARX-OE	Time Constant	(26.7, 36.8, 67.5, 117.6, 114.1, 1720.7)

transfer functions are very poorly estimated. In fact, the ARX-OE gain estimates and the DI time constant and gain estimates for the G_3 transfer functions are all significantly biased. Note that the poorly estimated parameters generally correspond to transfer functions that have incorrect time delay estimates. It is likely that part of the error in the time constant estimates is compensating for the incorrect time delay specification. Ultimately the majority of these estimates are too inaccurate for reliable controller tuning. This helps to explain the strange behavior observed in Figure 12.2 for outputs y_2 and y_3 after the controller is re-initialized. It is important to note that although the estimated parameters exhibit considerable inaccuracy it is entirely possible that multiple arrangements of parameter values for the complex process model can yield similar response behaviors. This is why it is important to compare the dynamic response of the process and not simply base the analysis off of the parameter bias. Next a similar experiment is performed except now the controller is updated using the DI estimates.

12.1.2 Test 1.2: Tuning with direct identification

The MPM detection algorithm successfully identifies the mismatch at the same point as the previous test, i.e., after 860 samples at $n_{ID} = 4860$ samples. Once again the PRBS excitation experiment begins at $n_{ID+\Delta} = 6460$. After the PRBS experiment the DI parameter estimates are used to re-tune the controller. The raw input profiles for this test are shown below in Figure 12.5. Again, these

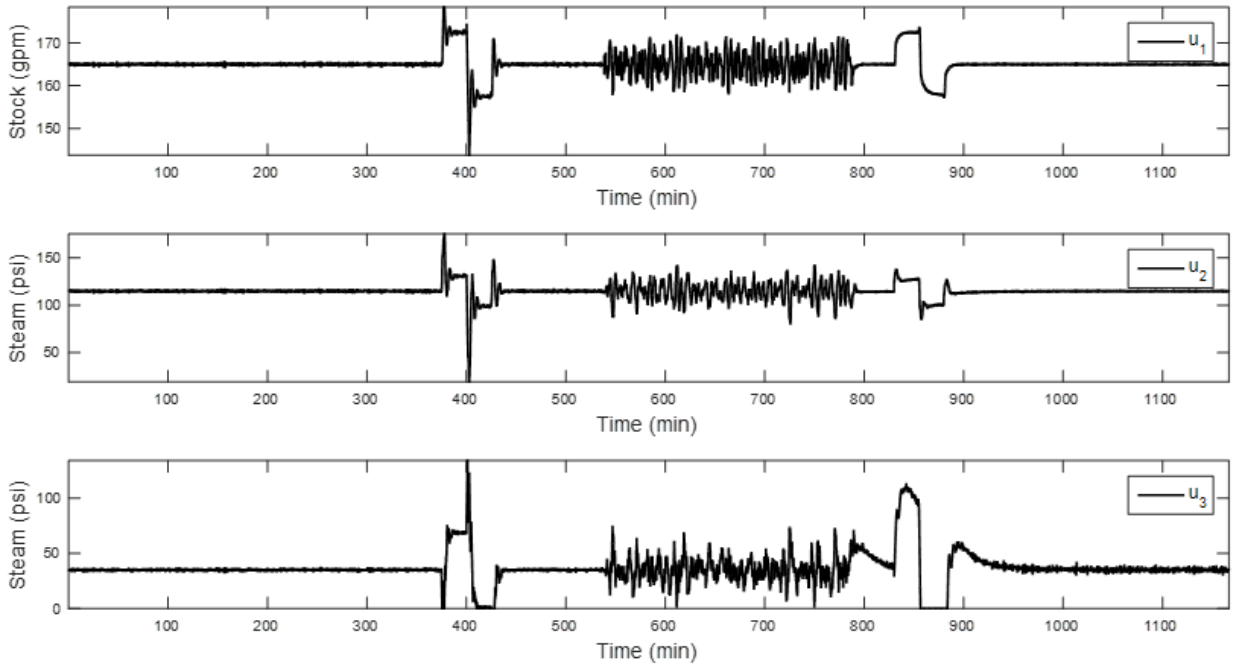

Figure 12.5: Raw input profiles for Test 1.2

input profiles show the set-point change in y_1 causing undesirable behavior in u_2 and u_3. Observing the excitation experiment between 550 and 800 minutes in Figure 12.5 shows a significant amount of PRBS excitation that becomes problematic as the input u_3 clearly exceeds acceptable boundaries. As shown in Figure 12.6, once the controller is tuned the output y_3 is unable to stabilize before the second set-point change. Recall, y_1 is the basis weight (BW) measured in lbs/(3000ft^2), y_2 is the press moisture (%) and y_3 is the reel moisture (%). The responses of the basis weight and press moisture improve after adaptive tuning but the reel moisture response to the set-point change is significantly worse after tuning.

The output responses from the first set-point change are presented more clearly in Figure 12.7. During this set-point change 2× gain MPM exists in G_{11} and the other controller parameters are 10% less than the true plant values. As before, the BW response to the set-point change displays

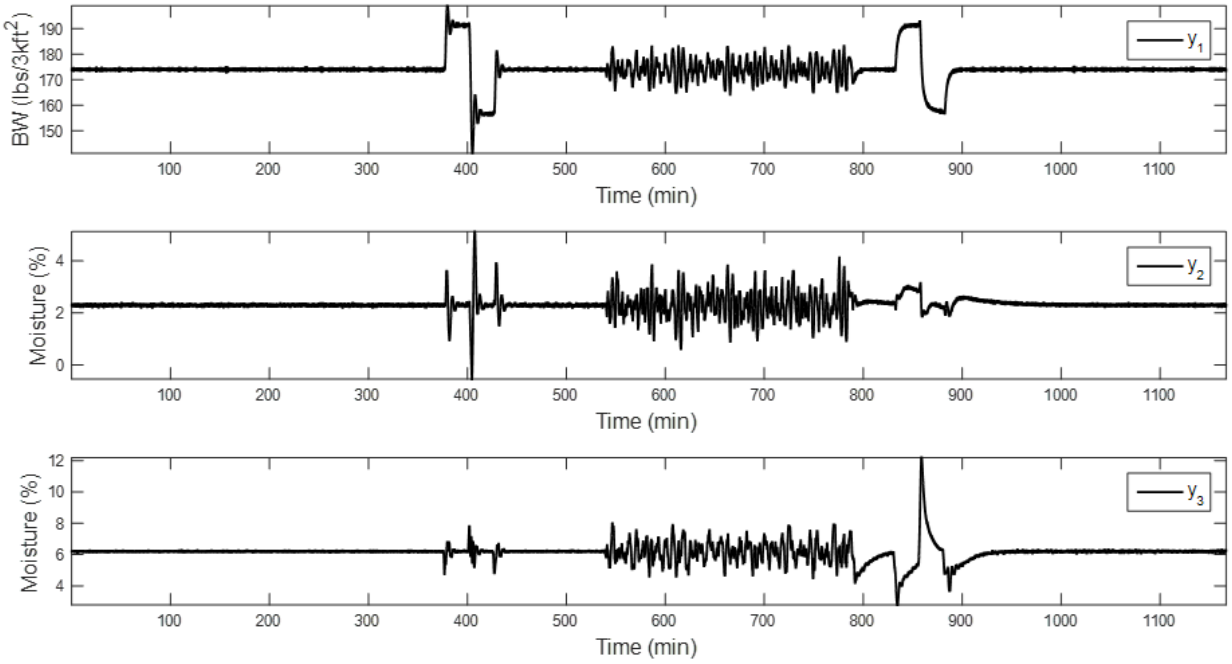

Figure 12.6: Raw output profiles for Test 1.2

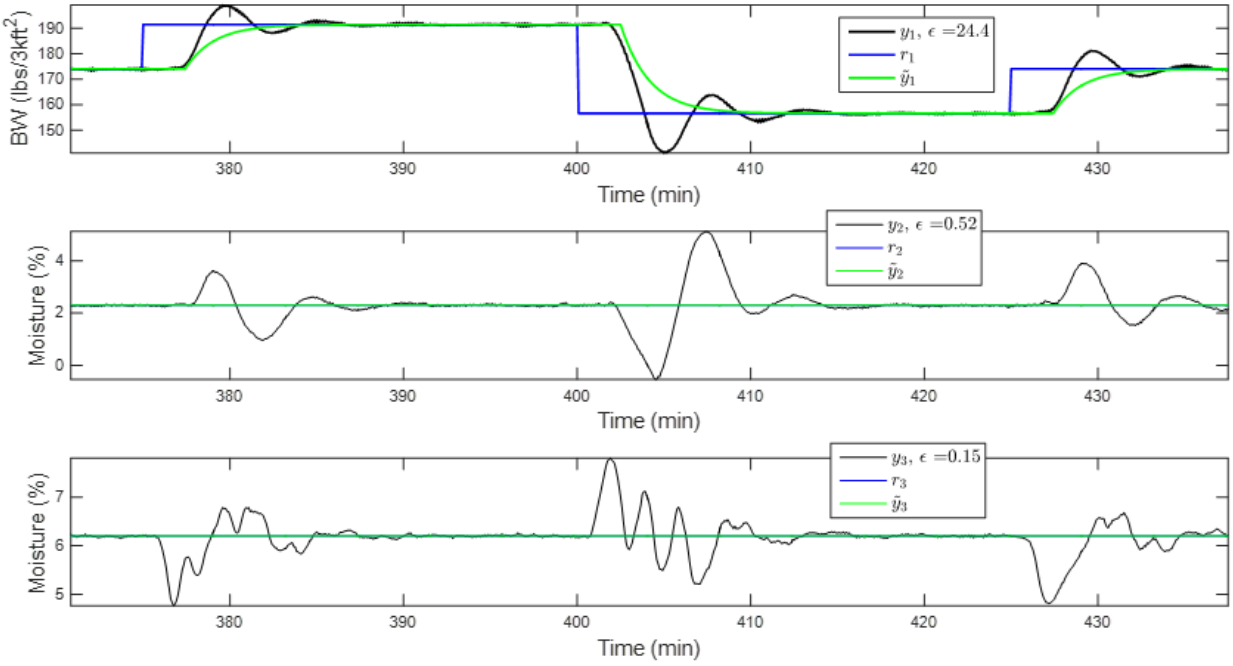

Figure 12.7: Dynamic response with MPM for Test 1.2

significant overshoot and oscillatory behavior resulting in a large MSE of $\epsilon = 24.4$. Due to the MIMO coupling the BW set-point change disrupts the press and reel moisture values as well.

After the controller is tuned using the DI estimate, another set-point change is performed. The

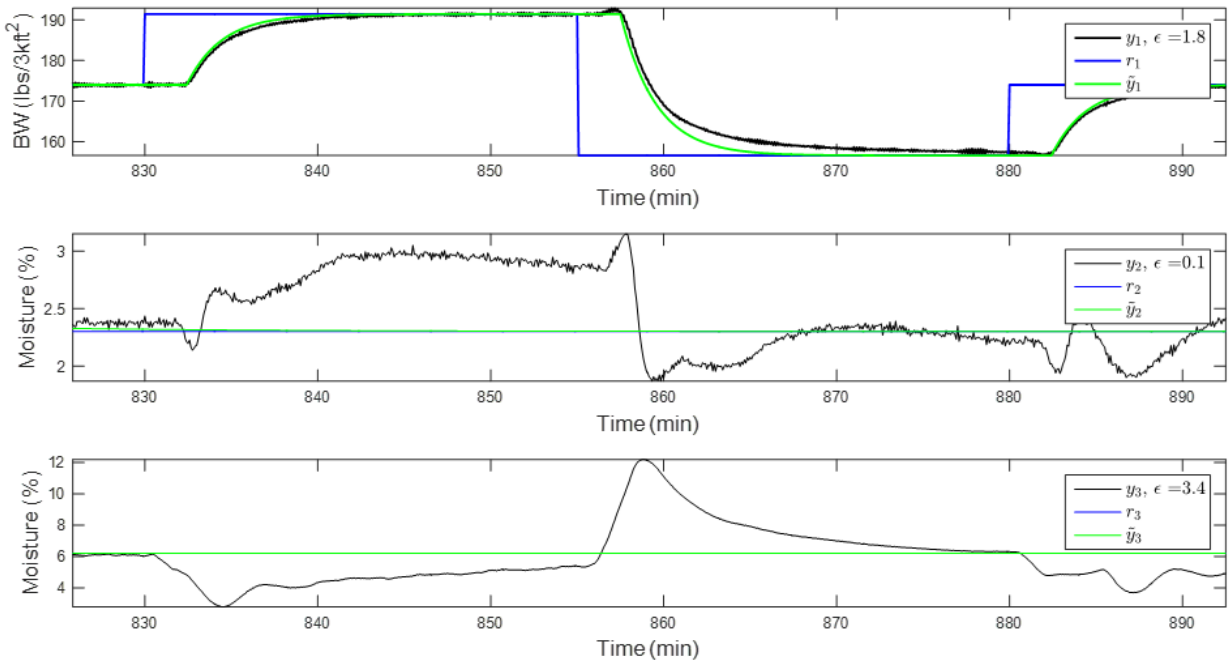

Figure 12.8: Dynamic response after tuning for Test 1.2

output response to this set-point change is shown in Figure 12.8. Comparing the responses before and after adaptive tuning shows once again that the adaptive tuning improves the BW response to set-point changes. The MSE of the BW dynamic response is reduced from 24.4 to 1.8 as a result of the adaptive tuning. Furthermore, the MSE of the press moisture is reduced after tuning but the MSE of the reel moisture increases substantially. The MSE values are presented in the legends of Figure 12.7 and Figure 12.8. Comparing Figure 12.8 to Figure 12.4 shows the effect of switching to DI from ARX-OE identification. It is clear both visually and by comparing the MSE values that all three of the dynamic output responses are better with DI than with ARX-OE identification for this trial.

Finally, the PE of the DI and ARX-OE identification parameter estimates are listed from transfer function G_{11} to G_{33} and presented in Table 12.2. As before, these PE values indicate the DI estimates

Table 12.2: Percent error of DI and ARX-OE estimates for Test 1.2

Method	Parameter	PE (%)
Both	Time Delay	(0, 0, 0, 11.1, 57.1, 26.3)
Direct	Gain	(13.7, 27.6, 107.8, 95.7, 26.5, 3.2)
Direct	Time Constant	(11.3, 29.5, 67.5, 149.0, 114.1, 688.4)
ARX-OE	Gain	(22.1, 34.1, 118.7, 58.0, 14.4, 75.2)
ARX-OE	Time Constant	(19.9, 34.2, 67.5, 79.5, 114.1, 1324.6)

are generally more accurate than the ARX-OE estimates. This helps to explain the reduction in MSE observed by using the DI method. Again, the PE of the time constant estimates are generally much higher if the time delay was incorrectly identified. Transfer function G_{22} and the G_3 transfer function parameter estimates exhibit particularly poor identification accuracy. In the following trial the PRBS excitation signal is replaced by the ID algorithm to determine whether or not better results can be obtained.

12.2 Trial 2: Input design, first order noise

The only difference between this trial and the previous one is that instead of using a PRBS excitation signal the ID algorithm described in Chapter 10 is used. Once again first order HP filter noise models are used and two tests are performed to compare the DI and ARX-OE identification methods.

12.2.1 Test 2.1: Tuning with ARX-OE identification

Only 740 samples after the gain mismatch is introduced, the MPM detection algorithm triggers the identification experiment which begins after a delay of 1600 samples at $n_{ID+\Delta} = 6340$. After the ID experiment an ARX-OE estimate is used to tune the controller. The raw input profiles for this test are shown in Figure 12.9. The input values during the excitation period (between 500 and 800 minutes) can be compared to the raw inputs from the previous trial to see the difference between using the ID algorithm or the PRBS excitation. The thick stock flow (u_1) variation during the excitation experiment ranges primarily between 160 gpm and 170 gpm and thus satisfies the required constraints of 165 ± 30 gpm. Variation of the press steam pressure u_2 is restricted to 115 ± 15 psi which is approximately satisfied for the duration of the excitation experiment. Finally, the dryer steam pressure (u_3) is constrained between 35 ± 8 psi but these constraints are violated a few times during the excitation experiment, particularly around time 550 min. Relative to using a PRBS as in the previous trial the dryer steam pressure is well-behaved. In general the PRBS excitation created larger and more frequent input constraint violations than the ID algorithm.

Similarly, each of the process outputs conform to the specified upper and lower bound constraints during the ID excitation experiment. This is in contrast to the PRBS excitation which occasionally resulted in violations of the output constraints. The raw output profiles for this test are shown in

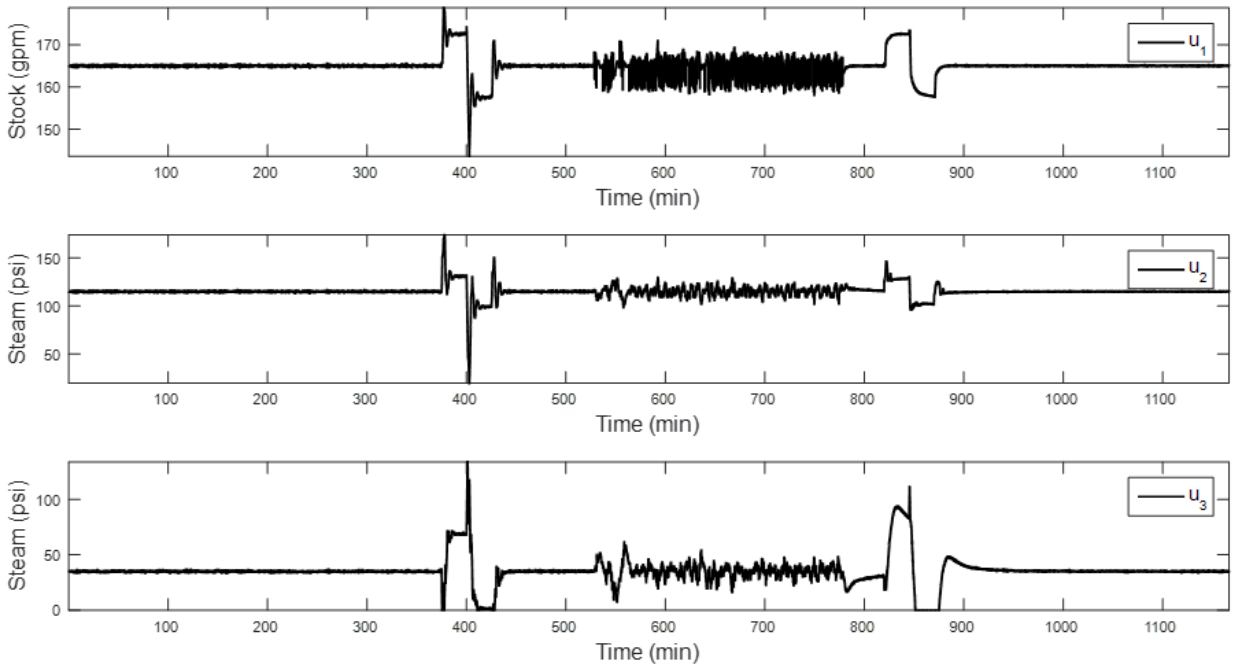

Figure 12.9: Raw input profiles for Test 2.1

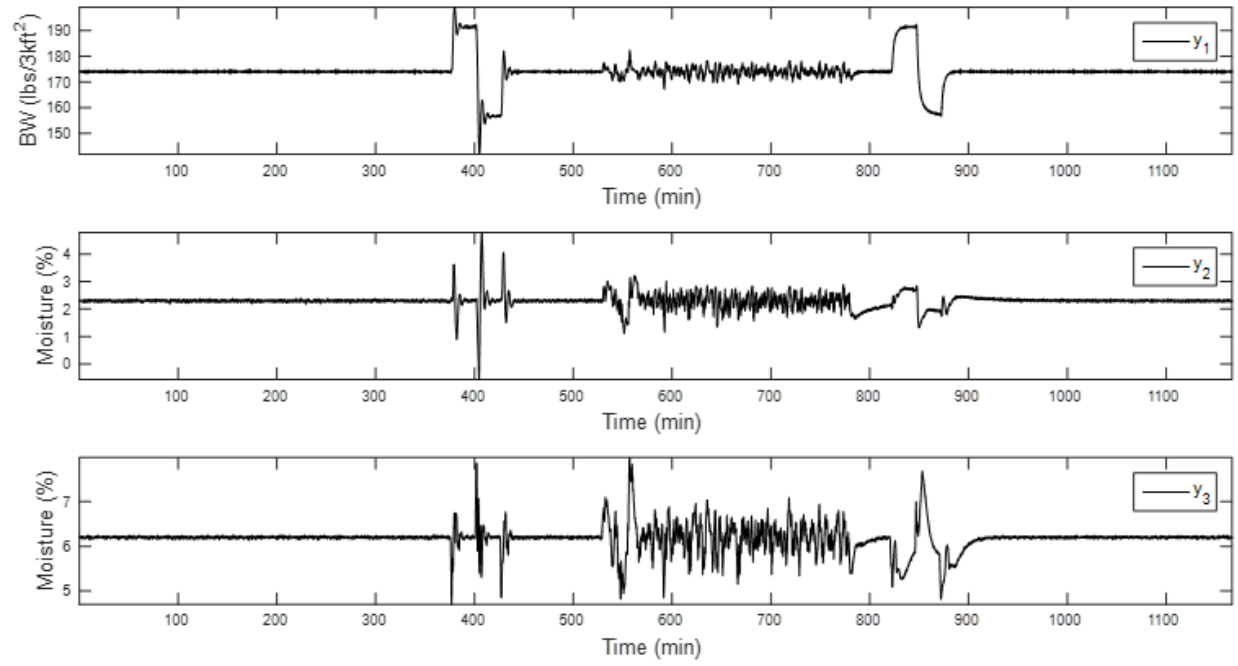

Figure 12.10: Raw output profiles for Test 2.1

Figure 12.10. Comparing the press moisture (y_2) profile before and after adaptive tuning shows the magnitude of the disruption due to the set-point change is reduced after adaptive tuning.

Zooming in on the process outputs during the first set-point change allows for a clearer assess-

ment of the dynamic output responses. Figure 12.11 shows the typical oscillatory and over-shoot behavior seen when MPM exists. After the controller is tuned using the ARX-OE model estimate,

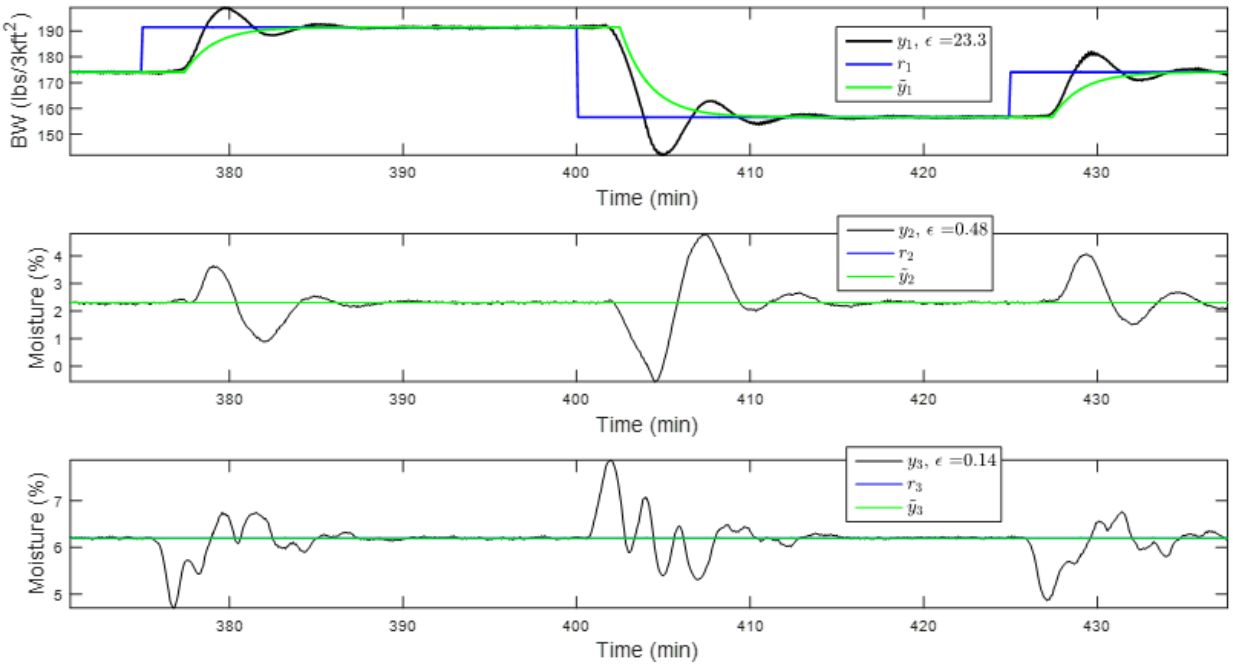

Figure 12.11: Dynamic response with MPM for Test 2.1

another set-point change is performed. The output response to this set-point change is shown in Figure 12.12. The BW response is much closer to the ideal response (shown in green) after the controller is tuned. Both the BW and press moisture (y_1 and y_2) have a lower MSE after tuning whereas the reel moisture is slightly increased. The MSE of the reel moisture is much smaller using the ID algorithm than it was in the previous trial with PRBS excitation.

The PE of the DI and ARX-OE parameter estimates are listed in Table 12.3 from transfer function G_{11} to G_{33}. Once again some of the time delay estimates are incorrectly estimated which appears to effect the time constant estimates, particularly for G_{31}. The ARX-OE gains for G_{22} and G_{31} are poorly estimated but otherwise the majority of the gain estimates for both the DI and ARX-OE methods are reasonably accurate. Comparing these values to those in Table 12.2 demonstrates that the ID algorithm provides significantly better estimates (in terms of PE) than the PRBS excitation. So far it appears as if the ID algorithm provides better identification estimates and improved controller performance than the PRBS excitation.

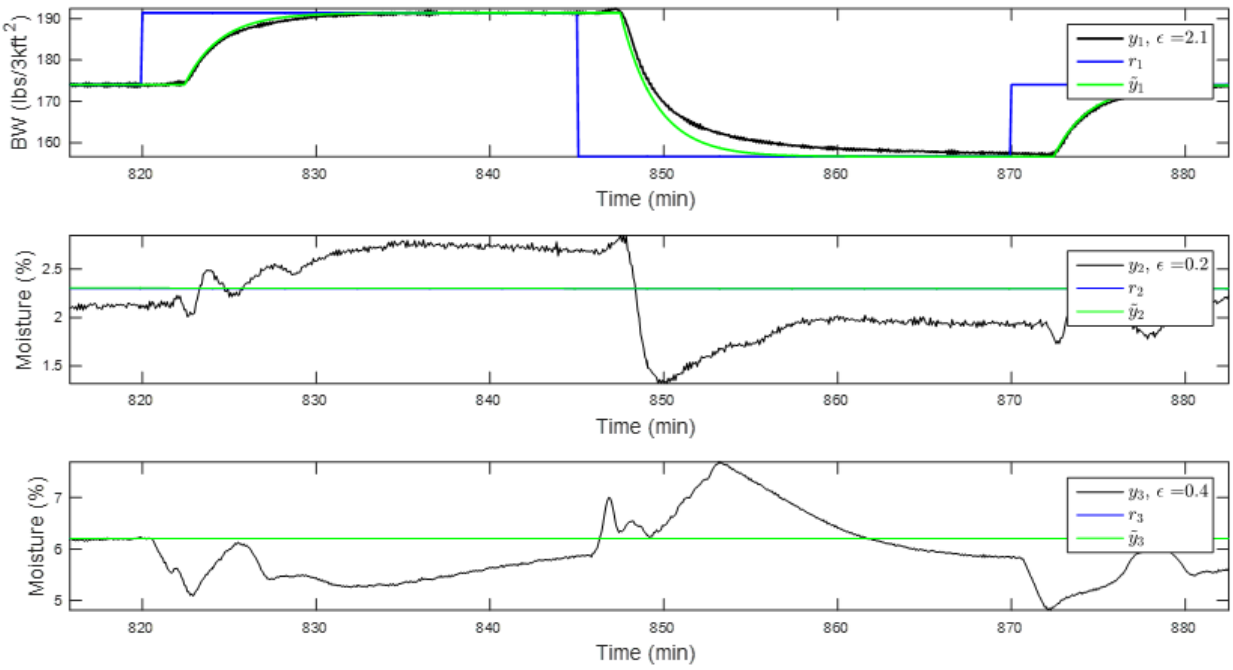

Figure 12.12: Dynamic response after tuning for Test 2.1

Table 12.3: Percent error of DI and ARX-OE estimates for Test 2.1

Method	Parameter	PE (%)
Both	Time Delay	(0, 0, 4.3, 22.2, 71.4, 0)
Direct	Gain	(6.9, 7.9, 7.6, 41.8, 37.1, 28.3)
Direct	Time Constant	(6.2, 13.3, 2.4, 87.1, 16.8, 11.9)
ARX-OE	Gain	(12.7, 58.2, 116.1, 97.7, 23.2, 1.5)
ARX-OE	Time Constant	(11.5, 72.9, 67.5, 192.2, 27.0, 17.6)

12.2.2 Test 2.2: Tuning with direct identification

The 2× gain MPM is detected by the PM algorithm 860 samples after it is introduced and the identification experiment begins at sample number $n_{ID+\Delta} = 6460$. As before the ID technique is used to perform the excitation but this time the DI method is used to tune the controller. The input profiles for this test are shown below in Figure 12.13. For the most part the excitation experiment appears very similar to the previous test. Roughly speaking the input constraints appear to be approximately met with the ID excitation. The only significant constraint violations are due to the set-point changes in y_1. Throughout the excitation experiment each of the output variables satisfies the upper and lower bound constraints. The raw output profiles are presented in Figure 12.14. Once again the press moisture (y_2) is much less disrupted by the set-point change after the adaptive tuning is performed.

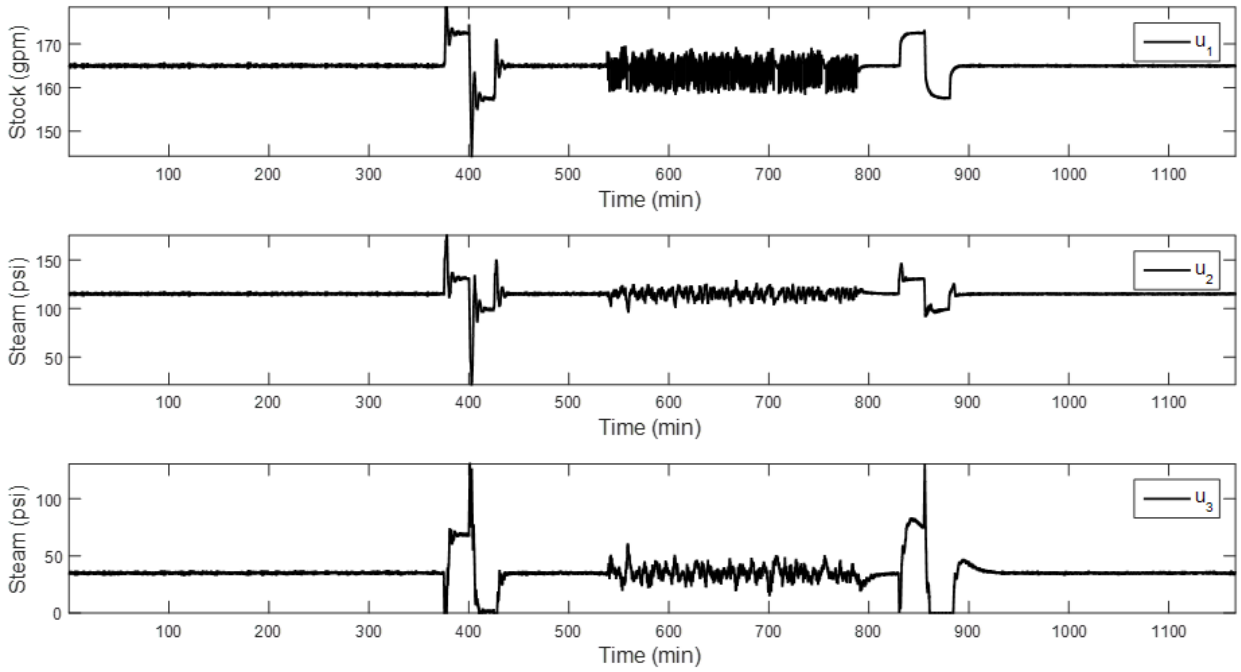

Figure 12.13: Raw input profiles for Test 2.2

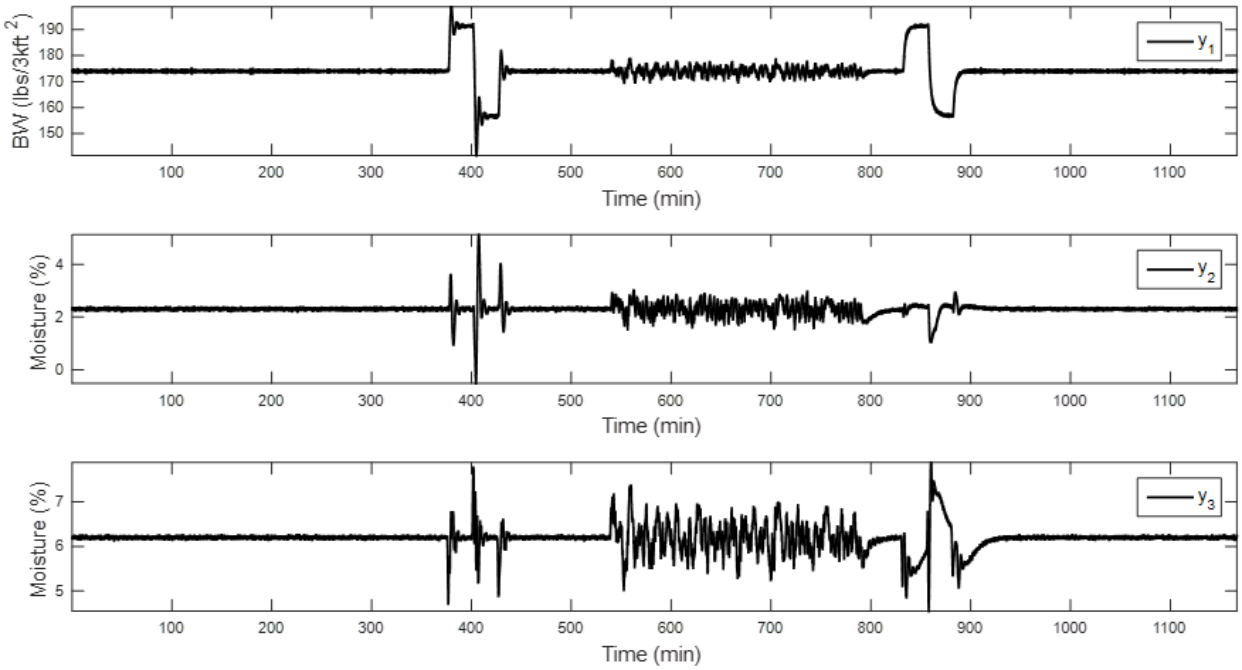

Figure 12.14: Raw output profiles for Test 2.2

A closer look at the dynamic response to the first set-point change (i.e., under MPM) is provided in Figure 12.15. The typical oscillatory response with over-shoot is observed once again, particularly for the BW profile. In each of the last four tests the output responses under MPM have been very

similar which is expected considering the simulation is equivalently configured up until the excitation experiment. After the controller is tuned using the DI estimates, another set-point change is carried

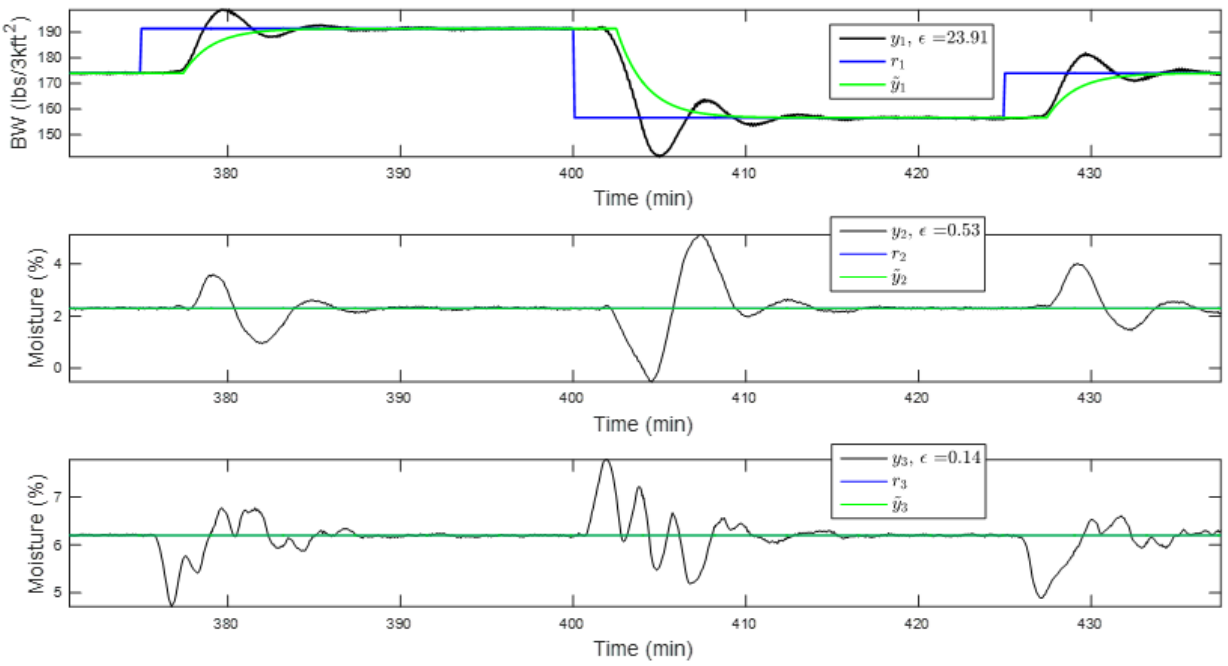

Figure 12.15: Dynamic response with MPM for Test 2.2

out. The output response to this set-point change is shown in Figure 12.16 and is significantly better than before adaptive tuning. Specifically the BW response is essentially equivalent to the ideal BW response, represented by the green profile. Comparing the BW response to the previous test (i.e., tuning with ARX-OE estimates) shows that the DI estimate has resulted in a slightly better dynamic response. The press moisture is also slightly improved with the DI estimate but the reel moisture is slightly worse. The difference between tuning with either the DI or ARX-OE identification estimates is fairly negligible but if anything the DI method appears slightly better.

Finally, the PE of the DI and ARX-OE identification parameter estimates are listed as (G_{11}, G_{21}, G_{22}, G_{31}, G_{32}, G_{33}) and presented in Table 12.4. Once again the time delays for some of the transfer functions are incorrectly estimated which appears to effect the corresponding time constant estimates. The DI gain estimates are fairly accurate and so are most of the ARX-OE identification estimates. So far the DI estimates have generally had lower PE values than the ARX-OE identification estimates and the ID excitation has resulted in lower PE values than the PRBS excitation. The following trial studies the effect of changing the order of the noise models.

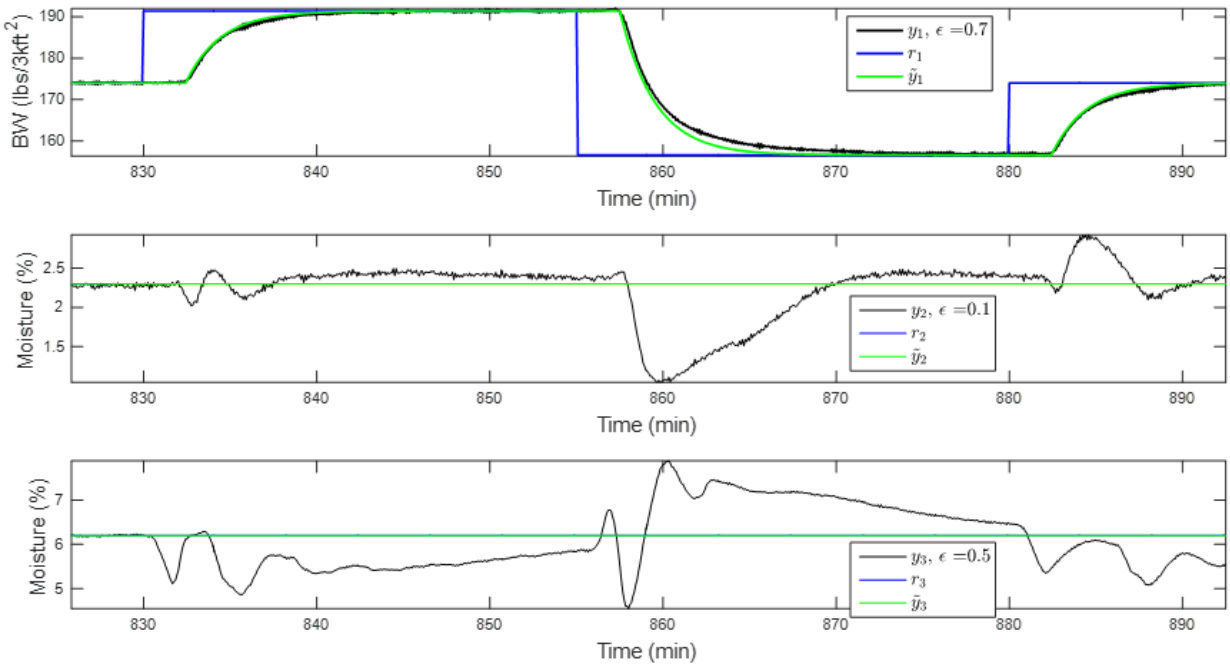

Figure 12.16: Dynamic response after tuning for Test 2.2

Table 12.4: Percent error of DI and ARX-OE estimates for Test 2.2

Method	Parameter	PE (%)
Both	Time Delay	(0, 0, 0, 22.2, 42.9, 15.8)
Direct	Gain	(8.6, 6.2, 16.3, 35.1, 52.6, 13.3)
Direct	Time Constant	(10.4, 11.7, 12.5, 192.2, 31.4, 193.3)
ARX-OE	Gain	(13.2, 59.2, 108.4, 97.7, 19.2, 81.8)
ARX-OE	Time Constant	(12.1, 77.7, 67.5, 192.2, 114.1, 447.0)

12.3 Trial 3: Input design, sixth order noise model

Once again the ID technique is applied in this trial however now the prediction horizon is changed from a length of four to a length of three. From the previous two trials it was determined that the ID technique provides a more controlled excitation and better system identification results than a simple PRBS excitation. Furthermore the DI method so far appears to have slightly out-performed the ARX-OE identification method. The previous trial implemented a first order noise model so it was expected for the DI method to perform well because the DI method correctly assumed the noise model was first order. In this trial the true noise models are changed to sixth order high pass filters while the DI method still assumes that the noise model is first order. Two tests are performed to tune the controller with both the DI method and the ARX-OE identification method.

12.3.1 Test 3.1: Tuning with ARX-OE identification

The PM algorithm successfully identifies the MPM after 800 samples and after a delay of 1600 samples the identification experiment begins at sample number 6400. After the ID experiment an ARX-OE estimate is used to tune the controller. The input profiles for this test, as shown below in Figure 12.17, demonstrate a relatively small magnitude of excitation relative to the previous tests. Apart from a brief disruption at around 540 minutes the inputs conform to their upper and lower bound constraints during the excitation experiment. The primary reason the excitation magnitude is lower is the shorter prediction horizon used by the ID algorithm in this trial. The response of both u_1 and u_2 to the set-point change are clearly better after adaptive tuning than with MPM.

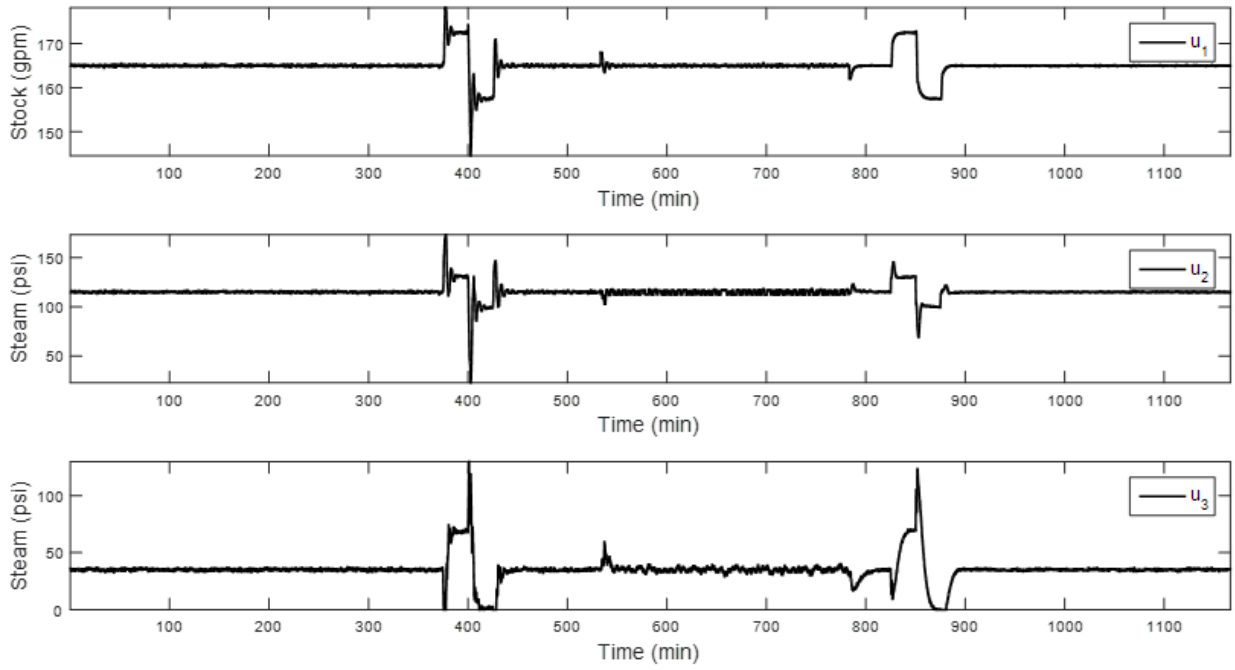

Figure 12.17: Raw input profiles for Test 3.1

Observing the raw output profiles in Figure 12.18 shows that the output variation during the excitation experiment is significantly less than the previous tests. Clearly the output upper and lower bound constraints are satisfied during the excitation experiments. The press moisture y_2 is far less disrupted by the set-point change after the adaptive tuning is conducted. Just before both 550 minutes and 800 minutes the process outputs experience a slight disturbance which corresponds with the beginning and end of the excitation experiment. The remainder of the excitation experiment exhibits only a small amount of variation in each of the outputs.

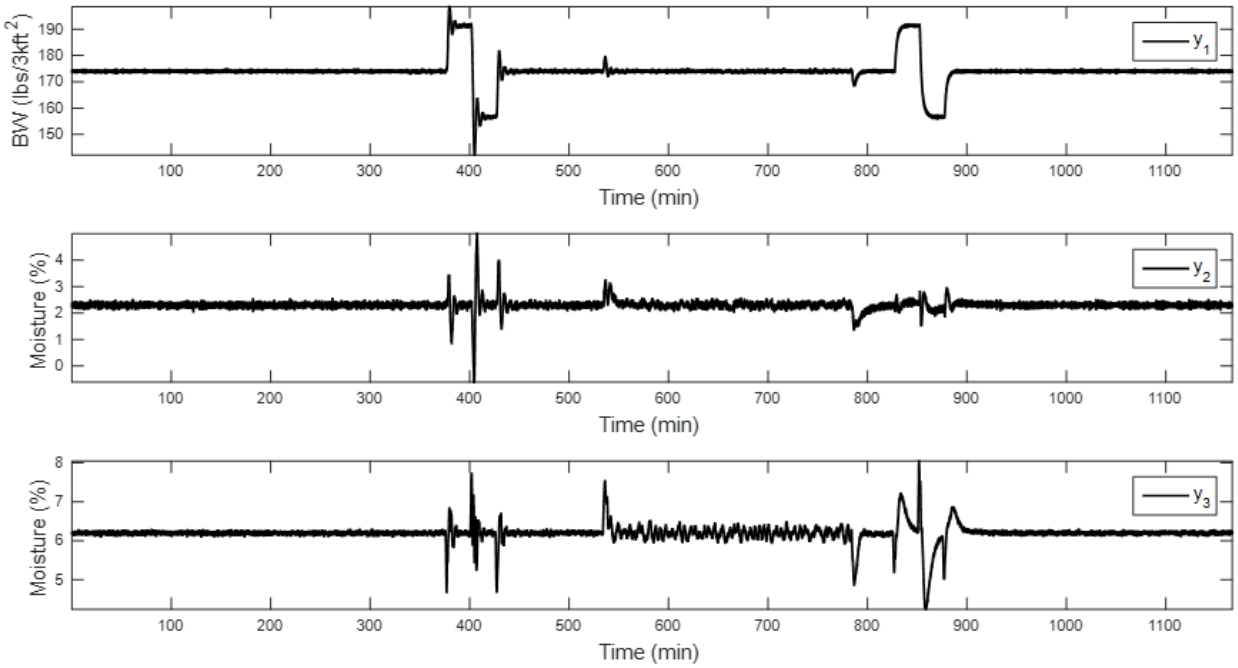

Figure 12.18: Raw output profiles for Test 3.1

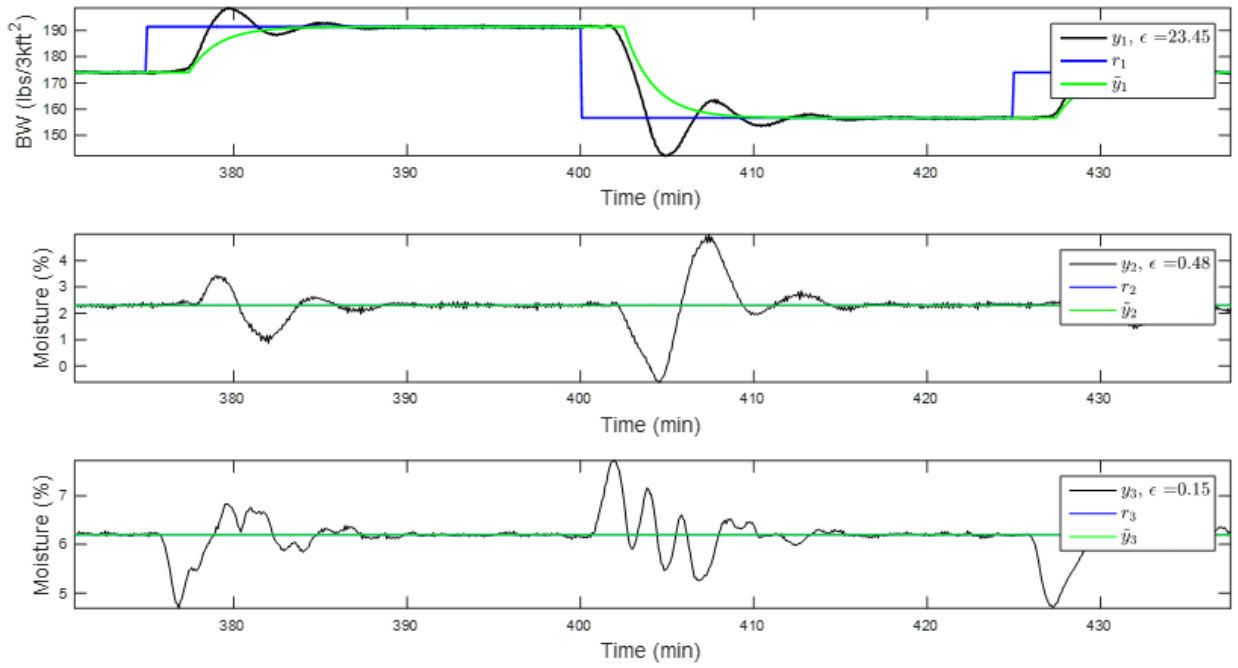

Figure 12.19: Dynamic response with MPM for Test 3.1

The dynamic response of the process outputs before adaptive tuning are shown in Figure 12.19. Again this response has the same oscillatory and over-shoot behavior as before. After the controller is tuned using the ARX-OE identification estimate, another set-point change is performed. The

output response of the re-tuned closed-loop system to this set-point change is shown in Figure 12.20. The same differences as before are observed between the dynamic responses, before and after adaptive tuning. The response after tuning is very smooth with essentially zero over-shoot, whereas before tuning the response has large over-shoot and lots of oscillations. Although the legend of y_2 shows an MSE of zero this is simply because the true value is rounded to the nearest two decimal points. Relative to the previous trial the output responses after tuning all have a lower MSE. It is possible that the sixth order noise models introduce more excitation to the process than the previous first order noise models, ultimately resulting in a more accurate identification of the process. The PE values are analyzed for further investigation.

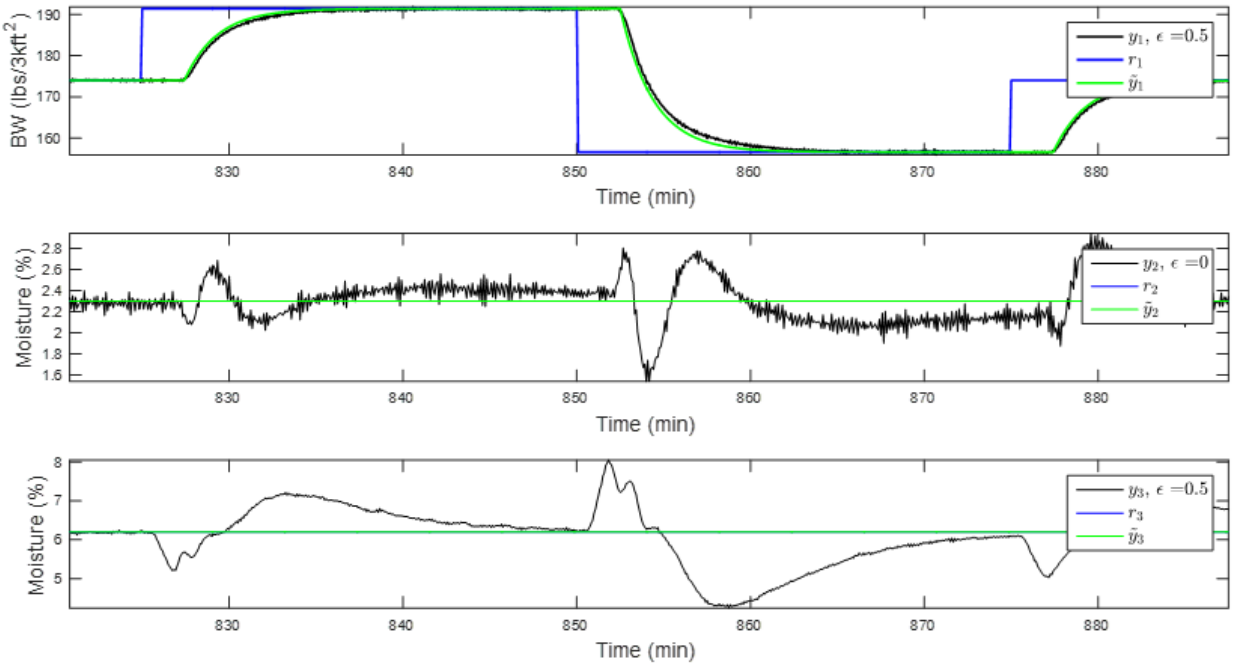

Figure 12.20: Dynamic response after tuning for Test 3.1

The PE values for the time delays, DI and ARX-OE estimates are presented for each transfer function in Table 12.5. Once again some of the time delay values are incorrectly estimated which appears to result in further inaccuracies with the time constant estimates. Relative to the previous tests both the DI and the ARX-OE identification gain estimates are significantly improved. The DI gain estimates are particularly accurate with the largest PE being 22.9%. This is surprising considering the DI method is incorrectly specified with a first order noise model. From the PE results it appears as if incorrect noise model order specification has not had a detrimental effect

on the DI method for the MIMO MD process. The final test aims to confirm this observation by analyzing the controller response after tuning with the DI estimates.

Table 12.5: Percent error of DI and ARX-OE estimates for Test 3.1

Method	Parameter	PE (%)
Both	Time Delay	(0, 0, 4.3, 0, 57.1, 5.3)
Direct	Gain	(8.6, 3.6, 0.4, 5.3, 19.3, 22.9)
Direct	Time Constant	(1.8, 3.9, 4.8, 62.6, 66.4, 126.4)
ARX-OE	Gain	(9.2, 11.6, 21.5, 2.2, 44.4, 38.5)
ARX-OE	Time Constant	(3.2, 6.9, 21.4, 59.4, 108.4, 116.1)

12.3.2 Test 3.2: Tuning with direct identification

For the final test the sixth order HP filter noise models are used with ID excitation but now the DI estimates are used to tune the controller. Just 800 samples after the mismatch is introduced it is successfully detected by the PM algorithm which triggers an identification experiment to begin at sample number 6400. The plant input profiles from the closed-loop system are shown in Figure 12.21. Once again the smaller prediction horizon has resulted in a lower magnitude of input variation

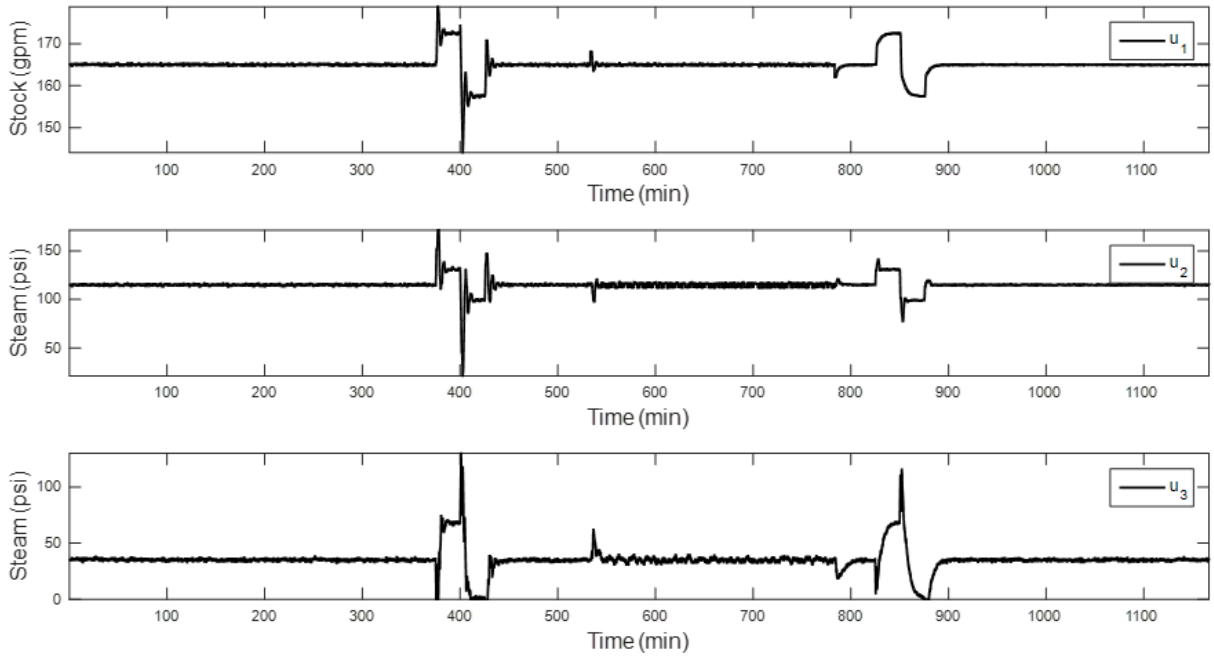

Figure 12.21: Raw input profiles for Test 3.2

during the identification experiments. Furthermore, both u_1 and u_2 are significantly less disrupted due to the set-point change after adaptive tuning than before. Similarly, the outputs presented

in Figure 12.22 show only a small amount of variation during the excitation experiment and thus clearly meet their respective constraints. From an initial glance it appears that all of the output responses to the y_1 set-point change are improved after adaptive tuning is conducted.

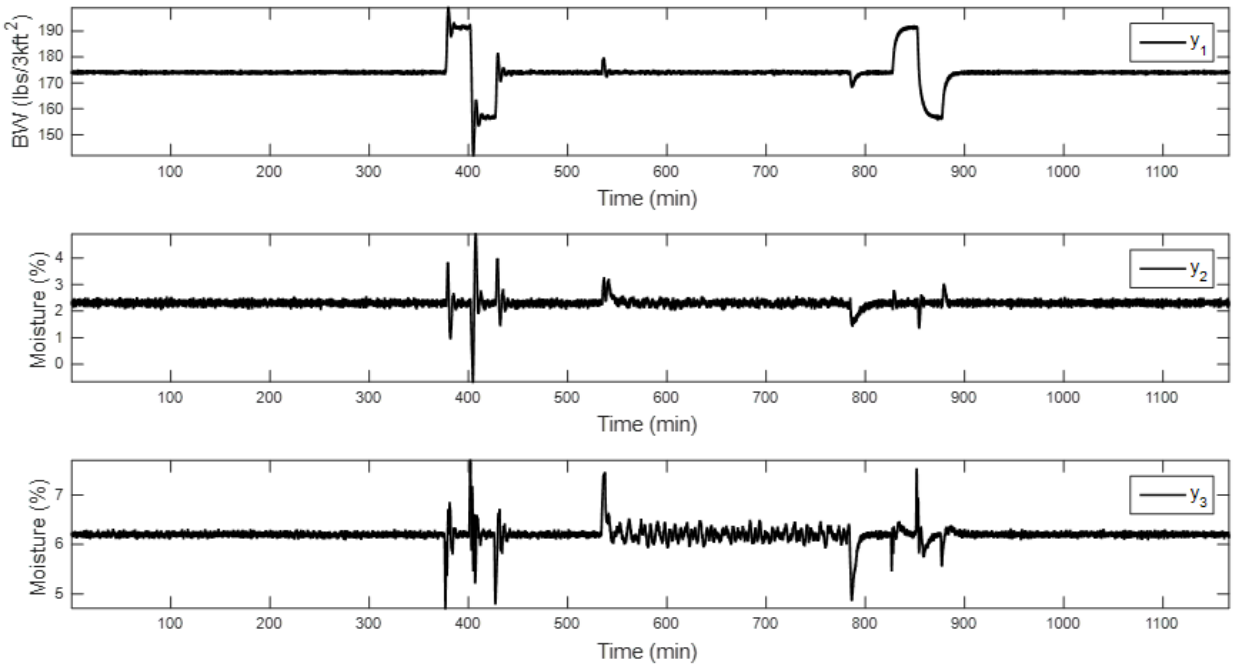

Figure 12.22: Raw output profiles for Test 3.2

The various output responses to the first set-point change are presented more clearly along with their MSE values in Figure 12.23. By now these dynamic responses under MPM have become redundant since the behavior is essentially the same for each test. The second set-point change is conducted after the controller is tuned using the DI estimate. The output response to this set-point change is shown in Figure 12.24. Now both moisture outputs (i.e., y_2 and y_3) portray an MSE of zero although this is only due to the fact that they are rounded to two decimal places. Comparing the dynamic responses before and after adaptive tuning demonstrates an improvement in each of the output profiles after the controller is re-tuned. The BW response which is the only one under significant MPM becomes significantly improved after the identification experiment is conducted. Comparing with the previous test demonstrates that the BW response was slightly better with the ARX-OE controller tuning and the reel moisture (y_3) response was slightly better with the DI controller tuning. Ultimately, the difference between tuning with the DI or ARX-OE identification is fairly negligible for this trial in terms of controller performance.

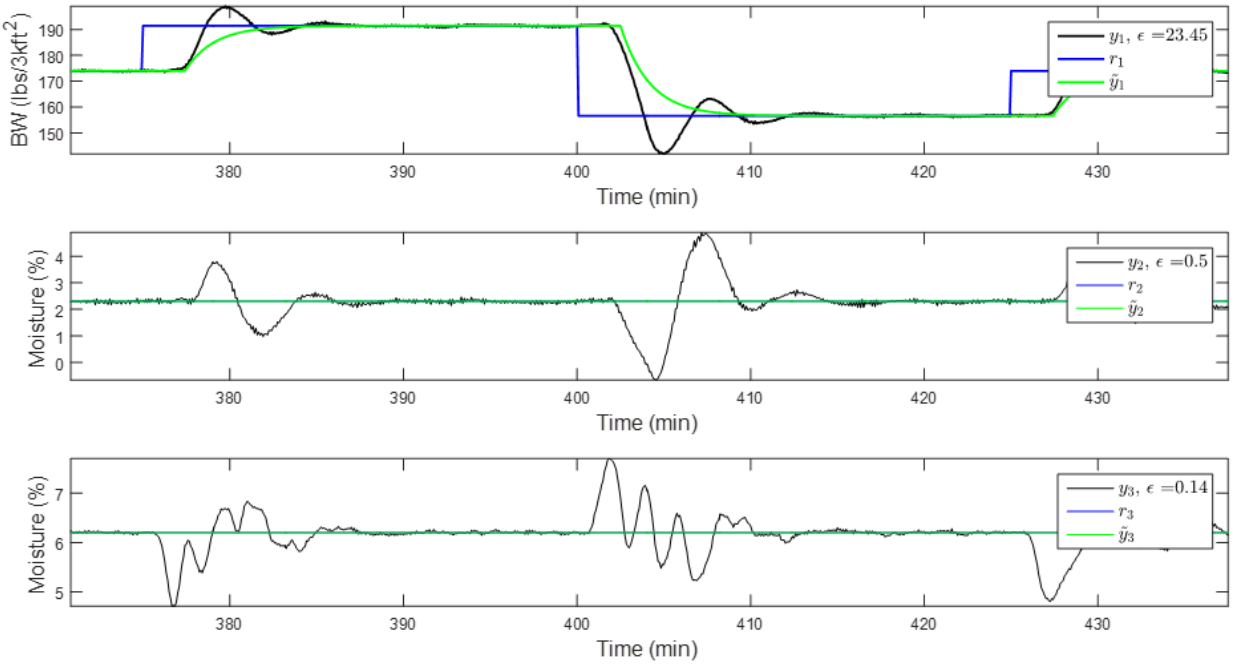

Figure 12.23: Dynamic response with MPM for Test 3.2

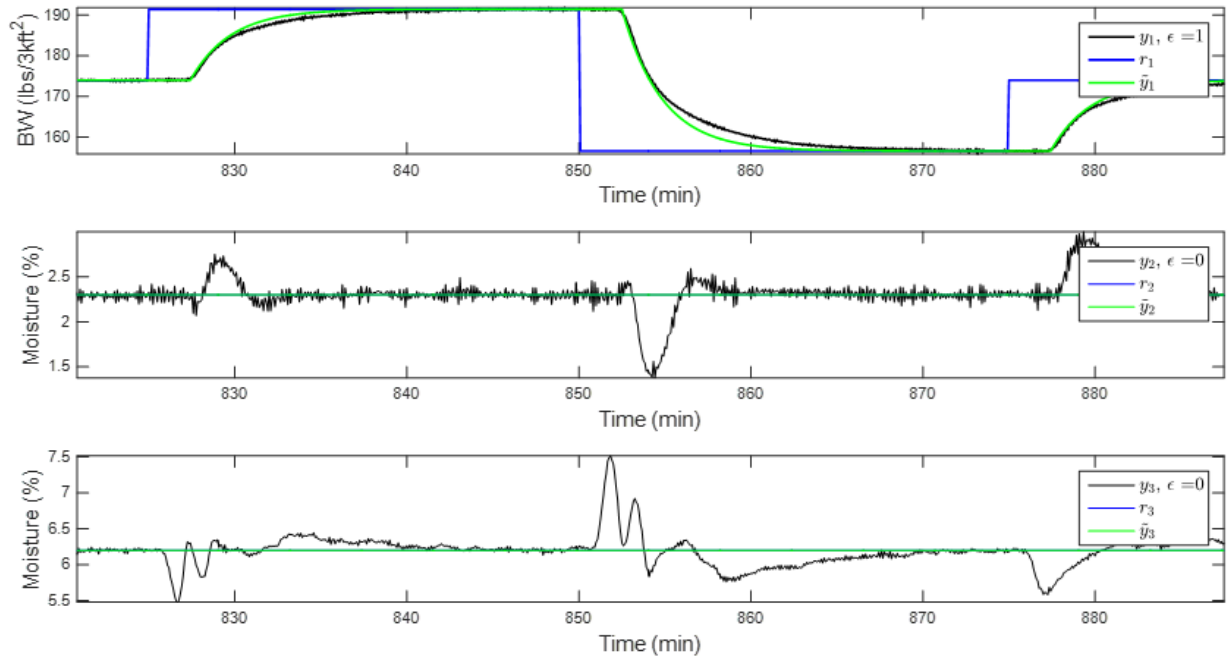

Figure 12.24: Dynamic response after tuning for Test 3.2

As a final analysis the PE values for the DI and ARX-OE estimate are presented in Table 12.6. Once again some of the time delay values are incorrectly estimated. Failure to correctly identify the time delays has been a recurring theme throughout these results. Typically the estimates are close

and most are correct but the time delays certainly represent a significant source of error within the current setup. Surprisingly the DI gain estimates and all but one of the DI time constant estimates are fairly accurate. It appears as if changing the noise model from first to sixth order did not have any detrimental effect on the DI estimates. The ARX-OE estimates are also fairly accurate with the exception of a couple of poorly estimated time constants (likely compensating for the incorrect time delay estimates). From the three different trials the experimental results demonstrate that the ID algorithm is superior to using a PRBS excitation and that the DI method provides slightly better results than the ARX-OE method even if the noise model order is incorrectly specified.

Table 12.6: Percent error of DI and ARX-OE estimates for Test 3.2

Method	Parameter	PE (%)
Both	Time Delay	(13.3, 0, 4.3, 22.2, 42.9, 10.5)
Direct	Gain	(17.4, 3.0, 1.0, 0.8, 8.4, 4.4)
Direct	Time Constant	(27.3, 5.6, 4.7, 8.3, 15.1, 102.8)
ARX-OE	Gain	(25.1, 13.9, 20.6, 2.0, 43.7, 37.6)
ARX-OE	Time Constant	(35.3, 6.1, 18.7, 49.0, 74.3, 145.4)

Chapter 13

Conclusion

To summarize the results from the previous chapter, the adaptive control framework is capable of consistently identifying MPM, generating an excitation signal that meets IO constraints and identifying a new process model. The third aspect of the adaptive control framework, i.e., system identification, is likely the part that could be improved upon the most. For the most part the parameter estimates are fairly accurate but there are certain transfer functions that are consistently more difficult to estimate than others. Furthermore, the SI technique correctly identifies most of the time delay values but inaccuracies in a few of the transfer function time delay estimates are problematic. One potential solution is to further analyze the scores from the MPM detection algorithm and only update the parameters from the transfer functions that exhibits MPM.

In all of the presented tests the output response to a set-point change of the transfer function that underwent MPM is significantly improved after the adaptive tuning. One fairly surprising result is that the DI method provides slightly better results than the ARX-OE method, even when the noise model order is incorrectly specified. Overall the results of the DI and ARX-OE identification estimates are very comparable, particularly the output response profiles. It is recommended that further studies be carried out on this adaptive control framework to determine ways to improve the closed-loop system identification. One potential solution could be to use a longer duration for the excitation experiment. Determining a better time delay estimation method, such as by incorporating available process data, could potentially provide improved identification results.

Comparing the results of the first two trials demonstrates that the ID method enables more accurate identification of the closed-loop system than a simple PRBS excitation. Both the dynamic response of the outputs and the PE values of the parameter estimates were improved by using the ID algorithm. Furthermore, the ID technique is able to provide better control of the input and output variation during the excitation experiment than the PRBS excitation. Particularly in the

final trial where the prediction horizon is set to three the ID algorithm is able to facilitate accurate SI results with only a small amount of IO deviation from the set-point. However, the change in noise model orders from first to sixth order could have influenced the excitation experiment as well. To follow up, the prediction horizon could be changed to four while using the sixth order noise models.

Each aspect of the adaptive control framework is capable of online implementation. Only the ID algorithm is computationally cumbersome which is why the prediction horizon is limited to values of four or less. Lowering the prediction horizon to three in the third trial appears to result in a lower magnitude of excitation but the parameter estimates are just as accurate regardless. It is important to note that online implementation of the ID technique is limited to MIMO systems of relatively small dimension since the number of computations scales exponentially with the dimension of the MIMO system. One beneficial quality of the ID algorithm is that it is capable of being solved through parallel computation. Dividing the expensive search of combinations between multiple processors could substantially increase the speed of each computation.

Ultimately, this report has successfully demonstrated the functionality and performance of the current implementation of the adaptive control framework. Further testing of the various adaptive control aspects is recommended to better understand the potential limitations and develop improvements. Recommendations for future studies include testing the MPM detection algorithm with different amounts of mismatch in each of the parameters for each of the transfer functions. Furthermore, it is recommended to study different variations of the excitation experiment such as altering the duration and the prediction horizon. Although there is room for further development, the contributions made in collaboration with fellow colleagues have resulted in a framework that can automatically detect MPM and re-tune an MPC controller while maintaining feedback control. This work provides a substantial contribution in developing a framework for the adaptive tuning of a MIMO model-based control system without requiring expensive open-loop identification experiments or operator intervention.

Bibliography

[1] Calculation and partitioning of variance using paper machine scanning sensor measurements. *TAPPI, TIP 1101-01*, 2005.

[2] Measurement systems and product variability. *TAPPI, Paper Machine Quality Control Systems (QCS)*, 1, 2005.

[3] George B Arfken and Hans J Weber. *Mathematical methods for physicists international student edition*. Academic press, 2005.

[4] S Aslani, MS Davies, GA Dumont, and GE Stewart. Detecting aliasing between cross and machine direction variations by variable sampling rate. *measurement*, 20(40):1, 2009.

[5] S Aslani, MS Davies, GA Dumont, and GE Stewart. Estimation of cross and machine direction variations using recursive wavelet filtering. *Pulp and Paper Canada*, pages 45–48, 2009.

[6] Abhijit S Badwe, Ravindra D Gudi, Rohit S Patwardhan, Sirish L Shah, and Sachin C Patwardhan. Detection of model-plant mismatch in mpc applications. *Journal of Process Control*, 19(8):1305–1313, 2009.

[7] Alexander Bastounis and Anders C Hansen. On the absence of the rip in real-world applications of compressed sensing and the rip in levels. *arXiv preprint arXiv:1411.4449*, 2014.

[8] WL Bialkowski. Newsprint uniformity and process control the untapped competitive edge. In *CPPA Conference, Montreal, Canada*, 1990.

[9] Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain, 2006.

[10] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[11] Anindya Chatterjee. An introduction to the proper orthogonal decomposition. *Current science*, 78(7):808–817, 2000.

[12] Lennart Ljung Tianshi Chen. What can regularization offer for estimation of dynamical systems? *IFAC Proceedings Volumes*, 46(11):1–8, 2013.

[13] S.C. Chen. Determination of cd and/or md variations from scanning measurements of a sheet of material, November 1 2012. US Patent App. 13/457,870.

[14] Shih-Chin Chen. Analysis of sheet variations–insights of two-dimensional variations.

[15] Shih-Chin Chen. Kalman filtering applied to sheet measurement. In *American Control Conference, 1988*, pages 643–647. IEEE, 1988.

[16] Danlei Chu, Cristian Gheorghe, Johan Backstrom, Michael Forbes, and Stephen Chu. *Model predictive control and optimization for papermaking processes*. INTECH Open Access Publisher, 2011.

[17] William T Cochran, James W Cooley, David L Favin, Howard D Helms, Reginald A Kaenel, William W Lang, GC Maling, David E Nelson, Charles M Rader, and Peter D Welch. What is the fast fourier transform? *Proceedings of the IEEE*, 55(10):1664–1674, 1967.

[18] James W Cooley, Peter AW Lewis, and Peter D Welch. The fast fourier transform and its applications. *IEEE Transactions on Education*, 12(1):27–34, 1969.

[19] K Cutshall. The nature of paper variation. *Tappi journal*, 73(6):81–90, 1990.

[20] Marco F Duarte and Yonina C Eldar. Structured compressed sensing: From theory to applications. *IEEE Transactions on Signal Processing*, 59(9):4053–4085, 2011.

[21] Guy A Dumont, Ivar Mar Jonsson, Michael S Davies, Fariborz T Ordubadi, Ye Fu, K Natarajan, Claes Lindeborg, and E Michael Heaven. Estimation of moisture variations on paper machines. *IEEE transactions on control systems technology*, 1(2):101–113, 1993.

[22] Stephen Duncan and Peter Wellstead. Processing data from scanning gauges on industrial web processes. *Automatica*, 40(3):431–437, 2004.

[23] Andrew P Featherstone and Richard D Braatz. Control-oriented modeling of sheet and film processes. *AIChE Journal*, 43(8):1989–2001, 1997.

[24] Urban Forssell and Lennart Ljung. Closed-loop identification revisited. *Automatica*, 35(7):1215–1241, 1999.

[25] Urban Forssell and Lennart Ljung. A projection method for closed-loop identification. *IEEE Transactions on Automatic Control*, 45(11):2101–2106, 2000.

[26] Nikolaos M Freris, Orhan Öçal, and Martin Vetterli. Recursive compressed sensing. *arXiv preprint arXiv:1312.4895*, 2013.

[27] Sylvain Gendron. Fundamental properties of scanner signals for the correct separation of paper machine cd and md variations. *PAPTAC 91st Annual Meeting*, 1(1):D297D302, 2005.

[28] Ivar Gustavsson, Lennart Ljung, and Torsten Söderström. Identification of processes in closed loopidentifiability and accuracy aspects. *Automatica*, 13(1):59–75, 1977.

[29] Alena Halousková, Miroslav Kárnỳ, and Ivan Nagy. Adaptive cross-direction control of paper basis weight. *Automatica*, 29(2):425–429, 1993.

[30] Li Han, Liguo Han, and Zhao Li. Inverse spectral decomposition with the spgl1 algorithm. *Journal of Geophysics and Engineering*, 9(4):423, 2012.

[31] Fredric J Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

[32] Thomas J Harris. Assessment of control loop performance. *The Canadian Journal of Chemical Engineering*, 67(5):856–861, 1989.

[33] E Michael Heaven, Ivar M Jonsson, T Max Kean, Michael A Manness, and Robert N Vyse. Recent advances in cross machine profile control. *IEEE Control Systems*, 14(5):35–46, 1994.

[34] ECPlaza Network Inc. Release paper and film coating machine, 1996.

[35] Mohieddine Jelali. An overview of control performance assessment technology and industrial applications. *Control engineering practice*, 14(5):441–466, 2006.

[36] Soroush Karimzadeh. *Online detection of picketing and estimation of cross direction and machine direction variations using the discrete cosine transform.* PhD thesis, University of British Columbia, 2008.

[37] Dave Lang and Calvin Fu. Variable speed scanning–a fundamentally better way to scan. *O Papel: revista mensal de tecnologia em celulose e papel*, 74(4):59–64, 2013.

[38] Ron Larson. *Elementary linear algebra.* Nelson Education, 2016.

[39] Canfor Library. Wet end of pulp machine showing fibre mat on the wire.

[40] C Lindeborg. A process model of moisture variations. *Pulp & paper Canada*, 87(4):42–47, 1986.

[41] Lennart Ljung. *System identification.* Wiley Online Library, 1999.

[42] Q Lu, RB Gopaluni, MG Forbes, PD Loewen, JU Backstrom, and GA Dumont. Model-plant mismatch detection with support vector machines. *Accepted in 2017 IFAC World Congress, Toulouse, France*, 2017.

[43] Q Lu, LD Rippon, RB Gopaluni, MG Forbes, and PD Loewen. A novel closed-loop arx output-error identication method. *Pending submission*, 2017.

[44] Qiugang Lu, Lee D Rippon, R Bhushan Gopaluni, Michael G Forbes, Philip D Loewen, Johan Backstrom, and Guy A Dumont. Cross-directional controller performance monitoring for paper machines. In *American Control Conference (ACC), 2015*, pages 4970–4975. IEEE, 2015.

[45] Alain Martel, Wissem MBarek, and Sophie DAmours. International factors in the design of multinational supply chains: the case of canadian pulp and paper companies. *Document de travail DT-2005-AM-3, Centor, Université Laval*, 10, 2005.

[46] Zoran Nesic, Michael Davies, and G Dumont. Paper machine data compression using wavelets. In *Control Applications, 1996., Proceedings of the 1996 IEEE International Conference on*, pages 161–166. IEEE, 1996.

[47] M Ohenoja. One-and two-dimensional control of paper machine: a literature review. october 2009. Technical report, ISBN 978-951-42-9316-0.

[48] Rohit S Patwardhan and R Bhushan Goapluni. A moving horizon approach to input design for closed loop identification. *Journal of Process Control*, 24(3):188–202, 2014.

[49] Don G Roberts, J Lethbridge, and H Carreau. *Changes in the global forest products industry*. Forest Sciences Centre, 2004.

[50] Maristela Oliveira Santos and Bernardo Almada-Lobo. Integrated pulp and paper mill planning and scheduling. *Computers & Industrial Engineering*, 63(1):1–12, 2012.

[51] JC Skelton, PE Wellstead, and SR Duncan. Distortion of web profiles by scanned measurements. *Pulp & Paper Canada*, 104(12):81–84, 2003.

[52] Sigurd Skogestad and Ian Postlethwaite. *Multivariable feedback control: analysis and design*, volume 2. Wiley New York, 2007.

[53] Siemens Industry Solutions. Plate rolling mill equipped by siemens, 2011.

[54] Gregory E Stewart, Dimitry M Gorinevsky, and Guy Albert Dumont. Feedback controller design for a spatially distributed system: The paper machine problem. *IEEE Transactions on Control Systems Technology*, 11(5):612–628, 2003.

[55] Parisa Towfighi. *Estimating paper sheet process variations from scanned data using compressive sensing*. PhD thesis, University of British Columbia, 2011.

[56] Paul Van den Hof. Closed-loop issues in system identification. *Annual reviews in control*, 22:173–186, 1998.

[57] Jeremy G VanAntwerp, Andrew P Featherstone, Richard D Braatz, and Babatunde A Ogunnaike. Cross-directional control of sheet and film processes. *Automatica*, 43(2):191–211, 2007.

[58] Xiaochun George Wang, Guy A Dumont, and Michael S Davies. Adaptive basis weight control in paper machines. In *Control Applications, 1993., Second IEEE Conference on*, pages 209–216. IEEE, 1993.

[59] Xiaochun George Wang, Guy A Dumont, and Michael S Davies. Estimation in paper machine control. *IEEE Control Systems*, 13(4):34–43, 1993.

[60] Xiaochun George Wang, Guy A Dumont, and Michael S Davies. Modelling and identifications of basis weight variations in paper machines. *IEEE Transactions on Control Systems Technology*, 1(4):230–237, 1993.

[61] Glenn Weigel. A strategic planning model for maximizing value creation in pulp and paper mills. *Université Laval, Québec*, 57, 2005.

[62] M Yousefi, LD Rippon, MG Forbes, RB Gopaluni, PD Loewen, GA Dumont, and J Backstrom. Moving-horizon predictive input design for closed-loop identification. *IFAC-PapersOnLine*, 48(8):135–140, 2015.

[63] Yucai Zhu. Multivariable process identification for mpc: the asymptotic method and its applications. *Journal of Process Control*, 8(2):101–115, 1998.

Appendix A

Additional Sheet Estimation Results

Table A.1: Simulation parameters for Trial 3, 100 CD bins

Parameter	Description	Value
w_{sheet}	sheet width (m)	6
v_{scan}	scan speed(s) (m/s)	1
t_{scan}	scan time (s)	6
v_{sheet}	sheet speed (m/s)	3
θ	traversing angle from MD (°)	18.4
$n_{CD} = n_B$	number of CD bins	100
n_S	number of scans (per scanner)	2
ω_{MD}	frequency of MD sinusoid(s) (Hz)	0.08, 0.16

Table A.2: Simulation parameters for Trial 3, 60 bins

Parameter	Description	Value
w_{sheet}	sheet width (m)	8
v_{scan}	scan speed(s) (m/s)	1
t_{scan}	scan time (s)	8
v_{sheet}	sheet speed (m/s)	1
θ	traversing angle from MD (°)	45
$n_{CD} = n_B$	number of CD bins	60
n_S	number of scans (per scanner)	2
ω_{MD}	frequency of MD sinusoid(s) (Hz)	0.062, 0.12

Table A.3: Simulation parameters for Trial 3, 60 CD bins, 4 scans

Parameter	Description	Value
w_{sheet}	sheet width (m)	8
v_{scan}	scan speed(s) (m/s)	1
t_{scan}	scan time (s)	8
v_{sheet}	sheet speed (m/s)	8
θ	traversing angle from MD (°)	7.1
$n_{CD} = n_B$	number of CD bins	60
n_S	number of scans (per scanner)	4
ω_{MD}	frequency of MD sinusoid(s) (Hz)	0.062, 0.12

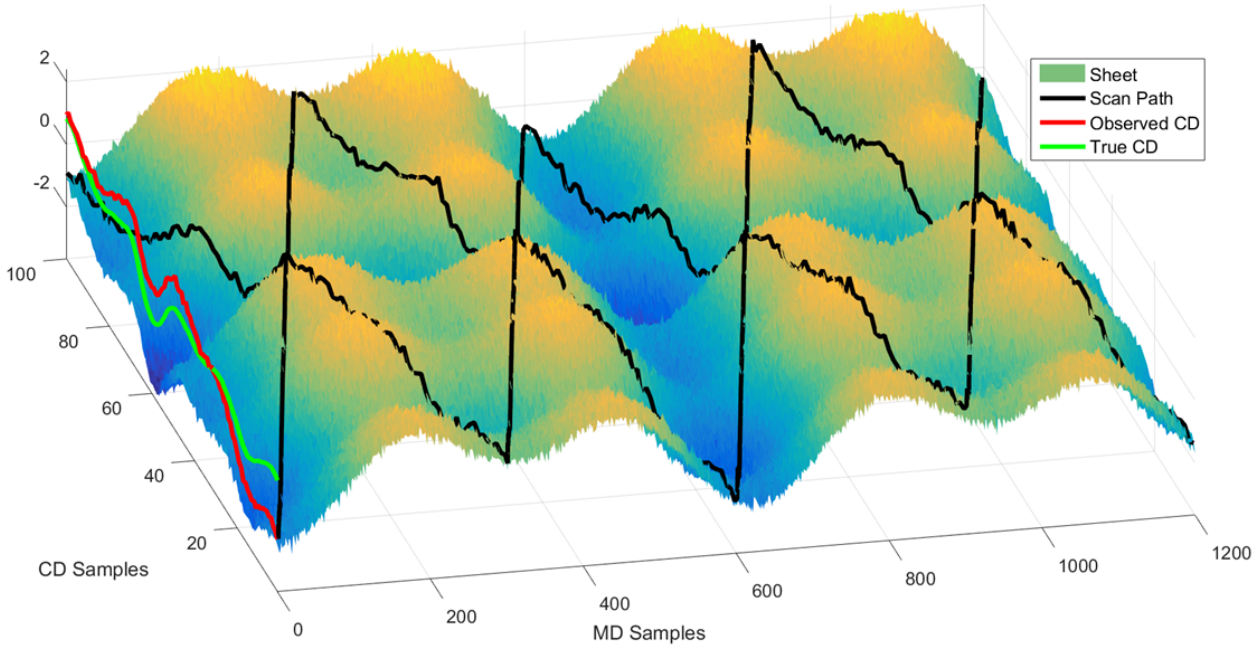

Figure A.1: Simulated sheet with 100 CD bins and 4 scans

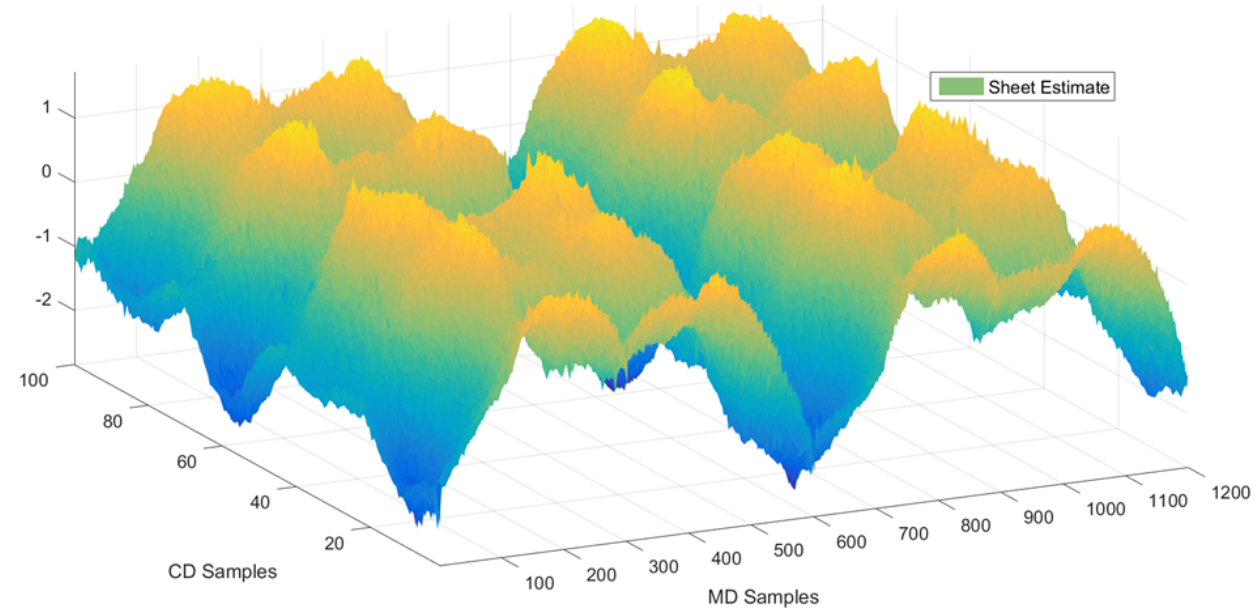

Figure A.2: Compressive sensing estimate with 100 CD bins and 4 scans

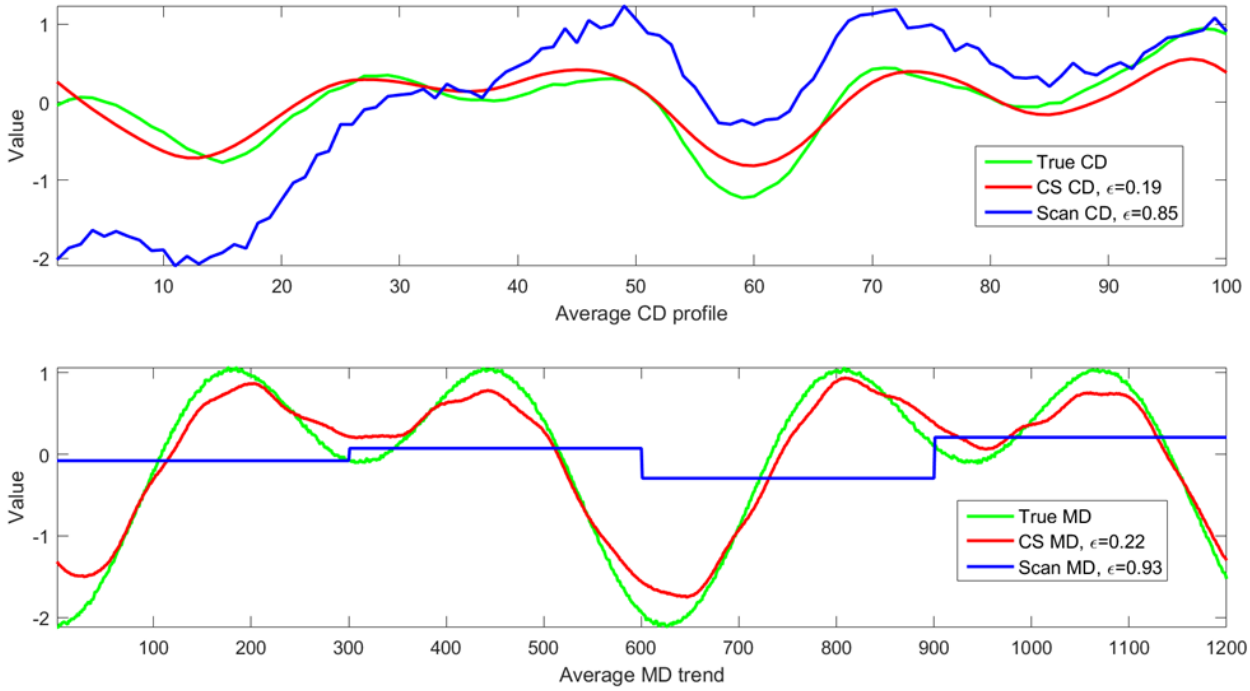

Figure A.3: MD-CD separation results with 100 CD bins and 4 scans

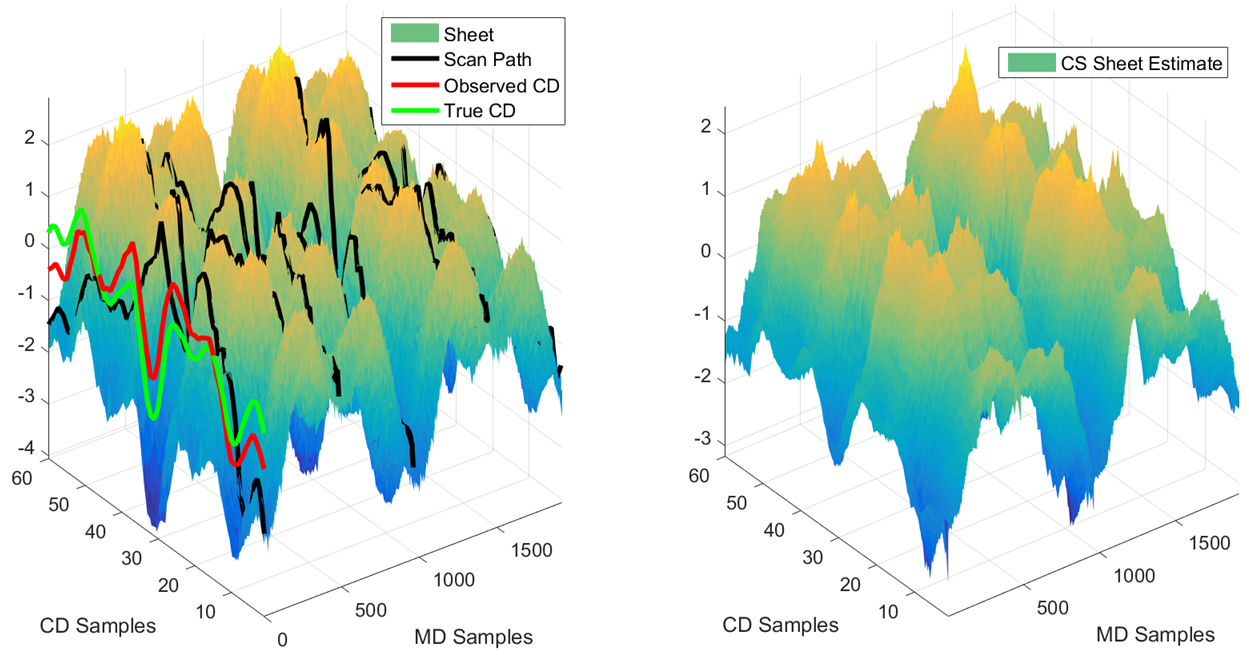

Figure A.4: Simulated sheet and CS estimate with 60 CD bins and 4 scans

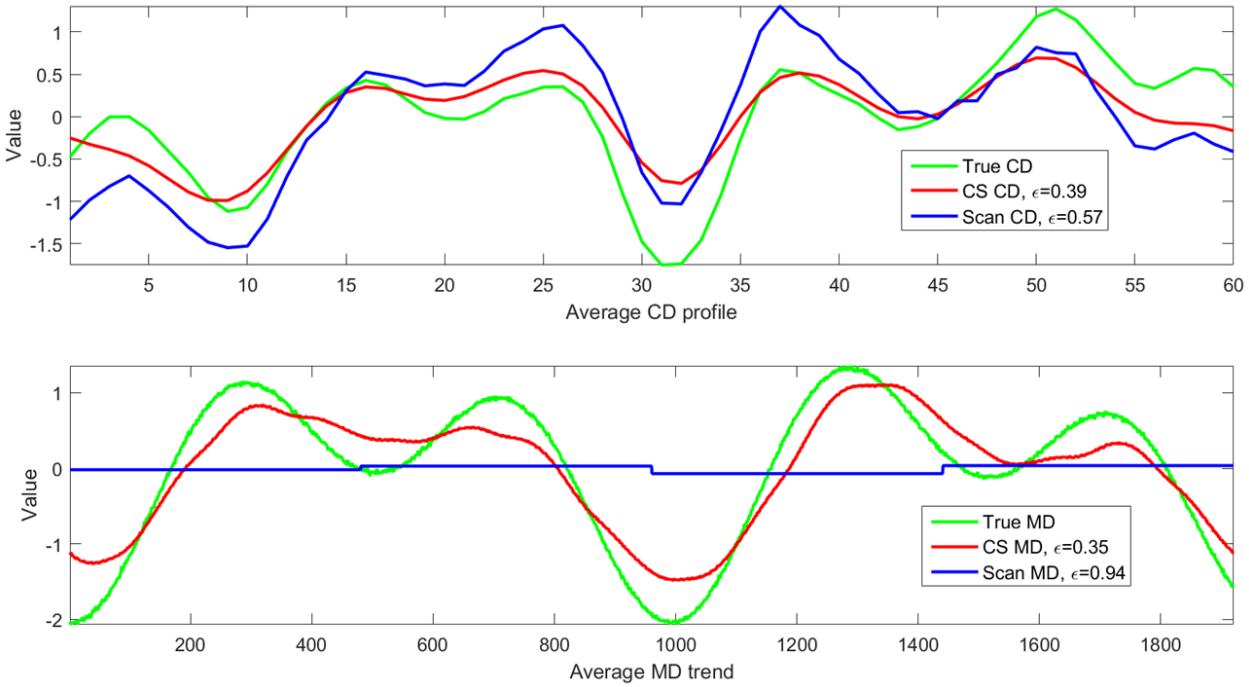

Figure A.5: MD-CD separation results with 60 CD bins and 4 scans