

Machine Direction Adaptive Control on a Paper Machine

Lee D. Rippon,^{*,†} Qiugang Lu,[†] Michael G. Forbes,[‡] R. Bhushan Gopaluni,^{*,†}
Philip D. Loewen,[¶] and Johan U. Backström[‡]

[†]*Department of Chemical and Biological Engineering, The University of British Columbia,
Vancouver, BC, Canada.*

[‡]*Honeywell Process Solutions, 500 Brooksbank Ave, North Vancouver, BC, Canada.*

[¶]*Department of Mathematics, The University of British Columbia, Vancouver, BC,
Canada.*

E-mail: leeripp@chbe.ubc.ca; bhushan.gopaluni@ubc.ca

Abstract

Control of industrial sheet and film processes involves separate controllers and actuators for minimizing both temporal variations along the machine direction (MD) and spatial variations along the cross direction (CD). Model-based control methods such as model predictive control (MPC) have gained widespread implementation for controlling both the MD and CD processes. One limitation of model-based methods is that changes in the true process pose significant identification challenges for operators which are often resolved with costly open loop identification experiments. The predominant industrial measurement technology acquires a signal of mixed MD and CD variations that requires separation. This work compares various model-free approaches for MD-CD separation as a prerequisite to effective MD control. To address the challenges of model-based control this paper introduces an adaptive control method for the

MD process. Closed loop identification experiments are conducted and compared to benchmarks on an industrial paper machine simulator.

Keywords: papermaking, model-based control, model-plant mismatch detection, input design, system identification, adaptive control

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. 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 control of sheet and film processes is of significant industrial importance.

Sheet and film processes are similar in 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), respectively¹. Moreover, outputs are often measured by a scanning sensor that traverses the moving sheet and samples a mixture of MD and CD variation.^{2,3} To consider the practical industrial complexities, this work focuses on an application with a paper machine. It is important to note that this work is relevant to all sheet and film processes involving a scanning sensor and a model-based control system.

A Fourdrinier paper machine, as shown in Figure 1, 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.⁴ 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.⁵

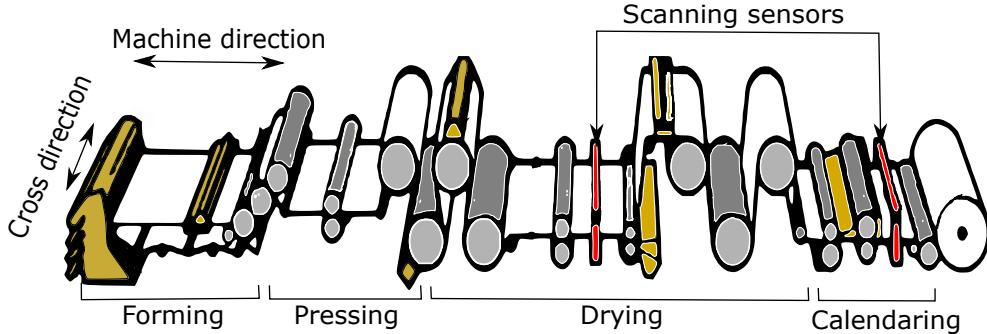


Figure 1: Schematic of a Fourdrinier paper machine (adapted with permission from⁶)

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. Dedicated MD and CD control systems are used to address temporal and spatial variations, respectively. To minimize temporal variations MD actuators (e.g., thick stock valve positioning and dryer steam pressure) are used as they have a relatively uniform effect on the spatial properties of the sheet.⁷ Spatial variations are minimized by manipulating spatially distributed CD actuators such as headbox dilution valves and induction profilers.⁸

To implement feedback control the MD and CD controllers require measured output profiles which are typically acquired using a scanning sensor. 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. The process of generating these separate controller inputs is referred to as MD-CD separation. One of the more challenging aspects of MD-CD separation that has yet to be sufficiently addressed is the presence of MD variations aliasing into the CD profile, i.e., MD aliasing. The procedure of MD-CD separation is a prerequisite not only for model-based feedback control, but also for performance monitoring and system identification. The use of MPC for MD control has become increasingly popular since it was introduced by Honeywell in 1994.⁹ Maintaining an accurate model of the process while the true plant changes according to the varying operating

conditions and physical characteristics is a problem of significant industrial interest.

This paper addresses two fundamental challenges of sheet and film process control with a traversing scanning sensor. Firstly, the MD-CD separation problem is introduced and a literature review is conducted to assess previous efforts. A set of model-free MD-CD separation techniques are selected for comparative analysis due to their practical appeal and flexibility. A novel technique based on compressive sensing (CS) is developed for implementation with a scanning sensor and a study with multiple scanners is presented. The results of the MD-CD separation have a direct impact on the second fundamental challenge addressed here, i.e., adapting the MD process model used by the MD MPC control system. Novel developments towards process monitoring (PM), input design (ID) and system identification (SI) are unified into an adaptive control framework. Integrated case studies are conducted to demonstrate the effectiveness of the unified adaptive control algorithm and compare with alternative identification techniques.

MD-CD separation

The comparative study of MD-CD separation methods conducted in this work centers around the problem of MD aliasing. As such, this section begins with some background material to introduce the practical industrial constraints. Afterwards, the findings of a thorough literature review on MD-CD separation methods is presented which motivates the selection of the candidate model-free methods for comparative analysis. These techniques, namely exponential filtering (EF), power spectral analysis (PSA) and CS, are introduced in greater detail with CS in particular being developed for implementation with a scanning sensor. The results of the comparative analysis are presented and a novel study into CS with multiple scanners is provided. Commentary on the relevance and importance of accurate MD-CD separation for adaptive MD control is provided before introducing the proposed adaptive control framework.

Background

Scanning sensors were first introduced around 1980 when they began to replace the single-point sensors which had a fixed CD position.¹⁰ The scanning sensor framework (as shown in Figure 2 below) 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).^{11,12} Some paper machines use scanning sensors at different locations as in Figure 1. One approach to address MD-CD separation is to use an array of stationary sensors that span the width of the sheet. However the cost of such a sensor array is generally prohibitive due to the expensive sensor technology.¹³ Until sheet spanning sensing mechanisms become commonplace in industry, the problem of MD-CD separation with a scanning sensor configuration remains relevant.¹⁴

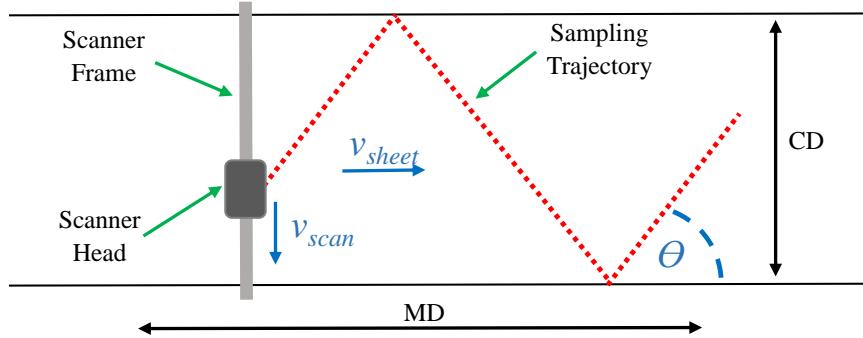


Figure 2: Schematic of a scanning sensor

The resulting sampling trajectory is described by the dotted red line in Figure 2 although in reality, the angle θ from the sampling trajectory to the MD axis is much smaller, typically $\theta \approx 1^\circ$. Specifically, θ can be roughly calculated as follows:¹⁵

$$\theta = 90 - \arctan\left(v_{\text{sheet}} \frac{t_{\text{scan}}}{w_{\text{sheet}}}\right), \quad (1)$$

where t_{scan} is the time required for a single traverse of the sheet, v_{sheet} is the MD speed of the sheet and w_{sheet} is the width of the sheet. It is worth noting that equation 1 ignores occasional stationary periods for sensor calibration and assumes a constant scan speed.

The acquired scanner signal, $\mathbf{y} \in \mathbb{R}^m$, needs to be processed such that it represents the two dimensional reality of the process, i.e., $\mathbf{Y} \in \mathbb{R}^{n_B \times n_S}$. Here m is equal to the number of measurements taken by the scanning sensor, n_B is the number of evenly spaced CD bins and n_S is equal to the number of MD scans. 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.² 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.¹⁶ To achieve $\mathbb{R}^m \mapsto \mathbb{R}^{n_B \times n_S}$ the measurements of each scan are collapsed into a column vector as shown by matrix \mathbf{Y} in Table 1.

Table 1: Two-dimensional representation of scanner signal

		Scans		\leftarrow	MD	\rightarrow
		Scan 1	\dots	Scan j	\dots	Scan n_S
Bins	Bin 1	$Y_{1,1}$	\dots	$Y_{1,j}$	\dots	Y_{1,n_S}
	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
\uparrow	Bin i	$Y_{i,1}$	\dots	$Y_{i,j}$	\dots	Y_{i,n_S}
CD	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
\downarrow	Bin n_B	$Y_{n_B,1}$	\dots	$Y_{n_B,j}$	\dots	Y_{n_B,n_S}

The sheet property matrix \mathbf{Y} can be decomposed as follows:

$$\mathbf{Y} = \bar{\mathbf{Y}}^{\text{MD}} + \bar{\mathbf{Y}}^{\text{CD}} + \mathbf{Y}^{\text{R}}, \quad (2)$$

where $\bar{\mathbf{Y}}^{\text{MD}}$, $\bar{\mathbf{Y}}^{\text{CD}}$ and \mathbf{Y}^{R} represent data associated with MD variations, CD variations and residual variations, respectively.¹⁷ More specifically, if the average of all data points across scan j is

$$\bar{Y}_j^{\text{MD}} = \frac{1}{n_B} \sum_{i=1}^{n_B} Y_{i,j} \quad (3)$$

then

$$\bar{\mathbf{Y}}^{\text{MD}} = \mathbf{q} \begin{bmatrix} \bar{Y}_1^{\text{MD}} & \dots & \bar{Y}_{n_S}^{\text{MD}} \end{bmatrix}, \quad (4)$$

where \mathbf{q} is a column vector of length n_B filled with all ones. In other words, the values along

the rows of a specific column of $\bar{\mathbf{Y}}^{\text{MD}}$ are fixed to the scan average and the matrix has a flat CD profile. Therefore, when $\mathbf{Y} - \bar{\mathbf{Y}}^{\text{MD}}$ is computed it is equivalent to removing the scan average from every data point along that particular scan, i.e., removing the MD variations. The definition of $\bar{\mathbf{Y}}^{\text{CD}}$ follows analogously with values along a particular row being fixed as the average over all scans resulting in a flat MD profile.

This treatment of scanner data ignores the geometry of the sampling trajectory. While the rows correctly represent the approximate spatial positioning of each measurement, the columns treat the temporal positioning of each measurement during a scan as if the sheet was stationary. 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.⁸ Subsequent scans are averaged or temporal filtering is applied to obtain the CD control profile. Ignoring the intra-scan MD position significantly limits the resolution of the MD trend and thus limits the information provided to the MD controller as well as any MD performance monitoring and system identification algorithms. Alternatively, the proposed MD-CD separation method uses CS to incorporate *a priori* knowledge of the sampling geometry to generate a more informative MD trend. Specifically, the CS method accounts for both the MD and CD positions of each measurement to effectively reconstruct a high resolution sheet estimate. The increased resolution of the MD trend provides direct benefits to the adaptive control framework with regards to early upset detection, improved performance monitoring and additional data for better system identification.

To show the effects of MD aliasing the following simplifying assumptions are made:

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.*

Although variable rate scanning has drawn recent interest it is still common in many applications for Assumption 1 to be true for extended periods of time. Ignoring occasional

calibration periods the scanner trajectory is reasonably approximated by Assumption 2. While Assumption 3 is not technically correct, it is true for the vast majority of CD variability over short periods of time.¹⁸ Controllable MD and CD frequencies can be determined to establish a lower bound on the required resolution for MD-CD separation methods. The highest cutoff frequency for MD control loops is for the headbox total head at around 0.05 Hz.¹⁹ The CD actuators may have up to 300 individual actuators and the actuator spacing (for a slice lip system) is from 7.0 cm to 20 cm with a bandwidth of around 15 cm.^{9,20}

According to the assumptions described above, the raw scanner data forms an even, periodic function with respect to the CD variations. Each time the scanner completes two scans back and forth across the sheet represents one period, i.e., $(2t_{\text{scan}})$. 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, aliasing of MD variations in the CD profile is most severe when the frequency of the MD variations occur at integer multiples of $(2t_{\text{scan}})^{-1}$.²¹ The effect of MD aliasing is apparent in Figure 3 by comparing the true CD profile (green) to the observed profile (red). 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 compromises profile resolution and controller response.¹⁴

Methods for MD-CD separation

Early developments in MD-CD separation focus on modeling the process outputs such as basis weight and moisture. A nonlinear MD moisture model was proposed in 1986 and a dual Kalman filter was introduced in 1988.^{3,22} In 1993 the moisture model was extended to the CD and consolidated with a basis weight model before testing with both simulated and industrial data.^{5,7,8} Although the tests presented are encouraging, these model-based MD-CD separation methods have not gained widespread industrial implementation. This is likely

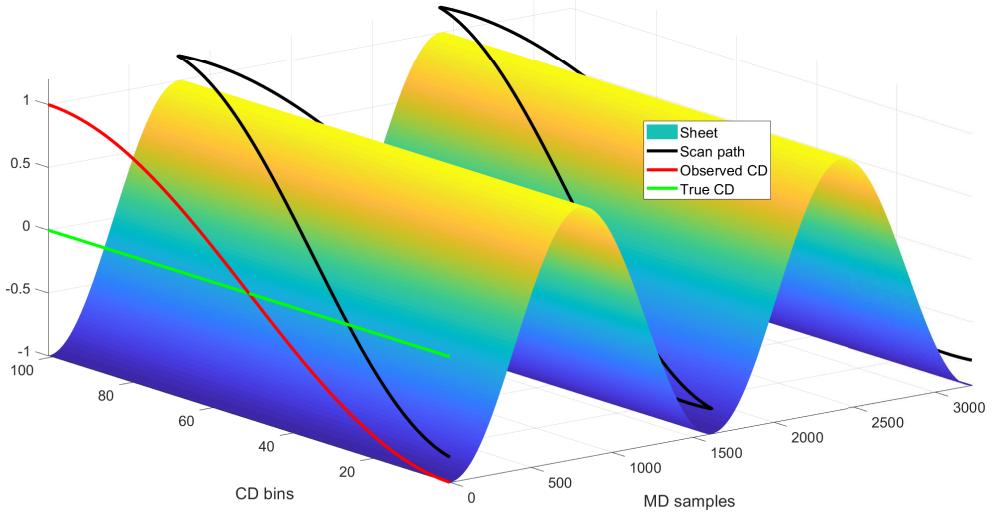


Figure 3: Four scans with MD variation at frequency $1/(2t_{\text{scan}})$

because model based methods rely on specification of parameters that are machine specific, vary with operating conditions and must meet restrictive constraints to ensure convergence.

Subsequent work in MD-CD separation introduces wavelet basis functions to successfully compress industrial data.²³ Early efforts failed to recognize the geometry of the scanner but in 2004 a technique was developed 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.¹ A causal approximation of the non-causal filter is proposed and combined with recursive wavelet denoising to estimate MD and CD profiles but the problem of MD aliasing is not addressed.⁴ In 2005 explicit attention is provided to MD aliasing.²¹ Similarly, in 2008 a discrete cosine transform (DCT) technique uses the fact that CD variations form an even periodic function in the time domain as a means to conduct MD-CD separation. This technique is compared to wavelet and EF methods where it is shown through simulations to have better performance.²⁴ Unfortunately this DCT technique does not distinguish between MD and CD variations that occur at integer multiples of $(2t_{\text{scan}})^{-1}$ and thus does not address severe MD aliasing.

Recent attempts to address MD aliasing have explored the use of variable speed scanning to minimize the impact of MD variations leaking into the CD profile. It is shown that if

the CD profile does not have any fast variations the support of the MD frequency components shift if the scan speed shifts.¹¹ In 2011 PSA with a fast Fourier transform (FFT) is introduced to improve variance partitioning.²⁵ A formal procedure for using PSA with two different scan speeds for MD-CD separation was patented by ABB in 2012.²⁶ The benefits of variable speed scanning for observing a spectral frequency shift in the MD components are confirmed by researchers with Metso.¹⁴ In 2011 the concept of using CS for MD-CD separation was introduced with promising initial results but implementation with the practical constraints of a scanning sensor was not considered.²⁷ This work extends the CS technique for implementation with a traversing scanning sensor and provides a comparative analysis with the PSA method and the longstanding industrially dominant EF method.

Exponential filtering

In EF the average of each scan is stored in a vector, \bar{Y}^{MD} , to generate a trend of measurements for the MD control system. 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.,

$$Y_{i,j}^{\text{CD,f}} = \alpha(Y_{i,j} - \bar{Y}_j^{\text{MD}}) + (1 - \alpha)Y_{i,j-1}^{\text{CD,f}}. \quad (5)$$

Here, $Y_{i,j}^{\text{CD,f}}$ is the filtered value for CD position i and MD position j . For EF the scan average is subtracted from the measured value at each CD bin position and multiplied by the filter factor, $\alpha \in [0, 1]$ which typically has a nominal value of 0.33.^{4,15} The effect of EF with $\alpha = 0.33$ is compared at the second, fortieth and eightieth scans as shown in Figure 4. By the eightieth scan the filtered profile is significantly less biased.

Power spectral analysis

By comparing the FFT of two consecutive scans at alternating scan speeds and observing the shift in the support of the spectral contents, dominant MD frequencies can be distinguished

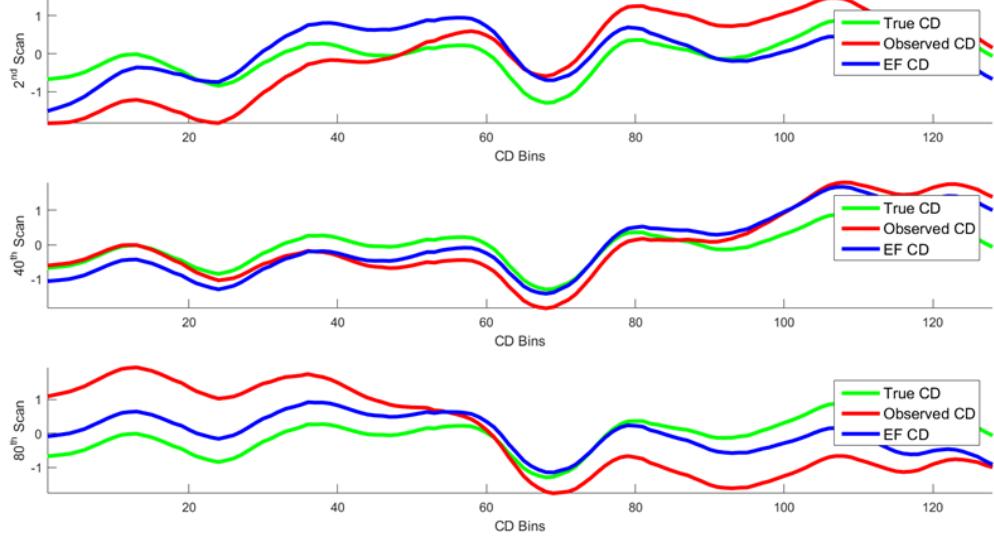


Figure 4: Exponential filtering after 2, 40 and 80 scans

from dominant CD frequencies.²⁶ To demonstrate the use of PSA for MD-CD separation, consider scanning a sheet of paper with a single dominant MD frequency and a single dominant CD frequency at alternating speeds. The top profile in Figure 5 shows the magnitude of the spectral content with respect to the spatial frequency whereas the bottom profile shows this spectral content plotted with respect to the temporal frequency. In the top plot

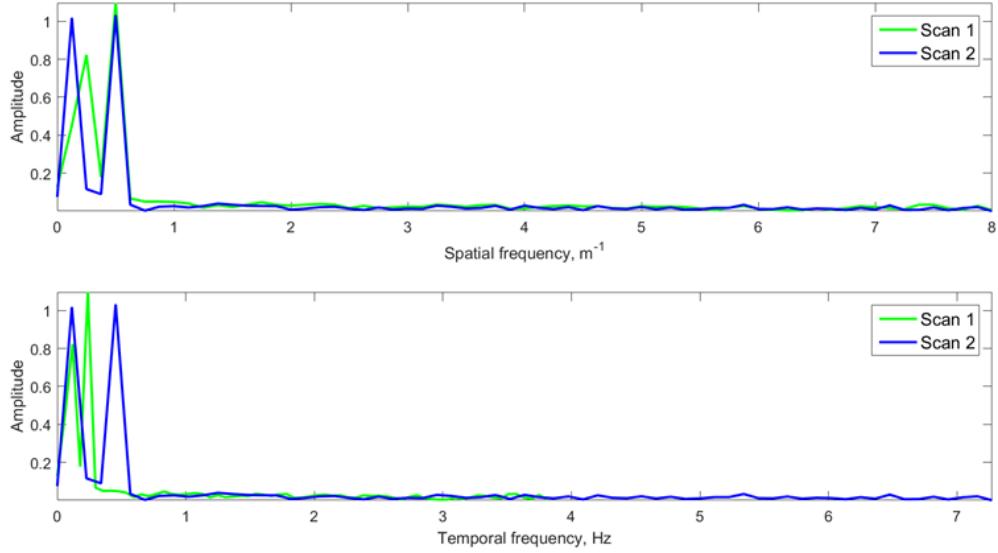


Figure 5: Spectral content of scans at different scan speeds

the dominant frequencies along each scan align at the frequency of the CD variation (0.5

m^{-1}) whereas in the bottom plot they align at the frequency of the MD variation (0.1 Hz). This shift in the spectral content provides the basis for which MD-CD separation can be conducted with PSA as described in Algorithm 1.

Algorithm 1 Power spectral analysis for MD-CD separation

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

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

for each scan **do**

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

 Set a threshold as $T_1 = c_1 \cdot \text{mean}(|x|)$ and set values of $|x| \leq T_1$ 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 ($T_2 = c_2 \cdot \text{mean}(|x|)$) to zero

 Set values of x at intersecting supports to zero

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

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

end for

Compressive sensing

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).^{28,29} 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. With CS, so 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.³⁰ In what follows the CS theory is first presented, followed by a summary of previous work using CS for MD-CD separation and finally the proposed technique for applying CS with a scanning sensor is presented.

Recall 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 $\mathbf{y} = \text{vec}(\mathbf{Y})$ with $\mathbf{y} \in \mathbb{R}^m$ where $m = n_B \times n_S$ represents the values from

each scan stacked upon one another. 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 (in this work $n_{CD} = n_B$). The number of columns n_{MD} represents the desired MD resolution of the reconstruction such that $n_{MD} \gg n_S$. The estimated sheet profile is originally reconstructed in vector form $\mathbf{f} = \text{vec}(\mathbf{F})$ with $\mathbf{f} \in \mathbb{R}^n$ 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 can be compressed and reconstructed accurately while storing few non-zero values. A signal, \mathbf{f} , is considered S -sparse in the representation basis Ψ if the coefficient sequence $\mathbf{x} \in \mathbb{R}^n$ has $S \ll n$ non-zero values and $\mathbf{f} = \Psi\mathbf{x}$.²⁸ 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 \mathbf{f} . A set of basis functions, Ψ , are orthogonal over the interval $[a, b]$ with respect to the weighting function $w(z)$ if for $i \neq j$

$$\int_a^b \psi_i(z)\psi_j(z)w(z)dz = 0. \quad (6)$$

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

$$\int_a^b \psi_i^2(z)w(z)dz = N_i^2. \quad (7)$$

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

$$\int_a^b \psi_i(z)\psi_j(z)w(z)dz = \delta_{ij} \equiv \begin{cases} 0, & \text{for } i \neq j \\ 1, & \text{for } i = j \end{cases} \quad (8)$$

where δ_{ij} is the Kronecker delta. A set of basis functions satisfying equation 8 are considered orthonormal basis functions, i.e., both orthogonal and unit normalized^{31, 32}. A set of orthogonal functions is complete if a piece-wise continuous function, $f(z)$ for $z \in [a, b]$, can

be represented by the series

$$f(z) = \sum_{n=0}^{\infty} x_n \psi_n(z) \quad (9)$$

such that the limit of the integrated squared error vanishes,³¹ i.e.,

$$\lim_{m \rightarrow \infty} \int_a^b \left[f(z) - \sum_{n=0}^m x_n \psi_n(z) \right]^2 w(z) dz = 0. \quad (10)$$

From equation 10 it is apparent that orthogonal completeness is an important property for ensuring the accuracy of the approximation of $f(z)$. In fact this property is critical for both spectral analysis and compressive sensing as it allows the function $f(z)$ to be written directly in the orthogonal basis, i.e.,

$$f(z) = \sum_{n=0}^{\infty} \langle f, \psi_n \rangle \psi_n(z). \quad (11)$$

Another condition of CS, referred to as coherence, relates to the projection matrix, $\Phi \in \mathbb{R}^{n \times n}$. The projection matrix describes how samples are taken from the signal domain, i.e.,

$$\mathbf{y} = \mathbf{R}\Phi\mathbf{f} + \nu, \quad (12)$$

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 \mathbf{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 $\mathbf{f} = \Psi\mathbf{x}$ into equation 12 to obtain

$$\mathbf{y} = \mathbf{R}\Phi\Psi\mathbf{x} + \nu = \mathbf{A}\mathbf{x} + \nu, \quad (13)$$

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

$$\mu(\Phi, \Psi) = \sqrt{n} \max_{1 \leq i, j \leq n} |\langle \phi_i, \psi_j \rangle|. \quad (14)$$

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., $\mu = 1$).^{30,33} The importance of sparsity and low coherence is shown in the following formula where the probability of exact reconstruction exceeds $1 - \delta$ if

$$m \geq C\mu^2(\Phi, \Psi)S\log(n/\delta). \quad (15)$$

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

To guarantee efficient and robust reconstruction 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

$$(1 - \delta_S)\|\mathbf{x}\|_2^2 \leq \|\mathbf{Ax}\|_2^2 \leq (1 + \delta_S)\|\mathbf{x}\|_2^2 \quad (16)$$

is not too close to one.³⁰ While the RIP is NP-hard to verify it is often satisfied by taking a sufficient number of samples for a given sparsity level (S), coherence (μ) and randomness of sampling specified by \mathbf{R} . Random coordinate sampling is often used to satisfy the RIP.

The objective in CS is to find the sparsest vector \mathbf{x} that yields the measurement vector \mathbf{y} . However, since the l_0 minimization is NP-hard an equivalent l_1 relaxation is solved with an inequality constraint to account for measurement noise, i.e.,

$$\begin{aligned} & \min \|\mathbf{x}\|_1 \\ \text{s.t. } & \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \sigma \end{aligned} \quad (17)$$

where σ is related to the noise level. This optimization is known as the basis pursuit with denoising (BPDN) problem. In this work the spectral projected gradient for l_1 (SPGL1) minimization algorithm is used to obtain a solution to the BPDN problem by solving a sequence of dual problems. The SPGL1 method is chosen because it is reported as a fast method for large scale sparse reconstruction.³⁴

Previous work showed the ability of CS with Fourier, Haar, Symlet and Daubechies representation bases to reconstruct paper profiles. For both the simulated and industrial trials the Fourier bases provide the lowest reconstruction error. The Fourier basis makes intuitive sense considering operations such as pressing and drying rollers are likely to result in sinusoidal variations. One limitation of this work is that the scanning geometry was not considered and instead random sampling was used which is not possible with a scanning sensor. For a high speed paper machine to satisfy the RIP condition it was proposed that either a sensor array or a large number of scanning sensors be used.²⁷

Based on previous work the Fourier basis is chosen as the representation basis for CS. Achieving the necessary number of samples to meet the RIP is likely not possible and regardless the requisite *a priori* sparsity knowledge is not available to verify the RIP. 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. To extend previous developments 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. To demonstrate CS with a scanning sensor under MD aliasing consider Figure 6.

Even with a relatively slow sheet 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.

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

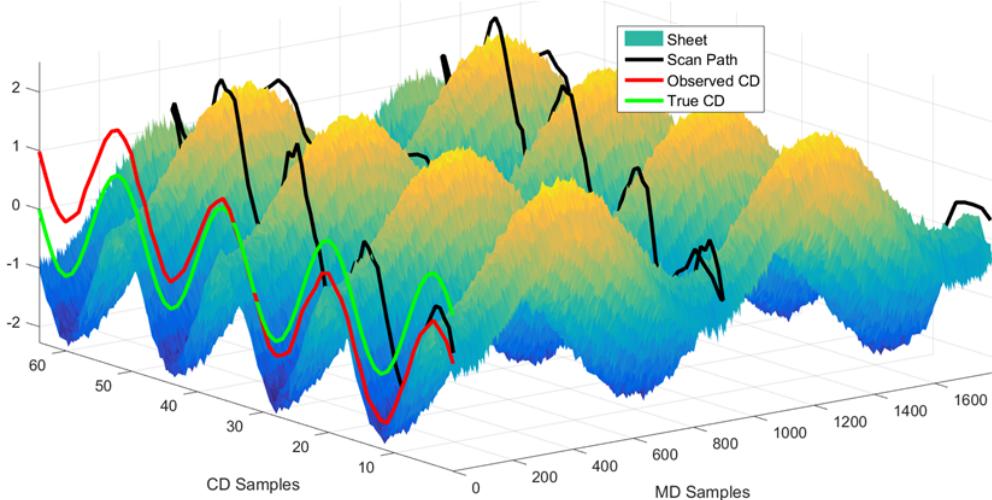


Figure 6: Simulated sheet with 64 CD bins and 4 scans

Note that this assumption is much less assertive than the assumption used by most MD-CD separation methods. 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. The measurement vector \mathbf{y} and CS matrix \mathbf{A} are updated to reflect these values. Next, CS is performed by solving the BPDN problem shown in equation 17. The sheet is reconstructed from the coefficients, i.e., $\mathbf{f} = \Psi\mathbf{x}$, and reshaped into the estimated sheet \mathbf{F} shown in Figure 7.

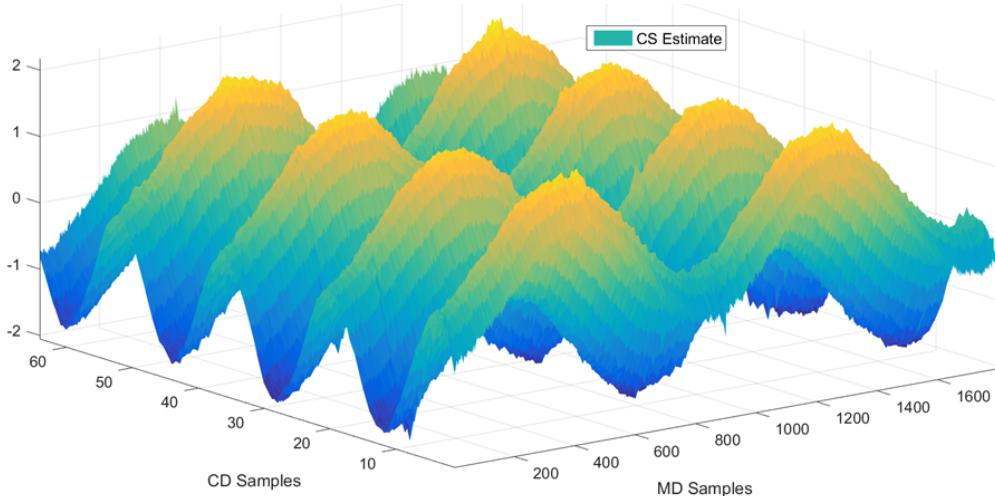


Figure 7: Simulated sheet reconstruction with compressive sensing

Comparative analysis

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 effective after a large number of scans have been processed whereas the PSA method compares two consecutive scans. Furthermore, the PSA method requires alternating scan speeds whereas the CS method performs better when the fastest scan speed is maintained. In the comparative analysis presented here special measures are taken to ensure each method is implemented as designed given the same simulated sheet and scanning setup.

Experimental setup

During industrial operation only the measurements obtained from the scanning sensor are available. Unfortunately since the entire true sheet is not measured verifying the CS reconstruction accuracy is challenging in practice. To determine the true sheet, post-production laboratory tests can be performed but these tests are limited in scope and expensive to perform.⁴ 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 8 and a compressed version of it (i.e., less CD bins) is used in the comparative analysis. To specifically study the ability of MD-CD separation methods to reduce MD aliasing, the sinusoidal MD variations are chosen to have frequencies near integer multiples of $(2t_{\text{scan}})^{-1}$. Zero mean normally distributed random noise with standard deviation σ_e generates the residual variation.

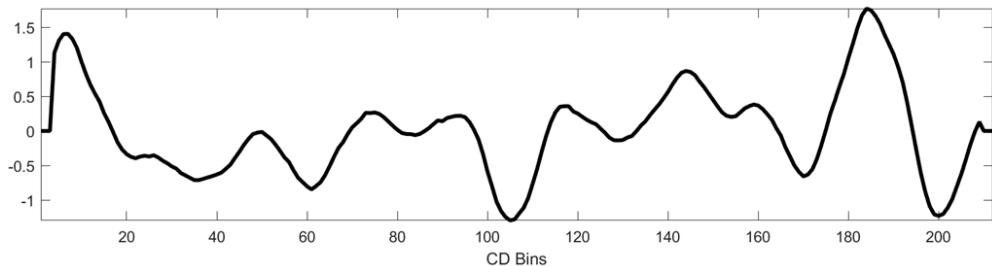


Figure 8: Realistic CD profile provided by Honeywell

Sheet profile estimates from each MD-CD separation method are compared based on how well they fit the true MD and CD variation profiles. Specifically, the root mean square error (RMSE) is calculated as follows:

$$\epsilon = \sqrt{\frac{1}{k} \sum_{i=1}^k (p_i - \hat{p}_i)^2}, \quad (18)$$

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 zero 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 interpolated to generate a signal of desired length. Simulation parameters are summarized in Table 2.

Table 2: Comparative analysis simulation parameters

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 ($^{\circ}$)	18.4, 45
$n_{\text{CD}} = n_B$	number of CD bins	100
$n_S^{\text{EF}}, n_S^{\text{CS}}$	number of scans for EF and CS	20, 8
ω_{MD}	frequency of MD sinusoid(s) (Hz)	0.062, 0.12

Results and discussion

For EF the sheet is simulated for twenty scans but the CS method is only performed on the final eight 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 9 demonstrates. The reverse scan speed of the PSA method is equivalent to the regular scan speed from the CS and EF methods. The final scan shares the same trajectory to allow direct comparison of the profile reconstruction.

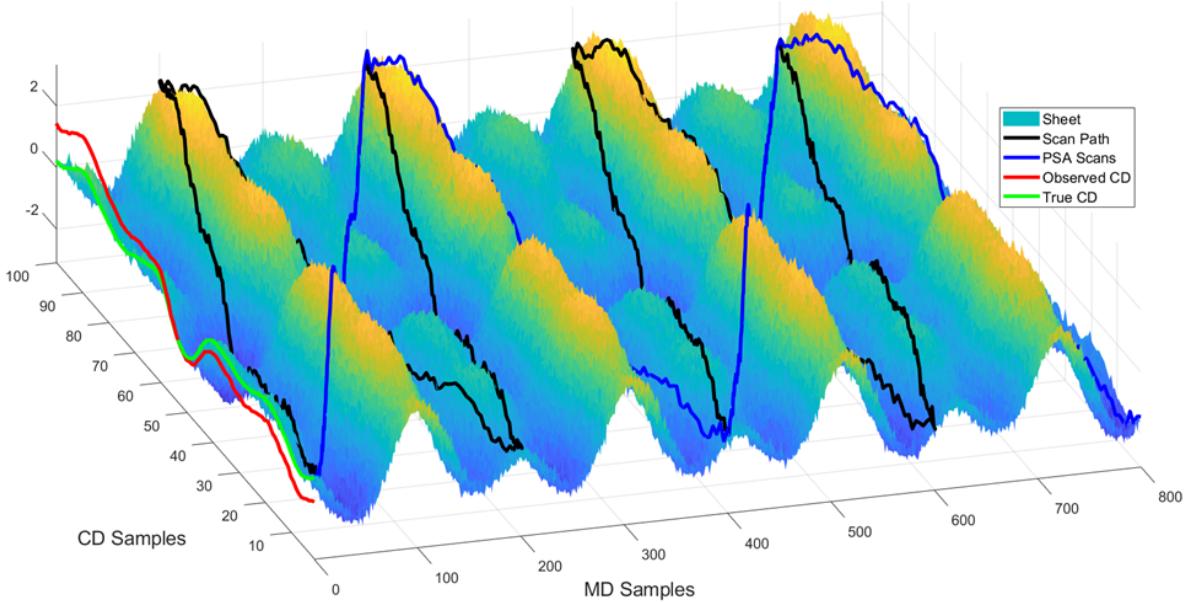


Figure 9: Simulated sheet for comparative analysis

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. Normally distributed zero mean random noise with standard deviation $\sigma_e = 0.15$ is added to the simulated sheet. All 20 scans of the original sheet are passed through an exponential filter with a filter factor $\alpha = 0.33$ to generate an EF estimate of the final scan. The PSA method is performed as described by Algorithm 1 with $c_1 = 1.5$ and $c_2 = 1.0$ to reconstruct the PSA estimate for the final scan. The eight scans shown in Figure 9 are used by the CS method to reconstruct the entire sheet with the help of Assumption 4. The estimated CD profile and MD trends are presented along with RMSE values in Figure 10.

The CS method provides the best reconstruction of the CD profile followed by the PSA technique and lastly the EF method. Although PSA removes part of the MD aliasing, the profile still over-estimates some of the CD variations, likely due to indistinguishable MD variations. Alternatively, the CS estimate captures the essential shape of the true CD profile but the amplitude is underestimated. Soft thresholding used by SPGL1 sparsifies the solution by removing small coefficients at the expense of reducing the amplitude of the remaining coefficients. This is apparent in the MD profile where the CS method once again gives the

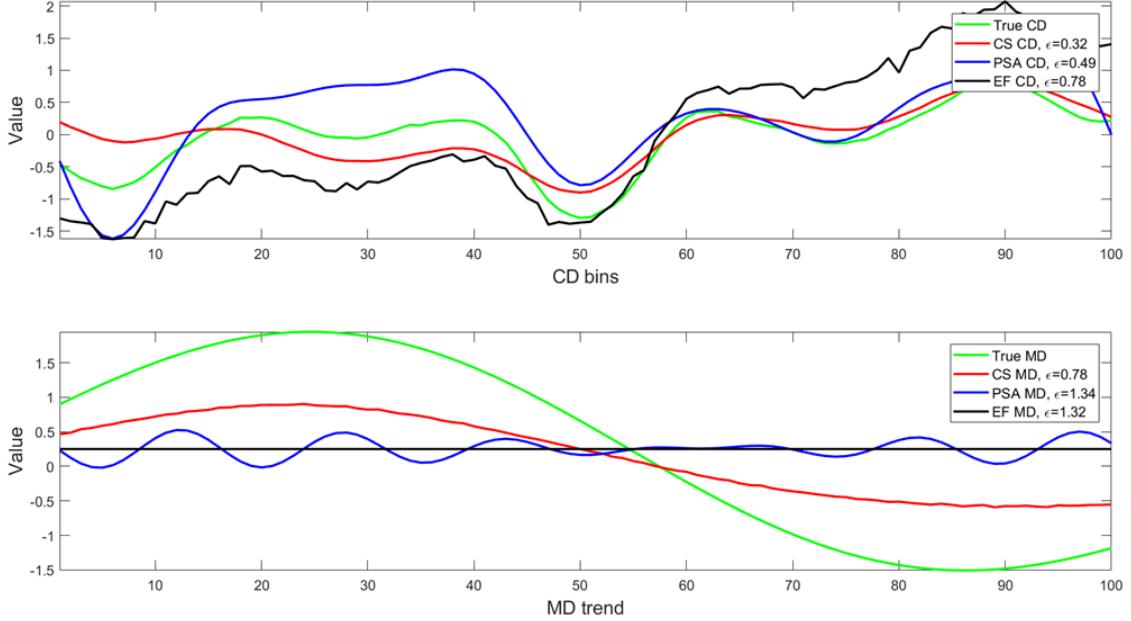


Figure 10: MD-CD separation comparative analysis results

best reconstruction. The PSA estimate of the MD trend has a higher RMSE than the scan average as it is not able to isolate and reconstruct the low frequency MD variation. This limitation could be addressed in future work by developing an adaptive threshold strategy.

Multiple scanners - looking deeper into CS

To better observe the potential of CS two scanners are implemented which represent a realistically achievable sensing framework and provide insight into the benefit of additional samples with CS. The following test involves a CD dimension of 100 bins and an 8 m wide sheet moving at 8 m/s (with $\theta = 7.1^\circ$). Each scanner performs four scans at a speed of 1 m/s with opposite trajectories. The MD, CD and residual variations are simulated as before to cause significant MD aliasing. In Figure 11 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. The CS reconstruction provides a relatively accurate representation of the entire true sheet especially considering it only uses the measurements along the trajectory of the scan paths (in black). Note the increment change in the z-axis indicates reduced magnitude in

the CS reconstruction. By using the geometry of the scanning sensors the CS method makes efficient use of the available scanner information to provide an informative sheet estimate.

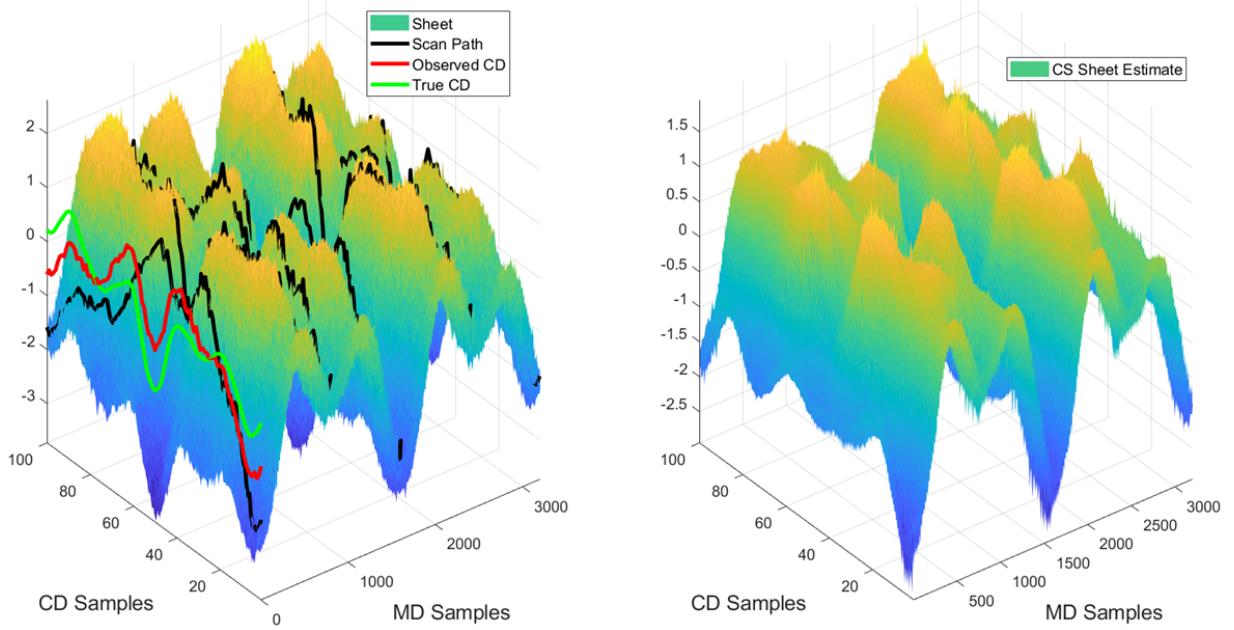


Figure 11: Simulated sheet and CS estimate with 100 CD bins

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 12. From the average CD profiles (top) the CS estimate (red) reduces a significant amount of the aliasing exhibited by the CD profile of the scan average (blue) which is evident numerically from the RMSE values in the legend. The average MD trends are shown in the bottom plot where the scanned values (blue) are the average from each scan. It is clear from these results that given the right circumstances the CS method is able to successfully reduce the detrimental effects of MD aliasing. Moreover, the MD trend in the CS reconstruction provides vastly more information than the scan average as in EF. This test demonstrates the strong potential of CS for reducing MD aliasing and improving the resolution and accuracy of the MD trend for subsequent control of the MD process.

In the following section the challenge of MD adaptive control is addressed. Prior to

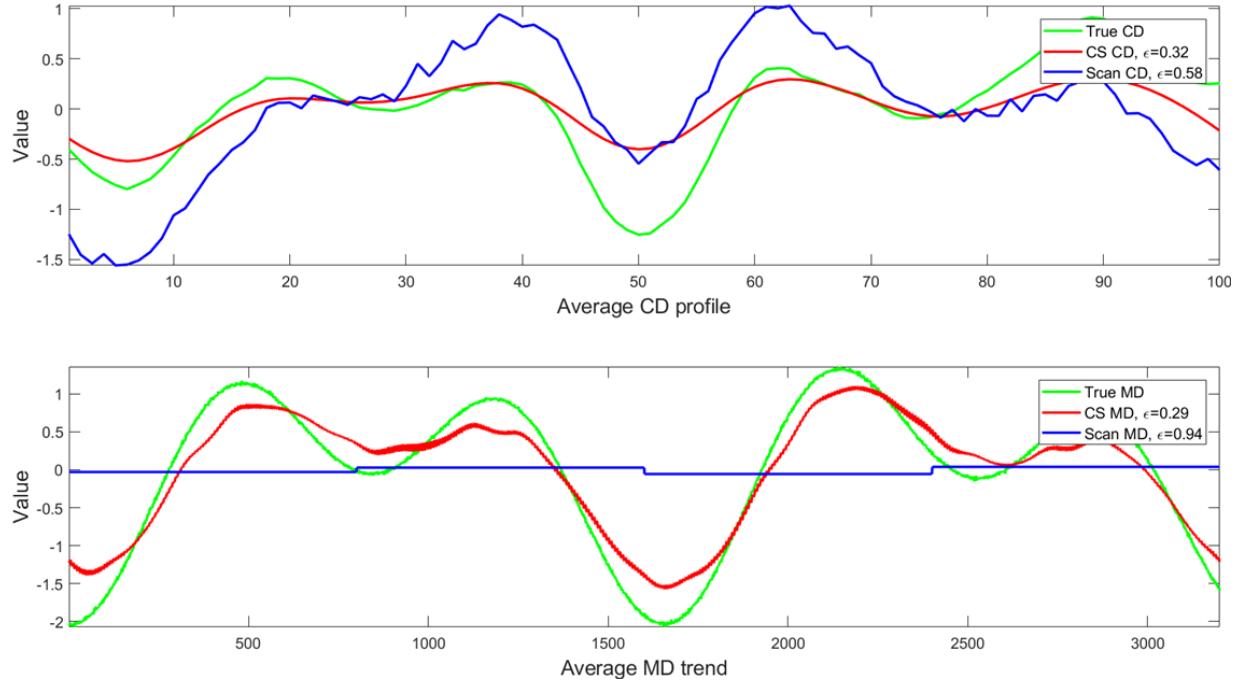


Figure 12: MD-CD separation results for CS with two scanning sensors

introducing the adaptive control framework it is important to elaborate on the significant connection between MD-CD separation and MD adaptive control. The best way to illustrate this connection is by referring to the MD trends in the bottom plot of Figure 12. The blue profile shows the four averaged measurements that would generally be available to an MD control system. Not only does the low resolution result in slow upset detection but it also necessitates longer identification experiments to collect data necessary for accurate system identification and performance monitoring. Alternatively, by using a method such as CS (shown by the red profile in the bottom of Figure 12) more than three thousand MD measurements can be provided to the MD control system over the same span of paper. The high resolution MD trend can facilitate improved feedback control, faster upset detection, better performance monitoring and more accurate system identification.

MD Adaptive Control

The adaptive control strategy consists of techniques developed for MD process monitoring (PM), input design (ID) and system identification (SI). By combining these sub-functions an adaptive control strategy for the MD process is developed. The adaptive framework automatically identifies MPM, triggers an identification experiment, designs an excitation signal, identifies a new process model and updates the controller. An illustrative schematic of the closed-loop adaptive control framework is provided in Figure 13.

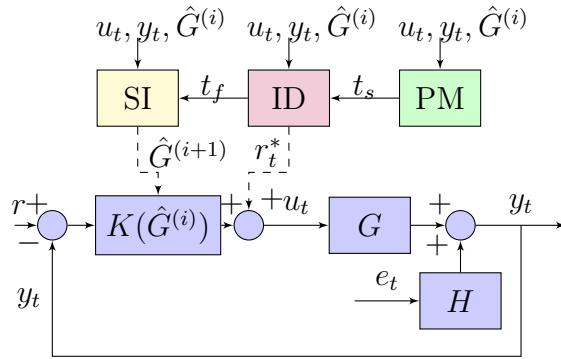


Figure 13: Adaptive framework for closed-loop MPC

During normal operation the model predictive controller, $K(\hat{G}^{(i)})$, controls the process, G , and the PM algorithm collects process input data, u_t , and output data, y_t . 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 calculated by the ID algorithm and added to the actuator movements, u_t , to provide additional process excitation for a specified experiment duration. Once the experiment ends (i.e., t_f is reached) the SI algorithm collects the input-output (IO) data from the experiment duration to determine a new process model estimate. The new estimate, $\hat{G}^{(i+1)}$, is implemented in the controller and normal operation resumes. Successful implementation allows the controller to respond to changes in the process while maintaining feedback control, without requiring operator intervention.

In what follows the MD process is described in greater detail after which each of the sub-

functions within the adaptive control framework are described separately. This is followed by a description of the set of experiments that are designed to test the adaptive control technique relative to benchmark model identification procedures. Upon analyzing the experimental results, conclusions and recommendations for further study are provided.

The MD process

The industrial simulator used in this work involves both the true process dynamics and the MPC control system. The MD process is modeled as a multi-input multi-output (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 an MPC system that determines the new process input values.

During normal operation the following expression is obtained from Figure 13:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = G(q) \begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{bmatrix} + H(q) \begin{bmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{bmatrix}, \quad (19)$$

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 $\sigma_{e_i}^2$.³⁵ Process outputs $y_1(t)$, $y_2(t)$ and $y_3(t)$ represent the measured basis weight (lbs/3000ft²), 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, $K(\hat{G})$, 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:

$$G(q) = \begin{bmatrix} G_{11}(q) & 0 & 0 \\ G_{21}(q) & G_{22}(q) & 0 \\ G_{31}(q) & G_{32}(q) & G_{33}(q) \end{bmatrix}. \quad (20)$$

Each G_{ij} in equation 20 represents an FOPDT transfer function which is defined in continuous time according to

$$G_{ij}(s) = \frac{b_{ij}}{a_{ij}s + 1} e^{-d_{ij}s}, \quad (21)$$

where b_{ij} , a_{ij} and d_{ij} are the continuous gain, time constant and time delay parameters, respectively.³⁶ The nominal model used by the controller, $\hat{G}(q)$, takes the 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 3. The

Table 3: 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

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

$$H(q) = \begin{bmatrix} H_1(q) & 0 & 0 \\ 0 & H_2(q) & 0 \\ 0 & 0 & H_3(q) \end{bmatrix}, \quad (22)$$

where each $H_i(q)$ can be either high pass or low pass filters of varying order.

Closed-loop model identification

In addition to updating the process model, SI is of critical importance for MPM detection. The predominant SI methods can be broadly categorized as belonging to either the prediction error family, subspace approaches or non-parametric correlation and spectral analysis methods. An important difference between open-loop and closed-loop data is the correlation between the noise and the input with closed-loop data as a result of feedback control. This correlation can be problematic for extending open-loop identification methods such as subspace and nonparametric methods to closed-loop data.³⁷ Therefore, the analysis presented here focuses on the prediction error method (PEM).

Depending on the feedback assumptions, closed-loop identification methods can be classified as either direct identification (DI), indirect identification or joint input-output identification.³⁸ For PM, it is desirable for closed-loop identification to be performed on routine operating data. Joint input-output identification methods are ruled out as they depend on external excitation. Furthermore, both indirect and joint input-output identification techniques can become overly complex in the presence of nonlinear feedback.^{37,39} As it is common for industrial MPC to display nonlinear dynamics, the indirect and joint input-output identification techniques are not ideal. The focus of this report is on comparing the direct identification (DI) method to a new method that utilizes both an autoregressive exogenous (ARX) model and an output-error (OE) model, i.e., closed-loop ARX-OE identification.

Both the DI and ARX-OE identification methods can be seen as varying parameterizations of the PEM which minimizes the squared 2-norm of the prediction error, i.e.,

$$V_N(G, H) = \frac{1}{N} \sum_{t=1}^N \hat{\epsilon}^2(t, \theta), \quad (23)$$

where $\hat{\epsilon}(t, \theta)$ represents the difference between the measured and predicted outputs, i.e., $\hat{\epsilon}(t, \theta) = y(t) - \hat{y}(t|\theta)$. Please note the distinction made between the actual process white noise sequence, $e_i(t)$, and the prediction error $\hat{\epsilon}(t, \theta)$.⁴⁰ For plant and noise models defined

by the parameter vector θ , the predicted output, $\hat{y}(t|\theta)$, is given by³⁵

$$\hat{y}(t|\theta) = \hat{H}^{-1}(q, \theta)\hat{G}(q, \theta)u(t) + [1 - \hat{H}^{-1}(q, \theta)]y(t), \quad (24)$$

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

$$\hat{\epsilon}(t, \theta) = \hat{H}^{-1}(q, \theta)[y(t) - \hat{G}(q, \theta)u(t)]. \quad (25)$$

Please note, with respect to MD-CD separation θ represents the angle of the scanning trajectory whereas in SI the symbol θ refers to the parameter vector.

Ultimately, the goal is to obtain θ_o such that the parameterized models are equal to the true system, i.e., $\hat{G}(q, \theta_o) = G(q)$ and $\hat{H}(q, \theta_o) = H(q)$. Minimizing equation 23 with respect to θ results in the following PEM estimate of the parameter vector

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta, Z^N), \quad (26)$$

where $Z^N = \{u(1), y(1), \dots, u(N), y(N)\}$ represents the provided IO data within a given time span $t = 1, \dots, N$.^{37,40} A polynomial representation of the MIMO MD system is used here, e.g., the expression for y_3 in equation 19 becomes

$$A_3(q)y_3(t) = \frac{B_{31}(q)}{F_{31}(q)}u_1(t - n_{k_1}) + \frac{B_{32}(q)}{F_{32}(q)}u_2(t - n_{k_2}) + \frac{B_{33}(q)}{F_{33}(q)}u_3(t - n_{k_3}) + \frac{C_3(q)}{D_3(q)}e(t), \quad (27)$$

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 equation 27 are defined as follows:

$$A(q, \theta) = 1 + \hat{a}_1 q^{-1} + \dots + \hat{a}_{n_a} q^{-n_a}, \quad (28)$$

$$B(q, \theta) = \hat{b}_1 + \hat{b}_2 q^{-1} + \dots + \hat{b}_{n_b} q^{-n_b}, \quad (29)$$

$$C(q, \theta) = 1 + \hat{c}_1 q^{-1} + \cdots + \hat{c}_{n_c} q^{-n_c}, \quad (30)$$

$$D(q, \theta) = 1 + \hat{d}_1 q^{-1} + \cdots + \hat{d}_{n_d} q^{-n_d}, \quad (31)$$

$$F(q, \theta) = 1 + \hat{f}_1 q^{-1} + \cdots + \hat{f}_{n_f} q^{-n_f}. \quad (32)$$

Where model orders n_a , n_b , n_c , n_d , n_f and n_k must be specified. 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.

ARX-OE identification

The ARX-OE method is a new technique for performing closed-loop identification using routine operating data without *a priori* knowledge of the noise model order. The ARX-OE method is carried out through the following two step procedure:

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 using a combination of line search algorithms including adaptive subspace Gauss-Newton, Levenberg-Marquardt least squares and steepest descent least squares.

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 equation 19 is represented with the following ARX model structure:

$$A(q, \theta)y(t) = B(q, \theta)u(t - n_k) + e(t). \quad (33)$$

Dividing both sides of equation 33 by $A(q, \theta)$ yields

$$y(t) = \frac{B(q, \theta)}{A(q, \theta)}u(t - n_k) + \frac{1}{A(q, \theta)}e(t) \quad (34)$$

and by relating back to equation 19, it is apparent that

$$\hat{G}(q, \theta) = \frac{B(q, \theta)}{A(q, \theta)} \quad \text{and} \quad \hat{H}(q, \theta) = \frac{1}{A(q, \theta)}. \quad (35)$$

The order of $A(q, \theta)$ is selected sufficiently high to capture the dynamics of the true noise model (e.g., $n_a \approx 15$). The Box-Jenkins model provides the flexibility necessary to vary the noise model order and compare the DI method with and without correct noise model order specification. A series of experiments are conducted to ensure closely matching impulse response (IR) coefficients for both the true and estimated inverse noise models.

Upon estimating the inverse noise model, the IO data is filtered with $A(q, \theta)$, i.e.,

$$u^f(t) = A(q, \theta)u(t) \quad \text{and} \quad y^f(t) = A(q, \theta)y(t). \quad (36)$$

Inserting IO expressions from equation 36 into equation 34 yields

$$\frac{y^f(t)}{A(q, \theta)} = \frac{B(q, \theta)}{A(q, \theta)} \frac{u^f(t - n_k)}{A(q, \theta)} + \frac{1}{A(q, \theta)} e(t) \quad (37)$$

and multiplying both sides of equation 37 by $A(q, \theta)$ yields the following OE model

$$y^f(t) = \frac{B(q, \theta)}{A(q, \theta)} u^f(t - n_k) + e(t). \quad (38)$$

Finally, this OE model can be solved via the PEM and equation 23 as described before.

For the MIMO system, equation 38 extends to a system of three output equations with the expression for $y_3^f(t)$ shown for reference as follows:

$$\begin{aligned} y_3^f(t) &= \frac{B_{31}(q, \theta)}{A_{31}(q, \theta)} u_1^f(t - n_{k_{31}}) + \frac{B_{32}(q, \theta)}{A_{32}(q, \theta)} u_2^f(t - n_{k_{32}}) \\ &\quad + \frac{B_{33}(q, \theta)}{A_{33}(q, \theta)} u_3^f(t - n_{k_{33}}) + e_3(t). \end{aligned} \quad (39)$$

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 OE expressions are obtained by substituting the first order model estimates for $A(q, \theta)$ and $B(q, \theta)$ into equation 39. The expressions for $y_1^f(t)$ and $y_2^f(t)$ are obtained in an analogous manner. 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.

The closed-loop DI method is implemented as a benchmark in this work. It begins by defining model orders and proceeds by solving equation 26 directly with process IO data. Although the noise model is unknown, an estimate of the noise model order is necessary for DI. The sensitivity of identification to an incorrect noise model is considered in the experimental investigation.

Implementation

Closed-loop SI is fundamental to both the PM and SI aspects of the adaptive control framework. 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 large orders of the first stage ARX model require estimation of many parameters which can result in high variance estimates due to the limited amount of 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. Regularization constants Λ and R are determined to reduce the variance of the model estimates and the regularized ARX model is solved, i.e.

$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^N \hat{\epsilon}^2(t, \theta) + \Lambda(\theta - \theta^*)^T R(\theta - \theta^*), \quad (40)$$

where θ^* is the mean of the prior distribution of θ which is assumed to be Gaussian.⁴¹

Since the PM algorithm uses small amounts of routine operating data the signal to noise ratio is improved by filtering out high frequency noise. This denoising step is performed after the IO data is filtered with the inverse noise model estimate. To improve the OE estimation the range of parameters are bounded using *a priori* knowledge of acceptable values. The excitation experiment has a duration of 3000 samples (i.e., ≈ 4.2 hours). All IO data from the excitation experiment is collected by the SI algorithm to perform both DI and ARX-OE identification. Both the OE model in the ARX-OE method and the DI model are estimated with first order numerator and denominator polynomials. The DI method assumes a first order numerator and denominator structure for the noise model. Once the identification results are determined and the controller update is complete, regular operation resumes.

Performance monitoring

As applications of model-based control methods increase, so too does the need for adequate performance assessment and model validation. 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.⁴² Controller PM is a well established field that has drawn increased interest since the development of the minimum variance benchmark by Harris in 1989.⁴³ Some other notable developments for MPC assessment include benchmarking with historical MPC objective function data, monitoring the prediction error to determine model quality and using principal component analysis (PCA) or partial least squares (PLS) to detect controller degradation.^{44,45}

Closed-loop prediction error monitoring presents challenges with respect to diagnosing large prediction errors as they can be due to MPM, set-point changes as well as measured and unmeasured disturbances.⁴⁶ Since both the noise model and the 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.⁴⁷ Since monitoring is performed almost constantly it is beneficial to employ a technique that passively monitors the process. In this work an innovative MPM detection algorithm using support vector machines (SVMs) on routine operating data is employed. For a thorough exposition of this MPM detection method please refer to the seminal work.^{48,50}

Model-plant mismatch detection with a one-class support vector machine

The MPM detection implemented in this work uses the ARX-OE identification method on routine operating data in combination with a one-class SVM in order to distinguish between changes in process models and noise models. Due to the presence of noise and the lack of significant excitation in routine operating data, directly comparing the ARX-OE estimates to the controller model does not provide a reliable MPM detection procedure. 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, a one-class 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.

To train the SVM while accounting for gain and time constant mismatch the model estimates are represented as finite impulse response (FIR) coefficients. The one-class SVM is formulated as⁴⁹

$$\begin{aligned} \min_{w, \xi, b} \quad & \frac{1}{2} \|w\|^2 - b + \frac{1}{vl} \sum_{i=1}^l \xi_i, \\ \text{s.t. } \quad & w^\top \Phi(x_i) \geq b - \xi_i, \quad \xi_i \geq 0. \end{aligned} \tag{41}$$

where w and b are the respective slope and offset of the separating hyper-plane in the high-dimensional feature space Φ generated with a Gaussian kernel $\kappa(x, y) = e^{-\|x-y\|/c}$. The parameters v and c tune the support vectors and the shape of the Gaussian kernel, respectively. The slack variable ξ permits local variations along the hyperplane boundary. Employing the kernel trick allows the dual optimization problem to be solved in the lower dimensional input space providing computational efficiency. Solving the dual problem pro-

vides the decision function $h(x)$ (where x is the FIR vector) which is used to define the MPM indicator that ultimately raises the MPM alarm according to a user-specified threshold.⁴⁸

The SVM implemented in this work 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. 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 and the FIR coefficients of the estimated process models create a reference cluster as shown in Figure 14. During the testing stage the FIR coefficients of current process model estimates are

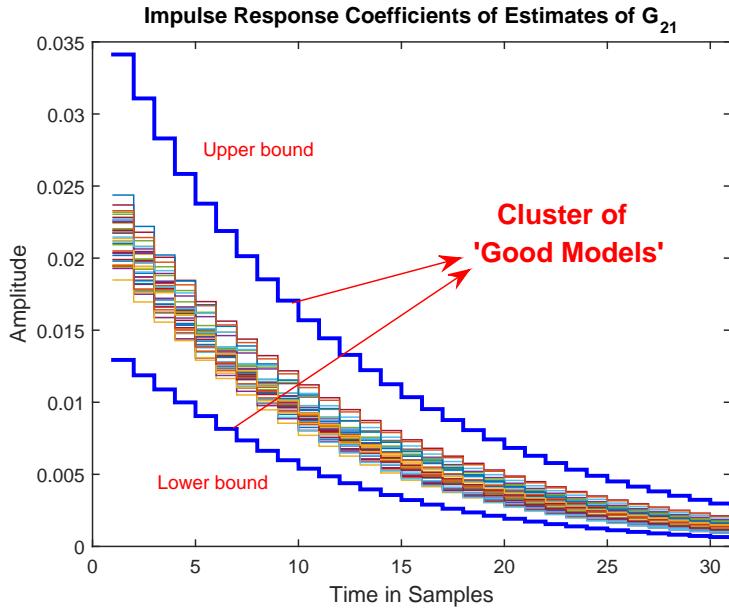


Figure 14: Training the SVM on FIR coefficients without MPM⁵⁰

compared with the reference cluster in the high dimensional feature. The decision function $h(x)$ generates a decision boundary that indicates whether the testing data belongs to the cluster of reference models or whether MPM has occurred. 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.

Implementation

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). The closed-loop ARX-OE training windows are 1440 samples long with a step size of 60 samples (i.e., 2 hour window length and 5 minute step size). The process and noise model estimates during training are used to generate the reference cluster considered free of MPM. During testing, the SVM algorithm takes current ARX-OE estimates and 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. The plant and noise model indicators are observed over the 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 SVM is retrained. If the noise model does not exhibit mismatch and the process model indicators reach the alarm threshold an identification experiment is triggered by designing an excitation signal, r_t^* , as described in what follows.

Input design

During the identification experiment an excitation signal, r_t^* , is added to the process input, u_t , to generate informative process IO data. Informative IO data has a relatively high signal to noise ratio which enables improved SI and more accurate parameter estimation. A common 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.⁵¹ 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.⁵²

In this work, the Fisher information matrix is maximized. The Fisher information matrix

is defined as the inverse of the parameter covariance matrix as follows $F = \text{cov}(\theta_N)^{-1} = \Psi\Psi'$ where the matrix Ψ is composed of a series of vectors ψ , i.e., $\Psi = [\psi_t, \dots, \psi_{t-N+1}]$. 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., $\psi_t = [y'_{t-1}, \dots, y'_{t-n_a}, u'_{t-d}, \dots, u'_{t-d-n_b}]'$, such that the vector ψ_t and the parameter vector θ model the process output as $y_t = \theta\psi_t$. For a process described by an ARX model the parameter vector θ is given by the expression $\theta = [a_1, \dots, a_{n_a}, b_1, \dots, b_{n_b}]$. The trace of the Fisher information matrix $R_u = \text{trace}(F)$ is

$$R_u = \sum_{p=1}^n \sum_{j=1}^{n_a} \sum_{i=j}^N (y_p(t-i))^2 + \sum_{q=1}^m \sum_{j=1}^{n_b} \sum_{i=j}^N (u_q(t-n_k+1-i))^2 \quad (42)$$

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.⁵³

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

$$\begin{aligned} & \max_{u_t} \tilde{R}_u \\ & \text{subject to} \quad y^L \leq y(t) \leq y^H, \\ & \quad u^L \leq u(t) \leq u^H, \\ & \quad |u(t) - u(t-1)| \leq \Delta u^H, \end{aligned} \quad (43)$$

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

$$\tilde{R}_u = \sum_{q=1}^m \left(\frac{1}{\Delta u_q^H} \text{var}(u_q^*) \right). \quad (44)$$

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 equation 42 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 equation 43 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 originally a solution based on visiting each vertex individually was considered computationally infeasible.⁵³ However, in this work the number of input variables is $m = 3$, so with a small prediction horizon the problem can be solved by 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 allowable step size are considered.

Implementation

Excitation signals are determined for 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. For efficient computation the prediction horizon is limited to a length of four steps after which the input value is held constant. The ID calculation determines the four step horizon at each new sample. Previous model estimates are used in combination with past and future input values to predict future outputs.

The maximum input movements are specified as $\Delta u_q^H = [3.0, 1.4, 0.6]$. Upper and lower bounds of input set-point deviation are set as $u_q^H = -u_q^L = [30, 15, 8]$ whereas upper and lower bounds for output set-point deviations are specified as $y_p^H = -y_p^L = [15, 1.0, 2.0]$. To account for both the time delay and the 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. The total prediction horizon has a length of 27 which is defined by the maximum time delay estimate plus a contingency factor of

5. Input sequences that meet the constraints are stored, values of \tilde{R}_u in equation 44 are calculated and the sequence that maximizes \tilde{R}_u is selected as the perturbation sequence.

Experimental setup

A series of experiments are conducted with an industrial MPC control system provided by Honeywell and the aforementioned MD process model. Specifically, a high fidelity paper machine simulator was provided by Honeywell that includes variable time delays according to changing machine speeds, a wide range of CD response shapes and nonlinear mass balance and dryer models accounting for consistency, retention and stock density. However, it is important to note that the specific configuration of the simulator implemented in this work is better characterized as medium fidelity since linear models were purposefully selected to demonstrate the proposed linear process model identification techniques.

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 updates the gain, time constant and time delay values of all process models.

A generalized illustration of the experimental simulation time-line is provided in Figure 15. One important common feature of all experimental simulations is that a $2 \times$ gain MPM exclusively in b_{11} is introduced at sample number $n_{MPM} = 4000$. 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 varies slightly across experiments 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 time to perform set-point changes and collect response data.

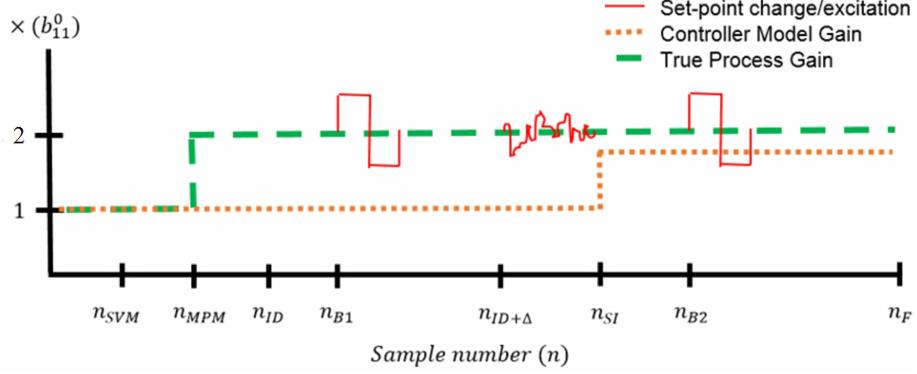


Figure 15: Illustrative time-line of the experimental simulations

After a specified waiting 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_1^0 . 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_1^0 . This series of set-point changes provides information on the dynamic response of the closed-loop system under MPM. An identical procedure is used at sample n_{SI} (i.e., after the controller update) to compare the dynamic closed-loop responses both with and without adaptive control.

A summary of the simulation parameters is provided in Table 4 below. Controller pa-

Table 4: 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^0	[174, 2.3, 6.2]	Nominal output set-points
u_q^0	[165, 115, 35]	Nominal input set-points
$\Delta y_p^0\%$	[10, 0, 0]	Magnitude of set-point change
$n_{ID+\Delta} - n_{ID}$	1600	Delay after MPM detection
$n_{B_1} - n_{MPM}$	500	Wait period before set-point change
$n_{\Delta y^0}$	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

rameters α^{sp} and α^d are used to tune the MPC for reference tracking performance and

disturbance rejection performance, respectively. The first study evaluates the ID algorithm relative to a PRBS perturbation benchmark of magnitude $u_q^{PRBS} = \pm [2.0, 1.0, 0.5]$ and the second study evaluates closed-loop ARX-OE method relative to a DI benchmark. The sensitivity of the DI method to noise model specification is evaluated with a sixth order noise model. The first and sixth order noise models are stable and inversely stable high pass (HP) filters. Both studies include an evaluation of the MPM detection and the closed-loop response to a set-point change before and after the adaptive update.

The dynamic response to a set-point change is evaluated qualitatively by observing the output response shape and quantitatively by comparing the observed response to the ideal response (based on the true process model). The mean square error (MSE) is used to determine the difference between the observed outputs y_i and the ideal outputs \tilde{y} , i.e.,

$$\epsilon = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2, \quad (45)$$

where n is the number of data points collected during the dynamic response test.

Experimental results

The first experiment involves using a PRBS for excitation to provide a reference for evaluating the ID algorithm. The $2\times$ gain MPM is detected by the PM algorithm after only 860 samples. 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 input profiles with PRBS excitation are shown below in Figure 16. These can be compared with the input profiles in Figure 17 where the proposed ID strategy is used. In Figure 17 the gain MPM is detected by the PM algorithm after 760 samples and the ID experiment begins at sample number 6340. The thick stock flow (u_1) variation during the ID excitation ranges primarily between 160 gpm and 170 gpm and thus satisfies the required constraints of 165 ± 30 gpm. Variation of

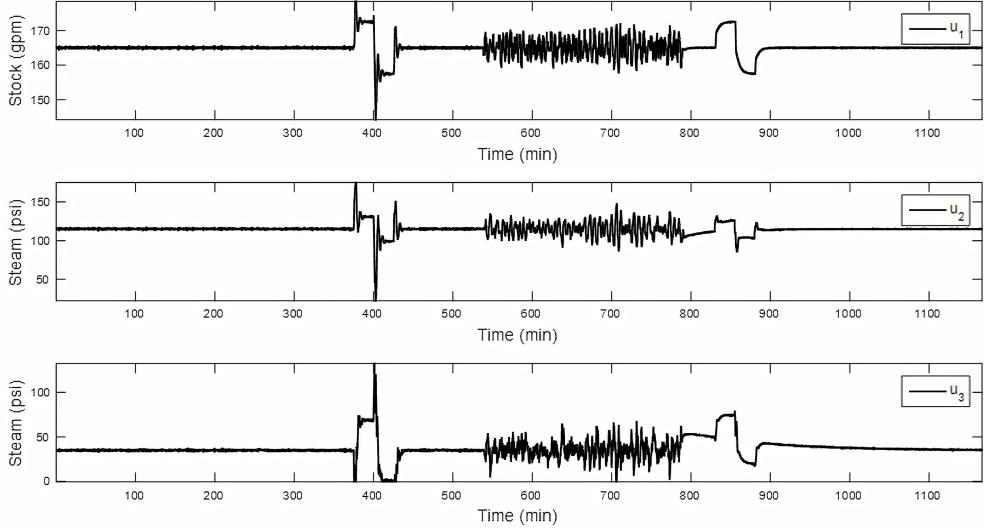


Figure 16: Input profiles with PRBS excitation

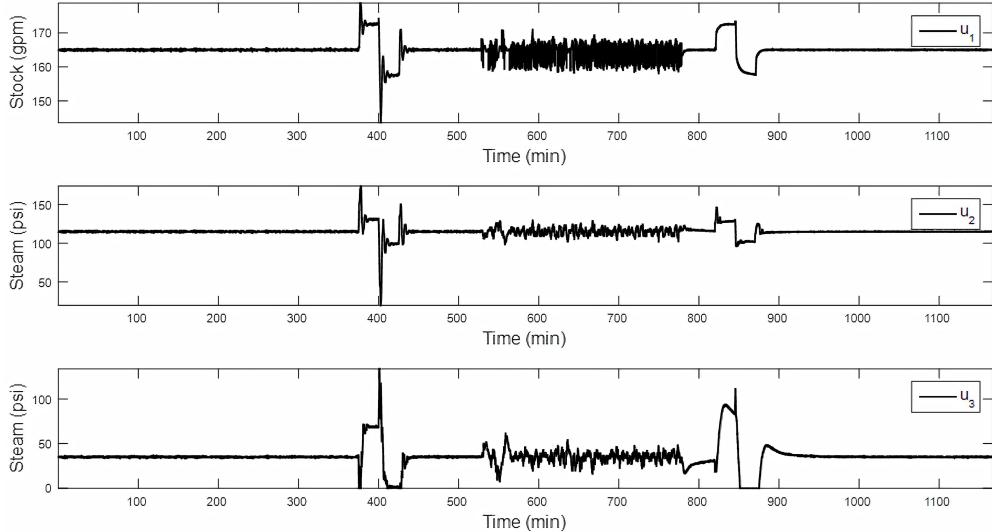


Figure 17: Input profiles with ID excitation

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 this is violated at around 550 min. In general the PRBS excitation creates larger and more frequent input constraint violations than the ID algorithm. 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 which ultimately degrades the output responses of y_2 and y_3 .

After the PRBS and ID experiments an ARX-OE estimate is used to update the con-

troller. The output profiles over the experiment duration with PRBS excitation and ID excitation are presented in Figure 18 and Figure 19, respectively. Recall, y_1 is the basis

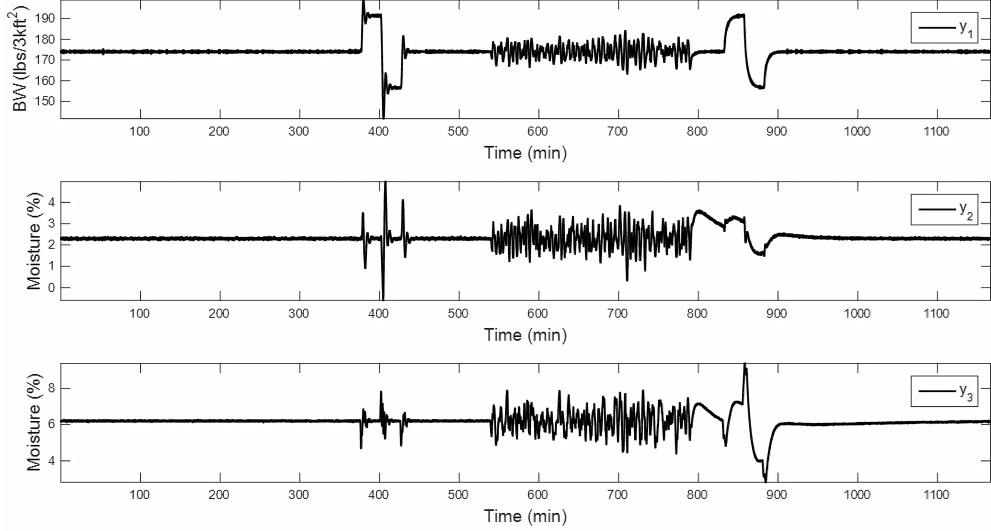


Figure 18: Output profiles with PRBS excitation

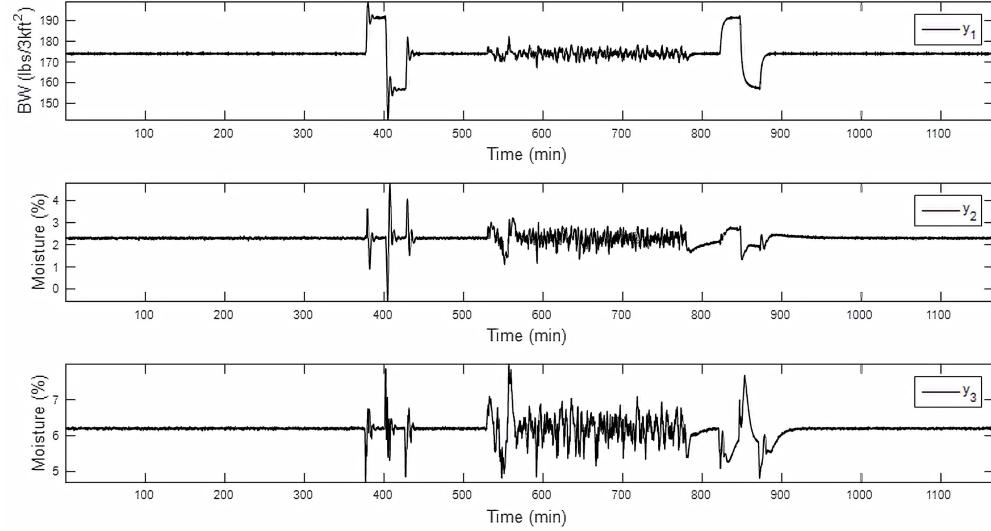


Figure 19: Output profiles with ID excitation

weight (BW) measured in $\text{lbs}/(3000\text{ft}^2)$, y_2 is the press moisture (%) and y_3 is the reel moisture (%). With ID excitation the press moisture and reel moisture exhibit a lower magnitude of disruption after the adaptive update. Alternatively, with PRBS excitation the y_1 set-point change creates a higher magnitude disruption in the reel moisture after the adaptive update.

Each of the process outputs obey the specified upper and lower bound constraints during the ID excitation experiment whereas the PRBS excitation results in occasional violations of the output constraints. The increased variation in the output profiles during the PRBS identification experiment is only desirable if it results in an improved adaptive update and subsequent dynamic response.

To evaluate the effect of the adaptive control update we observe the dynamic response to set-point changes before and after the adaptive update. Focusing on the process outputs during the first set-point change (i.e., before the adaptive update) yields Figure 20. It is

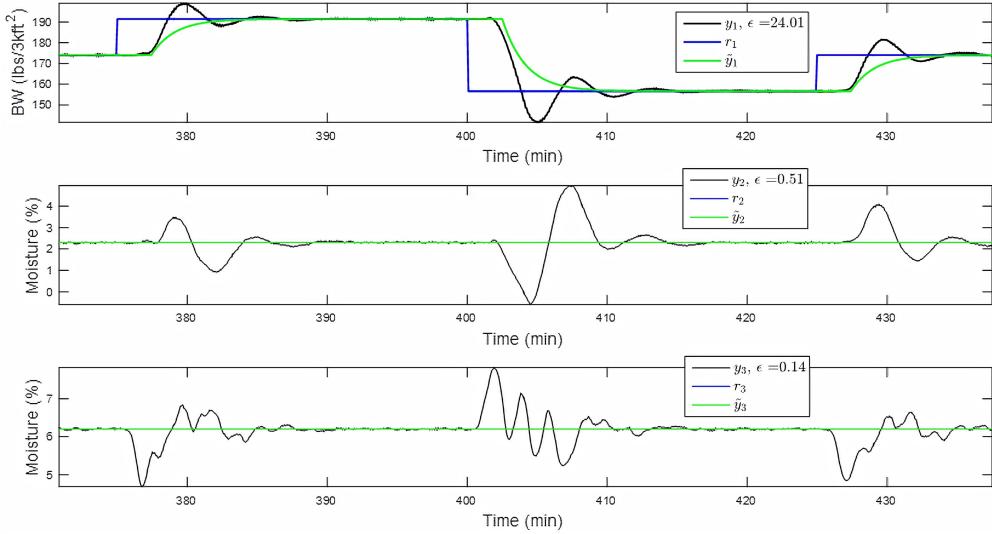


Figure 20: Closed-loop response to a 10% set-point change in y_1 with MPM

immediately clear that the closed-loop dynamic response under MPM exhibits undesirable over-shoot and oscillatory behavior. Due to the MIMO coupling the BW set-point change disrupts the press and reel moisture values as well. In all other experiments the closed-loop response to a set-point change exhibits the same behavior as the experimental configurations before the identification experiment are identical. Therefore, for concise presentation Figure 20 will serve as a qualitative reference for the closed-loop response to a set-point change with MPM. Quantitatively, over six trials the MSE (ϵ) before the adaptive update varies from 23.45 to 24.40 in y_1 , from 0.48 to 0.53 in y_2 and from 0.14 to 0.15 in y_3 .

The effect of the ARX-OE update is evaluated by comparing the closed-loop response

to a set-point change before and after the adaptive update. Moreover, the PRBS and ID techniques are evaluated based on their responses after the adaptive update as shown in Figure 21 and Figure 22, respectively. Regardless of whether a PRBS or ID signal is used,

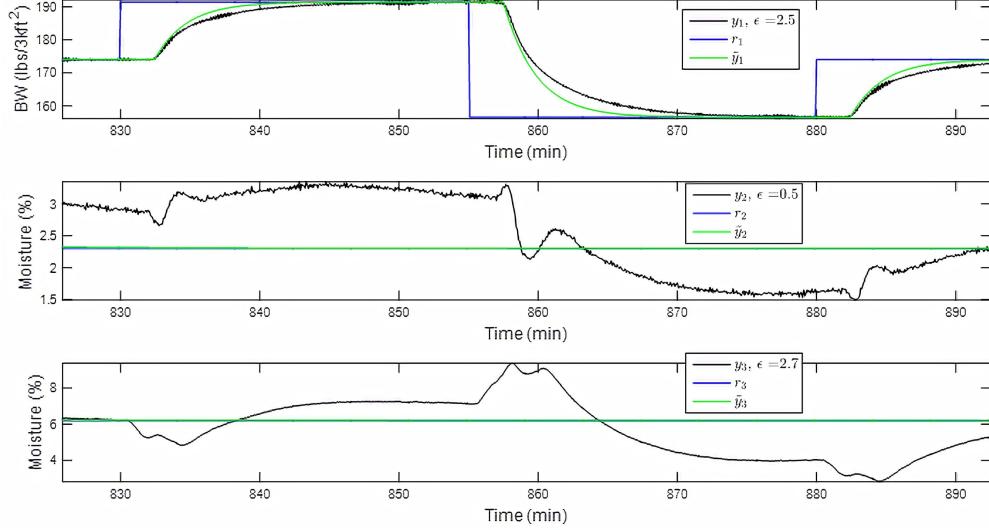


Figure 21: Closed-loop response after an ARX-OE update with PRBS excitation

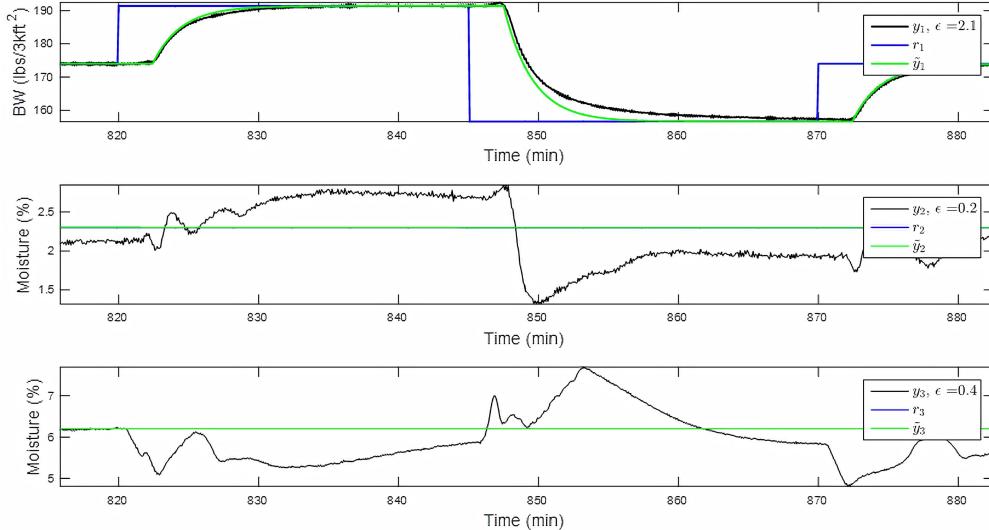


Figure 22: Closed-loop response after an ARX-OE update with ID excitation

the closed-loop response after the adaptive update exhibits considerable improvement. This is especially evident for the output y_1 that receives the set-point change as the MSE is reduced dramatically from $\epsilon = 24.01$ to $\epsilon = 2.5$ with PRBS or $\epsilon = 2.1$ with ID excitation. The MIMO

coupling introduces a higher MSE in y_3 after the adaptive update, particularly with a PRBS excitation. However, it is important to note that the transfer functions associated with y_3 did not exhibit MPM prior to the adaptive update but they were updated anyways. This issue will be addressed in future work by further analyzing the scores from the MPM detection algorithm and only updating the parameters from the transfer functions that exhibit MPM.

Evaluating the PRBS and ID excitation based on the dynamic response of the outputs yields a preference for the proposed ID technique. The ID excitation results in lower MSE values for all outputs but particularly for y_3 which has $\epsilon = 2.7$ for PRBS and $\epsilon = 0.4$ for ID excitation. Furthermore, the ID strategy is preferred because it adheres more strictly to the upper and lower bound constraints on the inputs and outputs during the excitation experiment. Improved control of process inputs and outputs during the excitation experiment is a significant benefit of the proposed ID method.

The second study of this report aims to evaluate the effectiveness of the ARX-OE adaptive update relative to a DI update. Since the DI method requires an estimate of the noise model order this study includes experiments where the noise model order is correctly and incorrectly specified. First, the closed-loop response to a set-point change after both ARX-OE and DI updates are compared with a correctly specified first order noise model. Afterwards the true noise model order is changed to six and the same analysis is conducted with the DI method still assuming first order noise. The proposed ID excitation strategy is conducted throughout the comparison of identification methods. Please note that the prediction length is decreased from four to three with the sixth order noise model for both the ARX-OE and DI methods so the relative comparison remains valid.

The closed-loop response with an ARX-OE update shown in Figure 22 can be compared to the closed-loop response with a DI update in Figure 23. In Figure 23 the $2\times$ gain MPM is detected by the PM algorithm after 860 samples and the identification experiment begins at sample number $n_{ID+\Delta} = 6460$. Since the same ID technique is applied the raw input and output profiles have been omitted as they are redundant. With a first order noise model

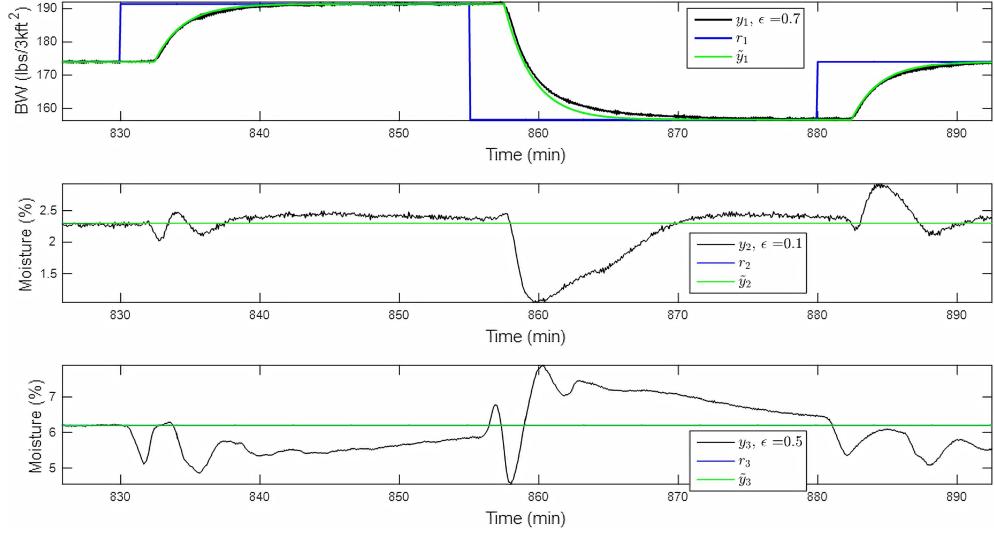


Figure 23: Closed-loop response after a DI update with ID excitation

the DI method provides a lower MSE for y_1 and y_2 but the MSE for y_3 is slightly higher relative to the ARX-OE method. These results are not surprising since the DI method assumes the correct noise model order whereas the ARX-OE method does not rely on such *a priori* knowledge. Even so the DI output response profiles are only marginally better than the ARX-OE response with the largest difference being the MSE in y_1 of $\epsilon = 2.1$ for the ARX-OE method and $\epsilon = 0.7$ for the DI method.

When a sixth order noise model order is used and the ID prediction length is dropped from four to three we obtain the closed-loop responses for ARX-OE and DI shown in Figure 24 and Figure 25, respectively. For both of these experiments 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. Once again the raw input and output profiles have been omitted as they are redundant considering the same ID technique was implemented.

With a sixth order noise model the ARX-OE update provides a lower MSE in y_1 than the DI method. Due to rounding to two decimal places and averaging over the number of samples the MSE values are shown as zero for y_2 in Figure 24 as well as y_2 and y_3 in Figure 25. In fact, comparing the results from first order and sixth order noise it is apparent that the MSE decreases in general with the exception of y_1 with the DI method. The sixth

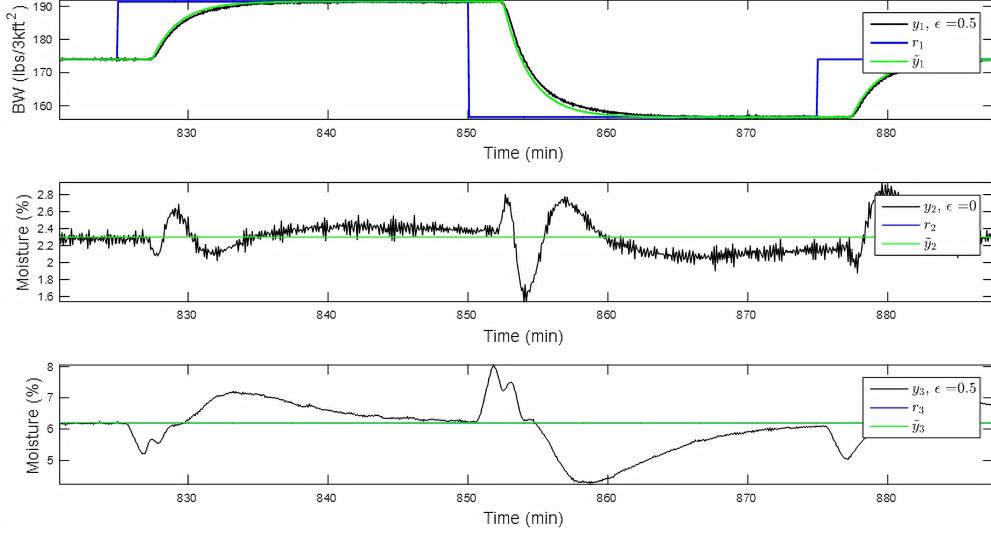


Figure 24: Closed-loop response after an ARX-OE update with sixth order noise

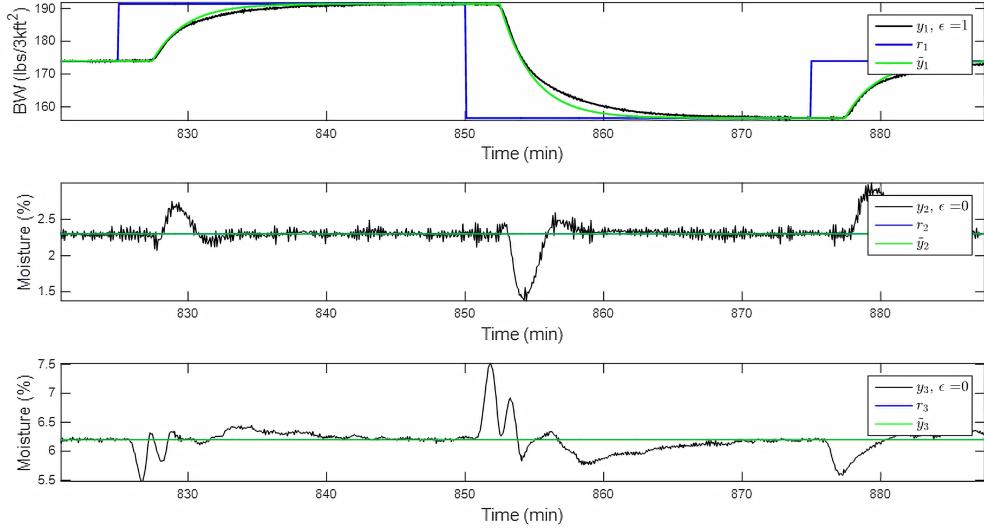


Figure 25: Closed-loop response after a DI update with sixth order noise

order noise models potentially introduce more excitation to the process resulting in a more accurate identification. Since the MPM occurs in G_{11} the closed-loop response of y_1 after the adaptive update is of particular interest. While the control performance of y_1 improves with ARX-OE method under a sixth order noise model the DI method reveals a slight decrease in control performance. This illustrates the benefit of ARX-OE when no *a priori* knowledge of the noise model order is available.

Ultimately, these results successfully illustrate the effectiveness of the unified adaptive

control strategy. The MPM was successfully detected by the PM algorithm in a timely fashion for every experiment. Moreover, the closed-loop response to a set-point change improves considerably with adaptive control. The proposed algorithm succeeds at eliminating the harmful over-shoot and oscillatory responses observed prior to the adaptive model update. A minor decrease in control performance can be observed in the variables that were not under MPM due to an unnecessary adaptive update. Future work will ensure the MPM detection algorithm is selectively updates only the transfer functions that undergo MPM. Another interesting avenue for future work which was brought to the authors attention is investigating an efficient single-step implementation of the ARX-OE method that treats the ARX parameters as hyper-parameters. This could yield improved accuracy at the cost of increased computation but it remains beyond the scope of this work.

Conclusion

In this paper complementary developments are presented towards MD-CD separation and MD adaptive control of a paper machine. Prior to implementing MD control the measured MD trend needs to be extracted through MD-CD separation. A novel MD-CD separation method based on compressive sensing is presented and demonstrated in comparison to two other model-free benchmarks with favorable results. The CS method is studied with two scanning sensors where it is shown to enable reconstruction of a high resolution MD trend for improved control of the MD process. One limitation is that the CS method needs to be implemented more efficiently for online estimation or significant computation resources are required. The CS method can be improved by implementing a debiasing strategy to mitigate the effects of soft-thresholding.

Once the MD profile is separated from the scanner signal it can be provided to the MPC system for the MD process. A more accurate MD trend can enable improved performance monitoring and identification of the MD process. The adaptive control framework

proposed in this work is capable of consistently identifying MPM, generating an excitation signal that meets IO constraints and identifying a new process model. An innovative MPM detection algorithm using an SVM provides fast and accurate detection of gain mismatch. The proposed input design strategy is compared to a PRBS excitation and shown to provide better controller performance as well as stricter adherence to IO constraints. Finally, a novel closed-loop model identification procedure is proposed and demonstrated to improve controller performance during a set-point change relative to the benchmark DI method.

Ultimately, this work has successfully demonstrated the functionality and performance of the proposed MD-CD separation and adaptive control techniques. Although there is room for further development, the contributions presented here have resulted in an algorithm that can automatically detect MPM and update 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.

It is important to note the direct impact of the MD trend determined through MD-CD separation on the ability to effectively harness the adaptive MD control framework. In this work we have proposed an MD-CD separation method that can increase the available MD measurements by at least an order of magnitude resulting in direct benefits for the PM, ID and SI algorithms. A unified experimental setup that combines MD-CD separation, adaptive MD control and adaptive CD control is desired for comprehensive closed-loop experiments.

Acknowledgement

The authors thank Honeywell Process Solutions, the National Science and Engineering Research Council of Canada and the Izaak Walton Killam Memorial Fund for funding this research through a collaborative research and development grant, a Canada Graduate Scholarship Master's award and a Killam Pre-Doctoral Memorial Fellowship.

References

- (1) Duncan, S.; Wellstead, P. Processing data from scanning gauges on industrial web processes. *Automatica* **2004**, *40*, 431–437.
- (2) VanAntwerp, J. G.; Featherstone, A. P.; Braatz, R. D.; Ogunnaike, B. A. Cross-directional control of sheet and film processes. *Automatica* **2007**, *43*, 191–211.
- (3) Chen, S.-C. Kalman filtering applied to sheet measurement. American Control Conference, 1988. 1988; pp 643–647.
- (4) Aslani, S.; Davies, M.; Dumont, G.; Stewart, G. Estimation of Cross and Machine Direction Variations Using Recursive Wavelet Filtering. *Pulp and Paper Canada* **2009**, 45–48.
- (5) Wang, X. G.; Dumont, G. A.; Davies, M. S. Estimation in paper machine control. *IEEE Control Systems* **1993**, *13*, 34–43.
- (6) Mijanovic, S. Paper machine cross-directional control near spatial domain boundaries. Ph.D. thesis, university of british columbia, 2004.
- (7) Wang, X. G.; Dumont, G. A.; Davies, M. S. Modelling and identifications of basis weight variations in paper machines. *IEEE Transactions on Control Systems Technology* **1993**, *1*, 230–237.
- (8) Dumont, G. A.; Jonsson, I. M.; Davies, M. S.; Ordubadi, F. T.; Fu, Y.; Natarajan, K.; Lindeborg, C.; Heaven, E. M. Estimation of moisture variations on paper machines. *IEEE transactions on control systems technology* **1993**, *1*, 101–113.
- (9) Chu, D.; Gheorghe, C.; Backstrom, J.; Forbes, M.; Chu, S. *Model predictive control and optimization for papermaking processes*; INTECH Open Access Publisher, 2011.
- (10) Wang, X. G.; Dumont, G. A.; Davies, M. S. Adaptive basis weight control in paper machines. Control Applications, 1993., Second IEEE Conference on. 1993; pp 209–216.

- (11) Aslani, S.; Davies, M.; Dumont, G.; Stewart, G. Detecting Aliasing Between Cross and Machine Direction Variations by Variable Sampling Rate. *measurement* **2009**, *20*, 1.
- (12) Halousková, A.; Kárny, M.; Nagy, I. Adaptive cross-direction control of paper basis weight. *Automatica* **1993**, *29*, 425–429.
- (13) Ohenoja, M. *One-and two-dimensional control of paper machine: a literature review*. October 2009.
- (14) Lang, D.; Fu, C. Variable Speed Scanning-A Fundamentally Better Way to Scan. *O Papel: revista mensal de tecnologia em celulose e papel* **2013**, *74*, 59–64.
- (15) Measurement Systems and Product Variability. *TAPPI, Paper Machine Quality Control Systems (QCS)* **2005**, *1*.
- (16) Heaven, E. M.; Jonsson, I. M.; Kean, T. M.; Manness, M. A.; Vyse, R. N. Recent advances in cross machine profile control. *IEEE Control Systems* **1994**, *14*, 35–46.
- (17) Lu, Q.; Rippon, L. D.; Gopaluni, R. B.; Forbes, M. G.; Loewen, P. D.; Backstrom, J.; Dumont, G. A. Cross-directional controller performance monitoring for paper machines. American Control Conference (ACC), 2015. 2015; pp 4970–4975.
- (18) Cutshall, K. The nature of paper variation. *Tappi journal* **1990**, *73*, 81–90.
- (19) Bialkowski, W. Newsprint Uniformity and Process Control The Untapped Competitive Edge. CPPA Conference, Montreal, Canada. 1990.
- (20) Stewart, G. E.; Gorinevsky, D. M.; Dumont, G. A. Feedback controller design for a spatially distributed system: The paper machine problem. *IEEE Transactions on Control Systems Technology* **2003**, *11*, 612–628.
- (21) Gendron, S. Fundamental Properties of Scanner Signals for the Correct Separation of Paper Machine CD and MD Variations. *PAPTAC 91st Annual Meeting* **2005**, *1*, D297—D302.

- (22) Lindeborg, C. A process model of moisture variations. *Pulp & paper Canada* **1986**, *87*, 42–47.
- (23) Nesic, Z.; Davies, M.; Dumont, G. Paper machine data compression using wavelets. Control Applications, 1996., Proceedings of the 1996 IEEE International Conference on. 1996; pp 161–166.
- (24) Karimzadeh, S. Online detection of picketing and estimation of cross direction and machine direction variations using the discrete cosine transform. Ph.D. thesis, University of British Columbia, 2008.
- (25) Chen, S.-C. Analysis of Sheet Variations—Insights of Two-Dimensional Variations. **2011**, 990–1002.
- (26) Chen, S. Determination of cd and/or md variations from scanning measurements of a sheet of material. 2012; <https://www.google.com/patents/US20120278032>, US Patent App. 13/457,870.
- (27) Towfighi, P. Estimating paper sheet process variations from scanned data using compressive sensing. Ph.D. thesis, University of British Columbia, 2011.
- (28) Duarte, M. F.; Eldar, Y. C. Structured compressed sensing: From theory to applications. *IEEE Transactions on Signal Processing* **2011**, *59*, 4053–4085.
- (29) Skelton, J.; Wellstead, P.; Duncan, S. Distortion of web profiles by scanned measurements. *Pulp & Paper Canada* **2003**, *104*, 81–84.
- (30) Candès, E. J.; Wakin, M. B. An introduction to compressive sampling. *IEEE signal processing magazine* **2008**, *25*, 21–30.
- (31) Arfken, G. B.; Weber, H. J. *Mathematical methods for physicists international student edition*; Academic press, 2005.

- (32) Chatterjee, A. An introduction to the proper orthogonal decomposition. *Current science* **2000**, *78*, 808–817.
- (33) Candès, E. J., et al. Compressive sampling. Proceedings of the international congress of mathematicians. 2006; pp 1433–1452.
- (34) Han, L.; Han, L.; Li, Z. Inverse spectral decomposition with the SPGL1 algorithm. *Journal of Geophysics and Engineering* **2012**, *9*, 423.
- (35) Ljung, L. *System identification*; Wiley Online Library, 1999.
- (36) Featherstone, A. P.; Braatz, R. D. Control-oriented modeling of sheet and film processes. *AIChE Journal* **1997**, *43*, 1989–2001.
- (37) Forssell, U.; Ljung, L. Closed-loop identification revisited. *Automatica* **1999**, *35*, 1215–1241.
- (38) Gustavsson, I.; Ljung, L.; Söderström, T. Identification of processes in closed loop—identifiability and accuracy aspects. *Automatica* **1977**, *13*, 59–75.
- (39) Van den Hof, P. Closed-loop issues in system identification. *Annual reviews in control* **1998**, *22*, 173–186.
- (40) Forssell, U.; Ljung, L. A projection method for closed-loop identification. *IEEE Transactions on Automatic Control* **2000**, *45*, 2101–2106.
- (41) Chen, L. L. T. What can regularization offer for estimation of dynamical systems? *IFAC Proceedings Volumes* **2013**, *46*, 1–8.
- (42) Jelali, M. An overview of control performance assessment technology and industrial applications. *Control engineering practice* **2006**, *14*, 441–466.
- (43) Harris, T. J. Assessment of control loop performance. *The Canadian Journal of Chemical Engineering* **1989**, *67*, 856–861.

- (44) Botelho, V.; Trierweiler, J. O.; Farenzena, M.; Duraiski, R. Methodology for detecting model–plant mismatches affecting model predictive control performance. *Industrial & Engineering Chemistry Research* **2015**, *54*, 12072–12085.
- (45) Botelho, V.; Trierweiler, J. O.; Farenzena, M.; Duraiski, R. Perspectives and challenges in performance assessment of model predictive control. *The Canadian Journal of Chemical Engineering* **2016**, *94*, 1225–1241.
- (46) Shah, S. L.; Patwardhan, R.; Huang, B. Multivariate controller performance analysis: methods, applications and challenges. AICHE Symposium Series. 2002; pp 190–207.
- (47) Badwe, A. S.; Gudi, R. D.; Patwardhan, R. S.; Shah, S. L.; Patwardhan, S. C. Detection of model-plant mismatch in MPC applications. *Journal of Process Control* **2009**, *19*, 1305–1313.
- (48) Lu, Q.; Gopaluni, R. B.; Forbes, M. G.; Loewen, P. D.; Backström, J. U.; Dumont, G. A. Model-Plant Mismatch Detection with Support Vector Machines. *IFAC-PapersOnLine* **2017**, *50*, 7993–7998.
- (49) Schölkopf, B.; Platt, J. C.; Shawe-Taylor, J.; Smola, A. J.; Williamson, R. C. Estimating the support of a high-dimensional distribution. *Neural computation* **2001**, *13*, 1443–1471.
- (50) Lu, Q. Adaptive model-predictive control and its applications in paper-making processes. Ph.D. thesis, University of British Columbia, 2018.
- (51) Zhu, Y. Multivariable process identification for MPC: the asymptotic method and its applications. *Journal of Process Control* **1998**, *8*, 101–115.
- (52) Patwardhan, R. S.; Gopaluni, R. B. A moving horizon approach to input design for closed loop identification. *Journal of Process Control* **2014**, *24*, 188–202.

- (53) Yousefi, M.; Rippon, L.; Forbes, M.; Gopaluni, R.; Loewen, P.; Dumont, G.; Backstrom, J. Moving-Horizon Predictive Input Design for Closed-Loop Identification. *IFAC-PapersOnLine* **2015**, *48*, 135–140.

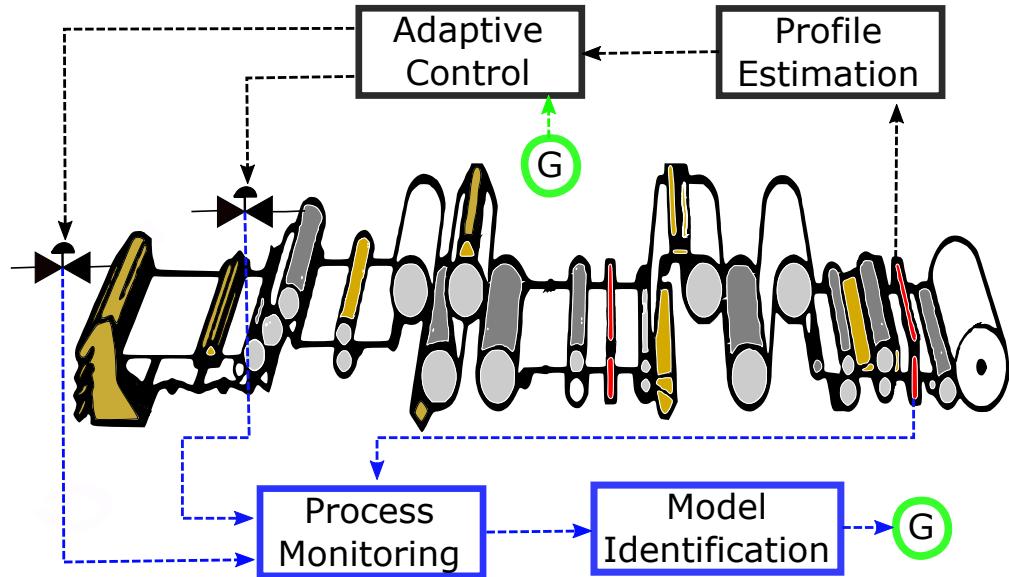


Figure 26: For table of contents only