

Paper Machine Data Analysis Using Wavelets

by Zoran Nestic

B.Sc. University of Belgrade, Yugoslavia, 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF ELECTRICAL ENGINEERING

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

February 1996

© Zoran Nestic, 1996

In presenting this thesis in partial fulfillment of the requirements of an advanced degree at the University of British Columbia, I agree that the library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical Engineering

The University of British Columbia

2356 Main Mall

Vancouver, BC

V6T 1Z4 Canada

Date: February 1996

Abstract

The thesis describes a new approach to paper machine process data analysis using one-dimensional and two-dimensional discrete wavelet transforms. These techniques have been adapted from a general theory that has been developed in recent years on the application of wavelets to signal analysis. Application areas in which the theory was first applied have included image processing and bandwidth compression for communications.

Two main applications of the discrete wavelet transform have been analyzed in this thesis. First, an analysis of the use of wavelets for processing scanned data representing basis weight and moisture variations on a paper machine has been carried out. It has been shown that wavelets are effective for the detection of process signals in noisy data, so leading to better estimation and visualization of the machine direction and cross machine variations in process data. The second main application of the method has been to allow significant compression of the process data without diminishing the ability to reconstruct accurate profiles. It has been shown that the compression method can be embedded into the estimation algorithm, producing excellent results without a major expense in computation time.

It has been shown that, in both applications, the new methods produce results superior to the industrially accepted procedures. For appropriate choice of wavelets, profile estimates are improved over those obtained using exponential filtering or other standard analysis methods. The data compression technique presents a new concept in paper machine data analysis and the author is not aware of any previous references to this subject. The ability to reduce data storage requirements is of importance in mill-wide process monitoring systems.

A comprehensive analysis of the proposed algorithms has been carried out on a variety of simulated data sets for which the true process variations are known. Industrial data has also been analyzed and it is apparent that the method had many desirable characteristics.

Table of Contents

Abstract	ii
List of Tables	v
List of Figures	vi
Acknowledgments	ix
1 Introduction	1
1.1 Measurement of Paper Properties	1
1.2 Review of Standard Estimation Methods	2
1.3 Motivation	3
1.4 Outline of Thesis	4
2 Wavelet Theory	5
2.1 Definitions	5
2.2 Wavelets as Basis Functions	6
2.3 Wavelet Properties	8
2.4 Implementation of the Discrete Wavelet Transform	9
2.5 Detection of Signals in Noise and Data Compression	11
2.5.1 De-Noising	12
2.5.1.1 Thresholding methods	13
2.5.1.2 Threshold Selection	14
2.5.1.3 De-Noising of Two-Dimensional Signals	15
2.5.2 Compression	16
3 Analysis of Paper Machine Process Data	19
3.1 Matlab™ Toolbox for Wavelet Analysis of Paper Machine Data	19
3.1.1 Padding of the Two-Dimensional Signals	21

3.2	Applying Wavelet De-noising to Simulated Scanned Data	21
3.2.1	Edge Effects and Step Changes	21
3.2.2	Smooth Surfaces	29
3.2.3	Grade Profile Change	33
3.2.4	Data Created by the Moisture Variation Model (MVM)	38
3.2.5	Bumps and Streaks	42
3.3	Wavelet Analysis of Industrial Data	47
3.3.1	Choice of an Estimator	47
3.3.2	Pulp Machine Moisture Content Data	47
3.3.3	Grade Change for a Paper Machine	56
4	Other Applications of Wavelets in Paper Making Process	59
4.1	Compression of Paper Machine Data	59
4.2	Wavelet Estimation for Control Purposes	73
5	Conclusion	75
	References	78
A	Exponential Multiple Scan Trending (EXPO) Algorithm	82
B	Estimation and Identification of Basis Weight and Moisture Content (EIBMC) Algorithm	83
B.1	Process Models	83
B.1.1	Moisture	83
B.1.2	Basis Weight	84
B.2	The EIBMC Algorithm	85
B.2.1	Moisture	85
B.2.2	Basis Weight	86
	References	87
C	Wavelet Filter Coefficients	88

List of Tables

3.1 Edge effects: thresholds and estimated noise at each resolution level.	25
3.2 Edge effects: summary of the results	26
3.3 Edge effects: best filter performances.	27
3.4 Smooth surfaces: thresholds and estimated noise at each resolution level.	29
3.5 Smooth surfaces: summary of the results	29
3.6 Smooth surfaces: best filter performances.	30
3.7 Grade change: thresholds and estimated noise at each resolution level.	34
3.8 Grade change: summary of the results	34
3.9 Grade change: best filter performances.	34
3.10 MVM test data: thresholds and estimated noise at each resolution level.	39
3.11 MVM test data: summary of the results	39
3.12 MVM test data: best filter performances.	40
3.13 Bumps and streaks: summary of the results	43
3.14 Bumps and streaks: best filter performances.	44
3.15 Filters used to analyze the industrial data.	48
4.16 Key table	59
4.17 Edge effects: summary of compression results	60
4.18 Smooth surfaces: summary of compression results	60
4.19 Grade change: summary of compression results	61
4.20 MVM test data: summary of compression results	61
4.21 Bump data: summary of compression results	62
4.22 Pulp machine moisture data: summary of compression results	69

List of Figures

1.1 Paper Machine	1
1.2 Sensor path	2
2.1 Block diagram of one-dimensional DWT.	9
2.2 Block diagram of one-dimensional inverse DWT.	10
2.3 Block diagram of two-dimensional DWT.	10
2.4 Signal in time-domain and its wavelet coefficients	11
2.5 Noisy signal in time-domain and its wavelet coefficients	12
2.6 Block diagram of a compressor/decompressor without a quantizer.	16
2.7 Block diagram of a compressor/decompressor with a quantizer.	17
2.8 Quantizer mapping function.	17
3.1 Wavelet filters used in the analysis.	20
3.2 Edge effects: CD+MD before and after the noise has been added.	23
3.3 Edge effects: estimated profiles.	24
3.4 Edge effects: estimation errors as functions of time (scan number).	25
3.5 Edge effects: estimation errors as functions of time (scan number) for a smooth estimator.	27
3.6 Edge effects: estimation results for a smooth estimator.	28
3.7 Smooth surfaces: composite profiles with and without noise.	31
3.8 Smooth surfaces: estimated profiles.	32
3.9 Smooth surfaces: estimation errors.	33
3.10 Grade change data set: CD+MD before and after the noise has been added.	35
3.11 Grade change test data set: estimated profiles.	36

3.12	Grade change test data set: estimated profiles (images).	37
3.13	Grade change test data set: estimation errors.	38
3.14	MVM test data: estimation errors.	40
3.15	MVM test data: estimated profiles.	41
3.16	MVM test data: CD profiles.	42
3.17	Bumps and streaks: estimated profiles.	45
3.18	Bumps and streaks: estimated profiles (images).	46
3.19	Pulp machine moisture data: estimated profiles.	49
3.20	Pulp machine moisture data: estimated profiles [1]	50
3.21	Pulp machine moisture data: estimated profiles [1] (images).	51
3.22	Pulp machine moisture data: estimated profiles [2]	52
3.23	Pulp machine moisture data: estimated profiles [2] (images).	53
3.24	Pulp machine moisture data: estimated profiles [3]	54
3.25	Pulp machine moisture data: estimated profiles [3] (images).	55
3.26	Paper machine basis weight data.	57
3.27	Paper machine basis weight data: estimated profiles.	58
4.1	MVM data set: energy distribution	63
4.2	MVM data set: distribution of nonzero wavelet coefficients.	64
4.3	2D DWT: coefficient storage.	64
4.4	MVM data set: compressed data.	66
4.5	MVM data set: compressed data (images).	67
4.6	Pulp machine moisture data: energy distribution	69
4.7	Pulp machine moisture data: distribution of nonzero wavelet coefficients.	70
4.8	Pulp machine moisture data: compressed data.	71

4.9 Pulp machine moisture data: compressed data (images)	72
4.10 Smooth surfaces: high resolution MD estimation	74
4.11 Pulp machine moisture data: high resolution MD estimation	74
B.1 Data indexing	83

Acknowledgments

I thank my thesis supervisor Professor Michael S. Davies for his support and guidance. I would also like to thank Professor Guy A. Dumont for his assistance and input during the course of this work

Thanks to all my friends at Pulp and Paper Centre and at the Department of Electrical Engineering for making my student days enjoyable, in particular, Anil Tuladhar, Ian McLeod, Gino Carrese, Leo Stocco, Lahoucine Ettaleb, Othman Taha, and Roger Shirt. Thanks to Brian McMillan for making the computers work.

Special thanks to Jahan Ghofraniha and Scott Morgan for their friendship and, of course, for sharing their rich libraries with me.

This thesis would have never been possible without the support and understanding of my friend and boss Professor Andy Black.

I am grateful to my wonderful wife, Maja Krzic, for her encouragement and patience during all those long years.

Chapter 1

Introduction

1.1 Measurement of Paper Properties

In the final stage of paper manufacturing, after pulping and bleaching have taken place, the mixture of water and fibre (stock) is delivered to the paper machine where moisture is first removed by drainage and mechanical pressing and then dried to produce a sheet of paper formed at the reel. A simplified diagram of a paper machine is given in Figure 1.1.

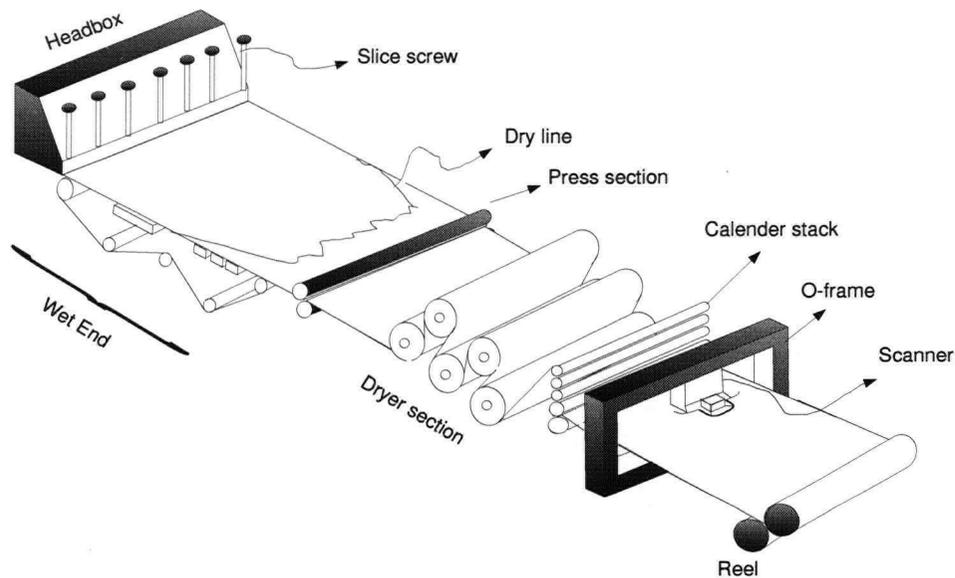


Figure 1.1: Paper Machine

To achieve uniformity and high quality in the paper produced, hundreds of control loops are needed as well as experienced operators. The most important control loops are those that maintain paper properties such as basis weight (mass of fibre per unit area) and moisture content (percent of water in the overall mass of paper sheet) close to the target values assigned for the grade of paper produced.

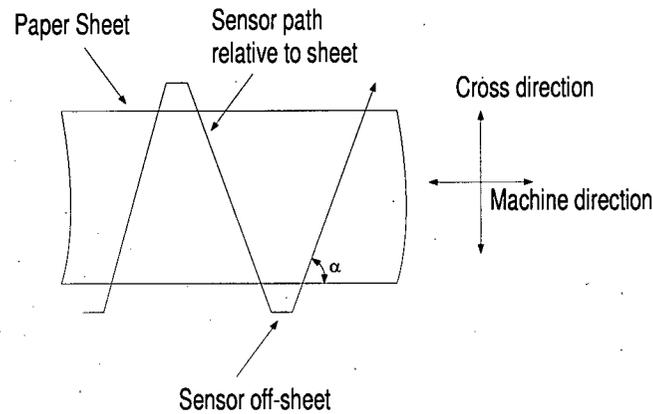


Figure 1.2: Sensor path

The sensor (see Figure 1.1) travels across the sheet (cross direction or CD). During one scan (one pass over the sheet) the sensor makes between 20 and 3000 uniformly spaced measurements (Morgan [30]). A scan period can vary from 10 to 60 seconds, depending on the installation. The speed of paper sheet is up to 25 m/s or more than a hundred times the sensor speed. The combination of the paper trajectory (in machine direction or MD) and the sensor movement (in CD) produces the zigzag pattern of measurements presented in Figure 1.2. This non-uniformly spaced sampling represents a major problem in estimation and control.

The MD variations, introduced by pressure fluctuations and consistency variations in the approach system and headbox are considered to be time-dependent, fast and independent of the CD position. Control action in MD direction is normally taken at least every scan. Changes in CD profiles, coming from nonuniformity in the headbox, slice lip, pressing and drying, are considered to be relatively slow, indeed nearly time-invariant. In practice, a CD control action is taken every 2 to 4 scans (Taylor [36], Jonsson [22]). Faster detection of MD upsets is increasingly desirable with changes in paper machine design.

It is obvious (see Figure 1.2) that the measured sequence of values contains information about both MD and CD variations. Decoupling the machine direction variations from the cross direction variations is not a trivial task.

1.2 Review of Standard Estimation Methods

Many estimation methods have been developed for processing scanned data since automatic

control of paper machines was introduced. Most industrial estimators are based on the Exponential Multiple Scan Trending (EXPO) algorithm (Dahlin [4]). Assuming slowly changing, zero mean CD profiles, this algorithm (see Appendix A) carries out exponential filtering of MD data at each CD position and produces CD profiles that are considered reasonably acceptable by the industry. The MD estimate, defined as the mean value of a scan, is available at the end of each scan.

Most recent algorithms (Lindeborg [25], Chen [2], Natarajan [32], Dumont [15], Jonsson [22], Morgan [30], Wang [37]) use stochastic models for basis weight and moisture variations. They can estimate CD (Least Squares filter) and/or MD profiles (Kalman filter), with the MD estimations updated at each data point [30, 37], leading to improved MD control bandwidth. The results of the estimation method described in this thesis are compared to the results obtained by the EXPO algorithm and the Estimation and Identification of Basis Weight and Moisture Content (EIBMC) algorithm [37, 30]. The details of the latter are given in the Appendix B.

1.3 Motivation

Recent trends in actuator design have allowed control systems for paper machines to involve an increased number of measurements and control signals. Increases in resolution for both sensors and actuators are the general trend. The increased resolution potentially leads to an improved overall control of paper machines if the multivariable control issues can be addressed effectively [36]. The increased amount of information causes problems in data processing (slows down the estimators and controllers), visualization of the process information, and in data storage. While the advances in computer technology tend to provide us with enough computing power, the visualization and data storage demands are becoming increasingly important. The machine operators benefit from a better presentation of the process data in order to be able to detect any change in the quality of final product. An 'enhanced' image of the two-dimensional (CDxMD) data is an important tool as part of the operator interface. The enhancements may include data filtering to provide 'smooth' images or to extract some other features, for example sudden changes in paper quality, or streaks. This type of data processing will lead to better diagnostics and feature recognition, first by the human operators and then, possibly, by the automated systems.

Quality control and production monitoring is increasingly important with the introduction of new standards (ISO-9001). Being able to record production and to trace back the events that caused a change in the paper quality has become essential for the paper companies. Data compression methods that provide an efficient data storage and retrieval method are desirable.

In this thesis a novel way of handling scanned paper machine data is introduced. An effort has been made to provide a framework that unites the estimation, visualizations, and data storage of the paper machine process data using wavelet analysis techniques. These techniques have been taken from the image processing literature and adopted to our purposes.

1.4 Outline of Thesis

An introduction to wavelet theory and its applications is given in the second Chapter. The third Chapter presents the results achieved by using wavelet filtering on simulated and industrial data. Applications of wavelets in paper machine process data compression and in the control of paper machines are presented in Chapter 4. The conclusions and the final remarks are given in Chapter 5. Models of basis weight and moisture variations, as well as the EIBMC and EXPO algorithms are given in appendixes.

Chapter 2

Wavelet Theory

In recent years wavelet theory has been established as a unified framework for a number of techniques used in different disciplines of science:

- Signal Processing (multiresolution analysis, subband coding, pyramid schemes)
- Applied Mathematics (multigrids for solving partial differential equations, approximation theory, grids)
- Physics (the theory of coherent states)
- Statistics (regression)

The roots of what is today known as wavelet theory can be found in the work done in the early 1930's (Littlewood-Paley techniques) or even earlier (1910, work done by A. Haar [17]). In the early 1980's, these different techniques were unified as wavelet theory. At that time, the word wavelet was suggested by Alex Grossman and Jean Morlet. In mid 1980's, a new orthogonal wavelet expansion was constructed by Pierre-Gilles Lemarié and Yves Meyer [24] and, by the end of the decade, Ingrid Daubechies [5] gave a method for the construction of wavelets, nonzero only on a finite interval and with arbitrarily high but fixed regularity. Most of the work carried out in applying wavelet theory in the detection of signals in noise, and in data compression has been published in the last five years.

In this Chapter an introduction to wavelet theory, one- and two-dimensional discrete wavelet transforms, wavelet filtering, and data compression is given, with an emphasis on the implementation issues.

2.1 Definitions

The *inner product* of two elements x, y of a real or complex linear space Y is denoted $\langle x, y \rangle$ and has the following properties:

- (i) $\langle x, y \rangle > 0$, whenever $x \neq 0$
- (ii) $\langle x, y \rangle = \overline{\langle y, x \rangle}$
- (iii) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$, α is a scalar
- (iv) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

The *norm*, denoted $\|y\|$, can be defined as $\|y\| = \langle y, y \rangle^{\frac{1}{2}}$. In an inner product space, two elements $x, y \in Y$ are defined to be orthogonal (denoted $x \perp y$) if $\langle x, y \rangle = 0$. Let $\{y_i\}$ be a set of mutually orthogonal ($\langle y_i, y_j \rangle = 0, i \neq j$) elements in an inner product space Y . Define $e_i = y_i/\|y_i\|$, then each e_i has unit norm and the set $\{e_i\}$ is called an orthonormal set.

A Hilbert space is a linear space on which an inner product has been defined, with a norm derived from the inner product and which is complete in this norm. L^2 is a Hilbert space with inner product

$$\langle x, y \rangle = \int_I x(t)y(t)dt \quad (2.1)$$

where the functions x, y are defined and integrable on a domain I . l^2 is a Hilbert space with inner product

$$\langle x, y \rangle = \sum_{i=1}^{\infty} x_i y_i \quad (2.2)$$

The condition number k of an operator T is defined as $k = \|T\| \cdot \|T^{-1}\|$. An operator is unitary if $T^* = T^{-1}$.

2.2 Wavelets as Basis Functions

Any signal in a Hilbert space can be approximated by a weighted sum of basis functions. The accuracy of the approximation will depend upon the class of signals, the choice of basis functions and the number of terms.

$$f(x) \approx \sum_{i=1}^N c_i \Psi_i(x) \quad (2.3)$$

Different sets of basis functions will lead to different sets of coefficients c_i . Two classes of representation are, for example, the sampling function and the sinusoid function. The signal representation by using sinusoids as the basis function is known as Fourier transform. This transform reveals information only about the signal's frequency content over the entire interval. On the other hand, the signal representation by a weighted sum of sampling functions will contain information only about the time domain behavior. This can be explained by comparing the functions' support (support is the interval over which a function is nonzero). The sampling function has infinitesimally small support and so produces precise time information but no frequency data. The sinusoids have infinite support and so provide spectral information but cannot establish time-domain variations. In

many cases a set of basis function that is a compromise between these two extremes can be used to track both spectral and time-domain disturbances. These functions should have finite support (good time resolution) of adjustable width [18] (good frequency resolution).

One such a set of basis functions is called wavelets. Wavelets are created by scaling and translating the same prototype function $\Psi(x)$ known as the *mother wavelet*. The scale factor is normally chosen to be a power of two, yielding the desired cascade octave band-pass filter structure. For the discrete wavelet transform, all the integer shifts of $\Psi(x)$ have to be considered. Therefore, the wavelet decomposition of the signal is given by

$$f(x) = \sum_j \sum_k c_{jk} \Psi_{jk}(x) \quad (2.4)$$

where

$$\Psi_{jk}(x) = 2^{\frac{j}{2}} \Psi(2^j x - k) = 2^{\frac{j}{2}} \Psi(2^j (x - 2^{-j} k)) \quad (2.5)$$

The coefficient $2^{-j} k$ is a dyadic point. The multiplier $2^{\frac{j}{2}}$ is needed to make the basis orthonormal.

The wavelet coefficients c_{jk} are computed by the *continuous wavelet transform*, which is the inner product of the signal $f(x)$ and the basis function $\Psi_{jk}(x)$.

Let the time and the frequency resolutions be defined (in rms sense) as

$$\Delta t = \sqrt{\frac{\int_{-\infty}^{\infty} t^2 |g(t)|^2 dt}{\int_{-\infty}^{\infty} |g(t)|^2 dt}} \quad (2.6)$$

$$\Delta \omega = \sqrt{\frac{\int_{-\infty}^{\infty} \omega^2 |G(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |G(\omega)|^2 d\omega}} \quad (2.7)$$

where $g(t)$ and $G(\omega)$ are basis function and its Fourier transform respectively. The denominators represent the energies of $g(t)$ and $G(\omega)$ respectively. The product of time and frequency resolutions is lower bounded [34] (Eq. 2.8).

$$\Delta t \Delta \omega \geq \frac{1}{2} \quad (2.8)$$

Eq. 2.8 is sometimes referred to as the uncertainty principle, or Heisenberg inequality. It implies that time and frequency resolutions cannot be controlled independently.

2.3 Wavelet Properties

Up to this moment, the choice of the mother wavelet $\Psi(x)$ has not been discussed. The choice of the function $\Psi(x)$ depends on the wavelet properties that are best suited to a particular problem. Some of the most important properties of wavelets are given below. For a more detailed discussion of wavelet properties see the work of Jawerth and Sweldens [20].

Orthogonality ($\langle \Psi, \Psi(\cdot - l) \rangle = \delta_l$) directly links the L^2 norm of a function to the norm of its wavelet coefficients by

$$\|f\| = \sqrt{\sum_i \sum_j c_{i,j}^2} \quad (2.9)$$

The fast wavelet transform, using orthogonal wavelets, is a unitary transformation. Orthogonality is also important in numerical calculations. An error present in the initial data will not grow under the transformation, and stable numerical computations are possible. In the multiresolution analysis orthogonality means that the projection operators into the different subspaces yield optimal approximations in the L^2 sense.

Compact support is needed to guarantee that the summations in the fast WT are finite (i.e. wavelet filters have finite impulse responses).

Symmetry is required for the generalized linear-phase filters (to prevent phase distortion).

Smoothness is a very important property for wavelet filtering (de-noising) and compression applications. A function is said to have M degrees of smoothness if its M^{th} derivative is continuous at all points. The reconstructed version of a function will be much smoother if the wavelet itself has a higher degree of smoothness. A higher degree of smoothness corresponds to better frequency localization of the filters.

Number of vanishing moments N for a sequence $\{g_k\}$ is given by Eq. 2.10 and is connected to the smoothness of the wavelet (and vice versa). A higher number of vanishing moments corresponds to a higher degree of smoothness.

$$\sum_k g_k k^p = 0, \text{ for } 0 \leq p < N \quad (2.10)$$

There is a trade-off among these properties and they are held to varying degrees by different families of wavelets. There are many researchers working on the wavelet design. Fundamental

and pioneering work has been done by Daubechies [6, 8, 5, 7], Meyer [29, 28], Cohen [3] and Wickerhauser [38].

Having chosen the appropriate wavelet, one can move towards the implementation of the discrete wavelet transform.

2.4 Implementation of the Discrete Wavelet Transform

The paper machine sensor provides a sequence of discrete measured values. Therefore, an efficient algorithm for the *discrete wavelet transform* (DWT) is needed. Mallat's pyramid algorithm [26] has been used in this thesis and it will be described in this section. Let the sequence s_k be

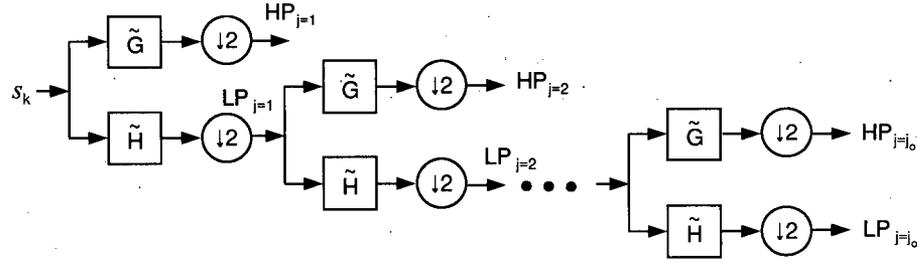


Figure 2.1: Block diagram of one-dimensional DWT.

of the length $n = 2^N$. The block diagram of a one-dimensional discrete wavelet transform (DWT) is given in Figure 2.1. It requires two properly designed wavelet filters: \tilde{G} — high pass filter and \tilde{H} — low pass filter, both of the length M . Examples of such filters are given in Figure 3.1 and Appendix C. These two filters are not independent since:

$$\tilde{g}_n = (-1)^{1-n} \tilde{h}_{1-n} \quad (2.11)$$

where \tilde{g}_n and \tilde{h}_n are filter coefficients of \tilde{G} and \tilde{H} respectively. The DWT is implemented by filtering the signal followed by down-sampling by a factor of 2 (Figure 2.1 and Eq. 2.12).

$$\begin{aligned} HP_k &= \sum_{i=0}^{M-1} \tilde{g}_i s_{2k-i}, \\ LP_k &= \sum_{i=0}^{M-1} \tilde{h}_i s_{2k-i} \end{aligned} \quad (2.12)$$

If the length of the original sequence s_k is n then the lowpass (LP) and the highpass (HP) component are each of the length $n/2$. This procedure of filtering and down-sampling can be repeated recursively

on the LP components of the signal. After N iterations the length of the lowpass and highpass components becomes equal to 1. The wavelet coefficients are given by:

$$\omega = \{LP_{j=N}, HP_{j=N}, HP_{j=N-1}, \dots, HP_{j=1}\} \quad (2.13)$$

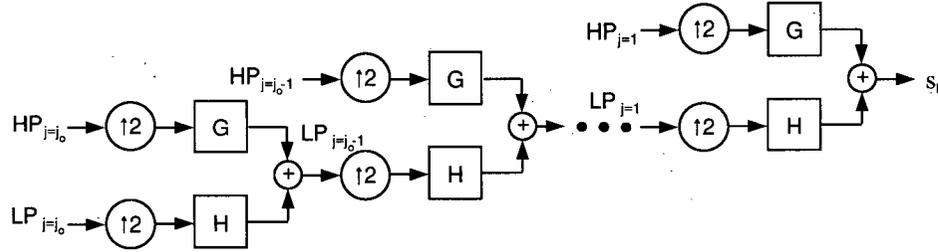


Figure 2.2: Block diagram of one-dimensional inverse DWT.

The inverse DWT is depicted in Figure 2.2. It is accomplished by upsampling of the highpass and lowpass components by inserting a zero after each element, and convolving the upsampled lowpass and highpass signals with the filters H and G respectively. The filter pair (H, G) is related to the forward DWT filter pair (\tilde{H}, \tilde{G})

$$\begin{aligned} g_n &= \tilde{g}_{-n} \\ h_n &= \tilde{h}_{-n} \end{aligned} \quad (2.14)$$

The fast DWT requires $O(n)$ operations. An actual algorithm in pseudo code is given in [20].

A two-dimensional DWT can be accomplished by two separate one-dimensional transforms [18] as presented in Figure 2.3.

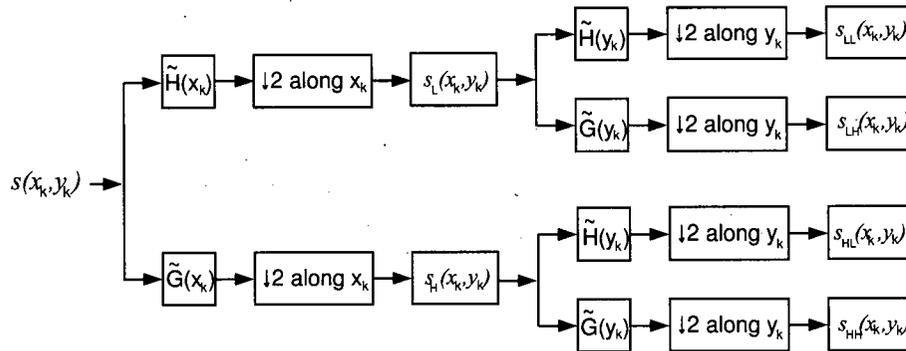


Figure 2.3: Block diagram of two-dimensional DWT.

When implementing the filtering of a signal by the filters H, G a choice has to be made regarding the signal padding at the ends of the sequence. A signal padding assumption is needed because both,

the filter and the signal lengths are finite. When the filter passes the last signal value (s_n)

$$\dots s_{n-4} \quad s_{n-3} \quad s_{n-2} \quad s_{n-1} \quad s_n \quad x_1 \quad x_2 \quad x_3 \quad \dots \quad \begin{matrix} h_1 & h_2 & h_3 & h_4 \end{matrix} \quad (2.15)$$

some values $\{x_1, x_2, x_3, \dots\}$ have to be added at the end of the signal. The simplest solution would be to pad zeros. Padding zeros introduces a discontinuity at that point and it is quite unlikely that zeros are a natural extensions of the signal. Knowing how sensitive the DWT is in registering discontinuities, which is a reason for using the DWT in the first place, it is likely that this artificial insertion will introduce a significant error.

A better approach is to consider the function to be periodic (i.e. $x_1 = s_1, x_2 = s_2, \dots$). Again, it will cause no error only if the behavior of the signal at s_1 and s_n do match. Other much more advanced approaches to this problem, are summarized in [20] (Meyer's boundary wavelets, Dyadic boundary wavelets).

2.5 Detection of Signals in Noise and Data Compression

The signal representation in wavelet domain can become very compact with only a few coefficients being needed to represent a complex signal (Figure 2.4). By contrast, broadband components associated with noise have their energy dispersed over the coefficients (Figure 2.5).

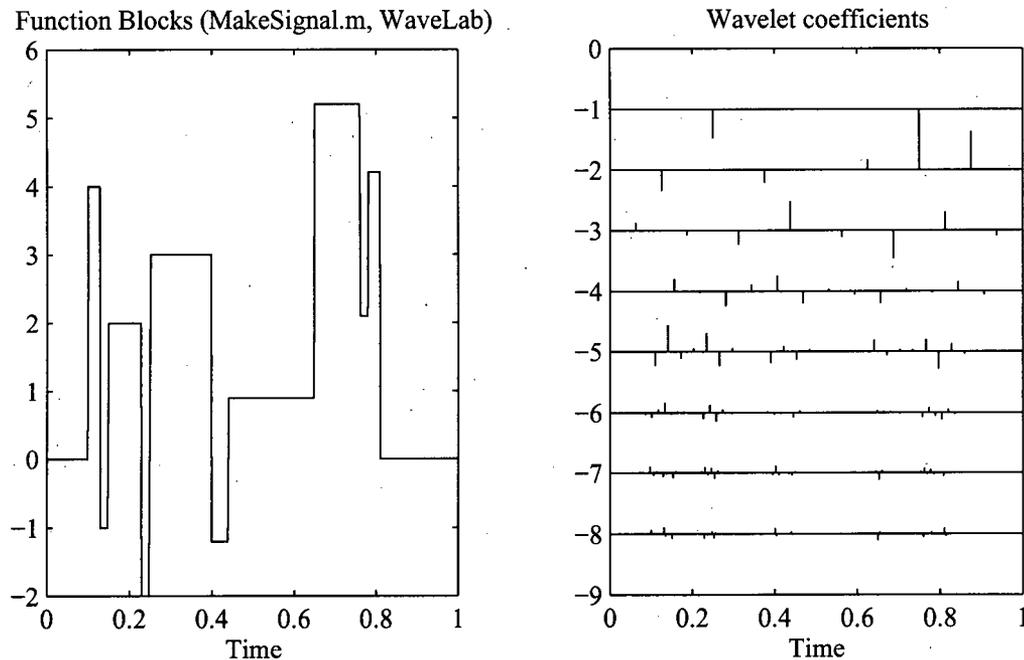


Figure 2.4: Signal in time-domain and its wavelet coefficients

This property makes the wavelet transform very suitable for estimation and data compression. By reducing (shrinking) the wavelet coefficients it is possible to remove the “less significant” features of a signal, most likely noise. Therefore an estimated, noiseless, signal can be obtained after the inverse DWT.

The fact that the signal energy is concentrated in a relatively small percentage of its wavelet coefficients, naturally led to the implementation of wavelets in data compression.

2.5.1 De-Noising

The term *De-Noising* describes, in an informal way, the various schemes which attempt to reject noise by damping or thresholding in the wavelet domain [11]. Different methods of wavelet filtering are very well documented in the literature. Most of the implementations are based on the work done by Donoho and Johnstone at Stanford University. They have shown [13, 12, 14, 10] that for the signal model:

$$y_i = f_i + \sigma \varepsilon_i, \quad i = 1, \dots, n \quad (2.16)$$

where $\varepsilon_i \stackrel{iid}{\sim} N(0, 1)$ is Gaussian white noise, the non-linear wavelet-coefficient thresholding produces estimates \hat{f} that are nearly optimal in the l^2 sense.

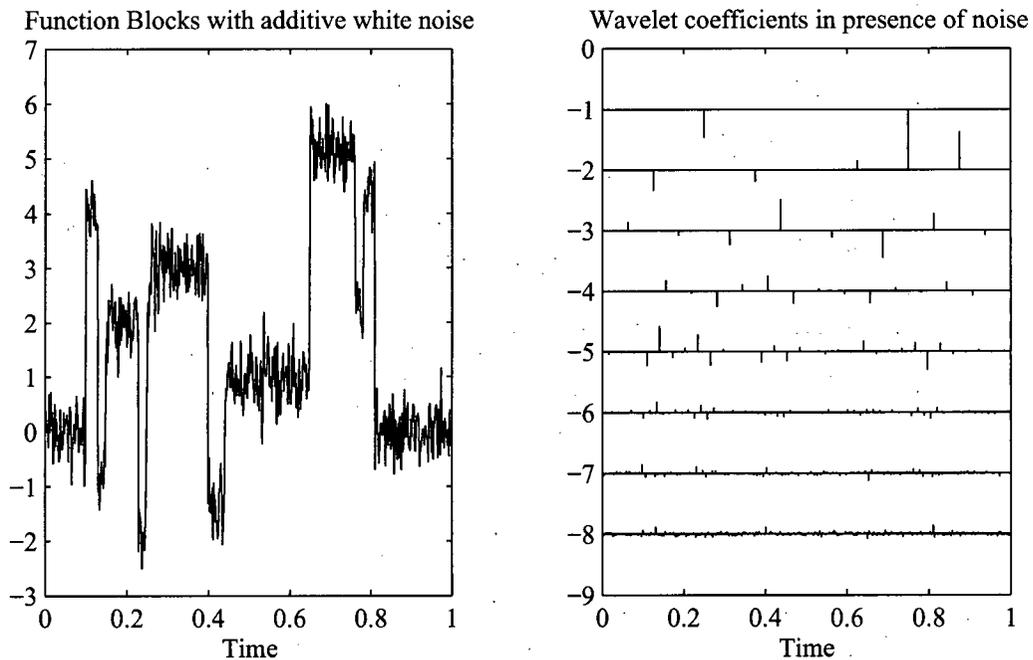


Figure 2.5: Noisy signal in time-domain and its wavelet coefficients

The wavelet regression estimator that is suggested by the Stanford group is based on a three-step procedure:

1. Find DWT of $y = (y_1, \dots, y_n)$

$$\omega = Wy \quad (2.17)$$

where operator W represents the wavelet transform;

2. Create a new set of wavelet coefficients ω^* by a non-linear modification of ω ;
3. Reconstruct an estimate of f by applying the inverse DWT to ω^* :

$$\hat{f} = W^T \omega^* \quad (2.18)$$

The step 2 is the most important part of the regression. Donoho and Johnstone have proposed thresholding of wavelet coefficients as the primary method of the modification. They have established that this method works because:

1. The DWT of Gaussian white noise is again white noise. Thus, it is evenly spread over all coefficients (every empirical wavelet coefficient contributes noise of variance σ^2 but only a very few wavelet coefficients contribute signal [14]).
2. The real signal f is represented by a small number of nonzero coefficients.

To achieve successful estimation, a threshold has to be found based on the estimated noise level.

2.5.1.1 Thresholding methods

There are two main types of thresholding: *Hard Thresholding* and *Soft Thresholding*.

Hard thresholding keeps all the coefficients that are greater than the threshold and sets the others to zero. Expressed in the standard MatlabTM notation, it is given by

$$\omega_{hard}^* = \omega \cdot (abs(\omega) > threshold) \quad (2.19)$$

Soft thresholding sets all the coefficients smaller than the threshold to zero and shrinks the others by the threshold value. It is given by

$$\omega_{soft}^* = sign(\omega) \cdot (abs(\omega) - threshold) \cdot (abs(\omega) > threshold) \quad (2.20)$$

It is very common to apply thresholding only on the higher-resolution coefficient levels, keeping the lower-frequency content of the signal intact.

It is obvious that these methods are highly non-linear. The soft thresholding is the l^2 optimal non-linear function in the wavelet domain to apply if one requires the resulting function to be at least as smooth as the original, noise-free one. Hard thresholding yields better l^2 performance but does not guarantee the smoothness property. In practice, hard thresholding may produce estimates containing “blips” and other irregularities that may not be visually appealing.

Another method of thresholding, proposed by Lang et al. [23], uses a shift invariant DWT for the noise reduction. Authors have shown that this method, when combined with hard thresholding, gives better noise suppression without oversmoothing of the details. This method calculates the DWT for all the shifts of the signal, followed by the usual hard thresholding and inverse DWT. Then, it averages the resulting estimates. When properly implemented, shift invariant DWT requires $O(N \log N)$ operations and storage space for $N \log N$ coefficients compared to $O(N)$ operations and storage space for N coefficients required for classical DWT.

2.5.1.2 Threshold Selection

Donoho and Johnstone have established [13, 10] several methods of threshold selection.

Universal Threshold or *VisuShrink* is given by

$$t_{univ} = \hat{\sigma} \sqrt{2 \log(n)} \quad (2.21)$$

where n is the sample size (2^N) and $\hat{\sigma}$ is the noise-level estimate. The authors have proposed *median absolute deviation* (MAD) method for the noise level estimation [13] from the finest scale of empirical wavelet coefficients (ω_f)

$$\hat{\sigma} = \text{Median}(|\omega_f|) / 0.6745 \quad (2.22)$$

This is a robust estimator which produces visually appealing (hence the name *VisuShrink*), very smooth, estimates. The same threshold is applied to all the levels. It has been shown that this method tends to underfit (oversmooth) the data [31].

SureShrink is an adaptive threshold-selection procedure [13]. A different threshold is applied at each resolution level leading to smoothness-adaptive estimates. It uses Stein's Unbiased Risk

Estimate for threshold choice. An implementation of *SURE* algorithm is given in the Matlab file: *ValSUREThresh.m* which is part of the wavelet toolbox developed at Stanford University [1]. At each level of resolution the algorithm checks whether there is enough signal to compute the *SURE* threshold. If there is, t_{sure} is used, otherwise t_{univ} is used. This method also tends to underfit.

There are some other methods, proposed by different authors. Nason [31] has suggested two methods. One uses *SURE* procedure on all the coefficients simultaneously. It is called *GlobalSure*. This procedure results in reasonable performance with respect to mean square error. The second method is based on *Cross-Validation*. The same paper presents a comparative study of these methods.

For all the threshold-selection procedures described above, the noise model is assumed to be as in (2.16). In case of correlated noise Johnstone and Silverman [21] have proposed a level-adaptive threshold:

$$t_j = s_j \sqrt{2 \cdot \log(n)} \quad (2.23)$$

where n is the data sample size and s_j is the standard deviation of the wavelet coefficients at j^{th} level. Another method that can be used in the case of correlated noise, if the noise structure is known, is to use a prewhitening transformation followed by universal thresholding. However, the wavelet decomposition of the signal in the original domain may possess sparsity properties that are lost in the prewhitening transformation, and the advantages of using wavelet de-noising on the prewhitened data would then be diminished [21].

2.5.1.3 De-Noising of Two-Dimensional Signals

The same methods described in the preceding sections can be used for two-dimensional signals. Donoho [10] proposed the same three step procedure described in 2.5.1, but applied to two-dimensional image data

$$\begin{aligned} y_{i_1, i_2} &= f_{i_1, i_2} + \sigma \varepsilon_{i_1, i_2} \\ i_1 &= 1, \dots, m_1 \\ i_2 &= 1, \dots, m_2 \end{aligned} \quad (2.24)$$

where $\varepsilon_{i_1, i_2} \stackrel{iid}{\sim} N(0, 1)$ is Gaussian white noise and the sample size n is given by $n = m_1 m_2$.

2.5.2 Compression

Data compression is one of the most common applications of wavelet theory. The ability of the DWT to compress the signal energy in a small number of coefficients makes it a natural choice for data compression. There are two basic types of data compression, lossless and lossy.

Lossless compression is used when an exact reconstruction of the signal is needed. In contrast, lossy compression allows small errors in the reconstruction, in an attempt to achieve higher compression ratios.

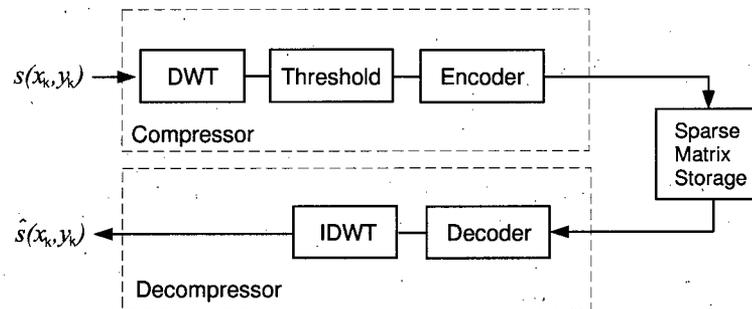


Figure 2.6: Block diagram of a compressor/decompressor without a quantizer.

In this thesis, only lossy compression will be considered. The reason for this should be apparent from the previous section (2.5.1). The goal is to try to reconstruct the signal from the noisy data. Most of the time it is only the reconstructed signal that is of interest and the “loss” should be the noise. The de-noising procedures reduce the number of nonzero elements, therefore making it an excellent candidate for compression.

There are two possible methods to implement data compression on the process data. One method is to use the procedure depicted in Figure 2.6. The *encoder's* role is to losslessly compress the sparse matrix of quantized coefficients. For the fast execution, simple *run-length* (Pratt [33]) coding of zero-valued coefficients has been proven very effective.

This simple procedure, when applied to the industrial data used later in this thesis, reduces the percentage of nonzero coefficients in the DWT to 5% (i.e. the signal is represented by 5% of its DWT coefficients). The remaining coefficients can be stored in a MatlabTM sparse matrix achieving the compression ratio of 13:1. Because the wavelet transform and de-noising procedure are an integral part of the estimator, no additional computing time is required.

Another method used in image compression is shown in Figure 2.7.

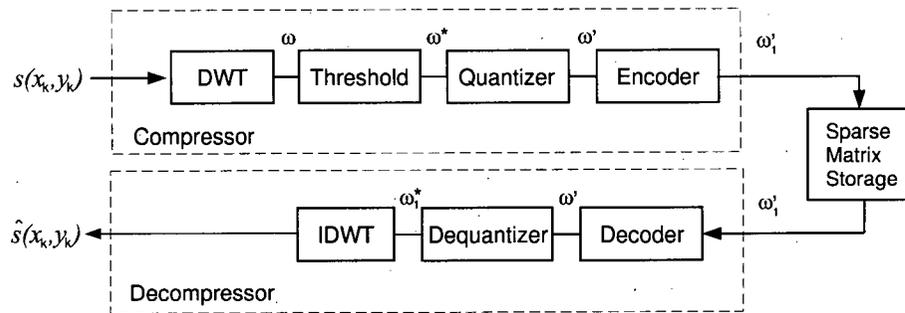


Figure 2.7: Block diagram of a compressor/decompressor with a quantizer.

This method adds one additional step, *quantization*, into the compression procedure. By quantization of nonzero DWT coefficients much higher compression ratios are achievable at the expense of an additional error. A *quantizer* is a function that maps many input values into a smaller set of output values. This mapping is generally a staircase function (Figure 2.8). It takes a continuous variable u and maps it into a discrete variable u^* which takes values from a finite set of *reconstruction levels* $\{r_1, r_2, \dots, r_L\}$ (Jain [19]).

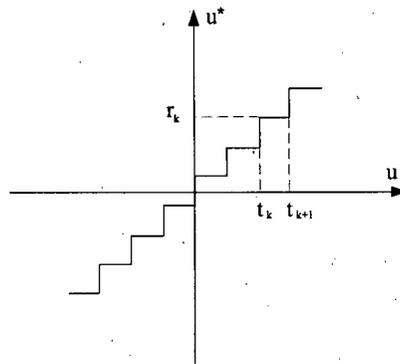


Figure 2.8: Quantizer mapping function.

If the value of u is between two *decision levels* $(t_k, t_{k+1}]$ the quantizer will map it into a single reconstruction level r_k (Figure 2.8). The quantizer mapping is irreversible — for the given reconstruction level (a quantizer output) the input value cannot be uniquely determined. Thus, an additional distortion of the original signal is introduced. Therefore, the quantizer should be designed in such a way that it minimizes the distortion. In the case of wavelet transform, where the significant information about the signal is carried by the high magnitude wavelet coefficients, it is obvious that the quantizer should have more reconstruction levels dedicated for those coefficients than for the low

magnitude ones. One design method, the *optimum mean square quantizer (LLOYD-MAX)* design, is given in [19].

The decompressor (Figure 2.7) applies decoding (unique) and dequantization (non-unique) to come up with a set of wavelet coefficients ω_1^* (note: $\omega_1^* \neq \omega^*$).

The compression of process data differs from the compression in image processing. The image processing compression should be optimized in a sense of accounting for the *Human Visual System* (Marr [27]) while the goal of process data compression is to preserve the information about the process.

For additional information on wavelet data compression one can refer to the overview given by Hilton et al. [18].

Chapter 3

Analysis of Paper Machine Process Data

The first part of this chapter will be used to introduce the application of wavelet estimation methods to paper machine data analysis. A Matlab™ toolbox has been developed so that different algorithms, described in the previous chapter, can be tested. Test data sets have been used to verify the algorithms and to compare their results to those obtained by using the industrially accepted estimators (EXPO, EIBMC).

3.1 Matlab™ Toolbox for Wavelet Analysis of Paper Machine Data

A set of Matlab™ functions and scripts has been developed to provide quick access to variety of different wavelet de-noising techniques and different test data sets. Even though this software is based on the Rice University Wavelet Toolbox [16] and WaveLab [1] only three functions from those toolboxes are used in the final version of the software (two-dimensional DWT and IDWT from the Rice Wavelet Toolbox and SURE threshold selection from the WaveLab). Most of the functions have been rewritten to adjust the algorithms to handle paper machine data (two-dimensional data with x and y dimensions of different, non-power of two, lengths). By selecting different options in the front-end interface procedures many different setups can be tried .

Two different thresholding methods can be used:

1. hard thresholding,
2. soft thresholding.

Four different methods of noise estimation and threshold selection are available:

1. standard deviation of the wavelet coefficients at the finest resolution multiplied by $\sqrt{2 * \log(n)}$,
2. VisuShrink procedure (see Chapter 2)
3. MultiMAD, resolution-dependent thresholding using MAD noise estimator (Eq. 2.22) to estimate the noise standard deviation at each resolution level,
4. SureShrink (see Chapter 2)

and filters created from four different mother wavelets can be used:

1. Haar wavelet,
2. Daubeshies orthogonal wavelets,
3. Coiflets,
4. Symmlets.

The above wavelet filters are depicted in Figure 3.1. For more information about the mother wavelets used the reader may refer to the original papers [8, 5, 9].

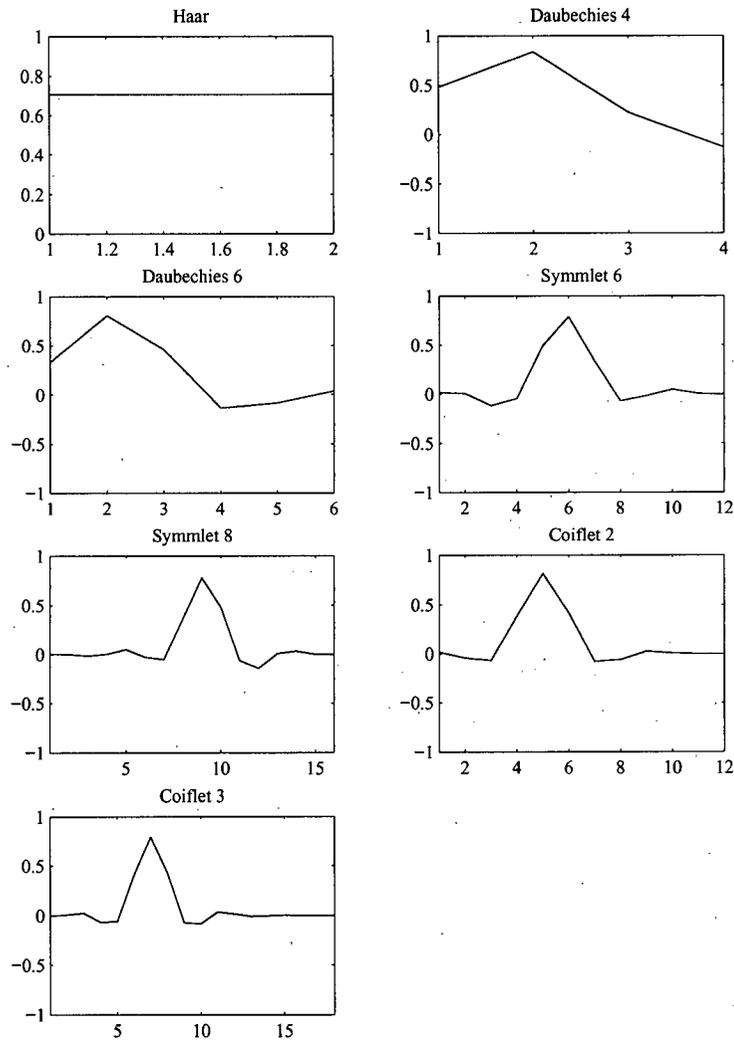


Figure 3.1: Wavelet filters used in the analysis.

The results of wavelet de-noising procedures are compared to the estimations obtained by the EXPO algorithm and, in some cases, to those obtained by the EIBMC algorithm. Only a few tests

were done using the EIBMC since, for a fair comparison, the EIBMC algorithm needs to be properly tuned and the test data sets have to be created by using the proposed models (see Appendix B). The verification process called for some extreme-case input signals. Hence, not all data sets used for the wavelet estimator testing were created by the given models. Note that both EIBMC and EXPO are recursive algorithms, whereas wavelet filtering is a batch operation at present.

3.1.1 Padding of the Two-Dimensional Signals

As mentioned in the previous chapter the fast DWT and IDWT procedures need input data sets of dimensions $D(2^{L_1}, 2^{L_2})$. Paper machines rarely give this kind of output. Therefore, some kind of extrapolation is needed so data can be padded up to the proper size. The padding by the repetition of the last data point in each particular direction has been chosen because it is simple to implement and it reduces the boundary errors due to the periodic nature of the input signals assumed by the DWT algorithm. The extra data points are removed after the de-noising. This procedure caused many of the boundary wavelet coefficients to be set to zero (signal becomes constant in that direction setting high-pass signal component to zero). Thus, the original MAD algorithm (Eq. 2.22) had to be modified so that all the zero coefficients are removed before the median is calculated.

3.2 Applying Wavelet De-noising to Simulated Scanned Data

Verification of the proposed algorithms has been conducted on the various sets of the simulated data. Data sets were designed in such a way that they provided signals with enough complexity so that the estimator could be properly evaluated. The test data also contained the features that commonly appear on a paper sheet (streaks, bumps and grade changes). All the data sets have been zigzag sampled and presented to the estimator as if they had been coming from a paper machine scanner.

The data processing has been done in batch mode. A matrix of data points of the size $D(n, m)$ is taken, padded to the size $D_1(2^N, 2^M)$, where $2^N \geq n$ and $2^M \geq m$, and sent to the estimator.

3.2.1 Edge Effects and Step Changes

The first data set has been created in such a way that the edge effects (errors from the signal periodicity assumed by DWT) and the effects of sharp (sudden) changes in profiles can be studied. The target values (noiseless data) are shown in Figure 3.3. This data set contains a very well

defined CD profile which differs greatly at the edge sensors (sensors #1 and #120) and it has a sharp discontinuity at the sensor #60. The MD profile is a sine function of time, with an amplitude of 1. At the MD position 29 the CD profile flattens up by 50% (sudden change in CD profile — the amplitude being changed from 1 to 0.5). The number of scans for all test data sets is 58 and the number of sensors is 120. Consequently, four data points were padded to each end of every CD profile and 3 scans were padded before and after the real signal.

This signal is corrupted with white noise $z_{i_1, i_2} \stackrel{iid}{\sim} N(0, \sigma^2)$ where $\sigma^2 = (0.2)^2$. The composite signal (CD + MD), before and after adding the noise, is shown in Figure 3.2.

This noisy signal has been processed by using different wavelet estimators and the results compared to those obtained by the EXPO. For a signal containing such abrupt changes it was expected that the Haar wavelet would perform very well. This has been confirmed by the experiments (see Tables 3.2 and 3.3). The estimated profiles as well as the real signals are given in Figure 3.3. The MD profiles are calculated as scan averages.

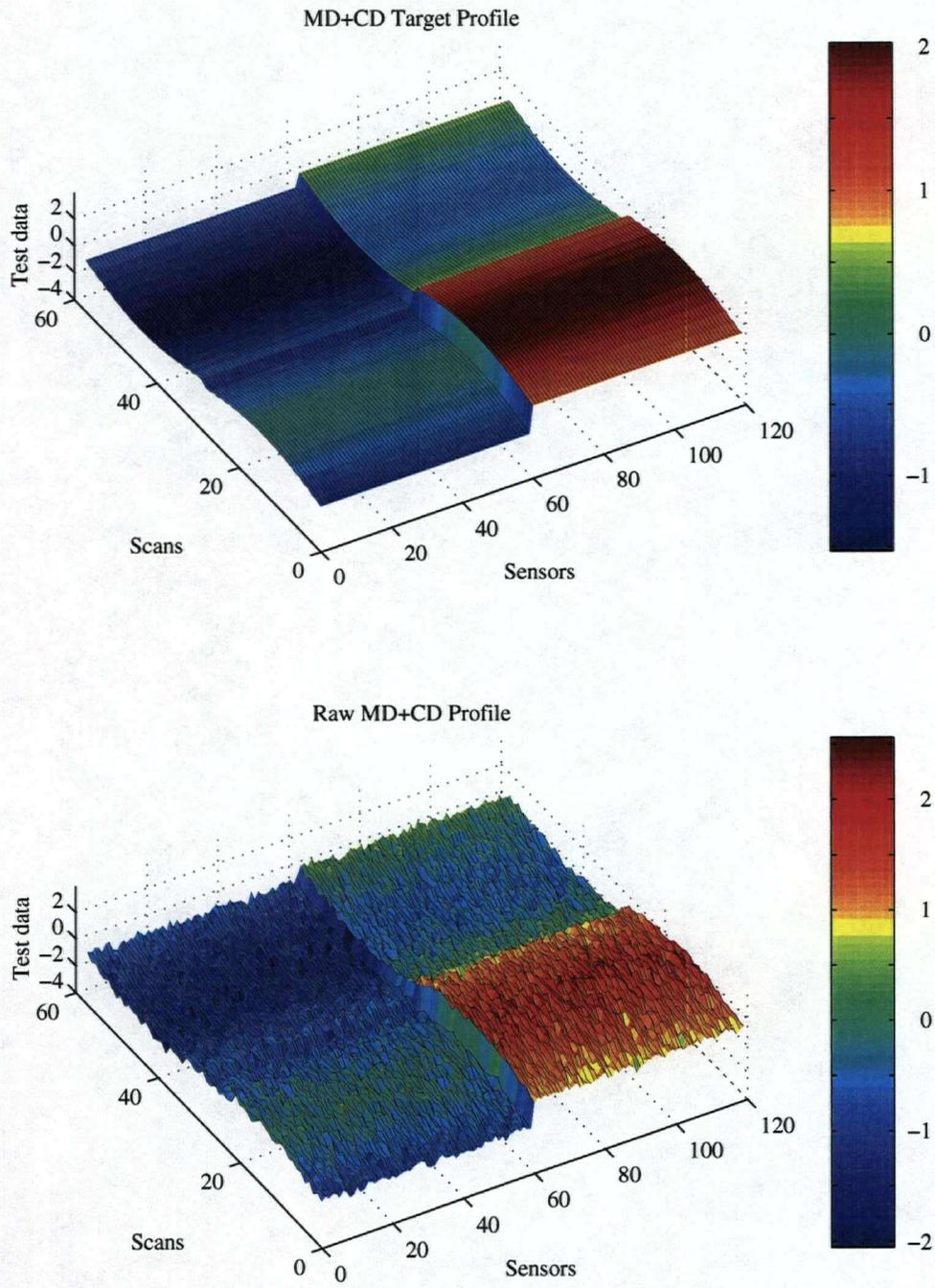


Figure 3.2: Edge effects: CD+MD before and after the noise has been added.

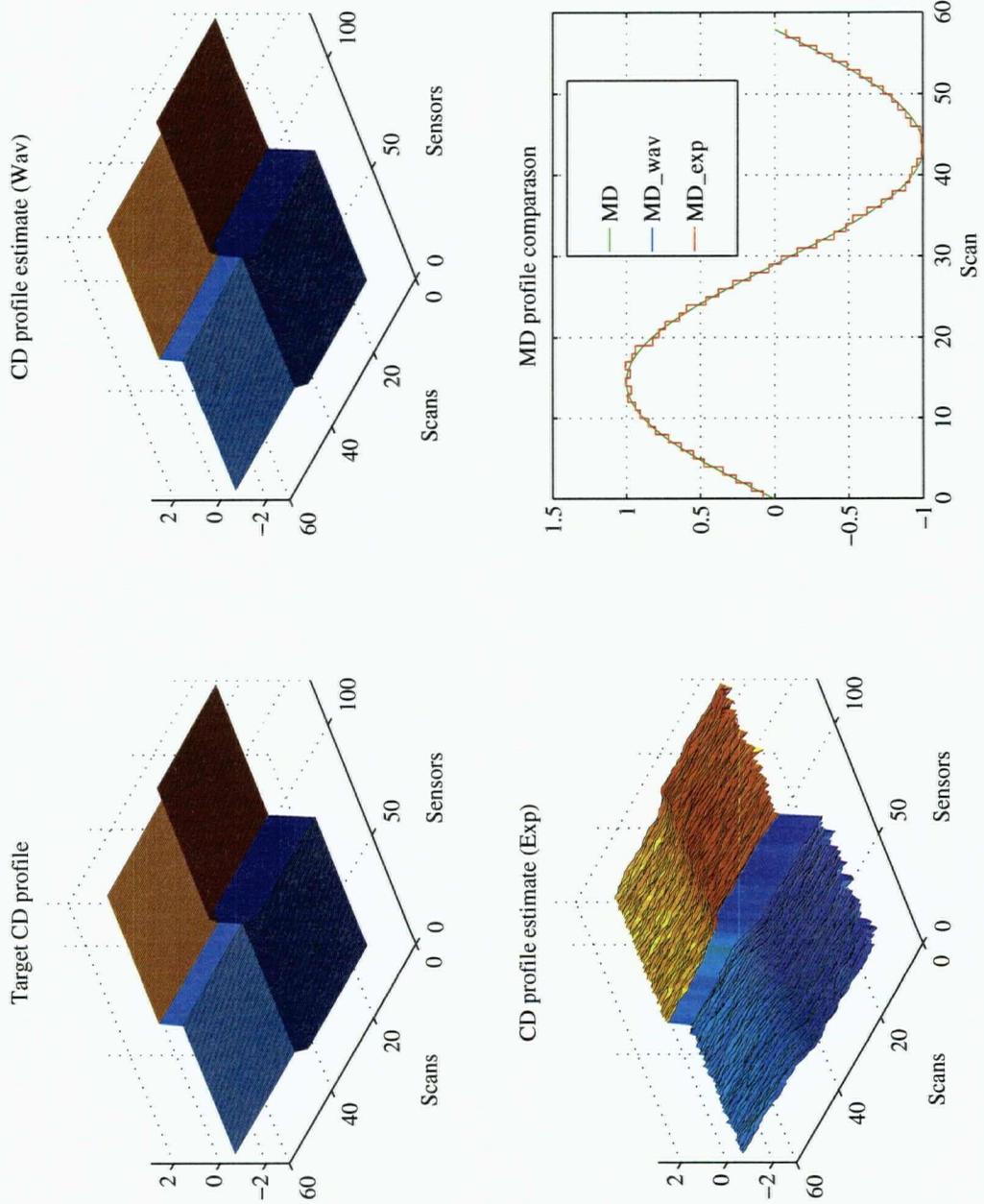


Figure 3.3: Edge effects: estimated profiles.

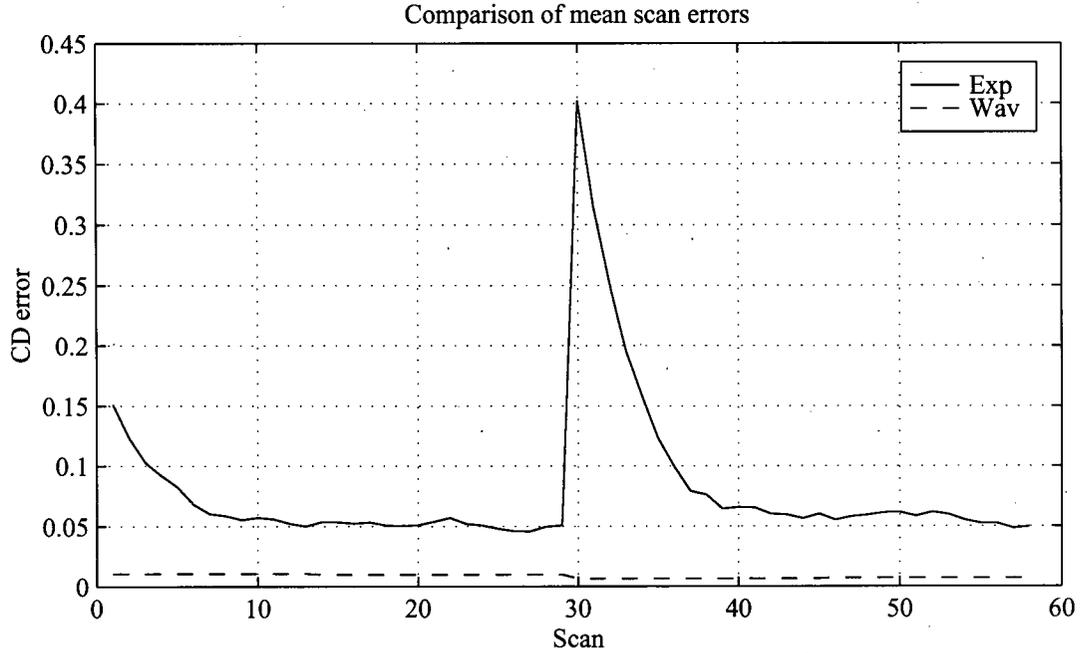


Figure 3.4: Edge effects: estimation errors as functions of time (scan number).

The Haar wavelet given in Table 3.3 has been used. The thresholds and the estimated noise standard deviations, used for de-noising, are given in Table 3.1. Figure 3.4 depicts the mean

Levels	Threshold	$\hat{\sigma}(\sigma = 0.2)$
1	0.75	0.19
2	0.64	0.18
3	0.63	0.20
4	0.50	0.19
5	0.25	0.12

Table 3.1 Edge effects: thresholds and estimated noise at each resolution level.

estimation error (per scan) as a function of time (scan position). The error calculation formula is given in Equation 3.1

$$CDerr_{i,j} = |CD_{i,j} - CDeestimated_{i,j}|, \quad i = 1 \dots n; j = 1 \dots m \quad (3.1)$$

where $CDerr_{i,j}$ is the absolute value of estimation error at the i^{th} scan and j^{th} sensor. For each scan the mean value has been plotted.

$$\overline{CDerr}_i = \frac{1}{m} \sum_{j=1}^m CDerr_{i,j} \quad (3.2)$$

$$J = \sqrt{\sum_j \sum_i CD^2 err_{i,j}} \quad (3.3)$$

The overall errors (Equation 3.3) were:

$$\begin{aligned} J_{wav} &= 0.86, \\ J_{expo} &= 9.6 \end{aligned} \quad (3.4)$$

and the mean error at each scan was less than 1% of the signal amplitude. Therefore, the signal was successfully recovered from the noisy data.

A number of tests using different methods of thresholding, threshold selections, and with a variety of different mother wavelets of different lengths, has been run for each data set. Some of the results are summarized in the Table 3.2. For the Symmlets and Coiflets the number in the fourth column represents the wavelet type not the actual length.

Two tables, summarizing the test results, are given for each data set. The first table always shows the results obtained by the same set of the wavelet filters and is used for comparison of the filter performances on different data sets. The second table displays the representatives of four different wavelet filters that have produced the best results, i.e. smallest errors, for that data set.

The Haar wavelet used in this example has produced the best results, as expected, but a number of different configurations have performed very well. When very smooth wavelet filters have been used, the estimation error would tend to go up with an increased number of resolution levels. This happens when thresholding starts to affect the low-frequency part of the signal (trying to smooth out the signal made of sharp edges).

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{expo}}$
Daubechies	soft	std	6	3	0.85
Daubechies	soft	MultiMAD	6	3	1.14
Symmlet	soft	std	6	3	0.85
Symmlet	soft	MultiMAD	6	3	1.19
Coiflet	soft	std	3	3	0.83
Coiflet	soft	MultiMAD	3	3	1.27
Haar	soft	MultiMAD	2	2	0.46

Table 3.2 Edge effects: summary of the results

It can be seen that the errors at the edges of the signal are very small. This may be attributed to the signal padding and to the wavelet length. For a filter length of two, only one data point has to be padded to the signal at any time — consequently only one point can be wrong.

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{expo}}$
Daubechies	hard	MAD	4	2	0.70
Symmlet	hard	MAD	6	3	0.72
Coiflet	hard	std	2	3	0.66
Haar	soft	MultiMAD	2	5	0.09

Table 3.3 Edge effects: best filter performances.

To contrast the results shown in Figures 3.3 and 3.4, in Figures 3.6 and 3.5 are shown the results of the estimation obtained by using the Coiflet filter described in Table 3.3. As expected, the filter with a higher degree of smoothness yielded much smoother estimates causing a decrease of 'sharpness' at the points of discontinuity (sensor #60 and scan #29). The overall estimation error was higher than in the first case but still significantly better than for the EXPO estimator. The filter length has also contributed to the edge effect errors.

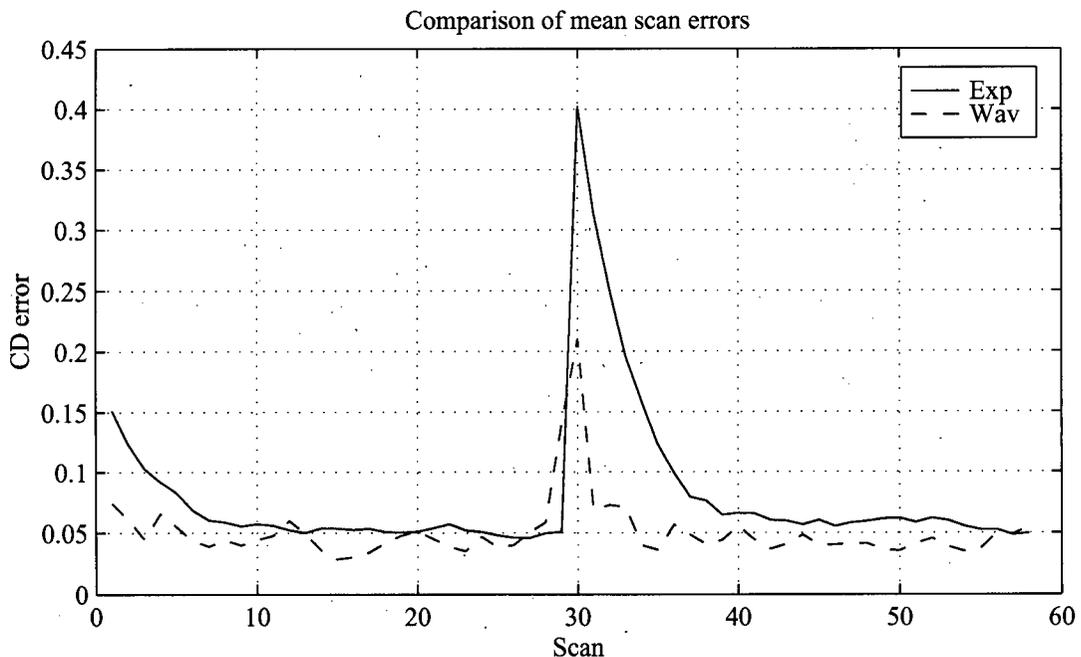


Figure 3.5: Edge effects: estimation errors as functions of time (scan number) for a smooth estimator.

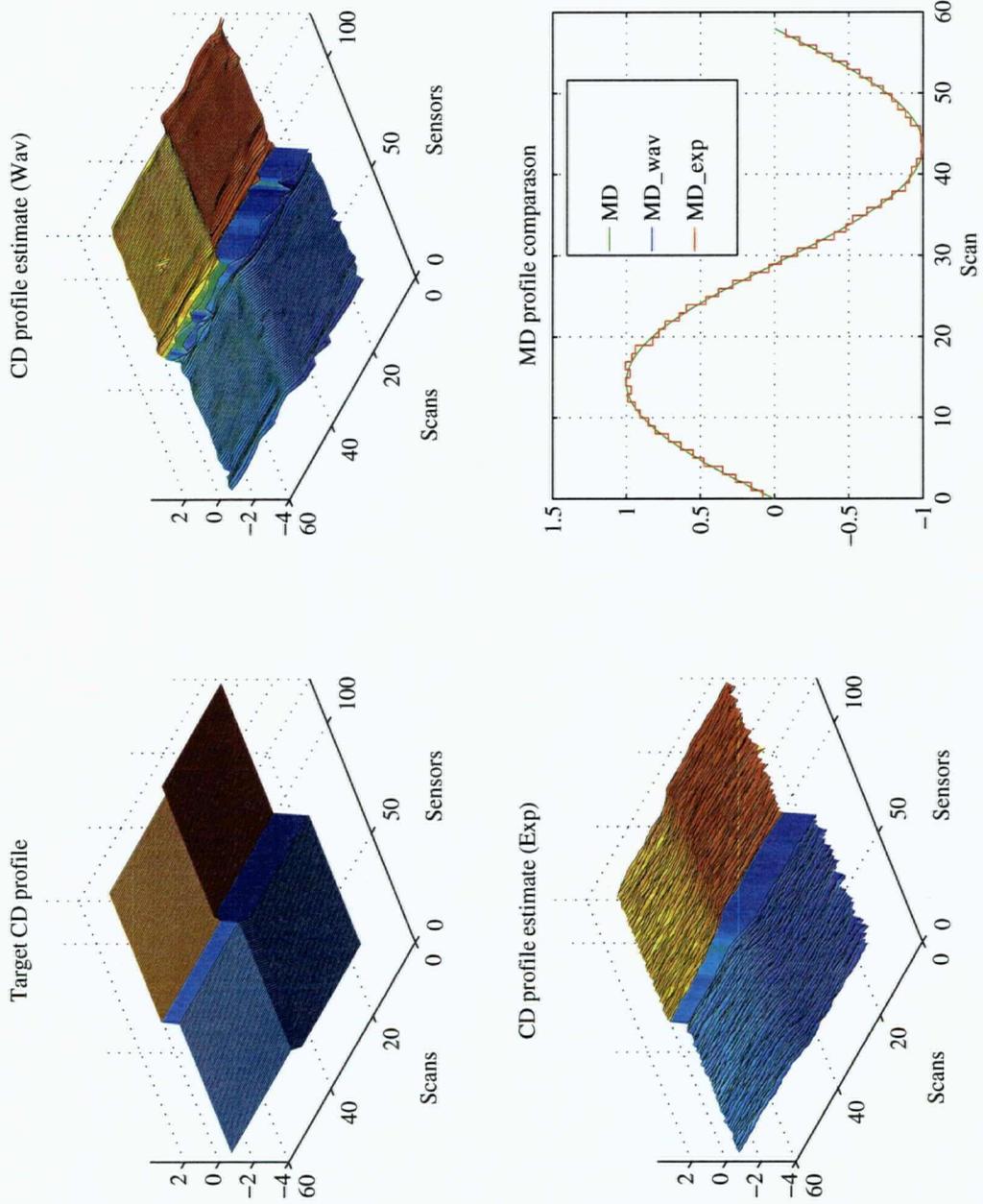


Figure 3.6: Edge effects: estimation results for a smooth estimator.

3.2.2 Smooth Surfaces

The second data set, depicted in Figure 3.7, has been created with the intention of testing the wavelet estimator on some smoother surfaces with a constant signal to noise ratio. The profiles

$$\begin{aligned} CD &= 1 \cdot \sin \theta_{CD}, & \theta_{CD} &\in [0, 6\pi] \\ MD &= 1 \cdot \sin \theta_{MD}, & \theta_{MD} &\in [0, 2\pi] \end{aligned} \quad (3.5)$$

were corrupted by the additive white noise $z_{i_1, i_2} \stackrel{iid}{\sim} N(0, \sigma^2)$ where $\sigma^2 = (0.5)^2$ which was much higher than in the first data set. The estimation has been done by the Symmlet filter using MultiMAD procedure (see Table 3.5). The thresholds and the estimated noise are given in Table 3.4. The estimation errors are shown in Figure 3.9 while the estimated profiles are given in Figure 3.8.

Levels	Threshold	$\hat{\sigma}(\sigma = 0.5)$
1	1.56	0.40
2	1.44	0.40
3	1.87	0.60

Table 3.4 Smooth surfaces: thresholds and estimated noise at each resolution level.

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{expo}}$
Daubechies	soft	std	6	3	0.51
Daubechies	soft	MultiMAD	6	3	0.51
Symmlet	soft	std	6	3	0.39
Symmlet	soft	MultiMAD	6	3	0.39
Coiflet	soft	std	3	3	0.42
Coiflet	soft	MultiMAD	3	3	0.42
Haar	soft	MultiMAD	2	2	0.95

Table 3.5 Smooth surfaces: summary of the results

Even though the signal to noise ratio was low, the estimated profiles were very close to the original ones and, due to the filter smoothness, very smooth. The overall errors were:

$$\begin{aligned} J_{wav} &= 6.22, \\ J_{expo} &= 16.03 \end{aligned} \quad (3.6)$$

or

$$J_{wav} = 0.39J_{expo} \quad (3.7)$$

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{expo}}$
Daubechies	soft	std	6	3	0.51
Symmlet	soft	MultiMAD	6	3	0.39
Coiflet	soft	MultiMAD	3	3	0.42
Haar	soft	MultiMAD	2	2	0.95

Table 3.6 Smooth surfaces: best filter performances.

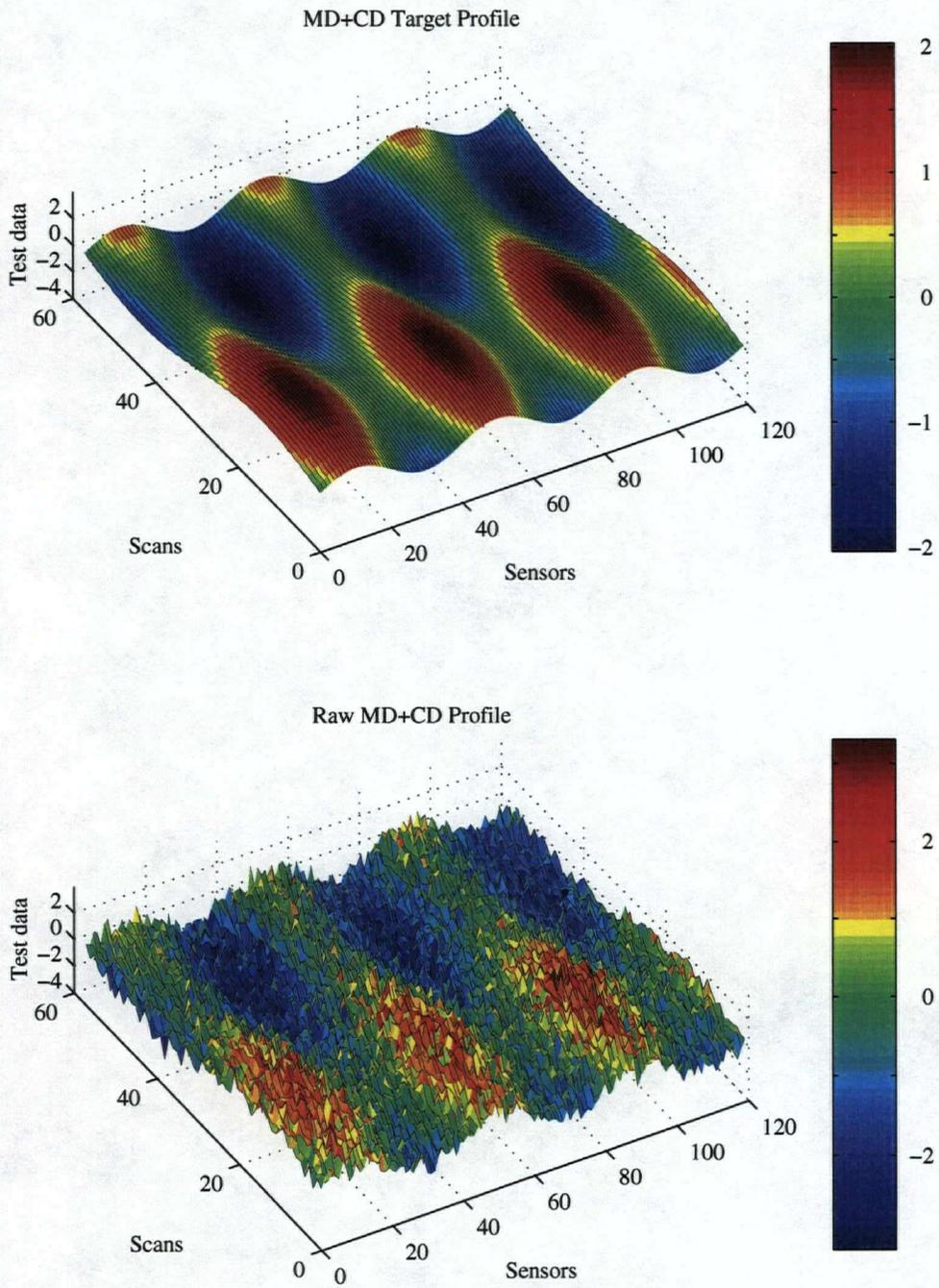


Figure 3.7: Smooth surfaces: composite profiles with and without noise.

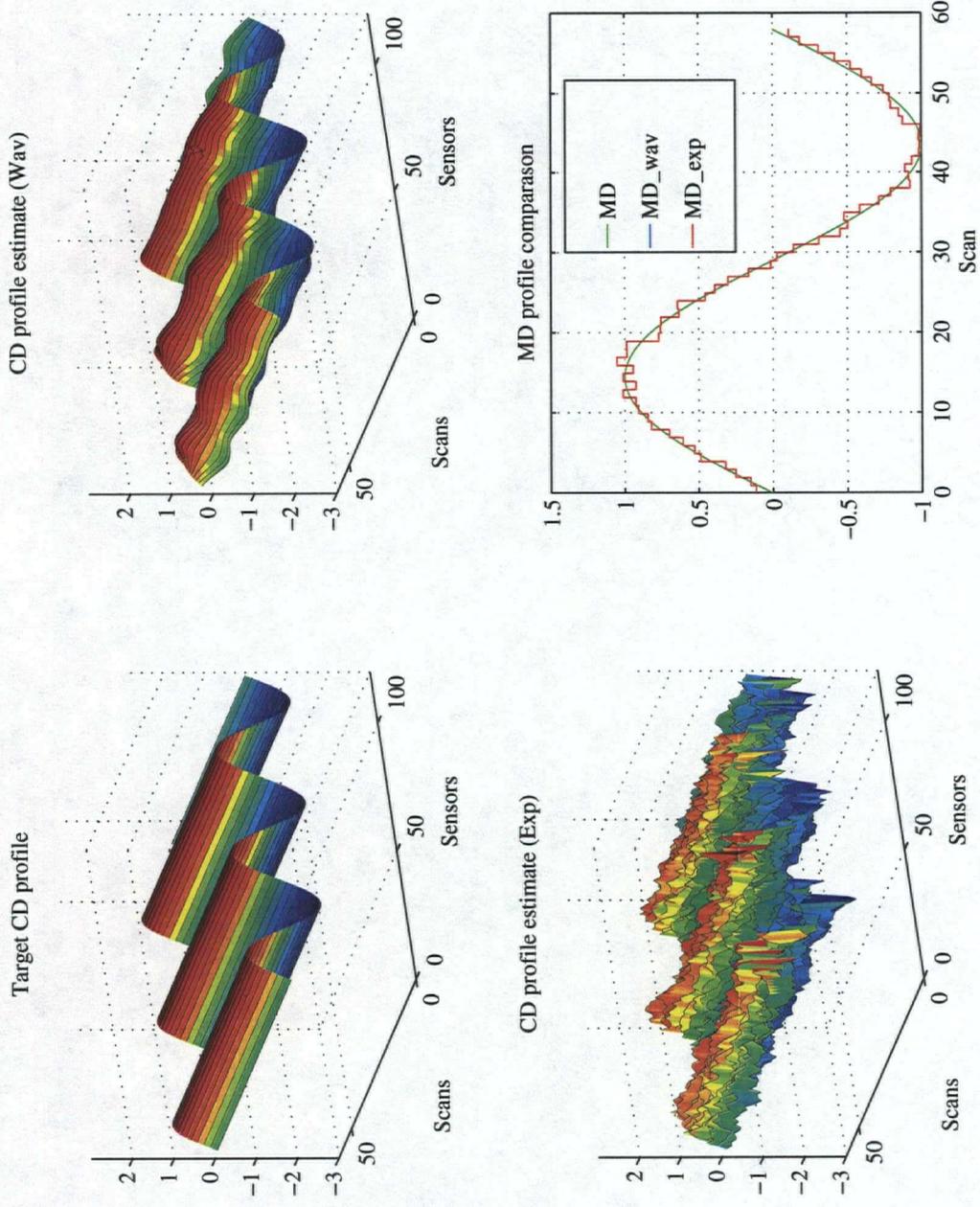


Figure 3.8: Smooth surfaces: estimated profiles.

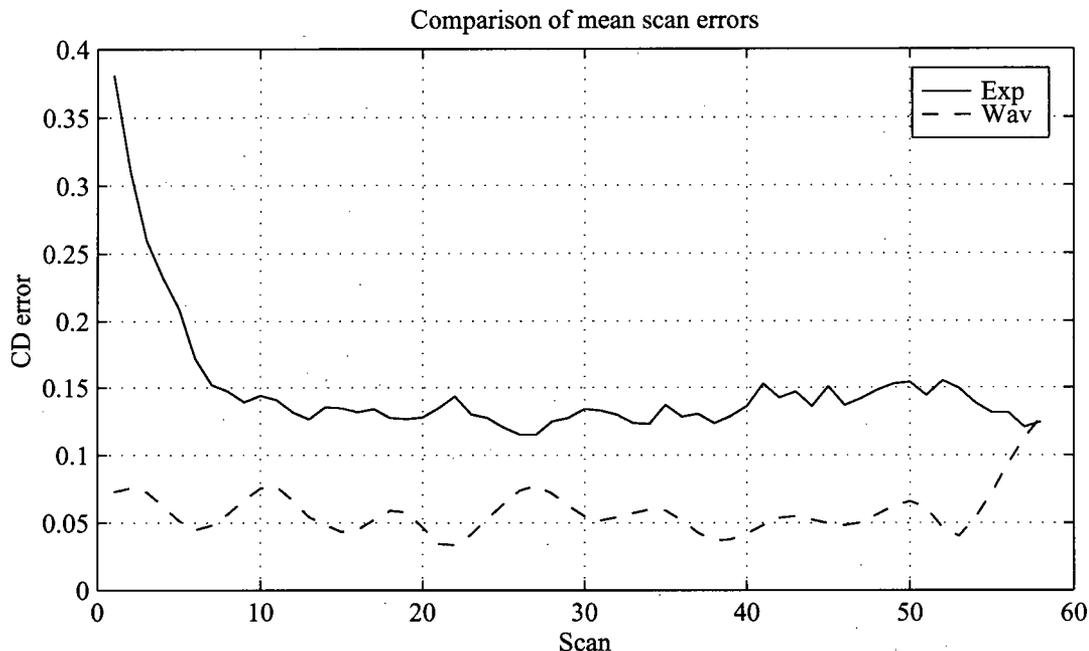


Figure 3.9: Smooth surfaces: estimation errors.

3.2.3 Grade Profile Change

This data set, shown in Figure 3.10, has been used with the wavelet estimator to see how the estimator behaves when a step change of amplitude 2 is introduced in the MD profile. The profiles

$$\begin{aligned}
 CD &= 0.3 \cdot \cos \theta_{CD}, & \theta_{CD} &\in [0, 2\pi] \\
 MD &= 0.3 \cdot \sin \theta_{MD}, & \theta_{MD} &\in [0, 2\pi]
 \end{aligned}
 \tag{3.8}$$

were corrupted by additive white noise with variance that was very high compared to the profile amplitudes: $\sigma^2 = (0.2)^2$. The estimation has been done using the Symmlet filter and MultiMAD procedure (see Table 3.9). The thresholds and the estimated noise are given in Table 3.7. It is interesting to note that the VisuShrink threshold that would apply to this case was 0.66. Table 3.7 shows that a threshold of this size was applied only at one resolution level. All the other threshold levels were lower. Therefore, a VisuShrink estimator would oversmooth the signal in this case. This agrees with the theory given in Chapter 2. The estimation errors are shown in Figure 3.13 while

Levels	Threshold	$\hat{\sigma}(\sigma = 0.2)$
1	0.61	0.16
2	0.60	0.17
3	0.66	0.21
4	0.49	0.19
5	0.38	0.19

Table 3.7 Grade change: thresholds and estimated noise at each resolution level.

the estimated profiles are given in Figures 3.11 and 3.12. The overall error for the EXPO filter was

$$J_{expo} = 6.42.$$

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{expo}}$
Daubechies	soft	std	6	3	0.49
Daubechies	soft	MultiMAD	6	3	0.39
Symmlet	soft	std	6	3	0.50
Symmlet	soft	MultiMAD	6	3	0.38
Coiflet	soft	std	3	3	0.51
Coiflet	soft	MultiMAD	3	3	0.38
Haar	soft	MultiMAD	2	2	0.71

Table 3.8 Grade change: summary of the results

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{expo}}$
Daubechies	soft	MultiMAD	6	4	0.20
Symmlet	soft	MultiMAD	8	5	0.17
Coiflet	soft	MultiMAD	2	4	0.19
Haar	soft	MultiMAD	2	3	0.50

Table 3.9 Grade change: best filter performances.

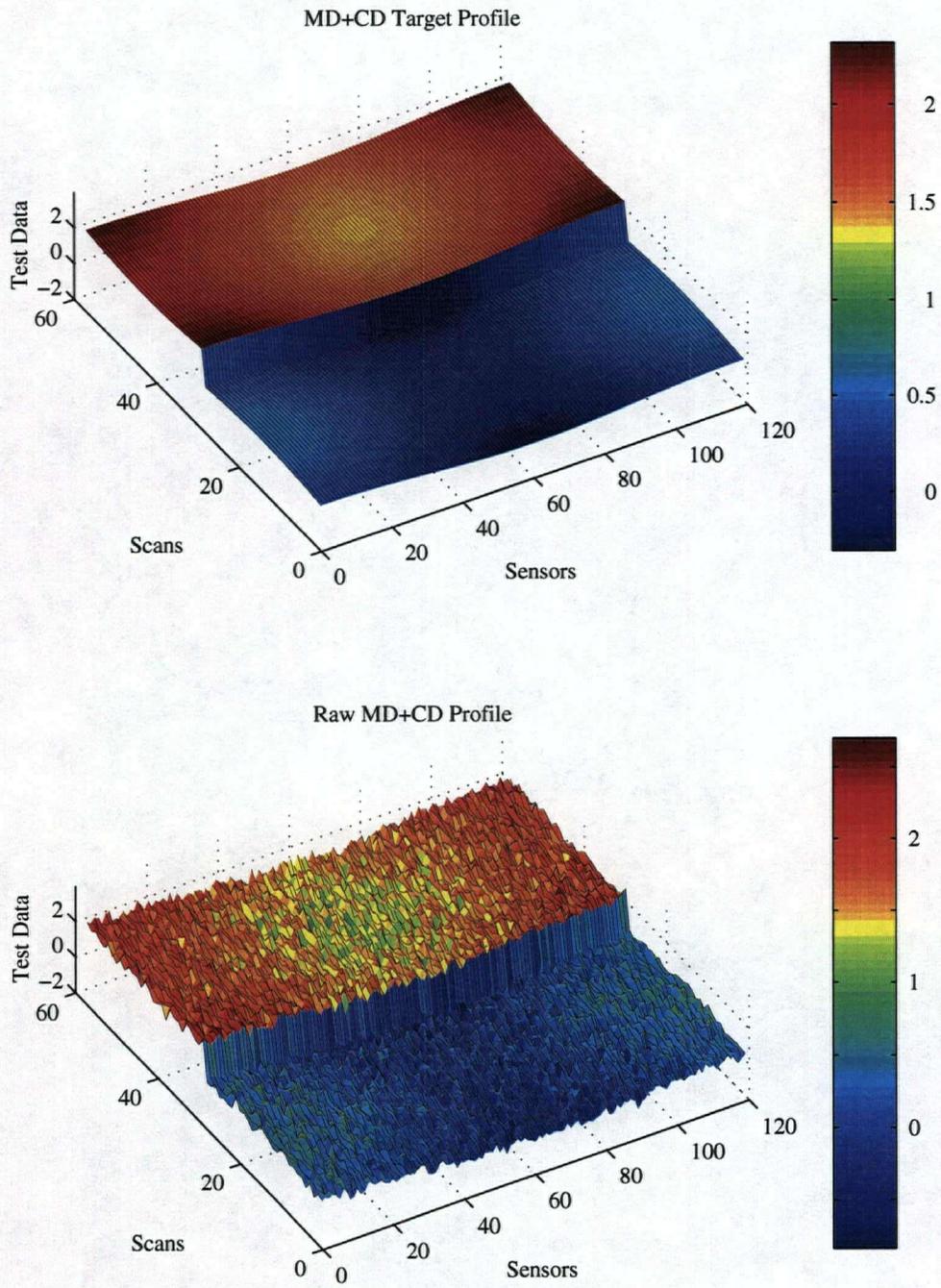


Figure 3.10: Grade change data set: CD+MD before and after the noise has been added.

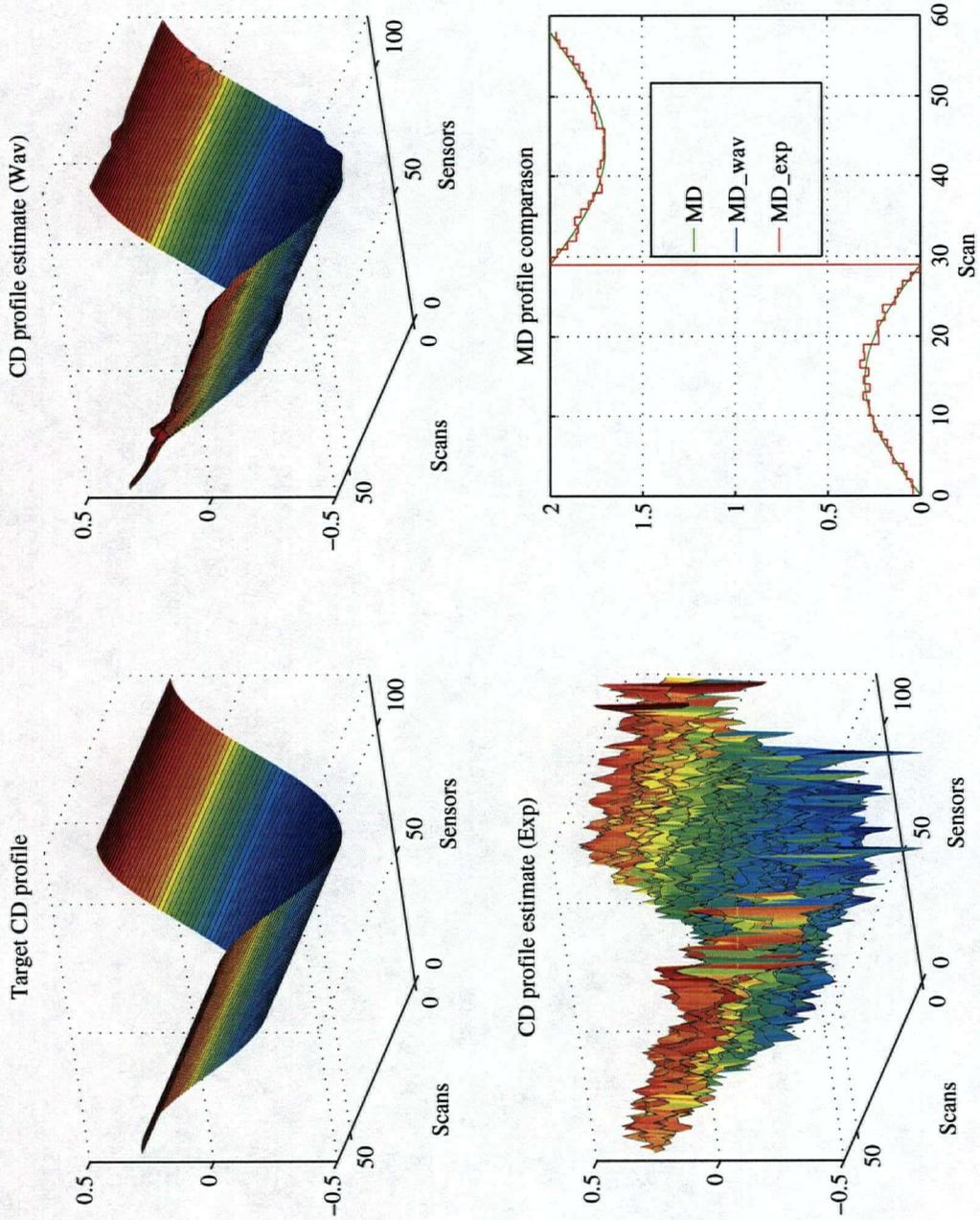


Figure 3.11: Grade change test data set: estimated profiles.

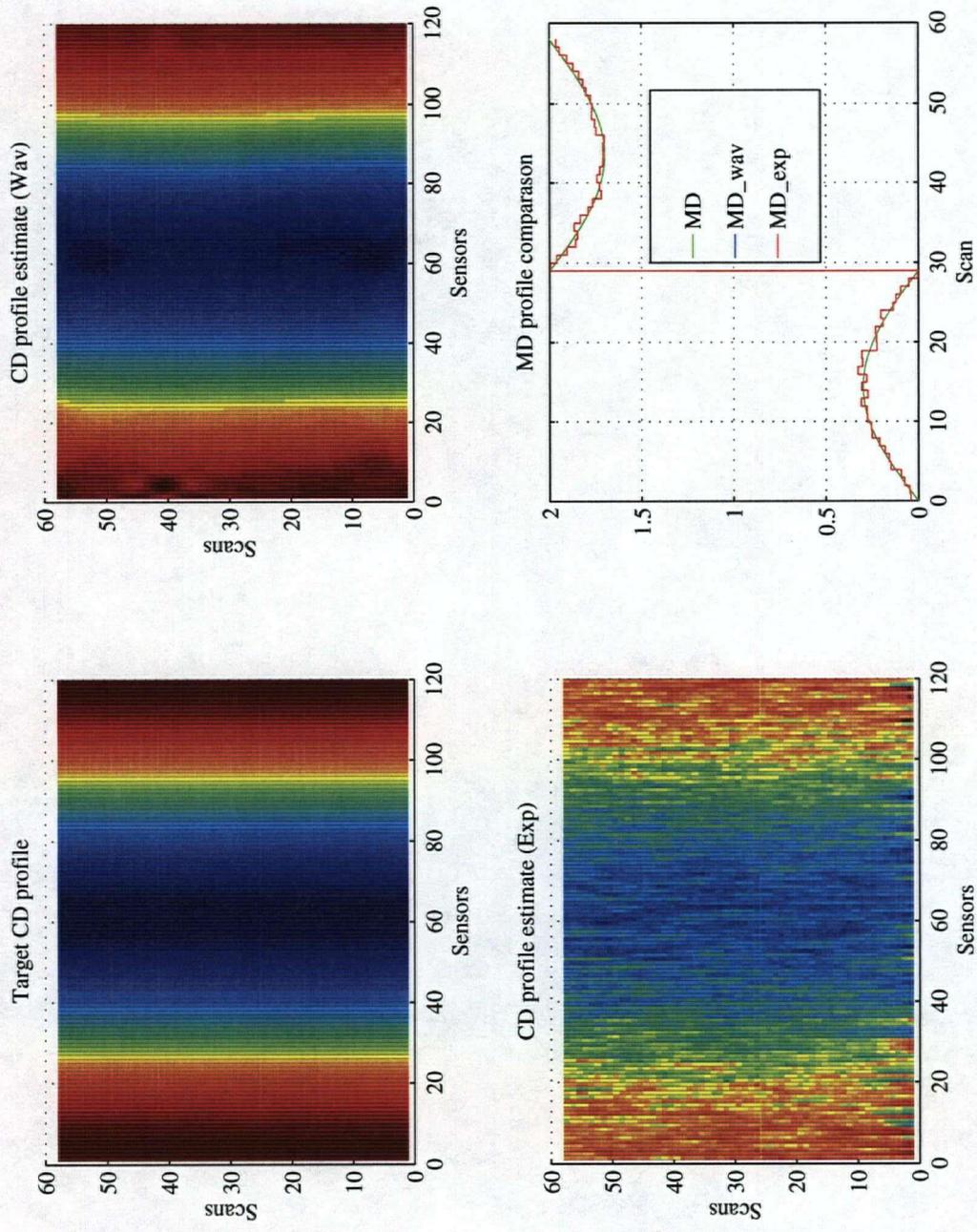


Figure 3.12: Grade change test data set: estimated profiles (images).

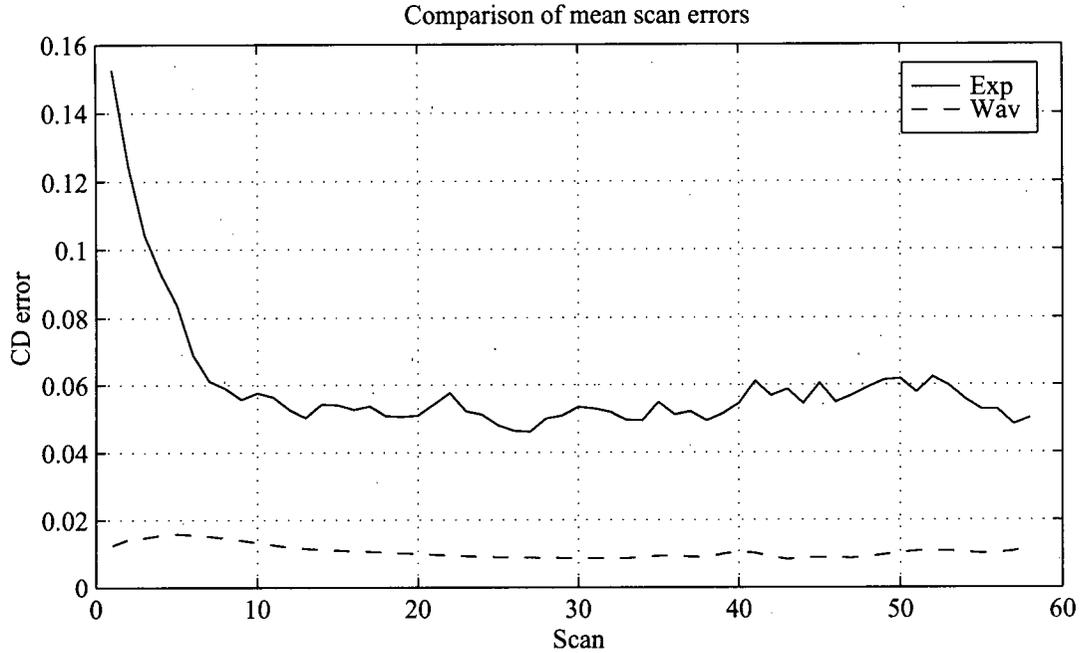


Figure 3.13: Grade change test data set: estimation errors.

3.2.4 Data Created by the Moisture Variation Model (MVM)

This data set (the profiles are shown in Figure 3.15) has been created by the MVM. The input parameters and the profiles are given by:

$$\begin{aligned}
 CD &= 1 \cdot \sin \theta_{CD} \exp(-0.03 * k) \\
 k &= 1 \dots 58, \quad (\text{scan}\#) \\
 \theta_{CD} &\in [0, 6\pi] \\
 MD &= \bar{u} + \frac{1}{1 - 0.8z^{-1}} \omega, \\
 \bar{u} &= 0.5, \\
 w &\overset{iid}{\sim} N(0, 0.1^2)
 \end{aligned} \tag{3.9}$$

The estimation has been done by the Coiflet filter using MultiMAD procedure (see Table 3.12). The thresholds and the estimated noise are given in Table 3.10. The VisuShrink threshold that would apply to this case was 0.27. Again, only one threshold (Table 3.10) used in this case was of that size. The threshold used on the lowest resolution level was much higher than the other two and was

the result of the estimator trying to filter out the coloured noise introduced in (3.9). The estimation

Levels	Threshold	$\hat{\sigma}$
1	0.20	0.05
2	0.31	0.09
3	0.65	0.21

Table 3.10 MVM test data: thresholds and estimated noise at each resolution level.

errors are shown in Figure 3.14 while the estimated profiles are given in Figure 3.15. The overall errors for the EXPO and the EIMBC estimators were

$$J_{expo} = 11.43 \quad (3.10)$$

$$J_{eibmc} = 0.94 \cdot J_{expo}$$

It can be seen, from the given results, that the wavelet estimator performed very well compared to the both EXPO and EIMBC algorithms regardless of the fact that a priori knowledge of the moisture model and coloured noise has not been used. A possible improvement would be to include the noise model in the threshold-estimation procedures by increasing or lowering the threshold, at different resolution levels, by prescheduled coefficients based on the noise model. The wavelet estimated profiles were much smoother and without the artificial CD shapes that can be seen on the EIMBC estimated results (see Figure 3.16). Therefore, if the estimated signals were to be used for control, the smooth profiles would prevent unnecessary CD control actions.

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{expo}}$
Daubechies	soft	std	6	3	0.78
Daubechies	soft	MultiMAD	6	3	0.69
Symmlet	soft	std	6	3	0.79
Symmlet	soft	MultiMAD	6	3	0.74
Coiflet	soft	std	3	3	0.76
Coiflet	soft	MultiMAD	3	3	0.67
Haar	soft	MultiMAD	2	2	0.89

Table 3.11 MVM test data: summary of the results

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{expo}}$
Daubechies	soft	std	6	5	0.66
Symmlet	soft	std	6	5	0.68
Coiflet	soft	MultiMAD	3	3	0.67
Haar	soft	MAD	2	5	0.76

Table 3.12 MVM test data: best filter performances.

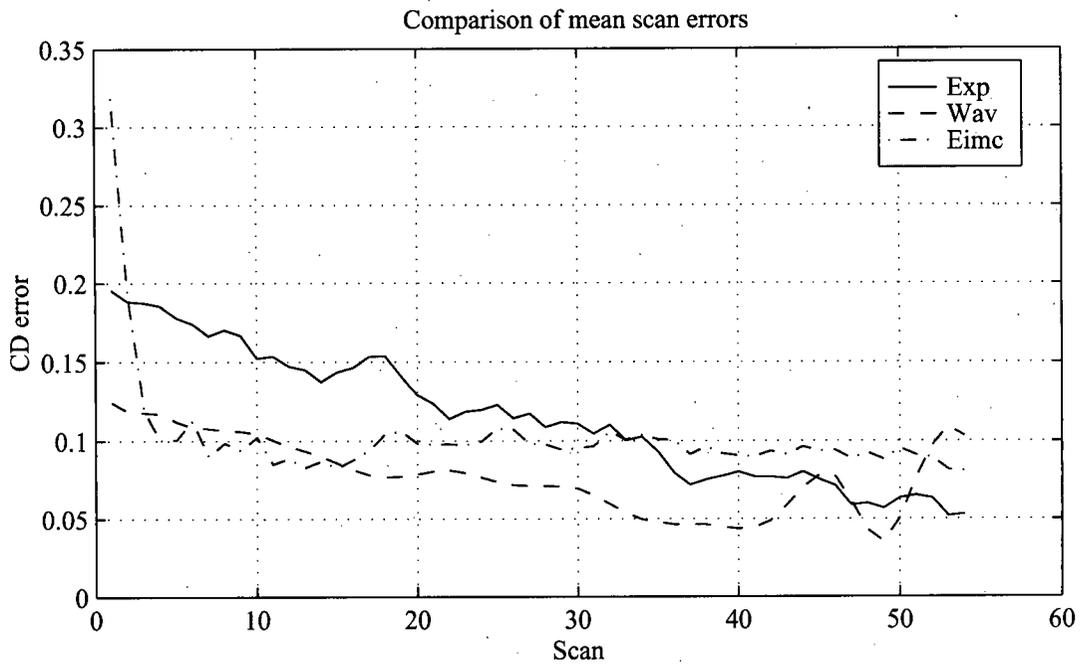


Figure 3.14: MVM test data: estimation errors.

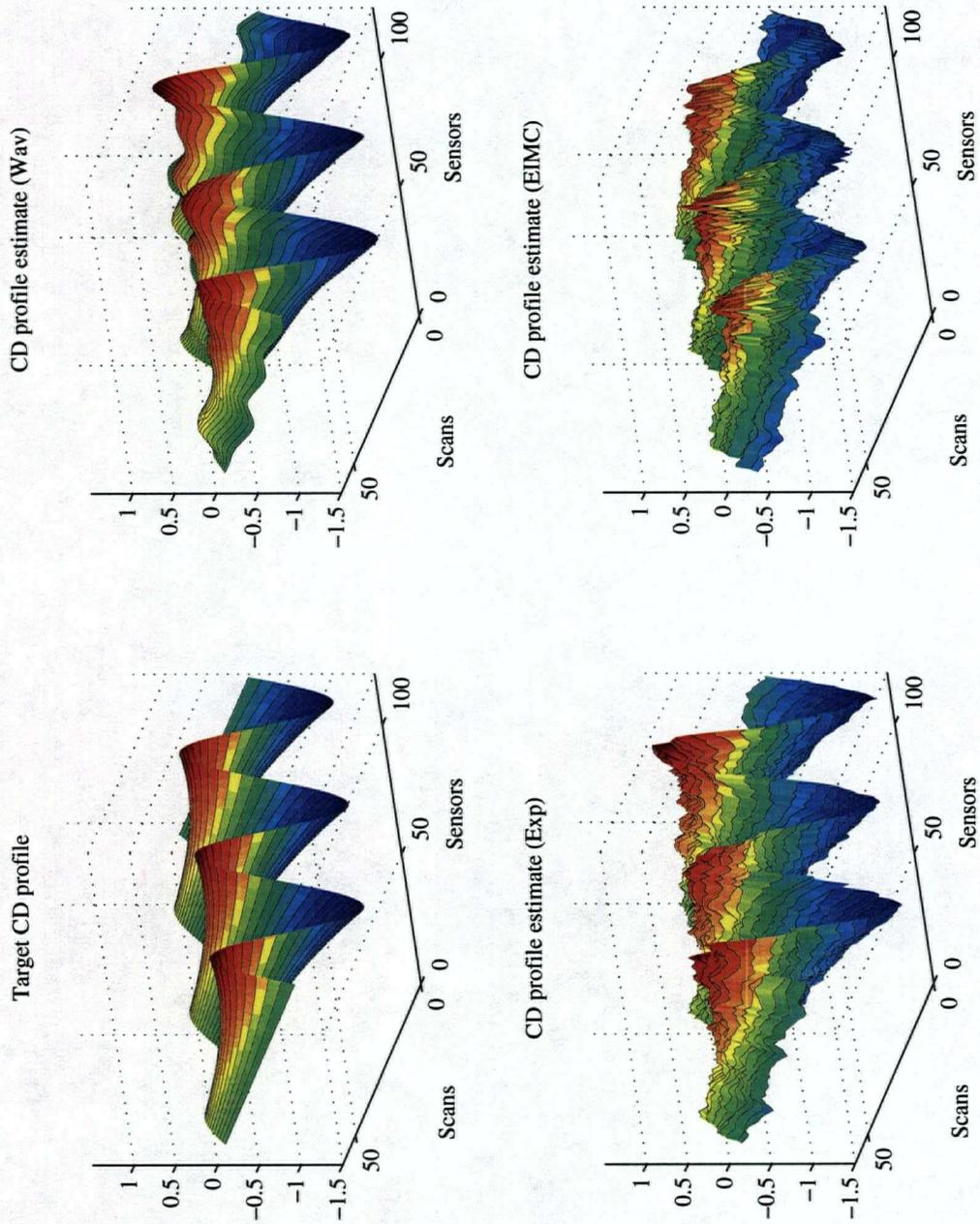


Figure 3.15: MVM test data: estimated profiles.

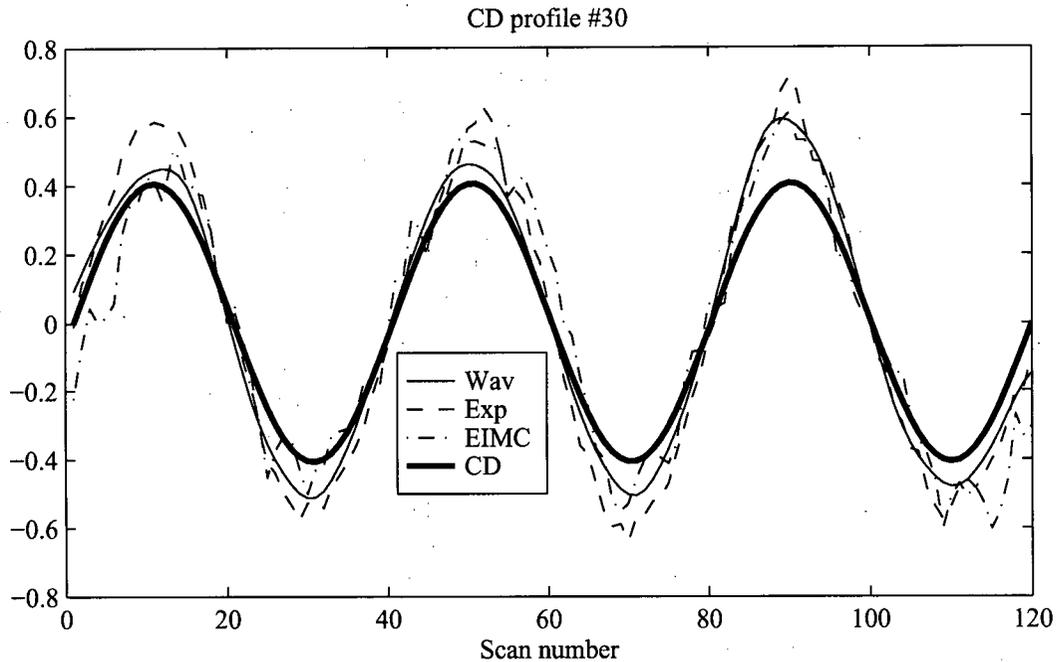


Figure 3.16: MVM test data: CD profiles.

3.2.5 Bumps and Streaks

The ability of the wavelet estimator to localize some characteristic patterns on a paper sheet has been tested using the data set given by Equation 3.11. This data set simulates a narrow streak in MD and two actuator bump changes.

The simulation results are given in Figures 3.17 and 3.18, and performances of different estimators and their respective errors of estimation are shown in Tables 3.13 and 3.14. Results shown in the Figures are obtained by the Coiflet estimator whose parameters are given in Table 3.14. Hard thresholding has been used with the threshold set at 0.754 and the estimated noise standard deviation was 0.18 (true value was 0.2). This signal was a very difficult test for the estimators because of its smoothness combined with some very short and/or narrow features. If the signal had contained flat profiles with the same bumps and the streak, then the Haar wavelet would have performed the best. In this case, as in the most of the other tests, the performance of the selected group of wavelet estimators is similar to the best performances for that data set. The wavelet estimator has successfully localized the test patterns of short duration. Estimation of the streak is not as effective as by the EXPO algorithm, although the contrast between smooth surfaces, the bumps, and the streak is impressive. EXPO is effective in detecting these variations since it carries on no CD smoothing.

$$CD = 0.2 \cdot \cos \theta_{CD}, \quad \theta_{CD} \in [0, 2\pi]$$

$$MD = 0.3 \cdot \sin \theta_{MD}, \quad \theta_{MD} \in [0, 2\pi]$$

$$\sigma = 0.2$$

1st bump :

$$\text{amplitude} = 1$$

$$\text{scan} \quad \#30 \dots 39$$

$$\text{sensor} \quad \#30 \dots 34$$

2nd bump :

$$\text{amplitude} = -1$$

$$\text{scan} \quad \#20 \dots 23$$

$$\text{sensor} \quad \#80 \dots 85$$

(3.11)

Streak :

$$\text{amplitude} = 0.6$$

$$\text{scan} \quad \#1 \dots 58$$

$$\text{sensor} \quad \#60 \dots 61$$

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{\text{expo}}}$
Daubechies	soft	std	6	3	0.82
Daubechies	soft	MultiMAD	6	3	0.99
Symmlet	soft	std	6	3	0.86
Symmlet	soft	MultiMAD	6	3	0.99
Coiflet	soft	std	3	3	0.85
Coiflet	soft	MultiMAD	3	3	1.00
Haar	soft	MultiMAD	2	2	1.04

Table 3.13 Bumps and streaks: summary of the results

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels	$\frac{J}{J_{exp0}}$
Daubechies	hard	MAD	6	3	0.74
Symmlet	hard	MAD	8	3	0.78
Coiflet	hard	std	2	3	0.73
Haar	hard	std	2	2	0.85

Table 3.14 Bumps and streaks: best filter performances.

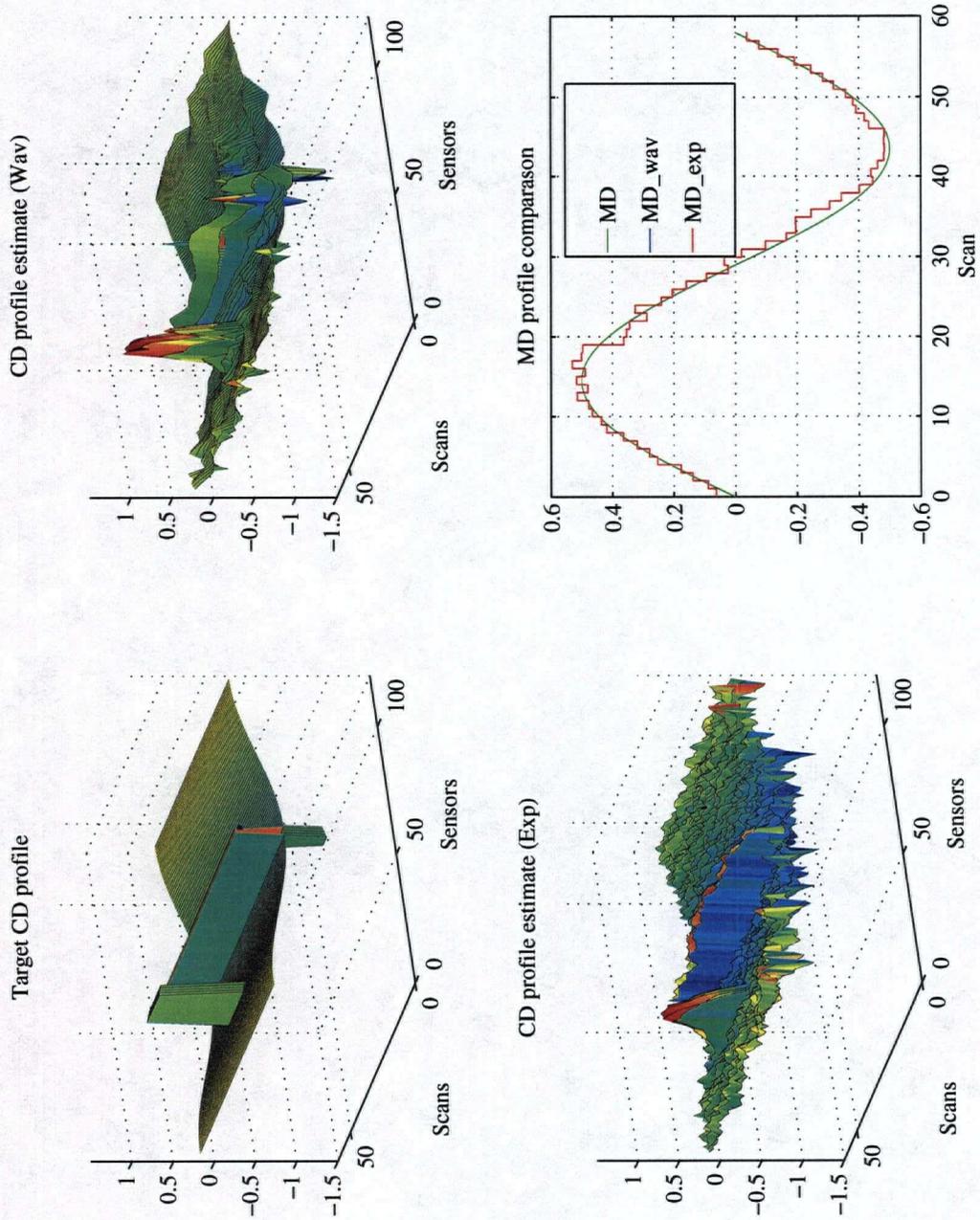


Figure 3.17: Bumps and streaks: estimated profiles.

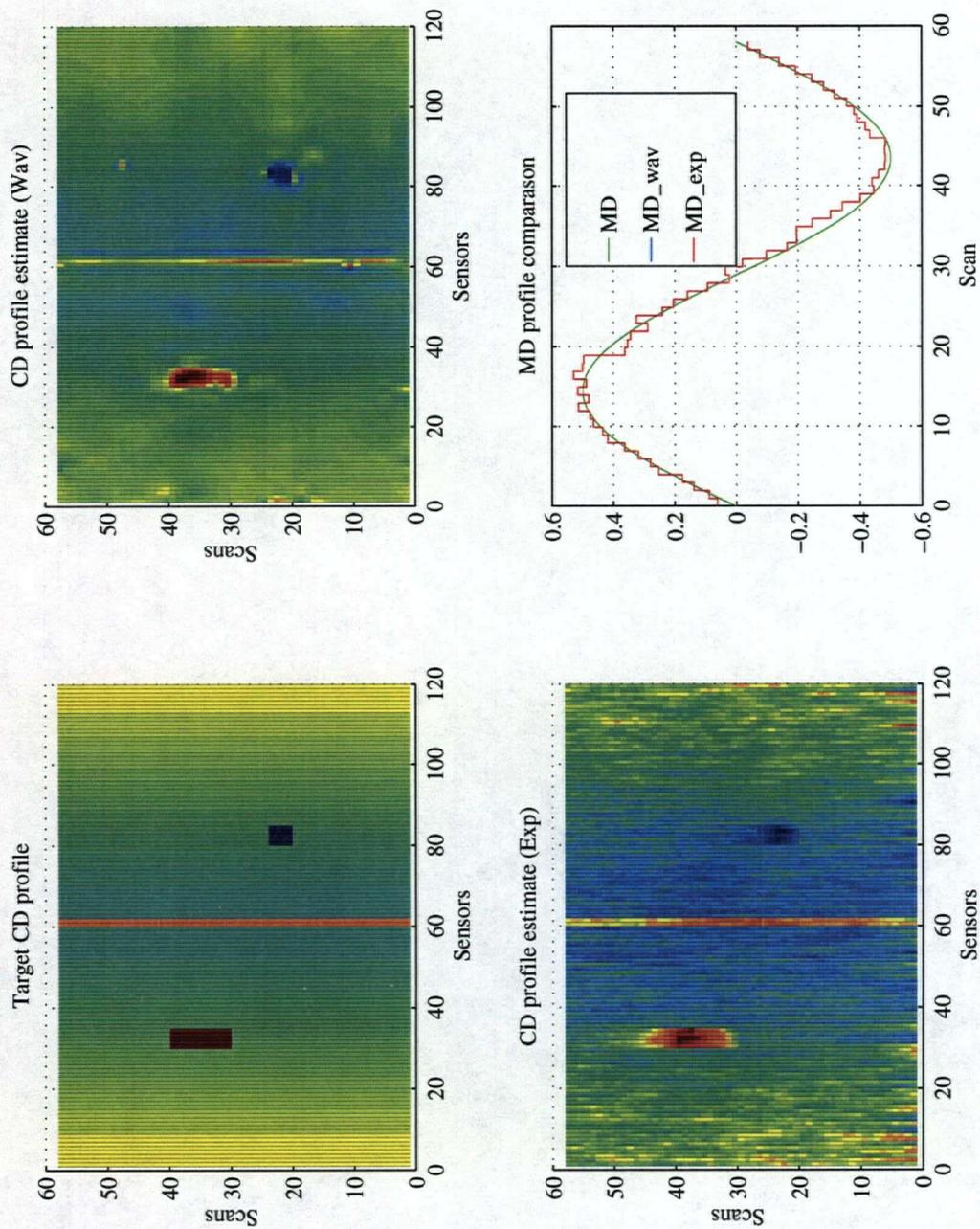


Figure 3.18: Bumps and streaks: estimated profiles (images).

3.3 Wavelet Analysis of Industrial Data

3.3.1 Choice of an Estimator

For almost all the tested cases, the selected group of wavelet estimators performed significantly better than the EXPO filter. The purpose of this testing was to show that even though we cannot choose a single optimal wavelet estimator for all the signals, one can be chosen from the selected group (given in the tables showing the summary results) and still achieve a good performance. When the features that have to be extracted are known (when looking for certain patterns on the sheet) an appropriate wavelet estimator can be selected from the table showing the best filters for each particular job. In this way a substantially better performance (compared to the conventional estimators) can be achieved, as has been shown for the given examples. This approach to the filter selection is now used when the estimators to be implemented on the industrial data are selected.

3.3.2 Pulp Machine Moisture Content Data

This data set is collected from a pulp machine manufacturing thick pulp [22] with the target value of 640.6 g/m^2 . Slice lip opening at the headbox was 15 cm. The headbox was 4.43 m wide with design speed of 184 m/min. The CD moisture control system was a *Devronizer* with 25 actuators each of a width of 15.24 cm. One scan took 34 seconds to complete. It has been recorded (from the actuator positions) that two CD bumps have occurred:

1. sensor #20, scans #19–31,
2. sensor #87, scans #13–30.

Three different wavelet estimators were used and their results were compared to the EXPO and EIBMC estimators. The chosen estimators are given in Table 3.15. The first estimator produced the results given in Figures 3.19, 3.20 and 3.21. This filter yielded very smooth results. The MD streak that can be seen in the scan #8 was picked up by the CD profile. It was concluded that this filter oversmooths data in the low-frequency range and consequently the number of resolution levels for the other two cases have been dropped to two. It has been observed, during the experiments, that the adaptive wavelet filtering methods work more reliably when the number of resolution levels has been kept low. It is well known [13, 10] that for adaptive filtering to work well, the number

of resolution levels used for de-noising must be much smaller than the maximum possible number of decomposition levels.

The other two filters (shown here for the comparison) have been more effective in the localization of bumps and the streaks on the paper sheet. The results obtained by the second filter have been shown in Figures 3.22 and 3.23 while Figures 3.24 and 3.25 demonstrate the results achieved by the Symmlet filter.

Wavelet	Thresholding method	Threshold estimation	Wavelet length/type	Resolution levels
Daubechies	soft	std	6	3
Daubechies	soft	MultiMAD	6	2
Symmlet	soft	MultiMAD	6	2

Table 3.15 Filters used to analyze the industrial data.

The wavelet filters have shown a superior performance compared to the conventional estimators when it came to the bump localization (see Figures 3.23 and 3.25). The size of the real bumps, based on the known positions of the actuators and their widths, was 13x4.8 and 18x4.8 for the first and the second bump respectively. Both, EXPO and EIBMC have shown signs of dispersing the bumps over a much greater number of scans although the width of the bumps was preserved. The wavelet estimators have detected the bump amplitude much better than the other two estimators.

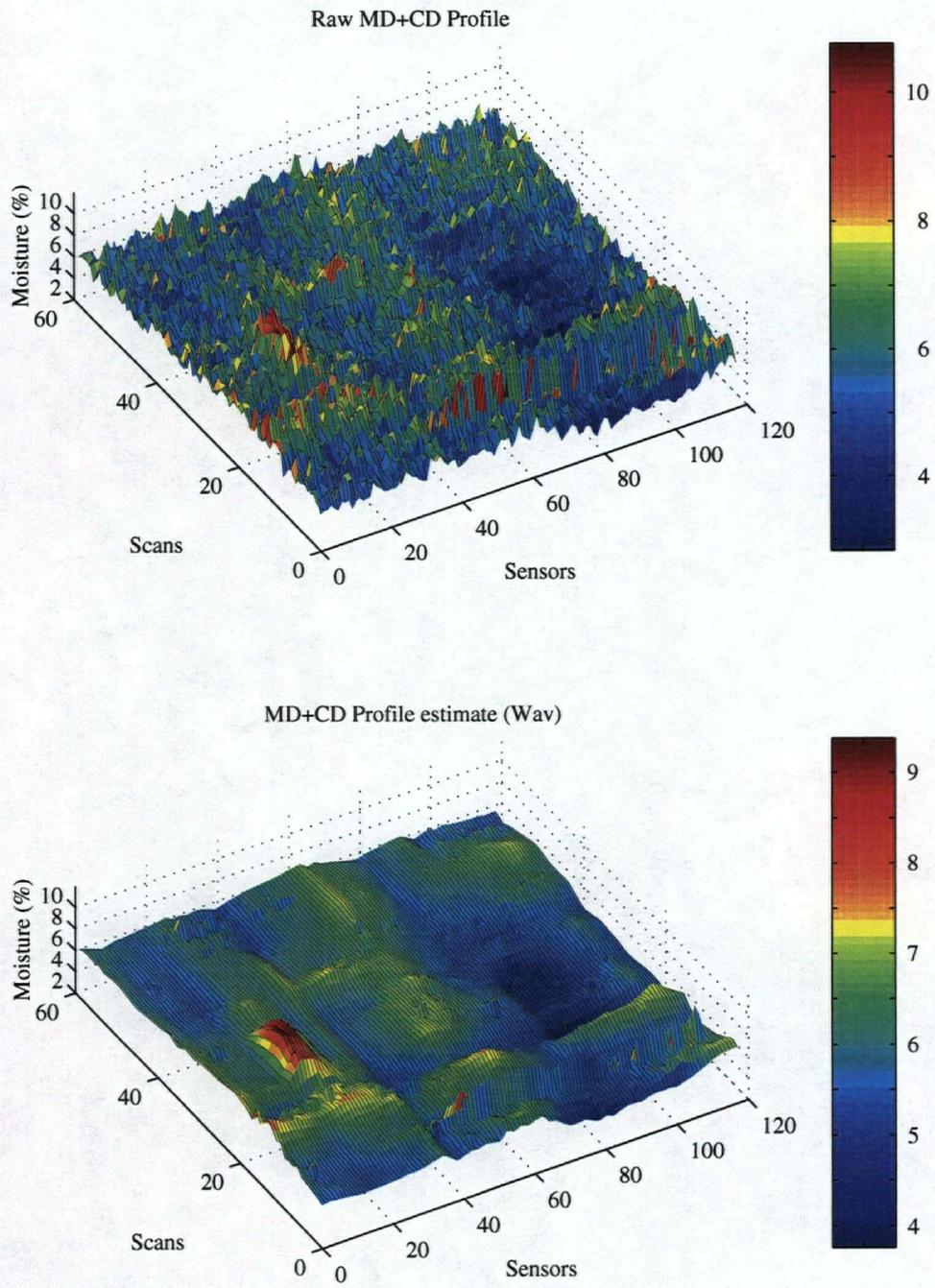


Figure 3.19: Pulp machine moisture data: estimated profiles.

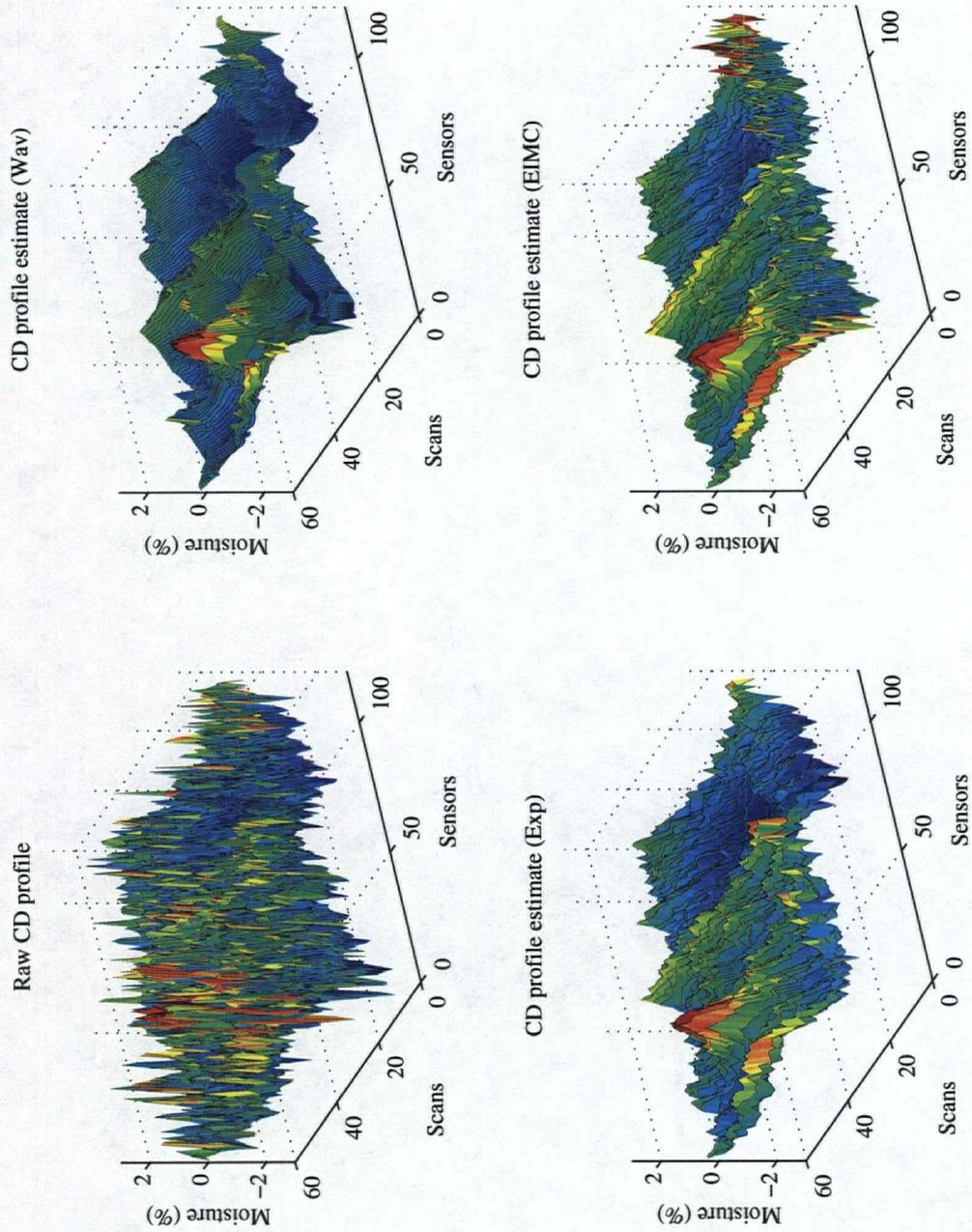


Figure 3.20: Pulp machine moisture data: estimated profiles [1]

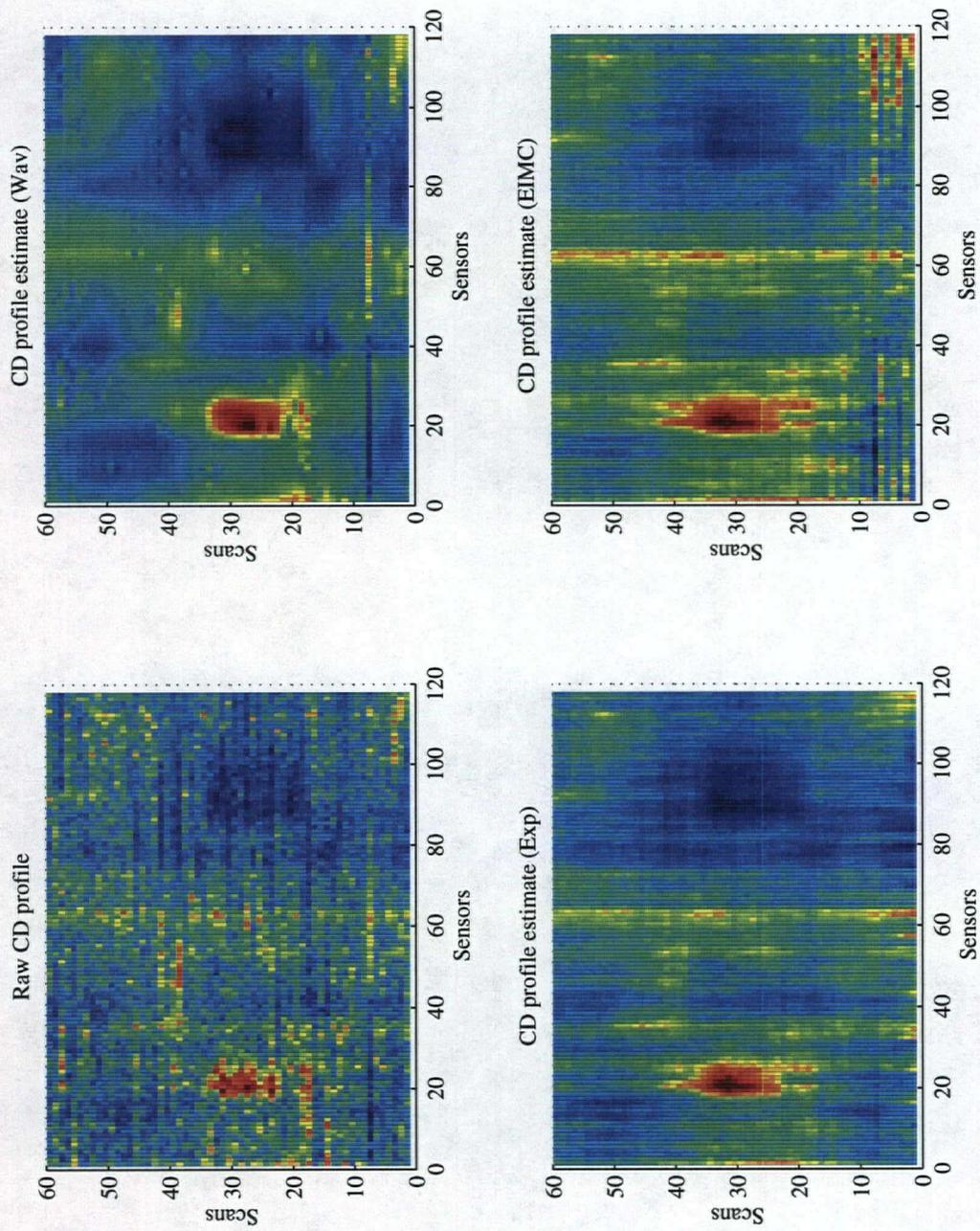


Figure 3.21: Pulp machine moisture data: estimated profiles [1] (images).

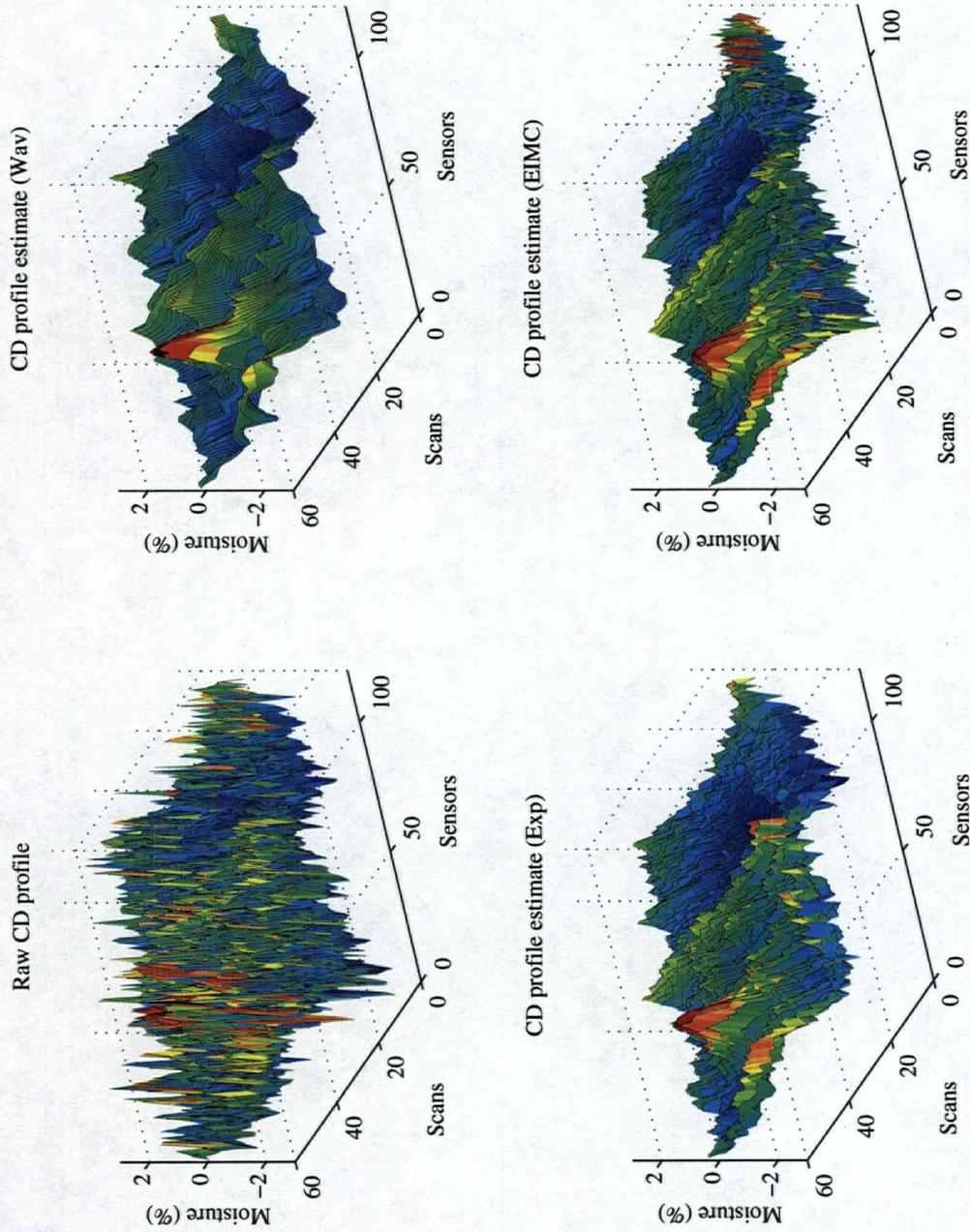


Figure 3.22: Pulp machine moisture data: estimated profiles [2]

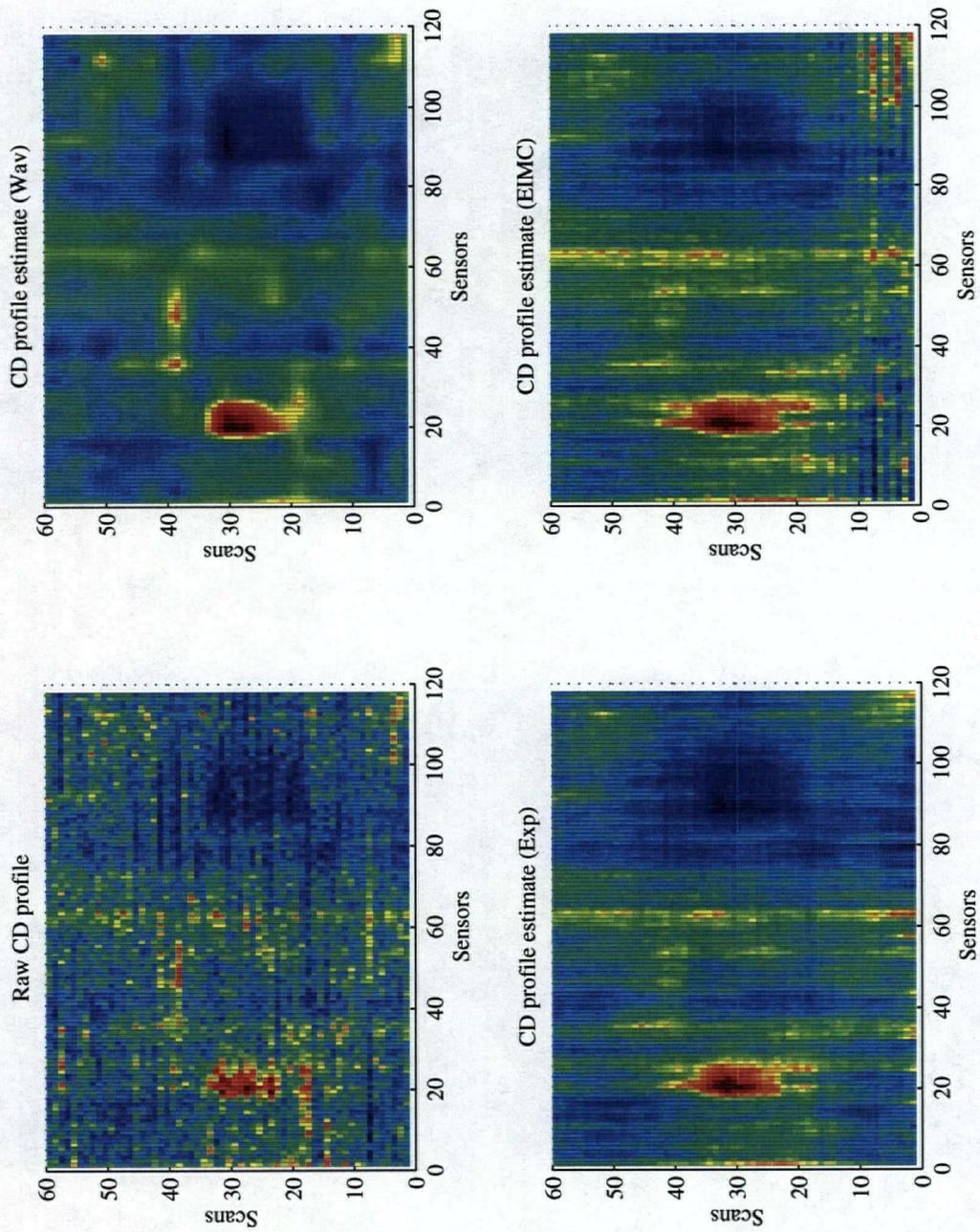


Figure 3.23: Pulp machine moisture data: estimated profiles [2] (images).

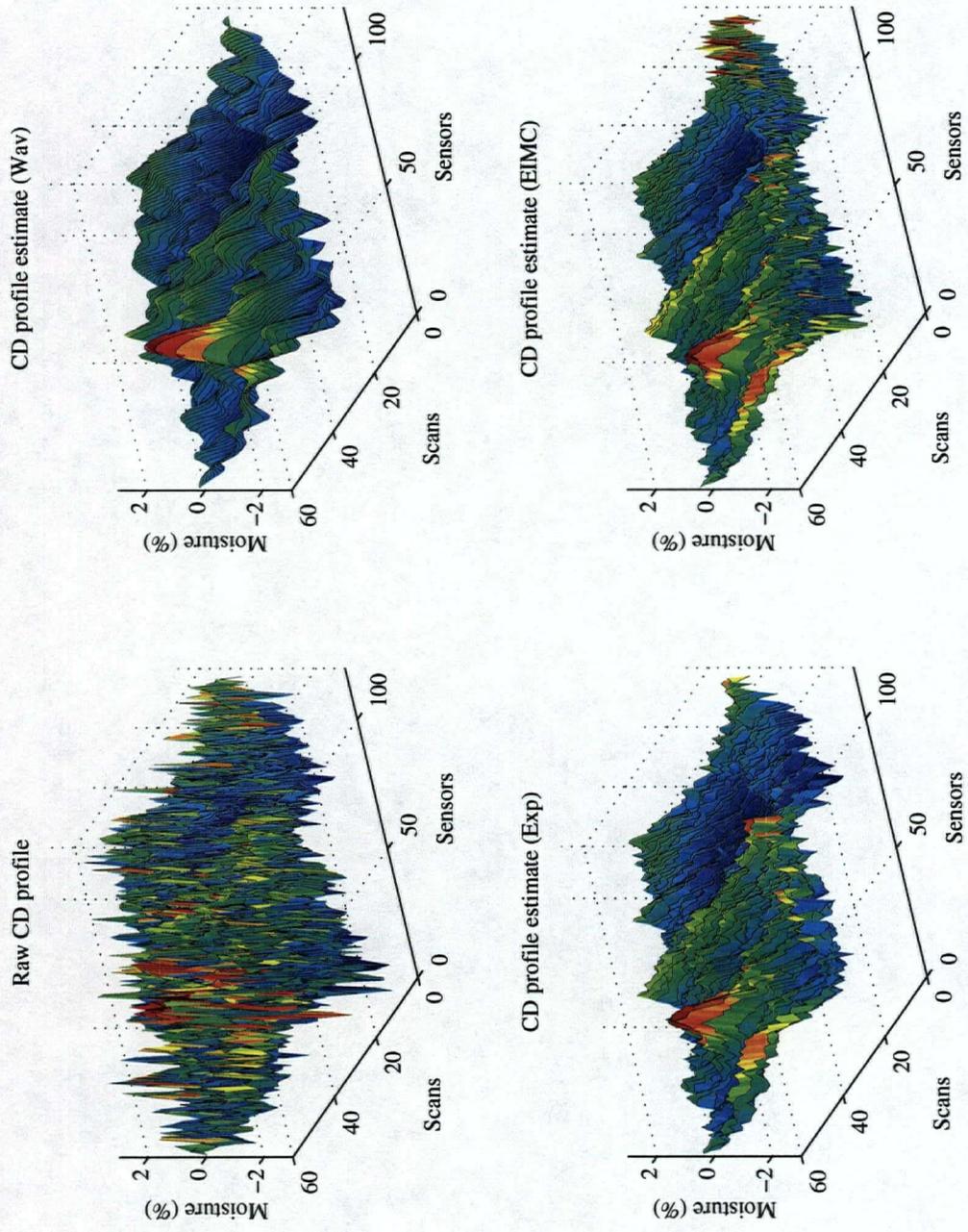


Figure 3.24: Pulp machine moisture data: estimated profiles [3]

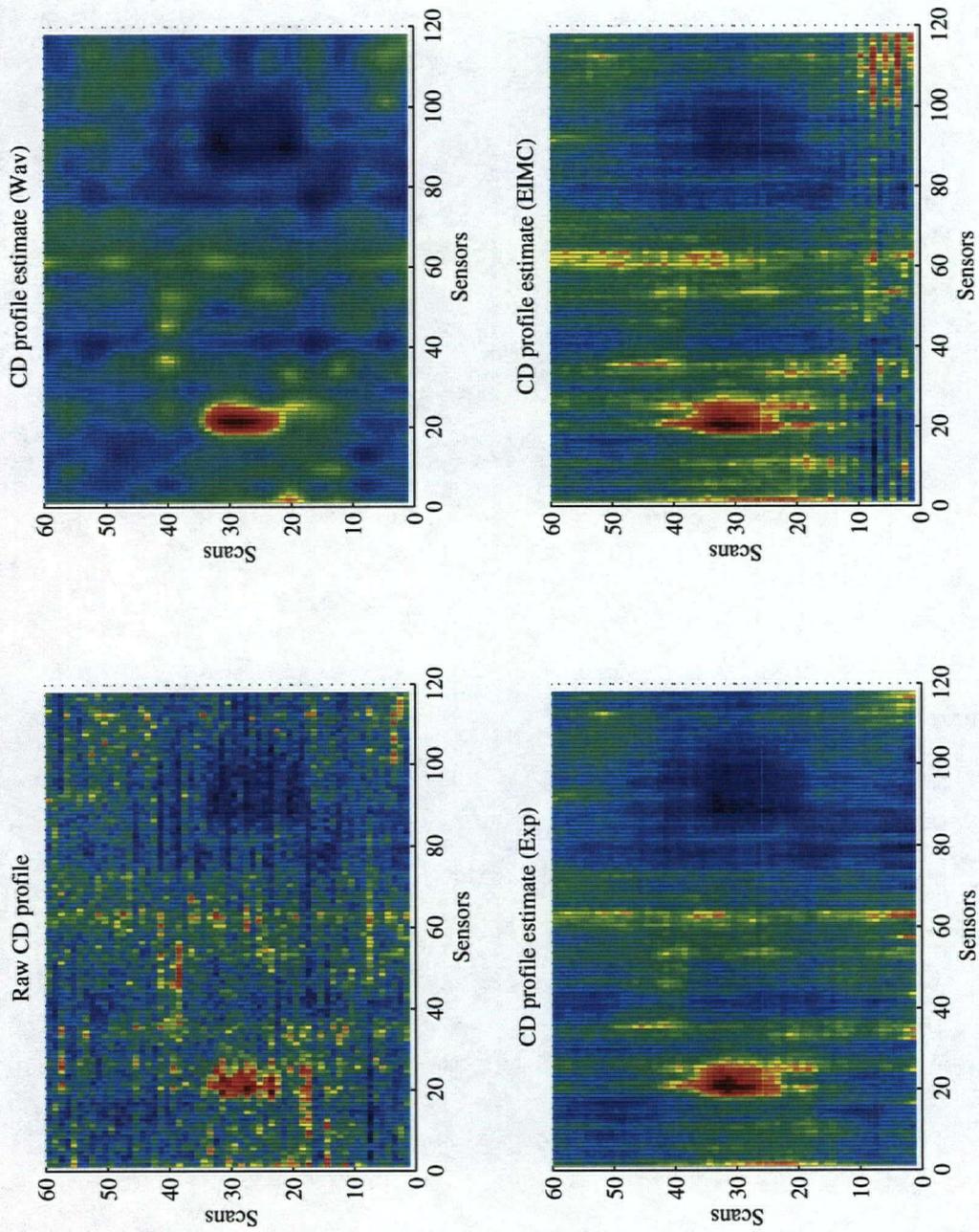


Figure 3.25: Pulp machine moisture data: estimated profiles [3] (images).

3.3.3 Grade Change for a Paper Machine

This data set is collected from a newsprint machine during a grade change with the dry weight control system in auto mode. Data used for the analysis was the basis weight in grams per square meter. There were 141 sensors and data was collected during 189 scans. This data set was much larger than those previously analyzed so that it was possible to select estimators that used more resolution levels for de-noising. A Coiflet3 filter has been used and soft thresholding has been applied on 5 resolution levels using a threshold $t = \hat{\sigma} \sqrt{2 * \log(n * m)}$ where the standard deviation has been estimated from the wavelet coefficients at the finest resolution level: $\hat{\sigma} = std(\omega_f)$. The results are shown in Figures 3.26 and 3.27. The estimated CD profiles look very smooth, and it is possible to see that some of the MD change has found its way into the CD profile. This has happened because the MD profiles were calculated (and updated) once per scan. An alternative method of MD profile estimation will be proposed in the next Chapter in order to avoid this difficulty.

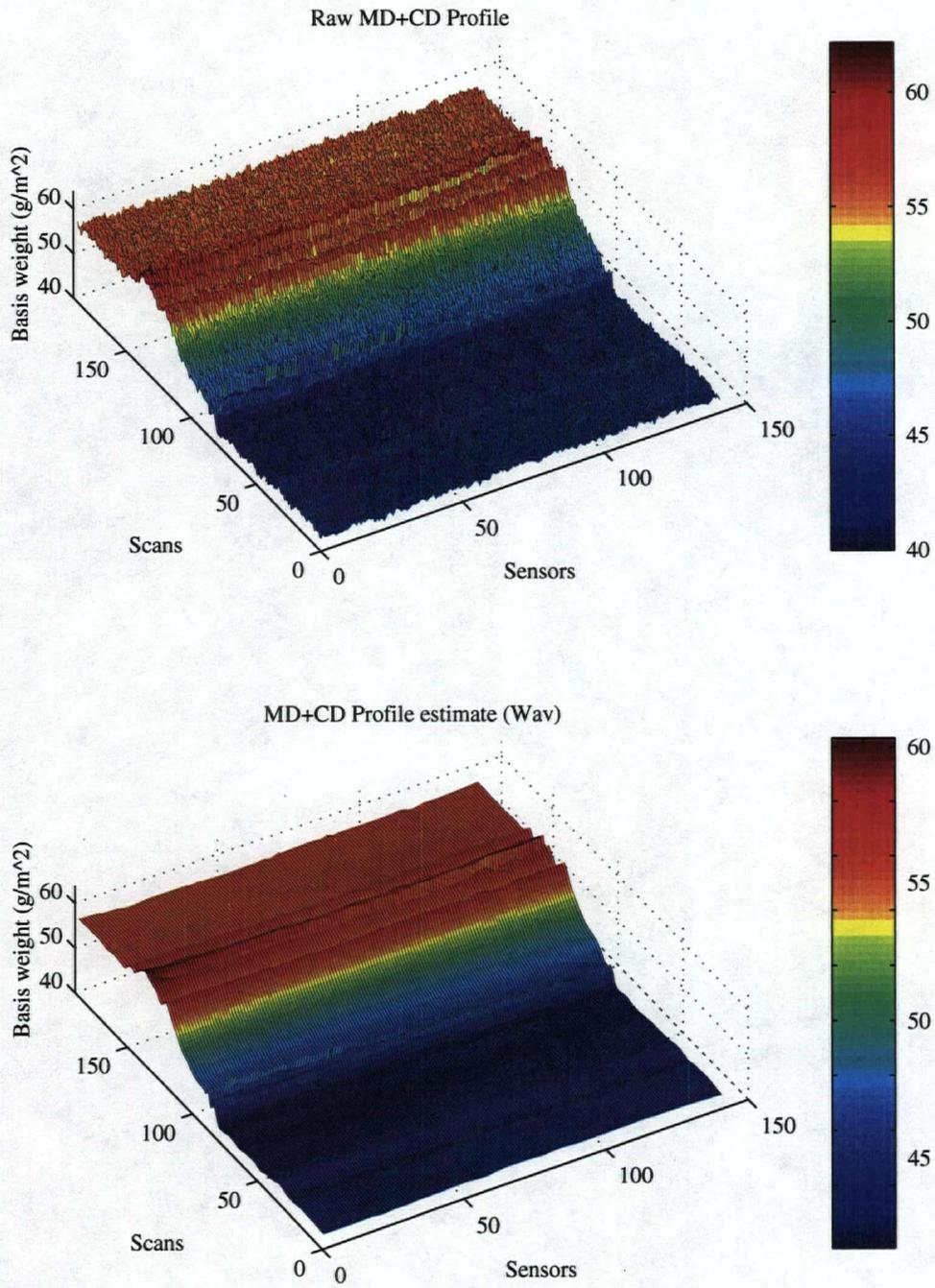


Figure 3.26: Paper machine basis weight data.

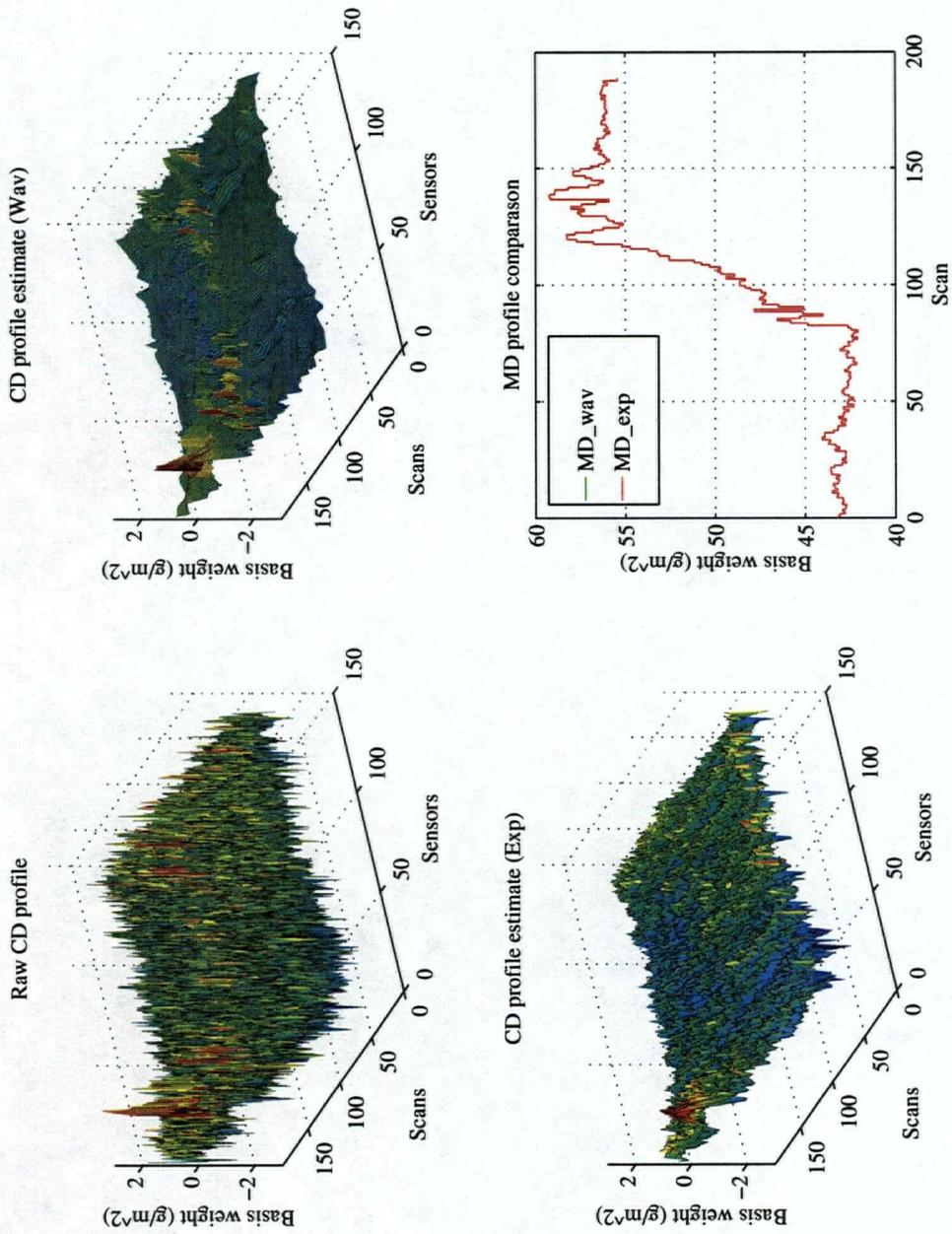


Figure 3.27: Paper machine basis weight data: estimated profiles.

Chapter 4

Other Applications of Wavelets in Paper Making Process

4.1 Compression of Paper Machine Data

The theoretical background needed for the understanding of the compression algorithms used in this chapter has been given in Section 2.5.2. In this chapter some of the more practical aspects of the process-data compression method, as well as the results achieved will be given. Both compression methods given in Section 2.5.2, (compression with and without quantizer) have been used. It has been assumed that the input data has been measured and stored in 16-bit format. All the calculations have been done using double precision (Matlab™'s default precision). The software has been written in such a way that quantizer resolution could assume any value. Only three resolutions have been used for the test purposes: 8-, 12- and 16-bit. The 16-bit resolution has been used as the reference, yielding the compression ratio of 1:1. The 12-bit resolution corresponded to a compression ratio of 1.33:1 and 8-bit resolution resulted in a compression factor of 2:1. All the above mentioned compression ratios are in addition to the compression achieved by the thresholding of wavelet coefficients and the encoding of data into sparse matrices.

Case	Wavelet	Threshold estimation	Wavelet type/length	Resolution levels
1.	Daubechies	std	6	3
2.	Daubechies	MultiMAD	6	3
3.	Symmlet	std	6	3
4.	Symmlet	MultiMAD	6	3
5.	Coiflet	std	3	3
6.	Coiflet	MultiMAD	3	3
7.	Haar	MultiMAD	2	2

Table 4.16 Key table

The examples used in Chapter 3 have been used to test the ability of the wavelet compressor to compress different types of signals. The results are summarized in Tables 4.17 to 4.21. The wavelet filters used here are given in Table 4.16.

Case	16-bit		12-bit		8-bit	
	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$
1.	30.00	0.85	40.00	0.85	60.00	1.17
2.	50.80	1.15	67.74	1.15	101.61	1.43
3.	28.06	0.85	37.42	0.85	56.13	1.02
4.	52.33	1.19	69.77	1.19	104.66	1.32
5.	27.08	0.82	36.11	0.83	54.16	0.85
6.	52.73	1.27	70.30	1.27	105.45	1.29
7.	13.57	0.46	18.09	0.46	27.13	0.58

Table 4.17 Edge effects: summary of compression results

Case	16-bit		12-bit		8-bit	
	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$
1.	50.43	0.51	67.25	0.51	100.87	0.66
2.	50.80	0.51	67.74	0.51	101.87	0.66
3.	49.36	0.39	65.82	0.39	98.72	0.51
4.	51.56	0.39	68.74	0.39	103.11	0.51
5.	49.01	0.42	65.35	0.42	98.03	0.43
6.	53.54	0.42	71.38	0.42	107.08	0.42
7.	13.57	0.95	18.09	0.95	27.13	0.96

Table 4.18 Smooth surfaces: summary of compression results

The actual quantization used here is given by this four-step procedure (for a data set D):

1. find $t_{min} = \min(D)$ and $t_{max} = \max(D)$
2. $\Delta t = (t_{max} - t_{min}) / (2^{bits} - 1)$
3. $D_1 = D - t_{min}$
4. $D_c = \text{floor}(D_1 / \Delta t)$

where the variable *bits* represents the resolution in the number of bits and the function *floor* rounds the given number to the first integer smaller or equal to the number.

The reconstruction (de-quantization) algorithm is given by

$$D' = D_c * \Delta t + t_{min} \quad (4.1)$$

Case	16-bit		12-bit		8-bit	
	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$
1.	19.39	0.49	25.85	0.49	38.77	0.51
2.	50.08	0.39	67.74	0.39	101.61	0.42
3.	16.93	0.50	22.58	0.50	33.87	0.53
4.	51.56	0.38	68.74	0.38	103.11	0.42
5.	17.66	0.51	23.55	0.51	35.33	0.53
6.	53.54	0.38	71.38	0.38	107.08	0.90
7.	13.57	0.71	18.09	0.71	27.13	0.71

Table 4.19 Grade change: summary of compression results

Case	16-bit		12-bit		8-bit	
	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$
1.	6.98	0.78	9.31	0.78	13.97	0.82
2.	48.00	0.69	64.00	0.69	96.00	0.74
3.	6.87	0.79	9.16	0.79	13.74	0.85
4.	44.38	0.74	59.18	0.74	88.77	0.80
5.	6.81	0.76	9.09	0.76	13.63	0.76
6.	45.96	0.67	61.28	0.67	91.91	0.67
7.	12.58	0.89	16.78	0.89	25.17	0.89

Table 4.20 MVM test data: summary of compression results

It can be concluded, from the given results, that very high compression ratios are achievable by the proposed algorithms. The additional errors introduced by the quantization of wavelet coefficients are very small even when the resolution is reduced to a half (8-bit). In most cases the increase in errors is negligible for the 12-bit resolution, and around 1 to 2 percent for the 8-bit resolution. A significant jump in errors sometimes happens when the compression ratios reach 80 to 100. It should be noted that in almost all the cases — even for the highest compression ratios, the overall relative

error stayed below one, suggesting that even for those cases the achieved accuracy of estimation after compression was better than obtained by the EXPO algorithm without compression.

Adaptive methods (MultiMAD) have given much better results (Table 4.20) due to their ability to adjust the thresholds at different levels. The method moved to higher thresholds at some resolution levels so setting more wavelet coefficients to zero (see Table 3.10 for an example).

Case	16-bit		12-bit		8-bit	
	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$	compress. ratio	$\frac{J_{compress}}{J_{expo}}$
1.	41.18	0.86	54.91	0.86	82.37	0.87
2.	51.56	1.06	68.74	1.06	103.11	1.07
3.	38.88	0.91	51.84	0.91	77.77	0.91
4.	51.56	1.06	68.74	1.06	103.11	1.07
5.	36.63	0.89	48.84	0.89	73.26	0.91
6.	53.13	1.07	70.84	1.07	106.26	1.08
7.	13.51	1.11	18.02	1.11	27.03	1.11

Table 4.21 Bump data: summary of compression results

The test data set created by the MVM is used to show the visual effects of the compression. This data set has been chosen because of its richness in MD and CD frequencies yielding relatively smaller compression ratios (Table 4.20). The results described are achieved by the filter #6 (compression ratio of 81.5:1). The energy distribution of the signal is shown in Figure 4.1(a), and is very close to normal distribution. Figures 4.1(b) and 4.1(c) depict, respectively, wavelet coefficients before and after the thresholding has been done (Note different scales in (a), (b) and (c)!). It is obvious, from the given diagrams, how wavelet compression operates. The wavelet coefficients are so concentrated in the quantization amplitude levels 50 to 75 that it had to be “zoomed in” to be able to see the low-frequency (high magnitude/energy) wavelet coefficients. The quantization level 62 corresponded to the wavelet coefficient of the zero magnitude. There were 2194 coefficients at this level. Even more impressive are the results after the thresholding was done. All the coefficients that used to be grouped around the zero value are now reduced to zero. There is only one peak at the level 62 and 8049 coefficients have that value. This compression resulted in only 143 nonzero coefficients.

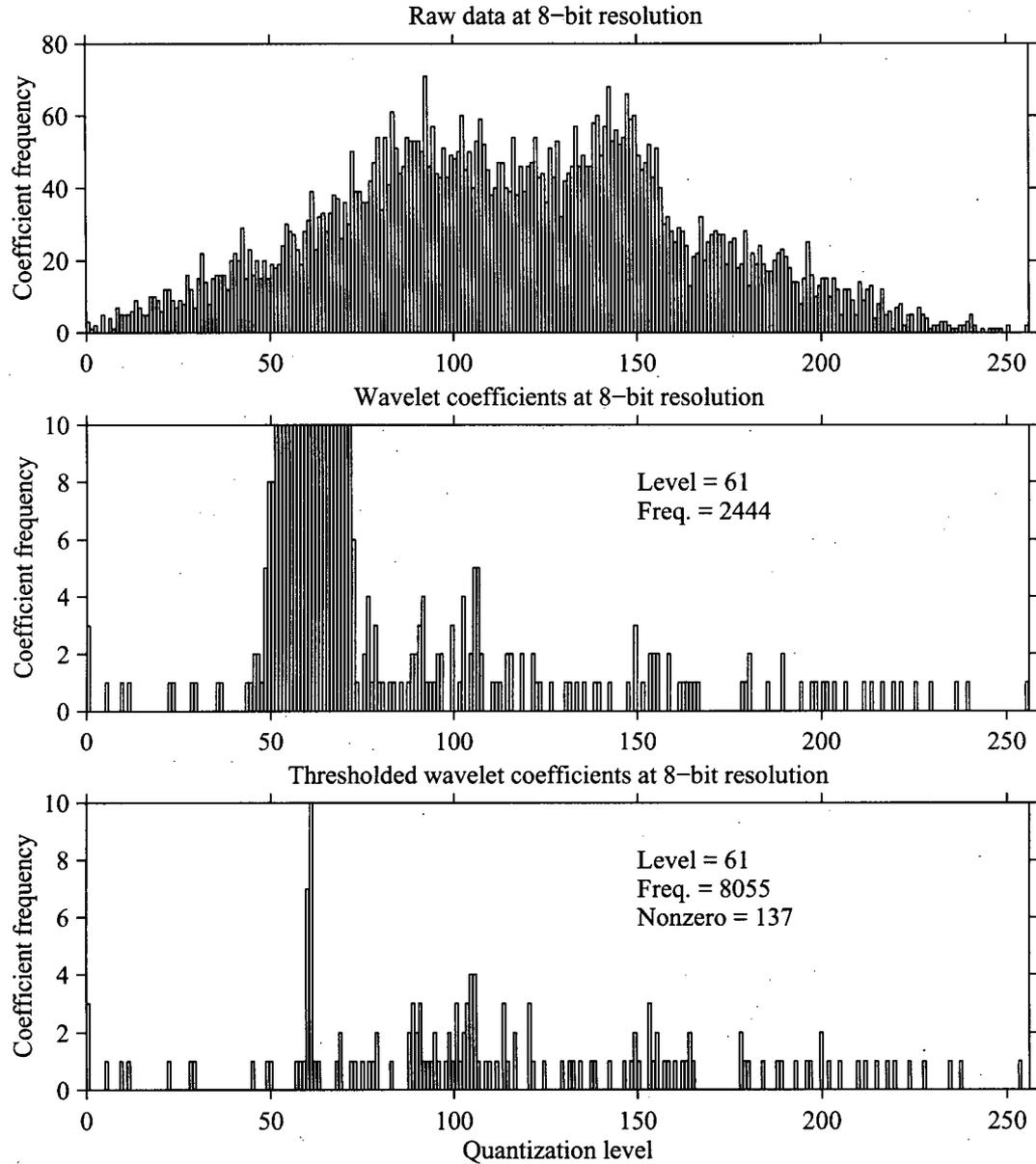


Figure 4.1: MVM data set: energy distribution

An alternative way to show the ability of wavelets to compress a given 2D signal can be seen in Figure 4.2. This figure depicts the wavelet coefficient distribution in a two-dimensional image. Each small black square represents one nonzero wavelet coefficient. The meaning of different regions, marked in Figure 4.2, can be understood by comparison to the diagram given in Figure 4.3 which is related to the calculation of the two-dimensional DWT (see Figure 2.3). The entire rectangle would be occupied by nonzero coefficients prior to compression.

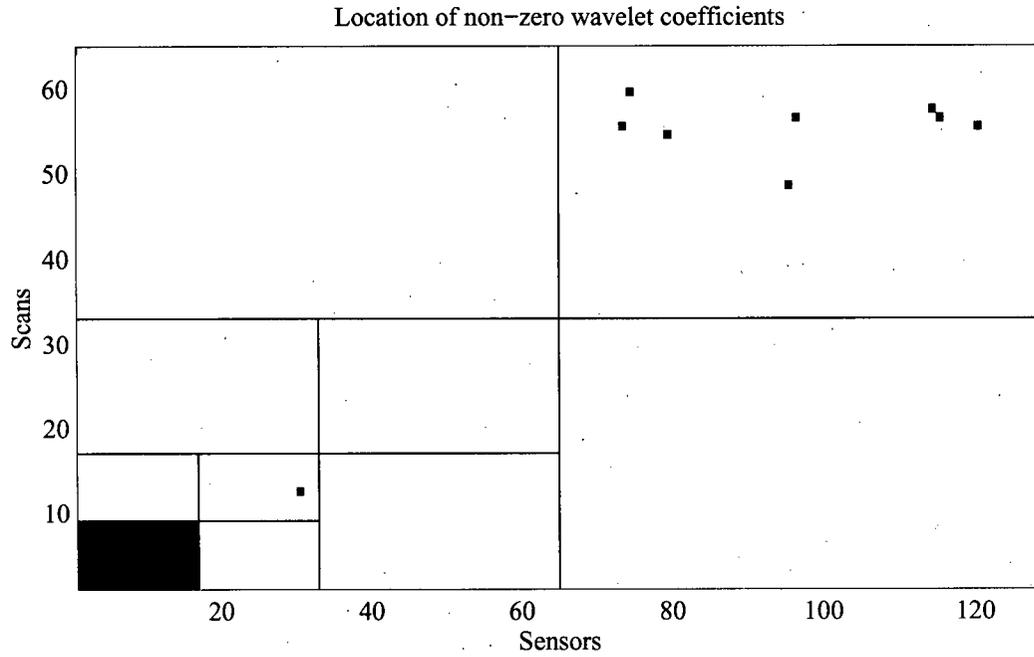


Figure 4.2: MVM data set: distribution of nonzero wavelet coefficients.

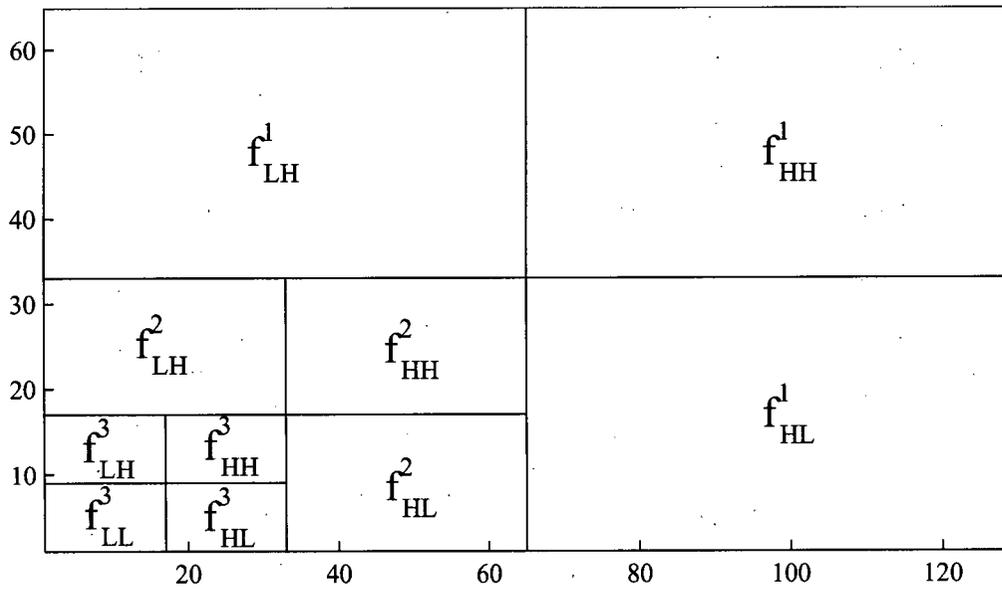


Figure 4.3: 2D DWT: coefficient storage.

The two-dimensional wavelet transform decomposes a signal into a low-frequency (average) signal (f_{LL}) and three high-frequency (detail) signals which are directionally sensitive [18]:

f_{LH} — emphasizes horizontal, CD-position dependant, features,

f_{HL} — emphasizes vertical, time-dependant, features,

f_{HH} — emphasizes diagonal, mutually-dependant, features.

This directional sensitivity is an artifact of the frequency ranges they contain.

It can be seen (Figure 4.2) that only a few nonzero coefficients lay outside of the f_{LL}^3 frequency component. It can be confirmed, from Figures 4.4 and 4.5, that the errors are indeed very small and visually they appear as a kind of fuzzyness surrounding the estimated profiles.

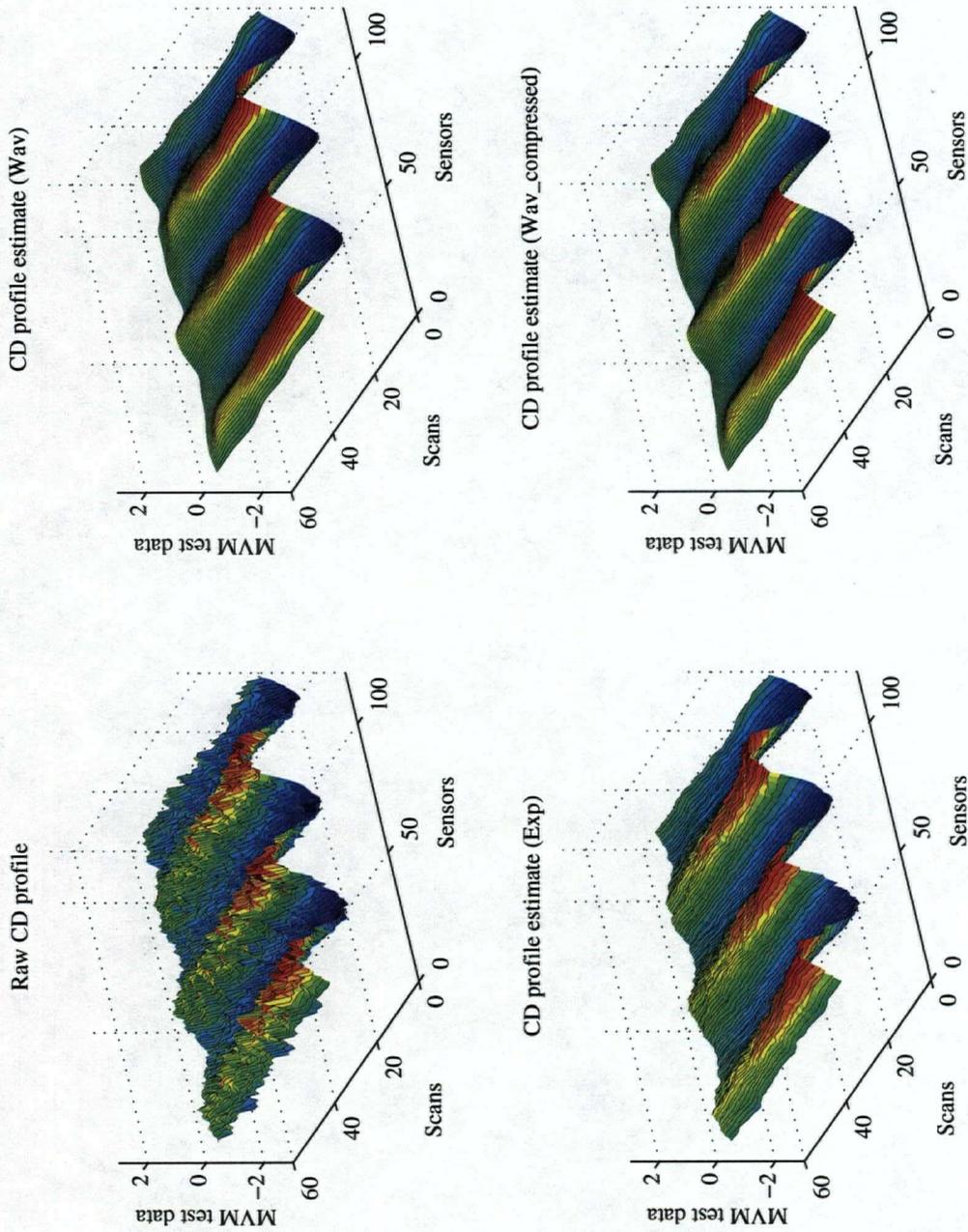


Figure 4.4: MVM data set: compressed data.

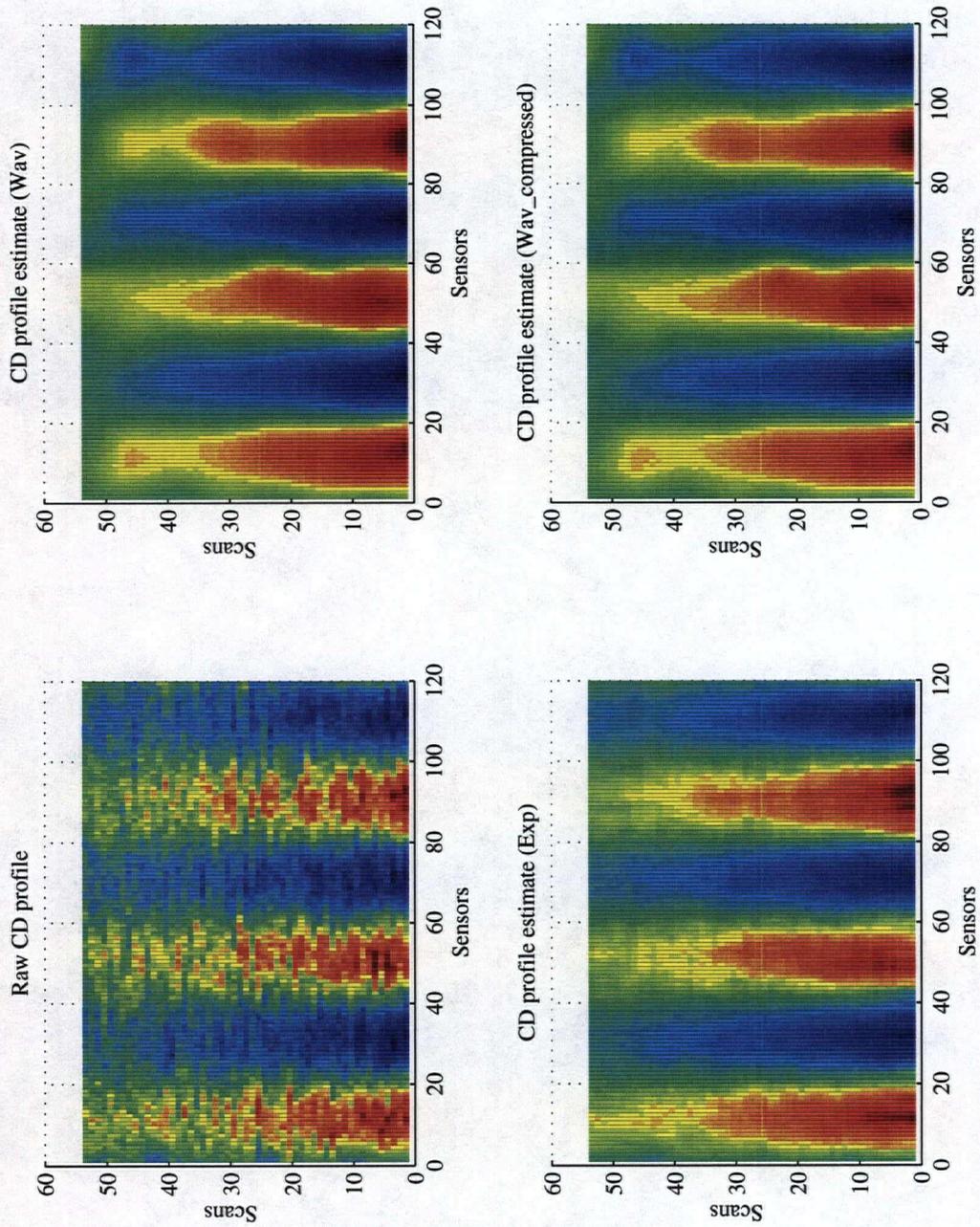


Figure 4.5: MVM data set: compressed data (images).

Compression has been also tested on the industrial data set given in Section 3.3.2. The compression ratios and the errors are given in Table 4.22. The filters are given in the same order as in Table 3.15. In this case the errors (Eq. 3.3) were calculated based on the estimated signal obtained by the same filter since the signal is not known. The errors were very small — even the largest one (case #1) corresponds to an average per sample error of $12.42/(60 * 118) = 0.0018$ in comparison to the standard deviation of the estimated signal which was around 0.7.

Figures 4.6 to 4.9 are obtained by the filter #3 from Tables 3.15 and 4.22 at the 8-bit resolution. It is impossible to see any difference between compressed and non-compressed images. There is only one wavelet coefficient (Figure 4.7) left outside of the l_{LL} region.

Based on the results shown, it is easy to see that the proposed compression method, which is a natural part of the estimation process, yields high compression with almost no additional errors. Different compression strategies may be developed to assure the smallest possible error and the highest compression ratio, based on the use to be made of the stored data. If data is to be used as a simple image-record of the production, good results may be achieved by simply selecting 2–5% of the largest wavelet coefficients before thresholding. This would produce a constant compression ratio. If the record showing the estimated profiles is to be kept, then the data processing described in this section can be used (estimation followed by compression, using the same wavelet coefficients). If there is a need to keep data with high-frequency components still present, thresholding may be done by selecting the threshold conservatively (i.e. one half of the one automatically selected by the estimator) and by using non-adaptive methods. This filtering would remove a portion of the noise but would still achieve a useful compression ratio.

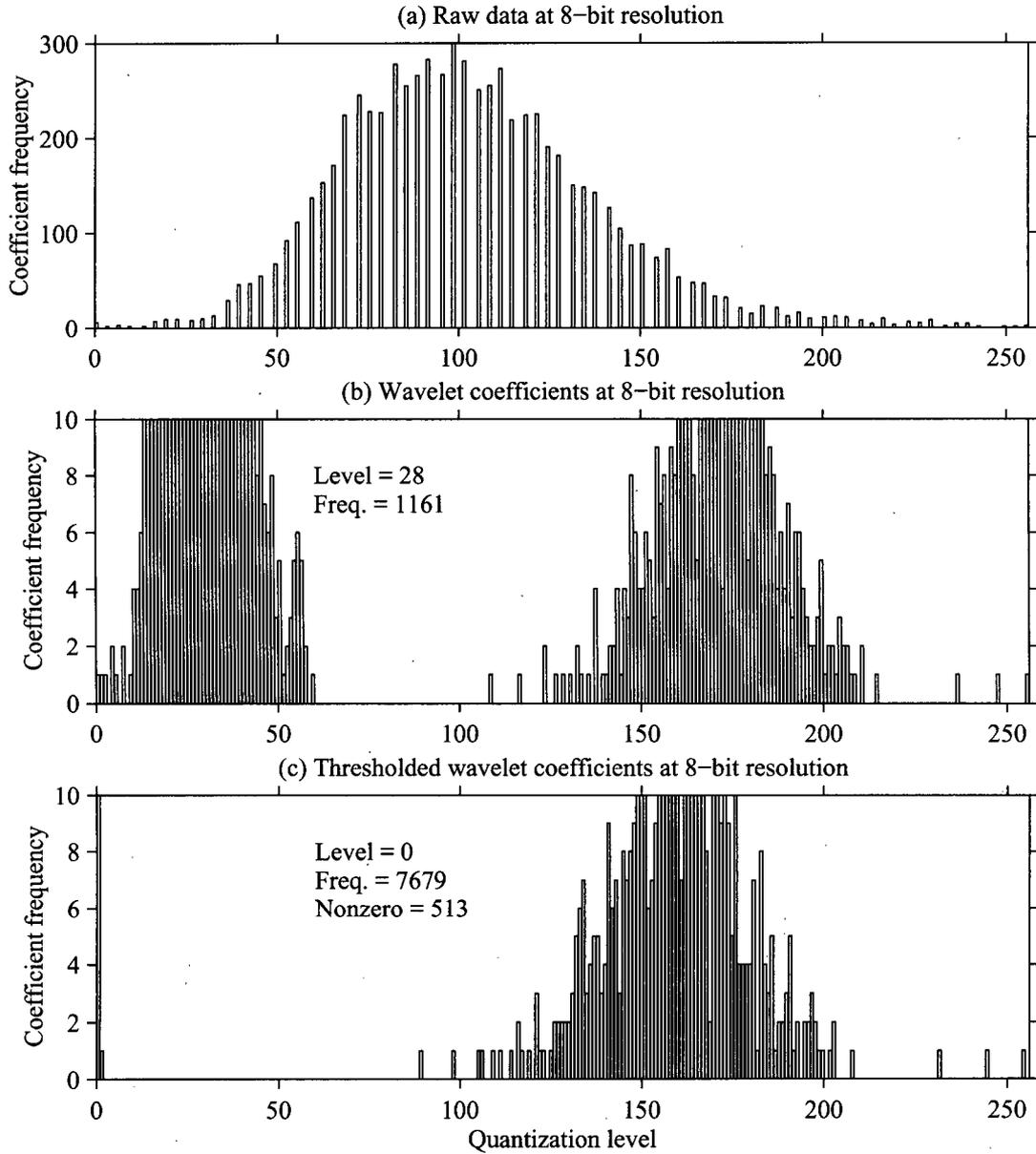


Figure 4.6: Pulp machine moisture data: energy distribution

Case	16-bit		12-bit		8-bit	
	compress. ratio	$J_{compress}$	compress. ratio	$J_{compress}$	compress. ratio	$J_{compress}$
1.	11.13	0.06	17.77	0.19	26.66	12.42
2.	13.62	0.01	18.16	0.42	27.24	2.37
3.	13.72	0.00	18.24	0.06	27.44	0.89

Table 4.22 Pulp machine moisture data: summary of compression results

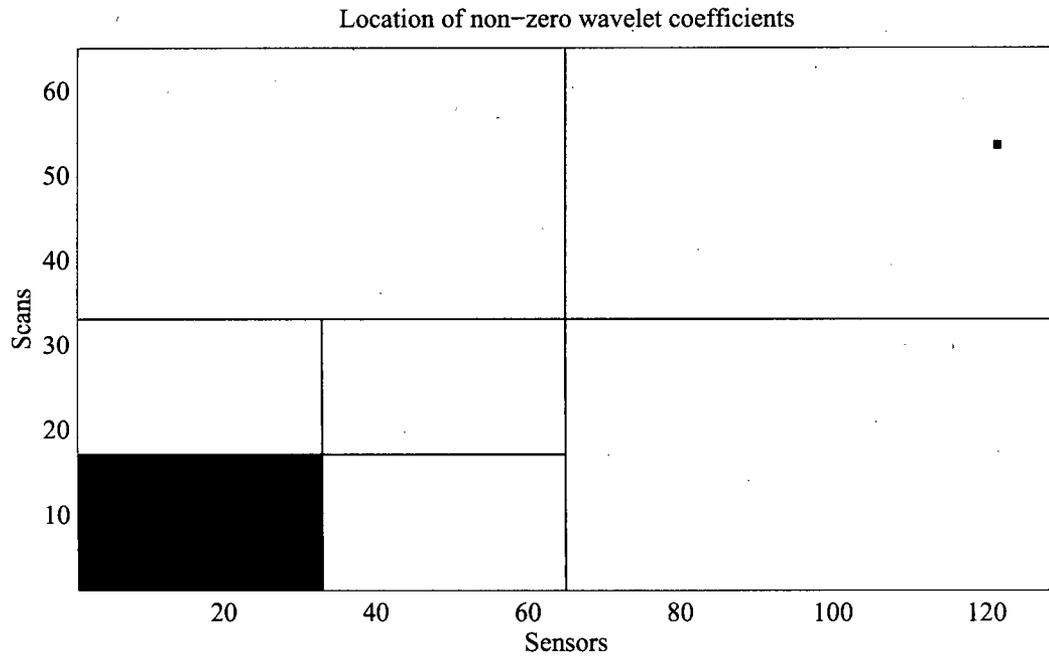


Figure 4.7: Pulp machine moisture data: distribution of nonzero wavelet coefficients.

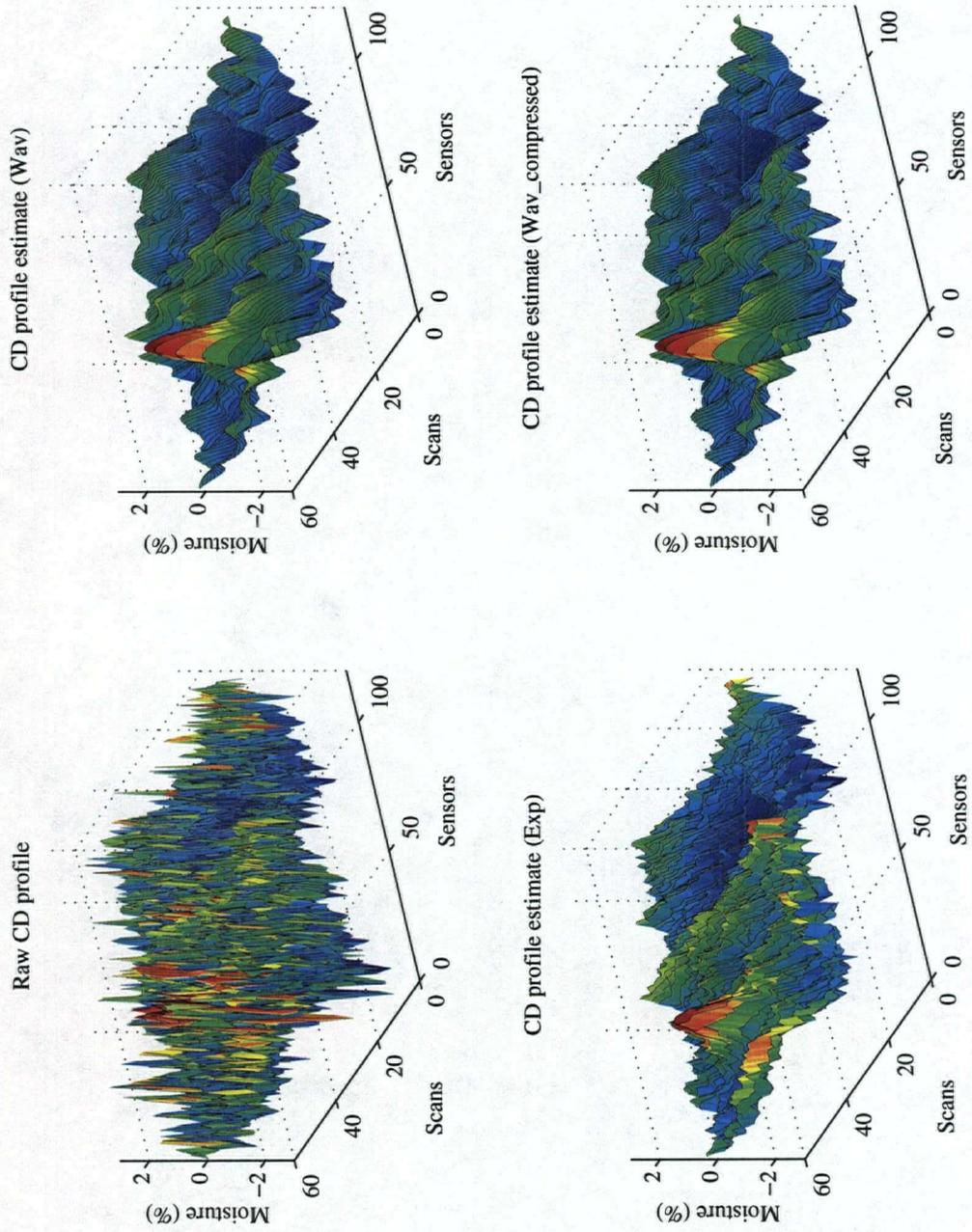


Figure 4.8: Pulp machine moisture data: compressed data.

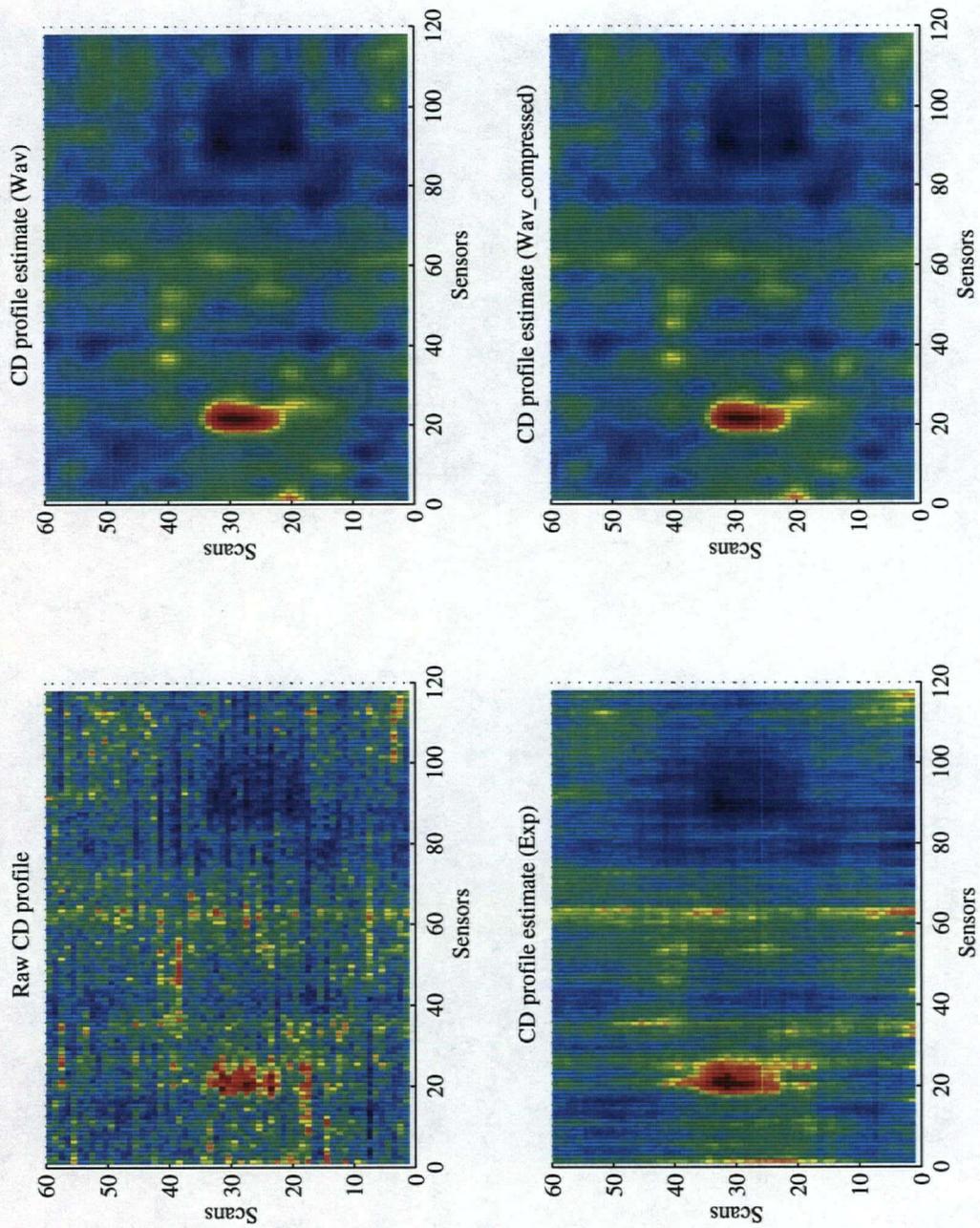


Figure 4.9: Pulp machine moisture data: compressed data (images).

4.2 Wavelet Estimation for Control Purposes

The wavelet estimator presented in Chapter 3 produced very good results with low estimation errors and at high execution speeds. Those estimates, in the form of two- and three-dimensional images can provide a paper machine operator with a much better view of the production quality. The operator acts on the basis of this improved information leading ultimately to better control and paper quality. The estimator, as given in Chapter 3, worked in the batch mode on a relatively large number of data samples (more than 50 scans and more than 110 sensor measurements per scan) and was quite suitable for the above purposes. Some modifications to the proposed algorithms are needed for the estimation to become more appropriate for use in real-time control. Those modifications concentrate on more practical issues such as the speed of MD profile updates and the handling of edge effects. This is mainly a topic for future research. One possible approach to this problem is given as follows.

The CD profiles should be estimated as shown in Chapter 3. The number of scans that need to be used for each CD profile calculation depends on the filter length and it should be at least in the range of 2 to 3 filter lengths. The most recent estimates of the signal are the most important because these are the samples needed for the controller to respond at the highest rate. Hence, better handling of those data samples may be needed (i.e. using Meyer's boundary wavelets or Dyadic boundary wavelets). After the CD profile is calculated (once per scan, batch mode) the MD profiles may be estimated by subtracting the CD profile from the raw data and using the result as an input to a one-dimensional wavelet de-noising procedure. This approach may lead to MD estimation at each sample point instead of once per scan. An example of the MD profile estimation at each sample and its comparison to the scan average method is given in Figure 4.10. The data set used in this example was given in Section 3.2.2 (Smooth Surfaces). The true MD profile, its scan average estimate, as well as the high resolution estimates achieved by the proposed method are shown in Figure 4.10(b). The raw MD profile is shown in Figure 4.10(a) for comparison. The same method was applied to the pulp machine data (Section 3.3.2). The results as shown in Figure 4.11(a)-(b).

The wavelet estimator will thus enable more frequent MD estimates. The MD controller will then be able to operate with reduced time delay and possibly with increased bandwidth.

A second possible benefit is in CD control, where the profile estimate, after decomposition into wavelet components, can be matched to CD actuator spacing and response characteristic.

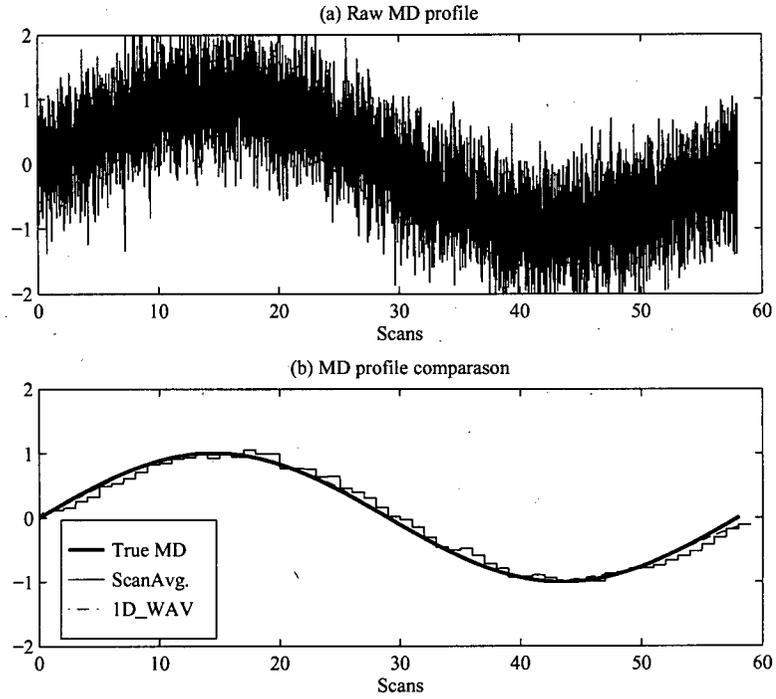


Figure 4.10: Smooth surfaces: high resolution MD estimation

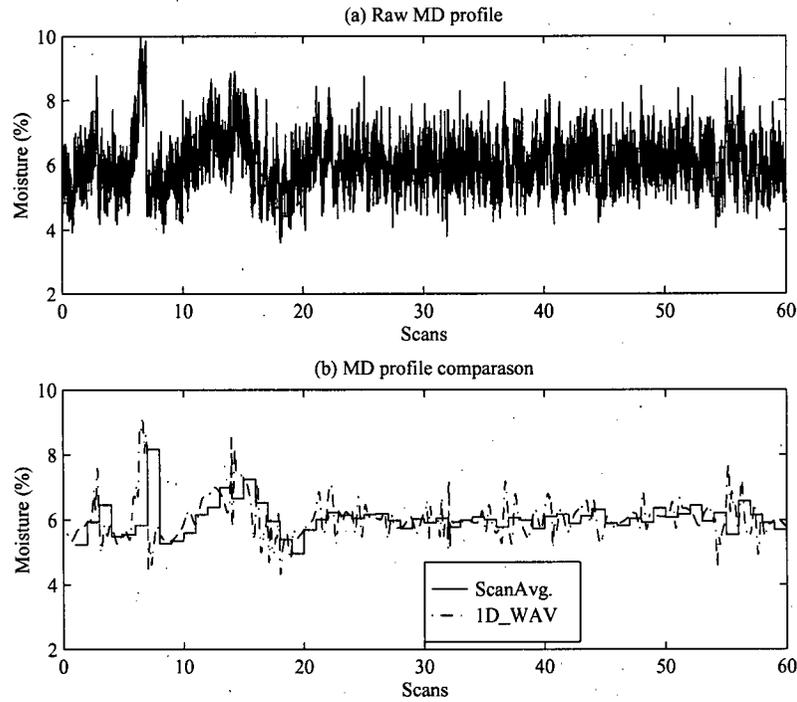


Figure 4.11: Pulp machine moisture data: high resolution MD estimation

Chapter 5

Conclusion

The thesis describes a new approach to paper machine process data analysis using one-dimensional and two-dimensional discrete wavelet transforms. These techniques have been adapted from a general theory that has been developed in recent years on the application of wavelets to signal analysis. Application areas in which the theory was first applied have included image processing and bandwidth compression for communications.

Two main applications of the discrete wavelet transform have been analyzed in this thesis. First, an analysis of the use of wavelets for processing scanned data representing basis weight and moisture variations on a paper machine has been carried out. It has been shown that wavelets are effective for the detection of process signals in noisy data, so leading to better estimation and visualization of the machine direction and cross machine variations in process data. The second main application of the method has been to allow significant compression of the process data without diminishing the ability to reconstruct accurate profiles. It has been shown that the compression method can be embedded into the estimation algorithm, producing excellent results without a major expense in the computation time.

It has been shown that, in both applications, the new methods produce results superior to the industrially accepted procedures. For appropriate choice of wavelets, profile estimates are improved over those obtained using exponential filtering or other standard analysis methods. The data compression technique presents a new concept in paper machine data analysis and the author is not aware of any previous references to this subject. The ability to reduce data storage requirements is of importance in mill-wide process monitoring systems.

A comprehensive analysis of the proposed algorithms has been carried out on a variety of simulated data sets for which the true process variations are known. Industrial data has also been analyzed and it is apparent that the method had many desirable characteristics.

The following properties have been shown for the wavelet estimator:

- The estimator reduces the error of estimation compared to the existing algorithms, when applied to test data.

- When applied to industrial data, the method presents an excellent visual interface to the paper making process. Operator data presentations are an important part of process control systems, and it is believed that wavelet analysis gives excellent visual insight into process variations, insight significantly better than is available in current systems.
- The wavelet decomposition can be closely related to cross machine actuator responses. It is expected that a cross machine control method based on an appropriate wavelet analysis will allow profile upsets to be matched to actuator settings.
- The algorithm is computationally fast, which allows real time analysis. The need for a recursive form may be avoided.
- The method has been shown to be applicable to basis weight and moisture estimation using data from a scanning sensor. It is to be expected that it could be used for the analysis of other process variables.
- Unlike many estimation methods based on least squares techniques, wavelet analysis is robust in its performance and does not require additional tuning once it has been properly installed.

The wavelet compression algorithm has been shown to have the following advantages:

- The method is very effective with compression ratios greater than 25:1 (up to 100:1) and with the error of estimation, after decompression, remaining lower than the corresponding error for an exponential filter based algorithm without compression.
- Data compression is an integral part of the estimator procedure, and therefore requires very little additional computation time.
- The algorithm provides an effective tool for the visualization of production variations over large time periods. It follows that it can be used to search for the possible causes of disturbances and faults.
- The algorithm can be adjusted to provide a minimum error or a constant compression ratio, allowing it to be tailored to particular applications.
- Data compression will become increasingly important with the introduction of new, high speed sensors that generate large volumes of data. A related advantage is the ability to reduce time for data transfers over the local area networks used in modern machines.

There are many possible extensions to the research presented here. Steps that might be taken to improve the effectiveness of the algorithms are described in the following:

- There is a need for better handling of the edge effects for the given DWT algorithms.
- For real time data analysis it is desirable that recursive versions of the algorithms replace the batch procedures.
- An organized procedure for incorporating knowledge of the process and noise models into the estimation procedures is needed.
- As the theory associated with wavelet decomposition advances, there is a need to evaluate new approaches. For example, there is a case for exploring new wavelet techniques such as wavelet packet analysis [38] and second generation wavelets [35].
- It is likely the use of wavelet estimation algorithms in automatic control applications will allow sensor and actuator spatial and temporal bandwidths to be matched.
- The existing threshold estimators that were used and play a critical role in data compression and noise elimination can still be improved. In some cases (for example the SURE technique), the results were not as expected and there is a need for further development of this part of the method.

References

- [1] J. Buckheit, S. Chen, D. Donoho, I. Johnstone, and J. Scargle. Wavelab 0.655, a library of matlab routines for wavelet analysis. <ftp://playfair.stanford.edu/pub/wavelab>, 1995.
- [2] S.-C. Chen. Kalman filtering applied to sheet measurement. In *7th American Control Conference*, pages 643–647, Atlanta, Georgia, 1988.
- [3] A. Cohen, I. Daubechies, and J. C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Applied Math.*, 1992.
- [4] E. Dahlin. Computational methods of a dedicated computer system for measurement and control on paper machines. In *24th Engineering Conference, TAPPI*, pages 62.1–62.42, San Francisco, California, Sep 1969.
- [5] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Applied Math.*, XLI(41):909–996, Nov. 1988.
- [6] I. Daubechies. Orthonormal bases of wavelets with finite support - connection with discrete filters. In J. M. Combes, A. Grossman, and P. Tchamitchian, editors, *Wavelets, Time-Frequency Methods and Phase Space*, pages 38–66. Springer-Verlag, Berlin, 1989. Proceedings of International Colloquium on Wavelets and Applications, Marseille, France, December 1987.
- [7] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis, Sept. 1990.
- [8] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992. Notes from the 1990 CBMS-NSF Conference on Wavelets and Applications at Lowell, MA.
- [9] I. Daubechies. Orthonormal bases of compactly supported wavelets II. Variation on a theme. *SIAM Journal of Mathematical Analysis*, 24(2):499–519, March 1993.
- [10] D. L. Donoho. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics*, volume 00, pages 173–205. American Mathematical Society, 1993.

- [11] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, 41(3):613–627, May 1995.
- [12] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. Technical Report 402, Stanford University, Department of Statistics, July 1992.
- [13] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Ass.*, 1993. To appear, Also Tech. Report 425, Department of Statistics, Stanford University, June, 1993.
- [14] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994. Also Tech. Report 400, Department of Statistics, Stanford University, July, 1992.
- [15] G. A. Dumont, M. S. Davies, K. Natarajan, C. Lindeborg, F. Ordubadi, Y. Fu, K. Kristinsson, and A. Jonsson. An improved algorithm for estimating paper machine moisture profiles using scanned data. In *30th IEEE Conference on Decision and Control*, Brighton, England, Dec 1991.
- [16] R. A. Gopinath, H. Guo, R. Hindman, M. Lang, J. E. Odegard, and D. Wei. Rice university wavelet toolbox, version 2.01. <http://www-dsp.rice.edu/software/RWT>, Apr 1994.
- [17] A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910. Also in PhD thesis.
- [18] M. Hilton, B. Jawerth, and A. Sengupta. Compressing still and moving images with wavelets. *Multimedia Systems*, 3(2), 1994. To appear.
- [19] A. K. Jain. *Fundamentals of digital image processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.
- [20] B. Jawerth and W. Sweldens. An overview of wavelet based multiresolution analyses. Technical Report 1993:1, Industrial Mathematics Initiative, Department of Mathematics, University of South Carolina, 1993. Submitted to SIAM Review.
- [21] I. M. Johnstone and B. W. Silverman. Wavelet threshold estimators for data with correlated noise. Technical report, University of Bristol, UK, Statistics Department, Sept. 1994.
- [22] I. M. Jonsson. Estimation and identification of moisture content in paper. Master's thesis, The University of British Columbia, Aug. 1991.

- [23] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, Jr. Noise reduction using an undecimated discrete wavelet transform. In *To appear in IEEE Signal Processing Letters*, 1995.
- [24] P.-G. Lemarié and Y. Meyer. Ondelettes et bases hilbertiennes. *Rev. Mat. Iberoamericana*, 2:1–18, 1986.
- [25] C. Lindeborg. A nonlinear algorithm for the estimation of moisture variation characteristics in the paper process. In *Proc. of 17th International Conference BIAS-81*, volume 3, pages 139–158, Milan, Italy, Oct 1981.
- [26] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Pattern and Machine Intelligence*, 11(7):674–693, 1989.
- [27] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, 1982.
- [28] Y. Meyer. Ondelettes, fonctions splines et analyses graduées. Lectures at Univ. of Torino, Italy, 1986.
- [29] Y. Meyer. Ondelettes: filtres miroirs en quadrature et traitement numérique de l'image. Technical report, Cahiers de Mathématiques de la Decision, Paris, France, 1989. Text from a conference on 22 February 1989 at the University of Autonooma of Madrid.
- [30] S. T. Morgan. Estimation and identification for machine direction control of basis weight and moisture. Master's thesis, The University of British Columbia, June 1994.
- [31] G. P. Nason. Wavelet regression by cross-validation. Technical Report 447, Stanford University, Statistics Department, Mar. 1994.
- [32] K. Natarajan, G. A. Dumont, and M. S. Davies. An algorithm for estimating cross and machine direction moisture profiles for paper machines. In *IFFAC/FORS Symposium*, pages 27–31, Beijing, PRC, Aug 1988.
- [33] W. K. Pratt. *Digital Image processing*. Wiley, New York, 1978.
- [34] O. Rioul and M. Vetterli. Wavelet and signal processing. *IEEE Signal Processing Magazine*, 8(4):14–38, Oct. 1991.

- [35] W. Sweldens. The lifting scheme: a construction of second generation wavelets. Technical Report 1995:6, Industrial Mathematics Initiative, Department of Mathematics, University of South Carolina, 1995.
- [36] B. F. Taylor. Optimum separation of machine-direction and cross-direction product variations. *Tappi Journal*, pages 87–92, Feb 1994.
- [37] X. G. Wang, G. A. Dumont, and M. S. Davies. Modelling and identification of basis weight variations in paper machines. *IEEE Transactions on Control System Technologies*, June 1993.
- [38] M. V. Wickerhauser. Smooth localized orthonormal bases. *CRAS*, 315, 1993.

Appendix A Exponential Multiple Scan Trending (EXPO) Algorithm

The equations describing the algorithm (Dahlin [1]) are:

$$\begin{aligned}\bar{Y}(m) &= \frac{1}{N} \sum_{n=1}^N D^n(m) \\ s^n(m) &= \rho(D^n(m) - \bar{Y}(m)) + (1 - \rho)s^n(m - 1)\end{aligned}\tag{A.1}$$

where:

$D^n(m)$ is the measured value (moisture content or basis weight) at scan m and CD position n ;

$\bar{Y}(m)$ is the mean value of scan m ;

N is the total number of samples (data boxes) in a CD;

$s^n(m)$ is the filtered CD profile of scan m at data box n ; and

ρ is the exponential filter weighting factor, $0 < \rho < 1$;

The weighting factor is usually chosen from the region $\rho \leq 0.3$. This eliminates the MD variations from the CD profiles but also slows down the CD profile estimation.

Appendix B Estimation and Identification of Basis Weight and Moisture Content (EIBMC) Algorithm

B.1 Process Models

Process model used to describe the moisture variation in the EIBMC algorithm is based on Lindeborg's [3] moisture variation model (MVM). It has been shown [6] that the same model, with some modifications, can be applied to basis weight. In this thesis only the necessary equations, describing the process models and the EIBMC algorithm are given, For the detailed discussion one can refer to the original papers [3, 6, 2, 4].

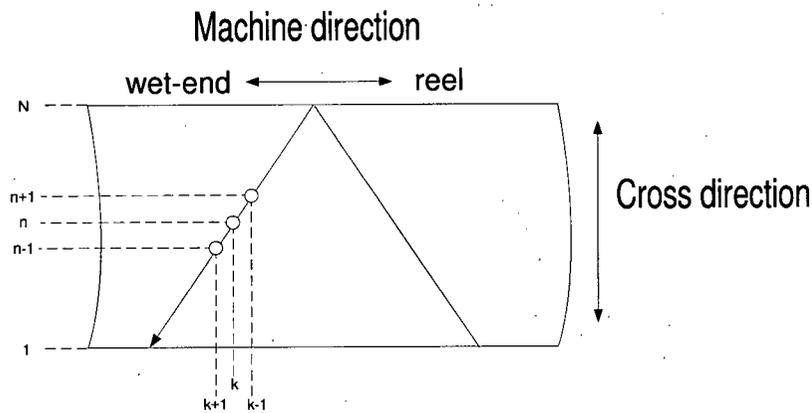


Figure B.1: Data indexing

B.1.1 Moisture

The MVM for dry-end variations is

$$y_k^n = p_k^n + (1 + Bp_k^n)u_k + v_k \quad (\text{B.1})$$

where

n is data box number (CD-position);

k is sample number or time instant (kT);

y_k^n is the difference between measured value at CD position n and time k , and the variable set point;

B is a measure of the non-linearity between the CD and MD variations;

p_k^n is pseudo-profile at position n and time k ;

u_k is MD value at time k ;

$v_k \sim N(0, R)$ is the measurement noise and model mismatch;

It is assumed that $\mathbf{P}_k = \{p_k^1, p_k^2, \dots, p_k^N\}$ and B change slowly [3]. The first order model of the MD variations is given [5] as

$$u_k = \bar{u} + \xi_k \quad (\text{B.2})$$

where

\bar{u} is the mean moisture content in the MD;

ξ_k is a zero mean stochastic process given by:

$$\xi_k = a\xi_{k-1} + w_k, \quad w_k \sim N(0, q) \quad (\text{B.3})$$

The parameter a and the variance q depend on the paper machine.

The model in state space form is given by

$$\begin{aligned} x_{k+1} &= Ax_k + W_k \\ y_k^n &= p_k^n + C_k^n x_k + v_k \end{aligned} \quad (\text{B.4})$$

where

$$\begin{aligned} x_k &= \begin{bmatrix} \bar{u} \\ \xi_k \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}, \quad W_k = \begin{bmatrix} 0 \\ w_k \end{bmatrix} \\ C_k^n &= [1 \quad 1](1 + Bp_k^n) \end{aligned} \quad (\text{B.5})$$

B.1.2 Basis Weight

It has been shown [6] that the previously described moisture model (B.1.1) can be modified to accommodate basis weight estimation by increasing the order of the disturbance (second order ARMA), setting $B=0$ and absorbing the state \bar{u} into the ARMA model. The ARMA model is given as

$$(1 - a_1q^{-1} - a_2q^{-2})u_k = (1 + b_1q^{-1} + b_2q^{-2})w_k \quad (\text{B.6})$$

Thus, the state space model for the basis weight becomes:

$$\begin{aligned} x_{k+1} &= Ax_k + W_k \\ u_k &= Cx_k + w_k \\ y_k^n &= p_k^n + u_k + v_k \end{aligned} \quad (\text{B.7})$$

where

$$A = \begin{bmatrix} a_1 & 1 \\ a_2 & 0 \end{bmatrix}, \quad W_k = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} w_k \quad (\text{B.8})$$

$$C = [1 \quad 0]$$

B.2 The EIBMC Algorithm

B.2.1 Moisture

The moisture version of the EIBMC algorithm is

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} \quad (\text{B.9})$$

$$\hat{u}_{k|k-1} = [1 \quad 1]\hat{x}_{k|k-1} \quad (\text{B.10})$$

$$\psi_k^n = 1 + \hat{B}\hat{u}_{k|k-1} \quad (\text{B.11})$$

$$\hat{y}_{k|k-1}^n = p_{k-1}^n + C_{k-1}^n \hat{x}_{k|k-1} \quad (\text{B.12})$$

$$V^n = \frac{1}{\lambda} V^n - \frac{\alpha(\psi_k^n V^n)^2}{1 + V^n (\psi_k^n)^2} + \beta - \delta(V^n)^2 \quad (\text{B.13})$$

$$\hat{p}_k^n = \hat{p}_{k-1}^n + \frac{\alpha V^n \psi_k^n (y_k^n - \hat{y}_{k|k-1}^n)}{1 + (\psi_k^n)^2 V^n} \quad (\text{B.14})$$

$$C_k^n = (1 + \hat{B}\hat{p}_k^n)[1 \quad 1] \quad (\text{B.15})$$

$$K_k^* = \frac{\Sigma_{k-1}(C_k^n)^T}{C_k^n \Sigma_{k-1} (C_k^n)^T + R} \quad (\text{B.16})$$

$$\Sigma_k = A \left(\Sigma_{k-1} - \frac{\Sigma_{k-1}(C_k^n)^T C_k^n \Sigma_{k-1}}{C_k^n \Sigma_{k-1} (C_k^n)^T + R} \right) A^T + Q \quad (\text{B.17})$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k^* (y_k^n - (\hat{p}_k^n + C_k^n \hat{x}_{k|k-1})) \quad (\text{B.18})$$

$$\hat{u}_{k|k} = [1 \quad 1]\hat{x}_{k|k} \quad (\text{B.19})$$

where

$$Q = E\{W_k W_k^T\} = \begin{bmatrix} q_1 & 0 \\ 0 & q \end{bmatrix} \quad (\text{B.20})$$

The equations are given in the order of their execution.

B.2.2 Basis Weight

The basis weight version of the EIBMC algorithm is

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} \quad (\text{B.21})$$

$$\hat{u}_{k|k-1} = C\hat{x}_{k|k-1} \quad (\text{B.22})$$

$$\hat{y}_{k|k-1}^n = p_{k-1}^n + \hat{u}_{k|k-1} \quad (\text{B.23})$$

$$V^n = \frac{1}{\lambda}V^n - \frac{\alpha(V^n)^2}{1+V^n} + \beta - \delta(V^n)^2 \quad (\text{B.24})$$

$$\hat{p}_k^n = \hat{p}_{k-1}^n + \frac{\alpha V^n (y_k^n - \hat{y}_{k|k-1}^n)}{1+V^n} \quad (\text{B.25})$$

$$K_k^* = \frac{\Sigma_{k-1}C^T}{C\Sigma_{k-1}C^T + R} \quad (\text{B.26})$$

$$\Sigma_k = A \left(\Sigma_{k-1} - \frac{\Sigma_{k-1}C^T C \Sigma_{k-1}}{C\Sigma_{k-1}C^T + R} \right) A^T + Q \quad (\text{B.27})$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k^* (y_k^n - (\hat{p}_k^n + C\hat{x}_{k|k-1})) \quad (\text{B.28})$$

$$\hat{u}_{k|k} = C\hat{x}_{k|k} \quad (\text{B.29})$$

where

$$Q = E\{W_k W_k^T\} = \begin{bmatrix} b_1^2 & b_1 b_2 \\ b_1 b_2 & b_2^2 \end{bmatrix} \sigma_w^2 \quad (\text{B.30})$$

The equations are given in the order of their execution.

References

- [1] E. Dahlin. Computational methods of a dedicated computer system for measurement and control on paper machines. In *24th Engineering Conference, TAPPI*, pages 62.1–62.42, San Francisco, California, Sep 1969.
- [2] G. A. Dumont, M. S. Davies, K. Natarajan, C. Lindeborg, F. Ordubadi, Y. Fu, K. Kristinsson, and A. Jonsson. An improved algorithm for estimating paper machine moisture profiles using scanned data. In *30th IEEE Conference on Decision and Control*, Brighton, England, Dec 1991.
- [3] C. Lindeborg. A process model of moisture variations. *Pulp and Paper Canada*, 4:T142–147, 1993.
- [4] S. T. Morgan. Estimation and identification for machine direction control of basis weight and moisture. Master's thesis, The University of British Columbia, June 1994.
- [5] K. Natarajan, G. A. Dumont, and M. S. Davies. An algorithm for estimating cross and machine direction moisture profiles for paper machines. In *IFAC/FORS Symposium*, pages 27–31, Beijing, PRC, Aug 1988.
- [6] X. G. Wang, G. A. Dumont, and M. S. Davies. Modelling and identification of basis weight variations in paper machines. *IEEE Transactions on Control System Technologies*, June 1993.

Appendix C Wavelet Filter Coefficients

Haar

0.70710678118655
0.70710678118655

Daubechies (4)

0.48296291314453
0.83651630373781
0.22414386804201
-0.12940952255126

Daubechies (6)

0.33267055295008
0.80689150931109
0.45987750211849
-0.13501102001025
-0.08544127388203
0.03522629188571

Symmlet (6)

0.01540410932734
0.00349071208433
-0.11799011114841
-0.04831174258600
0.49105594192764
0.78764114102879
0.33792942172824
-0.07263752278660
-0.02106029251270
0.04472490177075
0.00176771186440
-0.00780070832477

Symmlet (8)

0.00188995033290
-0.00030292051455
-0.01495225833679
0.00380875201406
0.04913717967348
-0.02721902991681
-0.05194583810788
0.36444189483596
0.77718575169975
0.48135965125920
-0.06127335906791
-0.14329423835105
0.00760748732528
0.03169508781035
-0.00054213233164
-0.00338241595136

Coiflet(2)

0.01638733646360
-0.04146493678196
-0.06737255472228
0.38611006682299
0.81272363544940
0.41700518442367
-0.07648859907867
-0.05943441864674
0.02368017194645
0.00561143481942
-0.00182320887071
-0.00072054944537

Coiflet(3)

-0.00379351286449
0.00778259642733
0.02345269614184
-0.06577191128186
-0.06112339000267
0.40517690240961
0.79377722262562
0.42848347637762
-0.07179982161931
-0.08230192710689
0.03455502757306
0.01588054486362
-0.00900797613666
-0.00257451768875
0.00111751877089
0.00046621696011
-0.00007098330314
-0.00003459977284