Applications of continuous wavelet methods in statistical modelling

SangYu Lee

Department of Statistics

University of Leeds

A thesis submitted for the degree of

Doctor of Philosophy

31st March 2021

Acknowledgements

First, I would like to express my sincere gratitude to my supervisors, Dr. Stuart Barber and Dr. Robert G. Aykroyd for providing the wonderful guidance to finish the long journey of Ph.D. They always helped me not to give up my research with kind advice and supports. I would like to thank them for giving me the amazing opportunity to do research and study at the University of Leeds.

Also, I am also grateful to Professor Mi Wang at the School of Chemical and Process Engineering to support research with brilliant tomography data.

Finally, to my family and friends, I would like to express my deep grateful to their support, love and caring. I am grateful to my parents for giving me the opportunity and experiences.

Abstract

The wavelet approach is an efficient time-frequency analysis tool to investigate stochastic data both in terms of time and frequency. Wavelet methods provide a decomposition of a signal using a wavelet function which is localised in both time and frequency. This thesis focuses on wavelet methods using the Haar wavelet function to develop a new approach to statistical modelling and data analysis.

We apply a regression model to the wavelet coefficients to classify the state of a gas-fraction for an engineering tomography dataset. In the previous research of Aykroyd *et al.* (2016), the model based on the wavelet coefficients, from the discrete resolution levels, classified the tomography data well. However, the model is fitted on a limited number of wavelet resolution levels of the Discrete Wavelet Transform (DWT). We expand the scale set to the continuous domain via the continuous wavelet transform (CWT) to see the effectiveness of flexible wavelet scale selection for similar tomography data analysis.

Apart from the modelling using the CWT, the locally stationary wavelet (LSW) method is introduced in Nason *et al.* (2000) as a model to investigate non-stationary process with wavelet functions. We will give an overview of the standard discrete LSW process and suggest an extension of the continuous LSW (CLSW) process to estimate frequency features. The standard LSW process is built on discrete resolution levels, but we will extend the LSW process to the continuous wavelet scales. However, due to the redundancy of the CWT, the estimation of the evolutionary wavelet spectrum (EWS) does not show

spectral densities matching true frequency characteristics. To cope with the problem, we apply the idea of the orthogonal matching pursuit (OMP) algorithm to select the best subset of continuous wavelet scales explaining data the most. We illustrate the modified CLSW process using the reflected doppler data and real tomography data and show the spectral estimate of them. Based on the improvement of spectral estimation, we fit a classification model on the estimate of EWS from the modified CLSW process using the real tomography data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Time-frequency analysis is an established area of interest in Statistics and Physics to understand the structure of data over both time and frequency. In particular, since data in signal processing usually has complicated structures, formed from various frequencies, time-frequency analysis can propose an approach to see the data over both domains, time and frequency. For example, the Fourier transform decomposes data over time by the combination of sine and cosine functions. The periodogram given by the decomposition of the trigonometric functions enables us to understand spectral properties of data. Like the Fourier transform, wavelet analysis, the main focus of the thesis, is another time-frequency analysis tool. The wavelet transform decomposes data using a wavelet function localised both in time and by wavelet resolution level or scale. The details of the wavelet method will be described in §2.2.

This thesis will develop statistical modelling based on continuous wavelet methods. The main objective of the research is to show the necessity of continuous scales in wavelet methods with applications. The discrete wavelet method restricts the number of wavelet resolution levels by the number of observations. Hence, the information based on the discrete wavelet transform is given by a few discrete wavelet resolution levels, which may lose valuable frequency information between the discrete levels. Meanwhile, the continuous wavelet transform allows us to include continuous scales which are not restricted by the length of the data.

Chapters 3 and 4 consider classification modelling using a subset of scales from the continuous wavelet transform and compare them to see how the intermediate scales, between discrete wavelet resolution levels, affect the performance of modelling. The data mainly covered here is tomography data where the aim is to see the state of the gas fraction in a liquid. The fitted model classifies the gas-liquid phase by the size of bubbles in liquid based on electrical conductivity data.

Aykroyd *et al.* (2016) fitted a classification model based on the coefficients from the discrete wavelet transform using the vertical flow tomography data. Chapter 3 fits a regression model to the same vertical flow tomography data from Aykroyd *et al.* (2016) and Chapter 4 is with the horizontal flow tomography data. The fitted models, based on the discrete wavelet coefficients of the electrical conductivity data, proposed the important levels to classify the state of gas-liquid phase. However, as the models are fitted from the limited discrete wavelet levels, we can consider expanding the variable domain to the continuous wavelet scales applying this to the vertical and horizontal flow tomography data.

The continuous wavelet transform (CWT) is obtained as an integral of continuous functions, but the continuous calculation is difficult to implement in practical computing. Also, most practical data are given as formations of vectors or matrices rather than continuous functions. Accordingly, we assume that the continuous calculation for the CWT are carried out as the discrete sum using a fine grid.

To allow fitting models using continuous wavelet methods, the locally stationary wavelet (LSW) process is developed based on the continuous wavelet method. The locally stationary wavelet process is proposed in Nason *et al.* (2000) as a tool for a specific type of non-stationary process. This method enables us to make a representation of data from the non-decimated wavelet transform and estimate spectral properties changing over time. The standard LSW process is based on the non-decimated discrete wavelet transform, but the intermediate levels between discrete levels may have some valuable time-frequency characteristics. Hence, we aim to expand the LSW process to the continuous wavelet method. Chapter 5 covers the definition of the LSW process and its application with example data and suggests a model built on the continuous wavelet scales. Since

the CLSW process is built based on a more redundant wavelet scale set than the existing LSW process, the estimated spectrum may not be close enough to the true spectral density. Therefore, there are more mathematical properties which need to be checked. To simplify this mathematical problem, we will consider a discretised continuous wavelet transform with sampled wavelet scales from a continuous wavelet scale set. This simpler approach can suggest a path towards understanding the fundamental mathematical assumptions and properties needed for our ultimate goal, the continuous LSW process. Along with the mathematical proof, we will show its application on real data to compare how it differs from the result of a discrete LSW process.

However, the CLSW process may have a distortion effect due to the redundant scale set from the discretised CWT. This thesis suggests the modified CLSW process through scale reduction by using the orthogonal matching pursuit (OMP) algorithm in Chapter 6 with the improvement of spectra analysis using example datasets. Based on the development of the modified CLSW process, Chapter 6 fit a classification model using the example dataset of vertical flow tomography data.

Before moving Chapter 2, we will review some background that will be used later in this thesis.

1.2 Preliminaries

1.2.1 Stationary Process

Stationarity is a fundamental concept in time series analysis. The stationarity is the statistical property with some assumptions of statistical behaviours not changing over time. For the degree to which statistical properties are assumed, the stationarity can be defined as ***strictly stationary*** or ***weakly stationary*** processes. First, the ***strictly stationary process*** is defined as a sequence of random variables, $\{x_{t_1}, \ldots, x_{t_n}\}$ which have identical probabilistic behaviour to the set $\{x_{t_1+k}, \ldots, x_{t_n+k}\}$ for any integer $k \leq 0$ and all the time points, t_i for

$i > 1$. This regularity for the probability in strict stationarity means

$$\mathrm{P}(x_{t_1} \leq c_1, \ldots, x_{t_n} \leq c_n) = \mathrm{P}(x_{t_1+k} \leq c_1, \ldots, x_{t_n+k} \leq c_n) \tag{1.2.1}$$

for any c_i for $i = 1, \ldots, n$.

However, stationarity in terms of probability is too strong to apply to practical time series analysis. To ease this strong regularity, the ***weakly stationary process***, x_t for time, t, and a lag, k, is defined in terms of mean and auto-covariance such that

(i) $\mathrm{E}(x_t) = \mu$

(ii) $\mathrm{Cov}(x_t, x_{t+k}) = \gamma_k$

where μ and γ_k are constants not depending on t for $k \in \mathbb{Z}$. The conditions mean that the first and second moments are independent with time, t and finite, but depend on a lag, k. In Statistics, a ***stationary process*** generally means a weakly stationary process.

1.2.2 Bayesian Information Criterion

The ***Bayesian Information Criterion (BIC)*** is a criteria often used to choose an optimal model which considers the number of parameters. The criteria is defined based on a likelihood for the model with k parameters such that

$$BIC = k \log n - 2 \log(\hat{L}) \tag{1.2.2}$$

where n is the number of observations and $\hat{L}$ is the maximized likelihood of the data, $\{x_1, \ldots, x_n\}$. The model minimising the criteria can be considered as the best model. Also, the number of observations, n, is also considered as a penalty to avoid overfitting.

If we assume that time series data, $\{x_1, \ldots, x_n\}$ are from a normal distribution, we can simplify the notation of **BIC** as

$$BIC = n \log \frac{\mathrm{RSS}}{n} + k \log n \tag{1.2.3}$$

where RSS is the residual sum of squares using the same notation as in (1.2.2).

1.2.3 Fourier Transform

Fourier transform (FT) is a time-frequency analysis tool commonly used to investigate frequency characteristics. Here, we will briefly explain the definition of the Fourier transform (FT) and its inverse function. The following explanations are mainly from Dyke (2014).

FT decomposes data into periodic components based on trigonometric functions. Assume that $f(x)$ is an integrable function for all $x \in \mathbb{R}$. Then, the FT is defined by

$$\hat{F}(w) = \int_{-\infty}^{+\infty} f(x) e^{-2\pi i x w} dx \tag{1.2.4}$$

where w is frequency. The FT, $\hat{F}(w)$ can be considered as a complex-valued function of $w \in \mathbb{R}$,

$$\hat{F}(w) = A(w) e^{i\psi(w)}. \tag{1.2.5}$$

The spectral information of $f(x)$ can be interpreted in terms of the coefficients of the FT. The squared coefficient, $|F(w)|^2$, is called the *spectral density*. This quantity suggests how the data, $f(x)$, is decomposed into frequencies. The total sum of the spectral density is connected to the energy of the data. The relationship is called *Parseval's theorem*.

Theorem 1.2.1. ***Parseval's Theorem***
If $f(x) \in L^2(\mathbb{R})$ has the FT, $F(w)$, then

$$\int_{-\infty}^{+\infty} |f(x)| dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{F}(w)| dw. \tag{1.2.6}$$

Discrete Fourier Transform

The standard FT is defined with a continuous function, $f(x)$, but we can perform the FT with a discrete stochastic process or sampled data of a real function as well. Let $X = (X_0, X_1, \ldots, X_{n-1})$ be a finite sequence from a discrete stochastic

process. Then, the discrete Fourier Transform (DFT) of X_t is defined by

$$\hat{X}_j = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} X_t e^{(-2\pi i f_j t)} \qquad j = 0, \ldots, n-1 \tag{1.2.7}$$

where f_j is the Fourier frequency, j/n.

The Inverse Fourier Transform

The FT coefficients of the standard FT and the DFT can be used to reconstruct the original data. The inverse can be shown as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{F}(w) e^{iwx} dw \tag{1.2.8}$$

for the standard FT. The process is called the inverse Fourier transform. The discrete FT also has the inverse Fourier transform,

$$X_t = \frac{1}{\sqrt{n}} \sum_{j} \hat{X}_{j=0}^{n-1} e^{2\pi i f_j t} \tag{1.2.9}$$

for $f_j = j/n$.

1.2.4 Inner Product

The inner product is a function of a pair of vectors. Let $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ be vectors in a vector space. Then, the inner product can be written as

$$< x, y >= \sum_{i=1}^{n} x_i y_i. \tag{1.2.10}$$

1.2.5 Entropy (Shannon's Entropy)

Entropy is the concept used in information theory to measure the amount of uncertainty in random variables. There are various formats or definitions of the entropy, but in general and particularly in Statistics, the entropy means *Shannon's Entropy*. Let x be a random variable from a distribution, P_x. Then, the entropy

of x is

$$\begin{aligned} H(x) &= \mathrm{E}\left(-\log P_x\right) \\ &= \sum_{x=x_i} -P(x_i)\log P_x(x_i) \end{aligned} \tag{1.2.11}$$

where $P_x(x_i)$ is the probability at $x = x_i$.

1.2.6 Big O notation

Big O notation is a notation to symbolise the asymptotic characteristics of a function, $f(x)$ as x goes to a limit of a particular value or infinity. This describes the rate of changes of the function, $f(x)$ in terms of the order of $g(x)$. We can define the Big O notation, $O(g(x))$ for the function, $f(x)$, as $x \to \infty$, if and only if there exists a constant c and x_n such that

$$|f(x)| \leq c|g(x)| \tag{1.2.12}$$

for all $x > x_n$. We can denote the definition as $f(x) = O(g(x))$ as $x \to \infty$.

Chapter 2

Introduction to Wavelets

2.1 Motivation

The purpose of time-frequency analysis is to investigate the structure of data over both time and frequency simultaneously. In signal processing, various datasets have complicated structures, both in terms of time and frequency. However, describing a signal as a function of time is localised in time but not frequency. Hence, this two-dimensional analysis enables us to expand our analyses of data into these two domains, time and frequency together rather than as separate analyses.

As mentioned earlier in §1.2.3, the Fourier transform (FT) decomposes data in terms of trigonometric functions. The standard FT defines the representation of data as the sum or integral of the periodic sine and cosine functions. While these functions used in the FT are localised in frequency, they cannot make a localised representation in time. Meanwhile, wavelet functions are known to be relatively well localised in both time and frequency compared to the standard trigonometric functions. Here, our main analysis tool is the wavelet transform to see the local time-frequency structure of data.

Wavelet methods have been developed to investigate time-frequency characteristics of stochastic data along wcith various analysing methods such as Fourier transform. In our thesis, we focus on wavelet methods using wavelet functions as basis functions rather than other methods. The wavelet transform can be divided into discrete and continuous wavelet transformations depending on the type of

data. The discrete wavelet transform is constructed over a limited number of discrete resolution levels, which relies on the number of data points in order to keep the relationship between levels. Meanwhile, the continuous wavelet transform allows us to include more flexible wavelet frequencies from the continuous real number domain. We keep both methods to compare the efficacy of statistical modelling for our real tomography data. However, real data is usually recorded as discrete sequence data instead of a continuous function, and therefore we assume that continuous wavelet transform is also calculated by a discrete summation.

First, we will introduce the basic idea of wavelets in §2.2 with a literature review. Various different forms of wavelet transforms can be used to decompose the data in terms of wavelet functions. Two particular transforms are the discrete wavelet transform (DWT) and the continuous wavelet transform (CWT), which break the signals down in to various wavelet frequencies used in the transformations. This chapter will describe a brief explanation of the DWT and CWT, illustrated by some example data for the development of statistical modelling for the next step. Most of explanations covered in this chapter are mainly based on Daubechies (1992), Nason (2010) and Vidakovic (1999).

2.2 What is a Wavelet?

The term, "wavelet" refers to the concept of a "small wave". The basis functions used in wavelet methods generally have small oscillations decaying to zero from the middle. There are various wavelet functions with different shapes such as the Haar function, the Morlet wavelet function and Shannon's wavelet function. The first function described in this thesis is the Haar wavelet function introduced in Haar (1911). In wavelet analysis, the discrete wavelet transform has the two different types of wavelet functions: mother wavelet functions and father wavelet functions. The detail of the notations will be covered in §2.3 with the discrete wavelet transform. The father Haar wavelet function, $\phi(x)$, and mother Haar

wavelet function, $\psi(x)$, are

$$\phi(x) = \begin{cases} 1 & \text{when } x \in (0,1] \\ 0 & \text{otherwise,} \end{cases} \qquad \psi(x) = \begin{cases} 1 & \text{when } x \in (0,1/2] \\ -1 & \text{when } x \in (1/2,1] \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.1}$$

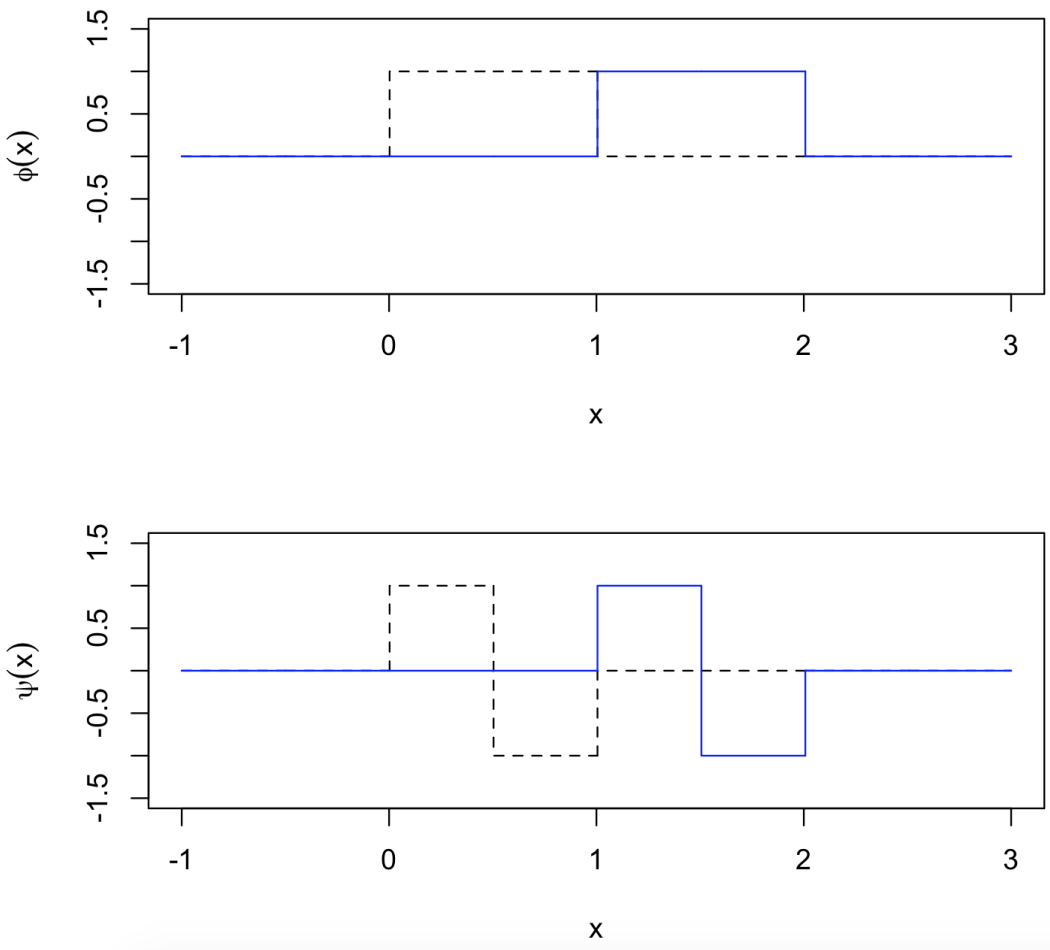


Figure 2.2.1: Haar wavelet functions: the father wavelet function, $\phi(x)$ (top) and the mother wavelet function, $\psi(x)$ (bottom). The dashed lines are the default wavelet function at the origin and the blue solid lines are the wavelet functions shifted to +1 units to the right.

Figure 2.2.1 shows line plots of the Haar wavelet functions defined in (2.2.1). They are square-shaped functions with discrete points and the width of non-zero

areas depends on the wavelet levels or scales. Although these functions are not continuous, they are advantageous for the development of modelling and computational techniques due to their simplicity. In addition to the Haar wavelet function, Daubechies wavelet functions have been highlighted in the discrete wavelet transform. The Daubechies' wavelet functions are compactly supported, which gives a limited period of small oscillations.

With these wavelet functions, wavelet methods are used to investigate the time-frequency characteristics in the domains of time and frequency simultaneously. The wavelet coefficients are transformed by projecting the data onto the corresponding wavelet function, which gives the contribution of that wavelet function to the data. The detail of wavelet transform and its application will be described in §2.3.

2.2.1 Literature Review

As mentioned previously, wavelet methods have been developed as a useful tool which operates simultaneously in both time and frequency domains. There are several books and papers on the subject. Daubechies (1990) describes two time-frequency analysis methods: the windowed Fourier transform and the wavelet transform. In both transformations, the data can be reconstructed stably with tight frames and this enables us to investigate local time-frequency behaviours by basis functions well localised in time as well. Also, the paper shows that the discrete wavelet transform with logarithmic scales can be a better way to analyse time-dependent signals such as acoustic data than the windowed Fourier transformation. Jawerth & Sweldens (1994) give an overview of the multiresolution analysis based on the space spanned by the discrete wavelet functions and the fast wavelet algorithm to decompose or reconstruct data by the relationship between levels. Daubechies (1992); Nason (2010); Vidakovic (1999) and Percival & Walden (2000) all explain the basic principle of general wavelet methods with mathematical accounts and computing techniques. Mallat (2008) also explains wavelet methods focusing on sparse representation. The book suggests the Matching Pursuit algorithm to select a sparse representation from a redundant set of basis functions. The algorithm has an important role for developing the

continuous locally stationary wavelet process which we propose in Chapter 5, so details will be explained in Chapter 6.

Wavelet methods have been used to explore data having time-varying complicated structure. The wavelet transformation decomposes data into corresponding wavelet functions so that the wavelet coefficients represent the contribution to the data at the resolution level. The transformation can often efficiently compress signals into a smaller number of wavelet coefficients enabling a sparse representation of many data sets. This characteristic of wavelet transform can be used for signal processing. For example, complex biological data such as electroencephalographic (EEG) and electromyographic (EMG) data has transient changes over high-time resolution Rafiee *et al.*, 2011. Therefore, wavelet methods, localised in both time and frequency, can be a useful tool to estimate their frequency behaviour.

Also, since most practical data for statistical modelling and signal processing has at least some noise present, we often need to denoise data to estimate the true time-frequency characteristics. Wavelet methods have been used as an effective tool to handle the noise problem. Antoniadis (2007) gives an outline of the general denoising algorithm using wavelet methods and some developments of the idea. Distributional assumptions on the noise will propagate through the transformation, and therefore the choice of denoising algorithm which modifies the noise in the wavelet coefficients are usually by a given thresholding or shrinkage rule. The denoised wavelet coefficients are used to reconstruct a denoised estimate of the true noiseless data by applying the inverse transform. Donoho *et al.* (1995) suggest the idea of curve estimation via the wavelet shrinkage and Nason (1996) applies the idea of cross-validation to choose the threshold. Also, the Bayesian approach has been used for wavelet shrinkage; Chipman *et al.* (1997) and Johnstone & Silverman (2005) among others use mixture distributions, to model the prior assumptions of sparsity in the wavelet coefficients.

2.3 Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) and the continuous wavelet transform (CWT) are distinguished by the domain of wavelet resolution levels or scales used in the transformations. First, the discrete wavelet transform is defined based on

the discrete wavelet resolution levels determined by the number of observations. Let the number of observations be N, and we assume that the length of data has the dyadic relationship, $N = 2^J$. The discrete wavelet resolution levels are given as $j = 0, \ldots, J-1$ and let the shifting parameter be k.

In discrete wavelet methods, there are the two types of wavelet functions: the mother wavelet functions and the father wavelet functions. Let the father wavelet function and the mother wavelet function be $\phi(x)$ and $\psi(x)$, respectively. The father wavelet function notated, $\phi(x)$, explains a local averaging and the mother wavelet function extract more detail information such as variation at the given wavelet level.

These wavelet functions are scaled by the wavelet resolution level, j and translated by a shift, k following the definitions,

$$\phi_{jk}(x) = 2^{\frac{j}{2}}\phi(2^j x - k) \quad \text{and} \quad \psi_{jk}(x) = 2^{\frac{j}{2}}\psi(2^j x - k), \tag{2.3.1}$$

for $j = 0, \ldots, J-1$ and $k = 0, \ldots, 2^j - 1$.

The wavelet resolution level, j, scales the wavelet function by the normalising factor, $2^{\frac{j}{2}}$ so that "energy" is conserved for wavelet functions over different levels, j, and shifting parameters, k. We can define the father wavelet coefficient, $c_{j,k}$ and the mother wavelet coefficient, $d_{j,k}$ of data, $f(x)$ to be

$$c_{jk} = \sum_x f(x)\phi_{j,k}(x) \quad \text{and} \quad d_{jk} = \sum_x f(x)\psi_{j,k}(x). \tag{2.3.2}$$

2.3.1 Multiresolution Analysis (MRA)

The discrete wavelet transform is carried out over multiple resolution levels, $j = 0, \ldots, J-1$. This multiscale method can be used to extract a smooth average from a low resolution level and to add more details from finer resolution levels. Based on this approach, Multi Resolution Analysis (MRA) provides a theoretical concept to understand the wavelet bases from scale, $j \in \mathbb{Z}$. Let V_j be the space spanned by the father wavelet function, $\{\phi_{jk}(x)\}_k$, and W_j be the space spanned by the mother wavelet function, $\{\psi_{jk}(x)\}_k$, at the wavelet resolution level, j.

Then, we can denote the spaces, V_j and W_j as

$$V_j = \{f \in \mathbb{L}^2(\mathbb{R}) | \, f(x) = \sum_k c_k \phi_{jk}(x)\} \tag{2.3.3}$$

$$W_j = \{f \in \mathbb{L}^2(\mathbb{R}) | \, f(x) = \sum_k d_k \psi_{jk}(x)\} \tag{2.3.4}$$

respectively (see Vidakovic, 1999, chap. 3).

The relationship of two different resolutions can be written as $V_j \subset V_l$ for $j < l$ so that the space based on the finer level of resolution has more details than the other. According to the relationship over multiple scales, we can write

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots . \tag{2.3.5}$$

We assume that a wavelet function, $\phi(x)$ is a orthonormal basis for V_0 so that the space, V_0, is spanned by the wavelet functions, $\{\phi(x-k)\}_k$. Therefore, by the linkage relationship of (2.3.5), $\phi(x)$ can be represented as

$$\begin{aligned} \phi(x) &= \sum_k h_k \phi_{1k}(x) \\ &= \sum_k h_k \sqrt{2} \phi(2x-k), \end{aligned} \tag{2.3.6}$$

for some $\{h_k\}$. That implies that the wavelet function, $\phi_{j,k}(x)$ is a orthonormal basis function of V_j.

Meanwhile, the space, W_j is based on the detail information not explained by the local average from $\phi_{j+1,k}(x)$. The relationship can be represented in terms of the difference between spaces as $W_j = V_{j+1} \ominus V_j$ (Daubechies, 1992; Vidakovic, 1999). Here, $\psi_{j,k}(x)$ is also an orthonormal wavelet function of W_j. This allows us to make the representation of data as

$$f(x) = \sum_k c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_k c_{j,k} \psi_{j,k}(x). \tag{2.3.7}$$

The wavelet resolution level, j_0, is often called the "primary resolution level". The representation consists of the local average part at the wavelet resolution

level, j_0 and the detailed information accumulated from the level, j_0 to infinity. In the same sense as $\phi(x)$, the mother wavelet function, $\psi_{j,k}(x)$ can be written as

$$\begin{aligned}\psi(x) &= \sum_k g_k \phi_{1k}(x) \\ &= \sum_k g_k \sqrt{2}\phi(2x-k),\end{aligned} \tag{2.3.8}$$

for some $\{g_k\}$ due to the linkage of spaces. The dilation is controlled by the coefficients, $\{h_k\}$ and $\{g_k\}$, depending on the kind of a wavelet function.

The relationships of (2.3.6) and (2.3.8) can be extened in Fourier domain. Let $m_0(w) = \frac{1}{\sqrt{2}}\sum_k h_k e^{-ikw}$ and $m_1(w) = \frac{1}{\sqrt{2}}\sum_k g_k e^{-ikw}$. Then, the Fourier transform of $\phi(x)$ and $\psi(x)$ become

$$\hat{\Phi}(w) = m_0\left(\frac{w}{2}\right)\Phi\left(\frac{w}{2}\right) \quad \text{and} \quad \hat{\Psi}(w) = m_1\left(\frac{w}{2}\right)\Phi\left(\frac{w}{2}\right) \tag{2.3.9}$$

where $\hat{\Phi}(w)$ and $\hat{\Psi}(w)$ are the Fourier transformation of $\phi(x)$ and $\psi(x)$. Here, the functions, $m_0(w)$ and $m_1(w)$ can be interpreted as the characteristic functions of h_k and g_k in the Fourier domain.

2.3.2 Non-decimated Wavelet Transform (NDWT)

Decimation

In §2.3.1, the linkage relationship between the discrete wavelet resolution levels is explained. Equation (2.3.6) can be extended to relationship between the level, j and $j+1$ as follows,

$$\begin{aligned}\phi_{j,k}(x) &= 2^{j/2}\phi(2^j x - k) \\ &= 2^{j/2}\sum_n h_n \phi_{1n}(2^j x - k) \\ &= 2^{j/2}\sum_n h_n \sqrt{2}\phi(2(2^j x - k) - n) \\ &= 2^{j+1/2}\sum_n h_n \phi(2^{j+1}x - (2k+n))\end{aligned}$$

$$= \sum_n h_n \phi_{j+1,2k+n}(x), \tag{2.3.10}$$

where k is the shift parameter. By the same principle, equation (2.3.8) can be generalised as

$$\psi_{j,k}(x) = \sum_n g_n \phi_{j+1,2k+n}(x). \tag{2.3.11}$$

Based on the generalised linkage relationship, the discrete wavelet coefficients can be obtained as

$$c_{j,k} = \sum_n h_n c_{j+1,2k+n} \quad \text{and} \quad d_{j,k} = \sum_n g_n c_{j+1,2k+n} \tag{2.3.12}$$

by applying (2.3.2) to (2.3.10) and (2.3.11). The above equations mean that the wavelet coefficients at the level, j, can be given from the information at the next finer level, $j+1$, with the filter coefficients, $\{h_n\}_n$ and $\{g_n\}$. Also, the wavelet function and coefficients, $\psi_{j,k}(x)$, $c_{j,k}$ and $d_{j,k}$, are derived from the wavelet values indexed by $2k+n$ at the next finest resolution level, $j+1$. This means that a single shift, $(k \to k+1)$, at level, j corresponds to two shifts at level, $j+1$. Nason & Silverman (1995) denote the even decimation operator, D_0, as

$$(D_0 x)_m = x_{2m}. \tag{2.3.13}$$

In the same spirit, the odd decimation to choose odd elements, D_1, is denoted as

$$(D_1 x)_m = x_{2m+1}. \tag{2.3.14}$$

We will follow these terms in this thesis with the wavelet filter operators, $\mathcal{H}$ and $\mathcal{G}$ defined over $\{h_n\}$ and $\{g_n\}$. Following this terminology, we can write the vectors of the wavelet coefficients at level j, c_j and d_j, as

$$c_j = D_0 \mathcal{H} c_{j+1} \qquad\qquad d_j = D_0 \mathcal{G} c_{j+1} \tag{2.3.15}$$

by equation (2.3.12). Nason & Silverman (1995) wrote the standard DWT as

$$c_j = D_0\mathcal{H}^{J-j}c_J \qquad\qquad d_j = D_0\mathcal{G}D_0\mathcal{H}^{J-j-1}c_J \tag{2.3.16}$$

in terms of the father wavelet coefficient vector, c_J at the finest level. Only the standard DWT through decimation gives an orthogonal transformation, resulting in a single basis set.

Application of the DWT to the doppler test function

Following the concept of the DWT and NDWT, we will apply the transformation to the doppler test function. The doppler test function is a example data set commonly used in signal processing including wavelet methods. As this data has time-depending frequency behaviours, it is useful to test a time-frequency analysing tool. Figure 2.3.1 shows the line plot of the doppler test function. The doppler test function tends to get slower, but the amplitude of the wave is increasing gradually.

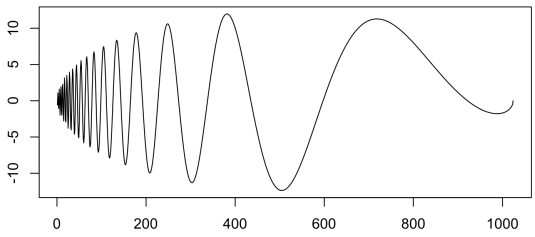


Figure 2.3.1: The line plot of the doppler test function.

Figure 2.3.2 displays the plots of wavelet coefficients from the standard DWT and non-decimated DWT using the doppler test function. Recall that the standard wavelet function depends on the origin and the number of wavelet functions and their coefficients are different at each resolution level by dyadic decimation in (2.3.12). As can be seen from Figure 2.3.2, the number of wavelet coefficients are different between the transformations. Although the standard DWT is useful to reconstruct the original data based on the coefficients, the non-decimated DWT is easier to understand the patterns of the frequency characteristics of the doppler function with the dense coefficients.

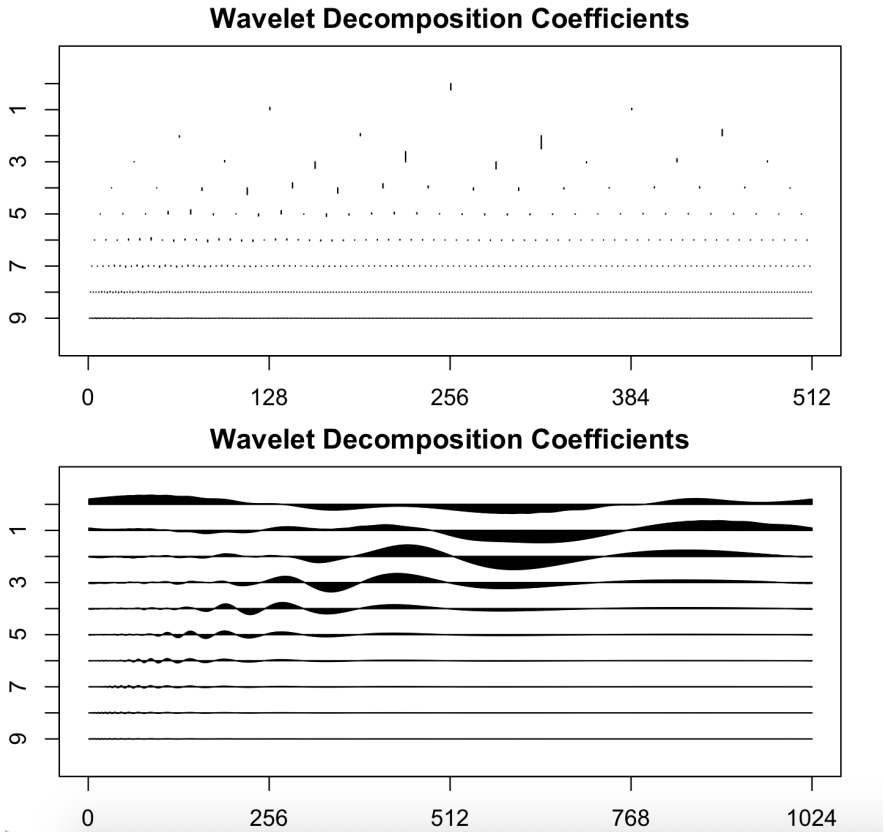


Figure 2.3.2: The plots of the wavelet coefficients from the standard DWT and non-decimated DWT using the doppler test function: DWT (upper) and NDWT (below).

Non-decimated Wavelet Transform (NDWT)

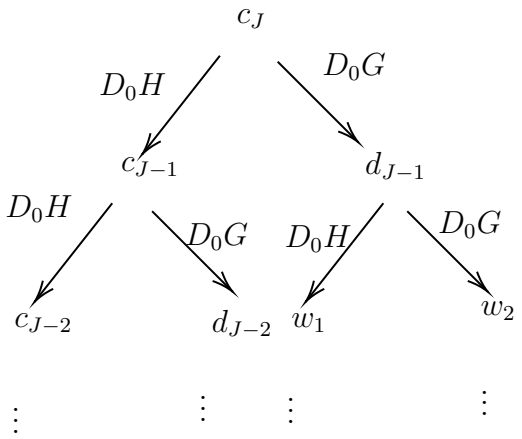
The standard discrete wavelet transform is defined using the binary decimation operator of (2.3.15). This decimation operator, D_0, chooses every even wavelet coefficients at the level, $j+1$, as the origin to obtain the wavelet coefficients at the next coarsest level, j. As the decimated wavelet transform requires to use only one decimation operator, the other wavelet coefficients, decimated by D_1, are not considered for the standard discrete wavelet transform. Recall that a single shift at a specific level, j, does not mean a single shift at the other levels in the standard DWT. That implies that the decimated wavelet transform can be different depending on an origin point.

The standard DWT has advantages in that the wavelet coefficients conserves

the energy of original data after the transformation with the orthogonal structure of wavelet basis functions. However, we may wish to retain all information to get an analysis which is not sensitive to the origin point. The non-decimated wavelet transform (NDWT) is based on this motivation. The NDWT includes all the wavelet coefficients and does not depend on the origin of data. The wavelet coefficients of NDWT at each resolution level keep both $D_0\mathcal{H}c_{j+1}$ and $D_1\mathcal{H}c_{j+1}$ for the level, j. Hence, the number of wavelet coefficients of the NDWT is always the same as the length of the data, regardless of the level of resolution. The computation cost is more expensive than the standard DWT, but the result of wavelet transform is translation-invariant and free from the origin of data. However, the NDWT is not an orthogonal transformation, and hence inverting the NDWT is more complicated than inverting the standard DWT. Also, we assume that the NDWT has a periodic boundary in our research.

2.3.3 Wavelet Packet Transform

In (2.3.16), we described the standard discrete wavelet transform in terms of the even decimation operator, D_0 and filter operators, $\mathcal{H}$ and $\mathcal{G}$. Although the wavelet coefficients depend on the type of operator filters, both of them are given by iterating $\mathcal{H}$. On the other hand, the wavelet packet transform includes all cases applying either $\mathcal{H}$ or $\mathcal{G}$ at each step.



Then, the wavelet packet functions are organised into the wavelet packet basis library. We can select the "best" basis functions from the wavelet packet library

by using cost functions such as Shannon's entropy. In the same spirit, the set of non-decimated wavelet functions is also considered as a wavelet basis library. Averaging or selecting basis functions can be used for the non-decimated wavelet transform as well, but we will not cover the details in the thesis.

Non-decimated Wavelet Packet Transform

The previously described wavelet packet functions are given by using the even decimation operator, D_0, equivalent to the standard discrete wavelet transform. This wavelet packet functions can be generalised by using both decimation operators, D_0 and D_1 with filters, $\mathcal{H}$ and $\mathcal{G}$. Therefore, the non-decimated wavelet packet transform applies the four combined operators, $D_0\mathcal{H}$, $D_0\mathcal{G}$, $D_1\mathcal{H}$ and $D_1\mathcal{G}$ to each packet at the one level before. The non-decimated wavelet packet transform also does not depend on origin, so are invariant to shift parameters.

2.4 Continuous Wavelet Transform

As we previously mentioned, the DWT is defined over a limited number of discrete wavelet resolution levels. However, the continuous wavelet transform allows us to choose wavelet scales more flexibly from a continuous domain. In this section, we will give an overview of continuous wavelet transform (CWT) with some mathematical properties which is needed in this thesis.

Wavelet functions described in § 2.3 are built based on the discrete wavelet resolution level set, $\boldsymbol{J} = \{0, 1, \ldots, J-1\}$ determined by the number of observations. Meanwhile, the continuous wavelet function, $\psi_{ab}(x)$ is defined as

$$\psi_{ab}(x) = \frac{1}{\sqrt{a}}\,\psi\Big(\frac{x-b}{a}\Big) \tag{2.4.1}$$

for the wavelet scale, $a \in \mathbb{R}^+$ and the shift, $b \in [0, T)$. Equivalently to the DWT, the continuous wavelet transform (CWT) of data, $f(x)$, is given as

$$CWT_f(a,b) = \int f(x)\psi_{ab}(x)dx = < f, \psi_{ab} > . \tag{2.4.2}$$

Although the CWT has a similar structure to the DWT, the selection of wavelet frequencies is more flexible and the domain of wavelet frequencies no longer depends on the length of data. Also, the CWT does not have a local averaging part which is represented by father wavelet functions. In general, the wavelet transformation of CWT means the decomposition of data by $\psi(x)$, which is a mother wavelet function of DWT.

The wavelet function of CWT is assumed to belong to $\boldsymbol{L}^2(\mathbb{R})$. Moreover, there is a fundamental condition for the CWT, *"admissibility condition"*, required for a continuous wavelet function. The admissibility condition is

$$C_\psi = 2\pi \int \frac{|\hat{\psi}(w)|^2}{|w|} dw < \infty, \tag{2.4.3}$$

where C_ψ is the *"resolution of identity"* and $\hat{\psi}(w)$ is the Fourier transform of $\psi(x)$. The condition implies $\int \psi(x)dx = 0$.

As previously mentioned in §2.3.2, the decimated DWT can provide an orthogonal transformation. On the contrary, it is hard to reconstruct original data from the wavelet coefficients of the CWT due to the redundant wavelet scale set. The resolution of identity, C_ψ, can explain the relationship between the coefficients and data as follows

$$f(x) = \frac{1}{C_\psi} \int \int CWT_f(a,b)\psi_{a,b}(x) \frac{da\, db}{a^2}. \tag{2.4.4}$$

The inverse relationship requires to satisfy the admissibility condition, (2.4.3) first. If we can assume the data, $f(x) \in \boldsymbol{L}^2(\mathbb{R})$, the squared norm of f(x) can converges as

$$\int |f(x)|^2 dx = \frac{1}{C_\psi} \int |CWT_f(a,b)|^2 \frac{da\, db}{a^2}. \tag{2.4.5}$$

The CWT is calculated by an integral, and accordingly we assume that data is continuous for the CWT. However, as practical data is generally given as a discrete sequence of data, it is difficult to evaluate a continuous wavelet transformation using real data. The data sets used in this thesis are also given as vectors of discrete observations, and therefore we need to get an approximation to the wavelet coefficients of the CWT. Here, we use a fine grid from the continuous

wavelet scale domain to get an wavelet coefficient. The approximated CWT can be denoted as

$$CWT_f(a,b) = \sum_x f(x)\psi_{ab}(x), \tag{2.4.6}$$

for the continuous wavelet scale, $a \in A_D$, where A_D is the discretised continuous wavelet scale set. To avoid confusion in terminology, although the formed CWT of (2.4.2) is defined as an integral, the CWT in this thesis means the discretised CWT in (2.4.6). The discretised continuous scale set, A_D, can be changed depending on the number of observations or a kind of wavelet function, so it will be defined individually at each case. Although the approximation to the wavelet coefficients of the CWT enables us to implement the transformation in practical computing, the inverse relationship, (2.4.4), is no longer satisfied. We also assume that the CWT is periodic in our research.

2.5 Frame

A ***frame*** is first introduced in Duffin & Schaeffer (1952) with a family of complex exponential functions.

Definition 2.5.1. *If all functions, $g_k(x)$ in a Hilbert space, $\mathbb{H}$ satisfy*

$$A||f||^2 \leq \sum_{k\in K} |\langle f, g_k\rangle|^2 \leq B||f||^2 \tag{2.5.1}$$

for $A \leq 0$ and $B < \infty$, a family of functions, $\{g_k\}_{k\in K} \in \mathbb{H}$ is a frame.

A and B are frame bounds, which control the tightness of a frame. We can apply the concept of the frame to the CWT and the discretised CWT which will be used in this thesis. We need a few extra conditions to set a frame bound with the CWT. Suppose that we restrict the range of the continuous wavelet scales, a between 0 and 1 and that the fine grids of continuous wavelet scales are regularly spaced over the range of a. Then, we have

$$c_a \sum_{a,b} |CWT_f(a,b)|^2 \approx \int |CWT_f(a,b)|^2 da\, db$$

$$\leq \int |CWT_f(a,b)|^2 \frac{da\, db}{a^2} = C_\psi ||f||^2, \tag{2.5.2}$$

where c_a is the distance between successive wavelet scales, a's, of the discretised CWT. Recall that the resolution of identity, C_ψ, is finite due to the admissibility condition. Then, we can re-write equation (2.5.2) as

$$0 \leq \sum_k |CWT_f(a,b)|^2 \leq \frac{C_\psi}{c_a} ||f||^2. \tag{2.5.3}$$

Therefore, we can claim that the discretised CWT is also a frame and the bound of the frame for the discretised CWT depends on the size of c_a and the resolution of identity, C_ψ. However, at the moment, since there is no details about the lower frame bound, we cannot guarantee how narrow the bound can be.

2.6 Comparison to Fourier Transform (FT)

2.6.1 Fourier Transform (FT)

Fourier Transform (FT) is considered as a tool of the most commonly used methods to understand the frequency characteristics of data. Although the wavelet and Fourier transformations are both tools to understand frequency characteristics of stochastic data, they are distinguished with respect to the basis functions used for the transformations. FT decomposes data into a combination of complex trigonometric functions, while the wavelet transformation decomposes data by wavelet functions. By the assumption such as the admissibility condition of (2.4.3), the wavelet function is localised for both time and frequency with fast decaying waves to zero at the edges. Figure 2.6.1 shows the line plots of example wavelet functions, the Shannon wavelet function and Mexican hat wavelet function.

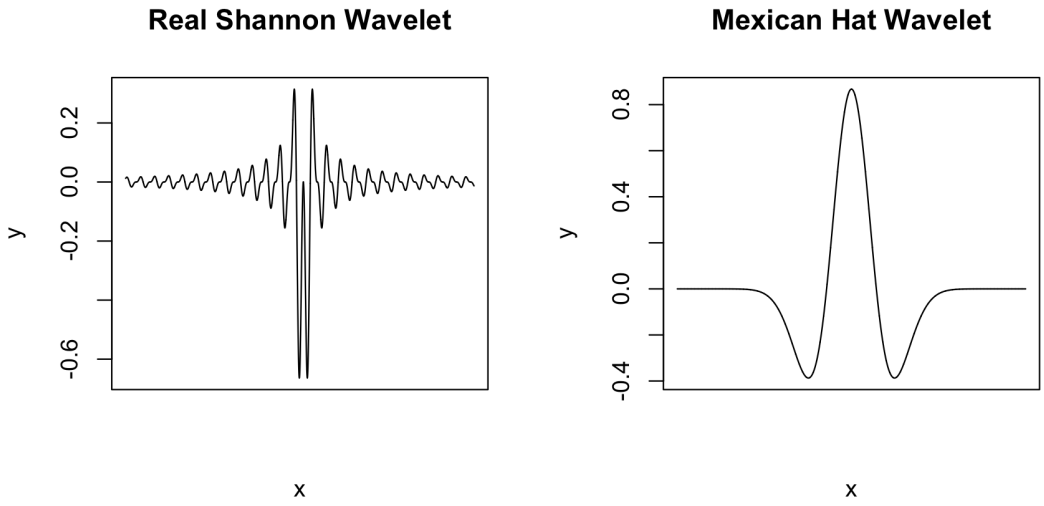


Figure 2.6.1: The examples of wavelet functions: Shannon's wavelet functions (left) and Mexican hat wavelet functions (right).

Meanwhile, the FT projects data onto the complex exponential basis functions. The transformed coefficients can be considered as a series of trigonometric functions.

The frequency of basis functions can be controlled by scale parameters for both a wavelet transform and FT, but the basis functions of the standard FT are not localised in time. To cope with the problem, there has been many developments of FT such as the short-time or windowed FT, but we will not cover details of them in our thesis. Here, we focus on wavelet methods to develop modelling and novel method to analyse signal using basis functions localised in both time and frequency.

Chapter 3

Wavelet Applications on the Vertical Tomography Data

3.1 Motivation

This chapter is about the data analysis of gas-fraction presence of liquid and gas in a pipe. The two-phase flow of gas and liquid can influence industrial processes, and hence, understanding the state of them has been an area of interest in engineering. As the phenomenon is complicated with various elements involved such as pressure, temperature and mass of liquid, a sophisticated modelling is needed to monitor the states.

The gas-flow regime can be categorised into two or more types based on the size or length of bubbles, which is known to rely on the velocities of gas and liquid in a pipe. Taitel *et al.* (1980) and Brauner & Barnea (1986) suggested transition curves describing how gas-flow states change based on the velocity of liquid and air. Also, Behkish *et al.* (2007) investigated the association of the bubble sizes in liquid with temperature and pressure. However, the flow map over superficial velocities is difficult to apply to the general properties of air-flow regime, as there are various elements affecting that (Rouhani & Sohal, 1983).

In addition to understanding the relationship between the parameters, we can approach this phenomenon by electrical tomography process. The tomography is used as the term to represent the cross section image of a target. Particularly,

electrical tomography used in our research are based on the electrical distribution or property occurring in some material. The approach is to determine the electrical tomography from the conductivity distribution related to the presence of bubble in liquid.

However, the image construction needs complicated algorithm and computation time for that. The main objective of our research is to classify the gas-flow regime based on the voltage data without visualising the cross-section of the targeted object. Prasser *et al.* (2002) suggests a modelling of bubble sizes based on the assumption of the linear dependence between the conductivity and gas-fraction distribution. The calculated gas-fraction measurement based on the conductivity of the two-phase flow is decomposed into the distribution model of bubble sizes.

Aykroyd *et al.* (2016) approached this topic from the difference of frequency characteristics between big and small bubbles and suggested a logistic regression model using wavelet coefficients. Goodwin *et al.* (2005) suggested signal process analysis using wavelets rather than image reconstruction from given data. They established logistic regression and discriminant analysis to classify the binary gas-flow type in wavelet coefficients and transformed variables from the coefficients.

3.2 Data Description

The data analysed in this chapter is the same as used in Aykroyd *et al.* (2016), which is simulated under the control of experiment. In this simulated data, the "bubble" and "churn" states are controlled by the input elements such as the inlet size and pressure. The voltage data of the simulated two-phase flow is collected from the pipe placed vertically in which the electrodes are installed along with the cross section by injected current. The voltage data is considered as the conductivity distribution on that vertical gas-liquid flow, which is known as being related with the physics of gas-fraction in liquid.

There are various of ways to measure voltage data by the design of electrode selection. The method used in this experiment is designed with one fixed reference electrode and electrodes for a drive and measurement flow of current in which repeated selection is allowed. In this simulation, currents are injected to the two

electrodes and the difference of currents between the drive and measurement flows is measured. Here, the simulation pipe is installed with eight electrodes and one of them is set for the reference. Therefore, the number of possible measurement pairs is 49 from the 7 remaining electrodes. Figure 3.2.1 shows one example of a drive and measurement flow with the reference electrode and selected electrodes.

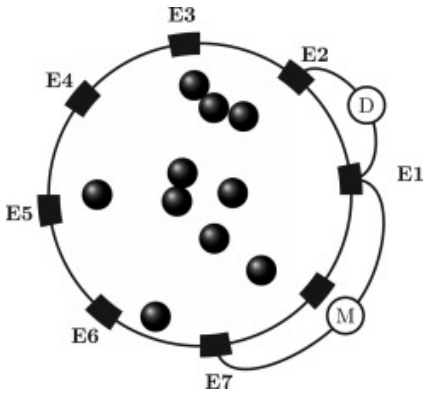


Figure 3.2.1: The diagram of a measurement circuit with selected electrodes (Aykroyd *et al.*, 2016).

In data collection, there are 8 electrodes built in a pipe to observe conductivity distributions. Along with one fixed reference electrode, a pair of remaining electrodes are used to record the flow of voltage over time. Therefore, the total number of electrode pairs are 49 from 7 electrodes except for the reference electrode.

They also simulated the conductivity data based on this environment by Maxwell's equation. The potential electrical field are given from the selected pair of electrodes and pipe between them. The simulated conductivity value can be calculated by a partial difference of Maxwell's equation and the numerical work were done by the software, EIDORS (Adler *et al.*, 2017) in MATLAB. Here, we focus on the simulated data from this process for modelling.

Figure 3.2.2 shows examples of bubble and churn data respectively. As can be seen from those plots, these two data seem to consist of different bandwidths of waveforms. For example, the bubble data tends to be relatively stationary, while the churn data has some rapid changing points and those changed values last long until getting recovered. It means the frequency feature of the churn

data is not stationary and changes over time. As described in the introduction chapter, the standard Fourier transform does not work well on data with local frequency changes compared to wavelets. Accordingly, we can guess the local time-frequency information encoded in the wavelet coefficients is distinct between "bubble" and "churn", and hence the wavelet coefficients can be used as explanatory variables in a regression model. Figure 3.2.3 shows the image of wavelet coefficients from continuous wavelet transform. It supports the distinction of frequency characteristics between the gas-flow regimes and the amplitude of wavelet coefficients changes over the time. The spectral changes seem to be associated with the increase or decrease of voltage on Figure 3.2.2. Therefore, we can expect to investigate the relationship between the physical structure of the two-phase gas-flow using the wavelet transform and its analysis of the voltage data.

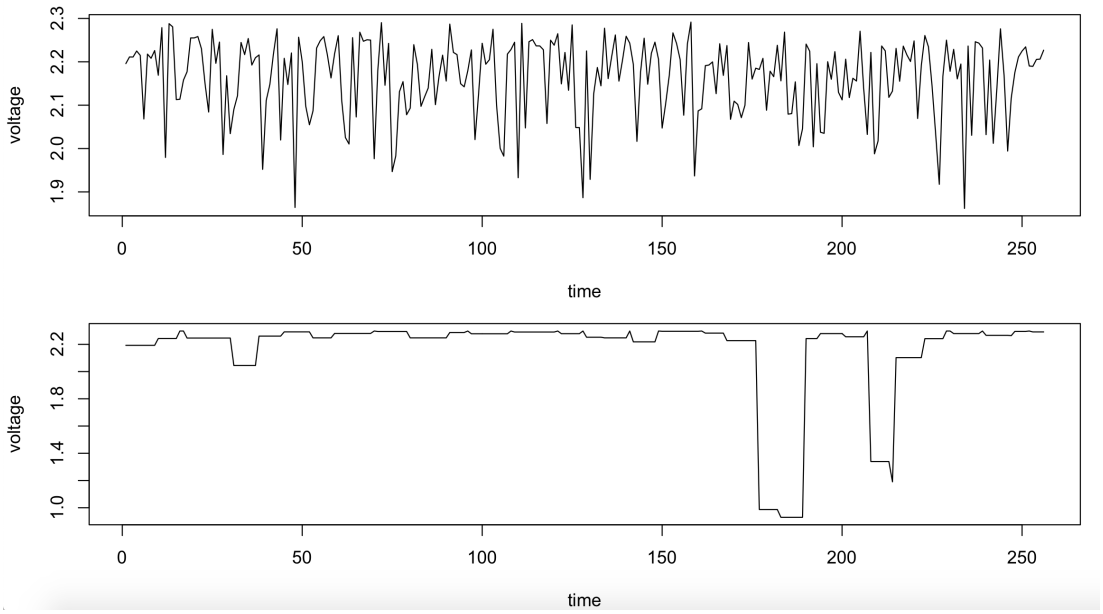


Figure 3.2.2: Plots of example voltage signals of bubble (top) and churn (bottom).

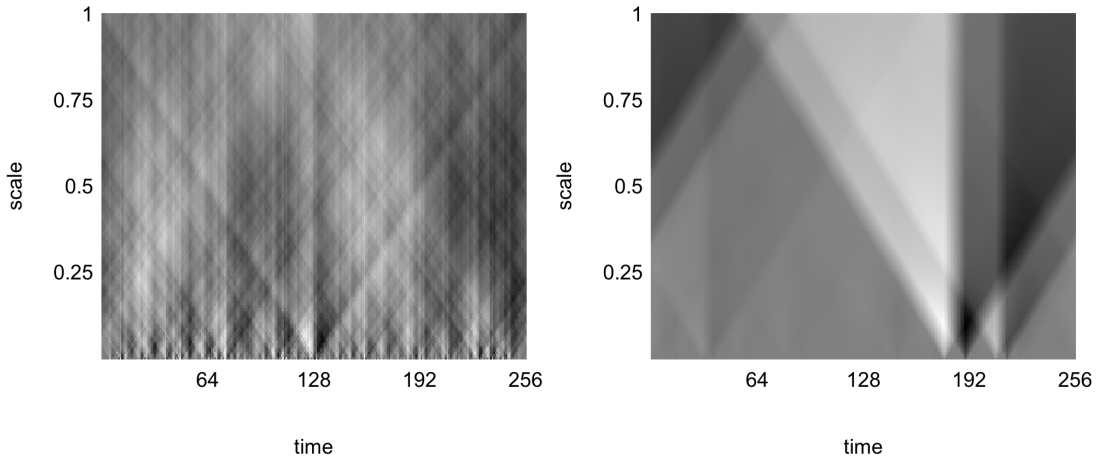


Figure 3.2.3: Plots of continuous wavelet transform from "bubble" (left) and "churn" (right).

3.3 Modelling

Here, we will establish a classification model based on the different frequency characteristics between the two different states. Figure 3.2.2 shows that the changes of bandwidth appear different between "bubble" and "churn" and over time. Time-frequency analysis can be a useful tool for understanding the gas-fraction regime with its frequency characteristics. Wavelet methods are especially used to detect localised behaviours based on the wavelet function being localised on both time and frequency, and thus we will fit a logistic classification model by analysing the wavelets at each scale.

Previous work by Aykroyd *et al.* (2016) established a regression model based on discrete wavelet transform (DWT) coefficients. However, wavelet levels are determined by the number of observations and the number of them is limited due to the dyadic structure between levels. Accordingly, the amount of information can be restricted due to the small number of wavelet levels. To deal with this problem, we can apply the continuous wavelet transform (CWT) to consider denser scales in a continuous domain than discrete wavelet resolution levels. McClusky *et al.* (2002) analysed the gas-flow data using CWT and showed how the wavelet spectrum changes at continuous scales between the different gas-flow

regimes. However, the fully continuous wavelet transform is the decomposition of data by integral over continuous scale and time. The calculation is not simple in practice particularly where data is given by a sequence of data points in time. Therefore, we will apply a sum over discrete values to quantify continuous wavelet transform with the intermediate levels between discrete wavelet resolution levels. In this chapter, we will only consider this discretised version of CWT and call that CWT.

In our modelling, we consider one specific wavelet function, Haar wavelet function to simplify the problem. The Haar wavelet function, Figure 2.2.1, has piecewise constant intervals under a compactly supported domain as one particular kind of Daubechies wavelet family. In theory, the CWT allows all continuous variables within the corresponding domain. However, the wavelet function in CWT is also given as discrete values to compute the coefficients with discrete data values. Accordingly, it is important to choose appropriate scales enabling to express the shape of original wavelet function in the discrete function values as well. The Haar wavelet is designed to have the same contributions of being positive and negative, and hence the discrete wavelet function value on the Haar wavelet should have the same number of positive and negative values. Here, we distinguish the Haar wavelet scales as the number of non-zero values which is the total number of positive and negative values.

Each simulated data consists of 256 observations from each of 49 pairs of electrodes, and hence the number of discrete wavelet resolution levels for the data is given as 8. Assuming that we restrict the range of wavelet scales to have the same positive and negative contribution within data for the Haar wavelet function, the maximum number of continuous wavelet scales is 128, the same as the number of even numbers until the number of observations, 256.

To compare the impacts of intermediate levels, we create 3 more scale sets from the discrete levels and full scales. In wavelet theories, it is already known few of coarse levels does not have valuable frequency information compared to other finer levels. Accordingly, the subset of full scales does not have any intermediate levels between the coarsest and second coarsest discrete levels to avoid interruption of model interpretation. In this principle, there are five sets of wavelet scales, “all”, “sel1”, “sel2”, “sel3” and “dcr” with the different degree of density. The “dcr”

has the same scales as DWT and the "all", "sel1", "sel2" and "sel3" include denser scales than the discrete levels. The "all" includes all possible discretised scales and the "sel1" is the next densest set among these scale sets. Figure 3.3.1 is plotted the scales on each group as the width of corresponding haar wavelet function.

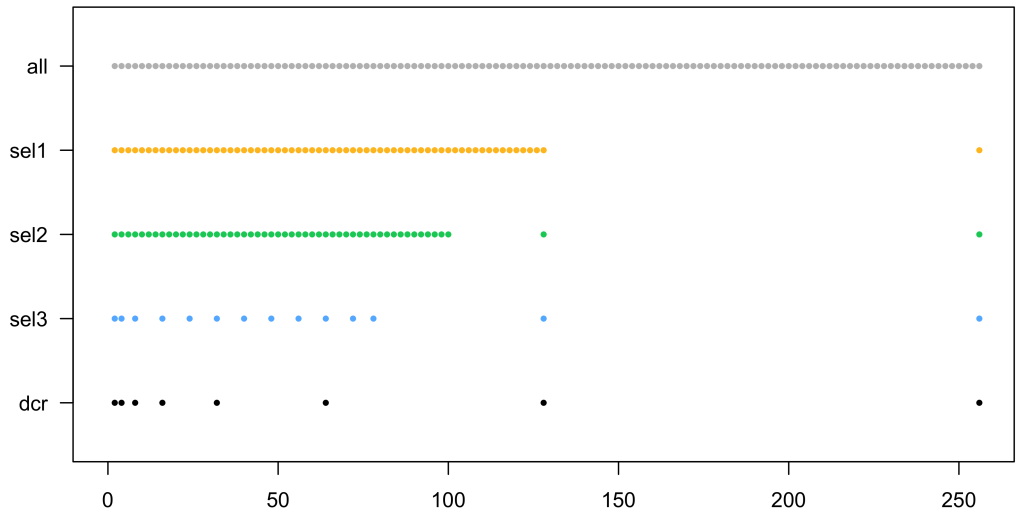


Figure 3.3.1: The width of Haar wavelet function on the scale subsets, "all" (128 scales), "sel1" (64 scales), "sel2" (50 scales), "sel3" (14 scales) and "dcr" (8 scales).

We can write the discretised CWT for a given fine grid of scales, $\boldsymbol{a} = \{a_1, \ldots, a_{|A_n|}\}$, where $|A_n|$ is the total number of scales for this transformation. The wavelet coefficient of discretised CWT is defined as the discrete summation as

$$d_{a_i,b} = \sum_x f(x)\psi_{a_i,b}(x), \tag{3.3.1}$$

where $a_i \in \boldsymbol{a}$ and $\boldsymbol{b}$, $i = \{1, \ldots, |A_n|\}$ and $x \in \{0, \ldots, N-1\}$. In Chapter 3, the number of observations, N, is 256 and $|A_n|$ varies from 8 to 128 depending on groups. The main objective of this modelling is to investigate frequency charac-

teristics associated with a flow regime based on various wavelet scales. However, as the voltage data is transformed separately for each pair of electrodes, the dimension of the resulting collection of wavelet coefficients is huge with 49 $|\boldsymbol{A_n}|$.

Therefore, in our research, we do not consider neighbouring structure of gas-flow in liquid achieved by the location of electrodes and make summary measures representing the information of each individual scale continued across all pairs of electrodes. Zhao *et al.* (2018) state that data aggregation before fitting a model can improve prediction quality. The data are aggregated by using an absolute value and windowing variance in the same way as Aykroyd *et al.* (2016). We call these aggregated values *activity measure* which are defined as

$$\begin{aligned} ms_{a_i,b} &= \sum_s |d_{a_i,b,s}| \text{ or} \\ ms_{a_i,b} &= \sum_s \text{var}(d_{a_i,(b-w+1),s}, \ldots, d_{a_i,b,s}), \end{aligned} \tag{3.3.2}$$

where s is the index of the sensor pair and $\{d_{a_i,b}\}$ are the wavelet coefficients from the CWT. In particular, the window variance provides the variation of wavelet coefficients within the specific width of a window averaging over all pairs of electrodes. Although the regression models are fitted with both, the absolute values and the window variance, the model based on absolute values shows a poorer classification result than that based on a window variance. Also, Aykroyd *et al.* (2016) showed that the window variance had higher classification rates especially with the window width, 15. Therefore, we focus on the result using the window variance with the width, 15 for this tomography data.

The response variable, y is the type of gas-flow regimes of the data. We assume that the variable, y is from a binary distribution classified as "bubble" and "churn" and the state of gas-flow stay steady within each data. For the vector, $\boldsymbol{y}$, we established a logistic regression modelling in the activity measure variables from (3.3.2), $\boldsymbol{MS} = \{ms_1 \ldots ms_{|A_n|}\}$.

3.4 Result

Models were fitted to the 100 simulated datasets individually. Each dataset consists of a training and test set, which are generated independently. Accordingly, the total number of models is 100 from each of training set and each fitted model is tested individually by the test set in the corresponding dataset. The data sets are given from the combined sequences of voltage signals of one "bubble" and one "churn". In a sequence of voltage data, the "churn" voltage signals are pasted after the "bubble" voltage signals.

For the training sets, there are two available options used for the wavelet transform. First, the "pure" setting calculates the wavelet coefficients and the activity measures based on the data within each of "bubble" and "churn" states separately. Meanwhile, the "mixed" setting considers the combined sequence as a vector for the wavelet transformation and generation of activity measures. This means that the "pure" setting does not consider the boundary effects during the transition between two different gas fraction states, "bubble" and "churn", while the "mixed" setting includes the transition for wavelet transform. Fitting a classification model can be on these two different settings, but testing is only with the "mixed" setting, which is close to real data, because the testing process aims to test our fitted models for future real situations. In addition to the "mixed" setting, the voltage data of each test set is shifted before transformation to avoid having the same change points between "bubble" and "churn" as the training set. The activity measures defined in (3.3.2) are calculated from the wavelet coefficients to get a represented value of each wavelet scale at a certain time point.

The number of scales are different between the scale subsets shown as Figure 3.3.1, but the number of wavelet scales is large especially in "all" with 128 scales and "sel1" with 64 scales. Therefore, our models still have a large number of variables after calculating the activity measures. We can reduce that to explain the wavelet frequency characteristics on the tomography data more clearly with regression. There are two criteria used for model selection : correct classification rate (CCR) and AIC (Akaike information criterion) both calculating using the training data by

$$CCR = \frac{\text{the number of correct predictions}}{\text{the total number of observations}} \tag{3.4.1}$$
$$= 1 - \text{training error}$$

$$AIC = 2p - 2\hat{l}(\theta), \tag{3.4.2}$$

where $l(\theta)$ is the log likelihood function and p is the number of variables of the fitted model. In AIC, $\hat{l}(\theta)$ is the maximum value of the log likelihood function for the model. Since the fitted models are exactly perfect in the training sets for almost all samples, they are difficult to choose a model based on the training errors. Therefore, for each case, we assess the preference of the models by the accuracy of their prediction on a test set.

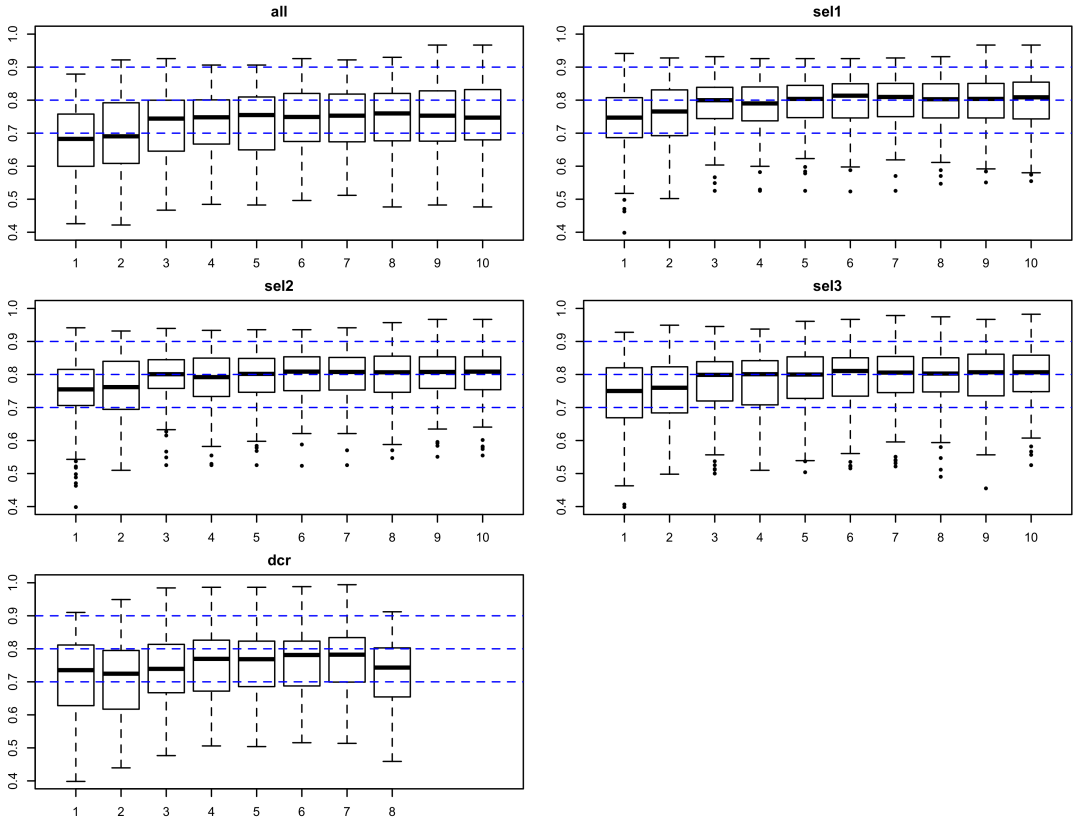


Figure 3.4.1: Boxplots of 100 classification rates from each of scale sets: "all", "dcr", "sel1", "sel2" and "sel3" over the number of variables when model selection is done by maximising classification rates.

The activity measures, which are the explanatory variables in our model, represent information at the corresponding wavelet frequencies, and variable selection process is to choose a set of wavelet frequencies for the classification model with fewer variables. We choose the set of variables from full activity measures minimising the AIC or maximising the classification rate and the selection algorithm continues to add more scales to the model until the chosen criteria cannot be improved fitting. A step-wise procedure is applied to select variables in our modelling. As the increasing number of variables leads to more accurate modelling to minimise training error, the selection algorithm would continue until all the scales were included if allowed to do so. We set the maximum number of variables as 10 to avoid overfitting. Figure 3.4.1 shows classification rates on the test sets from the different groups when using the classification rate (CCR) option. These boxplots show that the classification rates tend to increase until the three-variable model except for "dcr", while they do not change much after then. This means that the variable addition improves the performance of classification models at the beginning, but the effects is minimal after 3 for our tomography data analysis.

It indicates that the three-variable model has the best result considering both prediction accuracy and simplicity of the model. Although the "all" group has the worst result in terms of a classification rate, the "sel1", "sel2" and "sel3" results look better than the "dcr" set. These results indicate that the added intermediate levels available in "sel1","sel2" and "sel3" includes useful frequency characteristics for classification, which do not occur at discrete levels. Compared to the improvement of classification on other subsets of continuous wavelet scales, the "all" set has the worst classification result. The "all" group has a large number of wavelet frequencies, and, in particular, the coarse scales of this scale set are very dense. The models fitted from the dense scale set, "all", are likely to include relatively many coarse scales than the other scale sets and this may lead the poor classification results for test sets. It seems a few of coarse intermediate scales on the "all" set interrupt an accurate modelling.

On the contrary, a variable addition does not guarantee to enhance fitting in terms of AIC, as it includes the penalty term about the number of variables in 3.4.2. Figure 3.4.2 shows the distribution of the number of variables chosen from

each scale set. The majority of fitted models select two variables except for the "all" set. This corresponds to the finding of Aykroyd *et al.* (2016), where the models chosen by AIC tend to have a small number of variables.

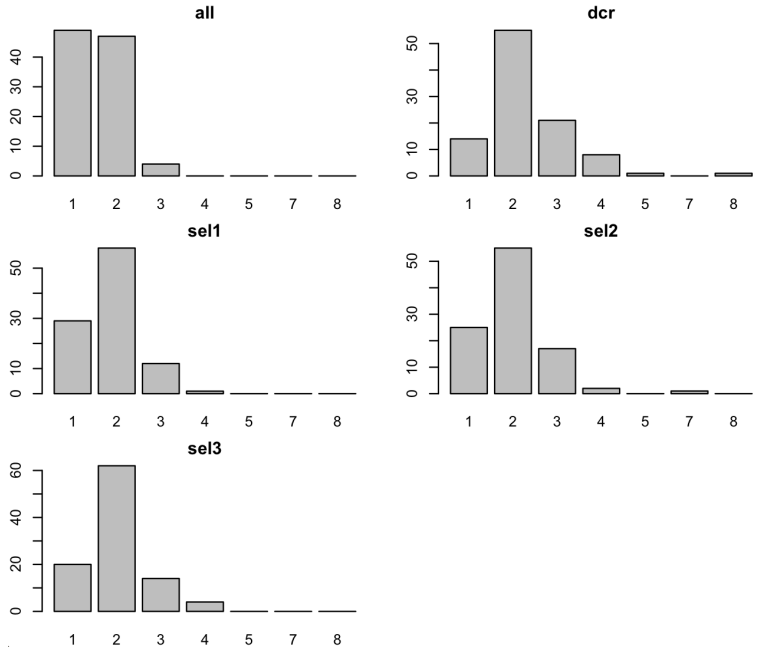


Figure 3.4.2: Distribution of the number of variables chosen in the fitted model using the AIC option.

Figure 3.4.3 shows the comparison of classification rates on test sets between the three-variable models selected by CCR and the final model by AIC. For both criteria, the "sel1", "sel2" and "sel3" subsets get higher classification rates compared to the "all" and "dcr" scale sets. From the aspect of prediction rate, the "all" group, which can use the entire possible range of wavelet scales, has the worst classification result. It seems that the redundant variables in the "all" cannot fit well on the test data despite a low training error, while the limited number of levels in the DWT also seems to cause a low classification rate with a relatively wide range of the boxplot.

The classification rates on the "all" scaling subset are higher for the CCR option than the AIC option. This is not surprising as the CCR option explicitly aims to optimise classification performance, while the AIC aims to optimise descriptive model fit based on likelihood. The medians of prediction rates on the all

groups of scales are significantly different between these two options, and the boxplots are generally narrower for the CCR option. In particular, the classification rates on the "all" are significantly worse than the CCR option with decreasing median to 0.65 from 0.75. It seems to be associated with the number of variables chosen by AIC in Figure 3.4.2. The majority of the "all" has a single variable in the fitted models and it risks relying on too small amount of information to apply to a general problem.

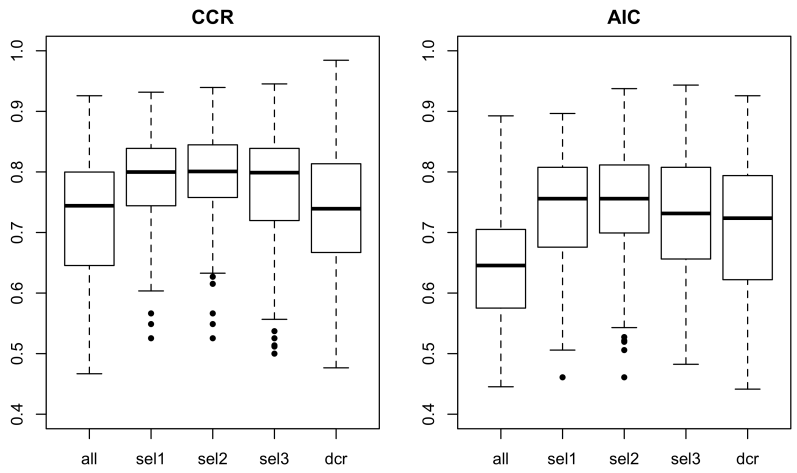


Figure 3.4.3: Boxplots of classification rates from the three-variable model using the CCR option and the final model using the AIC option.

This modelling is fitted based on wavelet coefficients transformed individually from the pure flow-regime data. Although we consider the different training set from the combined voltage sequence of "bubble" and "churn" equivalently to the test set. However, the classification rate are generally worse than using the training set from the pure data. The latter training set seems to misunderstand the frequency information on the transition between different flow-regime states. We cannot say the training set used in our research reflect a practical environment, but it gives more valuable information to investigate the relationship between wavelet coefficients and the flow-regime. Therefore, we will not deal with the result here.

3.5 Comments on the Selected Variables

Figure 3.4.1 shows the two-variable models have enough high median of classification rate for the CCR option, and Figure 3.4.2 shows that the models fitted by the AIC option prefer to choose a small number of variables. It implies that the information from the variables for the first few stages have a important role on classification. Therefore, we focus on the first two selected variables on 100 training datasets and find the patterns of these selection on each scale subset.

Figure 3.5.1 displays the selection of variables using the CCR option on each group which are coloured as "yellow", "green" and "blue" for the "sel1", "sel2" and "sel3" subsets respectively. The black dotted points mean the distribution of selected variables for the "dcr" set to compare that with the selection from the other sets. The size of a dotted point is given from the square root of the number of models having the corresponding scale variables. Also, the bar plots on the sides of axes represent the number of models having the corresponding variables on the each individual step.

As a variable increases, it means a finer scale. Hence, variables 1 and 8 correspond to the coarsest and finest possible scales, and integer values are the scales corresponding to the resolution levels available in the discrete wavelet transform. They are generally preferred to choose the variables between 2 and 4 on the first step of all the groups, while the finest variable selection is dominated on the second step. It implies that the wavelet coefficients and their activity measures from scales 2 to 4 are likely to include valuable information to classify the gas-flow regime. However, as the most preferred variable on the second step is the finest one, it seems to require more detailed information from the fine wavelets to reduce the classification error.

According to the scatter plots and bar plots, there are a considerable number of non-discrete levels chosen from these modelling. On the first step, all groups on Figure 3.5.1 tend to choose their variables from the similar interval between 2 and 4 corresponding to the second and fourth coarsest discrete resolution levels. The considerable number of "sel3" selects the coarsest or second coarsest level as their first variable, which seems to be from a relatively small number of variables on the set. This variable pattern implies that the frequency information between

2 and 4 have important information related with the flow-regime state. As shown in Figure 3.2.2, the "bubble" and "churn" have different frequency behaviours. The voltage signals from "churn" tend to have a few of long lasting wave forms. This frequency characteristic on "churn" seems to lead to coarse scale selection as the first variable. On the second step, majority of all groups equivalently choose the finest level, which gives detail information for classification.

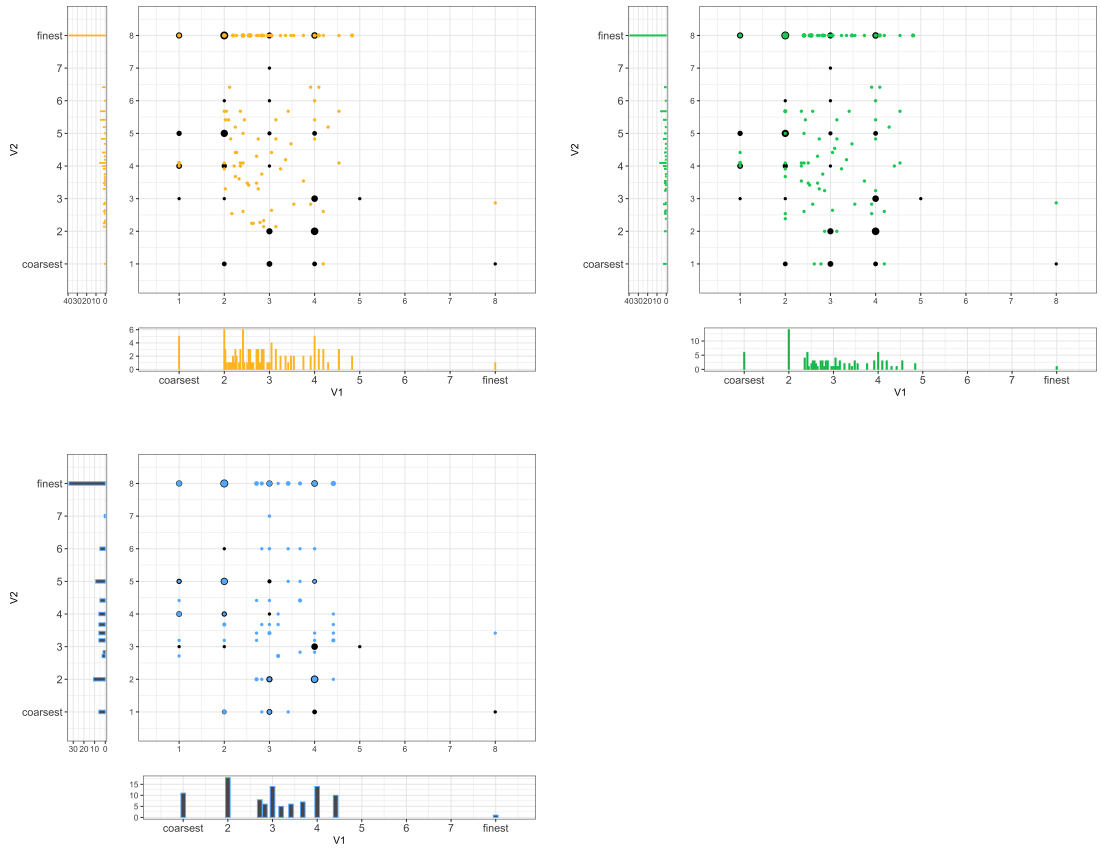


Figure 3.5.1: The best two-variable selection using CCR from 'sel1' (yellow), 'sel2' (green) and 'sel3' (light blue) and comparison to the 'dcr' (black).

Figure 3.5.2 shows scatter plots from the models fitted by AIC in the same way as Figure 3.5.1. The mainly chosen variables are placed in the similar area to the result done by the CCR, but the range of that is slightly widened from scales 2 to 5. However, the second variables are selected distinctly with the CCR models in Figure 3.5.1. About $20 - 30$ % of models stop fitting on the second

step without adding any further variables, while there are a small number of models having the finest variable on both variable selections. The remaining 70 % of models have scales spread between 2 and 5 equivalently to the first variable pattern. Accordingly, it indicates AIC prefers to decrease the number of variables than adding more details.

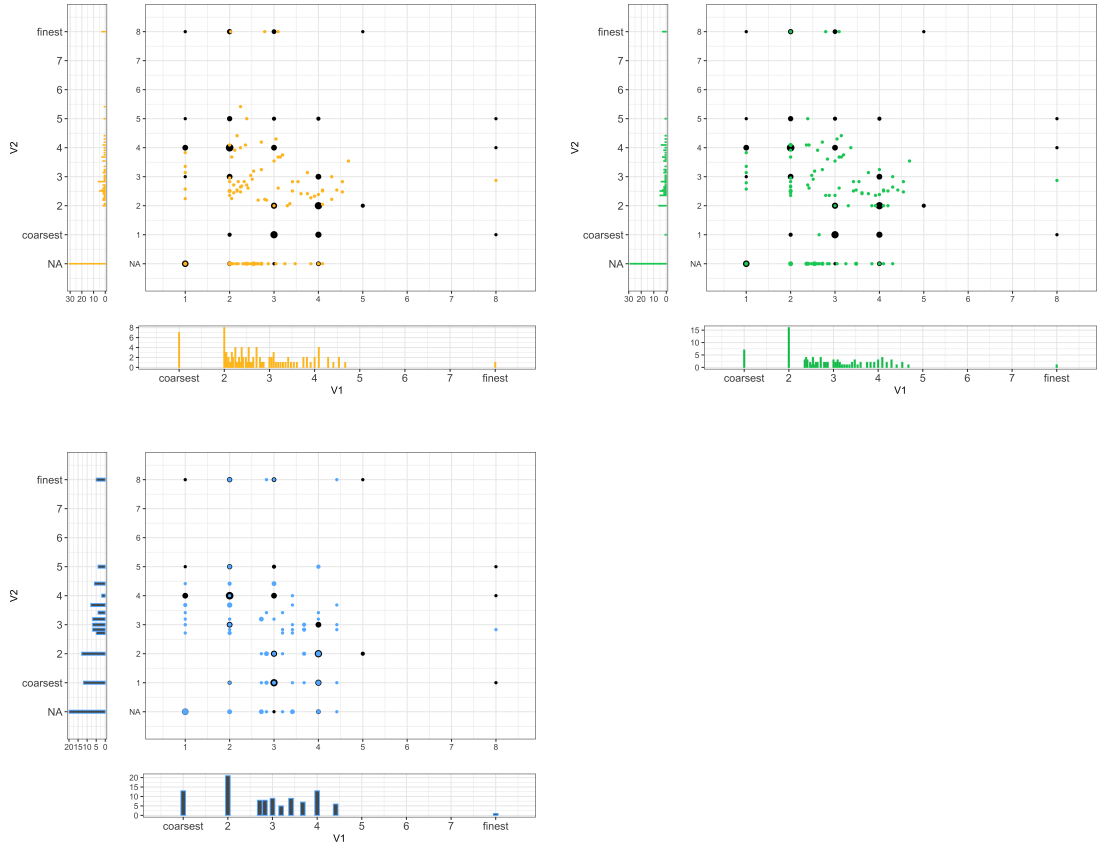


Figure 3.5.2: The best two-variable selection using AIC from "sel1" (yellow), "sel2" (green) and "sel3" (light blue) and comparison to the "dcr" (black).

3.6 Model Comparison on a Validation Set

The classification rates from the previously fitted model are fine, but the training set tends to be overfitted with a high classification rate, and it can lead to overestimate the accuracy of modelling on the training set. To overcome this

overfitting problem, we added a validation set to each of datasets. The newly generated validation set is not involved with both training and test sets, but is used to estimate the classification error on them to choose a model among the models fitted from the corresponding training set.

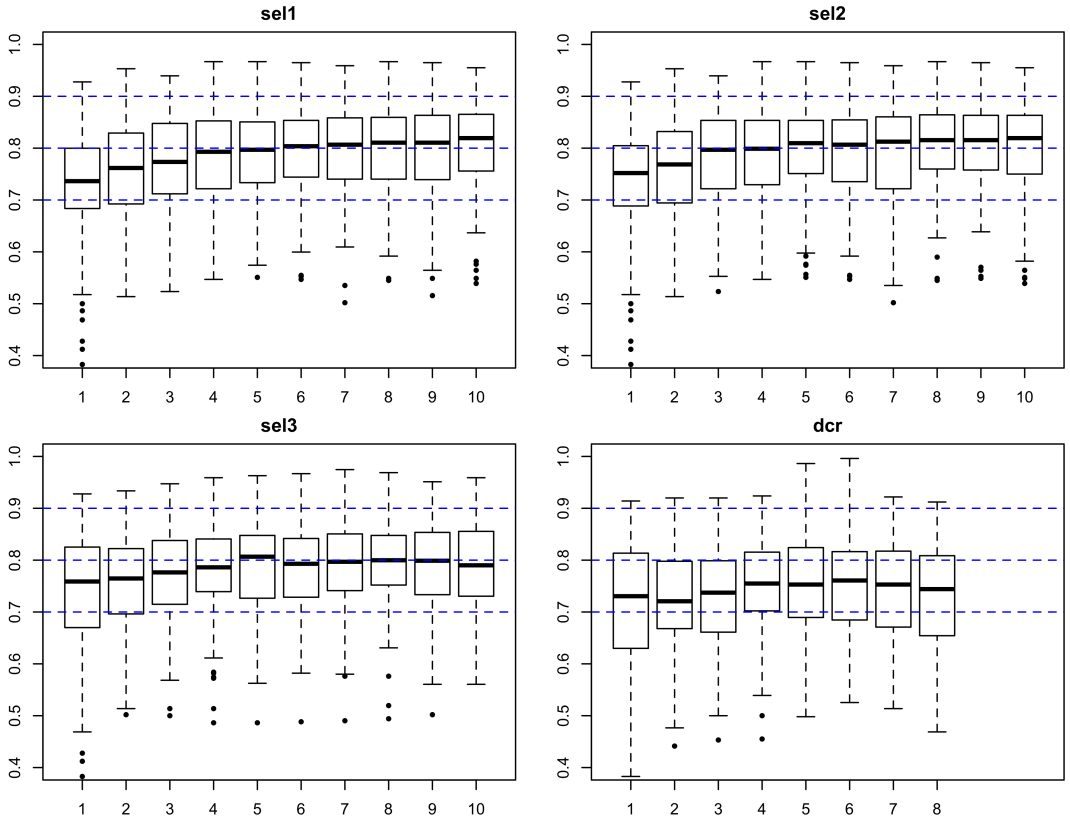


Figure 3.6.1: The boxplots of classification rates over the number of variables from the scale sets: 'dcr', 'sel1', 'sel2' and 'sel3' when model selection is done by maximising classification rates after validation.

The measure activities are generated independently from 100 pure data of "bubble" and "churn" individually in the same way as the training set. There are two changes in the validated modelling. First, we found that the "all" option has the worst classification rates with a wider dispersion in the previous section. Therefore, the "all" scale set is not included in this section. Also, the two of criteria, CCR and AIC, are used to select the best model in the previous modelling. However, the AIC measures the goodness of fit based on the likelihood function

of a model, and hence we need more consideration to derive the likelihood in a validation set given from the fitted model by a training set. Therefore, we only consider the CCR option in this section as the selection criteria.

Figure 3.6.1 shows the classification rate over the number of variables selected from each scale set after using a validation. In general, the median of classification rate changes within the range from 0.75 to 0.85 generally, while the "dcr" is shown a slightly lower classification rate than others. Although the degree of changes of medians is minimal on the most groups after 5, they still increase within a narrower range of boxplots except the "dcr" option. Accordingly, the fitted model using the validation step does not show a clear cut point of variable numbers.

The result from the "dcr" also stays constant after that time but decrease gradually. However, we need to consider the possible number of variables on each set. The discrete wavelet has 8 levels, and hence the accessible range of scale is limited. Therefore, at the last few steps, the discrete wavelet modelling is compelled to choose the remaining variables instead of choosing more informative variables for classification.

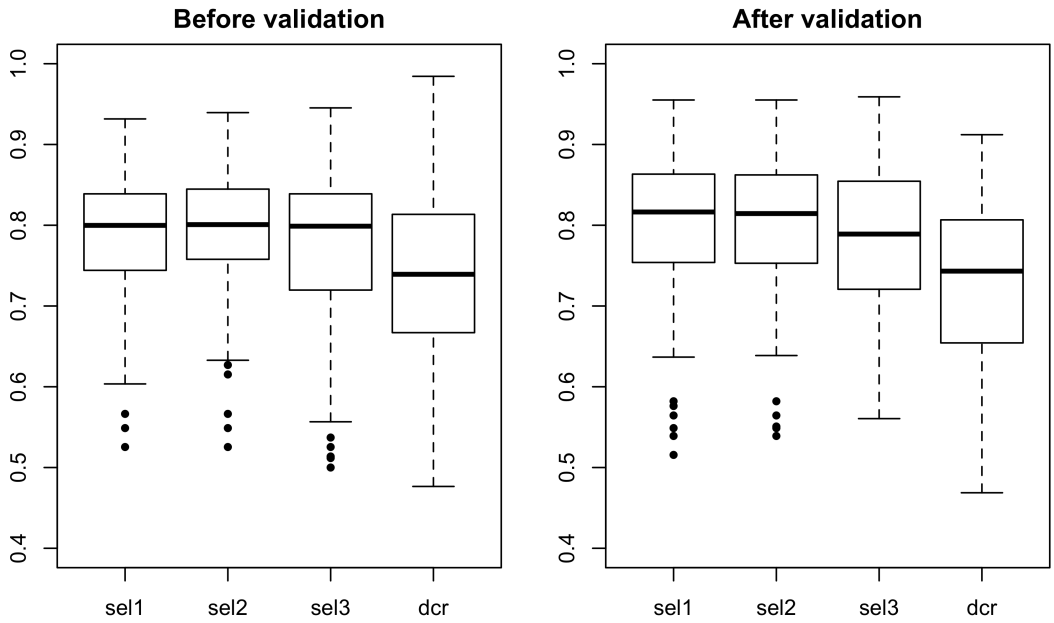


Figure 3.6.2: The boxplots of classification rates of the three-variable model before validation and the final model after validation on the CCR option.

Figure 3.6.2 shows the comparison of classification rate on each scale sets before and after the validation. Here, the fitted models without validation process are the three-variable model, as the classification rate can be considered to be enough stable at this step. Meanwhile, since the models determined through validation process continue to be improved with a varying degree of changes, the classification rates are plotted from the final models. The variable selection on the CCR is repeated until the validation set improves a classification rate or 10. The "dcr" set has 8 scales corresponding to the resolution levels, and hence the maximum step of their modelling is 8.

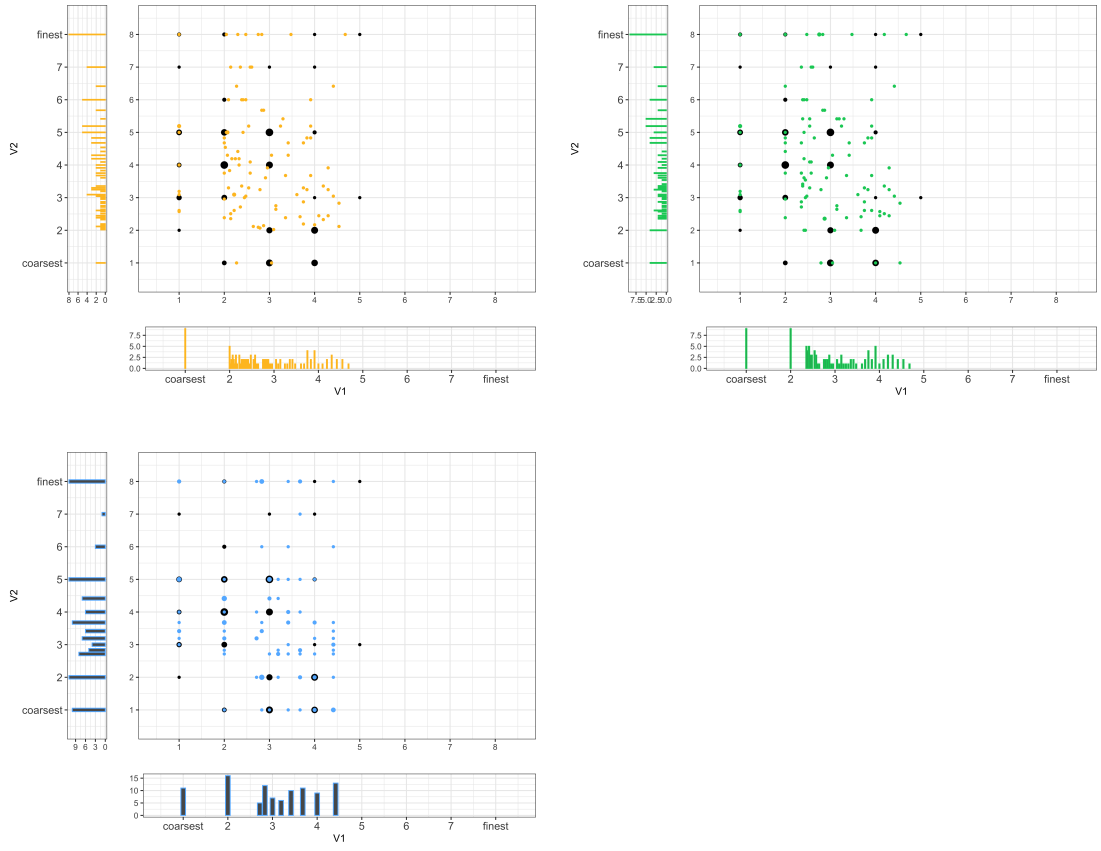


Figure 3.6.3: The best two-variable selection using a validation set and the CCR option from "sel1" (yellow), "sel2" (green) and "sel3" (light blue) and comparison to the "dcr" (black).

Figure 3.6.3 shows the distribution of variables selection on the first and sec-

ond steps. The scatter plots also support that the modelling through the validation step prefer to have intermediate scales to discrete wavelet levels. The first variables have the similar pattern to the previous modelling in that they are generally chosen from Scale 2 to 5. However, the second variables are spread over the entire range, while the previous modelling without the validation step has a large number of datasets selecting the finest scale as the second variable.

3.7 Conclusion

According to the classification results and the variable selection, there are a considerable number of datasets choosing the intermediate levels within a certain range of scales and our modelling approach makes an improvement in terms of the accuracy of classification. However, the "all" scale set, which includes all possible Haar wavelet scales for the given tomography data, generally shows a poorer classification result than the other intermediate scale sets and discrete resolution level set. This result shows that the intermediate levels are helpful to investigate the relationship between the gas-flow regimes and voltage signals, but a very dense scale set can misinterpret the frequency information by exaggerating the influence of coarse wavelet scales on the gas-fraction scheme. Also, the first variable tends to be selected from some specific wavelet scales between Scale 2 and 5. This selection pattern seems to be related to the lower frequency characteristics seen in "churn". While the training error is very small near zero, the prediction error is not satisfactory suggesting overfitting. To deal with the possible overfitting problem, we added the validation step to select the best model. The model after the validation process has, however, a higher median classification rates and the variation also increases. However, the criteria used in the validation step is only the CCR, since it is complicated to apply AIC based on a likelihood. In future research, we can apply the AIC option as well to choose a better modelling result.

Chapter 4

Gas-flow Regime Classification for Horizontal Flow Data

4.1 Introduction

When exploring the gas-fraction in two-phase flow, the direction of flow can be either horizontal or vertical. Chapter 3 established a regression model based on electric signal data from a vertical pipe. This chapter is equivalent to the previous data analysis in that the data modelling considers the voltage data recorded around the pipe on which the electrodes are installed. However, the data is collected using a totally different setting including the pipe orientation and the design of measuring the voltage is different with the previous vertical data. Here, the electrodes for injecting a current and measuring voltage are chosen as different consecutive pairs, while in the previous vertical data collection the possible pairs includes a fixed reference electrode. Also, the experiment is set up to have the independence between the current injection and the measurement by choosing non-overlapped electrodes for the pairs. Figure 4.1.1 shows one example of electrode selection for a single measurement.

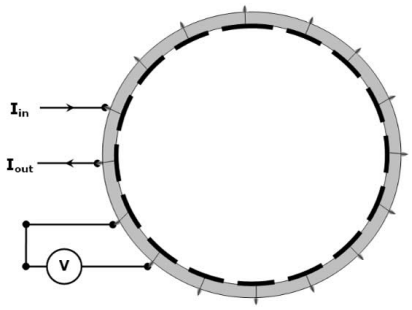


Figure 4.1.1: Cross-sectional view of the "adjacent strategy" for data collection (Taken from www.leeds.ac.uk/olil/tomographySensors.php.)

Figure 4.1.2 shows the flow map of gas-flow regime given from the velocities of gas and liquid in the simulation. The points marked with number means the data we have and their id numbers. Although the previous data analysis used the binary classification of the state between "bubble" and "churn", the horizontal data has the wider selection of gas-flow regime such as "bubble", "slug","plug", "annular" and "stratified". Along with these difference of the experiment environment, it is already known that the two-phase flow is a complicated phenomenon influenced by various parameters such as diameter of pipe, pressure and liquid velocities. Therefore, we need to re-build a classification model appropriate for the new data.

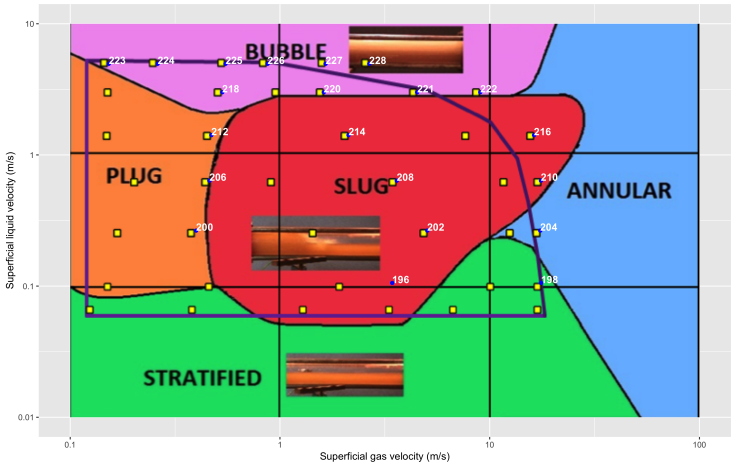


Figure 4.1.2: The flow map of gas-flow regime for this simulation (Corneliussen, 2005)

4.2 Data Description

The analysis approach for horizontal flow of the two-phase state is equivalent to the vertical data analysed in §3 in that the model is established based on the voltage data and its wavelet coefficients. However, from the process of collecting the data, we considered the independence between pairs of electrodes, which gives statistically a more stable condition for modelling. The "adjacent" design, which allows to choose two electrodes next to each other as a pair, restricts to assign an electrode to either the current injection or the measuring on each step. Here, there are 16 electrodes around the cross-sectional of pipe, accordingly the number of measurement from the "adjacent" design is $\frac{16 \cdot 13}{2} = 104$.

However, the current within a pipe tends to be weaker as the measuring electrode pair are farther from the current injection pair. Also, the horizontal data is likely to have gas-flow at the upper side of a pipe, while the vertical data can assume the randomness of gas-fraction locations. Accordingly, the locations of chosen electrodes can influence the measurement by the limited spatial structure of horizontal gas-flow. The collected conductivity measurement can be unbalanced, and hence we need to consider a representation to characterise them from all the electrode pairs at each time. Pradeep (2015) averaged all the voltage measurement from the electrodes pairs to get a representation at each time point. In a similar way, we will average the wavelet coefficients or some function of wavelet coefficients over the pairs of electrodes to get the representative values.

Figure 4.1.2 shows all the types of regimes in the experiment on a graph. However, although the velocities of liquid and gas are important parameter relating to the gas presence of the two-phase flow, the flow map cannot guarantee the state given from the flow-map is always true. As we mentioned previously, the two-phase flow is a complicated phenomenon relying on not only the velocities but other various elements such as temperature and pressure. Also, the data observed from the boundaries between two different states can be confusing to classify as one specific type. To avoid possible confusion of labelling, we need to compare the video data, as in Figure 4.2.1, to the flow-map. We have 21 datasets matching the flow map, only 14 datasets have video data. Two datasets out of the remaining 14 datasets have identical voltage measurement with different labels, which seems

to be saved over during data collection. Accordingly, we use the 12 datasets in this chapter having both voltage measurement and video dataset. Some datasets have the flow-map labels not matching the video data. For this case, the gas-flow state shown from video has a priority than the flow map because video shows the real image of flow on that time. When comparing the flow-map to the video data sets, there are 7 datasets agreed by both the flow-map and video. We will fit a regression model on the 7 data sets to get a more reliable classification.

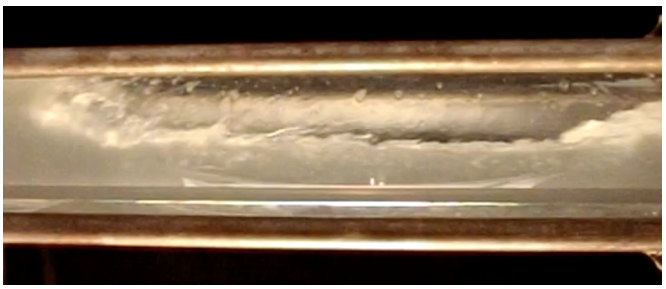


Figure 4.2.1: The flow map of a gas-flow regime.

Figure 4.2.2 shows that the examples of averaged voltage data over pairs of sensors from each gas-flow type and their wavelet coefficients given from the discretised continuous wavelet transform. At first, the "bubble" data has a remarkably lower amplitude than the other two types. Although we cannot say that the "bubble" data is stationary, it looks like having low and quick changing oscillations. In contrast, the "slug" and "plug" tend to change within wide ranges with little long-lasting patterns after a rapid and sharp spike. Also, the voltage data does not seem to be consistent in time except for the "bubble". These frequency behaviours are more clearly seen in the continuous wavelet transform on the lower image plots. The "bubble" has widely spread oscillations. Meanwhile, the two states, "slug" and "plug", seem to have a similar pattern in that both have significant peaks and troughs in the wavelet oscillations at specific time points. However, we need to consider that the changes of frequency characteristics appear more frequently in the "slug" than in "plug". Accordingly, the "slug" data consists of more changes in the continuous wavelet coefficients than the "plug". Also, the wavelet coefficients presumably have significant numerical differences between the states, "plug" and "slug", due to very high amplitude

spikes in the "slug" state despite the similar data structure. The wavelet coefficients from example data support that the horizontal tomography data also has different frequency behaviours over gas-flow states and time. Therefore, in the same spirit as the vertical tomography data analysis, wavelet methods can be a useful tool for fitting a classification model to investigate the local frequency structure over time from each state.

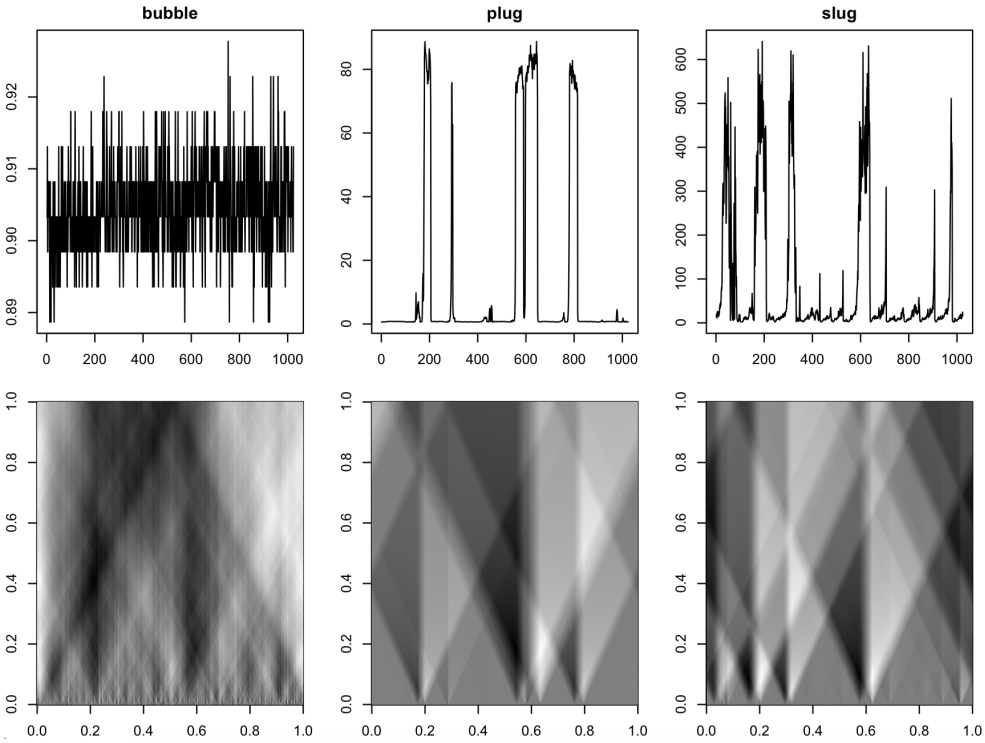


Figure 4.2.2: Examples of averaged voltage data (upper) and the corresponding continuous wavelet transform (lower).

The wavelet coefficients in Figure 4.2.2 are plotted at the centre point of the corresponding wavelet functions. For example, the Haar wavelet function, used in our research, is a piecewise constant function having the centre point between positive and negative values, and the calculated wavelet coefficients from a wavelet function are generally plotted at the corresponding center point. However, the modelling approach in our research is aimed at classifying the gas-flow type from the transformed wavelet coefficients until specific time points rather than center points. Therefore, the wavelet transform in this chapter is based on backward

facing data and calculated the corresponding past data from each time point. Figure 4.2.3 shows the wavelet coefficients calculated from the backward facing approach, which is used for modelling the gas-flow state at the corresponding time point.

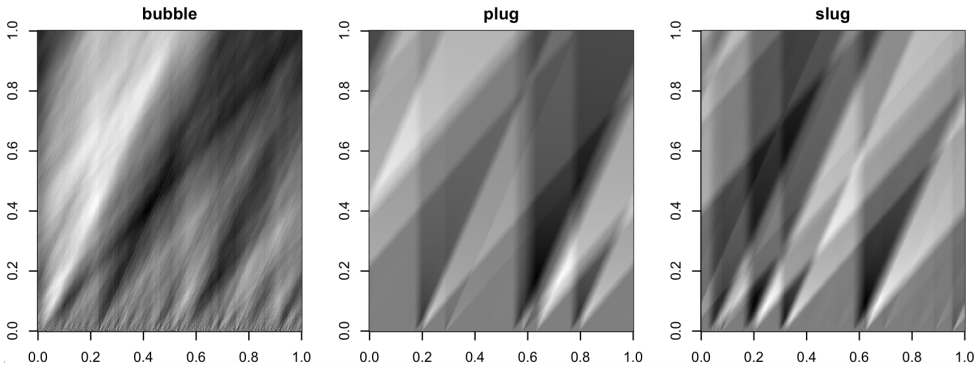


Figure 4.2.3: Examples of continuous wavelet transformation plotted for modelling.

4.3 Modelling of Local Frequency Characteristics

The data includes the three kinds of gas-flow regimes, "slug", "plug" and "bubble", while the previous modelling was performed using binary responses, "bubble" and "churn". There are two more types of states, "stratified" and "annular", but these states are not identically defined by the flow-map and video. Therefore, the horizontal data is analysed with the remaining three-level gas-flow states for more confident classification result.

The horizontal data model is fitted a general linear regression in the logit, log of odds ratio between different responses. The multinomial modelling sets one of the multinomial responses as the base line and establishes a generalised logistic regression model individually on the ratio with others.

The seven datasets consist of three "bubble", three "plug" and one "slug". In a similar way to the previous modelling, we split the data into the three different sets as the training, validation or test set. The first 1024 observations are sampled from each dataset for the training set and they are pasted in a row.

Therefore, the training set has 7168 observations from seven datasets. In the same way, remaining data from each dataset is split into the validation and test set individually and they are pasted one after another as well. The validation and test sets have 3228 and 3233 observations respectively. The validation step checks the estimated model of the training set and decides a model based on the prediction rates. The selected model through the validation step is verified by the other dataset which is independent with the training and validation set, so it enables models to reduce overfitting effect.

We found that the use of all available scales does not guarantee an improvement in the modelling from the previous result and the computation of the continuous wavelet transform is time consuming work using a dense scale set. In particular, a half of the continuous wavelet scales are placed between the two coarsest discrete resolution levels. Therefore, we will use a subset of the fully continuous wavelet scale for more efficient data analysis.

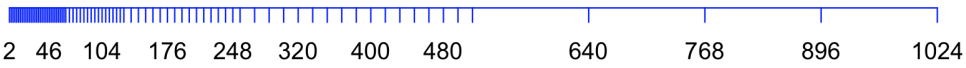


Figure 4.3.1: The scale sets used in the discretised wavelet transform. The labelled number means the width of non-zero area in the Haar wavelet function.

Figure 4.3.1 shows the scales used in our research. The label on the plot means the width of non-zero wavelet function in Haar wavelets. For consistent modelling and testing, the normalising constant on the wavelet function is set by the width plotted on Figure 4.3.1, not depending on the number of observations.

The datasets used in the modelling are given by combining the measurement calculated from the different conductivity data, and we can consider two different approaches of transformation, “pure” and “mixed”. First, “pure” transform is the transform calculated within one conductivity data, which means the activity measurement and its wavelet signal come from the same observation without any distortion from other observations or the change of gas-flow types. The calculated wavelet coefficients and their activity measurements from each data are combined as a string with the same order of original data. On the other hand, the “mixed” transform is calculated with one string data containing all

the conductivity data. It means we do not consider the change of gas-flow types or measurements. In practice, since we do not have any information about the current states, it seems to be reasonable to apply the mixed setting without any split for wavelet transformation or activity measurement calculation.

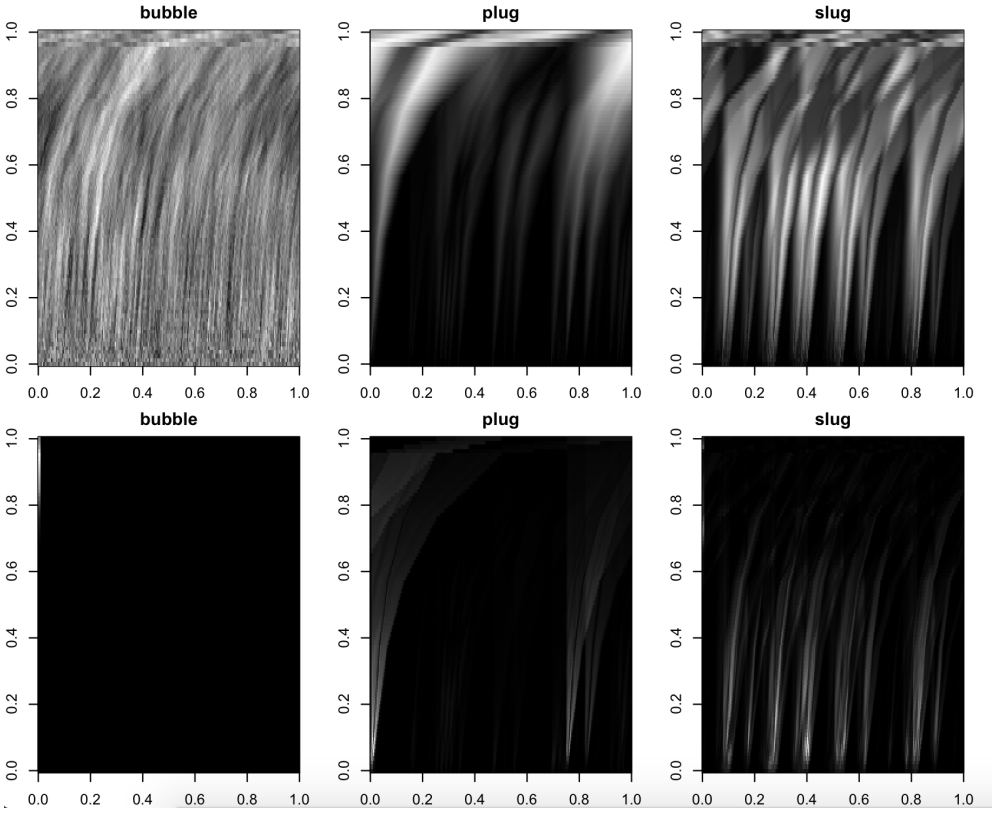


Figure 4.3.2: Examples of activity measures: the absolute value (upper) and window variance function (lower).

There are the 104 pairs of electrodes and the voltage data is measured from the pairs individually. The wavelet coefficients of the voltage data are computed individually from each pair, and hence we need to have a representation value over them by time points. Accordingly, the activity measures are defined as the mean of four different kinds of functions, original value, absolute value, window variance and window variance of absolute values,

$$ms_{a,b} = \sum_i M(d_{a,b}), \tag{4.3.1}$$

where d_{ab} is the wavelet coefficients calculated from the scale, a at the time b for $a \in \boldsymbol{a}$, $b = 0, 1, \ldots, N-1$ and M is the activity measure generating function such as mean. Here, the window variances are calculated from the past observations within the corresponding window, which is equivalent to the wavelet transform of the horizontal gas-flow data. Since the "slug" and "plug" show a similar structure in Figure 4.2.2, we set the window width as 10 to measure more sophisticated local variations than in the vertical tomography data analysis.

Figure 4.3.2 shows the activity measures of the example data. The absolute values have similar patterns to the wavelet coefficients described in Figure 4.2.2, while the window variances tend to emphasise the change in the wavelet transform. The calculated activity measures using the past data until the specific time point are fitted to classify the flow-regime responses. In the multi logistic modelling, a step-wise procedure is applied to select wavelet scales in our modelling.

4.4 Results

The horizontal flow models also have perfect classification results on almost all training sets, and therefore the classification rates on test sets are considered to measure the performance of classification models equivalently to the vertical tomography modelling in §3. In general, the prediction results using a pure training dataset is better in terms of the predicted classification rate. In particular, most activity measures using a pure training dataset have quite high classification rates of around 0.7 or 0.8, except for the "mean". Here, the activity measure, "mean", is calculated from the original wavelet coefficients without considering the sign of values. This causes some of the activity measures to be cancelled out through averaging, which can lead to misinterpretation of the wavelet signal used for the classification model.

Using mixed datasets has the advantage that all the datasets and their transform are close to the practical data environment in that the wavelet transform is performed at all the time points including transition points between different gas-flow types. However, the prediction rate plunged to lower than 0.6 for the window variance of both original and absolute values. Although it is clear that the pure dataset provides better classification modelling, it is impressive that the

absolute value appears to give a similar classification rate in the mixed training set as well. It implies that the absolute activity measure reflects well the frequency characteristics needed for classification on this horizontal data, and its classification result is not much affected by the type of training sets.

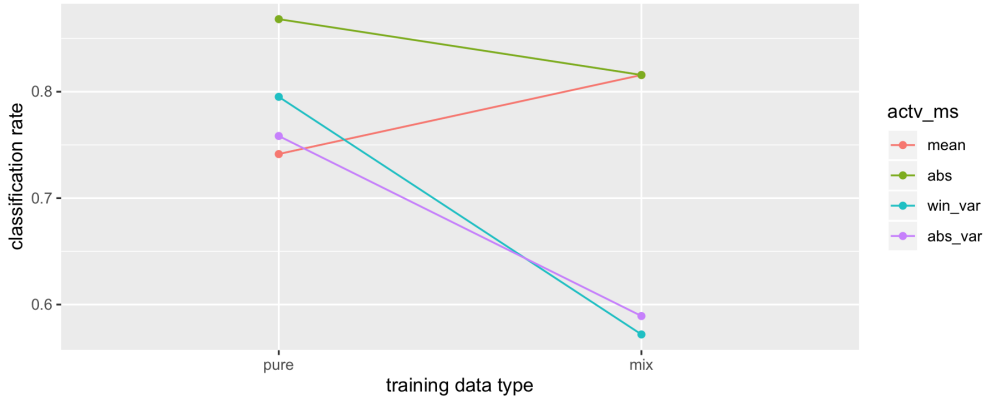


Figure 4.4.1: The result of classification results.

Tables 4.4.1 and 4.4.2 show the number of observations classified to each state by the two fitted models using the absolute and window variance activity measures from the pure training datasets. Here, we focus on the two models, because the pure data setting generally has the better classification result and the two models using the absolute value and the window variance show the two highest prediction rates in Figure 4.4.1.

The tables give the comparison of prediction and true labelling by the response types. According to the result, most observations in "bubble" data have true prediction, while "slug" data has a considerable number of mis-classified cases as "plug". The high classification rate for "bubble" seems to come from the distinct difference of amplitudes compared with the other types. Calculating the prediction rate each state individually, the absolute model shows 0.94, 0.97 and 0.35 from the "bubble", "plug" and "slug" respectively, while the window variance model has 0.9, 0.8 and 0.24.

		true		
		bubble	**plug**	**slug**
pred	**bubble**	1298	13	0
	plug	77	1331	316
	slug	3	24	171

Table 4.4.1: Prediction result using the absolute activity measure.

		true		
		bubble	**plug**	**slug**
pred	**bubble**	1242	118	26
	plug	135	1212	344
	slug	1	38	117

Table 4.4.2: Prediction result using the window variance activity measure.

4.4.1 Comments on Variable Selection

Here, the modelling is established on one common training dataset, and hence it is difficult to see the consistency of fitting. However, we are still able to observe the variable selection from different activity measures and compare the preference of the intermediate scales to the discrete wavelet resolution levels. Figure 4.4.2 shows the selected variables for the pure training dataset where the integers of the x-axis correspond to the discrete wavelet resolution levels. In general, the chosen scales are plotted on non-integer areas, which means that the intermediate scales, between discrete wavelet resolution levels, are used to classify the gas-flow regime. However, the range and pattern of those variables are different depending on the functions used for the activity measures. While the window variance and original value are spread from very coarse scales to fine scales without skewness, the absolute value and absolute window variance are fitted with a bunch of consecutive scales around a specific scale variable. From Figure 4.4.2, it can be seen that the model of absolute values is mainly selected around Scale 7 and the model using window variance of absolute values chooses variables around Scale 2-3. This implies that these two models are likely to

have the most informative frequency characteristics for classification around the certain scales. The model with the absolute window variance, however, is based on consecutive scales around 2, which are too coarse to react to a change between different types of gas-flow states. The coarse variables selected lead to a worse classification, although the chosen variables had a important role for the training and validation steps.

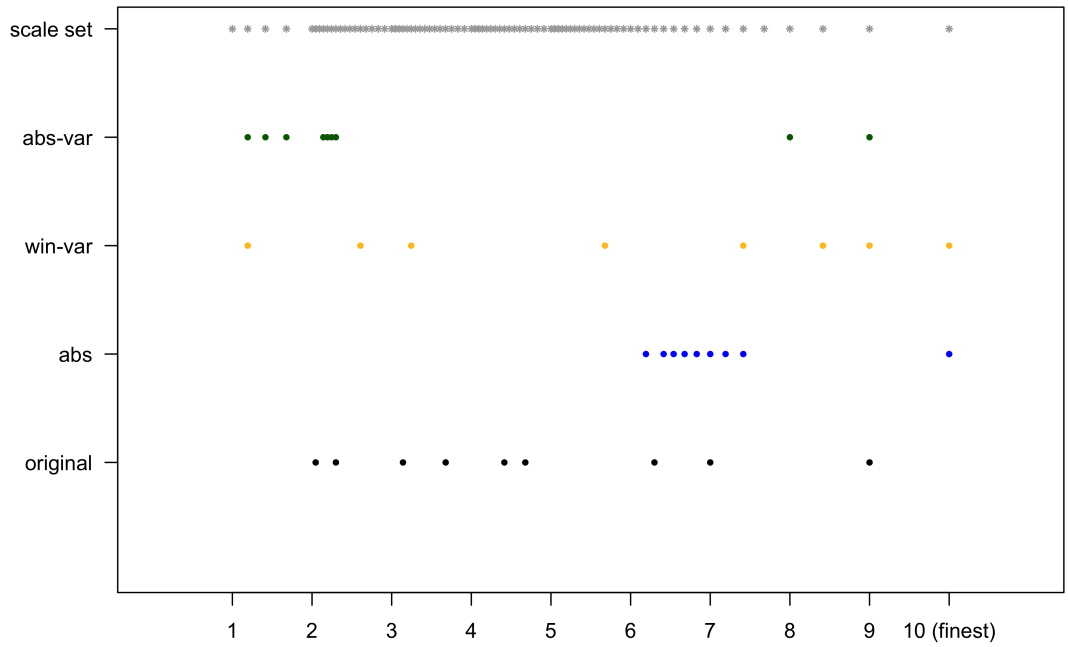


Figure 4.4.2: Selected variables with the pure train set using the original values, absolute values, window-variance and window-variance of absolute values.

Figure 4.4.3 shows the selected variables for the mixed training set. The patterns of variable selection using the mixed training dataset seem to be very different to the previous one using the pure training dataset. For example, the absolute value model using the pure training dataset includes adjacent scales within a specific range, while the equivalent model using the mixed training dataset chooses variables spread over all scales. Also, the absolute variance model selects rather fine scales, between scale 6-8, than Figure 4.4.2, and most variables are placed nearly each other. Likewise, the change of the training dataset leads to choose different wavelet scales for the classification model, as they have different

frequency characteristics during transition between individual data especially on coarse wavelet scales.

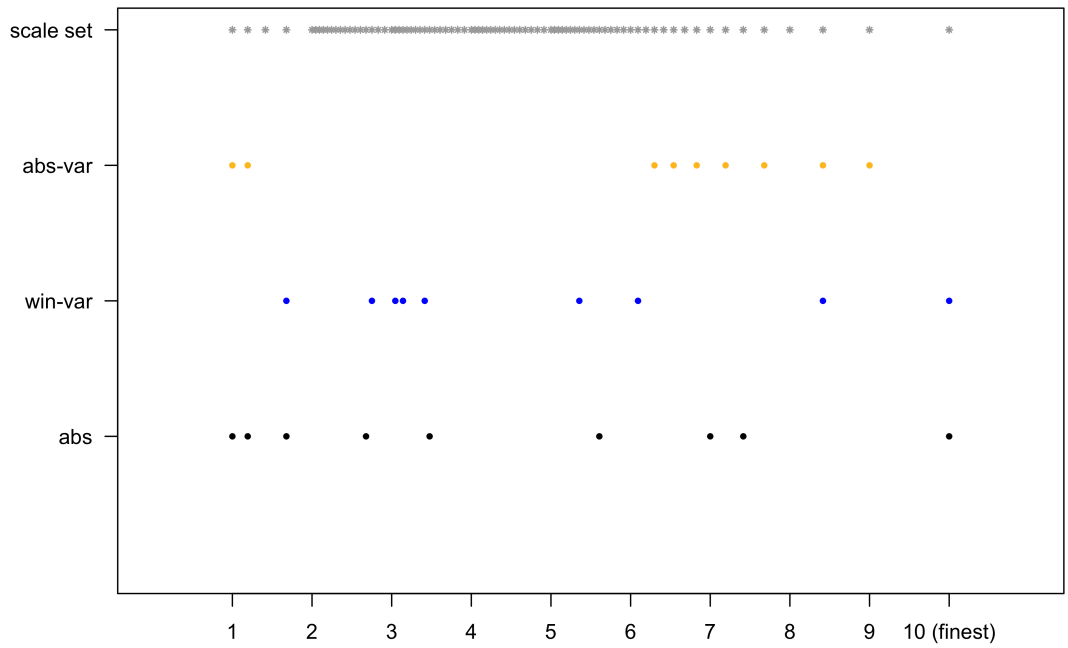


Figure 4.4.3: Selected variables with the mixed train set using original values, absolute values, window-variance and window-variance of absolute values.

4.4.2 Prediction Result on Test Sets

The prediction results can be seen individually on each dataset and gas-flow regime. Figures 4.4.4 and 4.4.5 show the change of predicted labelling with different colouring for the different activity measure models. Figure 4.4.4 shows the prediction result using the pure training dataset. The absolute value and window variance models seem to predict the result well as seen in the classification rate from Figure 4.4.1. Also, they react relatively quickly at the change of regimes compared to other activity measure models. In contrast, the mean model does not have a clear separation between different data, and hardly predicted "slug" state. This gives a poor classification result for the mean model using the pure training dataset. However, despite the good prediction result for the absolute value and window variance, the prediction quality for the slug dataset, which is

coloured green, is still poor. To improve a classification result, we need to find a way to get a better separation between “slug” and “plug”.

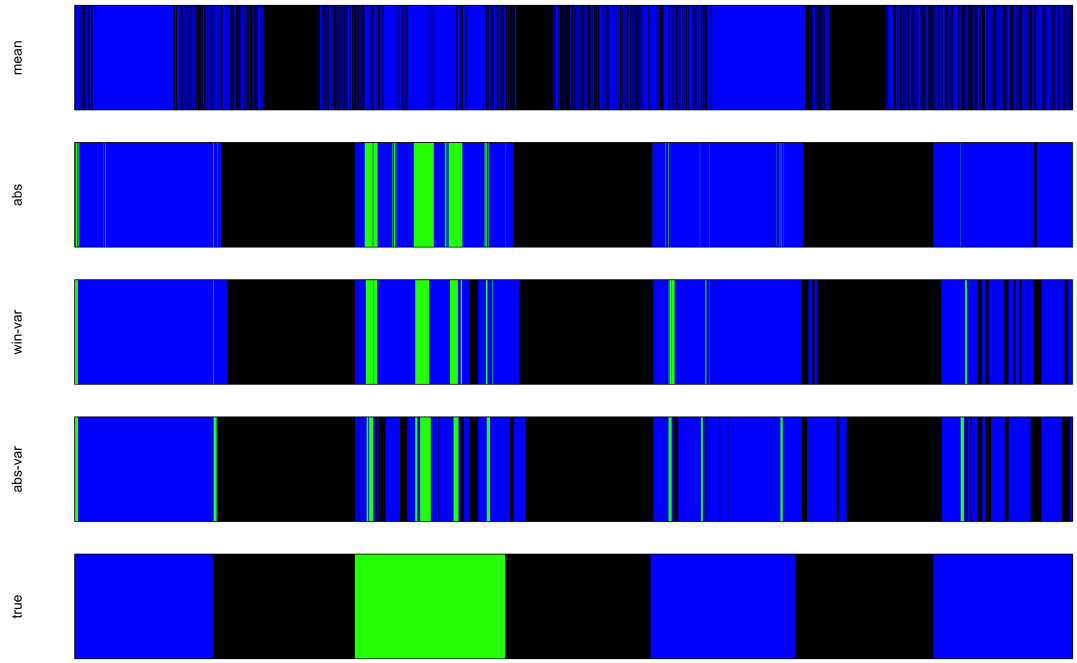


Figure 4.4.4: Predicted gas-flow regimes by the fitted regression model trained on the pure train set: “bubble” (black), “plug” (blue) and “slug” (green).

Figure 4.4.5 shows the prediction results for the mixed training set. The absolute window variance and window variance model have a poorer separation between “bubble” and “plug” compared to the models using the pure training dataset. In contrast, the absolute value model works well for the mixed training dataset, with clearly separated predictions except for the “slug” regime. It supports that absolute value does not show much change in classification rate between the pure and mixed training datasets. However, all the models using the mixed training set hardly predict the true “slug” state, and hence we need to consider to improve that.

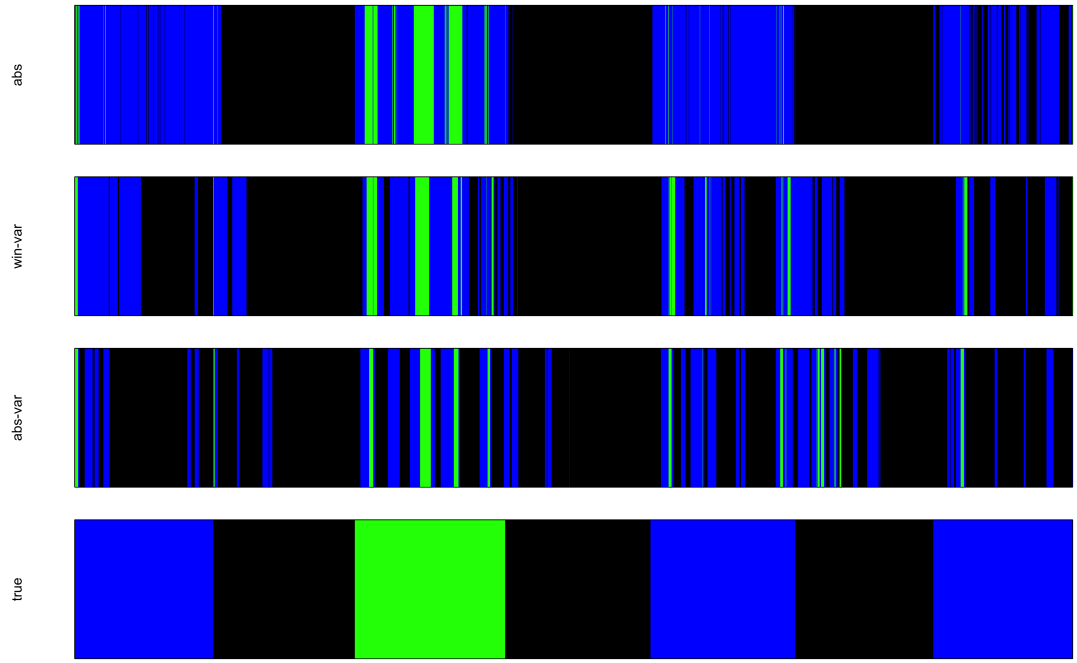


Figure 4.4.5: Predicted gas-flow regime by the fitted regression model trained on the mixed train set: “bubble” (black), “plug” (blue) and “slug” (green).

4.5 Prediction on non-labelled datasets

The training and testing in §4.3 use the seven datasets which have clearly defined labels according to both the flow map and video. Although the remaining data is not appropriate to train a model, in that the state on the flow-map and video data present different states, we can classify them using the model fitted on the previous seven datasets.

We can split the remaining data into the three types: non-labelled, not matched on the flow-map and transient data. For each, the gas-flow type is predicted using the previously fitted model. In the previous section, We found the pure training set, particularly with the absolute value and window variance, provides a stable prediction result. Accordingly, we apply these two models in this section with the three different examples. First, Dataset 198 does not have any information of label due to a mistake during data collection and Dataset 220

is classified as “slug” by the video, but does not match to the flow-map. Dataset 228 is a transient data on which the gas-flow regime changes over time.

Figure 4.5.1 shows the prediction result coloured by the label from the three different examples using the absolute value and window variance models. The coloured bar plots show which regime is predicted the most for each and how the prediction differs between the two models. For Datasets 198 and 220, the most classified regimes are the same as “slug” and “plug” respectively, although the individual classification over time does not exactly match between the models. Dataset 228 does not seem to have the equivalent result, but both absolute and window variance models are transient between “plug” and “slug”. Here, if the model reflects informative data characteristics well, we can classify the datasets as the most classified types.

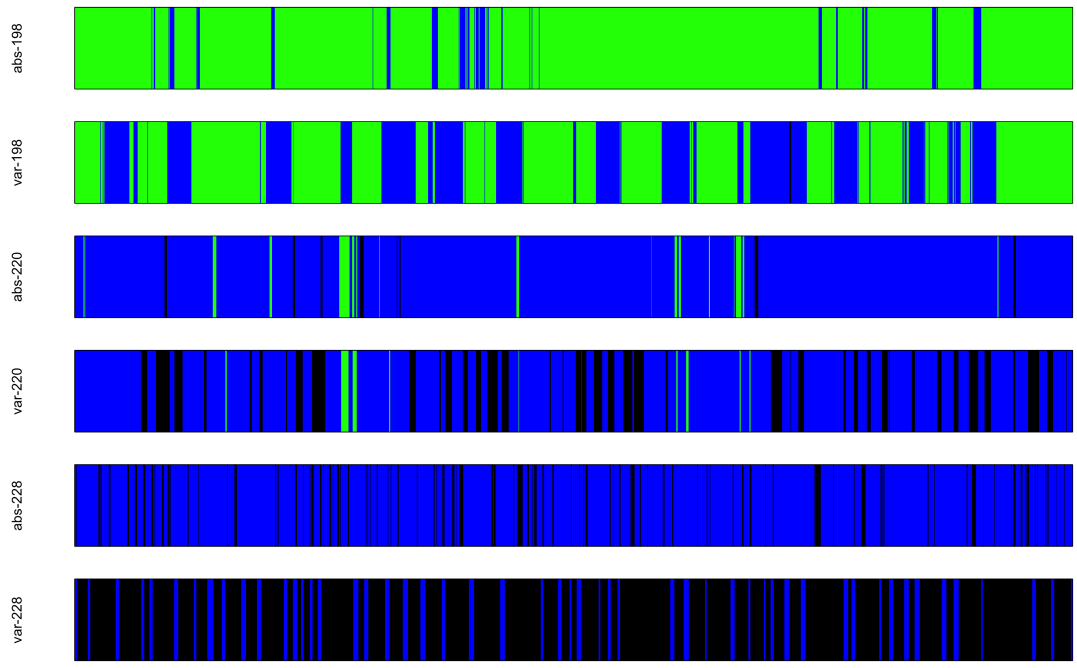


Figure 4.5.1: Predicted gas-flow regimes by the fitted regression model tested for Dataset 198 (non-labelled), Dataset 220 (slug) and Dataset 228 (transient) : “bubble” (black), “plug” (blue) and “slug” (green).

However, we need to consider the possibilities of mis-classification. First, practical data can be a case not belonging to the given gas-flow types. The

horizontal-flow data has two more kinds of the flow types: "stratified" and "annular", but we do not have data sets corresponding to these types. The model from our research are built on the other three gas-flow type, and hence it cannot predict the probabilities of the "stratified" and "annular" state. In addition, the modelling are influenced by the data used in training and testing. Since Dataset 220 does not have the same state from the video data and the flow-map, this non-steady state of the dataset can lead us to have a wrong classification. Last, we found that the classification between the "plug" and "slug" is not done clearly for Figure 4.4.4 and 4.4.5. Therefore, the mis-classification can be from the poor classification of the "slug" type.

4.6 Modelling Using an Adjusted Scale Set

The above models are established on the unequal-sized training, validation and test sets. The numbers of observations in the validation set and test set are almost half of the training set. Assuming we restrict the Haar wavelet function to be fitted within the dataset, the range of possible Haar wavelet scales are determined by the number of observations. In other words, the possible continuous wavelet scale set is narrower on the smaller number of observations.

In the previous modelling, we retained the fully continuous wavelet scales for the training step. The pure validation dataset should be transformed from the smaller number of observations than the training dataset, and hence it cannot have an identical scale set for the training step. In this section, we will discard a few of the coarse levels to have an identical scale set in the pure validation dataset as in the training dataset. The adjustment of scale set is aimed to comparison between the pure validation dataset and mixed validation dataset for classification, and also enables us to see how the scale adjustment changes classification result. Here, we remove the corresponding scales around the two coarsest discrete wavelet levels from the scale set shown in Figure 4.3.1. The scale used in this modelling are shown in Figure 4.6.1.

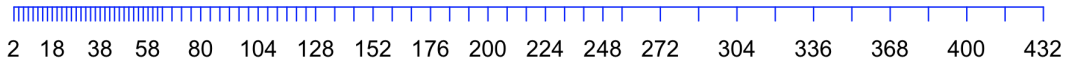


Figure 4.6.1: The adjusted scale set.

The activity measures are equivalently generated by the mean of the original wavelet coefficients, the absolute wavelet coefficients, the window variance of wavelet coefficients and the window variance of absolute wavelet coefficients. Also, the previous modelling showed that the pure training set leads to higher classification rates than the mixed training dataset except for the original wavelet coefficients. Also, it does not seem to be reasonable to apply the pure validation dataset after fitting a model using the mix training set. Therefore, here, we consider only the pure training dataset to simplify the problem.

4.6.1 Comments on Modelling from the Adjusted Scale Set

The purpose of the new modelling is to justify the method of generating the validation datasets and to compare the influence of the adjusted scale set focused on fine scales. Therefore, we will compare the prediction rates among the models using the pure validation dataset, the mixed validation dataset with the adjusted scale set in Figure 4.6.1 and the mixed validation dataset with the previous scale set in Figure 4.3.1. We define the models as Model 1, Model 2 and Model 3 for clear explanation. All models described in this section are fitted based on the same data.

In the same way as the previous modelling, the model variables are chosen by minimising classification rate and the maximum number of variables is fixed as 10. Figure 4.6.2 shows the classification rates on the test dataset from the four different activity measure models. As shown on the line plot, the order of accuracy is different between the models. First, the absolute value is the most accurate prediction for the model 2 and 3 with the mixed validation dataset, while it plunges to the worst classification rate below 0.5 on the mixed validation

dataset. Also, the classification results using window variance and absolute window variance are slightly better on both Model 1 and Model 2 than Model 3. In contrast, the mean model has poor result on the adjusted scale sets.

Overall, Model 2 and Model 3 show steadily good prediction results except for the original value model, and hence the mixed validation dataset seems to be reasonable for the classification model. Also, the adjustment of scales improves the classification rate on Model 2. It means that some coarse levels, which are chosen from the fully continuous wavelet scales, can explain the training dataset well, but it does not always match other practical data. For example, if the gas-flow regime changes within shorter time or more frequently than a training dataset, the choice of too coarse scales cannot reflect the transition on practical data.

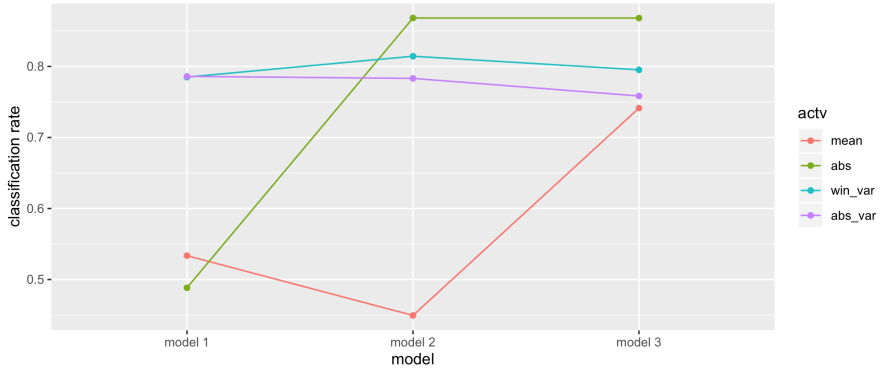


Figure 4.6.2: Classification rate for Model 1 (pure validation), Model 2 (mixed validation) and the Model 3 (mixed validation from the scale set in Figure 4.3.1).

However, since these test classification rates are calculated from only the 10-variable model, we need to consider that the test classification result can be different depending on the number of variables. Figure 4.6.3 shows the change of classification rates against the number of variables in the model. On Model 1 using the pure validation dataset, the absolute model has the highest prediction rate amongst the other one-variable models, but the classification rate plunges below 0.5 as another variables is added. In contrast, the window variance model with both the absolute wavelet coefficients and original wavelet coefficients fit well with a high classification rate around 0.8 and they tend to stay steady near 0.8. Also,

Model 2, which is fitted on the mixed validation set, seems to predict the test data well using the absolute value, the window variance of original and absolute wavelet coefficients. They tend to stay near 0.8 as the number of variables increases. In particular, the absolute value has the highest prediction rate of over 0.85 for the 10-variable model with a small increase each time a variable is added. The mean activity measure does not work well for this modelling, and the prediction rate does not have a monotone pattern.

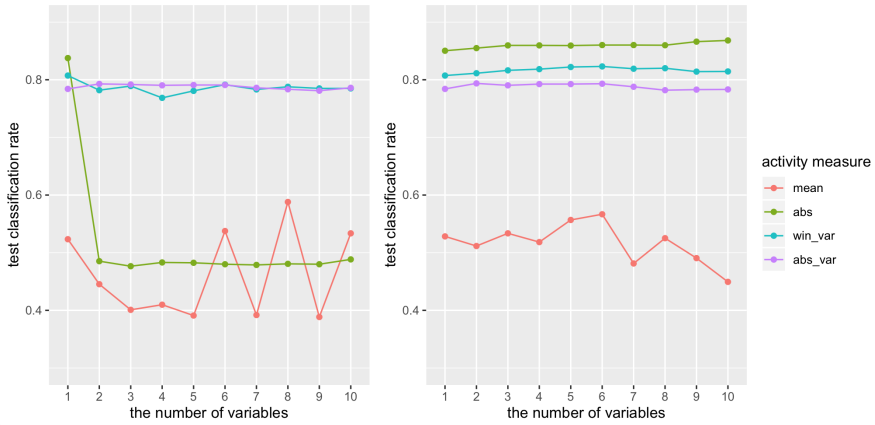


Figure 4.6.3: Classification rate against the number of variables: Model 1 (left) and Model 2 (right).

Accordingly, the above result shows that the absolute activity measure with a pure training set produces the best modelling using the mixed validation set. The prediction rate using the adjusted scale set does not change, but increases compared to the previously used scale set. This implies that the scale adjusting process can be beneficial by saving computation time while maintaining a similar level of accuracy.

Considering the change of accuracy against the number of variables, some models tend to stay steady without much changes. Although the stationarity means that variable addition is not very beneficial to classification, the classification rate remains at a similar level and can be useful especially for the gas-flow data modelling. As shown in the model results, the regression model, to classify the gas-flow type, relies a lot on training data especially in terms of variable selection. Also, the data modelling is fitted using only one training set, while

the vertical data analysis was performed using the 100 replicate modelling. The modelling on one specific training set does not give information on an ideal stopping time for variable selection, and hence it is better to choose a stable activity measure which is consistent with data or variable addition.

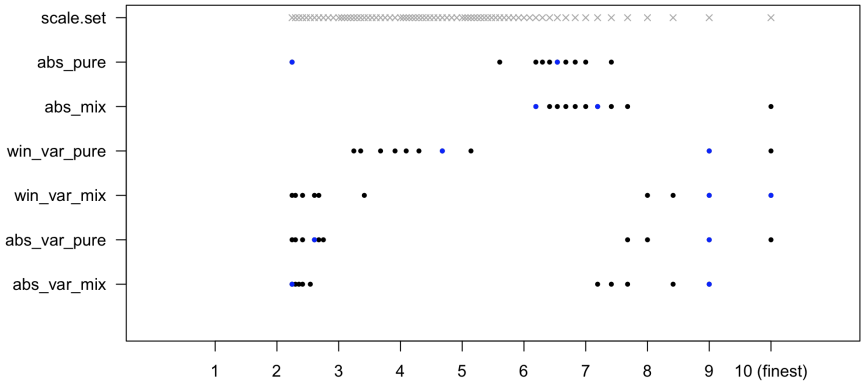


Figure 4.6.4: Variable selection on the adjusted scale sets. The first chosen two variables are highlighted by blue points.

Figure 4.6.4 shows the variable selection result for each activity measure type and validation set type. As the mean of the original value has poor classification rate on both validation types, it is not shown in this plot. The highlighted point with blue colour means the first two variables chosen, which is likely to have the most important information on classification. The variable selection patterns vary for each model. First, the absolute values prefer to choose variables between Level 6 and 8, and at least one variable out of the first two chosen variables belongs to the range. Also, the window variance activity measures choose the second finest variable within the first two steps on both validation processes, and therefore this fine scale has important wavelet information for the classification modelling. However, the window variance is calculated using the fixed width, 10, and hence the chosen scales are meant to the local variations of wavelet coefficients at the corresponding wavelet scales rather than the simple meaning of wavelet scales.

4.7 Conclusion

The horizontal modelling application shows the best result when using an absolute activity measure with the validation step, although it has stable result both on validated and non-validated models. In contrast, the previous vertical data application has the best performance on the window variance. We need to emphasise that the vertical data and horizontal data are observed under different conditions, including the direction of a pipe, and hence the generated data has different structures. According to the principle of the gas-liquid flow, the state of gas fractions is depending on not only conductivity, but other various elements such as a diagram of pipe, temperature and pressure. Therefore, we do not expect to have equivalent results, and we need to consider to establish individual models on each experiment.

It is satisfactory that the frequency characteristics are different between the different states and models for the wavelet coefficients were fitted reflecting the frequency information. However, the high classification rates are generally from the "bubble" and "plug" states. In particular, "bubble" is a considerably more easily classified variable with distinctly smaller amplitude than others. Also, the classification rate for "slug" is close to 0.33, that is equal to the random classification rate between three labels. Therefore, we need to consider more sophisticated modelling to investigate the information to distinguish between the "slug" and "plug" states. As one of the possible method, we can suggest a penalised logistic regression model. Our data analysis has done through finding the best subset of variables based on the validated prediction error, but penalised regression would suggest a sparse variable selection from the dense continuous scale set. The method is close to a shrinkage strategy than finding the best subset.

Also, the horizontal-flow data has the seven voltage datasets: three "bubble"s, three "plug"s and one "slug" to fit a regression model. This data has a relatively smaller number of observations for "slug" compared to the other two states, while the vertical-flow data has the equal number of observations from "bubble" and "churn". Conventional logistic regression modelling can struggle to get an accurate prediction from unbalanced samples, and therefore we need to consider

techniques to cope with the class imbalance. The mitigation of class imbalance is left for future work.

Chapter 5

Locally Stationary Wavelet (LSW) Process

5.1 Motivation

Stationarity is a fundamental concept in time series analysis. For a process to be stationary, we require that the mean is time-invariant and the auto-covariance is given as a function of lag rather than time. The detail of the stationary process was explained in §1.2.1. However, the conditions of the stationary process are too strict for practical data analysis in many cases, and hence more sophisticated modelling techniques are needed to analyse actual time series data. Accordingly, there have been many studies about non-stationary process to get realistic time series analysis. Here, we focus on the locally stationary process among the various kinds of non-stationary process modelling.

Locally stationary process is a specific kind of non-stationary processes with a few regularity conditions which are weaker than the stationary process. It is controlled in terms of the spectral density changing smoothly over time, while the stationary process has properties of the first and second moments independent with time. Dahlhaus (1997) and Dahlhaus (2000) defined the locally stationary process using the spectral density approach of Priestley (1965) :

$$X_t = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{+\pi} e^{(iwt)} A^0_{t,T}(w)\, d\xi(w), \tag{5.1.1}$$

where μ is the mean function of X_t and $A^0_{t,T}(w)$ is the transfer function. Also, this process is required to satisfy the following two conditions:

(i) $\xi(w)$ is a complex-valued Gaussian process on $[-\pi, \pi]$ with $\overline{\xi(w)} = \xi(-w)$, $\mathrm{E}(\xi\,(w))=0$ and $\mathrm{E}(d\xi_j(w_1)\; d\xi_k(w_2)) = \delta_{jk} \sum_l \delta\,(w_1 + w_2 + 2\pi l)\; d(w_1)\; d(w_2)$.

(ii) There exists a function $A(t/T, w)$ and a constant, $\boldsymbol{C_t}$ with $\sup|A^0_{t,T}(w) - A(t/T, w)| \leq \boldsymbol{C_t}/T$.

We assume that $\mu\,(t/T)$ and $A\,(t/T, w)$ are continuous in t/T. The second condition is designed to control the smoothness of the function, $A(t/T, w)$ approximately converging to the spectral density, $A^0_{t,T}(w)$. Dahlhaus (2012) describes the process as being close to a stationary process in a local sense with asymptotic properties.

Dahlhaus (2000) proposed a generalisation of the Whittle likelihood with approximation techniques for a locally stationary process. Also, Dahlhaus (2012) showed examples of locally stationary processes with the auto regressive (AR) and GARCH models and their uniquely defined spectral densities.

Mallat *et al.* (1998) proposed a model which is similar to the locally stationary process of Dahlhaus (1997) in that they describe a locally stationary process based on the spectral analysis approach. However, they focus on a short time interval by using a short-sized window function in which data are approximately stationary process rather than restricting the spectral density itself. Based on this approach, they establish a smooth supported function using cosine functions which are bounded by window functions. They investigate the time varying spectral characteristics using these windowed cosine functions, but we do not explain the details here, because we will follow the previously mentioned approach from Dahlhaus (1997).

Likewise, there has been a great deal of researches about the locally stationary process and its modelling for actual data not guaranteed to hold stationarity. Particularly, since signal processing has complicated data structures usually varying over time, the locally stationary process has been continuously highlighted in this field. For example, Qin *et al.* (2009) estimate the locally stationary spectrum from the modified version of the locally stationary process suggested in

Dahlhaus (1997) and fits a functional model based on the periodogram in an epileptic intracranial electroencephalogram (IEEG) dataset. They investigate the relationship between seizures and the sleep state based on the estimated time-varying spectra. Also, Birr *et al.* (2017) extended the locally stationary process to copulas using quantile approach with climate and financial example data.

Unlike the research mentioned above, we can consider wavelet functions as basis functions to construct a locally stationary process. The standard discrete wavelet transform is evaluated at the specific time points determined by the origin point and the corresponding levels. The decimation of the standard DWT enables us to have an orthogonal structure between discrete wavelet resolution levels. Meanwhile, the non-decimated wavelet transform, which is used for the discrete LSW process, allows the origin to be shifted to all possible time points for all discrete wavelet resolution levels. This non-decimated discrete wavelet transform has consistent results not relying on the origin of wavelet transformation, but the spaces spanned by the wavelet functions are redundant, and hence it is no longer orthogonal between the levels. It means that the derived wavelet coefficients cannot uniquely reconstruct the original data unlike the standard decimated DWT. Therefore, we should consider more conditions in order to have unique spectral densities for the locally stationary process with NDWT. The detailed definition and assumptions of that process was explained in Chapter 3.

5.2 The LSW Process with NDWT

Nason *et al.* (2000) introduced the locally stationary wavelet (LSW) process and showed the estimation of spectral densities using a few simple example functions and real ECG data. The LSW process is established using the NDWT following the definition of the locally stationary process from Dahlhaus (1997). The LSW process results in a spectral description of a non-stationary process using the concept of the locally stationary process in the wavelet frequency domain. In the same spirit as the locally stationary process, the amplitude is controlled by an assumption of smooth variation in the wavelet domain, and the assumption enables us to have an unique evolutionary wavelet spectrum.

5.2.1 Literature Review of the LSW Process

There has been continued research effort aimed to develop statistical theories using the LSW process and apply the LSW process to real data analysis. Knight *et al.* (2012) extend the LSW process and the estimation of its spectrum to the case with missing values. This article imposes the assumption of locally stationarity, and constructs the LSW process using the lifting scheme for irregular data or missing data. Cho & Fryzlewicz (2015) apply the LSW process to determine the consistency of their segmentation algorithm and suggest the model of multivariate LSW process. In addition to the aforementioned works with the theoretical progress using the LSW process, there are researches with applications to investigate the spectral characteristics for real non-stationary data assuming the LSW process. For example, actual data is difficult to assume stationarity, therefore less strict assumptions are needed for modelling than the stationary process. Wang *et al.* (2014) estimates the periodogram of the precipitation data by the LSW process. Wilson *et al.* (2019) also assumed that their acoustic data follows the multivariate LSW process of Park *et al.* (2014), and proposed an online dynamic classification model of the process to investigate irregularities.

5.2.2 The Standard Locally Stationary Wavelet (LSW) Process

Nason *et al.* (2000) defines the standard LSW process with the non-decimated discrete wavelet transform as follows.

Definition 5.2.1. *The standard* ***Locally Stationary Wavelet (LSW) process*** *is*

$$X_{t,T} = \sum_{j=-J}^{-1} \sum_{k=0}^{T-1} w^0_{j,k:T} \psi_{jk}(t) \xi_{j,k}, \tag{5.2.1}$$

where $\psi_{j,k}(t)$ is the t-th element of the non-decimated wavelet function vector, ψ_{jk}, shifted by k, and $\xi_{j,k}$ is a random orthonormal sequence for $t = 0, \ldots, T-1$. To fulfill the locally stationarity over the given wavelet basis functions, the following assumptions should be satisfied in terms of $w^0_{j,k:T}$ and $\xi_{j,k}$:

(a) $E\left(\xi_{j,k}\right) = 0.$

(b) $Cov\ (\xi_{j,k}) = \delta_{jl}\delta_{km}$.

(c) *There is a Lispchitz function,* $W_j(z)$ *satisfying the following properties for* $z \in (0,1)$ *and a Lipschitz constant,* L_j *corresponding to the function,* $W_j(z)$*, such that*

$$\sum_j |W_j(z)|^2 < \infty \quad and \quad \sum_j 2^{-j} L_j < \infty$$
$$w^0_{j,k:T} \to W_j(k/T) \qquad as\ T \to \infty. \tag{5.2.2}$$

Here, $\psi(x)$ is restricted to specific wavelet functions such as Daubechies' compactly supported wavelet functions. The amplitude, $w^0_{j,k;T}$, cannot be given as a unique value by itself due to the redundant basis functions. However, the condition *(c)* in (5.2.2) assumes that there exists the Lipschitz continuous function, $W_j(k/T)$ converging to the $w^0_{j,k;T}$. The Lipschitz continuity controls the first derivative of $W_j\,(k/T)$ not to change too quickly, and hence it allows to make an unique representation based on $W_j(k/T)$. This is the reason that the stochastic process, $X_{t;T}$ is called the LSW process.

The Lipschitz continuous function, $W_j(k/T)$, is defined over the rescaled variable, $k/T \in (0,1)$ instead of the time location, k. Accordingly, as more observations are added, that will improve the local structure by getting denser information over time.

Definition 5.2.2. *The* ***Evolutionary Wavelet Spectrum (EWS)*** *is the localised spectral density of the LSW process. It is defined by* $S_j(k/T) = |W_j(k/t)|^2$.

As mentioned before, the LSW process has the unique EWS as a representation of spectra. A proof of that can be found in Nason *et al.* (2000). The spectrum can be estimated from the wavelet coefficients of data.

Estimation of EWS

The wavelet coefficients of $X_{t,T}$ can be estimated through the non-decimated wavelet transform whose coefficients are defined as

$$d_{j,k} = \sum_t X_{t,T}\psi_{j,k}(t) \tag{5.2.3}$$

for the level, j and shift, k. Then, the wavelet periodogram, $I_{j,k}$, is defined as

$$I_{j,k} = |d_{j,k}|^2 \tag{5.2.4}$$

in terms of the wavelet coefficient, $d_{j,k}$ through NDWT. However, since the estimated wavelet periodogram is biased from the true EWS, $S_j(k/T)$, of the LSW process, we should consider a bias correction. The correction process is performed by the wavelet correlation function,

$$\Psi_j(\tau) = \sum_k \psi_{jk}(0)\psi_{jk}(\tau) \tag{5.2.5}$$

and the operator A,

$$A_{jl} = \sum_\tau \Psi_j(\tau)\Psi_l(\tau). \tag{5.2.6}$$

Let $I(z = k/T)$ be the vector of wavelet periodogram, $(I_{-1,k}, \ldots, I_{-J,k})$. Then, based on the correlation between wavelet levels, the corrected wavelet periodogram, $L(z)$, is given as

$$L(z) = A_J^{-1} I(z) \tag{5.2.7}$$

for $z = k/T$. The corrected spectrum is approximately unbiased to the EWS. Here, A_J is the operator A for the discrete LSW process at $T = 2^J$ and A^{-1} is the inverse matrix of this operator A_J.

5.3 The LSW Process with Other Wavelet Functions

The standard LSW process is defined in terms of the NDWT. In addition to the standard LSW process of §5.2.2, there are extensions of the LSW process to different types of wavelet functions or transformations. First, Cardinali & Nason (2017) suggest a model to expand the LSW process with wavelet packets. This process is defined in an equivalent way to the standard LSW process followed by the principles and terms used in Nason *et al.* (2000). However, wavelet packets

are given through all possible combinations of binary wavelet filters, while the non-decimated and decimated discrete wavelet transforms are performed through a selection of filters at each level. Therefore, the wavelet packet transform has a richer wavelet function dictionary than the NDWT used in the standard LSW process. To assess the redundancy of a locally stationary wavelet packet (LSWP) process, they apply a basis selection with cost functions. Also, Nelson & Gibberd (2016) extend the LSW process to the dual-tree complex wavelet function which does not belong to the Daubechies' wavelet family. Likewise, the principles and assumptions of the model are equivalently applied to the dual-tree wavelet transform. Although the wavelet functions are not placed in a compactly supported domain, the discrete wavelet transform is also built based on a binary wavelet filter, and hence the relationship between wavelet levels still can be explained by these filters on this process. In addition to the mathematically proven structure, the dual-tree wavelet transform includes directional information which cannot be given from the real-valued wavelet transform.

5.4 The Continuous LSW Process

5.4.1 Definition

Although the LSW process is useful to understand time-varying spectrum, as the number of discrete wavelet frequencies is determined by the number of observations, the spectral features given from the standard LSW process can be limited to a few discrete resolution levels. However, as we have seen in Chapter 3 and 4, there may exist valuable behaviour hidden between these discrete wavelet levels and this may be reflected in the spectrum. To cope with this problem, we propose the extension of the LSW process to the continuous wavelet scales for more flexible understanding of the spectral characteristics not limited to the discrete resolution levels. The range and number of scales in the CWT no longer rely on data, so we can choose any scales from the continuous domain in theory. In practice, since we work with the discretised CWT, we are limited to $n/2$ scales where n is the number of observations. We shall consider this in more detail later in section 5.4.2.

Definition 5.4.1. *Following the concept of the standard LSW process by (Nason et al., 2000), we define the* ***continuous locally stationary wavelet (CLSW) process****,* $\{X_t\}_{t=0,\ldots,T-1}$ *,* $T > 1$ *as*

$$\boldsymbol{C_t} = \int_{a \in \boldsymbol{A} \subset \mathbb{R}^+} \int_{b \subset \mathbb{R}} w_{a,b:T} \psi_{a,b}(t) \xi_{ab} \, dadb \tag{5.4.1}$$

where a *is the wavelet scale in the scale set,* $\boldsymbol{A}$*, on positive and continuous domain and* b *is the shifting parameter of the wavelet function,* $\psi_{a,b}(t)$*.*

We assume that

(a) $E\left(\xi_{ab}\right) = 0$

(b) $Cov\left(\xi_{ab}, \xi_{cd}\right) = \delta_{ab}\delta_{cd}$*.*

(c) There is the Lipschitz continuous function, $W_a(b/T)$ *satisfying the following properties for* $z = b/T \in (0,1)$ *and the Lipschitz constant,* L_a *corresponding to the function,* $W_a(b/T)$*.*

$$\begin{aligned} \sum_a \left|W_a\left(b/T\right)\right|^2 < \infty, \\ \sum_a a L_a < \infty \quad \text{and} \\ w^0_{a,b:T} \to W_a(b/T) \quad \text{as } T \to \infty \end{aligned} \tag{5.4.2}$$

In the same spirit as the LSW process using NDWT, local stationarity is imposed in terms of the amplitude, $w_{a,b:T}$. The Lipschitz continuity constrains the function, $W_a\left(b/T\right)$, converging to $w_{a,b:T}$, not to change too quickly. The CWT has a denser wavelet frequency set than the DWT, accordingly the CWT is also non-orthogonal like the non-decimated DWT. Therefore, the smoothness assumption, (5.4.2), is required to get an unique spectrum from the process. Also, by the re-scaled variable, $z = b/T$, as the number of time points increases, it implies that they have denser local information on the re-scaled domain, z.

Definition 5.4.2. *The* ***evolutionary wavelet spectrum (EWS)*** *of the CLSW process is defined as* $S_a(z) = \left|W_a\left(\frac{b}{T}\right)\right|^2$ *for* $z = b/T$*.*

Estimation of the EWS of the CLSW Process

The estimation and correction of the EWS are performed following the method of Nason *et al.* (2000). The coefficients given by the continuous wavelet transform of data, f, can be written as

$$d_{a,b} = \int f(t)\psi_{a,b}(t)dt \tag{5.4.3}$$

for the wavelet scale, a and the shifting parameter, b where $\psi_{a,b}(t)$ defined in equation (2.4.1). The EWS can be estimated by the squared coefficients for the CLSW process as

$$I_{a,b} = |d_{a,b}|^2. \tag{5.4.4}$$

However, as the periodogram is also biased from the true EWS, we should correct the periodogram to get an estimation close to the true spectrum. Following the property of the discrete LSW process, this biased estimator can be corrected by the local wavelet auto-correlation,

$$\Psi_a(\tau) = \int \psi_{ab}(0)\psi_{ab}(\tau)db \tag{5.4.5}$$

and the operator A,

$$A_{a_1,a_2} = \int \Psi_{a_1}(\tau)\Psi_{a_2}(\tau)d\tau. \tag{5.4.6}$$

Here, the wavelet auto-correlation and the operator A are defined as the integrals of the product of wavelet functions. Therefore, in the strict sense, we should consider an inverse of the operator A to correct the estimation. However, since we wish to apply a discretised CWT to data at discrete time points, the wavelet coefficients are evaluated as a discrete summation rather than an integral in our research, and in turn the wavelet auto-correlation and the operator A will be defined as

$$\Psi_a(\tau) = \sum_b \psi_{ab}(0)\psi_{ab}(\tau) \quad \text{and} \quad A_{a_1,a_2} = \sum_\tau \Psi_{a_1}(\tau)\Psi_{a_2}(\tau) \tag{5.4.7}$$

using discrete summations. Accordingly, the operator A will be a square matrix of size, $N_a \times N_a$, where N_a is the number of wavelet scales of the locally stationary

wavelet process, $\boldsymbol{C_t}$. Then, the inverse operator A is the inverse matrix of the operator A.

Based on these definitions, the corrected wavelet periodogram, $L(z)$, is

$$L(z) = A^{-1}I(z) \tag{5.4.8}$$

for $z = b/T$, where $I(z = b/T) = (CWT(a, b))_{a\in\boldsymbol{A}}$. The corrected spectrum is an approximately unbiased for the EWS in theory.

5.4.2 Unbiasedness of the Continuous EWS

The continuous wavelet frequency domain is denser than the discrete wavelet frequency domain. While the relationship between discrete wavelet levels is associated with the number of observations and the binary wavelet filter in the DWT, each successive level containing wavelets with twice the support of the one before, there is no such thing to explain the relationship between wavelet scales in the CWT. Also, as the fully continuous wavelet transform is calculated through an integral over continuous time, it must be approximated when using practical data usually given as a discrete sequence. Therefore, here we will evaluate a discrete version of the continuous wavelet transform and construct the corresponding CLSW process by discrete summation over a fine grid of continuous scales.

The Mean of the Estimated EWS

Proposition 5.4.1. $E\left[\hat{S}_{a_k}(z_k)\right] = \sum_{a\in\boldsymbol{A}} S_a(b_k/T)A_{a,a_k} + O(1/T)$

Suppose that the stochastic process, $\boldsymbol{C_t}$ is the CLSW process. Then, the process can be written as

$$\begin{aligned}\boldsymbol{C_t} &= \int_a\int_b w^0_{a,b;T}\psi_{a,b}(t)\xi_{ab}dadb \\ &\approx \sum_a\sum_b w^0_{a,b;T}\psi_{a,b}(t)\xi_{ab}\end{aligned} \tag{5.4.9}$$

by the definition of the CLSW process, (5.4.1) and its approximation using a discrete summation. For simplicity, the approximate CLSW process denoted via a

discrete summation is used as the formula of the CLSW process in the proof. The EWS of the CLSW process can be estimated from the squared wavelet coefficients, $|d_{a_*,b_*}|^2$, and therefore, the expectation of the estimated EWS, $\hat{S}_{a*}(z_* = b_*/T)$, is

$$\begin{aligned}
\mathrm{E}\left[\hat{S}_{a_*}(z_*)\right] &= \mathrm{E}\left[d^2_{a_*,b_*}\right] \\
&= \mathrm{E}\left[|< \boldsymbol{C}, \psi_{a_*,b_*} >|^2\right] \\
&= \mathrm{E}\left[\left(\sum_t \boldsymbol{C_t}\psi_{a_*,b_*}(t)\right)^2\right] \\
&= \mathrm{E}\left\{\sum_{t\in[0,T)}\sum_{a\in\boldsymbol{A}}\sum_{b\in[0,T)} w_{a,b;T}\psi_{a,b}(t)\psi_{a_*,b_*}(t)\xi_{ab}\right\}^2 \\
&= \sum_{a\in\boldsymbol{A}}\sum_{b\in[0,T)}(w_{a,b;T})^2\left\{\sum_{t\in[0,T)}\psi_{a,b}(t)\psi_{a_*,b_*}(t)\right\}^2, \qquad (5.4.10)
\end{aligned}$$

because $E\left(\xi_{a_i,b_j}\xi_{a_k,b_l}\right) = 0$ unless $i = k$ and $j = l$.

Substituting $b = c + b_*$,

$$\mathrm{E}\left[\hat{S}_{a_*}(z_*)\right] = \sum_{a\in\boldsymbol{A}}\sum_{c\in[-b_*,T-b_*)}(w_{a,c+b_*;T})^2\left\{\sum_{t\in[0,T)}\psi_{a,c+b_*}(t)\psi_{a_*,b_*}(t)\right\}^2 \qquad (5.4.11)$$

By assumption, (5.4.2), there exists the Lipschitz continuous function, $W_a\{(c+b_*)/T\}$, approximately converging to $w_{a,c+b_*;T}$, and by the definition of the Lipschitz continuous function, $|W_a\{(c+b_*)/T\} - W_a\{(b_*)/T\}| = O(c/T)$. Therefore,

$$\begin{aligned}
\mathrm{E}\left[\hat{S}_{a_*}(z_*)\right] &= \sum_{a\in\boldsymbol{A}}\sum_{c\in[-b_*,T-b_*)}\left[W_a\{(c+b_*)/T\}^2 + O(1/T)\right]\left\{\sum_{t\in[0,T)}\psi_{a,c+b_*}(t)\psi_{a_*,b_*}(t)\right\}^2 \\
&= \sum_{a\in\boldsymbol{A}}\sum_{c\in[-b_*,T-b_*)}\left\{W_a(b_*/T)^2 + O(c/T)\right\}\left\{\sum_{t\in[0,T)}\psi_{a,c+b_*}(t)\psi_{a_*,b_*}(t)\right\}^2 + O(1/T)
\end{aligned}$$

$$= \sum_{a \in \boldsymbol{A}} \sum_{c \in [-b_*, T-b_*)} W_a(b_*/T)^2 \left\{ \sum_{t \in [0,T)} \psi_{a,c+b_*}(t) \psi_{a_*,b_*}(t) \right\}^2 +$$
$$\sum_{a \in \boldsymbol{A}} \sum_{c \in [-b_*, T-b_*)} O(c/T) \left\{ \sum_{t \in [0,T)} \psi_{a,c+b_*}(t) \psi_{a_*,b_*}(t) \right\}^2 + O(1/T). \tag{5.4.12}$$

The Lipschitz continuous function, $W_a\{(c+b_*)/T\}$ can be replaced by $W_a(b_*/T)+O(c/T)$ based on the distance between these two function values. Also, since the squared sum, $\left\{\sum_{t\in[0,T)} \psi_{a,c+b_*}(t)\psi_{a_*,b_*}(t)\right\}^2$, is finite due to the compact support of the Haar wavelet function used in our research, the approximation term can be summed as the $O(1/T)$ independently with the discrete sum in (5.4.12). Therefore,

$$\mathrm{E}\left[\hat{S}_{a_*}(z_*)\right] = \sum_{a \in \boldsymbol{A}} \sum_{c \in [-b_*, T-b_*)} W_a(b_*/T)^2 \left\{ \sum_{t \in [0,T)} \psi_{a,c+b_*}(t) \psi_{a_*,b_*}(t) \right\}^2 + O(1/T).$$

Since the evolutionary wavelet spectrum, $S_a(b_*/T) = W_a(b_*/T)^2$, is independent of c,

$$\begin{aligned} \mathrm{E}\left[\hat{S}_{a_*}(z_*)\right] &= \sum_{a \in \boldsymbol{A}} S_a\{(b_*)/T\} \left\{ \sum_{c} \sum_{s \in [0,T)} \sum_{t \in [0,T)} \psi_{a,c+b_*}(t) \psi_{a_*,b_*}(s) \psi_{a,b_*}(t) \psi_{a_*,b_*}(s) \right\} + O(1/T) \\ &= \sum_{a \in \boldsymbol{A}} S_a\{(b_*)/T\} \left\{ \sum_{s \in [0,T)} \sum_{t \in [0,T)} \Psi_a(s-t) \Psi_{a_*}(s-t) \right\} + O(1/T) \end{aligned} \tag{5.4.13}$$

where $\Psi_a(t) = \sum_b \psi_{a,b}(0)\psi_{a,b}(t)$ is the wavelet correlation. Finally, substituting $v = s - t$, the expectation is

$$\mathrm{E}\left[\hat{S}_{a_*}(z_*)\right] = \sum_{a \in \boldsymbol{A}} S_a(b_*/T) \left\{ \sum_{v} \Psi_a(v) \Psi_{a_*}(v) \right\} + O(1/T)$$

$$= \sum_{a \in \boldsymbol{A}} S_a(b_*/T) A_{a,a_*} + O(1/T). \tag{5.4.14}$$

Let $A^{-1}_{a_*,a}$ be the element of the inverse A between two wavelet scales, a_* and a. According to the proof, (6.3.11), if the operator A is invertible, the corrected spectrum, $\sum_a A^{-1}_{a_*,a} \hat{S}_{a_*}(z_*)$, is approximately unbiased as the number of time points, T, is large enough.

The Invertibility of the Operator A

Here, we focus on building the CLSW process with the Haar wavelet function which is the simplest of Daubechies' compactly supported wavelet functions.

The wavelet correlation can be considered as a matrix given from the discretised continuous wavelet scale set.Therefore, the operator A in (5.2.6) can be written as a matrix calculation as $A = \Psi^T \Psi$, where Ψ is a $N_\tau \times N_A$ wavelet correlation matrix for the number of τ, N_τ and the dimension of scale set, N_A. The matrix A is a Gram matrix, and hence the operator A is

$$x^T A x = x^T \Psi^T \Psi x = (\Psi x)^T (\Psi x) \geq 0, \tag{5.4.15}$$

for any vector x. However, we would like the matrix to be positive definite, $(\Psi x)^T(\Psi x) > 0$ for any $x \neq 0$, so that the operator A is invertible. Therefore, we can claim the existence of an inverse A by proving $(\Psi x)^T(\Psi x) \neq 0$ unless $x = 0$.

Suppose that there is a CLSW process, $\boldsymbol{C_t}$ with the length of time series data, T. By the definition of the wavelet correlation, $\Psi(\tau) = \sum_b \psi(0)\psi(\tau)$, and the formula of the Haar wavelet function in (2.2.1), the wavelet correlation of the Haar wavelet function can be written as

$$\Psi(\tau) = \begin{cases} 1 - 3|\tau| & \text{for } |\tau| \in (0, 1/2] \\ |\tau| - 1 & \text{for } |\tau| \in (1/2, 1]. \end{cases} \tag{5.4.16}$$

In the CLSW process, we assume that the wavelet functions are compactly supported and fitted within the data, $(0, T]$. Also, since the Haar wavelet function has the same width of positive and negative values (as can be seen in Figure 2.2.1), a discretised sequence of points representing the Haar wavelet function

should have an equal number of points in the positive and negative section to keep the balanced properties of the Haar wavelet function. Accordingly, the largest continuous wavelet scale set, $\boldsymbol{a}$ for the CLSW process is given by the descretised the Haar wavelet functions of lengths $2, 4, \ldots, T$. Here, we define the largest scale set as the fully continuous wavelet scale set.

Then, the wavelet correlation for the scale, a is

$$\Psi_a(\tau) = \begin{cases} 1 - 3|\frac{\tau}{a}| & \text{for } |\tau| \in [0, a/2) \\ |\frac{\tau}{a}| - 1 & \text{for } |\tau| \in [a/2, a) \\ 0 & \text{otherwise,} \end{cases} \tag{5.4.17}$$

by the definition of a continuous wavelet function, $\psi_{ab}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)$.

First, the range of possible τ is from $(-\infty, +\infty)$ in general wavelet transform, but as our CLSW process with the Haar function has a compactly supported domain fitted in data, the range of wavelet correlation, Ψ_a, is also restricted within $(-T, T)$. Then, we can approach the proof algebraically with the matrix, $\boldsymbol{\Psi}$. Define the i-th row of the wavelet correlation matrix, Ψ as $\Psi[i,]$. Then, the first row consists of correlation values at $\tau = T - 1$, the largest possible τ. Here, only the first element of $\Psi[1,]$ has non-zero value at the coarsest scale, $a = T$, by (5.4.17), and hence x_1 satisfying $\Psi[1,]x = 0$ is 0 where $x = (x_1, \ldots, x_{N_A})$.

Next, move to the next rows, 2, having the first non-zero value for the second column. Then, x_2 should be zero as well when $\Psi[2,]x = 0$, because we already proved that $x_1 = 0$ and other elements are all zeroes except for the first and second elements by (5.4.17). In the same principle, we can observe each row by the order of columns from the coarsest to finest and $\Psi x = 0$ only if $x = 0$. Therefore, $xAx \geq 0$ for any $x \neq 0$, and hence the operator A of the CLSW process is invertible for the Haar wavelet function.

Apart from the Haar wavelet function, other Daubechies' wavelet functions can prove the invertibility of the operator A because they have compact supports. However, there are two practical difficulties of other Daubechies' wavelet functions. First, Daubechies' wavelet functions are constructed by through Mallat's pyramid algorithm at discrete resolution levels, and hence the functions do

not have a closed formula except for the Haar wavelet function. To evaluate the CWT with other functions, another technique is needed to get approximate wavelet function between discrete wavelet resolution levels. Also, the compact support is wider than the Haar wavelet function. Therefore, we should consider a larger number of $\tau's$ for the CLSW process.

5.5 Application of the CLSW Process

In this section, we will apply the CLSW process to a few example datasets. The aim of the applications is to estimate the EWS of the CLSW process and to compare the result to the standard discrete LSW process.

5.5.1 Reflected Doppler Data Analysis

The first example data used in this section is the reflected doppler data constructed based on the doppler function. As the doppler data has the interesting frequency characteristics changing over time, it is useful to compare the estimated spectral density from our model with the true frequency information. The reflected doppler data is created by flipping over the standard doppler function at the centre to prevent sudden discontinuous boundary effect in the spectral representation. Figure 5.5.1 shows the line plot of the reflected doppler data.

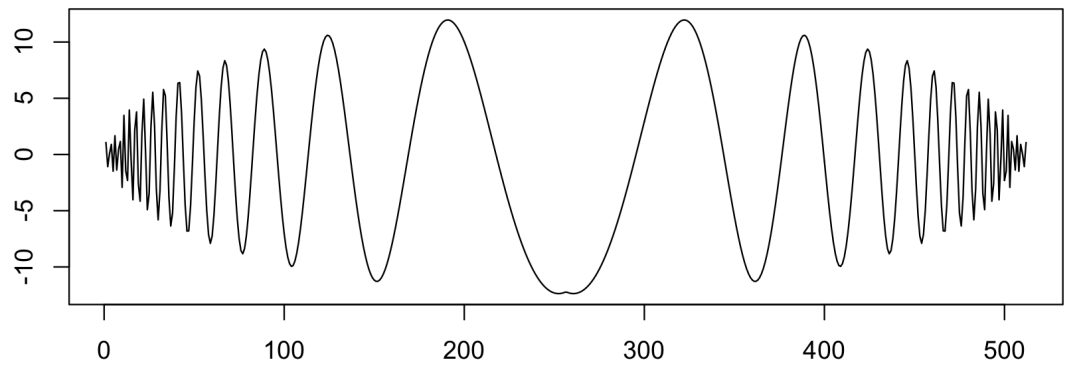


Figure 5.5.1: The reflected doppler data.

In our applications, we apply the Haar wavelet function to simplify the problem. Recall that the Haar wavelet function is the square shaped function with a discontinuous changing point at the middle. As mentioned previously, given data on a set of discrete time points, the CWT is replaced by a discrete calculation with a fine grid, which approximates the continuous calculation. Therefore, we need to consider which wavelet scales can perform well without distortion of their wavelet function structure.

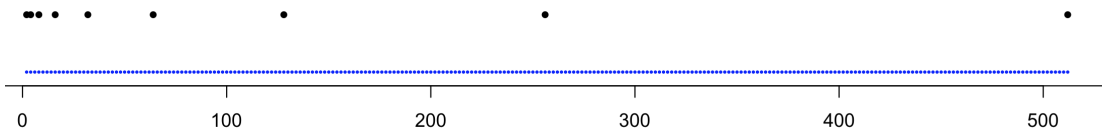


Figure 5.5.2: The comparison of discrete resolution levels (black dots) and fully continuous Haar wavelet scales (blue dots) used for the reflected doppler data in Figure 5.5.1.

Here, we use the fully continuous wavelet scale set for the CLSW process. The reflected doppler data used in this section has 512 observations, which correspond to the 9 discrete wavelet resolution levels and 256 continuous wavelet scales for the Haar wavelet function. These scale sets are shown in Figure 5.5.2.

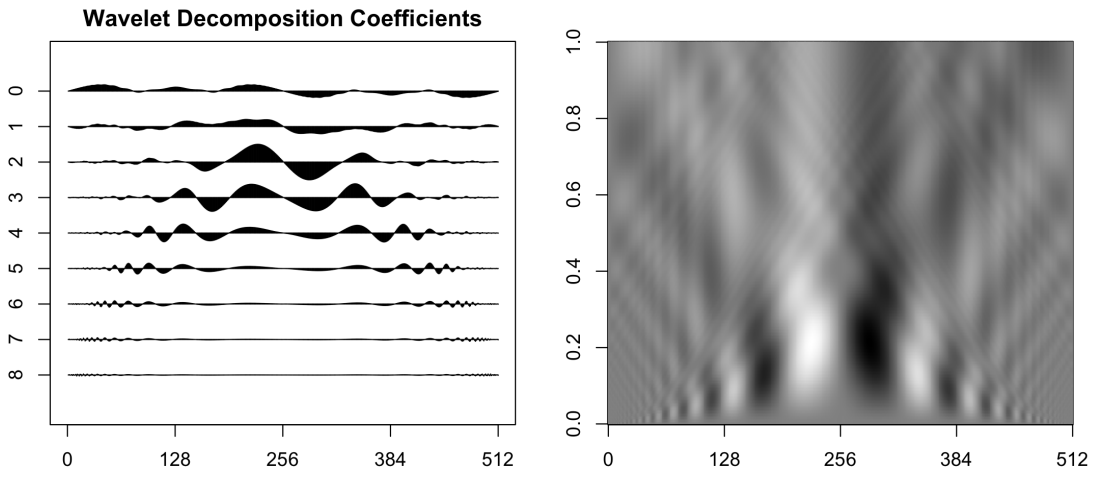


Figure 5.5.3: The comparison of wavelet coefficients from the reflected doppler data: the discrete wavelet coefficients (left) and the continuous wavelet coefficients (right).

All computing process for the DWT and discrete LSW process are performed by the R package, "wavethresh" (see Nason, 2016). Figure 5.5.3 shows the wavelet coefficients from the reflected doppler data on DWT and CWT. Although they are plotted in different formats, both show the equivalent patterns in that the main activity in the wavelet spectra tends to move to the middle where the peak point exists, as the scales get coarser.

The wavelet spectral density of the CLSW process can be estimated by the squared wavelet coefficients, but it needs to be corrected by the operator A. In §5.4.2, we proved that the operator A based on the Haar wavelet function is invertible, and hence we can correct the wavelet periodogram by (5.4.8). Figure 5.5.4 shows the comparison of the corrected estimate of the EWS between the discrete LSW process and the CLSW process. As the objective of our research is to find frequency information which is not captured well at the discrete wavelet resolution levels, they do not need to show equivalent results. However, it seems to be evident that the estimate EWS from the CLSW process does not follow the true frequency information that we expected from the data structure in Figure 5.5.1. In particular, there are some strong line effects moving to the right bottom corner from the middle, but the trends do not show in the discrete LSW process.

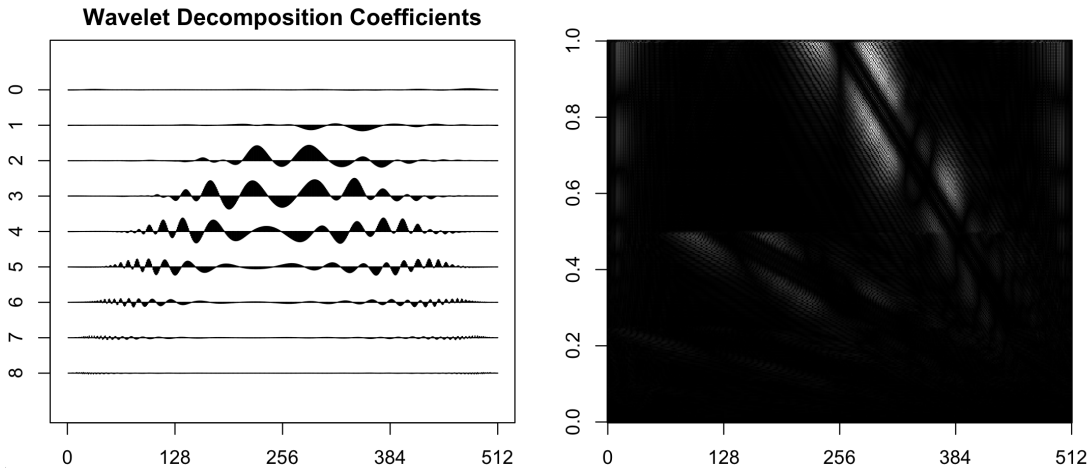


Figure 5.5.4: The comparison of estimated corrected EWS from the reflected doppler data : the discrete EWS (left) and the continuous EWS (right).

To investigate the reason of unsatisfactory results using the CLSW process,

we compare the operator A and its inverse between the discrete LSW and CLSW processes. Figure 5.5.5 shows the image plots of operator A's and their inverse matrices. The operator A created from the DWT and CWT have equivalent structures in that the values of A get larger when closer to the diagonal elements and at coarse scales or levels. Meanwhile, the inverse matrices show completely different patterns in that the CLSW process has larger diagonal elements as the scale gets coarser, in contrast to the trend of the discrete LSW process. The different structures are likely to affect the correction process as well.

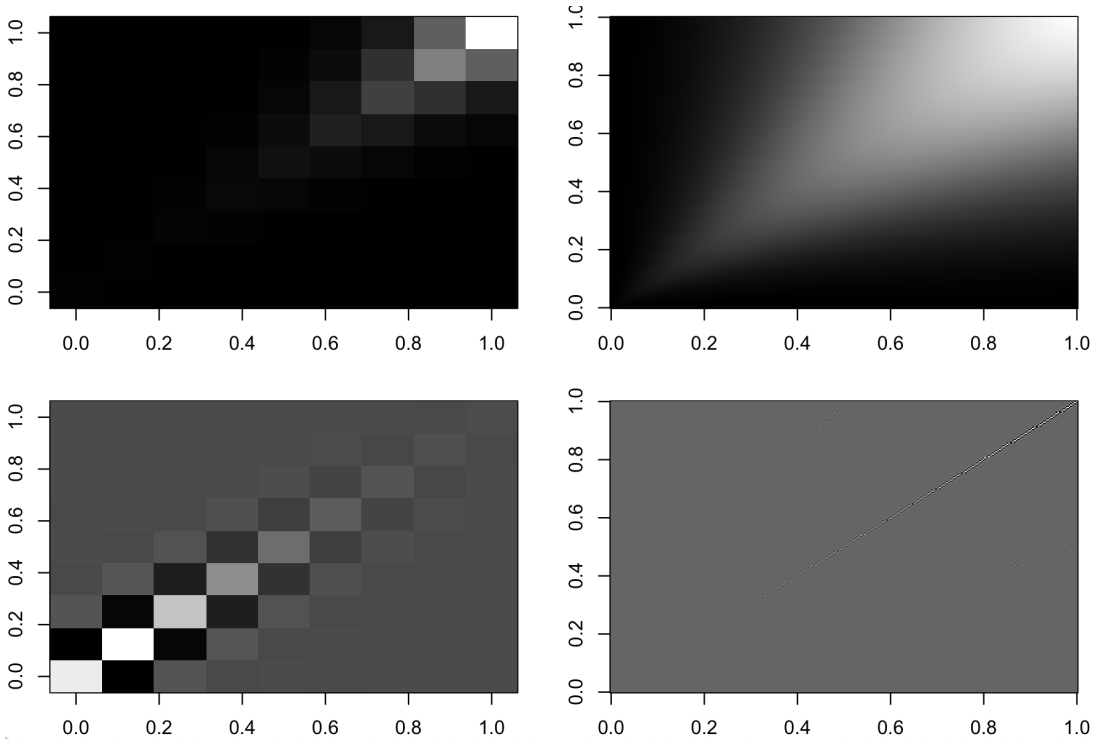


Figure 5.5.5: The comparison of the operator A and their inverse matrices for $N = 512$: the discrete wavelet operator A (top left), the continuous operator A (top right), the inverse of discrete wavelet operator A (bottom left) and the inverse of continuous operator A (bottom right).

5.5.2 Applications of the CLSW Process Using the Tomography Data

We will apply the idea of the CLSW process to the horizontal tomography data and compare that to the standard LSW process in terms of the estimated EWS. The horizontal tomography data is the same as the data used in §4 to establish a classification model based on wavelet coefficients. The details of data characteristics was explained at §4.

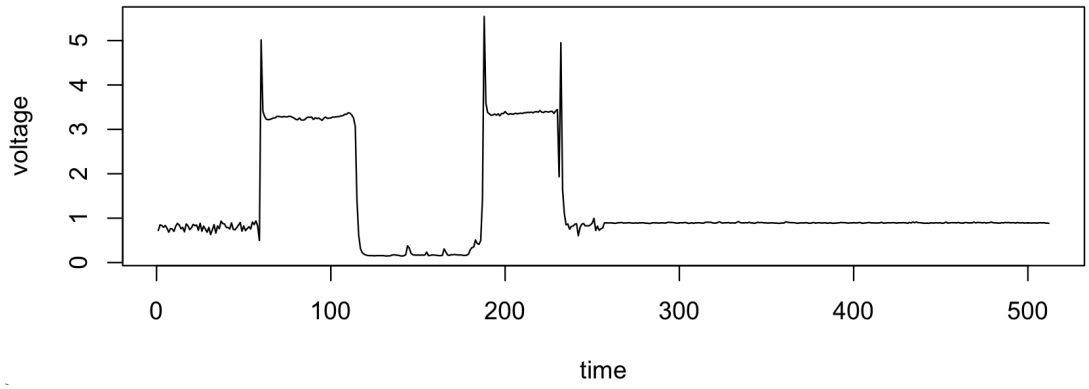


Figure 5.5.6: The plot of the horizontal tomography data.

The example data in Figure 5.5.6 shows the transition from "plug" flow, big and long lasting bubbles, to "bubble" flow, small and short lasting bubbles. The first half of the data has relatively wider lengths of oscillations, whereas the other half seems to stay steady around 1. Figure 5.5.7 shows the wavelet coefficients and the square root of the corrected estimate of the EWS from the discrete LSW and the CLSW processes. The wavelet coefficients indicate that the discrete wavelet levels, 2 and 3 have the highest contributions in representing the tomography data. The continuous wavelet transform also has significant power at scales between 0.1 and 0.3 which correspond to the discrete resolution levels, 2 and 3. When assuming they are from the LSW process, the estimated spectral densities from DWT still has high spectra at levels between 2 and 3, but their oscillating patterns over time are adjusted after correction. Meanwhile, it is difficult to find any frequency characteristics from the estimate of the EWS from our CLSW

process except a few vertical lines. This implies that our CLSW process do not generate appropriate spectral densities especially after correction.

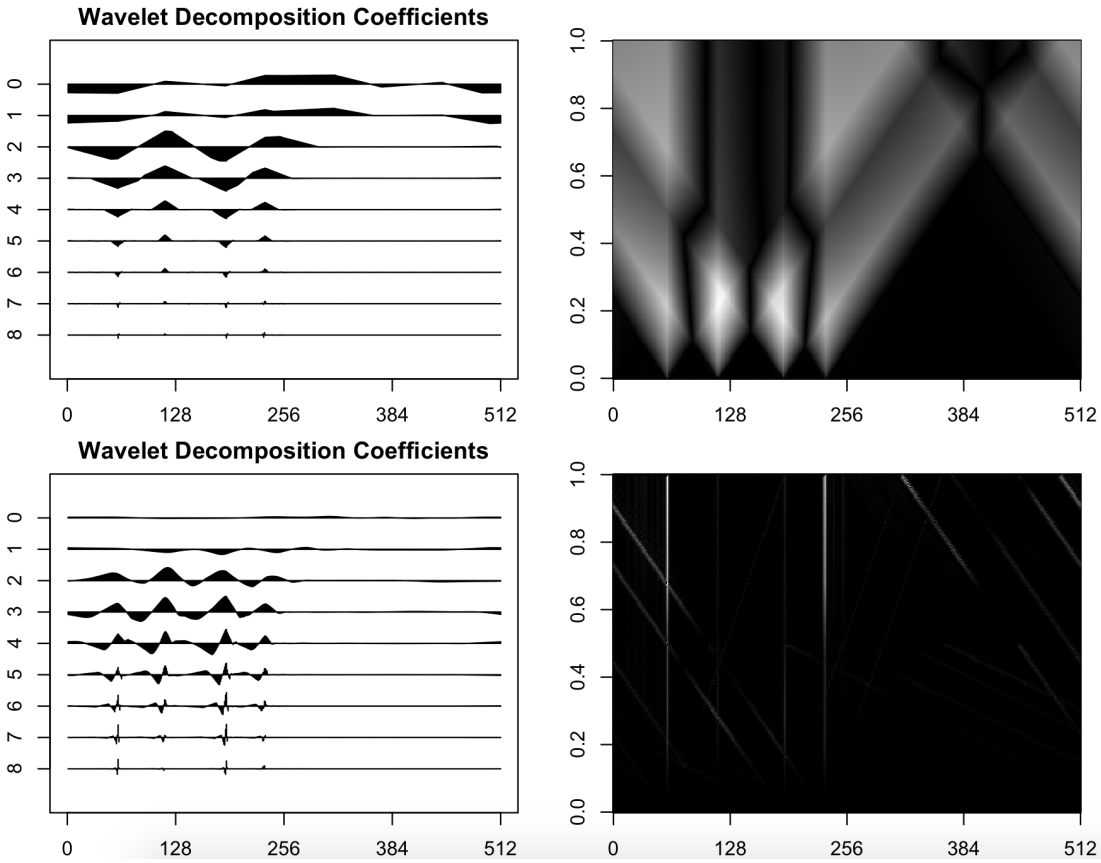


Figure 5.5.7: Locally stationary spectral analysis of one example of tomography data trace using the CLSW process: the discrete wavelet coefficients (top left), the absolute continuous wavelet coefficients (top right), the square root of the corrected estimate of the EWS from the discrete LSW process (bottom left) and the continuous LSW process (bottom right).

5.5.3 The Simulated CLSW Process Based on the Haar MA Process

Nason *et al.* (2000) used an example based on the moving average (MA) process with the Haar wavelet functions to generate the LSW process. In the same way, we generated a CLSW process based on the Haar MA process derived from not only the discrete wavelet resolution levels but also intermedi-

ate wavelet levels. The Haar MA process from the order, n, can be written as $x_t(n) = \frac{1}{\sqrt{2n}}(\epsilon_t + \ldots + \epsilon_{t-n+1} - \epsilon_{t-n} - \ldots - \epsilon_{t-2n+1})$, where ϵ is an i.i.d random variable with mean, 0, and variance, 1. Note that taking $n = 2^j$ for $j = 0, 1, \ldots$ corresponds to the Haar MA LSW process on discrete wavelet resolution levels.

We concatenated a series of the four Haar MA processes, $x_t(2), x_t(7), x_t(10)$ and $x_t(32)$ with 128 observations from each process and estimate spectral densities from 100 replicates. The Haar MA process can be considered as a stationary process when it is generated from a consistent frequency. However, here, since we pasted four individual MA processes derived from different frequencies, the entire stochastic process, which is a connected series of Haar MA processes, is non-stationary. Figure 5.5.8 shows the average of concatenated series from the Haar MA process, $x_t(2), x_t(7), x_t(10)$ and $x_t(32)$ with 128 observations from each process over 100 replicates. As can be seen from the plot, the frequency characteristics appear differently depending on the order of Haar MA processes.

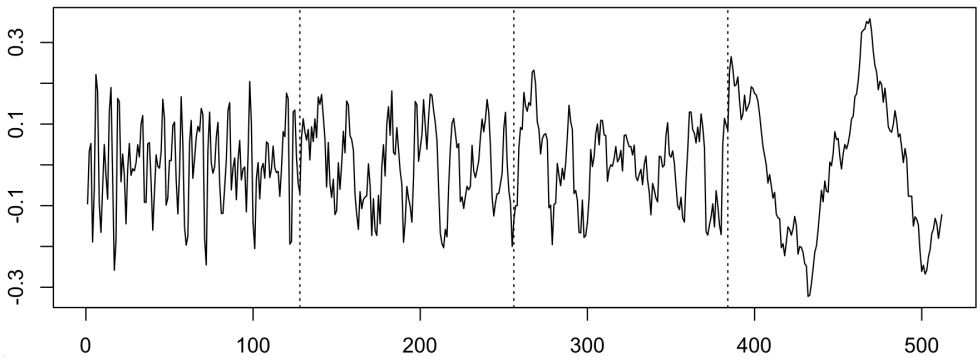


Figure 5.5.8: The concatenated series of the Haar MA process, $x_t(2), x_t(7), x_t(10)$ and $x_t(32)$ averaged over 100 replicates.

Figure 5.5.9 shows the square root of estimates of corrected EWS for the concatenated Haar MA process. The process has high estimated spectra on certain coarse wavelet scales, but these strong spectral estimates seem to be affected by the boundary effects of data rather than its original frequency features. Although the frequency orders for Haar MA process increase over time, this characteristic is not shown in the estimates except for the boundary effects.

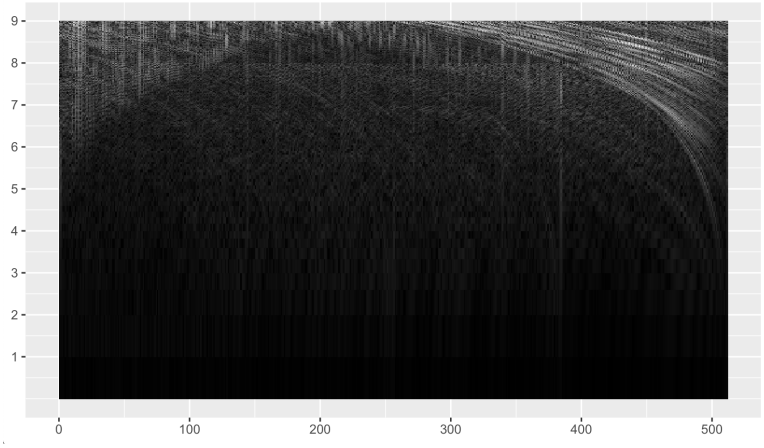


Figure 5.5.9: The averaged corrected estimate of evolutionary wavelet spectra over 100 replicates.

5.6 Limitation of the Process

5.6.1 The Change of Structure of the Inverse of the Operator A

In the DWT, the number of discrete wavelet resolution levels available, J, is given by the base-2 logarithm of the number of observations, $J = \log_2 N$. Therefore, as the number of observations increases, elements are added to the wavelet autocorrelation vector and the operator A next to the existing terms. The added information does not make much difference to the general structure of the inverse matrix, because the discrete wavelet functions are not strongly correlated between levels. Meanwhile, the continuous wavelet scale set can be considered as adding intermediate scales between discrete wavelet levels. As the wavelet correlation and the operator A are individually calculated at each scale, the intermediate scales seem to add local information between discrete wavelet levels without changing the general structure. However, to invert the operator A includes calculation between scales, and hence the highly correlated continuous wavelet scale set influences the structure of the inverse of the operator A. The changed structure results in a different spectral density of the CLSW process which does not look reasonable in §5.5.

5.6.2 The Lack of Assumptions over Scales

The scale set of the CWT is much denser than the NDWT or NDWP and the intermediate scales between discrete wavelet levels are highly correlated. However, the regularity conditions in our CLSW process do not consider the relationships between the dense wavelet scales. Also, the expectation of the estimated EWS in Proposition 5.4.1 is proven with the approximate error terms summed over the scales. The dense scale set of the CLSW process affects the asymptotic calculation by changing an $O(1/T)$ term to $O(c/T)$ in the expectation of the uncorrected spectral estimate in equation (5.4.12) when compared to the discrete LSW process. To identify the redundant scales and spectral estimate, we may need to add stronger assumptions over scales than the current CLSW process or to consider another technique to decrease correlation between scales.

Also, we can consider the connection between the redundancy of scales and the frame of the discretised CWT in (2.5.3). In our research, the CWT is computed through a discrete summation over regularly spaced fine grids and the scales are chosen from 0 to 1 to keep the structure of the Haar wavelet function. Let $\{A^{-1}\}_{a_1,a_2}$ be the element of the inverse A at the corresponding scales, a_1 and a_2. In the CLSW process, we can write the absolute sum of corrected estimates of EWS,

$$\sum_{a1}\sum_{a2,b}|\{A^{-1}\}_{a_1,a_2}\{CWT_f(a_2,b)\}^2| \tag{5.6.1}$$

$$\approx \frac{1}{c_a^2}\int_{a_1,a_2\in\boldsymbol{A}}\int_{b\in\mathbb{R}}|\{A^{-1}\}_{a_1,a_2}\{CWT_f(a_2,b)\}^2|da_1\,da_2\,db \tag{5.6.2}$$

where c_a is the distance between a's of fine grids of continuous wavelet scales. As we can assume from (5.6.2), that the corrected spectrum, (5.6.1), of the discretised CLSW process has a boundary relying on the operator function, $A^{-1}(a,b)$. In the rich scale set of CWT, the range of inverse operator A is very wide, and therefore it may have a value not close to the second norm of data, $||f^2||$. This shows that the larger scale set of CWT causes the wide frame bounds, and that may be related to the unexpected estimation from §5.5.

5.7 Conclusion

We have suggested a method to expand the locally stationary wavelet process with more flexible wavelet frequency. However, the applications to example datasets showed the imprecise representation of time-dependent frequency patterns compared to the standard LSW process despite the expensive computation costs. The results indicates that the current CLSW process cannot replace the discrete LSW process. The possible reasons of low resolution of the estimated EWS may be linked to the excessively rich scale set of CWT. The dense set of continuous wavelet scales can cause the limitations mentioned in §5.6; the change of structure of the inverse operator A and the lack of assumptions over scales. For these reasons, we need to add more conditions to modify the current version of the CLSW process, and we shall investigate one possibility in the following chapter.

Chapter 6

The Modified CLSW Process

6.1 Motivation

The previous applications of the CLSW process in §5.5 to the real and artificial example datasets show that our CLSW process with a fully continuous wavelet scale set derive spectra which do not match the data structure. The conditions assumed for the process do not fully reflect the redundancy of the high dimension of wavelet scales in the CWT, and therefore cannot correct the periodogram appropriately. It means that we can no longer guarantee the uniqueness of the evolutionary wavelet spectrum and the approximate unbiasedness of its estimator.

Accordingly, we should modify the CLSW process to obtain a more appropriate spectral analysis. We can consider two approaches to improve our CLSW process design. First, we can consider a decrease in the number of wavelet scales before fitting a CLSW process. This scale selection allows us to adjust the level of redundancy of the continuous wavelet scale set, and reduce the distortion effect between scales which seems to happen during the correction process. Alternatively, we can impose stronger assumptions on the smoothness of variation in the amplitudes, $w_{a,b:T}$ and $W_a(b/T)$ over not only time but also wavelet scales. The stronger regularities may adjust the smoothness of the spectral densities more strictly, and improve the correction process.

In our research, we focus on the first approach, using scale selection to alter the current CLSW process. This scale selection can be carried out in various ways based on a range of criteria and treating the full set of wavelet functions, $\psi_{a,b}$

as a basis dictionary from which to select the most "useful" elements. Coifman & Wickerhauser (1992) established an entropy-based approach to select a basis function from a waveform dictionary localised in time and frequency such as wavelet packets. The best basis selection is based on minimising the information cost of the given data based on Shannon's entropy. Apart from Shannon's entropy, other cost functions can be considered as a measure of basis selection such as log energy and thresholds (Wang *et al.*, 2011). Cardinali & Nason (2017) suggest cost functions to select the wavelet function library from the non-decimated wavelet packets to build a locally stationary wavelet packet process with a smaller number of basis functions.

Here, we modify the orthogonal matching pursuit algorithm to select the "best" continuous wavelet scales adaptively chosen to explain the data structure. We will modify the scale set for the CLSW process to get a subset of continuous wavelet scales adaptive to the data being analysed.

6.2 Matching Pursuit (MP) Algorithm

Our aim is to use scale selection to establish a CLSW process with a smaller subset of wavelet scales while retaining as much of the important frequency information. Therefore, although it is important to reduce the number of scales in the CLSW process, we should choose to retain wavelet scales which are the most strongly related to the data structure. There are various ways to investigate the contribution of scales to data structure, but here we apply a modified Matching Pursuit (MP) algorithm to reduce the scale dimension.

The Matching Pursuit (MP) algorithm was first introduced in Mallat & Zhang (1993). The algorithm suggests a way to choose a subset of basis functions from an complete and orthogonal function dictionary by minimising the l_2 norm. It selects the best subset of functions from a specific basis dictionary. The chosen functions can be used to make a linear representation of data. A detailed description of the MP algorithm will be given in §6.2.2.

6.2.1 Literature Review of the MP Algorithm

Based on the standard MP algorithm introduced in Mallat & Zhang (1993), there has been a great deal of research developing the algorithm and its applications. The MP algorithm is helpful when we wish to understand time-frequency features especially in signal processing data which contains various types of non-stationary time-frequency structure. Bergeaud & Mallat (1995) applied the MP algorithm to image data, showing that they can reconstruct images effectively from the selected waveform functions. Also, the MP algorithm extracts useful time-frequency information which is related to the contribution of the selected basis functions to the data, and this relationship can be used to fit a statistical model. In the application to the tomography data in Chapters 3 and 4, the variables are chosen from the continuous wavelet scales on the grounds of their effectiveness in the classification models. That variable selection process is similar in spirit to the MP algorithm in that they are chosen in terms of contribution of data to the state of gas-fraction. Christov *et al.* (2006) defined descriptors of their ECG data based on the waveform functions chosen by the MP algorithm and fitted a classification model to determine heartbeat states. Aharon *et al.* (2006) proposed a generalised K-means clustering method called the K-SVD model based on the MP algorithm. Based on their K-SVD model, Elad & Aharon (2006) showed how effectively MP-based classification modelling can be applied to denoise image data. Chu *et al.* (2009) analysed noise-like environmental sound data using the MP algorithm to investigate the time-frequency structure of their data. Also, the MP-based approach is used to build a HMM model in Bicego *et al.* (2004).

Along with various applications, there are many approaches to expanding the standard MP algorithm to a wide range of dictionaries. First, Pati *et al.* (1993) introduced the orthogonal matching pursuit (OMP) algorithm to perform the MP algorithm outside of an orthogonal basis function dictionary. Tropp (2004) and Tropp & Gilbert (2007) give mathematical proofs about the effective usage of the OMP algorithm for signal recovery. We shall modify the OMP algorithm to take an important role for scale selection process in our research. The OMP algorithm will be described after the standard MP algorithm. Chen *et al.* (2001) also built the representation of a linear combination from a subset of an over-complete

dictionary, but it is chosen by l_1 norm, while the standard MP algorithm uses the l_2 norm. Vincent & Bengio (2002) expanded the MP algorithm to a kernel-based dictionary from general time-frequency analysis, such as a wavelet transform. Their approach deals with the over-complete dictionary problem as well by using a pre-fitting method.

6.2.2 Definition of the MP Algorithm

The MP algorithm finds a signal representation as a linear expansion of waveform functions. The algorithm selects a subset of waveform functions out of a redundant dictionary, choosing those functions which best match the data. This is helpful to investigate the distribution of frequency information of the data.

The standard MP algorithm assumes that the basis dictionary used for the pursuit algorithm consists of orthogonal basis functions. Let the basis dictionary used in the MP algorithm be denoted by $\boldsymbol{D}$ and assume that basis functions, $x_i \in \boldsymbol{D}$, are orthonormal. The algorithm finds a linear decomposition of the data, f by projecting f on to p waveform functions from $\boldsymbol{D}$. Denote the kth function chosen as $x_k^* \in \boldsymbol{D}$ and let $R_k f$ be the residual at the kth step in the MP algorithm. Then, the first step in the MP algorithm is to decompose the data, f, by a function, x_1^* such that

$$f = R_1 f + < f, x_1^* > x_1^* \tag{6.2.1}$$

implicitly defining the first residual, $R_1 f$. Taking norms on each side of equations (6.2.1), we obtain

$$||f||^2 = ||R_1 f||^2 + |< f, x_1^* >|^2, \tag{6.2.2}$$

since x_1^* is orthogonal to the residual, $R_1 f$. Here, the first function chosen can be considered as the function best matching to the data, f by minimising the norm of the residual, $||R_1 f||^2$. By (6.2.2), minimising residual is the same process as maximising the norm of the inner product, $||< f, x_1 >||^2$, such that

$$|< f, x_1^* >| = \sup_{x_i \in \boldsymbol{D}} |< f, x_i >|. \tag{6.2.3}$$

In the dictionary, $\boldsymbol{D}$, we assume that all functions are orthogonal, and hence the l_2 norm conserves energy after the decomposition so that

$$||f||^2 = ||R_k f||^2 + \sum_{i=1}^{k-1} |< R_{i-1} f, x_i^* > \ x_i^*|^2 \tag{6.2.4}$$

$$= \sum_{x_i^* \in \boldsymbol{D}} |< R_{i-1} f, x_i^* > \ x_i^*|^2, \tag{6.2.5}$$

where we take $R_0 f = f$. Over p iterations, we can choose p waveform functions, $\{x_1^*, x_2^*, \ldots, x_p^*\}$, maximising $|< R_{k-1} f, x_i >|$ at the kth step, and this subset of $\boldsymbol{D}$ is the function set which best "matches" the data, f, as chosen by the MP algorithm.

6.2.3 Orthogonal MP (OMP) Algorithm

The standard MP algorithm explained in §6.2.2 assumes an orthogonal and complete dictionary. This means that all different waveform functions in $\boldsymbol{D}$ are linearly independent with zero inner products and all of the data can be expressed as a linear span of the functions in the dictionary, $\boldsymbol{D}$. Accordingly, the p waveform functions chosen by the standard MP algorithm, $\{x_1^*, \ldots, x_p^*\}$ are orthogonal to the pth residual of f, $R_p f$. However, in practice, there are various situations where we may wish to use non-orthogonal and over-complete function dictionaries. Define an over-complete dictionary as $\boldsymbol{D}_o$ and suppose the waveform functions, $x_i \in \boldsymbol{D}_o$ are not orthogonal for $i = 1, \ldots, n_1$ where n_1 is the number of functions in $\boldsymbol{D}_o$. Therefore, the part of f explained by waveform functions from $\boldsymbol{D}_o$ can overlap in the sense that we can no longer guarantee that the residual, $R_{k-1} f$ and the kth chosen function, x_k, are orthogonal. That means that the decomposition by the next function, $< R_k f, x_{k+1}^* >$ may have some correlation with the previous decompositions over the previous k steps. Accordingly, the

sum of $|< R_{i-1}f, x_i^* > \ x_i^*|^2$, is no longer a representation of data, f for $x_i^* \in \boldsymbol{D}$. This relationship should be considered to avoid choosing highly correlated basis functions, in which later selections do little to improve our understanding of the data.

To deal with the limitation of the MP algorithm, which does not support a non orthogonal function space, Pati *et al.* (1993) suggest the Orthogonal Matching Pursuit (OMP) algorithm. Denote the i-th function chosen by the OMP algorithm as x_i'. In the OMP algorithm, the linear expansion of data can be written as similarly to (6.2.1)

$$f = R_k f + \sum_{i=1}^{k} \alpha_i^k x_i', \tag{6.2.6}$$

with the new residual and coefficients for the chosen i functions, $\{x_1', \ldots, x_k'\} \in \boldsymbol{D_o}$ at the k-th step. But, the coefficient, α_i^k, no longer simply means the inner-product of the function, x_i', and data, f, or the i-th residual, $R_i f$.

For $k = 1$, the OMP algorithm is equivalent to the standard MP algorithm in that it chooses the function maximising the norm of the initial projection of the data. For $k \geq 2$, however, as the functions chosen through the OMP algorithm may be linearly dependent, the coefficients, α, need to be adjusted. These coefficients, α_i', are recalculated based on the geometry of the functions, x_i' selected over the previous k steps, and hence the term, α_i^k, is indexed by both i and k. Pati *et al.* (1993) dealt with the over-complete problem by interpreting the function added at the i-th step, x_i' in terms of the previously added functions, $\{x_1', \ldots, x_{i-1}'\}$. Let the space, V_i be the space linearly spanned by the functions until the ith step, W_i be the orthogonal space to V_i, and $V_i(x)$ and $W_i(x)$ be projection functions on to the spaces, V_i and W_i, respectively.

Suppose that there are $(k-1)$ functions already chosen by OMP. Then, at the kth step, the new variable, x_k', can be divided into two elements; the one part explained by the previously chosen variables, $\{x_1', \ldots, x_{i-1}'\}$ and the other part not related to them. It can be written as

$$x_k' = V_{k-1}(x_k') + W_{k-1}(x_k)$$

$$= \sum_{i=1}^{k-1} \beta_i^{k-1} x'_i + W_{k-1}(x_k) \tag{6.2.7}$$

by projection on to the spaces, V_{k-1} and W_{k-1}. The transformation in (6.2.7) adjusts the contribution of the function, x'_k to the data with regard to the linear dependencies with the previously chosen functions, $\{x'_1, \ldots, x'_{k-1}\}$. Here, $W_{k-1}(x_k)$ can be considered as the unique characteristics of x_k not explained by the previous function, $\{x'_1, \ldots, x'_{k-1}\}$. Based on the transformation of x'_k, the α in (6.2.8) is updated using

$$\alpha_i^k = \alpha_i^{k-1} - < R_{k-1} f, x'_k > \beta_i^{k-1} \tag{6.2.8}$$

for $i = 1, \ldots, k-1$ and

$$\alpha_k^k = \frac{< R_{k-1} f, x'_k >}{< W_{k-1}(x'_k), x'_k >}. \tag{6.2.9}$$

The k-th function is chosen by minimising the residual norm in equation, (6.2.6) based on the updated coefficients.

6.3 Scale Selections for the CLSW Process

6.3.1 The Modified CLSW process

We now present a modified CLSW process. The main difference between the modified CLSW process and the CLSW process described in Chapter 5 is that the modified process is defined with a subset of wavelet scales, while the previous CLSW process uses the full set of available continuous wavelet scales.

Denote the fully continuous wavelet function dictionary as $\boldsymbol{D_c}$ and let $\boldsymbol{A}$ be the set of the scales, a of all available wavelet functions in $\boldsymbol{D_c}$. Then, $\boldsymbol{D_c} = \{\psi_{ab}; a \in A, b = 0, \ldots, T-1\}$ The CLSW process defined in Chapter 5 is

$$\boldsymbol{C_t} = \int_{a \in \boldsymbol{A}} \int_{b \in [0,T)} w_{ab:T} \psi_{ab}(t) \xi_{a,b} da\, db, \tag{6.3.1}$$

where $\xi_{a,b}$ is the random element for the corresponding wavelet scale and shift parameter. The CLSW process are given from the full scale set, A, under the same time period, T. On the contrary, the modified CLSW process reduces the dimension of continuous wavelet scales by choosing a subset of A, and hence simplifies the structure of the process. Here, we will choose the subset of wavelet scales adaptively to best represent the data using the idea based on the OMP algorithm.

The OMP algorithm selects one individual function at each step, which means that the algorithm will suggest the best subset of individual wavelet functions rather than conducting scale selection that we require. However, our aim is to construct a locally stationary model which can be used to understand time-varying spectrum for all time points. Hence we require a process, where the activity at scale a, can be defined over all values the shift parameters, $b \in [0, T)$.

First, we could use the linear decomposition of data, $f(t)$ in terms of the wavelet functions chosen through the OMP algorithm given by

$$f(t) = \sum_{(a^*,b^*)\in(\boldsymbol{A_M},\boldsymbol{B_M})} \alpha_{a^*,b^*;T}\psi_{a^*,b^*}(t) + R_{n_a}f \tag{6.3.2}$$

at the stage, n_a. Here, $(\boldsymbol{A_M}, \boldsymbol{B_M})$ is the set of pairs of indices, (a^*, b^*) for the chosen functions, ψ_{a^*,b^*} and α is the updated contributions to the data from the corresponding wavelet functions through (6.2.8) at the final step, n_a.

The function selection by the OMP algorithm is likely not to cover the full time domain in each selected scale. Therefore, the shift parameter set, $\boldsymbol{B_M}$ may be sparse over time at each selected scale, $a^* \in \boldsymbol{A_M}$, and hence the spectra given from the sparse set are probably very rough due to the limited temporal information. Hence, to improve temporal resolution and give more accurate analysis over time, we will include all values of the shift parameter, $b \in [0, T)$ for each scale chosen through the OMP algorithm, $a' \in \boldsymbol{A_M} \subset \boldsymbol{D_c}$.

Definition 6.3.1. *The modified CLSW process, M_t is defined as*

$$\boldsymbol{M_{t;A_M}} = \sum_{a'\in\boldsymbol{A_M}} \sum_{b\in\{0,\ldots,T-1\}} w_{a'b:T}\psi_{a'b}(t)\xi_{a',b}, \tag{6.3.3}$$

with the subset of scales, $\boldsymbol{A_M}$ selected by the OMP algorithm.

Accordingly, the modified CLSW process has a smaller number of wavelet scales than (6.3.1), but still allows us to include non-discrete wavelet levels chosen to best match the data. The subset of scales relies on the data structure, therefore, the number of scales in $\boldsymbol{A_M}$ is not determined by the number of observations or time like the standard LSW and CLSW processes.

The regularity conditions for the modified CLSW process, (6.3.3), are equivalent to those for the fully CLSW process. The random elements, $\xi_{a',b}$ has the following statistical properties.

(a) $\mathrm{E}(\xi_{a',b}) = 0, \qquad \mathrm{Cov}(\xi_{a'_1,b_1}, \xi_{a'_2,b_2}) = \delta_{a'_1,a'_2}\delta_{b_1,b_2},$

for $a'_1, a'_2 \in \boldsymbol{A_M}$ and $b_1, b_2 \in [0, T)$. From the relations, the covariance of the modified CLSW modified process depends on elements only for the same paired indices, (a', b).

(b) There is a Lipschitz continuous function, $W_{a'}(b/T)$ satisfying

$$\begin{aligned} \sum_{a' \in \boldsymbol{A_M}} |W_{a'}(b/T)|^2 &< \infty \\ \sum_{a' \in \boldsymbol{A_M}} a'\, L_{a'} &< \infty \\ w_{a',b:T} \to w_{a'}(b/T) \quad \text{as } t \to \infty & \end{aligned} \tag{6.3.4}$$

for $b/T = z \in (0, 1)$ and the Lipschitz constant, L_a corresponding to the function, $W_a(b/T)$. In the same way as for the LSW process, the modified CLSW process controls the change of amplitude, $w_{a,b:T}$, over time by assuming the existence of the Lipschitz function, $w_{a,b:T}$ converging to $W_a(b/T)$.

The wavelet correlation, $\Psi_a(\tau)$, is also defined equivalently to that of the discrete and continuous LSW process as

$$\Psi_{a'}(\tau) = \sum_b \psi_{a'b}(0)\, \psi_{a'b}(\tau). \tag{6.3.5}$$

The correlation is the multiplication of two wavelets from the subset of full wavelet scales chosen through modification. As the discrete sum is calculated over all time points for the corresponding scales, the wavelet correlation can be considered as the subset of wavelet correlation from the full continuous wavelet domain. Based on the wavelet correlation, the operator A is also given as

$$A_{a_1',a_2'} = \sum_{\tau} \Psi_{a_1'}(\tau)\Psi_{a_2'}(\tau). \tag{6.3.6}$$

6.3.2 Evolutionary Wavelet Spectrum (EWS) and its Estimation

The evolutionary wavelet spectrum (EWS) from the modified CLSW process is defined in the same way as for the other LSW process followed by Nason *et al.* (2000). The evolutionary wavelet spectrum (EWS), $S_a(z)$, of the modified CLSW process is defined as $|W_a(b/T)|^2$. The spectrum can be estimated by the wavelet periodogram, $d^2_{a,b} = |< \boldsymbol{M}, \psi_{a,b} >|^2$, by the wavelet coefficients of the data with respect to the modified wavelet scale set, $\boldsymbol{A_M}$.

However, since the modified CLSW process, $\boldsymbol{M_{t;A_M}}$ of (6.3.3), also has random elements, $\xi_{a,b}$, the expectation of its estimate should be considered to obtain the statistical evidence as a useful spectral measurement. Through the proof, we will be able to see how close the estimate is to the true spectral characteristic of the CLSW process. The proof of the expectation of the wavelet periodogram here follows the proof for the discrete LSW process in Nason *et al.* (2000).

Proposition 6.3.1. $E\left[\hat{S}_{a_k}(z_k)\right] = \sum_{a\in \boldsymbol{A_M}} S_a(b_k/T)A_{a,a_k} + O(1/T)$.

Assume that the stochastic process, $\boldsymbol{M_{t:A_M}}$, is the modified CLSW process with the wavelet scale set, $\boldsymbol{A_M}$ through the scale selection process. According to the definition of the modified CLSW and process, (6.3.3), and the estimate of its EWS, $d^2_{a,b} = |< \boldsymbol{M}, \psi_{a,b} >|^2$, the expectation of the estimated EWS for the modified CLSW process for the given scale, a_k, and shift, b_k, can be written as

$$\begin{aligned} \mathrm{E}\left[\hat{S}_{a_k}(z_k)\right] &= \mathrm{E}\left[d^2_{a_k,b_k}\right] \\ &= \mathrm{E}\left[|< \boldsymbol{M}, \psi_{a_k,b_k} >|^2\right] \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}\left[\left(\sum_{t} \boldsymbol{M}_{\boldsymbol{t};\boldsymbol{A_M}} \psi_{a_k,b_k}(t)\right)^2\right] \\
&= \mathrm{E}\left\{\sum_{t\in[0,T)} \sum_{a\in\boldsymbol{A_M}} \sum_{b\in[0,T)} w_{a,b;T}\psi_{a,b}(t)\psi_{a_k,b_k}(t)\right\}^2 \\
&= \sum_{a\in\boldsymbol{A_M}} \sum_{b\in[0,T)} (w_{a,b;T})^2 \left\{\sum_{t\in[0,T)} \psi_{a,b}(t)\psi_{a_k,b_k}(t)\right\}^2 . \qquad (6.3.7)
\end{aligned}
$$

because $E(\xi_{a_i,b_j}\xi_{a_k,b_l}) = 0$ unless $i = k$ and $j = l$.

Substituting $b = c + b_k$,

$$
\mathrm{E}\left[\hat{S}_{a_k}(z_k)\right] = \sum_{a\in\boldsymbol{A_M}} \sum_{c\in[-b_k,T-b_k)} (w_{a,c+b_k;T})^2 \left\{\sum_{t\in[0,T)} (\psi_{a,c+b_k}(t)\psi_{a_k,b_k}(t)\right\}^2 . \qquad (6.3.8)
$$

By assumption, (6.3.4), there exists a Lipschitz continuous function, $W_a\{(c+b_k)/T\}$, approximately converging to $w_{a,c+b_k;T}$. Then, the expectation can be written in terms of the Lipshitz function, $W_a\{(c+b_k)/T\}$ as

$$
\begin{aligned}
\mathrm{E}\left[\hat{S}_{a_k}(z_k)\right] &= \sum_{a\in\boldsymbol{A_M}} \sum_{c\in[-b_k,T-b_k)} \left[W_a\{(c+b_k)/T\}^2 + O(1/T)\right] \left\{\sum_{t\in[0,T)} \psi_{a,c+b_k}(t)\psi_{a_k,b_k}(t)\right\}^2 \\
&= \sum_{a\in\boldsymbol{A_M}} \sum_{c\in[-b_k,T-b_k)} \left\{W_a(b_k/T)^2 + O(c/T)\right\} \left\{\sum_{t\in[0,T)} \psi_{a,c+b_k}(t)\psi_{a_k,b_k}(t)\right\}^2 + O(1/T) \\
&= \sum_{a\in\boldsymbol{A_M}} \sum_{c\in[-b_k,T-b_k)} W_a(b_k/T)^2 \left\{\sum_{t\in[0,T)} \psi_{a,c+b_k}(t)\psi_{a_k,b_k}(t)\right\}^2 \\
&\quad \sum_{a\in\boldsymbol{A_M}} \sum_{c\in[-b_k,T-b_k)} O(c/T) \left\{\sum_{t\in[0,T)} \psi_{a,c+b_k}(t)\psi_{a_k,b_k}(t)\right\}^2 + O(1/T). \qquad (6.3.9)
\end{aligned}
$$

The function, $W_a\{(c+b_k)/T\}$ can be substituted as $W_a(b_k/T)+O(c/T)$ based on the distance between these two function values. Also, since the squared sum,

$\left\{\sum_{t\in[0,T)} \psi_{a,c+b_k}(t)\psi_{a_k,b_k}(t)\right\}^2$, is finite due to the compact support of the Haar wavelet function used in our research, the approximation term can be summed as the $O(1/T)$ independently with the discrete sum in (6.3.9).

$$\mathrm{E}\left[\hat{S}_{a_k}(z_k)\right] = \sum_{a\in\boldsymbol{A_M}} \sum_{c\in[-b_k,T-b_k)} W_a(b_k/T)^2 \left\{\sum_{t\in[0,T)} \psi_{a,c+b_k}(t)\psi_{a_k,b_k}(t)\right\}^2 + O(1/T).$$

Since the evolutionary wavelet spectrum, $S_a(b_k/T) = W_a(b_k/T)^2$, is independent of c,

$$\begin{aligned}\mathrm{E}\left[\hat{S}_{a_k}(z_k)\right] &= \sum_{a\in\boldsymbol{A_M}} S_a\left\{(b_k)/T\right\} \left\{\sum_{c}\sum_{s\in[0,T)}\sum_{t\in[0,T)} \psi_{a,c+b_k}(t)\psi_{a_k,b_k}(s)\psi_{a,b_k}(t)\psi_{a_k,b_k}(s)\right\} + O(1/T) \\ &= \sum_{a\in\boldsymbol{A_M}} S_a\left\{(b_k)/T\right\} \left\{\sum_{s\in[0,T)}\sum_{t\in[0,T)} \Psi_a(s-t)\Psi_{a_k}(s-t)\right\} + O(1/T)\end{aligned} \tag{6.3.10}$$

where the wavelet correlation, $\Psi_a(t) = \sum_b \psi_{a,b}(0)\psi_{a,b}(t)$. Finally, substituting $x = s - t$, the expectation is

$$\begin{aligned}\mathrm{E}\left[\hat{S}_{a_k}(z_k)\right] &= \sum_{a\in\boldsymbol{A_M}} S_a(b_k/T)\left\{\sum_{v} \Psi_a(v)\Psi_{a_k}(v)\right\} + O(1/T) \\ &= \sum_{a\in\boldsymbol{A_M}} S_a(b_k/T)A_{a,a_k} + O(1/T),\end{aligned} \tag{6.3.11}$$

where the operator A, $\boldsymbol{A}_{a_1,a_2} =< \Psi_{a_1}, \Psi_{a_2} >$. Supposing that we have a large enough number of observations, T, we can ignore the terms related to the approximation, $O(c/T)$ and $O(1/T)$ in (6.3.7). However, the approximation of the expectation is still biased from the true EWS, $S_a(b_k/T)$ with the elements of the operator A. Hence, the bias in the estimate can be corrected using the operator A to give an unbiased estimate, $\tilde{S} = A^{-1}|d^2|$.

Also, as the modified process, $\boldsymbol{M_{t,A_M}}$ is defined from the smaller scale set, $\boldsymbol{A_M}$, the sum of these approximation terms is likely to have less variation around

the expectation. This implies that the periodogram of $\boldsymbol{M}_{t;\boldsymbol{A}_M}$ may converge more stably to the unbiased EWS compared to that of the fully CLSW process, $\boldsymbol{C}_t$ after appropriate correction.

6.4 The Applications of the Modified CLSW Process

6.4.1 The Reflected Doppler Data

The reflected doppler test function was applied to the fully CLSW process in §5.5.1, but the estimated spectra did not seem to reflect the frequency characteristics of the data. We now compute the modified CLSW process for the same reflected doppler test function in this section. The default scale set and the wavelet functions are defined equivalently to the previous applications in §5.5.1. The modified scale set is chosen by using the OMP algorithm from the fully continuous wavelet scale set in Figure 5.5.2 with the Haar wavelet functions. The OMP algorithm is computed using the R package, "Rfast" (Papadakis *et al.*, 2020).

The OMP algorithm allows us to choose the basis functions which best match a given dataset, and the scale used in the CLSW process is determined from the wavelet functions selected by the OMP algorithm. Therefore, the scale set used in the modified CLSW process is considered as being adaptive to the data structure.

In the "Rfast" package, there are two options of criteria to optimise in the OMP algorithm, SSE and BIC. The number of iterations is controlled by a tolerance value. The criteria, BIC, is explained in §1.2.2. Here, the tolerance value means the minimum difference value between two successive steps for the algorithm to continue. In our research, we will apply both criteria to compare which one gives a better representation of the spectral characteristics.

We present three different models defined by the choice of optimisation criteria and tolerance values to compare the performance. The first model is the BIC model with the tolerance value of 2, and the second and third models are chosen by SSE and BIC respectively with the tolerance value, 0. The SSE model with the tolerance value, 2 is not shown as in that case no function were chosen for

the reflected doppler test function. Figure 6.4.1 shows the scale-location wavelet functions chosen by the OMP algorithm for the reflected doppler test function.

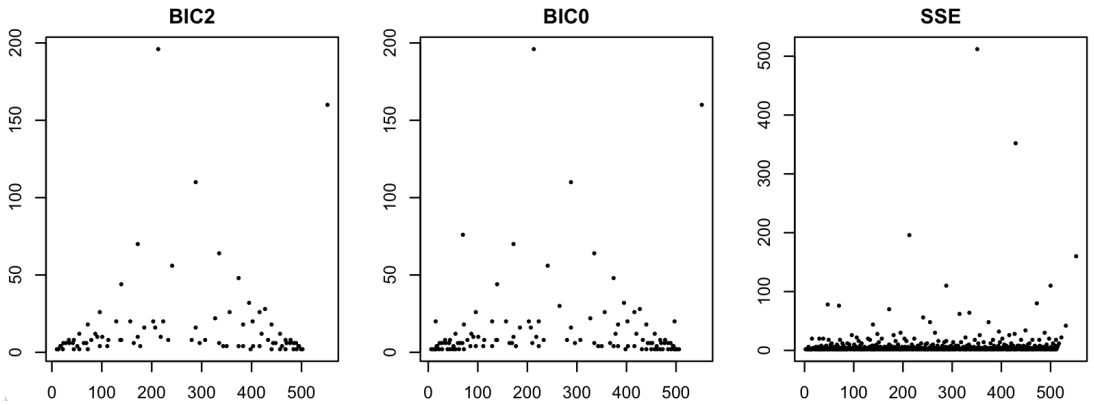


Figure 6.4.1: The scale-location pairs chosen by the OMP algorithm for the reflected doppler test function.

First, the SSE model tends to choose the functions mainly from very fine scales and they are spread over the entire time. On the contrary, the BIC-based models seem to select the functions in the time coordinates where the signal has the corresponding frequency characteristic. However, as the number of the chosen wavelet functions is different in each model, we cannot compare the patterns only by the function locations. The SSE-based model chooses 510 functions, while the Model BIC2 and BIC0 have 89 and 114 functions respectively, which are relatively sparse compared to Model SSE. This implies that the SSE-based model is more inclined to add functions than the BIC-based models and that leads to include a wider range of time particularly on fine scales. The BIC-based models generally tend to be more sensitive to the data structure with a smaller number of functions. For example, the reflected doppler test function has high-frequency oscillations at the beginning and end of data, and this data structure seems to affect the BIC-based scale selection at the fine scales. Based on the individual function selection using the OMP algorithm, the modified scale sets, which includes all the time points for the corresponding wavelet scales, are shown in Figure 6.4.2.

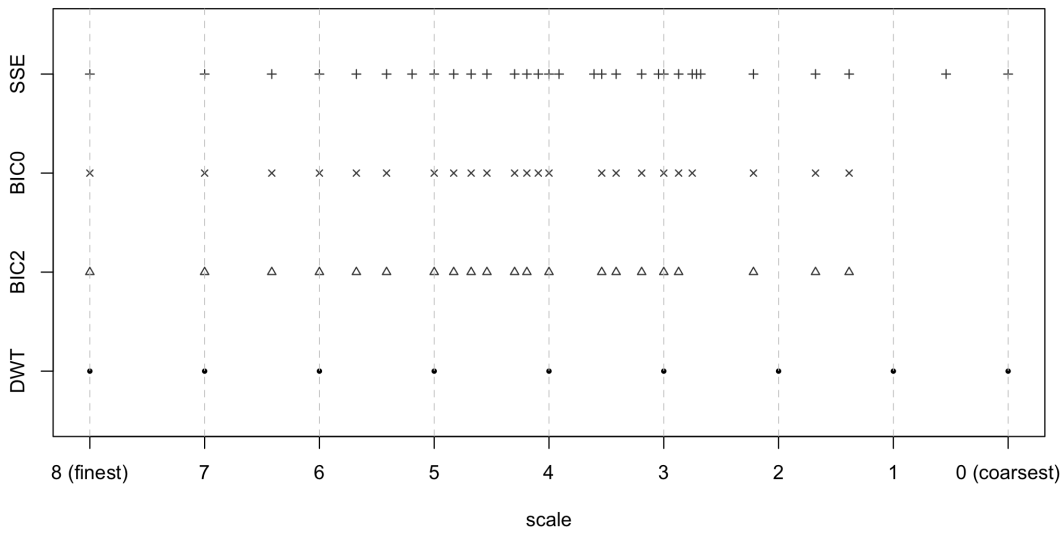


Figure 6.4.2: The scale chosen by the OMP algorithm from the reflected doppler test function.

The numbers of scales are 21, 31 and 23 respectively for Model BIC2, SSE and BIC0. The scale set on the bottom in Figure 6.4.2 is the discrete wavelet resolution levels to compare the scale selection for our modified CLSW models. The SSE-based model has a wider range of scales than the BIC-based models with the coarsest scale. The difference between Model BIC2 and BIC0 is the size of tolerance values for the stopping time of the OMP algorithm, and hence it is logical to have very similar scale sets in these two models. Although there are little differences between the three models, the modified scales sets include intermediate scales for the CLSW process.

Figure 6.4.3 shows the corrected estimate of the EWS for the reflected doppler test function. The plotted periodograms are transformed as the square root of the absolute estimate of the EWS to see the image plot more clearly. In this thesis, we plot the EWS estimates on the same colour scale where comparison across multiple LSW processes is drawn. The estimated EWS is not as smooth as the continuous wavelet coefficients in Figure 5.5.3, but the oscillations shown in the estimated spectra are matched to the doppler test function structure better than the fully CLSW process. In general, the corrected spectral estimates tend to get stronger when closer to the centre, as the scale gets coarser. In particular,

the models have distinctly high spectral estimates between the scale, 2.5 and 3.5, which corresponds to the low-frequency components in the middle.

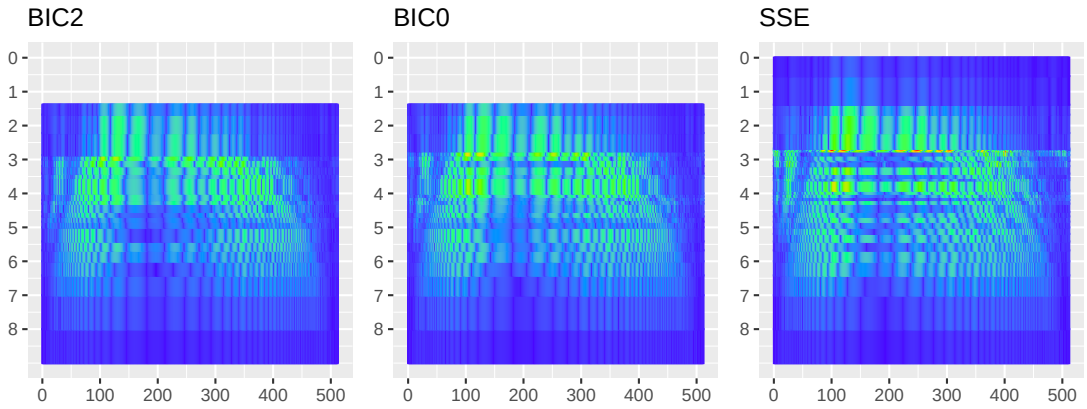


Figure 6.4.3: The image plot for the corrected estimate of the EWS from the modified CLSW process for the reflected doppler test function. They are plotted as the squared root of corrected periodogram.

The image plots are helpful to see the general patterns of the corrected estimate of the EWS visually, but it is hard to see subtle changes without inspecting. To see more detailed patterns of spectral estimates, Figure 6.4.4 and 6.4.5 show line plots of the corrected estimate of the EWS on each scale from one example model. Model BIC2 has relatively smooth spectral densities with a small number of wavelet scales, so this model will be used for the line plots. The corrected estimates of the EWS are plotted by different scaling. Figure 6.4.4 is the spectral estimate standardised by the one scale factor chosen over the entire scales, while Figure 6.4.5 shows the spectral estimate standardised by the scale factor chosen from each scale. Equivalently to the pattern seen from Figure 6.4.3, the overall spectral densities tend to be stronger, when closer to the center and coarser scales.

The notable scales on the lines plots are Scale 56, 64 and 70 with very high spectral densities. Meanwhile, the fine scales from 2 to 12 are densely chosen for Model BIC2, but the relative power of these scales are weak compared to the other scales. Also, the spectral estimates on the first three coarsest scales, Scale 110, 160 and 196, are not that strong, but it can be considered as a spectral representation from the low frequency characteristics in the middle.

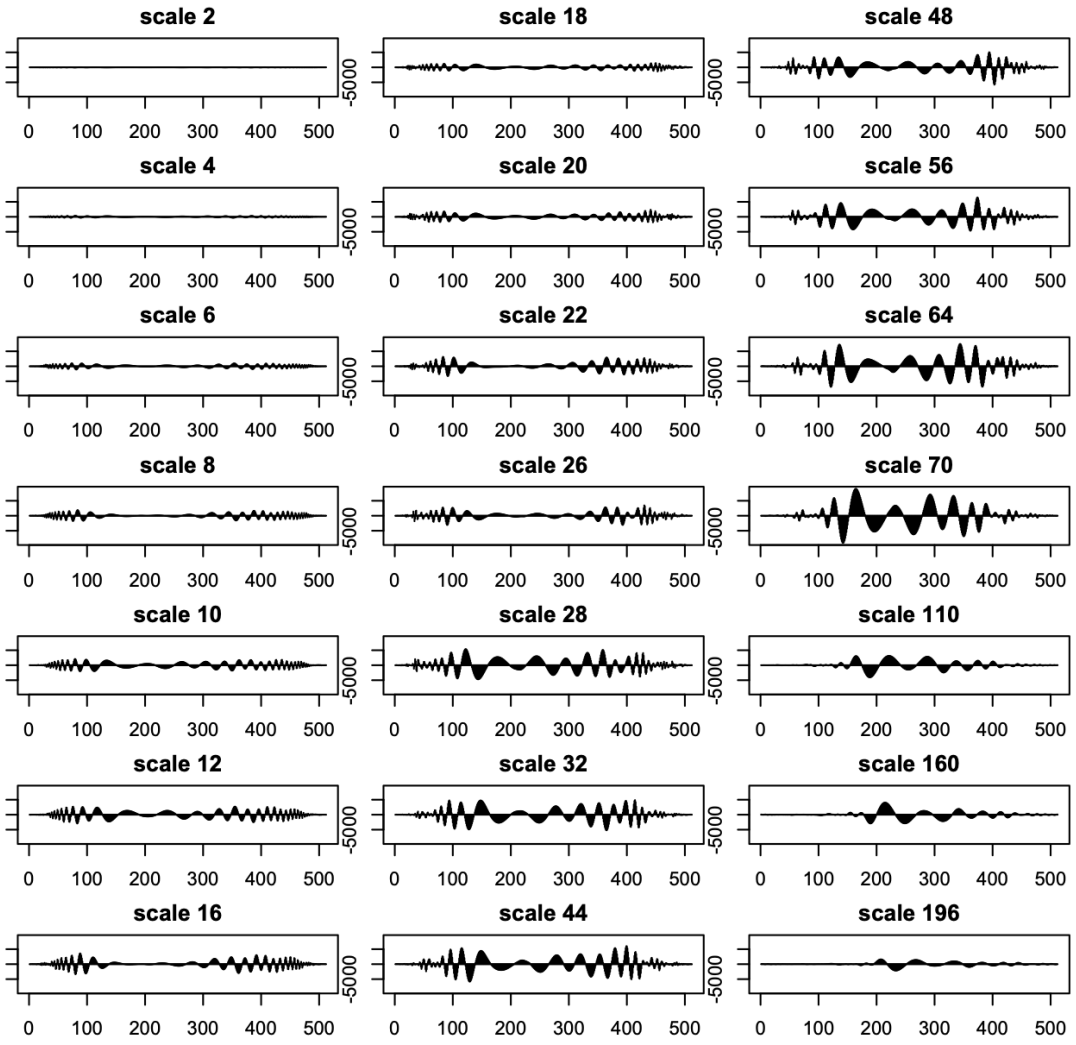


Figure 6.4.4: The global-scaled line plots for the corrected estimate of the EWS from the modified CLSW process for the reflected doppler test function using the BIC2 model.

The plots, standardised by each scale in Figure 6.4.5, shows how the estimated spectra change within each scale. The fine scales between Scale 2 and 6 have relatively low spectral densities compared to the other scales in Figure 6.4.4 due to the low amplitudes, but show rapid oscillations of the estimated spectrum in Figure 6.4.5. There are some unexpected patterns in the middle on those fine scales, but the absolute values are relatively small compared to the spectra of Scale 110 and 160.

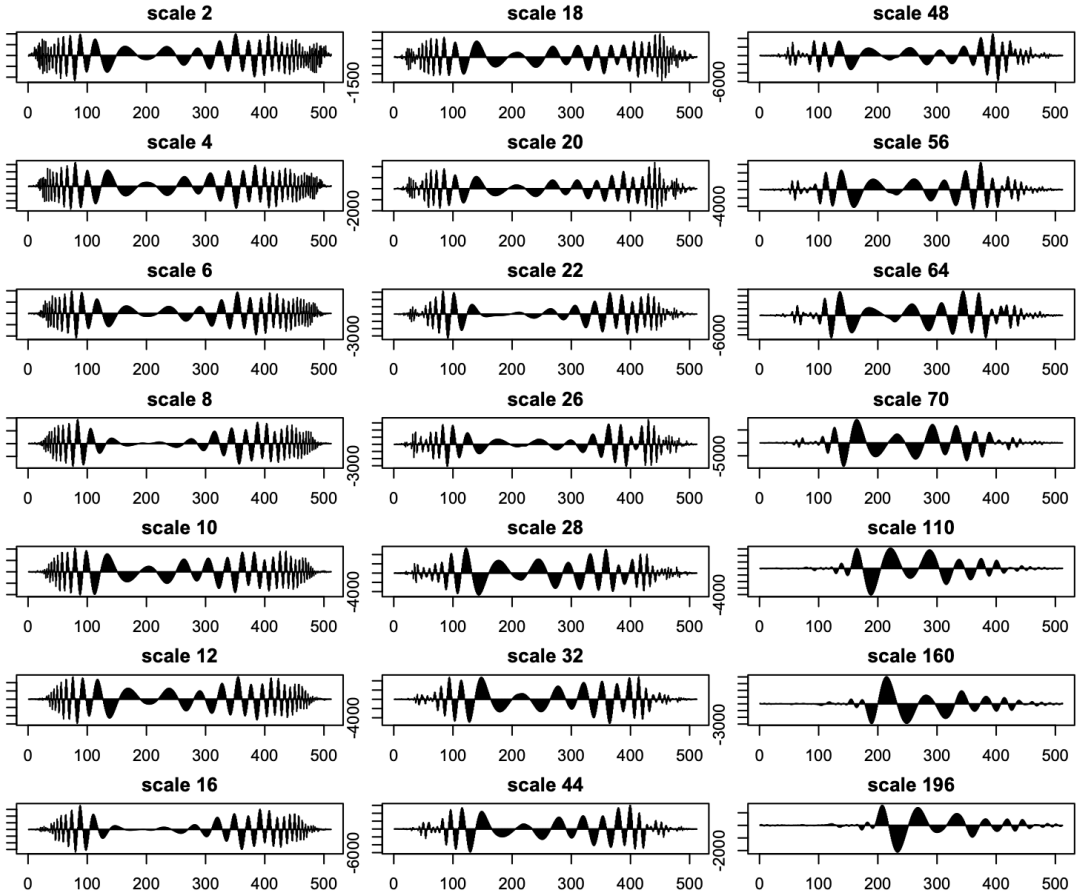


Figure 6.4.5: The level-scaled line plots for the corrected estimate of the EWS from the modified CLSW process for the reflected doppler test function using the BIC2 model.

6.4.2 Limiting the Number of Scales

The modified scale set of the CLSW process in (6.3.3), $\boldsymbol{A_M}$, seems to resolve a part of distortion effects within the dense and redundant default scale set during the correction process. Based on the improvement from the modified scale set for the CLSW process, we can consider the size of a modified scale set as another option. In the applications of the modified CLSW process in §6.4, the number of scales is determined by the different criteria options and tolerance values in the OMP algorithm. However, the CLSW process includes a correction process to get an approximately unbiased estimator of the EWS. The operator A used

for the correction is calculated based on the relationship between two scales, and therefore the complexity and dimension of the operator A is related to the number of scales in $\boldsymbol{A_M}$.

The operator A of the modified CLSW process is the subset of the default operator A from the CLSW process, but the inverse matrix changes depending on the choice of scales. As the number of scales increases, the structural changes in the inverse of the operator A can cause to have more complicated correction process with distortion effect between scales. Also, it may affect the asymptotic elements in the expectation of the periodogram in equation (6.3.7) equivalently to the fully CLSW process. Therefore, we limit the number of scales using the same data in §6.4 and compare the estimated spectral densities.

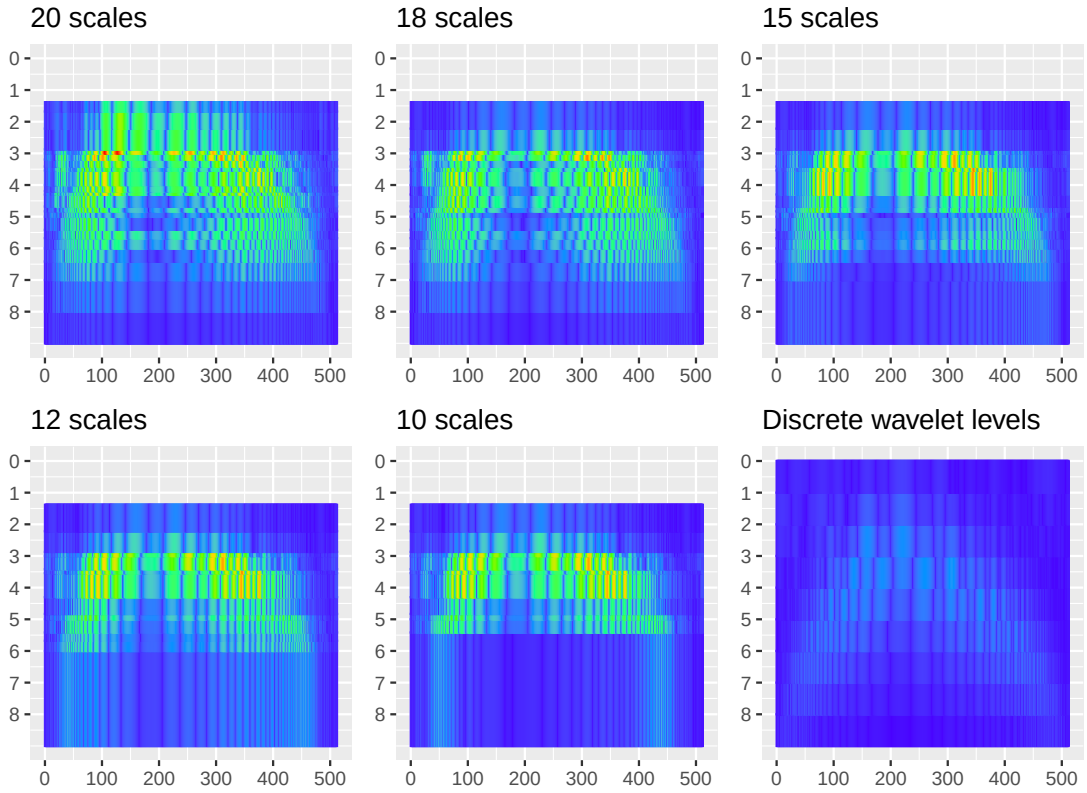


Figure 6.4.6: The image plots for the corrected estimate of the EWS from the modified CLSW process (with 20, 18, 15, 12 and 10 wavelet scales) and the discrete LSW process for the reflected doppler test function.

Here, the five different models are fitted for the reflected doppler test function:

20, 18, 15, 12 and 10-scale models. We do not consider the kind of criteria for this modelling, as the scales are added by the same order until 20-scale model using the SSE and BIC. Figure 6.4.6 shows the image plots from these five models and the discrete LSW process. As could be seen from the plots, the modified CLSW processes show high spectral estimates between Scale 3 and 6 in all the scale set. In particular, they tend to have very strong power of spectra in the middle near Scale 3, which are coloured as yellow and red. Also, as we decrease the scale dimension for the modified CLSW process, the reflected doppler test function tends to retain spectral power at the wavelet scales near Scale 3 and excludes very fine scales from the modified process. The estimates from the modified CLSW processes show the different time-varying frequency characteristics at each scale and their patterns seem to match the data structure of reflected doppler data and the discrete EWS. However, compared to the modified CLSW process, the discrete wavelet resolution levels have lower power of spectral estimates. Also, since a small number of wavelet scales makes the modified CLSW process simpler, the dimension reduction allows us to have the smoother spectral estimate.

In the similar way to the earlier analysis, we will see the line plots of the corrected estimate of the EWS to investigate more detailed pattern in each scale. The 18-scale model seems to have the smooth spectral estimates without much loss of information in Figure 6.4.6. Therefore, we will use the model as a example. Figure 6.4.7 show the line plots of the corrected estimate of the EWS scaled by the entire values from all chosen scales. These plots are similar to the results of Model BIC2 in Figure 6.4.4, but there is a little difference after controlling the number of scales. The Scale 20 and 26 have slightly increased spectra than Model BIC2, which seems to be due to the absence of the Scale 22 and 160 in the modified 18-scale CLSW process. On the contrary, the Scale 110 has the lower spectral estimates than before.

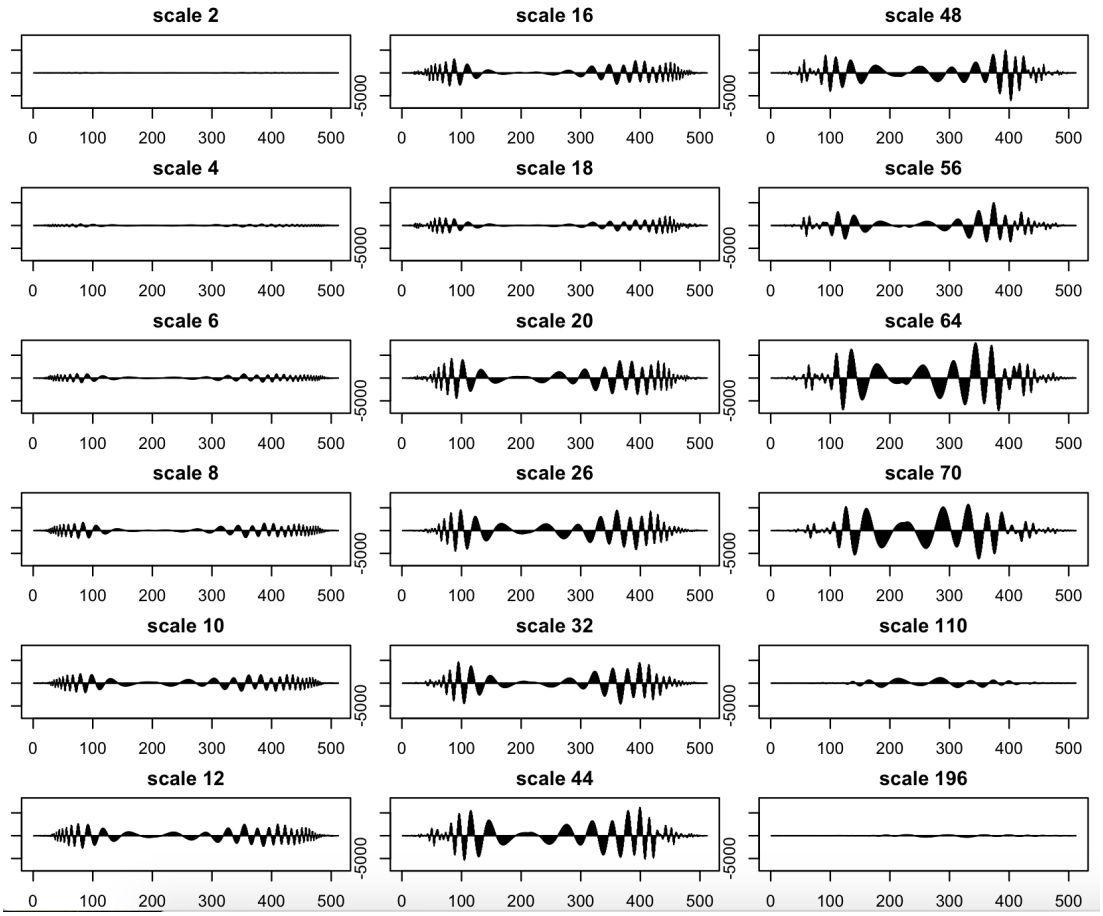


Figure 6.4.7: The global-scaled line plots for the corrected estimate of the EWS from the modified CLSW process for the reflected doppler test function using the 18-scale model.

The line plots of Figure 6.4.8 shows the spectral estimate standardised by the scale factor from each scale to compare with Figure 6.4.5. They show more sophisticated change within each scale after excluding the Scale 22, 28 and 160 from Model BIC2. As there is no change of fine scales at the first column, the estimates look similar to the result of Model BIC2. However, on the second and third column, there are some shrunk spectra especially in the middle in the middle on Scale 20, 26, 32 and 56. Scale 110 showed the decreasing spectral estimates in Figure 6.4.7, but the changed estimate centered in the middle seems to match better the structure of the reflected doppler test function. These changes allow us to have the better resolution in the image and line plots.

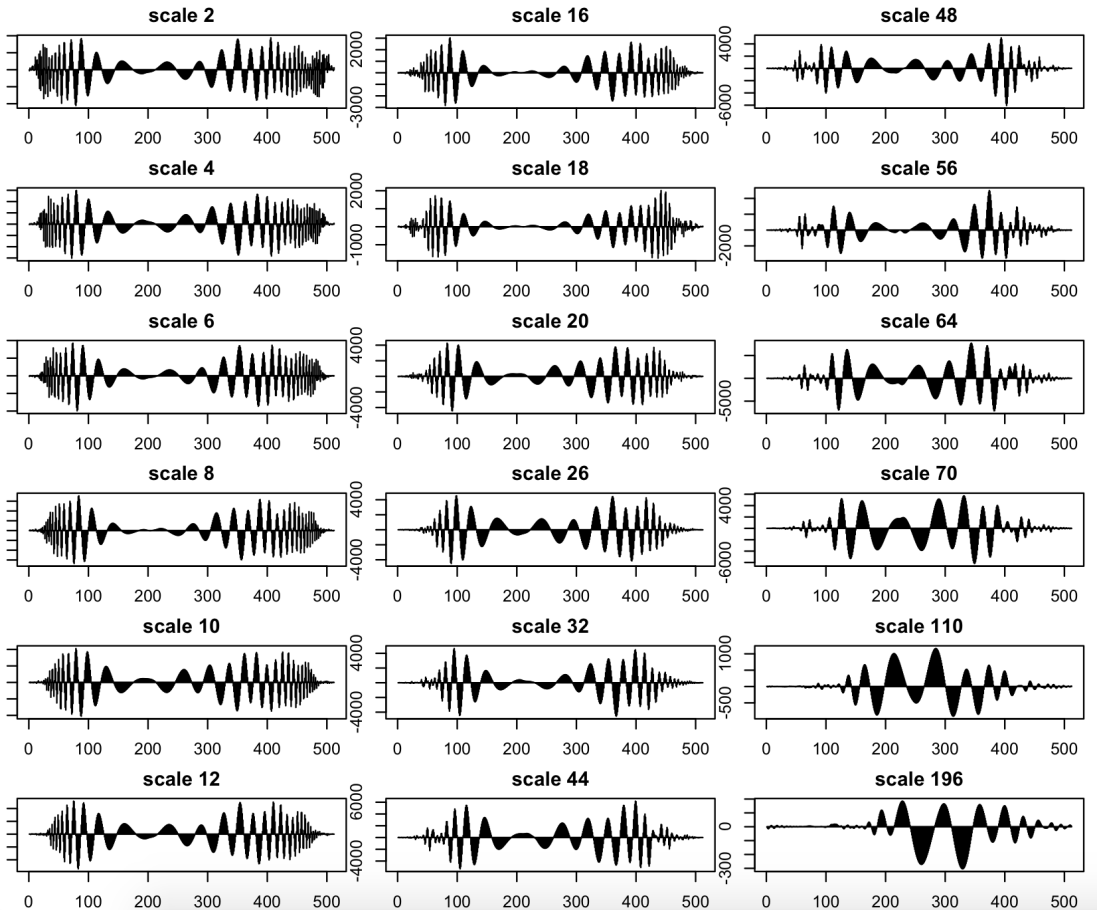


Figure 6.4.8: The level-scaled line plots for the corrected estimate of the EWS from the modified CLSW process for the reflected doppler test function using the 18-scale model.

6.4.3 The Application to the Real Tomography Data

In §5.5.2, we investigated the spectral characteristics using the real tomography data modelled by our CLSW process. We will use the same example data to compare the spectral estimates from the modified CLSW process with the previous result. The example data consists of the segments of two different flow types, "plug" flow and "bubble" flow, which was shown in Figure 5.5.6. In theory, the tomography data has significantly different frequency characteristics between the two states, but the estimated EWS of the fully CLSW process in Figure 5.5.7 does not display the frequency properties we expect from the continuous wavelet

coefficients.

Here, we modify the scale set for the CLSW process to use a smaller number of wavelet scales. As described in §6.3, the scales of the modified CLSW process are chosen adaptively to the structure of the tomography dataset by using the OMP algorithm. The criteria for scale selection are the same as the previous reflected doppler test function analysis and the models are labelled equivalently to them as well. Figure 6.4.9 shows the functions selected by the OMP algorithm.

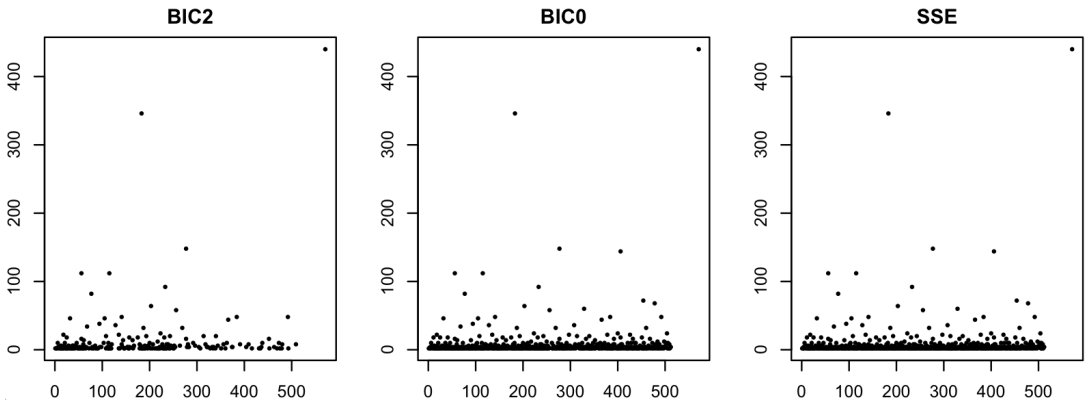


Figure 6.4.9: The wavelet function selection of the tomography data by the OMP algorithm

Model BIC has the smallest the number of functions due to the lower tolerance value, 2 than the other models. In general, the chosen functions are gathered at fine scales smaller than 50. Model BIC tends to have a denser group on the first half compared to the other models. The "plug" data from the first half has relatively wide fluctuations with a few spikes, while the "bubble" data from the other half has many fine oscillations but does not change much. Therefore, the "plug" data needs a wider range of wavelet scales to explain the frequency characteristics than the "bubble" data. Meanwhile, the SSE and BIC0 models choose wavelet functions spread over the two different states.

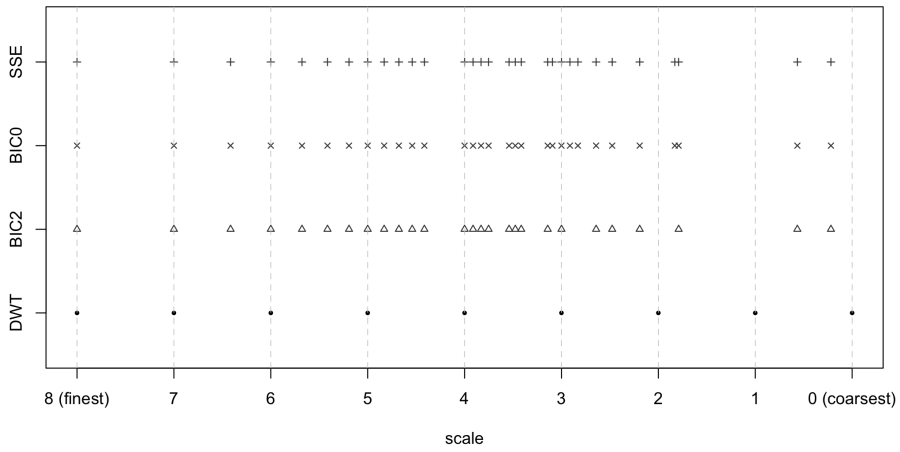
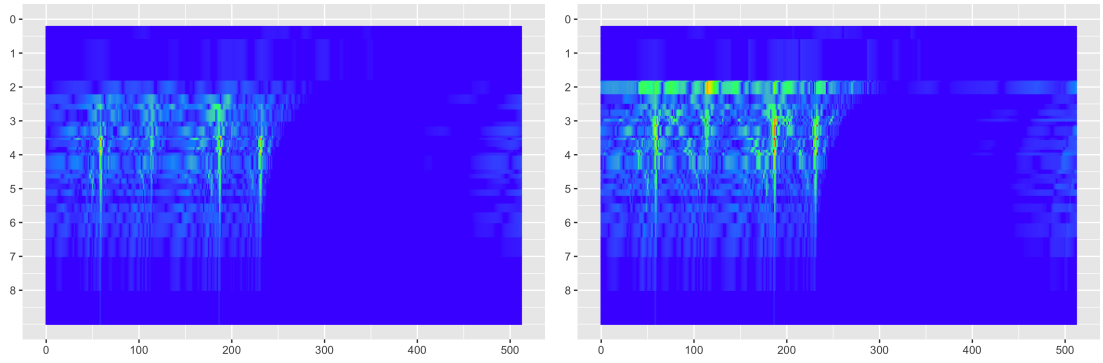


Figure 6.4.10: The scale selections from the BIC, BIC0 and SSE models with a comparison to the discrete wavelet resolution levels.

Based on the function selection, we can retain the scales for which at least one function is selected by the OMP algorithm. Figure 6.4.10 shows the scale selection and compares them to the discrete wavelet resolution levels. Model BIC2 has 27 scales and Model BIC0 and SSE have the identical 31 scales. Therefore, we will refer to the BIC model as Model 1, and the BIC0 and SSE models as Model 2 to simplify the interpretation.



(a) Model 1 (b) Model 2

Figure 6.4.11: The corrected estimate of EWS from the modified CLSW process of the example tomography data

Figure 6.4.11 is the image plots of the corrected estimate of the EWS from Model 1 and 2 using the example tomography data. These plots show distinct

characteristics between the "plug" and "bubble" states. There are four short intervals having sparse and high spectra on the "plug" state, which seem to match the four spikes seen in the original data. Generally, Model 1 and 2 have similar patterns of the spectral estimates, but the denser scale set between Scale 2-4 of Model 2 suggests more detailed frequency characteristics about which scales are related to the corresponding spikes.

The modified CLSW process can be adjusted by the number of scales as proposed in §6.4.2. Figure 6.4.12 shows the estimate of the corrected EWS from the modified CLSW process with the various scale numbers including the discrete model. The modified CLSW processes tend to have high spectra between Scales 2 and 4. The 27-scale model has significantly high spectral estimates at the spikes around Scale 4, the square roots of which are about 50. The 20-scale model in Figure 6.4.12 retains the high spectral densities at Scale 3 and 5 equivalently near the spikes, but the square roots of them are about 25. In the same spirit, as the number of scales decreases, the range of EWS estimates tends to be narrower. Since the image plots are scaled by the same colour scales to compare the estimated EWS across the different scale sets, the 15-scaled and discrete models seem to have very low spectra due to the extremely high values in the other denser models. However, they still have clear separations between two different states, "bubble" and "plug", and the very high spectra near the spikes, which are shown in the 27-scale model, seem to spread to the neighbourhood. This allows us to investigate frequency characteristics in a wide range of scales for the spikes. In addition, the dimension reduction seems to control the smoothness of corrected spectrum. In the modified CLSW process, the smaller the scale set is, the smoother the corrected spectrum is given.

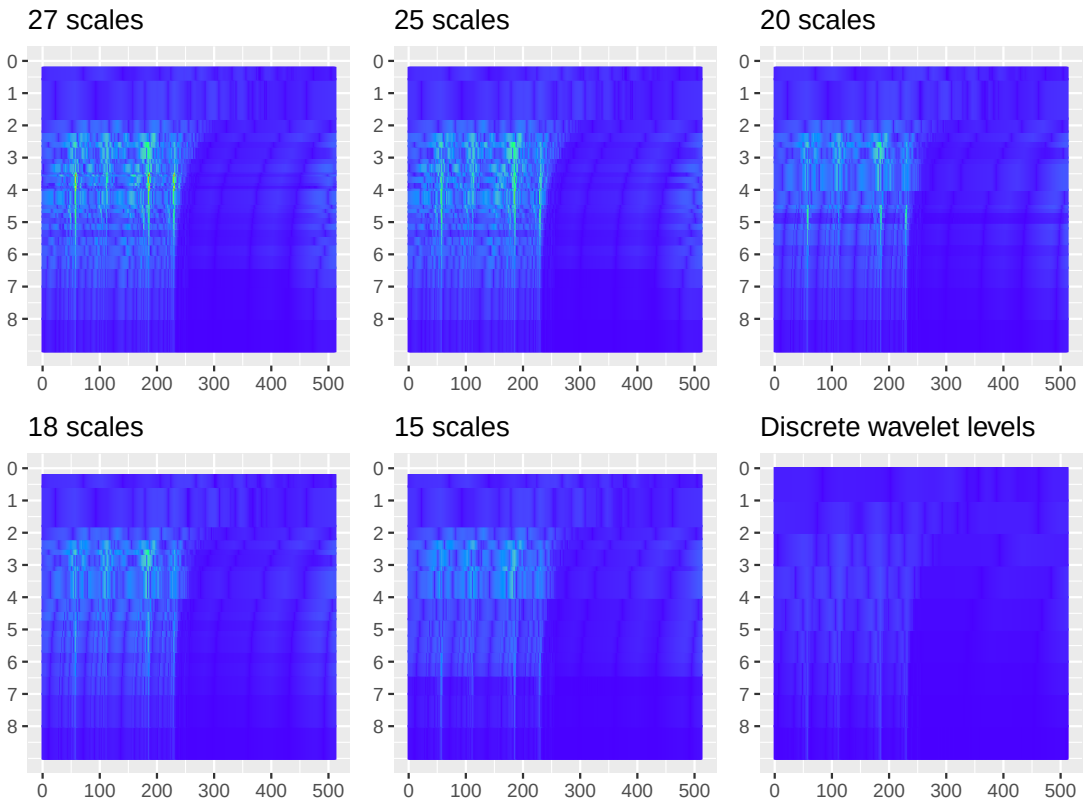


Figure 6.4.12: The comparison of the corrected estimate of EWS from the 27-scaled, 25-scaled, 20-scaled, 18-scaled, 15-scaled, 12-scaled and discrete models.

In comparison to the other above models, the 20-scaled model seems to explain the frequency characteristics of the "plug" states without much loss of information. Here, we will see the detailed trends in spectra within each scale by line plots. The sparse and high spikes are mainly shown between Scale 4 and 5, and the area between spikes are explained by the spectrum around Scale 3. Here, the Scale 3 show some spectra between the spikes, which seems to be related to the wide waveforms of the "plug" state in Figure 5.5.6. Also, the scale chosen for the 20-scaled modified CLSW process is fitted with the 15 intermediate wavelet scales and 5 discrete wavelet resolution levels.

Hence, we can conclude that there are useful spectral information from the intermediate wavelet scales to explain the frequency characteristics of the real tomography data. Also, limiting the number of wavelet scales can result in the better resolution of spectra for the real tomography data.

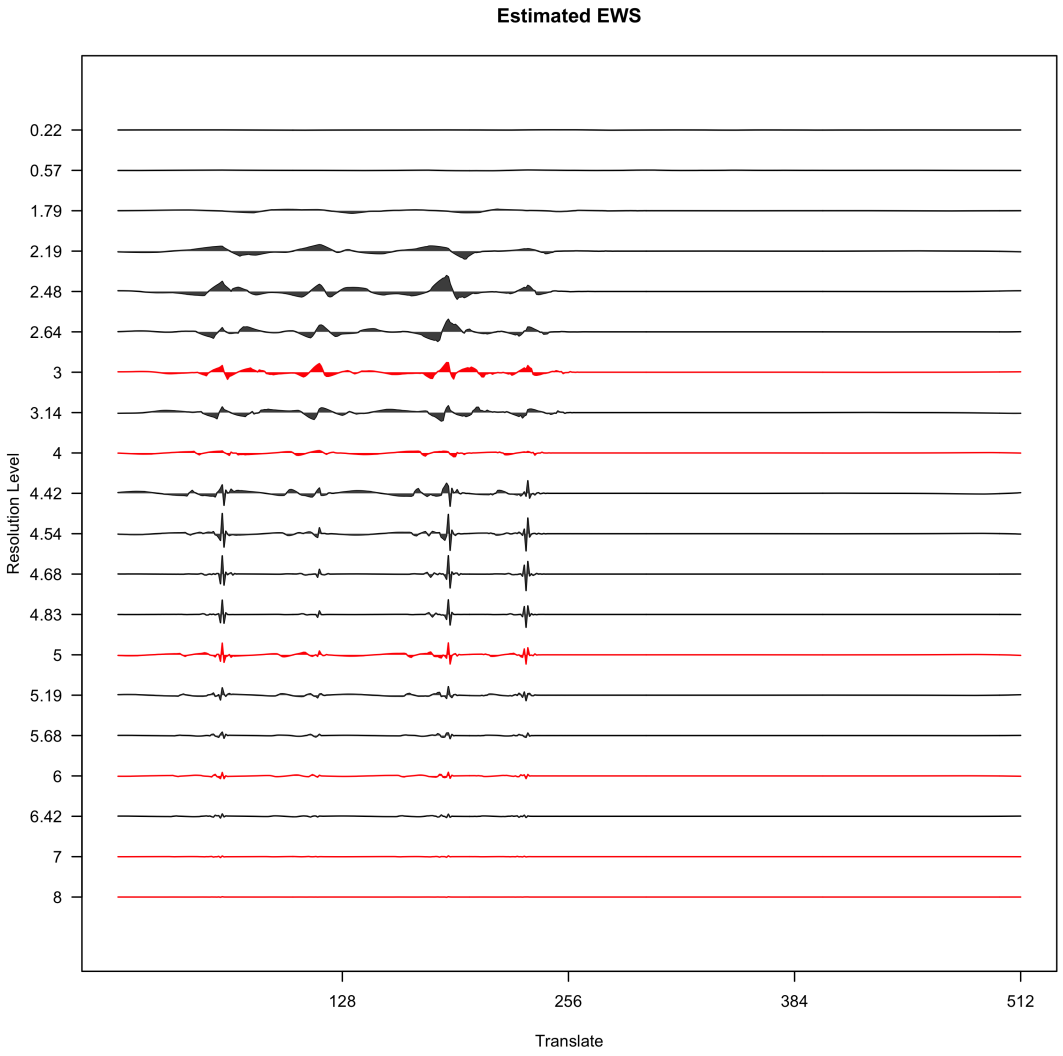


Figure 6.4.13: The line plots of the corrected estimate of EWS from the 18-scaled CLSW process of example tomography data. The red coloured lines are the estimate from the scale corresponding to the discrete wavelet resolution levels.

6.4.4 The Simulation of modified CLSW Process Based on the Haar MA process

In §5.5.3, 100 replicates of the concatenated Haar MA processes were used to estimate the EWS from the CLSW process. However, the corrected estimate of EWS in Figure 5.5.9 did not show a clear frequency feature. We updated the

wavelet scale set of the CLSW process following the definition of the modified CLSW process in §5.4.1 and estimated the EWS from the modified process. Figure 6.4.14 is the collection of image plots of the square root of corrected estimates of EWS from various numbers of wavelet scales.

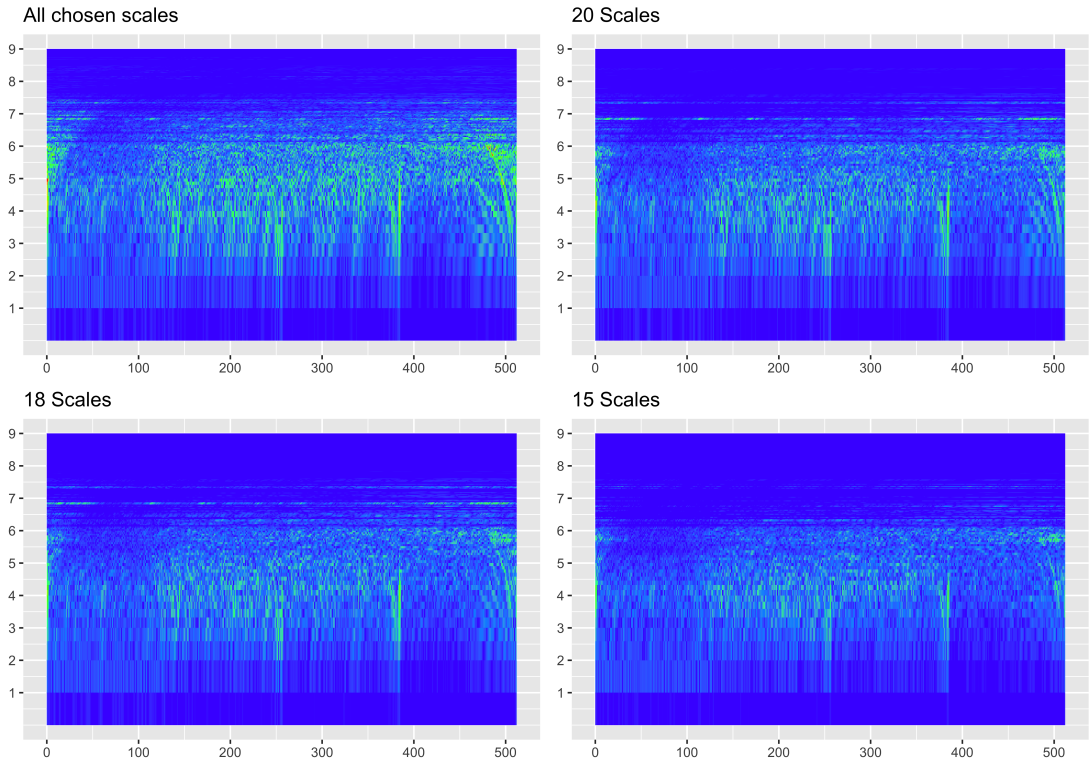


Figure 6.4.14: The averaged corrected estimate of evolutionary wavelet spectra from the modified CLSW processes over 100 replicates.

The orders of Haar MA processes, 2, 7, 10 and 32, correspond to the wavelet resolution levels, 2, 3.8, 4.32 and 6, on the y-axis. All the estimated spectral densities in Figure 6.4.14 were plotted under the same rainbow colour scale to compare the estimates across the image plots.

There are some high spectral estimates coloured as bright green or red at both edges of some coarse wavelet scales between 4 and 6. These high spectra seem to be from the cone-shaped boundary effects, which are seen in other wavelet transformations as well. In addition to the boundary effect, the image plots show some useful frequency information not presented in Figure 5.5.9. For example,

the first interval corresponding to the Haar MA process, $x_2(t)$, has relatively high spectral estimates near Scale 2 compared to other wavelet scales and the high spectra tend to be long lasting within the interval. In the same manner, the scale, which has the steadily active and high spectral power at each interval, seems to be near the wavelet scale corresponding to the order of the Haar MA process. Also, although the number of wavelet scale set does not affect significantly the main spectral feature within each process, the boundary effects tend to be weaker in a smaller number of wavelet scales.

However, the wavelet frequency information depicted from the image plot does not perfectly match the order of each process. For example, in the second interval with the Haar MA process, $x_7(t)$, although the wavelet scales near 3.8 show relatively high spectral estimates, it is not easy to investigate the true order of this Haar MA process from the image plot of the estimated EWS alone.

Also, the entire spectral power tends to decrease as the number of wavelet scales for the modified CLSW process gets smaller. Our CLSW process and modified CLSW process are constructed from the idea to create a representation of data as a summation of wavelet signals. However, the continuous wavelet transform is computed independently in each scale, and therefore, the total sum of squared wavelet coefficients increases as more scales are added to the modified wavelet scale set. Also, since the operator A has more elements in a denser scale set, constructing its inverse is more complicated and the range of elements in the inverse of the operator A is wider. The dependence of spectral energy on the number of wavelet scales needs to be considered for future work.

6.5 Smoothing Wavelet Spectra

Nason *et al.* (2000) and von Sachs *et al.* (1997) demonstrate that the periodogram of the LSW process is not a consistent estimator of EWS. To get a good estimate, they applied two types of correction, using the inverse of the operator A and smoothing to be consistent. In the same principle as the discrete LSW process, the continuous LSW process is also not consistent as the asymptotic variance of wavelet periodogram does not vanish. Hence, the wavelet periodogram of the CLSW process also needs to be smoothed.

In general wavelet methods, various smoothing techniques have been developed to cope with noisy data. The most commonly used and simplest thresholding technique is the universal threshold introduced by Donoho & Johnstone (1994). The threshold, λ, is defined by $\sigma\sqrt{2\log n}$ where σ is the noise level and n is the length of data. The noise level, σ can be estimated by the median absolute deviation of the wavelet coefficients from the finest wavelet scale. To shrink a wavelet coefficient, d, based on the universal threshold, λ, soft and hard thresholding functions are defined as

$$d^*_{hard} = dI(d \leq \lambda) \tag{6.5.1}$$

$$d^*_{soft} = \text{sgn}(d)(|d| - \lambda)I(d \leq \lambda) \tag{6.5.2}$$

by Donoho & Johnstone (1994). Under the assumption of gaussian process for the error, it is known that the universal threshold can denoise the error with high probability (Vidakovic, 1999).

In addition to the universal threshold, we can consider the use of Bayesian methods. Johnstone & Silverman (2005) introduced the empirical Bayes selection of wavelet threshold implemented in their R package, "Ebayesthresh" (Silverman & Johnstone, 2005).

The smoothing for the modified CLSW process will be applied to the reflected doppler test function and the real tomography data in §6.4. Since the fully CLSW process does not generate meaningful spectral characteristic due to the distortion effects from the redundant scale set, we only consider smoothing wavelet periodogram for the modified CLSW process with the scale set chosen adaptively to the data structure. The wavelet periodogram is smoothed first to denoise, and we then correct the smoothed wavelet periodogram by the inverse of the operator A to get the approximately unbiased EWS.

The reflected doppler test function in Figure 5.5.1 is generated from the doppler function without any noise. But, to smooth the wavelet periodogram, we add Gaussian noise to the reflected doppler test function and average the estimated EWS over 50 iterations. Figure 6.5.1 shows the mean of the smoothed spectral estimate of EWS from the modified CLSW process over 50 iterations. In each process, the scale set modified by using the OMP algorithm is different

depending on noise of individual data. Hence, the mean of spectral densities at each scale can be associated to how many datasets include the corresponding scale. However, the change of spectra over time are independent with averaging within each scale, and therefore we can still compare the mean of the smoothed estimate of corrected EWS to the average of spectral characteristics for the data.

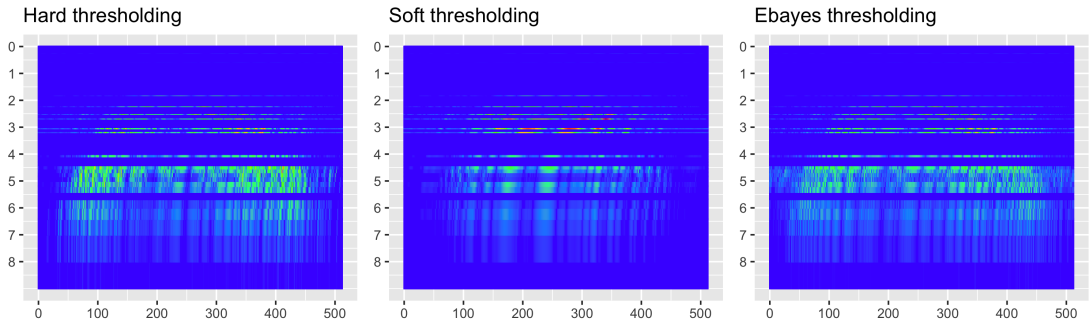
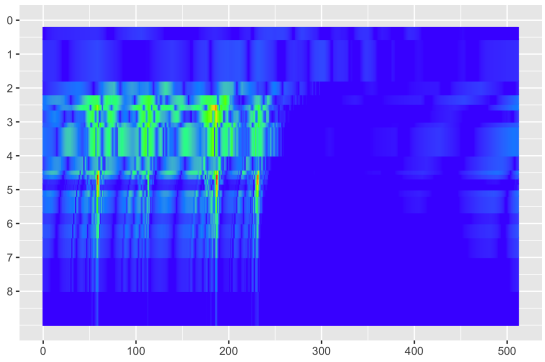


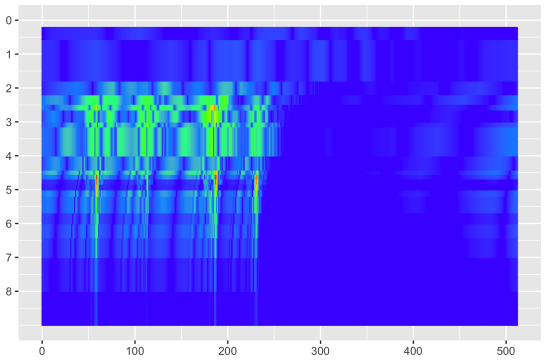
Figure 6.5.1: The mean of the corrected estimate of EWS from the modified CLSW process over 50 iterations.

Figure 6.5.2 shows the smoothed estimate of the corrected EWS from the 20-scaled modified CLSW process using the real tomography data. Since the tomography data is real data with noise, we do not consider to add more noise, while the reflected doppler data added some Gaussian noise before smoothing the wavelet spectra. The wavelet periodogram is smoothed by the universal and Ebayes thresholds equivalently to the previous smoothing example.

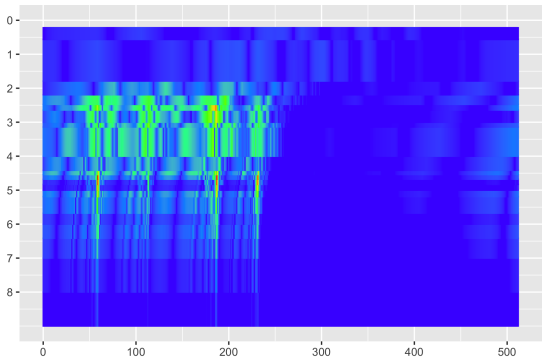
As can be seen from the plots, it is difficult to find significant difference between these different threshold techniques. Here, we retain the hard universal threshold and Ebayes threshold to see whether there are any more subtle difference revealed by the line plots in Figure 6.5.3. They still do not have any clearly different patterns between the universal and Ebayes thresholds. Also, compared to the estimate before smoothing in Figure 6.4.13, the smoothing does not make a significant change in the wavelet periodogram for the real tomography data. This could be because the noise level of the real tomography data is quite small, so the noise may not have much influence on the wavelet periodogram.



(a) Soft universal



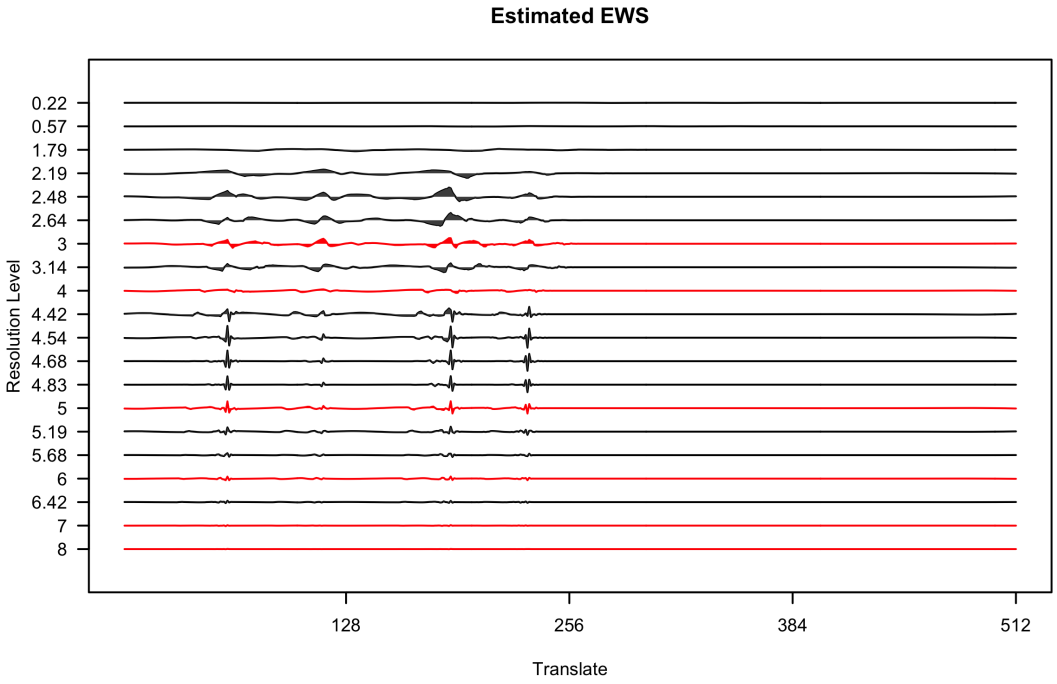
(b) Hard universal



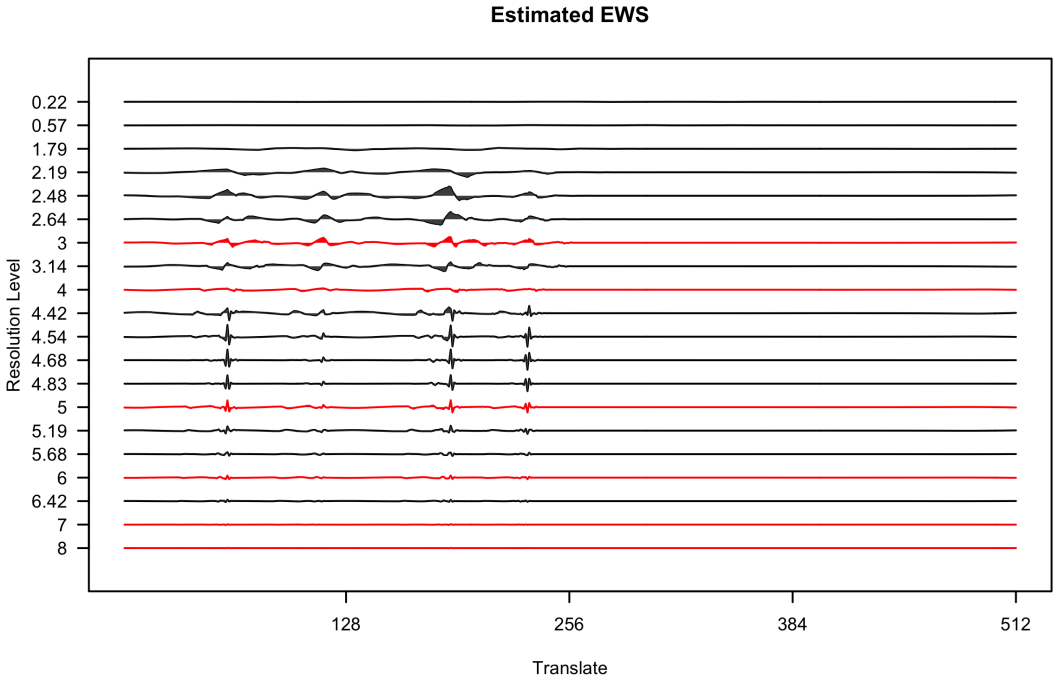
(c) Ebayes

Figure 6.5.2: The image plots of the corrected estimate of EWS from the 20-scaled CLSW process.

Although we only applied the universal threshold and Ebayes threshold for the modified CLSW process, Nason *et al.* (2000) suggests using the translation invariant denoising by the LSuniversal threshold explained in von Sachs *et al.* (1997). The non-linear threshold of the discrete wavelet method enables the discrete LSW process to estimate more statistically stable wavelet periodogram with approximately vanishing variance, but we need to verify that the theories can be applied to the dense continuous scale set. Also, Fryzlewicz & Nason (2006); Fryzlewicz (2003) and Nason (2010) advocate using the idea of using wavelet Fisz transform for smoothing the wavelet periodogram of the LSW process. We can also consider investigating smoothing with the wavelet Fisz transform as future research.



(a) Hard universal threshold



(b) Ebayes threshold

Figure 6.5.3: The line plots of the corrected estimate of EWS from the 20-scaled CLSW process.

6.6 Conclusion

The modified CLSW process is aimed to decrease distortion effect between scales through selection of a subset of scales from the fully continuous wavelet scale set. The scale reduction in the modified scale set has a smaller number of approximation terms in the discrete sum over scales, by which we can expect to narrow the boundary of approximation for the expected value of estimated EWS. Applications of the modified CLSW process on the reflected doppler test function and real tomography data show the improvement in the spectral estimation after correction and limiting the number of scales can also revise the spectrum.

The estimated spectrum is not as smooth as that of the discrete LSW process, but true spectrum is not always smooth or simple. Hence, if data has more noisy spectral characteristics from various scales in between the discrete wavelet resolution levels, the modified CLSW process can suggest more sophisticated spectral analysis with the scale selection adaptive to data structure.

In our research, we considered the two tolerance value options, 0 and 2, for the BIC option in the OMP algorithm. According to the BIC interpretation scheme in Raftery (1995), an alternative model has a positive evidence to improve performance compared to a null model when the BIC difference is larger than 2. On the other hand, the BIC difference under 2 implies a weak evidence for an alternative model and the alternative model is not worth to mention. However, in our research, the tolerance value, 0, was also used to choose all wavelet functions which have any small contribution to explaining data. In §6.4, Model BIC0 showed the very similar or identical wavelet scale sets to Model BIC2. Therefore, based on the conventional BIC interpretation scheme and our application in the modified CLSW process, the tolerance value, 2, may derive an appropriate number of wavelet scales matching the structure of data, and we can consider this tolerance value option as the default option to construct the modified CLSW process. One possible direction for future work would be to investigate the effect of using other tolerance values.

Although the modification of CLSW process showed the improved spectral estimates in the example data analysis, there are still some points to be investigated further. First, the modified CLSW process is given by the modified scale

set, $\boldsymbol{A_M}$ which we obtained from the OMP algorithm. In addition, the number of scale set in the modified CLSW process can influence on the correction by the operator A and we compared the spectral estimates over the different models in §6.4.2. But, it is difficult to indicate which model shows spectral density more clearly. Therefore, we need to find a way to choose the most appropriate number of scales to use in representing a given data set. Also, there is still some noise in the unexpected area especially on fine scales. The wavelet functions on fine scales are easily included to explain details of data through the OMP algorithm, but the dense set of fine scales still has risk to change the shape of spectral estimation.

Chapter 7

Classification Model with the CLSW process

7.1 Introduction

Chapter 3 and 4 built classification model based on the wavelet coefficients of the DWT and CWT. They show that measurements derived from the wavelet coefficients can support a classification to investigate the state of gas-fraction in the air-liquid phase and the flexible scale selection of the CWT also enables us to improve the accuracy of the modelling. Along with the result, Chapter 6 suggested the locally stationary wavelet process on the CWT modifying via the matching pursuit algorithm. Section 6.4 shows that the application on the modified CLSW process gives better spectral resolution from intermediate frequencies between discrete resolution levels. Here, we will apply the modified CLSW process to the classification modelling problem, using an example of the vertical tomography data. The main objective of this chapter is to to give an example of modelling using the CLSW process. Therefore, we will fit a model using the spectrum on one example of vertical tomography data rather than doing a complete modelling over entire datasets.

7.2 Modelling on the Vertical Tomography Data

7.2.1 Variables

The scale is chosen through the adapted OMP algorithm described in §6.3. We denote the subset of scales chosen by our algorithm from the fully continuous domain as A_M. The basic principle of the modelling is equivalent to that used in the modelling in Chapter 3. Recall that the objective of modelling is to classify the current state of gas-fraction based on the past conductivity data from the sensors. Therefore, the information available for modelling at time t consists of all the data up to that time. First, denote the wavelet coefficients used for the modelling as

$$d^i_{ab} = \sum_{t\in[0,T)} f_i(t)\psi_{ab}(t) \tag{7.2.1}$$

for the corresponding scale, a and shift, b where i is the id of a pair of sensors. In our research, we restrict the wavelet function for modelling and construction of the CLSW process to the Haar wavelet function with a compact support. Therefore, the edge of the Haar basis function is set to be at $t = b$ so that the wavelet function, $\psi_{ab}(t)$ (7.2.1) targets the data before the time, $t = b$.

For simplicity, the mean of the absolute values of the wavelet coefficients is used as an activity measure for the classification modelling. The activity measure can be written as

$$x_a(b) = \sum_i \frac{|d^i_{ab}|}{N_I}, \tag{7.2.2}$$

where N_I is the total number of pairs of sensors for the experiment.

7.2.2 Models

To see the efficiency of the CLSW process and our scale selection approach for modelling, we can compare the models fitted from different wavelet signals and subsets of scales. With the simplest model of the discrete wavelet coefficients, the estimated spectra of the discrete LSW process and modified CLSW process are fitted for the example tomography data. Also, for the modified CLSW process, §6.4.2 shows that using different cut-off thresholds during scale selection can

present better resolution for the estimated EWS. Therefore, different numbers of scales will be considered for the model comparison. Models are defined as follows:

- Model 1 : Wavelet coefficients of DWT
- Model 2 : Estimated EWS of the discrete LSW process
- Model 3 : Estimated EWS of the modified CLSW process (32 scales)
- Model 4 : Estimated EWS of the modified CLSW process (20 scales)
- Model 5 : Estimated EWS of the modified CLSW process (15 scales).

Model 3 has 32 scales, which is the maximum number of scales chosen through the orthogonal matching pursuit algorithm.

Figure 7.2.1 shows the data trace of one sensor pair from the example dataset used for this chapter. In the same spirit as the modelling in Chapter 3, the first 256 observations are recorded from the "bubble" state and the other half is from the "churn" state. The example used for plotting and the following data exploration is the mean value of conductivity data over the 49 pairs of sensors, but modelling will be done through an individual wavelet transformation of data from each pair of sensors.

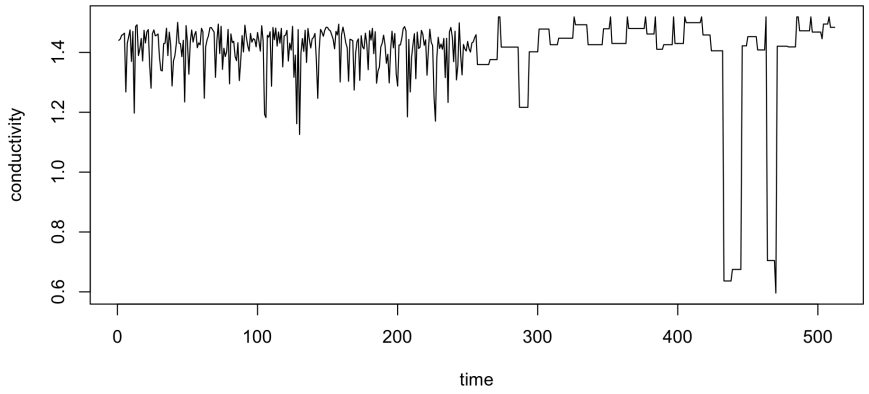


Figure 7.2.1: The vertical tomography dataset for a train set.

Based on the structure of this example data and continuous wavelet functions, the scale set for the CLSW process is modified by using the OMP algorithm. Figure 7.2.2 shows which wavelet frequencies are chosen in each model.

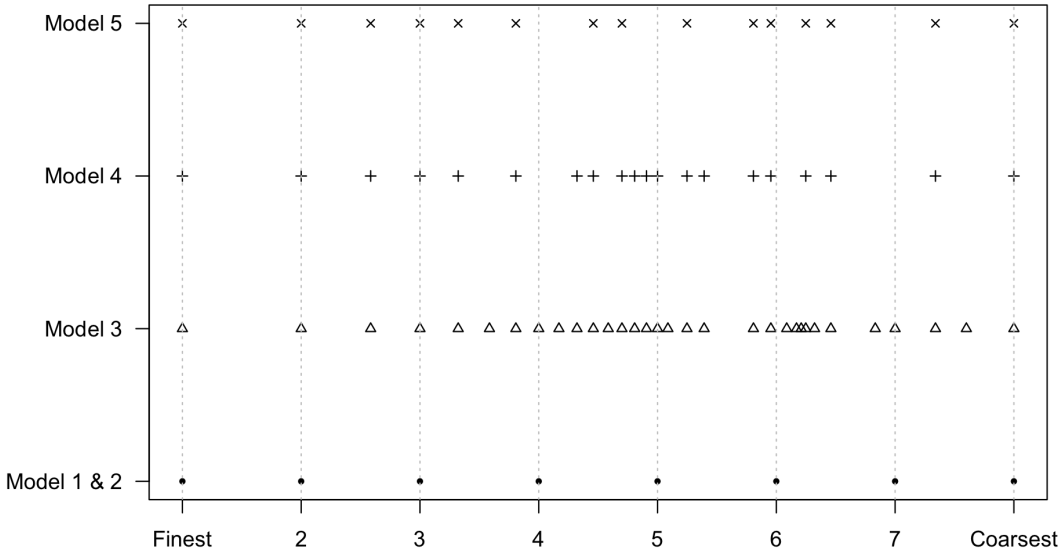


Figure 7.2.2: The scale selection for Models 1-5.

Figure 7.2.3 is the image plot of the absolute wavelet coefficients from discrete wavelet resolution levels. As we could guess from a few of rapid changes of Figure 7.2.1, there are some significantly high wavelet coefficients between $t = 400$ and 500.

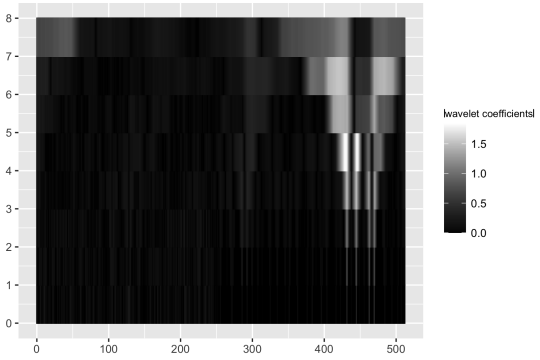
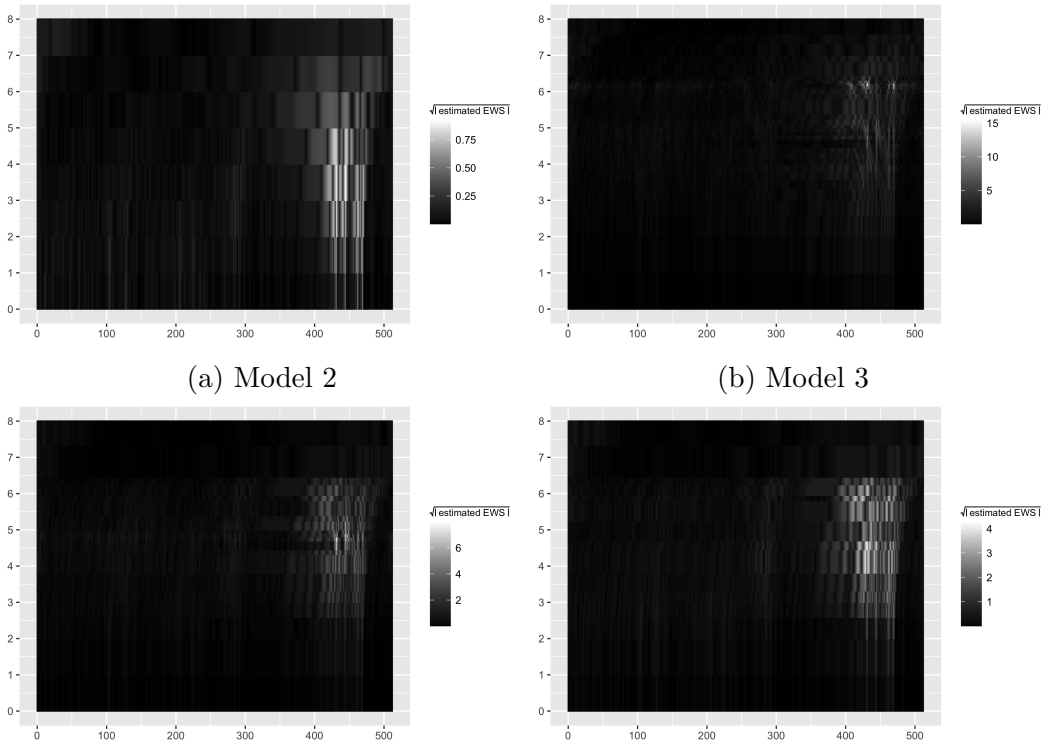


Figure 7.2.3: The image of the absolute wavelet coefficients used for Model 1.

Following the discrete wavelet coefficients, Figure 7.2.4 shows the image plots of the corrected estimate of the EWS based on the discrete LSW and modified CLSW processes. These image plots were not scaled over the models due to very small values of the estimated EWS in Model 2. The estimated spectrum from

the discrete LSW process looks similar to the standard DWT, but the spectrum estimated from the LSW process shows more active spectral features at the finer wavelet levels compared to the wavelet coefficients. Meanwhile, the spectra based on the modified CLSW processes, particularly at Model 3, have the dominant spectral features around Scale 6. However, as the number of scales decreases, the spectral density is prone to spread over the scales.



(a) Model 2

(b) Model 3

(c) Model 4

(d) Model 5

Figure 7.2.4: The image of the square root of the estimated EWS for Models 2-5.

7.3 Result

Logistic classification models were fitted to the wavelet coefficients from Model 1 and the corrected estimate of the EWS from the other models. We fitted each model to the activity measures from the training set and tested that by using the test set in one example dataset. Validation process is not considered in this modelling due to the complexity of data generation in the LSW processes. The

wavelet signals for activity measures were transformed individually from each sensor pair and then averaged. Figure 7.3.1 shows the classification rates over all fitted models. As can be seen, Model 1, which is fitted with the absolute wavelet coefficients, has the highest accuracy among the five models. Also, more variables give better results in general except for the sixth step. Here, as one specific example dataset was used to see the efficiency of LSW processes in the tomography data analysis, the fitted models might struggle to find a general pattern of classification rates. However, based on the rapid fall of the classification rate at the sixth step in Model 1, the variable added at this step seems to have different characteristics relating to the gas-fraction states between the training and test set. Meanwhile, the others fitted using the LSW process show poor classification rates below 0.5. Model 2 has slightly better results than others after the fifth step, but the classification rates are still very low and they seem not to be improved by the addition of variables.

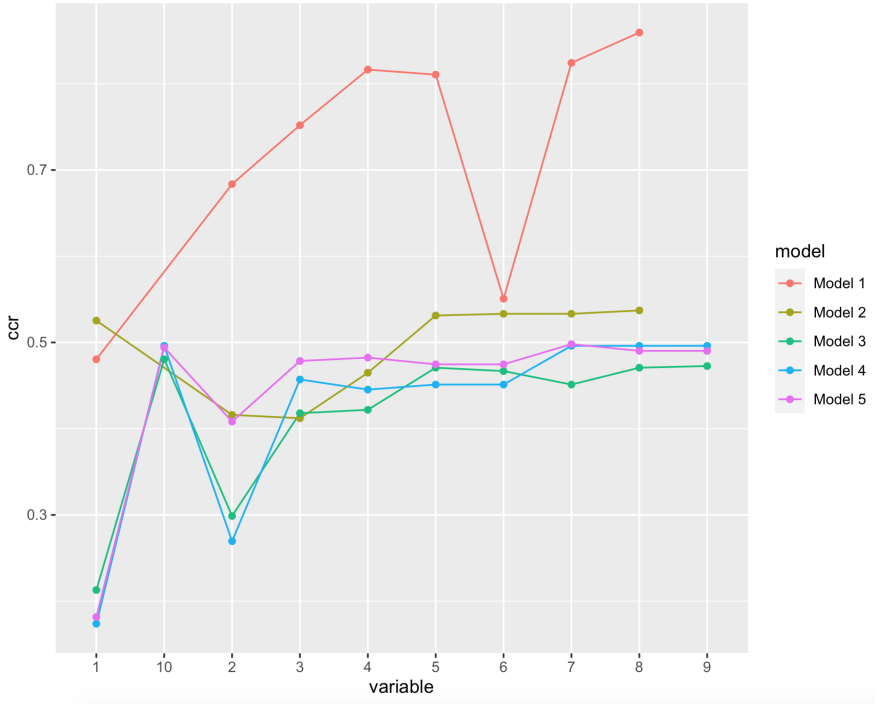


Figure 7.3.1: The comparison of classification rates on the test set.

The results mean that the discrete and modified continuous LSW process,

do not seem to fit well for the example tomography data. The simple wavelet coefficients from DWT predict the state of gas-fraction better compared to the estimated EWS. The poor result may be caused from the structure of tomography data. The data has different frequency features at the "bubble" and "churn" states with a few large spikes. So, the frequency characteristic can be considered as a time-changing factor depending on the state of gas-fraction. However, in particular, the "churn" state is likely to have a few large spike period of rapidly changing within a short time like Figure 7.2.1. In this case, the activity of the estimated EWS tends to be concentrated into the spikes. This can be useful to investigate the changing point of frequency features or the time points and scales related to the spikes, but does not work with the classification of the vertical tomography data. Also, Aykroyd *et al.* (2016) mentioned that the existing classification models are often based on an entire data set Krzemieniewska *et al.* (2014) and Fryzlewicz & Ombao (2009) rather than a segment of data, and therefore it can be not appropriate to classify the state at the specific time point for the tomography data.

7.4 Conclusion

This chapter fitted logistic classification models to the vertical tomography data using different estimated evolutionary wavelet spectra as explanatory variables. As the modelling is only carried out for one specific example dataset, it is difficult to get a general conclusion of regression with the CLSW process. However, we can have a glimpse of the modelling and its limitation.

The classification models were built over various choices of the scales and the kinds of wavelet signal. We compared the classification result from the wavelet coefficients to those from the estimated EWS. In general, the models with the LSW process are not as good as the models with the discrete wavelet coefficients. These poor classification rates imply that the LSW process does not seem to provide good explanatory variables for use in a regression model for the tomography data.

With the low accuracy of prediction, there are a few of limitations of modelling with the CLSW process. Since the standard CLSW process has a very redundant

set of wavelet functions, we decided to alter the set of scales by using the adapted OMP algorithm. The scale reduction can be considered as a variable selection step, but the modelling requires another variable selection step. The objectives of these two steps are different in that the first is to find the best subset of wavelet scales matching to the stochastic data and the other is to find the best variables to explain the relationship between the stochastic data and responses. However, the processes can be considered as duplicated work. One possibility for future work would be to investigate a single variable selection step which attempts to satisfy both criteria.

The computational cost is also another problem. The modified CLSW process reduces the scale dimension by the OMP algorithm, whose computational cost is related to the number of basis functions. The tomography example dataset is relatively small having 512 observations giving 128 scales, so the computational effort is not very expensive. However, the horizontal tomography data in Chapter 7 has a long sequence of data, over 3000 observation giving 1024 scales. Accordingly, the number of scales needed to check for the OMP algorithm is over 3,000,000, which is very computationally expensive with a big memory usage. The current R package is not ideal for such a big data problem, so we should devise more advanced computing techniques to save the cost and to expand the process into bigger data.

In conclusion, although the modelling with the CLSW process did not give a satisfactory result on this tomography data, the result shows that the choice of wavelet signal can be important depending on the structure of data and the objective of modelling. Also, the limitation found in the modelling will enable us to consider the future direction of the research with the computation costs.

Chapter 8

Conclusion

8.1 Summary

The thesis focused on the application of the continuous wavelet methods and the development of statistical modelling tool using the wavelets.

In Chapter 3 and 4, we fitted a logistic regression model to the tomography data based on wavelet methods. Time-frequency methods have been used in the signal processing field to understand frequency characteristics that change over time or are non-stationary (Papandreou-Suppappola, 2018) and the wavelet method is one useful example of time-frequency methods due to the localised natural wavelet functions. The frequency properties of tomographic conductivity data changes over time depending on the state of the gas-fraction, so we expected that the wavelet methods would be suitable to investigate the time-changing frequency features of signals. Aykroyd *et al.* (2016) had shown that a classification model based on the discrete wavelet transform (DWT) generated meaningful results. Although the data used in Chapter 3 is the same as Aykroyd *et al.* (2016), we expanded the scale set to the continuous domain to consider the benefit of non-discrete wavelet levels for analysing the tomography data. We created the five different sets of wavelet scales in varying degrees of denseness and compared the classification rates to see the impact of different initial scale-selection processes before subsequently choosing scales by variable selection when fitting a model. In general, the models fitted using the discrete wavelet levels and fully continuous

wavelet scales tended to produce lower classification rates than other scale sets with initial scale selection.

The classification model selected scale variables by the given criteria, "BIC" and "CCR". In particular, the first two variables chosen from the scale set are considered to play a significant role for classifying the gas-flow states by the largest change of the criteria. The scatter plots of the first two variables showed that a considerable number of variables represent the activity at non-discrete scales. That means that the continuous wavelet scales can give better understanding of the tomography data than using just the discrete wavelet resolution levels. Also, the use of an initial scale-selection process can improve the classification model using a smaller number of scales.

Chapter 4 fitted a classification model to the vertical flow tomography data. In the tomography data, the direction of the pipe and liquid flow can affect the mechanism of bubble formation. To supplement the previous modelling using the vertical tomography data analysis, we considered models based on the DWT and CWT. The modelling method and variable selection process used for the horizontal data are equivalent to the vertical tomography data modelling, but the pre-selected scale set was used for fitting according to the result in Chapter 3. The horizontal tomography data modelling also showed that the choice of intermediate variables, between the discrete wavelet resolution levels, which implied that the non-discrete wavelet frequencies, can be helpful to describe the frequency features of the horizontal tomography data. However, the low accuracy of classification in the "slug" regime needs to be improved in the future research.

This study aimed to extend the locally stationary wavelet (LSW) process with the continuous wavelet transform (CWT) in order to see the spectral features of data from intermediate levels between discrete wavelet resolution levels. The continuous LSW (CLSW) process designed in Chapter 5 was built on the fully continuous wavelet scale set. In the same spirit as the standard LSW process (Nason *et al.*, 2000), we imposed the Lipschitz continuity assumption to require the spectral characteristics of data not to change too quickly over time. In theory, the estimated evolutionary wavelet spectrum (EWS) can be corrected to be an approximately unbiased statistic by the operator A calculated from the wavelet correlation, but the application in §5.5 showed that the calculated result was

different to the true spectrum that we expected. A possible reason of poor resolution of the estimated EWS may be the different structure of the inverse A and the increasing sum of asymptotic terms due to the dense and highly correlated scale set.

Therefore, Chapter 6 suggested the selection of a subset of scales to modify the CLSW process using the idea of the orthogonal matching pursuit (OMP) algorithm. The principle of the matching pursuit (MP) algorithm is to choose the combination of basis functions matching the data structure. While the MP algorithm is for an orthogonal basis function space, the OMP algorithm enables the use of the algorithm to work for non-orthogonal function spaces. This extension can be applied to the continuous wavelet functions as well. We modified the CLSW process so that they were defined over the scale set chosen from the OMP algorithm. In the application with the example datasets in §6.4 and §6.4.3, the reduced scale set showed a clearer estimation of spectral density than the default CLSW process from the fully continuous scale set. The performance of the estimation of the EWS can be adjusted by the number and the choice of scales included in the modified CLSW process, but there is no definitive criteria yet to determine an optimal cut-off threshold of the number of scales to retain.

Based on the good performance of classification modelling in Chapters 3 and 4, and the improvement of spectral estimation in the modified CLSW process, we fitted the classification models using the explanatory variables based on the modified CLSW process. Chapter 7 showed that the model using the CLSW-derived variables from an example dataset of the vertical tomography data had a lower classification rate compared to the DWT based the models already investigated. As the tomography data seems to be non-stationary with a few large spikes within short intervals, we speculate that this characteristics may not be well explained by the CLSW process. As in many modelling context, we need to assess which method is better for data we have.

8.2 Future Work for the Tomography Data Analysis

Our logistic regression modelling for the vertical and horizontal tomography data showed the necessity of continuous wavelet scales to understand the frequency structure of the tomography data. In this section, we will suggest some ideas as future work for the tomography classification modelling.

The ultimate goal of the tomography modelling is to predict the state of a gas-fraction in real time. In our modelling, the backward facing option was used for the wavelet transformation to investigate the relationship between the frequency features and gas-fraction states from the data until specific time points. With the backward facing data, we can consider an algorithm for online inference for the future research.

Also, the horizontal tomography data used in §4 has different portions of gas-flow states, "bubble", "slug" and "plug". The unbalance of data between classes can derive a biased classification result. King & Zeng (2001) showed that the maximum likelihood estimate is biased in the logistic regression of unbalanced data. To cope with this problem, they suggested the correction methods based on the prior information in a population and the proportion of each class in samples. These correction methods can be considered in our modelling to mitigate the influence of class imbalance on the accuracy of classification.

In our research, we assume that the tomography data is independent over time. However, in practice, the voltage data seems to be correlated over time and the gas-fraction state at a certain time point is likely to be affected by the neighbouring data. To handle the correlation issue, we can consider to incorporate appropriate dependence of time points to the future model.

8.3 Future Work for the CLSW process

The modified CLSW process has the improvement of resolution in spectral estimation, but there are still some limitations. Here are a few ideas for future work to develop the results of this thesis.

8.3.1 The Assumption between Scales

The CWT has more redundant information than the non-decimated DWT, and therefore the CLSW process requires extra work to cope with this problem. This thesis has suggested dimension reduction imposing the OMP algorithm in Chapter 6 as a remedy for the redundancy. Since the redundancy comes from the more dense scale set of the CWT, we could consider more strict assumptions over scales. A correction process, considering how quickly activity can change over both time and scale, may improve spectral resolution for the EWS.

8.3.2 Theoretical Foundation for the CLSW Process

Chapter 6 suggested the modifed CLSW process with the scale set updated by using the OMP algorithm. The modification reduces the dimension of wavelet scales for the CLSW process, and the estimated spectra seem to be more reasonable than the estimates from the standard CLSW process in that the spectral power is shown near where it is expected to be based on the frequency structure of data.

However, statistical theories for the CLSW and modified CLSW processes are not established to justify the use of EWS for spectral analysis. For example, in the discrete LSW process, the motivation of estimating spectral characteristics through the EWS is based on the relationship between the EWS, $S_j(k/T)$, and the auto-covariance, $\mathrm{Cov}(X_t, X_{t+k})$, proved in Nason *et al.*, 2000. However, in this thesis, we did not consider the auto-covariance of the CLSW process.

Therefore, an important direction for future work is to consider theoretical arguments for the future research including the relationship between the auto-covariance and EWS to get a justification to estimate spectral characteristics through the EWS, $S_a(b/T)$. However, the theoretical foundation of the discrete and continuous LSW processes are based on the asymptotic behaviours derived from the assumption of Lipschitz continuity of $W_j(k/T)$ and $W_a(b/T)$. As previously mentioned in §5.6, the dimension expansion of wavelet scales and their high correlations in the CLSW process may affect the asymptotic computation, and therefore we may need stronger assumptions.

8.3.3 Faster Computing Technique

Calculations of the modified CLSW process are in *R*. In particular, selecting a scale set is done via the package, "Rfast", Papadakis *et al.* (2020) in R. This *R* package efficiently chooses the basis functions adaptively to data that we have, but as the number of functions increases, so does the computational load. The horizontal tomography data in Chapter 4 has 3072 observations with 1024 possible Haar wavelet scales for classification modelling. For this case, the total number of basis functions is 3,145,728, so the "Rfast" package cannot work with the high dimensional matrix of basis functions. To speed up the computing process and enable analysis of larger data sets, we could consider implementation in fast complied languages such as *C++*. Otherwise, it may be possible to develop improved algorithms that work faster for scale selection.

8.3.4 The Applications of Different Wavelet Functions

In this thesis, only the Haar wavelet function has been used for the construction of the CLSW process. The Haar wavelet function is the simplest wavelet function and is advantageous in terms of easy interpretation. However, as the Haar wavelet function has discontinuity, the CWT with the Haar wavelet function may not work very well depending on the wavelet scale set that we use. Therefore, we can consider other wavelet functions for the CLSW process. There are various of wavelet functions available to use in the DWT and CWT. For example, Mexican Hat wavelet and Daubechies' other wavelet functions can be considered for the real-valued wavelet transform. However, our CLSW process assumes that a wavelet function is compactly supported, especially for the invertible operator A. Therefore, we should select a wavelet function to have a compact support.

The first approach that we can consider is the approximation of compactly supported wavelet function used in the DWT. Daubechies (1988) constructed orthonormal and compactly supported basis functions with strict regularity conditions. However, they are based on the multi-resolution analysis using the concept of filters. In the CWT, as there no longer exists a relationship between scales such as the multi-resolution analysis, the Daubechies' wavelet function is difficult to generate with continuous scales. Instead, we can get an approximate function

for non-discrete wavelet scales by a linear interpolation. This allows us to estimate the wavelet function values from intermediate wavelet levels based on the given discrete wavelet function values. Suppose that we use the Daubechies 4 tap wavelet function for the CLSW process. Figure 8.3.1 shows the Daubechies tap 4 wavelet function with the length of 16 at the level, $j = 2$. Then, we can use the linear interpolation values as approximations at the intermediate x points.

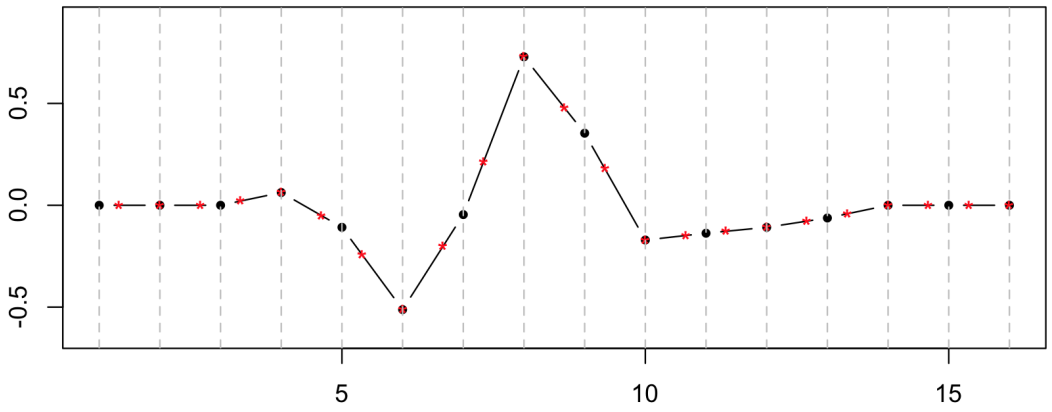


Figure 8.3.1: The example of linear interpolation with the Daubechies 4 tap wavelet function: given values (black dots) and approximations (red dots).

But, we need to re-consider the scaling constant for the generation of a wavelet function because of the relationship of the continuous wavelet functions,

$$\psi_{a,b}(x) = \sqrt{\frac{1}{a}}\left(\frac{1}{a}(x-b)\right). \tag{8.3.1}$$

The scaling constant can be determined depend on the model function used for approximation. Figure 8.3.2 shows the example of approximation of a wavelet function to the intermediate wavelet levels. The model function on the left is the wavelet function at the level, $j = 2$ with 1024 observations. The discrete wavelet resolution level corresponds to the continuous level, $a = 256/1024$. By a linear interpolation, the line plot on the right shows the approximate wavelet function at $j = 1.415$ or $a = 384/1024$ considering a scaling constant.

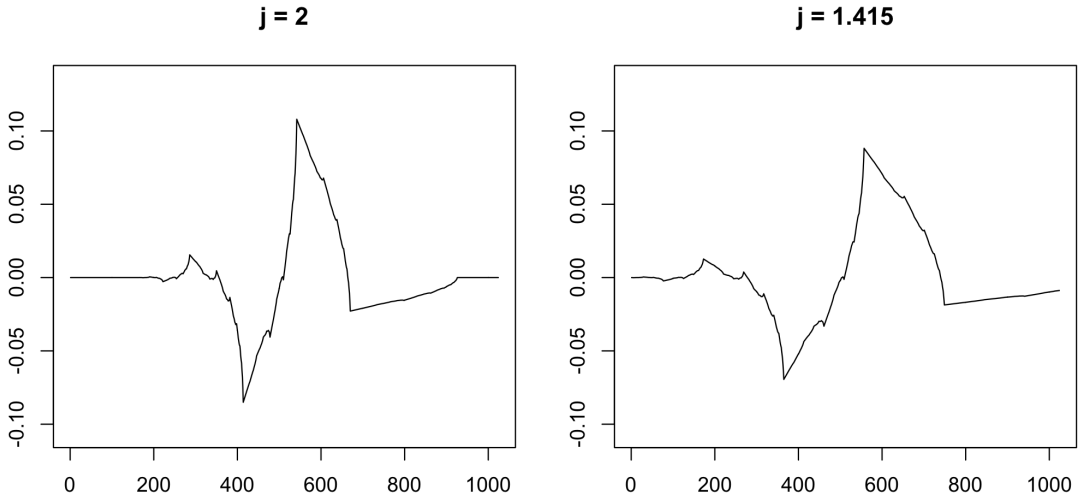


Figure 8.3.2: The Daubechies 4 tap wavelet function.

The second approach is to truncate a wavelet function to have a compact support. For example, the Mexican Hat wavelet function is the second derivative of the Gaussian function. This function is known to be useful for sensitive signal data such as seismic data. The wavelet function does not have a compact support, and therefore we cannot guarantee the existence of the invertible operator A for the correction of bias. To overcome the limitation, we can store a portion of the true wavelet function that we want and truncate the rest to zero. However, the conditions for the CWT, such as admissibility, are no longer valid for those two cases. However, in actual data analysis, if we are careful to keep the original structure of a wavelet function as much as we can, it may not make much effect on results.

References

Adler, A., Boyle, A., Braun, F., Crabb, M.G., Grychtol, B., Lionheart, W.R., Tregidgo, H.F. & Yerworth, R. (2017). EIDORS Version 3.9. *ELECTRICAL IMPEDANCE TOMOGRAPHY*, **21**, 63. 27

Aharon, M., Elad, M. & Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, **54**, 4311–4322. 94

Antoniadis, A. (2007). Wavelet Methods in statistics: some recent developments and their applications. *Statistics Surveys*, **1**, 16–55. 12

Aykroyd, R.G., Barber, S. & Miller, L.R. (2016). Classification of multiple time signals using localized frequency characteristics applied to industrial process monitoring. *Computational Statistics & Data Analysis*, **94**, 351–362. ii, 2, 26, 27, 29, 32, 36, 133, 135

Behkish, A., Lemoine, R., Sehabiague, L., Oukaci, R. & Morsi, B.I. (2007). Gas holdup and bubble size behavior in a large-scale slurry bubble column reactor operating with an organic liquid under elevated pressures and temperatures. *Chemical Engineering Journal*, **128**, 69–84. 25

Bergeaud, F. & Mallat, S. (1995). Matching pursuit of images. In *Proceedings., International Conference on Image Processing*, vol. 1, 53–56, IEEE. 94

Bicego, M., Murino, V. & Figueiredo, M.A. (2004). Similarity-based classification of sequences using hidden markov models. *Pattern Recognition*, **37**, 2281–2291. 94

Birr, S., Volgushev, S., Kley, T., Dette, H. & Hallin, M. (2017). Quantile spectral analysis for locally stationary time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 1619–1643. 70

Brauner, N. & Barnea, D. (1986). Slug/churn transition in upward gas-liquid flow. *Chemical Engineering Science*, **41**, 159–163. 25

Cardinali, A. & Nason, G.P. (2017). Locally stationary wavelet packet processes: basis selection and model fitting. *Journal of Time Series Analysis*, **38**, 151–174. 73, 93

Chen, S.S., Donoho, D.L. & Saunders, M.A. (2001). Atomic decomposition by basis pursuit. *SIAM review*, **43**, 129–159. 94

Chipman, H.A., Kolaczyk, E.D. & McCulloch, R.E. (1997). Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, **92**, 1413–1421. 12

Cho, H. & Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **77**, 475–507. 71

Christov, I., Gómez-Herrero, G., Krasteva, V., Jekova, I., Gotchev, A. & Egiazarian, K. (2006). Comparative study of morphological and time-frequency ecg descriptors for heartbeat classification. *Medical Engineering & Physics*, **28**, 876–887. 94

Chu, S., Narayanan, S. & Kuo, C.C.J. (2009). Environmental sound recognition with time–frequency audio features. *IEEE Transactions on Audio, Speech, and Language Processing*, **17**, 1142–1158. 94

Coifman, R.R. & Wickerhauser, M.V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, **38**, 713–718. 93

Corneliussen, S. (2005). *Handbook of Multiphase Flow Metering*. Norwegian Society for Oil and Gas Measurement. 46

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 1–37. 68, 69, 70

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Annals of Statistics*, 1762–1794. 68, 69

Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of Statistics*, vol. 30, 351–413, Elsevier. 69

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, **41**, 909–996. 140

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on Information Theory*, **36**, 961–1005. 11

Daubechies, I. (1992). *Ten lectures on wavelets*, vol. 61. SIAM. 9, 11, 14

Donoho, D.L. & Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 425–455. 121

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 301–337. 12

Duffin, R.J. & Schaeffer, A.C. (1952). A class of nonharmonic Fourier series. *Transactions of the American Mathematical Society*, **72**, 341–366. 22

Dyke, P.P. (2014). *An introduction to Laplace Transforms and Fourier Series*. Springer. 5

Elad, M. & Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, **15**, 3736–3745. 94

Fryzlewicz, P. & Nason, G.P. (2006). Haar–Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 611–634. 123

Fryzlewicz, P. & Ombao, H. (2009). Consistent classification of nonstationary time series using stochastic wavelet representations. *Journal of the American Statistical Association*, **104**, 299–312. 133

Fryzlewicz, P.Z. (2003). *Wavelet techniques for time series and Poisson data*. Ph.D. thesis, University of Bristol. 123

Goodwin, D., Aykroyd, R. & Barber, S. (2005). Modelling and predicting flow regimes using wavelet representations of ert data. In *Proceedings of the 4th World Congress on Industrial Process Tomography*, 904–909. 26

Haar, A. (1911). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, **71**, 38–53. 9

Jawerth, B. & Sweldens, W. (1994). An overview of wavelet based multiresolution analyses. *SIAM review*, **36**, 377–412. 11

Johnstone, I.M. & Silverman, B.W. (2005). Empirical bayes selection of wavelet thresholds. *Annals of Statistics*, 1700–1752. 12, 121

King, G. & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, **9**, 137–163. 138

Knight, M.I., Nunes, M.A. & Nason, G.P. (2012). Spectral estimation for locally stationary time series with missing observations. *Statistics and Computing*, **22**, 877–895. 71

Krzemieniewska, K., Eckley, I.A. & Fearnhead, P. (2014). Classification of non-stationary time series. *Stat*, **3**, 144–157. 133

Mallat, S. (2008). *A wavelet tour of signal processing: the sparse way*. Academic Press. 11

Mallat, S., Papanicolaou, G. & Zhang, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Annals of Statistics*, 1–47. 69

Mallat, S.G. & Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, **41**, 3397–3415. 93, 94

McClusky, H.L., Holloway, M.V., Beasley, D.E. & Ochterbeck, J.M. (2002). Continuous wavelet transforms of instantaneous wall pressure in slug and churn upward gas-liquid flow. *Journal of Fluids Engineering*, **124**, 625–633. 29

Nason, G. (2010). *Wavelet methods in Statistics with R*. Springer Science & Business Media. 9, 11, 123

Nason, G. (2016). *wavethresh: Wavelets Statistics and Transforms*. R package version 4.6.8. 84

Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 463–479. 12

Nason, G.P. & Silverman, B.W. (1995). The stationary wavelet transform and some statistical applications. In *Wavelets and statistics*, 281–299, Springer. 16, 17

Nason, G.P., Von Sachs, R. & Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**, 271–292. ii, 2, 70, 71, 72, 73, 75, 76, 87, 101, 120, 123, 136, 139

Nelson, J.D. & Gibberd, A.J. (2016). Introducing the locally stationary dual-tree complex wavelet model. In *Image Processing (ICIP), 2016 IEEE International Conference on*, 3583–3587, IEEE. 74

Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., Borboudakis, G., Burkardt, J., Zou, C., Lakiotaki, K. & Chatzipantsiou., C. (2020). *Rfast: A Collection of Efficient and Extremely Fast R Functions*. R package version 1.9.9. 104, 140

Papandreou-Suppappola, A. (2018). *Applications in Time-Frequency Signal Processing*. CRC press. 135

Park, T., Eckley, I.A. & Ombao, H.C. (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Transactions on Signal Processing*, **62**, 5240–5250. 71

Pati, Y.C., Rezaiifar, R. & Krishnaprasad, P.S. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, 40–44, IEEE. 94, 97

Percival, D.B. & Walden, A.T. (2000). *Wavelet methods for time series analysis*, vol. 4. Cambridge university press. 11

Pradeep, C. (2015). *Tomographic approach to automatic and non-invasive flow regime identification*. Ph.D. thesis, Telemark University College. 47

Prasser, H.M., Krepper, E. & Lucas, D. (2002). Evolution of the two-phase flow in a vertical tubedecomposition of gas fraction profiles according to bubble size classes using wire-mesh sensors. *International Journal of Thermal Sciences*, **41**, 17–28. 26

Priestley, M.B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, **27**, 204–229. 68

Qin, L., Guo, W. & Litt, B. (2009). A time-frequency functional model for locally stationary time series data. *Journal of Computational and Graphical Statistics*, **18**, 675–693. 69

Rafiee, J., Rafiee, M., Prause, N. & Schoen, M. (2011). Wavelet basis functions in biomedical signal processing. *Expert Systems with Applications*, **38**, 6190–6201. 12

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163. 125

Rouhani, S. & Sohal, M. (1983). Two-phase flow patterns: A review of research results. *Progress in Nuclear Energy*, **11**, 219–259. 25

Silverman, B.W. & Johnstone, I. (2005). Ebayesthresh: R programs for empirical bayes thresholding. *Journal of Statistical Software*, **12**. 121

Taitel, Y., Bornea, D. & Dukler, A. (1980). Modelling flow pattern transitions for steady upward gas-liquid flow in vertical tubes. *AIChE Journal*, **26**, 345–354. 25

Tropp, J.A. (2004). Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, **50**, 2231–2242. 94

Tropp, J.A. & Gilbert, A.C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, **53**, 4655–4666. 94

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. Wiley. 9, 11, 14, 121

Vincent, P. & Bengio, Y. (2002). Kernel matching pursuit. *Machine Learning*, **48**, 165–187. 95

von Sachs, R., Nason, G.P. & Kroisandt, G. (1997). Adaptive estimation of the evolutionary wavelet spectrum. *University of Kaiserslautern, Germany, and University of Bristol, UK. October*, **15**. 120, 123

Wang, D., Miao, D. & Xie, C. (2011). Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Systems with Applications*, **38**, 14314–14320. 93

Wang, H., Killick, R. & Fu, X. (2014). Distributional change of monthly precipitation due to climate change: comprehensive examination of dataset in southeastern united states. *Hydrological Processes*, **28**, 5212–5219. 71

Wilson, R.E., Eckley, I.A., Nunes, M.A. & Park, T. (2019). Dynamic detection of anomalous regions within distributed acoustic sensing data streams using locally stationary wavelet time series. *Data Mining and Knowledge Discovery*, **33**, 748–772. 71

Zhao, X., Barber, S., Taylor, C.C. & Milan, Z. (2018). Classification tree methods for panel data using wavelet-transformed time series. *Computational Statistics & Data Analysis*, **127**, 204–216. 32