

WAVELET METHODS FOR TIME
SERIES WITH BIVARIATE
OBSERVATIONS AND IRREGULAR
SAMPLING GRIDS



Jean Sanderson
School of Mathematics

September 2010

A DISSERTATION SUBMITTED TO THE UNIVERSITY OF BRISTOL
IN ACCORDANCE WITH THE REQUIREMENTS OF THE DEGREE
OF DOCTOR OF PHILOSOPHY IN THE FACULTY OF SCIENCE

Abstract

This thesis explores the use of wavelet techniques to estimate the spectral properties of nonstationary time series and, in particular, the dependence between pairs of nonstationary signals.

The first part of this thesis addresses the problem of estimating the dependence between two regularly sampled time series. A new dependence measure is proposed, derived from a bivariate locally stationary wavelet time series model. Since wavelets are localised in both time and in scale, this approach leads to a natural, local and multiscale estimate of nonstationary dependence. The methodology is illustrated by application to an example from neuroscience.

Next we consider the use of wavelet lifting to compute the periodogram of signals that are not observed at regular sampling intervals. This work is motivated by application to palaeoclimate time series, which are intrinsically unevenly sampled. The relationship between wavelet lifting scale and Fourier frequency is investigated and the use of empirical mode decomposition is also considered, providing an interesting comparison to the lifting methodology.

Finally, extension of the wavelet lifting methodology to provide measures of coherence and phase between two irregularly sampled time series is considered. A lifting scheme that produces complex valued coefficients is proposed, and this scheme is used to define the wavelet lifting cross-periodogram. The methodology is demonstrated with respect to bivariate irregularly sampled palaeoclimate series and compared to the methodology of the first two chapters using a simulated example with regularly sampled observations.

Acknowledgements

Firstly, I would like to thank my supervisor, Jonty Rougier, for all his help and support over the last couple of years. Special thanks also go to Piotr Fryzlewicz, for his supervision during the earlier days of my PhD, and continued support thereafter.

Thanks to Matt Jones, Michel Crucifix and Matt Nunes for their help with the application of this work

I would also like to thank the rest of the Maths Department for providing a supportive and friendly atmosphere throughout my time in Bristol. And a final thanks to friends elsewhere, and to my family for looking after me as always.

Author's Declaration

I declare that the work in this thesis was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree. Any views expressed in the dissertation are those of the author and do not necessarily represent those of the University of Bristol. The thesis has not been presented to any other university for examination either in the United Kingdom or overseas.

Jean Sanderson

Date: September 2010

Contents

Abstract	iii
Acknowledgements	v
Author's Declaration	vii
1 Introduction	1
2 Literature review	4
2.1 Wavelet theory	4
2.1.1 Review of Fourier analysis	4
2.1.2 Wavelets	6
2.1.3 Multiresolution analysis	12
2.1.4 Discrete wavelet transform	15
2.1.5 Nondecimated wavelet transform	19
2.1.6 Extensions to classical wavelet methods	21
2.1.7 Statistical applications of wavelets	22
2.2 The wavelet lifting scheme	23
2.2.1 Computational approach to wavelet lifting	24
2.2.2 One coefficient at a time wavelet lifting	28
2.3 Time series analysis	35
2.3.1 Bivariate methods	37
2.3.2 Spectral analysis	39
2.3.3 Allowing for nonstationarity	41

2.3.4	The locally stationary wavelet model	43
3	Locally stationary wavelet coherence	49
3.1	The bivariate LSW time series model	51
3.2	The evolutionary wavelet cross-spectrum and coherence	55
3.3	Estimation	58
3.4	Relationship between the LSW and Fourier coherence	66
3.5	Phase estimation	71
4	Locally stationary wavelet coherence: application	73
4.1	Practical considerations	73
4.1.1	Stability of the estimator	73
4.1.2	The effect of J^*	81
4.1.3	The effect of smoothing over time	82
4.1.4	Other issues	83
4.2	Applications	84
4.2.1	Simulated example	84
4.2.2	Application to neuroscience data	84
4.3	Summary and further work	91
5	Periodogram estimation for irregular situations using wavelet lifting	93
5.1	The nondecimated lifting transform	96
5.2	The nondecimated lifting periodogram	98
5.3	Relationship to Fourier frequency	103
5.4	Unblurring the periodogram	110
5.5	Relationship between the NLT periodogram and the EWS	114
5.6	Empirical mode decomposition	115
5.7	Application to palaeoclimate time series	123
5.7.1	Characteristics of palaeoclimate data	124
5.7.2	Orbital forcing data	126

5.7.3	Analysis	128
5.8	Method comparison	134
5.9	Summary	135
6	Bivariate spectral analysis for irregular situations	138
6.1	The complex one coefficient at a time lifting scheme	139
6.1.1	Filters for the complex lifting scheme	144
6.1.2	The complex nondecimated lifting periodogram	148
6.2	The bivariate complex nondecimated lifting transform	149
6.2.1	Relationship to frequency and filter design	153
6.2.2	Example on simulated data	156
6.2.3	Physical meaning of phase	161
6.2.4	Choice of filters	165
6.3	Application to regularly sampled time series	167
6.4	Application to real data	171
6.5	Summary and further work	173
7	Conclusions	176

List of Tables

6.1	Relationship to frequency for the different bivariate CNLT periodogram components: periodogram ($C I_{x_k}$), co-periodogram (c_{x_k}) and quadrature periodogram (q_{x_k}). The cross periodogram, phase and coherence are formed using combinations of these elements. Here we use the value of α that maximises the wavelet coefficients to denote the relationship to frequency.	154
6.2	Approximate relationship between wavelet scale, Fourier frequency and wavelet lifting scale. The scale-frequency relationship for the wavelet lifting scheme depends on the chosen prediction method and so we show the relationships for two different lifting schemes; The univariate lifting scheme using linear prediction with one neighbour either side ($\alpha = \log_2(P/2)$), and the bivariate scheme based on this filter ($\alpha = \log_2(P/3)$).	169

List of Figures

2.1	Haar mother wavelet, $\psi_{1,1}(x)$, and translated/dilated Haar mother wavelet, $\psi_{3,3}(x)$	8
2.2	Daubechies Extremal phase (top row) and Least Asymmetric (bottom row) wavelets with N vanishing moments, plotted using Wavethresh (Nason et al., 2008).	11
2.3	Diagram of the decomposition of a function using the DWT through successive applications of the high and low pass filters g and h	17
2.4	Schematic of stage n of the general lifting transform.	25
2.5	Stage n of the one coefficient at a time lifting scheme using linear prediction with one neighbour either side of the removal point. The first figure shows the prediction of point $j_n = 4$ and the second figure shows the process of updating the neighbours and integrals.	34
3.1	Example of the construction of a bivariate LSW process. a) EWS of $X^{(1)}$ and $X^{(2)}$ b) LSW coherence between the series, c) simulation of $X^{(1)}$ (black) and $X^{(2)}$ (red).	54
3.2	Autocorrelation wavelets and corrected autocorrelation wavelets for Haar (panels a and c) and Daubechies least asymmetric $N = 6$ (panels b and d).	67

4.1	Piecewise constant Fourier coherence. The true coherence structure is shown in black and the estimated coherence for one realisation is shown in blue.	79
4.2	Comparison of stabilisation methods: 1) uncorrected 2) corrected 3) corrected with regularisation 4) corrected with truncated correction matrix 5) corrected and scale-smoothed. The true coherence structure is shown by the dashed line.	80
4.3	Estimated stationary LSW coherence for different choices of J^* . Here we plot the mean over 200 simulations and the choice of J^* is indicated by the length of the corresponding estimate. The underlying coherence structure is indicated by the dashed line.	82
4.4	Effect of the width of the smoothing window on the variance of the periodogram of a white noise process. Choices of $M = 0, 10, 25, 100, 200$ and T are considered.	83
4.5	Results showing the true coherence in grey, and estimated coherence in black from the simulation example. Subplots a) to f) correspond to scales -1 to -6 respectively.	85
4.6	Trial averages for choice runs (a) choice forced-turn runs (b) scales -1 to -5 using Daubechies extremal phase wavelet with $N = 8$ vanishing moments.	87
4.7	LSW coherence for scale -4 with means before and after the reference point plotted in bold. The associated confidence intervals are plotted with dashed lines.	88
4.8	Trial-averaged central-arm coherence, reproduced from Jones and Wilson (2005, Figure 7b). Central-arm coherence is subdivided into forced-turn (grey) and choice (red) directions. Dashed line marks 95 % confidence level, with shaded band thickness corresponding to jackknife error bars (estimated over trials and nine tapers).	90

4.9	Trial-averaged coherence for the choice trials (a) and forced-turn trial (b) using the short time Fourier transform with a window width of 0.5 seconds, run at 0.005 second increments.	90
5.1	Relationship between wavelet lifting scale and removal order for the example illustrated in Figure 2.5. When the point x_4 is removed at the first stage of the algorithm it is associated with a small integral as shown in subplot a. When the point is removed after points x_2, x_3 and x_5 , it is associated with a larger integral, as shown in subplot b.	98
5.2	Simulated time series with irregular spacing. The mean sampling rate is 2 and the series contains oscillations with periods 25 and 80. The middle section between the vertical lines contains oscillations at both frequencies.	102
5.3	Estimating the periodogram ordinates for a given time point, x_{200} , using a) the average in each group b) nonparametric regression with splines. Estimation of the periodogram using c) moving average d) splines. All estimates are smoothed over time using a smoothing window of 30.	102
5.4	Relationship between the integral and associated detail coefficients using one neighbour either side linear prediction and a sine wave of period 10. Removing the point marked with the blue cross it is clear that the largest coefficient will be obtained by considering an integral of 5.	105
5.5	Effect of uneven sampling on the form of the resulting wavelet. Here demonstrated for the removal of a point with an integral of 5. The resulting wavelet for $\beta = 1$ is shown in blue and for $\beta = 10$ in red.	107

5.6 Average value of the squared detail coefficients, $\bar{d}(\alpha)^2$, against scale, α , for a sinusoid with period P . The different lines represent an increasing degree of asymmetry in the sampling: $\beta = 1, 1.5, 2, 3, 4$ 107

5.7 Lifting periodogram of the simulated time series illustrated in Figure 5.2, decomposed according to the degree of asymmetry of the neighbours (β) involved with production of each coefficient: a) all detail coefficients b) detail coefficients that satisfy $\beta \leq 1.45$ c) coefficients that satisfy $1.45 < \beta \leq 2.45$ and d) coefficients that satisfy $\beta > 2.4$ 108

5.8 Average value of the squared detail coefficients, $\bar{d}(\alpha)^2$, against scale, α , obtained with a linear prediction scheme for different sizes of neighbourhood. 109

5.9 Different methods for unblurring the asymmetric periodogram: a) setting a cut-off value of $\beta_{\text{cut}} = 2.5$ b) favouring the removal of symmetric points c) searching for symmetric neighbours d) value of β for initial points. 113

5.10 Example of EMD decomposition using a small section of the orbital forcing data introduced later in Section 5.7.2. a) Estimation of the upper and lower envelopes, b) Difference between the data and the mean of the envelopes. This is not an IMF so the sifting process is repeated. c) Final IMF decomposition. . . 119

5.11 Time series (left) and wavelet periodograms (right) for the orbital forcing and carbon dioxide signals. 129

5.12 IMF decomposition (left) and associated lifting periodograms (right) for the orbital forcing data. 130

5.13 IMF decomposition for the Carbon Dioxide data using a tolerance level of $0.175^2\sigma_X$ (red) and a tolerance of $0.125^2\sigma_X$ (blue). 131

5.14 Lifting periodograms of the IMF components for the carbon dioxide data, decomposed using a tolerance level of $0.175^2\sigma_X$. . 132

5.15 Lifting periodograms of the IMF components for the carbon dioxide data, decomposed using a tolerance level of $0.125^2\sigma_X$. . . 133

6.1 The complex OCAAT lifting scheme. The black lines correspond to the prediction and update steps of the standard OCAAT lifting scheme and the blue lines indicate the extra prediction step required for the complex scheme. 143

6.2 An alternative complex OCAAT lifting scheme, consisting of two separate branches with updating occurring in both. Starting with the same data vector \mathbf{c}_n , the coefficients at later stages, \mathbf{c}_{n-i} , \mathbf{c}_{n-i}^* , may be associated with different scales. 143

6.3 Relationship between the integral and maximum value of the associated detail coefficients for the prediction schemes defined by filters \mathbf{L} (blue) and \mathbf{M} (red), here shown for a sine wave of period 10. The integrals are shown by the double solid lines and the detail coefficients by the dashed lines. For scheme \mathbf{L} the coefficients are maximised by considering an integral of $P/2$ and for scheme \mathbf{M} the coefficients are maximised by considering an integral of $P/4$ 146

6.4 Relationship between detail coefficients and integral for $m_3 = \frac{l+1}{\sqrt{3}}$. This ensures that combining the coefficients by addition and multiplication results in the same relationship to frequency, as shown by the overlap of the red and green lines. 155

6.5 Relationship between detail coefficients and integral for $m_3 = 1$. Using this weighting the coefficients $\lambda_{x_k}^p$ and $\mu_{x_k}^p$ have the same maximum value. However combining the coefficients by addition or multiplication leads to different relationships to frequency, as shown by the different positions of the red and green lines in this example. 155

6.6 Simulated bivariate time series with irregular sampling. $X^{(1)}$ is shown in black and $X^{(2)}$ in red. The vertical line indicates the location where the time lag is introduced. 157

6.7 Estimation of coherence using the real-valued bivariate lifting scheme. a) periodogram of $X^{(1)}$ b) periodogram of $X^{(2)}$, c) cross periodogram and d) coherence. 159

6.8 Estimation of coherence and phase using the complex lifting scheme. a) periodogram of $X^{(1)}$ b) periodogram of $X^{(2)}$, c) co-periodogram d) quadrature periodogram e) coherence and f) phase. 160

6.9 Phase shift between $X^{(1)}$ and $X^{(2)}$ from example 1 as a function of scale. The phase is estimated for $\tau = 0, 1, 2, 5$ and averaged over time. 162

6.10 Coherence between $X^{(1)}$ and $X^{(2)}$ from example 1 as a function of scale. The coherence is estimated for $\tau = 0, 1, 2, 5$ and averaged over time. 162

6.11 Coherence between $X^{(1)}$ and $X^{(2)}$ from example 2 as a function of scale, for various choices of τ . The vertical lines correspond to Fourier periods of 30 and 60. 164

6.12 Phase between $X^{(1)}$ and $X^{(2)}$ from example 2 as a function of scale, for various choices of τ 164

6.13 Phase at scale $\log_2(30/3)$, equivalent to a Fourier period of 30, as a function of the time delay between the series. The red points show the resulting estimates from the simulations and the black line indicates the observed trend. 165

6.14 Coherence (plots a-d) and phase (plots e-h) for different degrees of dependence between the filters; $d_f = 0$ (plots a and e), $d_f = 0.5$ (plots b and f), $d_f = 0.75$ (plots c and g) and $d_f = 0.9$ (plots f and h). 166

6.15 Data for simulated example: $X^{(1)}$ (black) and $X^{(2)}$ (red) 169

6.16	Estimated coherence for the data in Figure 6.15. a) LSW coherence b) wavelet lifting coherence using 6 artificial scales c) wavelet lifting coherence using 12 artificial scales d) wavelet lifting phase using 12 artificial scales.	170
6.17	Time series of isotopic temperature proxies: a) deuterium ($\delta\mathbf{D}$) and b) oxygen ($\delta^{18}\mathbf{O}$).	171
6.18	Coherence (a) and phase (b) between the $\delta\mathbf{D}$ and $\delta^{18}\mathbf{O}$ series. .	172

Chapter 1

Introduction

Time series arise from many diverse and different fields, from economics to climatology. Time series analysis is therefore an important topic for many different applications. For series that satisfy certain properties, such as stationarity and regular spacing of observations, there is much literature on well-established analysis methods. This thesis is primarily concerned with analysis methods for nonstationary time series with characteristics that deviate from the standard assumptions. We consider the analysis of bivariate signals, and time series that are observed with irregular sampling intervals. We are interested in describing the variation within the series and, in the case of bivariate series, also the dependence between the series.

Wavelets can be thought of as localised, oscillatory basis functions with several attractive properties for function representation. They are localised in both time and frequency, providing sparse multiscale representations for many signals. Due to their time localisation, wavelets provide natural ‘building blocks’ for nonstationary series. Classical wavelet methods also suffer from some limitations; For example, they are restricted to cases where the data is sampled at regular time intervals. Second generation wavelets provide a generalisation of classical wavelets, allowing for greater flexibility in applications.

Chapter 2 reviews the literature in the area of time series analysis using wavelets. It summarises the basic concepts from wavelet theory, including

the discrete wavelet transform (DWT), and describes the locally stationary wavelet (LSW) time series model of Nason et al. (2000). The LSW model uses classical nondecimated wavelets as building blocks, and is therefore a useful tool for modelling multiscale signals with time varying characteristics. This model provides the foundations for our work in Chapters 3 and 4. The *lifting scheme* is also introduced, providing a method of constructing second generation wavelets and wavelet transforms, suitable for situations involving departures from the assumptions of the DWT. The chapter includes a review of the *one coefficient at a time* lifting scheme of Jansen et al. (2001). This scheme is a key component of the irregular design methodology of Chapters 5 and 6.

Throughout this thesis our work is motivated by its usefulness in its application to specific scientific issues. Chapter 4 considers the problem of estimating the dependence between observations in two functionally and anatomically connected areas of a rat's brain: the hippocampus and the prefrontal cortex. The estimated dependence provides an indication of the extent to which activities in the two areas are coordinated. Later, in Chapters 5 and 6, the analysis of palaeoclimate time series is considered. These signals are inherently irregularly spaced, with the sampling becoming coarser as we look further back in time.

Chapter 3 addresses the problem of estimating the dependence between two regularly sampled nonstationary time series. The bivariate LSW model is introduced, providing an extension to the univariate LSW model (Nason et al., 2000). Stemming from this formulation, we propose a novel measure of wavelet coherence termed 'locally stationary wavelet coherence', which provides a direct measure of the linear dependence between two nonstationary series. Chapter 4 addresses several issues relating to the practical implementation of our methodology, discussing several options for ensuring stability. The method is then demonstrated on a simulated example as well as real experimental data from neuroscience. The work of this chapter also appears in the

paper by Sanderson et al. (2010).

The work of Chapter 5 is motivated by the problem of estimating the spectral characteristic of irregularly spaced time series. The chapter first reviews the work of Knight and Nason (2008) who propose a nondecimated lifting transform for irregular data, based on the one coefficient at a time lifting scheme (Jansen et al., 2001). This methodology can also be used to define a periodogram (Knight et al., 2010). Our contribution to this area arises from our chosen application in climatology. A relationship between wavelet lifting scale and Fourier frequency is established, aiding the physical interpretation of our results. We also discuss some of the limitations of the methodology and propose a technique for improving the clarity of the resulting periodogram estimates.

Chapter 6 extends the ideas of Chapter 5 to the bivariate setting, presenting a method that enables the estimation of the dependence between two irregularly sampled time series. To provide an estimate of phase between the series, a complex valued extension to the standard one coefficient at a time scheme of Jansen et al. (2001) is introduced. The work of this chapter is demonstrated by estimating the coherence and phase between two different temperature proxy records.

Chapter 7 concludes with a summary of our work and some interesting ideas for future research.

Chapter 2

Literature review

2.1 Wavelet theory

This chapter provides an overview of aspects of the literature which are essential to the remainder of this thesis. Firstly we provide a brief background to Fourier theory, which serves as a useful reference for the non-stationary methods that are introduced later on. Wavelets are then introduced, providing ‘building blocks’ which are naturally suited to non-stationary data. Section 2.2 reviews the wavelet *lifting scheme* which allows the principles of wavelet analysis to be extended to more general situations. This includes a description of the *one coefficient at a time lifting scheme* that is implemented in later chapters. Finally, Section 2.3 introduces the topic of time series analysis and in particular the locally stationary wavelet model.

2.1.1 Review of Fourier analysis

Before introducing wavelets, we start by first reviewing some relevant concepts from Fourier theory. An understanding of these methods provides motivation for the use of wavelets, since certain signals cannot be represented efficiently using Fourier sines and cosines. Fourier methods also provide a useful benchmark for comparison with the local methods introduced in later chapters. Our

review follows the description in Priestley (1981, Chapter 4) and the original text can be consulted for a more detailed discussion. Thorough descriptions of Fourier analysis and its applications can be found in e.g. Stade (2005) and Folland (2009).

In Fourier analysis, sine and cosine waves can be used to form bases for functions in $L^2(\mathbb{R})$, that is the space of *square integrable functions* defined by $g \in L^2(\mathbb{R})$ iff $\int_{-\infty}^{\infty} g(t)^2 dt < \infty$. We start by recalling the definition of Fourier series.

Definition 2.1.1. Let g be periodic with period 2π so that $g(\omega) = g(\omega + 2\pi)$ and be square integrable over the interval $[-\pi, \pi)$. Then the Fourier series representation of g is:

$$g(x) = \frac{a_0}{2} + \sum_n \left(a_n \cos(nx) + b_n \sin(nx) \right), \quad n \in \mathbb{Z},$$

where the Fourier coefficients are calculated from

$$a_n = \frac{1}{\pi} \int_0^{2\pi} g(x) \cos(nx) dx, \quad b_n = \frac{1}{\pi} \int_0^{2\pi} g(x) \sin(nx) dx.$$

The magnitude of the Fourier coefficients a_n and b_n reveals the frequency content of the signal. The Fourier sinusoids $\{\cos(nx), \sin(nx)\}_{n \in \mathbb{N}}$ form an orthonormal basis and hence can be thought of as basic ‘building blocks’ from which periodic functions of various forms may be constructed.

The decomposition of periodic functions into Fourier series representations can be extended to the case of non-periodic functions in $L^1(\mathbb{R})$, the space of *absolutely integrable functions* defined by $g \in L^1(\mathbb{R})$ iff $\int_{-\infty}^{\infty} |g(t)| dt < \infty$. The Fourier transform is defined as follows:

Definition 2.1.2. The Fourier transform of a function $g \in L^1(\mathbb{R})$ is given by

$$\hat{g}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(x) e^{-i\omega x} dx.$$

If \hat{g} is the Fourier transformation of g , and $\hat{g}, g \in L^1(\mathbb{R})$ then the inverse Fourier transform is given by

$$g(x) = \frac{1}{\sqrt{2\pi}} \int \hat{g}(\omega) e^{i\omega x} d\omega. \quad (2.1)$$

Here ω is the angular frequency, and ω is related to the Hertz frequency f by the relationship $\omega = 2\pi f$. Unlike the Fourier series decomposition of a periodic function which involves a discrete set of frequencies, in the Fourier integral representation, frequency varies on a continuous scale.

When representing a series by a combination of basis functions, it is generally desirable that the representation be sparse, in the sense that there are only a few non-zero coefficients. This is beneficial for understanding the structure of the signal and also leads to better signal compression. Sparse decompositions are achieved by using basis functions with similar properties to the function they are representing. Although Fourier sinusoids are localised in frequency, they are *non-local in time*. Therefore Fourier sinusoids are suitable for representing smooth functions which exhibit periodicity but do not provide economical representations for series with local features such as sharp changes and discontinuities. In order to represent functions of this type it is desirable to use basis functions which themselves are local. This requirement is met by wavelets, which we now introduce.

2.1.2 Wavelets

Wavelets can be thought of as small, localised oscillations. Unlike Fourier series, locality can be achieved in both the time and frequency domains simultaneously, providing a natural foundation for representing nonstationary functions. For an in-depth discussion of wavelets see Daubechies (1992), Mallat et al. (1989), or Vidakovic (1999). Following Vidakovic (1999, Chapter 3), we define a wavelet, $\psi \in L^2(\mathbb{R})$, to be any function which satisfies

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \quad (2.2)$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(x)$. Condition (2.2) is referred to as the *admissibility condition* and ensures that the wavelet is localised in frequency. The admissibility condition also implies that $\Psi(0) = 0$ so that

$$\int_{-\infty}^{\infty} \psi(x) dx = 0, \quad (2.3)$$

which ensures $\psi(x)$ is localised in time (as it implies $\psi \in L_1(\mathbb{R})$) and is oscillatory. Given a ‘mother’ wavelet, ψ , a sequence of wavelets can be constructed by taking translations and dilations of ψ . For dilation parameter $a \in \mathbb{R} \setminus \{0\}$ and translation parameter $b \in \mathbb{R}$ we define

$$\psi_{a,b}(x) = |a|^{-1/2} \psi\left(\frac{x-b}{a}\right). \quad (2.4)$$

The normalisation parameter in this representation, $|a|^{-1/2}$, ensures that the L^2 norm of ψ is independent of a and b : $\|\psi_{a,b}\| = \|\psi\|$. As described later, it is possible to choose a and b so that the collection $\{\psi_{a,b}\}_{a,b}$ forms an orthonormal basis.

An important property of a wavelet, useful for function representation, is the number of vanishing moments. The wavelet ψ is said to have $n + 1$ vanishing moments if

$$\int_{-\infty}^{\infty} x^k \psi(x) dx = 0, \quad \text{for all } k \in \{0, 1, \dots, n\}. \quad (2.5)$$

The vanishing moments property of a wavelet implies that the wavelet coefficients of polynomials of degree n or less are annihilated in a decomposition on such a wavelet basis. This property therefore has important implications for selecting a wavelet basis that will lead to economical representations of a given function.

Haar wavelets (Haar, 1910) are perhaps the simplest example of a wavelet system. The Haar mother wavelet is defined by

$$\psi^H(x) = \begin{cases} 1 & \text{for } x \in [0, 1/2), \\ -1 & \text{for } x \in [1/2, 1], \\ 0 & \text{otherwise.} \end{cases}$$

Using equation (2.4) the translations and dilations for $a \in \mathbb{R}^+, b \in \mathbb{R}$ of the Haar mother wavelet are given by

$$\psi_{a,b}^H(x) = \begin{cases} \frac{1}{\sqrt{a}} & \text{for } x \in [b, a/2 + b), \\ \frac{-1}{\sqrt{a}} & \text{for } x \in [a/2 + b, a + b), \\ 0 & \text{otherwise.} \end{cases}$$

The Haar mother wavelet is shown in Figure 2.1. It can be seen that the Haar wavelet has one vanishing moment and so the associated coefficients will be zero only for constant functions.

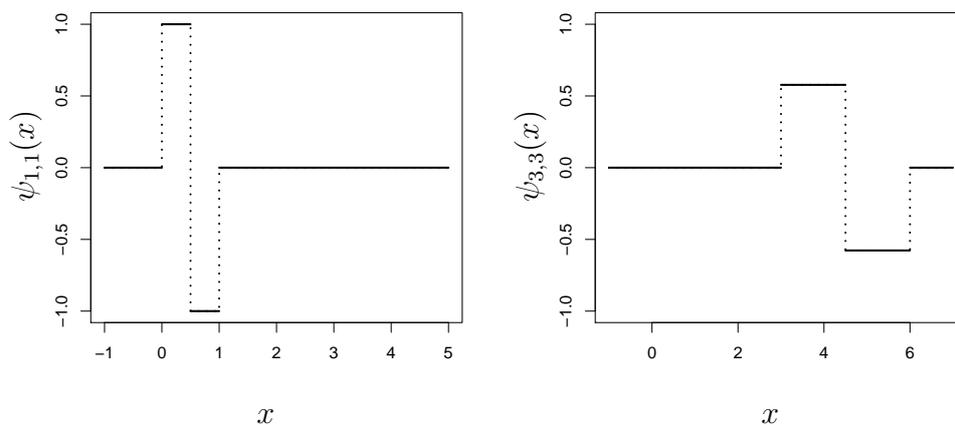


Figure 2.1: Haar mother wavelet, $\psi_{1,1}(x)$, and translated/dilated Haar mother wavelet, $\psi_{3,3}(x)$.

The Continuous wavelet transform

For a function $g(x) \in L^2(\mathbb{R})$, the continuous wavelet transform is defined as

$$CWT_g(a, b) = \langle g, \psi_{a,b} \rangle = \int g(x) \overline{\psi_{a,b}(x)} dx, \quad (2.6)$$

for $b \in \mathbb{R}$, $a \neq 0$ and where $\overline{\psi_{a,b}(x)}$ is the complex conjugate of $\psi_{a,b}(x)$. Assuming ψ satisfies the admissibility condition, it is possible to find the inverse continuous wavelet transformation via a relation which is known as the *resolution of the identity*:

$$g(x) = \frac{1}{C_\psi} \int_{\mathbb{R}^2} CWT_g(a, b) \psi_{a,b}(x) \frac{dad b}{a^2}, \quad (2.7)$$

where C_ψ is the quantity defined in equation (2.2). There are many wavelets that can be used here, including the Haar wavelet introduced earlier. For further examples of wavelets for the CWT see Vidakovic (1999, Section 3.1.2). Equation (2.7) shows that any $g \in L^2(\mathbb{R})$ can be approximated by a superposition of wavelet basis functions. Comparing this to the Fourier transform representation in equation (2.1), we see that in the Fourier decomposition the coefficients provide information about the amplitude associated with each frequency. In the wavelet decomposition the coefficients provide information on the amplitude of the wavelet both at a given scale (equivalent to frequency) and also location. Further information on the CWT can be found in Daubechies (1992, Chapter 2).

Discretisation

The work of Chapters 3 and 4 concentrates on discrete wavelet constructions rather than continuous representations as outlined above. The CWT is a redundant transform (i.e. unlike the Fourier sinusoids, the CWT wavelet bases are not orthonormal). However by discretising a and b it is possible to construct transformations with less redundancy. Given suitable choices of ψ and

restricting $a = 2^{-j}$, $b = k2^{-j}$ for $j, k \in \mathbb{Z}$, the so called dyadic translations and dilations given by

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \quad (2.8)$$

can be made to form an orthonormal basis for $L_2(\mathbb{R})$. This choice of a and b is known as the *critical sampling rate* (Vidakovic, 1999, Chapter 3). A finer choice of sampling rate will lead to a transformation with redundancy, while a coarser sampling rate will not maintain the invertibility of the transform.

Another common choice of a, b is given by $a = 2^{-j}$, $b = k$, which leads to a so called *non-decimated transform* (Vidakovic, 1999):

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j(x - k)). \quad (2.9)$$

Here wavelets exist at all integer locations rather than with dyadic spacing and the transformation is therefore *translation equivariant* (as will be discussed further in Section 2.1.5). The nondecimated transform is redundant.

Before discussing the implementation of these transformations, we first provide some further examples of wavelets which will be relevant to the later chapters of this thesis.

Examples of wavelets

The Daubechies Extremal Phase family of wavelets (Daubechies, 1992) are compactly supported with the minimum support possible for a given number of vanishing moments. These wavelets are indexed by the number of vanishing moments, n , and include the Haar wavelet system, specified when $n = 1$. The Daubechies Least Asymmetric family of wavelets are also of compact support and indexed by the number of vanishing moments but are, as the name suggests, constructed so that the resulting wavelets are closer to symmetric. These families of wavelets do not have a closed-form analytic representation in the time domain. A rigorous exploration of the properties and construction of these wavelets is provided in Daubechies (1992, Chapter 6,7) and Percival

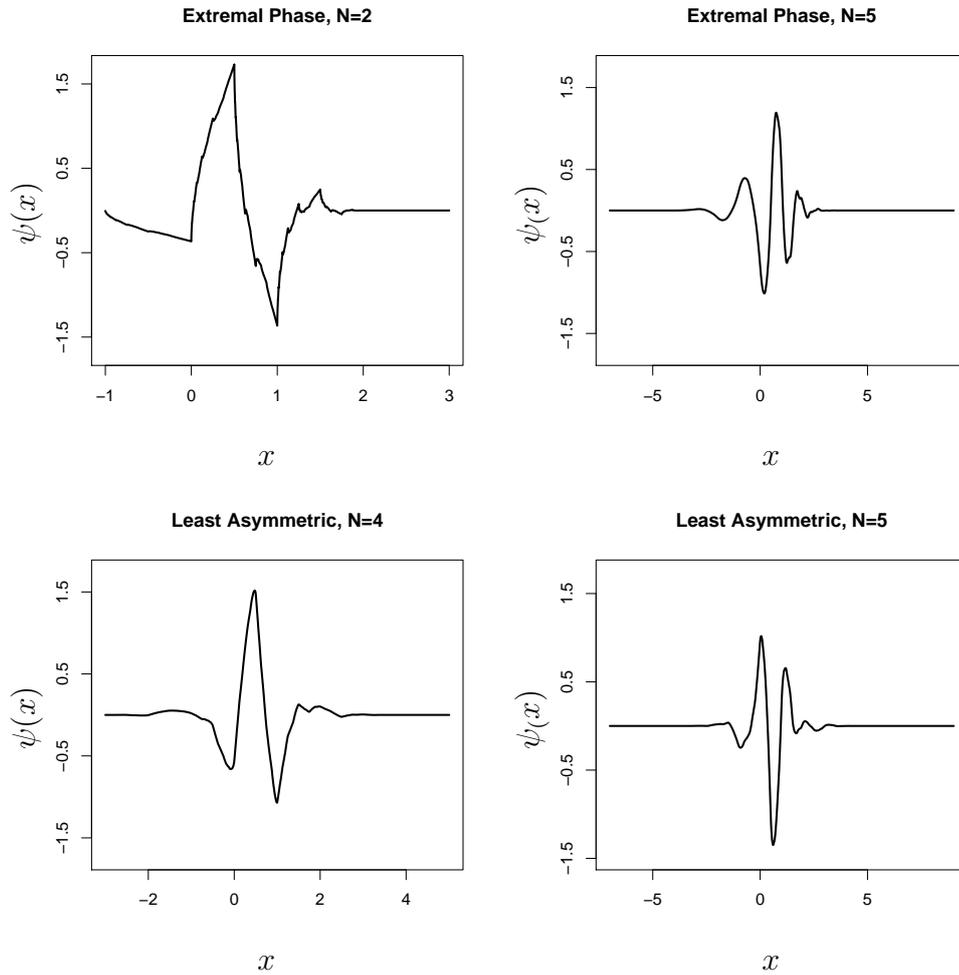


Figure 2.2: Daubechies Extremal phase (top row) and Least Asymmetric (bottom row) wavelets with N vanishing moments, plotted using Wavethresh (Nason et al., 2008).

and Walden (2000, Chapter 4). Examples of the Daubechies Extremal Phase and Least Asymmetric families of wavelets are shown in Figure 2.2.

The *Shannon wavelet* is compactly supported in the Fourier domain, but has infinite support in the time domain. Following Chui (1997), the Shannon mother wavelet and its Fourier transform are given by

$$\begin{aligned}\psi(x) &= \frac{\sin(2\pi x) - \cos(\pi x)}{\pi(x - 1/2)}, \\ \hat{\psi}(\omega) &= -\exp(-i\omega/2)\mathbb{I}_{[-2\pi, -\pi) \cup (\pi, 2\pi]}(\omega).\end{aligned}$$

Although rarely used in practice, its compact support in the Fourier domain means that the Shannon wavelet is often useful in a theoretical setting. As noted in Chui (1997), Daubechies wavelets with high numbers of vanishing moments ‘imitate’ the Shannon wavelet and so the Shannon wavelet can be thought of as the limiting wavelet of this family.

2.1.3 Multiresolution analysis

First proposed by Mallat et al. (1989) and Meyer (1992), a multiresolution analysis (MRA) is a mathematical structure which allows signals to be examined at different scales; ‘zooming out’ to look at the signal in an overall sense or ‘zooming in’ to study finer details. MRA is important in the construction of wavelet bases and also leads to the discrete wavelet transform. A multiresolution analysis is defined as follows:

Definition 2.1.3. A *multiresolution analysis* of $L^2(\mathbb{R})$ is a system of closed subspaces $\{V_j\}_{j \in \mathbb{Z}}$ of $L^2(\mathbb{R})$ such that

1. They lie in a containment hierarchy: $\dots V_{-1} \subset V_0 \subset V_1 \subset \dots$
2. $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$, i.e the union is dense in $L^2(\mathbb{R})$. Here the overbar denotes closure.
3. $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$, i.e the intersection is trivial.
4. $f(x) \in V_j \iff f(2x) \in V_{j+1}, \forall x \in \mathbb{R}$.
5. $f(x) \in V_0 \iff f(x - k) \in V_0 \forall k \in \mathbb{Z}, x \in \mathbb{R}$.
6. There exists a *scaling function* $\phi \in V_0$, such that $\{\phi(x - k)\}_{k \in \mathbb{Z}}$ is an orthonormal basis in V_0 .

Conditions (4) (5) and (6) imply that $\{\phi_{j,k} := 2^{j/2}\phi(2^j x - k)\}_{k \in \mathbb{Z}}$ is an orthonormal basis for the space V_j (Vidakovic, 1999). We also have $V_0 \subset V_1$

and so the function $\phi(x) \in V_0$ can be represented as a linear combination of functions from V_1 :

$$\phi(x) = \sum_k h_k \sqrt{2} \phi(2x - k), \quad (2.10)$$

for some coefficients $h_k, k \in \mathbb{Z}$. This equation is known as the *scaling equation* and is fundamental in the construction of wavelets. The coefficients $\{h_k\}_{k \in \mathbb{Z}}$ are referred to as the *low-pass filter* associated with ϕ .

The scaling function gives a series of pictures of the signal, each resolution level differing by a factor of two (Hubbard, 1996). For example, the Haar scaling function is given by $\phi(x) = \mathbb{I}_{[0,1]}$, the indicator function taking the value 1 on the interval $[0, 1)$ and zero elsewhere. It is clear that the integer translates form an orthonormal basis since they are indicator functions on disjoint unit intervals. Moving to the next coarsest level, the first two integer translates (i.e. $k = 0, 1$) of the scaling function are given by

$$\begin{aligned} \phi_{-1,0}(x) &= \frac{1}{\sqrt{2}} \phi(x/2) = \frac{1}{2} \mathbb{I}_{[0,2)}, \\ \phi_{-1,1}(x) &= \frac{1}{\sqrt{2}} \phi(x/2 - 1) = \frac{1}{2} \mathbb{I}_{[2,4)}, \end{aligned}$$

thus ‘zooming out’ on the signal. An approximation of a function, f , at resolution level j is given by

$$f_j(x) = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j,k}(x) = P_j f,$$

where P_j is the projection operator of f onto the space V_j and $c_{j,k} = \langle f, \phi_{j,k} \rangle$. This shows that every function in $L^2(\mathbb{R})$ can be approximated by elements of the subspaces V_j . As j increases, the precision of this approximation increases, so that $\lim_{j \rightarrow \infty} P_j f = f$ for all $f \in L^2(\mathbb{R})$ (Daubechies, 1992, Chapter 5). As we now detail, wavelets encode the difference (i.e. the detail lost) when moving from one scale to the next coarsest level.

Derivation of the wavelet function

When a sequence of subspaces satisfies the MRA properties, there exists an orthonormal basis for $L^2(\mathbb{R})$, given by

$$\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \text{ for } j, k \in \mathbb{Z}\},$$

where $\psi_{j,k}(x)$ are the integer translations and dilations of $\psi(x) = \psi_{0,0}(x)$, known as the *wavelet function* or *mother wavelet*. To demonstrate this, consider the detail information that is lost by moving from one resolution space, V_{j+1} , to a coarser space V_j . This is quantified by the detail space W_j , which is the orthogonal complement of V_j in V_{j+1} :

$$V_{j+1} = V_j \oplus W_j, \quad (2.11)$$

This is true for all j , and so successive application of this relationship gives

$$V_{j+1} = V_0 \oplus \bigoplus_{i=0}^j W_i. \quad (2.12)$$

Furthermore, remembering that $\bigcup_{j \in \mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$ we obtain

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j. \quad (2.13)$$

Thus an L^2 -function can be decomposed into mutually exclusive subspaces, W_j , containing information at different scales. Also, the detail subspaces inherit the scaling property from Definition 2.1.3:

$$f(x) \in W_j \iff f(2x) \in W_{j+1}.$$

Therefore if a function, $\psi(x)$, is found such that its integer translations form an orthonormal basis of W_0 , then through dyadic dilations and translations, $\{\psi_{j,k}(x)\}_{k \in \mathbb{Z}}$ is an orthonormal basis for the difference space W_j , and $\{\psi_{j,k}(x)\}_{j,k \in \mathbb{Z}}$

provides an orthonormal basis for $L^2(\mathbb{R})$. We have $\psi(x) \in W_0 \subset V_1$, and so the wavelet function can be expressed as

$$\psi(x) = \sqrt{2} \sum_k g_k \phi(2x - k). \quad (2.14)$$

Here the coefficients $\{g_k\}_{k \in \mathbb{Z}}$ are, due to their action in the frequency domain, called the *high-pass filter*. One possible solution to this equation (Daubechies, 1992, Chapter 5) is to set

$$g_k = (-1)^k h_{1-k}, \quad (2.15)$$

where h_k is the low-pass filter introduced in equation (2.10). Filters linked with this relation are called quadrature mirror filters. Note also, that the filters satisfy the orthogonality relation (Vidakovic, 1999, Section 3.3):

$$\sum_{n \in \mathbb{Z}} h_n h_{n-2k} = \delta_{k,0} \quad \text{and} \quad \sum_{n \in \mathbb{Z}} g_n g_{n-2k} = \delta_{k,0}. \quad (2.16)$$

2.1.4 Discrete wavelet transform

The cascade algorithm for computing the discrete wavelet transform (DWT) was proposed by Mallat et al. (1989). Based on multiresolution analysis, it provides an efficient scheme for performing a discrete, wavelet-based transformation.

Consider the following *refinement relationships*, which can be shown by substitution of indices in the scaling equations (2.10) and (2.14):

$$\phi_{j-1,k}(x) = \sum_{l \in \mathbb{Z}} h_{l-2k} \phi_{j,l}(x) \quad \text{and} \quad \psi_{j-1,k}(x) = \sum_{l \in \mathbb{Z}} g_{l-2k} \phi_{j,l}(x). \quad (2.17)$$

Denoting the coefficients associated with $\phi_{j,k}$ and $\psi_{j,k}$ as $c_{j,k}$ and $d_{j,k}$ respectively, use of the refinement equations leads to the following recursive relation

between the coefficients:

$$\begin{aligned}
 c_{j-1,k} &= \langle f, \phi_{j-1,k} \rangle, \\
 &= \left\langle f, \sum_{l \in \mathbb{Z}} h_{l-2k} \phi_{j,l} \right\rangle \\
 &= \sum_{l \in \mathbb{Z}} h_{l-2k} \langle f, \phi_{j,l} \rangle, \\
 &= \sum_{l \in \mathbb{Z}} h_{l-2k} c_{j,l}.
 \end{aligned} \tag{2.18}$$

And similarly for the detail coefficients we obtain

$$d_{j-1,k} = \sum_{l \in \mathbb{Z}} g_{l-2k} c_{j,l}. \tag{2.19}$$

Hence, denoting $\mathbf{c}_j = \{c_{j,k}\}_{k \in \mathbb{Z}}$ and $\mathbf{d}_j = \{d_{j,k}\}_{k \in \mathbb{Z}}$, if we have the coefficients \mathbf{c}_j and \mathbf{d}_j at a particular resolution level, we can obtain the coefficients at the next level, \mathbf{c}_{j-1} and \mathbf{d}_{j-1} , by application of equations (2.18) and (2.19). Thus it is possible to represent a function in $L^2(\mathbb{R})$ by:

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x). \tag{2.20}$$

The coefficients $c_{j,k}$ and $d_{j,k}$ are often referred to as the *scaling coefficients* and the *detail (or wavelet) coefficients*, respectively. The detail coefficients, $d_{j,k}$, provide local information about the function at scale j , where the scaling coefficients provide information on the ‘smooth’ global behavior.

Given a function, f , observed at 2^J equally spaced time points $n = 1, \dots, 2^J$, if we have the scaling coefficients at the finest scale, \mathbf{c}_J , it is possible to construct the detail and scaling coefficients for all scales $j < J$. Since we only know the function values at the observed locations $n = 1, \dots, 2^J$, it is not possible to calculate the scaling coefficients at the finest scale exactly; we need to approximate them in some way. It is assumed that the finest scaling coefficients

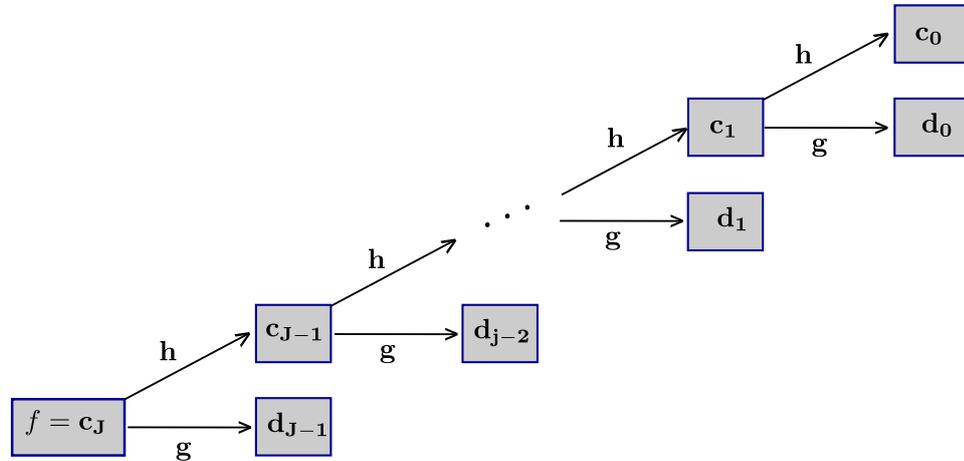


Figure 2.3: Diagram of the decomposition of a function using the DWT through successive applications of the high and low pass filters g and h .

can be approximated by the function values, giving:

$$c_{J,n} = f_n. \quad (2.21)$$

Thus using the low and high pass filters h_k and $g_k = (-1)^k h_{1-k}$ (specified by the wavelet), the expansion coefficients can be calculated recursively for each scale $j < J$, using the decomposition relations (2.18) and (2.19). Each application of the filter g and h produces a new sequence of half the length. Starting with the data, \mathbf{c}_J , of length T , one application of each filter provides the new sequences \mathbf{c}_{J-1} and \mathbf{d}_{J-1} of length $T/2$. The wavelet coefficients \mathbf{d}_{J-1} are stored as these provide the finest level details. The vector \mathbf{c}_{J-1} is then used in the next stage of the algorithm to produce the next level detail and smooth coefficients, and so on until the desired level of decomposition has been met. This gives the transformed data vector:

$$\text{DWT}(\mathbf{f}) = (\mathbf{c}_{j_0}, \mathbf{d}_{j_0}, \mathbf{d}_{j_0+1}, \dots, \mathbf{d}_{J-1}),$$

where the vector \mathbf{c}_{j_0} represents the smooth coefficients at the *primary resolution level*. A full decomposition of a series of length T is conducted so that the primary resolution level $j_0 = 0$ and the vector of smoothing coefficients at this level c_0 is of length 1.

The DWT can also be recursively reversed. The scaling coefficients at each scale can be obtained from the scaling coefficients at the previous coarser scale using the following relationship:

$$c_{j,k} = \sum_{l \in \mathbb{Z}} c_{j-1,l} h_{k-2l} + \sum_{l \in \mathbb{Z}} d_{j-1,l} g_{k-2l}. \quad (2.22)$$

Example decomposition using Haar wavelet

For a clear illustration of the DWT in practice, we provide an example of using the Haar wavelet to decompose a four element vector. For the Haar wavelet the the low and high pass filters are given by

$$h_0 = \frac{1}{\sqrt{2}}, \quad h_1 = \frac{1}{\sqrt{2}} \quad \text{and} \quad h_k = 0 \quad \text{otherwise.}$$

$$g_0 = \frac{1}{\sqrt{2}}, \quad g_1 = \frac{-1}{\sqrt{2}} \quad \text{and} \quad g_k = 0 \quad \text{otherwise.}$$

We start by defining the finest level scaling coefficients as the function values:

$$\mathbf{c}_2 = (c_{2,0}, c_{2,1}, c_{2,2}, c_{2,3}) = (3, 4, 2, 2).$$

Then by applying equations (2.18) and (2.19) we obtain the new sequences \mathbf{c}_1 and \mathbf{d}_1 of length 2:

$$c_{1,0} = \sum_l h_l c_{2,l} = 1/\sqrt{2} \cdot 3 + 1/\sqrt{2} \cdot 4 = 7/\sqrt{2},$$

$$c_{1,1} = \sum_l h_{l-2} c_{2,l} = 1/\sqrt{2} \cdot 2 + 1/\sqrt{2} \cdot 2 = 4/\sqrt{2},$$

$$\begin{aligned}
d_{1,0} &= \sum_l g_l c_{2,l} = 1/\sqrt{2} \cdot 3 - 1/\sqrt{2} \cdot 4 = -1/\sqrt{2}, \\
d_{1,1} &= \sum_l g_{l-2} c_{2,l} = 1/\sqrt{2} \cdot 2 - 1/\sqrt{2} \cdot 2 = 0.
\end{aligned}$$

Giving $\mathbf{c}_1 = (c_{1,0}, c_{1,1}) = (7/\sqrt{2}, 4/\sqrt{2})$ and $\mathbf{d}_1 = (d_{1,0}, d_{1,1}) = (-1/\sqrt{2}, 0)$. Continuing to the next coarsest scale using the same formula, we obtain

$$\begin{aligned}
c_{0,0} &= \sum_l h_l c_{1,l} = 1/\sqrt{2} \cdot 7/\sqrt{2} + 1/\sqrt{2} \cdot 4/\sqrt{2} = 11/2, \\
d_{0,0} &= \sum_l g_{l-2} c_{1,l} = 1/\sqrt{2} \cdot 7/\sqrt{2} - 1/\sqrt{2} \cdot 4/\sqrt{2} = 3/2,
\end{aligned}$$

and so we obtain the transformed data vector $(c_{0,0}, d_{0,0}, d_{1,0}, d_{1,1})$. The original data vector can be recovered from the collection of wavelet coefficients and coarsest level smooth using the inverse DWT defined in equation (2.22).

Boundary issues

Issues occur with the DWT when the support of the wavelet filter used in the decomposition extends beyond the length of the input vector. There are several approaches to dealing with boundary problems. As detailed in Nason and Silverman (1994), options include assuming symmetry so that the function data can be reflected at the endpoints to extend further than the original sampled function vector. The function could also be assumed to be periodic on the range of the data, and the range of the original series extended accordingly. Another method is to pad the observed series with zeros outside of the observation range.

2.1.5 Nondecimated wavelet transform

An undesirable property of the DWT is that it is not translation invariant. This means that starting the transform from a different point in the series could

lead to a different decomposition of the data. Motivated by this, Coifman and Donoho (1995) and Nason and Silverman (1995) discuss the translation invariant or nondecimated wavelet transform which computes wavelet coefficients at all possible scales and locations.

To describe the NDWT, we continue with the toy example from the previous section: $\mathbf{c}_2 = (c_{2,0}, c_{2,1}, c_{2,2}, c_{2,3}) = (3, 4, 2, 2)$. Recall that for the DWT, the level 1 detail and smooth coefficients were computed by comparing the pairings $(c_{2,0}, c_{2,1})$ and $(c_{2,2}, c_{2,3})$. If an extra data point was to be added to the beginning of the series, then we would be considering an entirely different set of pairings. The NDWT avoids this problem by considering *all* pairs of neighbours. So that for this example (and assuming periodic boundary conditions), the level 1 smooth coefficients are given by

$$\begin{aligned} c_{1,0} &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \cdot (c_{2,0}, c_{2,1}) = \frac{7}{\sqrt{2}}, \\ c_{1,1} &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \cdot (c_{2,1}, c_{2,2}) = \frac{6}{\sqrt{2}}, \\ c_{1,2} &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \cdot (c_{2,2}, c_{2,3}) = \frac{4}{\sqrt{2}}, \\ c_{1,3} &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \cdot (c_{2,3}, c_{2,0}) = \frac{5}{\sqrt{2}}. \end{aligned}$$

The first level wavelet coefficients are calculated similarly, but replacing the low pass filter with the high pass filter $(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$ to give $\mathbf{d}_1 = \frac{1}{\sqrt{2}}(-1, 2, 0, -1)$. Unlike the DWT, the new sequences \mathbf{c}_1 and \mathbf{d}_1 both have the same length as the original data. Continuing to the next level, zeros are inserted between each two elements of the wavelet filters. The next level smooth coefficients are

$$\begin{aligned} c_{0,0} &= \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0\right) \cdot (c_{1,0}, c_{1,1}, c_{1,2}, c_{1,3}) = \frac{11}{2}, \\ c_{0,1} &= \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0\right) \cdot (c_{1,1}, c_{1,2}, c_{1,3}, c_{1,0}) = \frac{11}{2}, \\ c_{0,2} &= \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0\right) \cdot (c_{1,2}, c_{1,3}, c_{1,0}, c_{1,1}) = \frac{11}{2}, \\ c_{0,3} &= \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0\right) \cdot (c_{1,3}, c_{1,0}, c_{1,1}, c_{1,2}) = \frac{11}{2}. \end{aligned}$$

The wavelets coefficients are calculated similarly to give $\mathbf{d}_0 = \frac{1}{2}(-3, 1, -3, 1)$.

2.1.6 Extensions to classical wavelet methods

There have been many extensions to the classical wavelet methods introduced in this chapter. In this section we provide only a brief summary of some of the most common extensions. Further methods and more detailed explanations can be found in e.g. Vidakovic (1999) and Nason (2008).

Multiresolution analysis and the construction of wavelet bases can also be generalised to spaces of higher dimensions. Mallat et al. (1989) proposed a multivariate version of the DWT, by considering wavelets $\psi \in L^2(\mathbb{R}^d)$.

Wavelet packets were introduced by Coifman and Wickerhauser (1992). They are collections of linear combinations of wavelet functions, providing a generalisation of standard orthonormal wavelet bases. The collection of their dilations and translations forms a ‘library’ of all packet functions. Since wavelet packets provide greater flexibility than classical wavelet methods they are capable of providing more efficient representations. When using wavelet packets to decompose a signal, both the low and high pass filters are applied at every step of the transform, to the scaling coefficients and also to the detail coefficients. Coifman and Wickerhauser (1992) also introduce a method for searching subsets of coefficients for the ‘best basis’ representation. There are also other examples of basis libraries, for example the smoothed localized complex exponentials (SLEX) library of Ombao et al. (2002).

First introduced by Cohen et al. (1992), biorthogonal wavelet bases utilise different wavelets for the decomposition and reconstruction steps. The orthogonality constraint is relaxed and instead these wavelets satisfy biorthogonality criteria. This approach leads to more flexibility allowing greater symmetry whilst maintaining compact support. The lifting scheme, which is reviewed in Section 2.2, provides a method of creating biorthogonal wavelets.

The unbalanced Haar wavelet basis was introduced by Girardi and Sweldens (1997). Unlike traditional Haar wavelets, jumps in the basis functions do not necessarily occur in the middle of their support and therefore avoid the

restriction of jumps occurring at dyadic locations.

Lina and Mayrand (1995) provide a detailed description of complex-valued Daubechies wavelets. The complex valued wavelet with $N = 3$ was also derived by Lawton (1993). Complex valued wavelet decompositions can also be obtained using the dual tree complex wavelet transform (CWT) introduced by Kingsbury (2001). The CWT produces the real and imaginary coefficients using two different real valued filters which are orthogonal to each other. This method is described in more detail in Chapter 6, when we consider estimating the dependence and phase between two signals using wavelet lifting.

2.1.7 Statistical applications of wavelets

As seen throughout this chapter, wavelets have desirable properties which make their use attractive in many applications. In particular, due to their natural localisation wavelets can provide sparse representations for certain functions that cannot be represented efficiently using Fourier series. Also, unlike in Fourier analysis where there is only one fixed basis, there are several families of discrete wavelets with varying properties to choose from. This means that the basis can be selected to suit the properties of the particular function that is being analysed. In this section we briefly mention the main uses of wavelets in statistics. Further descriptions can be found in Nason (2008), Antoniadis (2007), Abramovich et al. (2000), Antoniadis (1999), and Vidakovic (1999).

This thesis is concerned with the application of wavelets in time series analysis (see Section 2.3 for an introduction to time series). A review of some of the core concepts and applications of wavelets in time series analysis can be found in Nason and von Sachs (1999). In Section 2.3.4 we review the use of wavelets in estimating the time dependent spectra of nonstationary processes using the locally stationary wavelet model of Nason et al. (2000). This method, and the associated methods introduced in later chapters, provide techniques for visualising the features of complicated time series and aiding interpretation.

Other applications of wavelets in time series analysis include change point detection (Davis et al. (2006) and Cho and Fryzlewicz (2010)), estimation of self-similarity parameters of long-memory time series (Wang et al., 2001), forecasting (Fryzlewicz et al., 2003) and classification (Fryzlewicz and Ombao, 2009).

Wavelets are often used for nonparametric regression. The basic idea is to denoise the signal by modifying the noisy wavelet coefficients according to some rule. The inverse DWT is performed on the modified coefficients, providing an estimate of the noise free signal. Donoho and Johnstone (1994) proposed the non-linear estimation technique of *thresholding*. There have been many subsequent extensions to the the original wavelet thresholding methodology, for example Barber and Nason (2004) introduce various thresholding techniques using complex wavelets, Coifman and Donoho (1995) introduce *translation invariant denoising* based on the NDWT and Nason (1996) proposes using cross validation to select the threshold.

The estimation of probability density functions using wavelets is another popular application, see e.g. Hall and Patil (1995), Donoho et al. (1996), Antoniadis et al. (1999). Wavelet estimators provide accurate estimates for densities with sharp changes or discontinuities. However unlike kernel density function estimates, wavelet estimators do not guarantee non-negative density estimates. A modification to wavelet methods, *Slepian semi-wavelets*, were introduced by Walter and Shen (2005) to overcome the problem of negative estimates.

2.2 The wavelet lifting scheme

As described in the previous section, wavelets provide a useful tool for function decomposition and, through the use of the DWT (or NDWT), a sampled signal can be represented efficiently as a set of wavelet coefficients. Although classical wavelet techniques have proved useful in many settings, there are also

limitations of these methods. In particular, the classical DWT and NDWT rely on regular spacing of observations.

Techniques for dealing with irregular data often involve pre-processing the data to a regular grid. One disadvantage of pre-processing methods is that they typically result in a loss of information at high frequencies, an effect that we would like to avoid. This, and further solutions, will be discussed in more detail in Chapter 7 in the context of periodogram estimation. The solution to this problem that we present arises from a new way for building wavelets known as the *lifting scheme*. The lifting scheme was introduced by Sweldens (1996) to handle more general settings than the standard wavelet transformations and provides a natural approach to extending classical wavelet techniques to irregular grids. Wavelet functions obtained through the lifting scheme are known as *second generation wavelets*. Unlike the *first generation wavelets* introduced in Section 2.1.2, second generation wavelets are not necessarily translates and dilates of the same function.

In what follows, we review wavelet lifting from a computational perspective. A more rigorous exposition, including a description of multiresolution analysis from the second generation viewpoint, can be found in the original text of Sweldens (1996). We start by introducing the general lifting approach, before describing one coefficient at a time wavelet lifting, a particular type of lifting that can be used for problems with irregular sampling.

2.2.1 Computational approach to wavelet lifting

We begin by introducing the general specification of the wavelet lifting scheme. Assume we have observations f_1, \dots, f_n of a function f , observed at time points, x_1, \dots, x_n . For now we assume that the observation times are equally spaced. At the first stage, n , of the decomposition, define $c_{n,k} = f_k$ so that the finest level scaling coefficients interpolate the data. A single stage of the lifting transform is composed of the following 3 steps:

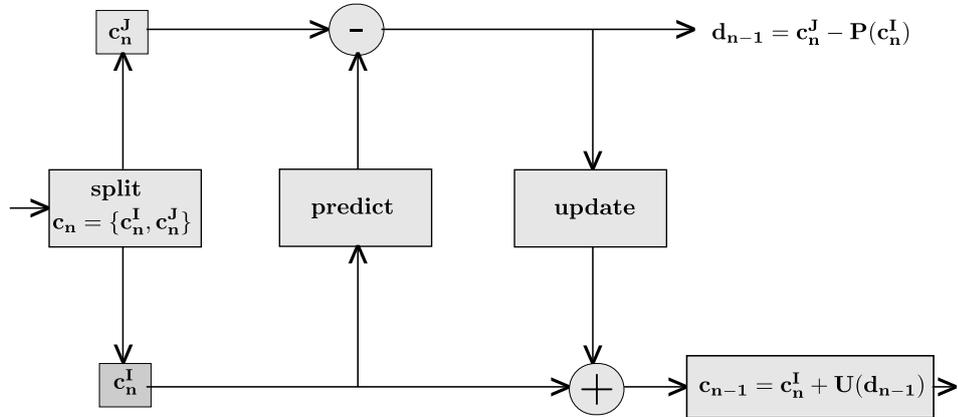


Figure 2.4: Schematic of stage n of the general lifting transform.

- **Split:** The data is divided into two disjoint subsets, with I and J denoting the indexing sets of the two groups so that $I \cup J = \{1, \dots, n\}$. The definition of subsets depends on the application but one example is to take I and J to be the sets of odd and even indices. We then denote \mathbf{c}_n^I the vector $(c_{n,i}, i \in I)$ and \mathbf{c}_n^J the vector $(c_{n,j}, j \in J)$.
- **Predict:** \mathbf{c}_n^I is used to give a prediction of \mathbf{c}_n^J , denoted $\tilde{\mathbf{c}}_n^J = P(\mathbf{c}_n^I)$. Here P is the prediction function, specific to the particular lifting scheme. The difference between the observed and predicted values provides the detail coefficients of this step:

$$\mathbf{d}_{n-1} = \mathbf{c}_n^J - \tilde{\mathbf{c}}_n^J.$$

- **Update:** \mathbf{c}_n^I is then updated according to the calculated detail coefficients, and the update function U specified by the lifting scheme. This provides the next level smooth:

$$\mathbf{c}_{n-1} = \mathbf{c}_n^I + U(\mathbf{d}_{n-1}).$$

The update step serves to preserve some quantity from the initial signal, such as the mean value. The procedure is repeated until the signal has been decomposed to the desired level, giving the vector $(\mathbf{c}_{j_0}, \mathbf{d}_{j_0}, \mathbf{d}_{j_0+1}, \dots, \mathbf{d}_{n-1})$, where j_0 is the primary resolution level of the decomposition. The finest scale wavelet coefficients are denoted \mathbf{d}_{n-1} , the next finest scale are denoted \mathbf{d}_{n-2} and so forth until the coarsest level detail coefficients, which are denoted \mathbf{d}_{j_0} .

The lifting algorithm can also be easily reversed, obtaining the vector \mathbf{c}_n from the coarser level smooth \mathbf{c}_{n-1} and the associated detail coefficients, \mathbf{d}_{n-1} , as follows:

- **Undo update:**

$$\mathbf{c}_n^I = \mathbf{c}_{n-1} - U(\mathbf{d}_{n-1}).$$

- **Undo predict:**

$$\mathbf{c}_n^J = \mathbf{d}_{n-1} + \tilde{\mathbf{c}}_n^J.$$

- **Merge:**

$$\mathbf{c}_n = \text{merge}(\mathbf{c}_n^I, \mathbf{c}_n^J),$$

The inverse transform uses the same operators, P and U , as with the forward transform, but with a reversal of sign at each step.

Haar transform using wavelet lifting

To give a specific example of the lifting transform in practice, we consider the Haar transform in a wavelet lifting setting. Using the same example as in Section 2.1.4, we start with the data vector $\{f_k\}_{k=1,\dots,4} = (3, 4, 2, 2)$ and set $n = 2$, defining the level which contains the data. The finest level smooth is assumed to interpolate the data, giving

$$\mathbf{c}_2 = (c_{2,1}, c_{2,2}, c_{2,3}, c_{2,4}) = (3, 4, 2, 2).$$

The data is split into two sets, I and J corresponding to odd and even indices.

$$\mathbf{c}_2^I = (3, 2) \quad \mathbf{c}_2^J = (4, 2)$$

In the case of the Haar transform, the prediction is carried out by estimating each odd indexed element with the next even indexed element. So we have $\tilde{c}_{2,1}^I = c_{2,1}^J = 4$ and $\tilde{c}_{2,2}^I = c_{2,2}^J = 2$. The details are then constructed by taking the difference between the observed and predicted values:

$$\begin{aligned} d_{1,1} &= c_{2,1}^I - \tilde{c}_{2,1}^I = 3 - 4 = -1, \\ d_{1,2} &= c_{2,2}^I - \tilde{c}_{2,2}^I = 2 - 2 = 0. \end{aligned}$$

The set of even indexed coefficients, \mathbf{c}_2^J , are then updated to form the next level smooth:

$$\begin{aligned} c_{1,1} &= c_{2,1}^J + \frac{1}{2}d_{1,1} = \frac{1}{2}(c_{2,1}^J + c_{2,1}^I) = \frac{7}{2}, \\ c_{1,2} &= c_{2,2}^J + \frac{1}{2}d_{1,2} = \frac{1}{2}(c_{2,2}^J + c_{2,2}^I) = 2. \end{aligned}$$

So we have $\mathbf{c}_1 = (7/2, 2)$ and $\mathbf{d}_1 = (-1, 0)$. Repeating the steps at the next level we obtain

$$\begin{aligned} d_{0,1} &= c_{1,1}^I - \tilde{c}_{1,1}^I = 7/2 - 2 = 3/2, \\ c_{0,1} &= c_{1,1}^J + \frac{1}{2}d_{0,1} = \frac{1}{2}(c_{1,1}^J + c_{1,1}^I) = 11/4. \end{aligned}$$

Comparing this to the example presented in Section 2.1.4, we see that the obtained detail coefficients are rescaled versions of that obtained by the DWT cascade algorithm. In this example we have used the split, predict and update steps which correspond to the Haar wavelet but, as shown in Daubechies and Sweldens (1998), all classical filter banks can be decomposed into a sequence of lifting steps.

As in the first generation wavelet setting, boundary considerations are necessary. Particularly if a high order prediction scheme has been used, there may

be insufficient data points when lifting near the end points of the signal. At these points, predictions can be made by extrapolating the signal using copies of the neighbouring data points.

Many variations on the general wavelet lifting split-predict-update procedure are possible. In the presented example with the Haar wavelet, the data was split into groups of odd and even indices. Other criteria for splitting the data are also possible, for example Jansen et al. (2001) introduce a method that involves selecting just one coefficient at each step. Nunes et al. (2006) propose an adaptive lifting transform which, based on the one coefficient at a time lifting scheme, allows for different method of prediction at each stage.

2.2.2 One coefficient at a time wavelet lifting

The one coefficient at a time procedure (OCAAT), first introduced by Jansen et al. (2001), is a variation on standard wavelet lifting which is suitable for application to irregularly spaced data. With this type of lifting, the split, predict and update procedure is followed as with conventional wavelet lifting, but with just one data point being selected in the ‘split’ step. In order to account for irregular sampling, the concept of the integral associated with the scaling function of each point is introduced. Following Jansen et al. (2001), we now summarise the one coefficient at a time lifting procedure.

Initial definition of points

We start with observations f_1, \dots, f_n of a function f observed at time points x_1, \dots, x_n where the time points are permitted to be *irregularly sampled*. As previously, the scaling functions at the finest level are defined as the characteristic functions of the intervals associated with each point. This gives $\phi_{n,k}(x_i) = \delta_{i,k}$, setting $c_{n,k} = f_k$.

The sampling of the data is described by means of the integral of the scaling function of each point. Intervals associated with each gridpoint can

be constructed by using the midpoints between successive observations as the interval endpoints. The scaling function is assumed to be zero outside the support of the interval. The integral of point x_k at stage n in the algorithm is denoted

$$I_{n,k} = \int_{\mathbb{R}} \phi_{n,k}(x) dx.$$

This is a definite integral with respect to some suitable measure. If the series is sampled at regular unit intervals then at the first stage of the algorithm, before any points have been removed, all points will have an integral of 1. It is also necessary to consider the definition of the intervals associated with the first and last observations. Following Nunes et al. (2006), the end intervals are reflected, so that the first and last data points are in fact the midpoints of their associated intervals.

The algorithm

At the **first stage** (n) of the lifting algorithm, the set of indices of scaling coefficients S_n , and the set of indices of wavelet coefficients, D_n , are given by

$$\begin{aligned} S_n &= \{1, \dots, n\}, \\ D_n &= \emptyset. \end{aligned}$$

At the **next stage** ($n - 1$) we choose the index of the point to be lifted and denote this index j_n . This point is removed from the set of scaling coefficients and converted into a detail coefficient so that the new set of indices of scaling coefficients and wavelet coefficients is

$$\begin{aligned} S_{n-1} &= \{1, \dots, n\} \setminus \{j_n\}, \\ D_{n-1} &= \{j_n\}. \end{aligned}$$

The point to be lifted, (x_{j_n}, c_{n,j_n}) , is often chosen according to the size of integral of the scaling function. Removing the point with the smallest integral,

$I_{n,k} = \min_k \int_{\mathbb{R}} \phi_{n,k} dx$, means that we are removing the point with the finest detail, since the integral accounts for the length of the sampling interval between points. Regions where the function is sampled more densely will correspond to smaller integral values.

The set of neighbours, J_n , of the point j_n are also identified. Note that the set of neighbours is indexed by n as the choice will depend on the removal stage (via the points remaining at that stage).

Once the point to be removed has been identified, it is removed from the set of scaling coefficients and converted into a detail coefficient using the prediction and update steps, familiar from standard wavelet lifting:

Prediction step: The detail coefficient is found by taking the difference between the actual value and that which is predicted by its neighbours

$$d_{j_n} = c_{n,j_n} - \sum_{i \in J_n} a_i^n c_{n,i}. \quad (2.23)$$

The choice of prediction coefficients, a_i^n is dependent on the chosen prediction method. For a function that is constant over the region of the neighbours, the wavelet coefficient should be zero and so the weights should satisfy $\sum_j a_j^n = 1$.

Update step: The scaling coefficients of the neighbouring points are then updated accordingly:

$$c_{n-1,i} = c_{n,i} + b_i^n d_{j_n} \quad \text{for } i \in J_n, \quad (2.24)$$

where b_i^n are the update weights which will be discussed in more detail later in the section. For points which are not neighbours, we have:

$$c_{n-1,i} = c_{n,i} \quad \text{for } i \notin J_n.$$

The integrals of the scaling functions also need to be updated. This is because

there is a smaller number of scaling functions to account for the same interval. Only the integrals of the set of neighbours are affected so the updating is conducted following:

$$I_{n-1,i} = I_{n,i} + a_i^n I_{n,j_n} \quad \text{for } i \in I_n. \quad (2.25)$$

Calculation of updating weights

In the update stage we want to preserve the overall mean of the data, and so it is necessary to ensure

$$\sum_{k \in \{1, \dots, n\}} c_{n,k} I_{n,k} = \sum_k c_{n-1,k} I_{n-1,k}.$$

Since the only integrals to change in the updating step are those of the neighbours, this can be re-written as

$$c_{n,j_n} I_{n,j_n} + \sum_{i \in J_n} c_{n,i} I_{n,i} = \sum_{i \in J_n} c_{n-1,i} I_{n-1,i}.$$

Using the relationships provided by the updating step and rearranging gives

$$\sum_{i \in J_n} b_i^n I_{n-1,i} = I_{n,j_n}.$$

The solution to this equation proposed by Jansen et al. (2004) is:

$$b_i^n = \frac{I_{n,j_n} I_{n-1,i}}{\sum_{i \in J_n} I_{n-1,i}^2}. \quad (2.26)$$

At each step of the algorithm the scaling and wavelet functions are recursively constructed from the scaling functions at the coarser level:

$$\begin{aligned}\phi_{r-1,i}(x) &= \phi_{r,i}(x) + a_i^r \phi_{r,j_r}(x), & \text{for } i \in J_n, \\ \phi_{r-1,i}(x) &= \phi_{r,i}(x), & \text{for } i \notin J_n, \\ \psi_{j_r}(x) &= \phi_{r,j_r}(x) - \sum_{i \in J_n} b_i^r \phi_{r-1,i}(x).\end{aligned}$$

Their values depend on the prediction and updating weights as well as the sampling of the data (through the initial definition of the integrals) and the removal order of the points. This can be contrasted with the wavelets used in the Classical DWT where, for a given choice of wavelet family, the wavelet vector ψ_j is completely defined by the scale j . With the OCAAT lifting scheme, the wavelet vector used at each scale is indexed according to the removal point. The concept of scale is different in this context and will be described later in this section.

Example

To illustrate the OCAAT lifting scheme we provide an example of the removal of one coefficient using linear prediction with one neighbour either side. Starting with the data vector $\{f_i\}_{i=1,\dots,6} = (4, 2, 2, 6, 4, 4)$, sampled at irregularly spaced time points $\{x_i\}_{i=1,\dots,6} = (1, 2, 4, 5, 8, 9)$, we set the finest level smooth to be the data:

$$\mathbf{c}_n = (4, 2, 2, 6, 4, 4).$$

and we have $S_n = \{1, 2, 3, 4, 5, 6\}$ denoting the indices of the scaling coefficients and $D_n = \emptyset$ denoting the indices of the detail coefficients. The integrals are given by $\{I_{n,i}\}_{i \in S_n} = (1, 1.5, 1.5, 2, 2, 1)$.

Split: The point $j_n = 4$ is selected to be removed. This observation will be converted to a detail coefficient and so we have $S_{n-1} = \{1, 2, 3, 5, 6\}$, $D_{n-1} = \{4\}$.

Predict: As defined by the prediction scheme, the set of neighbours is given by $J_n = (3, 5)$. The detail coefficient is calculated following

$$d_4 = c_{n,4} - \sum_{i=3,5} a_i^n c_{n,i},$$

where $a_3^n = 0.75$ and $a_5^n = 0.25$, giving a detail coefficient of $d_4 = 3.5$.

Update: The integrals of the neighbours are updated according to equation (2.25) giving $\{I_{n-1,i}\}_{i \in S_{n-1}} = (1, 1.5, 3, 2.5, 1)$. Notice that the overall sum of the integrals remains the same. The updating weights can then be calculated using equation (2.26) to give $b_3^n = \frac{2 \times 3}{3^2 + 2.5^2} = 0.39$, $b_5^n = \frac{2 \times 2.5}{3^2 + 2.5^2} = 0.33$. These weights are then used to update the data vector according to equation (2.24), giving $\mathbf{c}_{n-1} = (4, 2, 3.4, 5.4, 4)$. The prediction and update steps of this example are illustrated in Figure 2.5.

The steps of the one coefficient at a time lifting scheme are repeated until the specified number of points have been removed. After points j_n, j_{n-1}, \dots, j_r have been removed, the initial function can be written represented as

$$f(x) = \sum_{i \in S_{r-1}} c_{r-1,i} \phi_{r-1,i}(x) + \sum_{k \in \{n, n-1, \dots, r\}} d_{j_k} \psi_{j_k}(x).$$

Definition of scale

From Jansen et al. (2004) we have that a useful measure of scale for the wavelet ψ_{j_r} is the integral of the scaling function for site j_r at the last stage before the observation is removed from future consideration. This scale is denoted α_{j_r} and is defined as

$$\alpha_{j_r} = I_{r-1, j_{r-1}}. \tag{2.27}$$

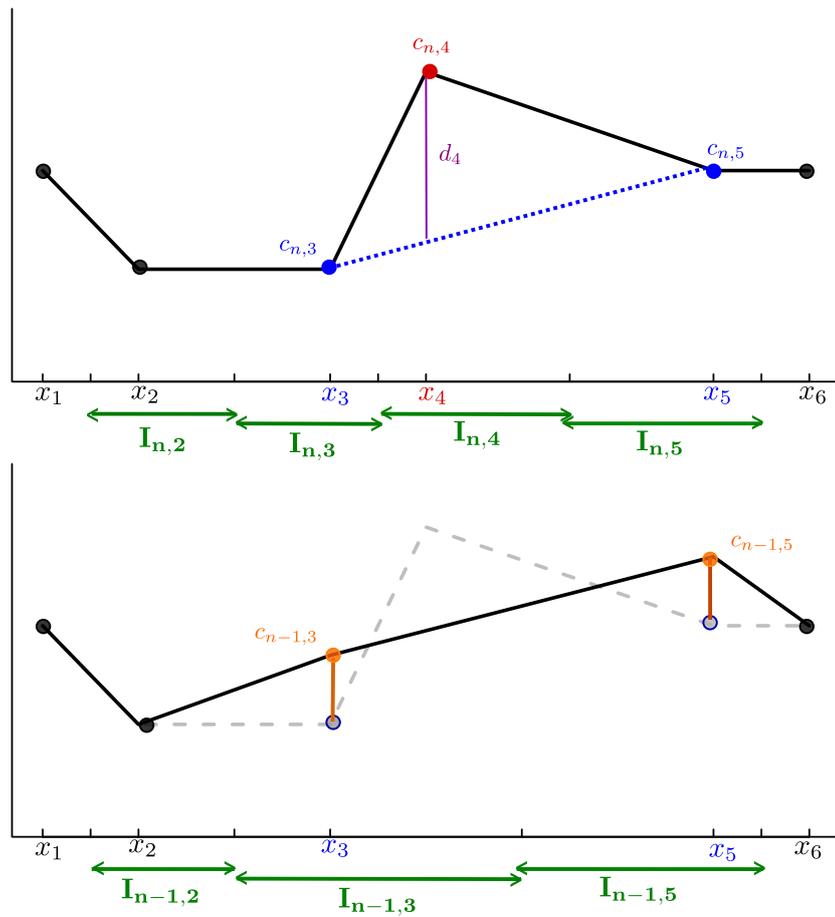


Figure 2.5: Stage n of the one coefficient at a time lifting scheme using linear prediction with one neighbour either side of the removal point. The first figure shows the prediction of point $j_n = 4$ and the second figure shows the process of updating the neighbours and integrals.

As seen later in Chapters 5 and 6, in practice it is common to work with $\log_2(\alpha_{j_r})$. Unlike the DWT, in the one coefficient at a time lifting setting scale is now a continuous measure. Assuming that the prediction weights, a_j are strictly positive, the scales α_{j_r} are a monotonic function of the index r and the order in which the lifting scheme determines the coefficients (Jansen et al., 2004). However, as noted by Nunes et al. (2006), for certain choices of prediction scheme it is possible that the prediction weights can take negative values. This implies that when updating the integrals according to equation (2.25) the integrals will become smaller rather than larger.

2.3 Time series analysis

A discrete time series is a family of random variables, $\{X_t, t \in T\}$, where the index set T is often a set of time points. An observed realisation of the process, x_t , is also referred to as a time series. We are predominantly interested in discrete time series; however it is also possible to obtain continuous time series, obtained when observations are recorded continuously over some time period. For continuous time series we use the notation $x(t)$. A key characteristic of time series data is that successive observations are (generally) not independent. The observations taken over time may have an internal structure, such as autocorrelation or season variation. Examples of time series arise from diverse and numerous fields such as economics (e.g. share prices on successive days) and meteorology (e.g. measurements of rainfall, temperature). In particular, we shall be considering the application of time series methods to data from an experiment in neuroscience, and palaeoclimatic data extracted from ice-cores. Interesting reference books on time series include that of Priestley (1981), Chatfield (2003), Hamilton (1994) and Brockwell and Davis (2006).

In order to make inferences on the characteristics of a time series, it is necessary to impose assumptions on its evolution. In some situations it is possible to assume that the time series is stationary. Intuitively, a time series

is said to be stationary if there is no systematic change in characteristics of the time series over time. More formally we can define a time series in terms of *weak stationarity* or *strict stationarity*. A time series is said to be strictly stationary if the joint distribution of observations is not a function of time, i.e. the joint distributions of $(X_{t_1}, \dots, X_{t_k})'$ and $(X_{t_1+h}, \dots, X_{t_k+h})'$ are the same for all positive integers k and for all $t_1, \dots, t_k, h \in \mathbb{Z}$. Often this assumption is relaxed to that of weak (or covariance) stationarity. For a zero mean time series X_t , its auto-covariance function is given by $c_X(t, t + \tau) = \mathbb{E}(X_t X_{t+\tau})$. X_t is said to be weakly stationary if $E|X_t|^2 < \infty$ and $c_X(t, t + \tau)$ for $t, \tau \in \mathbb{Z}$, is dependent only on the time lag, τ , and not the value of t . In this case, the time independent autocovariance is denoted $c_X(\tau)$.

Any weakly stationary, discrete time series $\{X_t\}$ with mean zero can be decomposed using the Cramér representation:

$$X_t = \int_{-\pi}^{\pi} A(\omega) \exp(i\omega t) dZ(\omega), \quad (2.28)$$

where $A(\omega)$ is the amplitude and $\{Z(\omega)\}$ is a random process with orthonormal increments i.e. $dZ(\omega) = Z(\omega) - Z(\omega-)$, $\mathbb{E}(dZ(\omega) \overline{dZ(\omega')}) = \delta_{\omega, \omega'}$ (see e.g. Brillinger (1975, Section 4.6.)). The quantity $f(\omega) = |A(\omega)|^2$ is called the spectrum, or spectral density function and quantifies the contribution to variance given by frequency ω . X_t can be thought of as a linear combination of Fourier sinusoids oscillating at different frequencies.

Different realisations of the process involve the same deterministic function $A(\omega)$, but different realisations of the random process $\{Z(\omega)\}$. The Cramér representation can be compared to the representation in equation (2.1) for deterministic series. The major difference between the two is that in equation (2.28), for each frequency, ω , $dZ(\omega)$ is a random quantity and the integral is a stochastic integral. The representation is therefore understood in a mean square sense.

The autocovariance function of X_t has a Fourier representation in terms of

the spectrum, given by

$$c_X(\tau) = \int_{-\pi}^{\pi} f(\omega) \exp(i\omega\tau) d\omega. \quad (2.29)$$

This relationship is invertible, with the Fourier transform of the covariance function giving the spectrum.

2.3.1 Bivariate methods

Often we are interested not in the individual characteristics of a single time series but in the dependence between two time series. Given two stationary, zero mean, time series, $\{X_t^{(1)}\}$ and $\{X_t^{(2)}\}$, their cross-covariance is given by $c_X^{(1,2)}(\tau) = \mathbb{E}(X_t^{(1)} X_{t+\tau}^{(2)})$. The cross covariance function has a Fourier representation in terms of the cross-spectral density function, $f^{(1,2)}(\omega)$:

$$c_X^{(1,2)}(\tau) = \int_{-\pi}^{\pi} f^{(1,2)}(\omega) \exp(i\omega\tau) d\omega. \quad (2.30)$$

The cross-covariance function provides a natural estimate of the relationship between the two series in the time domain, while the cross-spectral density function can be used similarly in the spectral domain (Brillinger, 1975). Similarly to the univariate case, the relationship between the cross-covariance and cross-spectrum is invertible, with the cross spectrum defined as the Fourier transform of the cross-covariance.

The cross correlation function is derived by normalising the cross-covariance by the individual variances of the two processes. The cross-correlation function lies in the range $[-1, +1]$ with a value of zero indicating a complete lack of correlation between the two series at time delay τ . The cross-correlation typically depends on the parameters of the processes $X^{(1)}$ and $X^{(2)}$.

Similarly, the coherence function is derived by normalising the cross-spectrum by the individual spectra and, roughly speaking, measures the correlation be-

tween the signals as a function of frequency.

$$K(\omega) = \frac{|f^{(1,2)}(\omega)|^2}{\sqrt{f^{(1)}(\omega)f^{(2)}(\omega)}}. \quad (2.31)$$

In practice, the *squared coherence function*, $K(\omega)^2$ is often used. By the Cauchy-Schwarz inequality, the function satisfies $0 \leq K(\omega)^2 \leq 1$ for $-\pi \leq \omega \leq \pi$ (Brockwell and Davis, 2006, Section 11.6). A coherence of $K(\omega) = 0$ indicates that the signals are uncorrelated at frequency ω and a value near 1 indicates a strong linear relationship. Unlike the cross-correlation, the coherence does not depend on the parameters of the serial autocorrelation within each process. This can be demonstrated by re-writing equation (2.31) as $K_{12}(\omega) = \mathbb{E}(dZ_1(\omega)\overline{dZ_2(\omega)})$ where $Z(\omega)$ is the orthonormal increment process in equation (2.28) (see Brockwell and Davis (2006, Section 11.6)). It effectively measures the correlation between the orthonormal increments corresponding to the frequency components of $X^{(1)}$ and $X^{(2)}$, and thus provides different information about the dependence between the processes than that provided by the cross-correlation function.

As well as the coherency, the cross spectrum can be used to derive several other quantities which are useful in interpreting the cross spectrum. The cross-spectrum is a complex quantity and so can be written in the form

$$f^{(1,2)}(\omega) = c(\omega) - iq(\omega), \quad (2.32)$$

where the real-valued term, $c(\omega)$, is called the *co-spectrum* and the imaginary term, $q(\omega)$, is called the *quadrature spectrum*. Alternatively, the cross spectrum can be written in the form

$$f^{(1,2)}(\omega) = \alpha^{(1,2)}(\omega)e^{i\phi^{(1,2)}(\omega)}, \quad (2.33)$$

where $\alpha^{(1,2)}$ and $\phi^{(1,2)}(\omega)$ are the *cross-amplitude spectrum* and the *phase spec-*

trum, given by:

$$\alpha^{(1,2)} = \sqrt{c(\omega)^2 + q(\omega)^2}, \quad \phi^{(1,2)}(\omega) = \tan^{-1}\left(\frac{-q(\omega)}{c(\omega)}\right). \quad (2.34)$$

2.3.2 Spectral analysis

As highlighted by the previous sections, the spectral density can provide a useful description of the properties of a time series. In later chapters we will be interested in estimating the spectrum and cross-spectrum of nonstationary processes using wavelets. It is therefore useful to have an understanding of the corresponding Fourier methods for the spectral estimation of stationary series. The spectrum is commonly estimated by first estimating a quantity known as the *periodogram*, which can be described as the ‘sample spectral density’ of the series (Shumway and Stoffer, 2000). The periodogram is based on the discrete Fourier transform of the series which we now introduce.

Definition 2.3.1. Given a discrete, regularly spaced time series, $\{X_n\}_{n=1}^N$, following Shumway and Stoffer (2000, Chapter 4), the discrete Fourier transform (DFT) is given by

$$\xi(\omega_j) = \frac{1}{\sqrt{N}} \sum_{n=1}^N X_n e^{-it\omega_j}, \quad \text{for } j = 0, \dots, N/2.$$

where $\omega_j = \frac{2\pi j}{N}$ are the *Fourier frequencies* of the series.

Definition 2.3.2. Given a discrete, regularly spaced time series, $\{X_n\}_{n=1}^N$, the periodogram is given by

$$I_N(\omega_j) = |\xi(\omega_j)|^2, \quad \text{for } j = 0, \dots, N/2,$$

where $\omega_j = \frac{2\pi j}{N}$ and $\xi(\omega_j)$ is the DFT of $\{X_n\}$ at ω_j .

Note that for a series sampled at unit time intervals, the highest frequency that can be detected in the data (termed the *Nyquist frequency*) is at frequency

$\omega = \pi$. The lowest frequency that can be detected corresponds to completing one full cycle in the time period. This lowest frequency is given by $\omega = \frac{2\pi}{N}$ (Chatfield, 2003, Chapter 6.). For series not sampled at unit time intervals, but with constant sampling interval Δt , the corresponding frequencies can be found by replacing π in the previous equations by $\frac{\pi}{\Delta t}$. Also note that, as illustrated by the neuroscience experiment detailed in Section 4.2.2, it is common to work in Hertz frequency f rather than with the angular frequency ω . Since $\omega = 2\pi f$, the lowest and highest frequencies in hertz (for unit sampling) correspond to $f_1 = \frac{1}{N}$ and $f_{N/2} = \frac{1}{2}$.

In the case of bivariate observations, the cross-spectrum can be estimated from the *cross-periodogram*.

Definition 2.3.3. Given two discrete, regularly spaced, time series, $\{X_n^{(1)}\}_{n=1}^N$, $\{X_n^{(2)}\}_{n=1}^N$ the cross-periodogram is given by

$$I_N^{(1,2)}(\omega_j) = \xi^{(1)}(\omega_j) \overline{\xi^{(2)}(\omega_j)}, \quad \text{for } j = 0, \dots, N/2 ,$$

where $\xi^{(1)}(\omega_j), \xi^{(2)}(\omega_j)$ are the DFTs of $X^{(1)}$ and $X^{(2)}$.

Though the periodogram and cross-periodogram are unbiased estimators of the spectrum and cross-spectrum, the variance of the estimators does not decrease as we increase the length of the time series. In order to provide consistency, the estimators need to be smoothed in some way. Reviews of linear smoothing techniques can be found in Brillinger (1975) or Shumway and Stoffer (2000). In cases where the spectral density exhibits a large degree of irregularity, these linear smoothing techniques are incapable of achieving the optimal mean-square rate of convergence (Fryzlewicz et al., 2008). Nonlinear smoothing techniques include Davies and Kovac (2004) and Fryzlewicz et al. (2008).

2.3.3 Allowing for nonstationarity

The assumption of stationarity is extremely restrictive and it is intuitive that many observed time series will not be stationary. For example, we later consider a problem from neuroscience in which recordings were made from two areas of a rat's brain. Since the rat's behaviour, sensory environment and the excitability and rhythmicity of its neuronal networks can vary over sub-second timescales, it is likely that the resulting time series will also be non-stationary.

An extension to Fourier analysis, to allow for non-stationarity, is the windowed, Gabor transform (Gabor, 1946) which estimates the spectrum over time as well as frequency by applying the Fourier transform to a localised time window that slides along the time axis; the data within each time window is presumed stationary. The Gabor transform is given by

$$\hat{G}(b, \omega) = \int_{\mathbb{R}} x(t) \exp(-i\omega t) g(t - b) dt, \quad (2.35)$$

where $g(t)$ is the localising function. This approach is also commonly known as the short time Fourier transform (STFT). One drawback to this approach is that the window width is constant over all frequencies. Unless the window width is chosen to be sufficiently small, the time localisation at high frequencies is poor and it is possible to miss fine scale features. It is desirable for the time window to change so that we use a larger window to observe more information for lower frequencies, and a smaller window for more precise time resolution at high frequencies.

Stemming from the Cramér representation in equation (2.28), there have been several modifications that can be made to allow for non-stationarity. Introduced by Priestley (1965), one possibility is to replace the amplitude function, $A(\omega)$, with a time varying version, $A_t(\omega)$, leading to a class of non-stationary processes termed oscillatory processes. In oscillatory form, the Fourier transform of a process is concentrated about frequency ω_0 and the process then behaves like a sine wave with conventional frequency ω_0 , modu-

lated by a smoothly varying amplitude function. Here smoothly varying means that the Fourier transform is concentrated around zero (Priestley, 1988, Section 6.3). This also leads to the definition of a time dependent evolutionary spectrum, which describes the frequency content of the process over time.

Continuing with the concept of a time varying amplitude, Dahlhaus (1997) introduces the class of locally stationary Fourier (LSF) processes. The process X_t is modelled as a triangular stochastic array $\{X_{t,T}\}_{t=0}^{T-1}$ such that

$$X_t = \int_{-\pi}^{\pi} A_{t,T}^0(\omega) \exp(i\omega t) dZ(\omega), \quad (2.36)$$

and there exists a transfer function $A : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}$ which is continuous in the first argument and 2π periodic in the second, such that

$$\sup_{t,\omega} |A_{t,T}^0(\omega) - A(t/T, \omega)| \leq \frac{K}{T}. \quad (2.37)$$

The consideration of asymptotics is difficult in a nonstationary setting since future observations may not contain any information on the structure of the process at the present time. This problem is overcome in the LSF process formulation of Dahlhaus (1997), due to the introduction of the important concept of *rescaled time*, $z = t/T \in (0, 1)$. By modelling the amplitudes as a function of rescaled time, it means that as the length of the time series increases there is more information about the local structure of $A(u, \omega)$, permitting asymptotic estimation of the model structure.

Dahlhaus and Polonik (2002) introduce time varying empirical spectral processes for locally stationary processes (see also Dahlhaus and Polonik (2006) and Dahlhaus and Polonik (2009)). Van Bellegem and Dahlhaus (2006) discuss fitting autoregressive models with time-varying parameters to locally stationary processes.

Ombao et al. (2002) propose a representation for non-stationary processes

in terms of SLEX (Smooth Localized Complex Exponential) functions which can be thought of as localised versions of the Fourier exponential functions. Using the Auto-SLEX method (Ombao et al., 2001), the data is automatically segmented into approximately stationary dyadic blocks, allowing estimation of the time-varying spectra.

Nason et al (2000) introduce the class of locally stationary wavelet (LSW) processes which also adopts the rescaled time principle but replaces the Fourier exponentials with non-decimated wavelets. This forms the basis of our work in Chapters 3 and 4 and will be detailed later.

Finally, we briefly mention empirical mode decomposition (EMD), an adaptive decomposition method suitable for non-stationary time series (Huang et al., 1998). EMD is based on the local extrema of the signal and imposes no assumptions on stationarity. The method differs from the other approaches described in this section in that the basis functions are dependent on the data. EMD is described in further detail in Section 5.6, where it is introduced as an interesting comparison to wavelet lifting techniques.

2.3.4 The locally stationary wavelet model

The bivariate LSW process considered in this thesis is based on the univariate case of Nason et al. (2000). In this section we recall the definition of a LSW process as well as several related quantities. Firstly we describe non-decimated wavelets, the building blocks of the LSW model.

Discrete wavelets and associated functions

Let $\{g_k\}$ and $\{h_k\}$ denote the high and low pass filters, satisfying the quadrature mirror relation of equation (2.15). Following Nason et al. (2000), the discrete wavelet vectors $\psi_j = (\psi_{j,0}, \dots, \psi_{j,(L_j-1)})$, for scale $j < 0$, are obtained

using the formulae

$$\begin{aligned}\psi_{-1,n} &= g_n, & \text{for } n = 0, \dots, L_{-1} - 1, \\ \psi_{j-1,n} &= \sum_k h_{n-2k} \psi_{jk}, & \text{for } n = 0, \dots, L_{j-1} - 1,\end{aligned}$$

where $L_j = (2^{-j} - 1)(N_h - 1) + 1$ is the length of support of the wavelet and N_h is the number of non-zero elements of $\{h_k\}$. The notation $j = -1$ denotes the finest scale wavelet, $j = -2$ the next finest scale and so forth.

The quantity $\psi_{j,k}$ is the k^{th} element of the vector ψ_j . Also shifts of the vector by τ for $\tau \in \mathbb{Z}$ are denoted $\psi_{j,k}(\tau)$, where $\psi_{j,k}(\tau) = \psi_{j,k-\tau}$ is the $(k - \tau)^{\text{th}}$ element of the vector ψ_j . Given the vectors ψ_j , the collection of non-decimated wavelet vectors $\psi_{j,k}(t)$, for $t = 0, 1, \dots, T - 1$ is found by shifting the vectors, ψ_j to all integer locations k ;

$$\psi_{j,k}(t) = \psi_{j,k-t}. \quad (2.38)$$

Non-decimated wavelets are an overcomplete collection of shifted vectors; however this ensures that the resulting transformation is translation-invariant.

We also define, from Nason et al. (2000), the discrete autocorrelation wavelets which are used in representing the autocovariance functions of LSW processes. The autocorrelation wavelets $\Psi(\tau)$ are defined for $j < 0$, $\tau \in \mathbb{Z}$ as

$$\Psi_j(\tau) = \sum_{k \in \mathbb{Z}} \psi_{j,k} \psi_{j,k-\tau}. \quad (2.39)$$

Although the summation here is given over $k \in \mathbb{Z}$, the product $\psi_{j,k} \psi_{j,k-\tau}$ is zero outside the range $k = \max\{0, \tau\}$ to $L_j - 1 + \min\{0, \tau\}$. The autocorrelation wavelets are compactly supported on the interval $[1 - L_j, \dots, L_j - 1]$ and symmetric about $\tau = 0$ with $\Psi_j(0) = 1$ and $\sum_{\tau} \Psi_j(\tau) = 0$ for all j . The

cross-scale autocorrelation wavelets are given by

$$\Psi_{i,j}(\tau) = \sum_k \psi_{i,k} \psi_{j,k-\tau}. \quad (2.40)$$

We define from Nason et al. (2000) the autocorrelation wavelet inner product matrix. As we see later, this is required in order to form unbiased estimates of the evolutionary wavelet spectra. The $J \times J$ autocorrelation wavelet inner product matrix A_J is defined by

$$A_J = (A_{i,j})_{i,j=-1,\dots,-J}, \quad (2.41)$$

where

$$A_{i,j} = \sum_{\tau} \Psi_i(\tau) \Psi_j(\tau). \quad (2.42)$$

Further properties of this quantity and its computation can be found in Eckley and Nason (2005). Another useful property of the autocorrelation wavelets is that $\sum_j 2^j \Psi_j(\tau) = \delta_{0,\tau}$, as shown in Fryzlewicz et al. (2003). This also leads to the result $\sum_j 2^j A_{i,j} = 1$.

LSW Model specification

Definition 2.3.4. (Nason et al., 2000) The LSW process $\{X_{t,T}\}_{t=0,\dots,T-1}$ for $T = 2^J \geq 1$ is a triangular stochastic array with mean-square representation

$$X_{t,T} = \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} w_{j,k,T} \psi_{j,k-t} \xi_{j,k}, \quad (2.43)$$

where $\{\psi_{k,t}\}$ are discrete, real valued, compactly supported, non-decimated wavelet vectors with scale and location parameters $j \in \{-1, -2, \dots\}$ and $k \in Z$ respectively. $\{\xi_{j,k}^{(i)}\}$ are orthonormal identically distributed random variables with mean zero and $\{w_{j,k,T}\}$ is a set of amplitudes. For each $j < 0$, we assume there exists a Lipschitz continuous function $W_j(z)$ for $z \in (0, 1)$ which fulfils

the following properties:

- $\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty$, uniformly in $z \in (0, 1)$.
- The Lipschitz constants, L_j , are uniformly bounded in j and satisfy $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$.
- There exists a sequence of constants C_j , such that for each T .

$$\sup_{k=0, \dots, T-1} |w_{j,k;T} - W_j(k/T)| \leq C_j/T, \quad (2.44)$$

and $\{C_j\}$ fulfills $\sum_{j=-\infty}^{-1} C_j < \infty$.

The parameters $w_{j,k;T}$ can be thought of as time and scale dependent transfer functions while the non-decimated wavelet vectors, ψ_j , can be thought of as building blocks analogous to Fourier exponentials in a spectral domain representation. LSW processes include all stationary processes with absolutely summable autocovariance (Nason et al., 2000). In this case stationarity is characterised by an evolutionary wavelet spectrum which is constant over time, so that $S_j(z) = S_j$ for all $z \in (0, 1)$.

The evolutionary wavelet spectrum

The classical spectrum, $f(w)$, can be used to quantify the contribution to the variance in a stationary series over frequency, ω . Similarly, the *evolutionary wavelet spectrum* (EWS) quantifies the contribution to variance within a LSW process over scale, j , and rescaled time, $z = k/T$. It is defined as:

$$S_j(z) = |W_j(z)|^2. \quad (2.45)$$

A measure of local covariance may be associated with the EWS. The autocovariance function of an LSW process $X_{t,T}$ at lag τ and rescaled time location z is given by

$$c_T(z, \tau) = \mathbb{E}(X_{zT,T}, X_{zT+\tau,T}). \quad (2.46)$$

Nason et al. (2000) show that c_T tends asymptotically to a local autocovariance, $c(z, \tau)$, defined as

$$c(z, \tau) = \sum_{j=-\infty}^{-1} S_j(z) \Psi_j(\tau), \quad (2.47)$$

for $\tau \in \mathbb{Z}$ and $z \in (0, 1)$. This produces a multiscale decomposition of the dependence structure over time. Since $\Psi_j(0) = 1$, the localised variance is given by $c(z, 0) = \sum_{j=-\infty}^{-1} S_j(z)$.

The relationship between the evolutionary wavelet spectrum and local autocovariance is invertible, with the inversion formula given by

$$S_j(z) = \sum_l A_{j,l}^{-1} \sum_{\tau} c(z, \tau) \Psi_l(\tau). \quad (2.48)$$

Estimation of the EWS

The basic estimator for the EWS is the set of squared wavelet coefficients, termed the *wavelet periodogram* by analogy to the classical Fourier periodogram. First we define the empirical wavelet coefficients.

Definition 2.3.5. (Nason et al., 2000) For the process $X_{t,T}$, constructed using the wavelet system ψ , the empirical nondecimated wavelet coefficients are given by

$$d_{j,t,T} = \sum_s X_{s,T} \psi_{j,s-t}. \quad (2.49)$$

Definition 2.3.6. (Nason et al., 2000) The wavelet periodogram of $X_{t,T}$ at scale j is given by

$$I_{j,t,T} = |d_{j,t,T}|^2. \quad (2.50)$$

Each wavelet periodogram ordinate is the squared wavelet coefficient of a zero mean Gaussian time series. The wavelet periodogram therefore follows a

scaled χ^2 distribution with 1 degree of freedom.

Proposition 2.3.7. (*Nason et al., 2000*) *The expectation of the wavelet periodogram is given by*

$$\mathbb{E}(I_{j,t,T}) = \sum_{i=-\infty}^{-1} S_i(t/T)A_{ij} + 2^{-j}O(T^{-1}). \quad (2.51)$$

Also, assuming $X_{t,T}$ is Gaussian, then the variance is given by

$$\text{Var}(I_{j,t,T}) = 2 \left(\sum_{i=-\infty}^{-1} S_i(t/T)A_{i,j} \right)^2 + 2^{-j}O(T^{-1}). \quad (2.52)$$

Equation (2.51) implies that the wavelet periodogram is a biased estimator of the evolutionary wavelet spectrum. This bias is due to the redundancy in the NDWT and causes the power to be spread across the scales. An unbiased estimate of $S_j(z)$ can be obtained by setting $\hat{S}_j(z) = \sum_{i=-J}^{-1} I_{j,t,T}A_{i,j}^{-1}$.

From equation (2.52) we see that the wavelet periodogram has non-vanishing variance and so needs to be smoothed to obtain consistency. Further details of smoothing the wavelet periodogram are given in Nason et al. (2000). Associated techniques include that of Fryzlewicz et al. (2008) who introduce a new estimation method, based on the Haar-Fisz transform, for time series with piecewise constant evolutionary wavelet spectra. Also Van Bellegem and von Sachs (2008) introduce a locally adaptive estimation procedure for evolutionary wavelet spectra, allowing for processes whose spectral density functions contain abrupt changes.

Chapter 3

Locally stationary wavelet coherence

This chapter addresses the problem of estimating the dependence between two non-stationary time series. The motivation for this stems from a problem in neuroscience where population local field potential (LFP) recordings were made from two areas of a rat's brain as it performed a maze-based task designed to invoke spatial working memory and decision-making. Estimating the time-varying dependence structure between these signals can provide direct insight into the timing and fundamental nature of interactions between brain regions (Varela et al., 2001). This chapter contains theoretical results regarding the proposed dependence estimator, while Chapter 4 considers the practical implementation, including application to the described neuroscience problem.

As described in Section 2.3.1, if the series are stationary then the cross correlation function provides a useful measure of dependence in the time domain, while the coherence provides a useful measure of dependence between the series in the Fourier domain. For nonstationary series, the short time Fourier transform (Gabor, 1946) can be used to provide a localised measure of coherence, however this method suffers from the same shortcomings as with the univariate application; the time localisation is not ideal as the same window width is

used at all frequencies, hence other methods are preferable for the representation of bivariate nonstationary signals. Dahlhaus (2000) introduces a method for modelling multivariate processes based on the locally stationary process model (Dahlhaus, 1997). Priestley and Tong (1973) provide an extension to their univariate methodology (Priestley, 1965) to define the evolutionary cross spectra and coherence, however this resulting coherence is not dependent on time. Ombao and Van Bellegem (2008) introduce a coherence estimator based on time localised linear filtering. The coherence is estimated within a local time window which is automatically selected and adaptive in the sense that it is permitted to change over time.

The SLEX approach can also be used in the bivariate setting to estimate the time varying coherence (Ombao et al., 2001). As with wavelets, the SLEX basis functions are localised in both time and frequency and so provide a natural foundation for nonstationary signals. However, unlike the LSW model, the SLEX model suffers from the constraint of dyadic segmentation and in the bivariate SLEX setting it is necessary for each of the signals to follow the same segmentation. The bivariate SLEX model does have the advantage that, due to the complex nature of the basis functions, it is possible to estimate the phase between the signals. In the bivariate LSW model introduced in this chapter we do not incorporate phase estimation but the topic is discussed in further detail in Section 3.5.

Due to the natural localisation in both time and scale, wavelets are a popular tool for modelling the dependence between two non-stationary series. Unlike time resolved Fourier coherence, the wavelet transform uses shorter windows for higher frequencies, which leads to more ‘natural’ localisation. The concept of the wavelet cross spectrum, in terms of the continuous wavelet transform, was introduced by Hudgins et al. (1993), and has since been applied to fields including climatology (Maraun and Kurths (2004) and Grinsted et al. (2004)) and neuroscience (Lachaux et al., 2002). Wavelet cross-covariance and correlation has also been defined based on the maximal overlap discrete wavelet

transform (MODWT) (see Whitcher et al. (2000) and Serroukh and Walden (2001)). Their formulation assumes that the d 'th order backwards differences of the series can be modelled as a stationary process. A similar concept to that of cross-wavelet analysis is that of Hilbert wavelet pairs (Whitcher et al., 2005). Using this approach, the coherence and phase between the signals can be estimated with less redundancy than with the CWT.

In this chapter we propose a novel measure of wavelet coherence termed locally stationary wavelet coherence. This is derived from the locally stationary wavelet time series model of Nason et al. (2000). Following the work of Dahlhaus (1997), the model adopts the rescaled time principle, replacing the exponentials in the Fourier representation by a system of non decimated wavelets. An important difference between the locally stationary wavelet model and previous wavelet coherence measures lies in the particular bias correction implied by the locally stationary wavelet model. The locally stationary wavelet coherence provides a measure of the dependence between the innovations of each process, and is therefore uncontaminated by within-process dependence of each series. Furthermore, our formulation provides a model that is theoretically tractable and which can be estimated efficiently by means of the nondecimated wavelet transform.

3.1 The bivariate LSW time series model

We begin by defining the joint locally stationary wavelet (LSW) process model. Assuming that both series can be modelled as LSW processes, the model is defined as follows.

Definition 3.1.1. The joint LSW process $(X_{t,T}^{(1)}, X_{t,T}^{(2)})_{t=0,\dots,T-1}$ is a triangular stochastic array with representation

$$\begin{aligned}
 X_{t,T}^{(1)} &= \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_j^{(1)}(k/T) \psi_{j,t-k} \xi_{j,k}^{(1)}, \\
 X_{t,T}^{(2)} &= \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_j^{(2)}(k/T) \psi_{j,t-k} \xi_{j,k}^{(2)},
 \end{aligned}$$

where $\{\psi_{j,k}\}$ are discrete, real valued, compactly supported, non-decimated wavelet vectors with scale index and location parameters $j \in \{-1, -2, \dots\}$ and $k \in Z$, respectively. For each $j \leq -1$, the functions $W_j^{(i)}(k/T)$ are assumed to be Lipschitz continuous with Lipschitz constants, L_j , uniformly bounded in j and satisfying $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$. The functions are defined on *rescaled time* $z = k/T$ for $k = 0, \dots, T - 1$. Also, $\xi_{j,k}^{(i)}$ are zero mean orthonormal identically distributed random variables with the following properties

- $\text{cov}(\xi_{j,k}^{(i)}, \xi_{j',k'}^{(i)}) = \delta_{j,j'} \delta_{k,k'}$,
- $\text{cov}(\xi_{j,k}^{(1)}, \xi_{j',k'}^{(2)}) = \delta_{j,j'} \delta_{k,k'} \rho_j(k/T)$,

where $\delta_{i,j}$ is the Kronecker delta function, giving $\delta_{i,j} = 1$ for $i = j$ and 0 otherwise. The functions $\rho_j(k/T)$ are also assumed to be Lipschitz continuous with Lipschitz constants, R_j , satisfying $\sum_{j=-\infty}^{-1} 2^{-j} R_j < \infty$.

Note that the LSW formulation presented in this chapter is a simplification of the model given in definition 2.3.4, replacing the more general amplitudes, $w_{j,k,T}$, with $W_j^{(i)}(k/T)$. For the class of locally stationary Fourier processes (Dahlhaus, 1997) the general transfer function is required to ensure that the class includes time-varying AR models. In the LSW formulation the benefits of allowing a general transfer function are not clear and so we adopt the simplified formulation. As in the univariate case, the parameters $W_j^{(i)}(k/T)$ can be thought of as scale-dependent transfer functions of rescaled time and the non-decimated wavelet vectors, ψ_j , can be thought of as building blocks analogous to Fourier exponentials in the spectral domain representation. The

notation $j = -1$ denotes the finest scale wavelet, $j = -2$ the next finest scale and so forth. Throughout the thesis we assume that the errors are normally distributed. In principle other distributions could be used, however this assumption is made to ensure theoretical tractability. The formulation in Definition 3.1.1 parallels the univariate case of Nason et al (2000), but in extending this to the bivariate setting we must allow for a potential correlation structure between the two series, given by $\rho_j(k/T)$. The scale dependent quantity $\rho_j(k/T)$ is a direct measure of the dependence between the innovation sequences of each process at scale j . Therefore, unlike other measures of dependence such as the cross-correlation function, $\rho_j(k/T)$ does not depend on the other parameters of the model. This dependence measure is our main quantity of interest within the model. It is analogous to the classical coherence, $K(\omega)$, in spectral domain setting and quantifies the dependence between the series over scale and time. Though in principle other distributions could be used, throughout this thesis we assume that the innovations, $\xi_{j,k}^{(i)}$, are normally distributed, ensuring theoretical tractability.

The LSW coherence, $\rho_j(k/T)$, together with the evolutionary wavelet spectra, $S^{(i)}(z) = W^{(i)}(z)^2$ for $i = 1, 2$, provides a description of the joint process. This can be demonstrated by means of a brief example. Suppose that for both series the spectrum is zero at all scales apart from $j = -2$. At this scale, both spectra are constant at $S_{-2}^{(1)}(z) = S_{-2}^{(2)}(z) = 5$. The processes are assumed to have zero coherence at all scales apart from $j = -2$, and at this scale the LSW coherence is 0 for the first third of the series, then increases linearly to 1 throughout the second third of the series, then remains at 1 for the last segment of data. The EWS and coherence are plotted in Figure 3.1, along with one simulation of time series with these properties. The resulting series are independent until time $T/3$, exactly the same for the latter half of the series (since the coherence is specified to be 1 at the only non-zero scale) and for the middle section of the data, the series show a transition from independence to complete dependence.

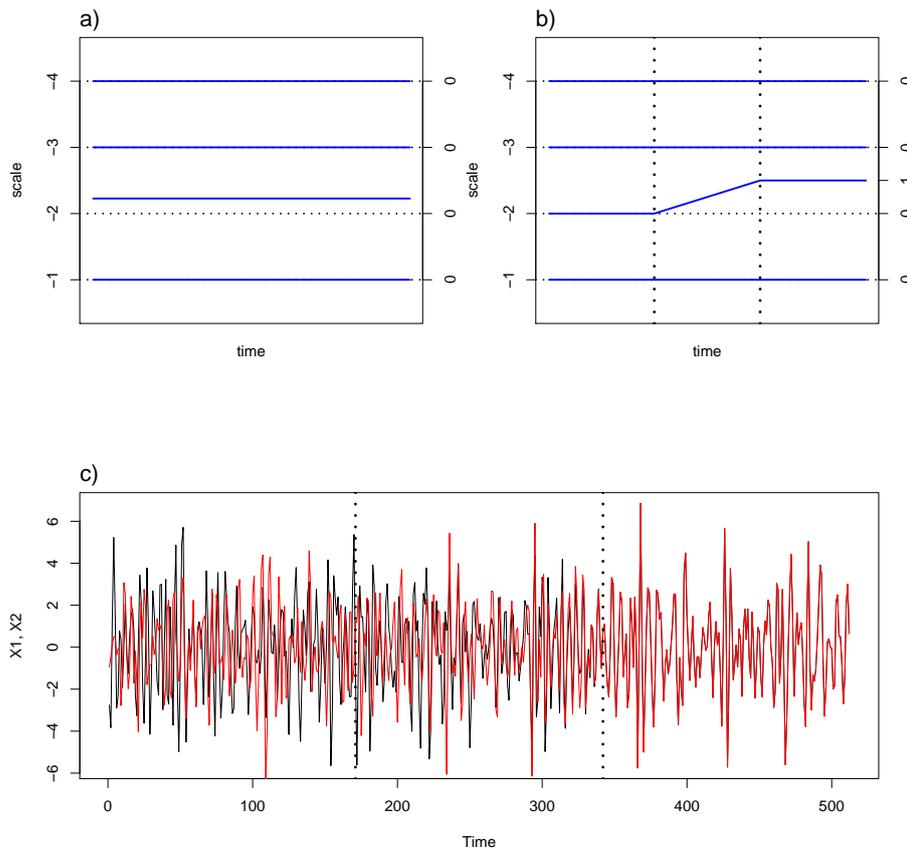


Figure 3.1: Example of the construction of a bivariate LSW process. a) EWS of $X^{(1)}$ and $X^{(2)}$ b) LSW coherence between the series, c) simulation of $X^{(1)}$ (black) and $X^{(2)}$ (red).

3.2 The evolutionary wavelet cross-spectrum and coherence

The *evolutionary wavelet cross-spectrum* provides a measure of the dependence between two time series. It is analogous to the EWS in the univariate setting and provides a natural stage in estimating the LSW coherence.

Definition 3.2.1. The *evolutionary wavelet cross-spectrum* of the LSW processes $X_{t,T}^{(1)}$ and $X_{t,T}^{(2)}$, is given by

$$C_j(z) = W_j^{(1)}(z)W_j^{(2)}(z)\rho_j(z), \quad \text{for } j \in \{-1, -2, \dots\}, \quad z \in [0, 1]$$

A measure of local cross-covariance, introduced in Definition 3.2.2, may be associated with the evolutionary wavelet cross-spectrum. As shown by Proposition 3.2.3, the LSW process cross-covariance given by $c_T^{(1,2)}(z, \tau) = \text{cov}(X_{zT,T}^{(1)}, X_{zT+\tau,T}^{(2)})$, asymptotically tends to the local cross-covariance.

Definition 3.2.2. The local cross-covariance of the LSW processes $X_{t,T}^{(1)}$ and $X_{t,T}^{(2)}$ at lag $\tau \in \mathbb{Z}$, is given by

$$c^{(1,2)}(z, \tau) = \sum_{j=-\infty}^{-1} C_j(z)\Psi_j(\tau), \quad (3.1)$$

where $\Psi_j(\tau) = \sum_k \psi_{j,k}\psi_{j,k-\tau}$ is the autocorrelation wavelet.

Proposition 3.2.3. *Assuming there exists a constant, C , such that for all $j \in \{-1, -2, \dots\}$, $|W_j^{(i)}(z)| \leq C2^{j/2}$. Then for $\tau \in \mathbb{Z}$ and $z \in (0, 1)$, $|c_T^{(1,2)}(z, \tau) - c^{(1,2)}(z, \tau)| = O(T^{-1})$.*

The assumption of Proposition 3.2.3 is satisfied if, for example, $X_{t,T}$ is a white noise process. Before verifying this proposition, we first introduce the following lemma which provides a useful result of the properties of products of Lipschitz continuous variables.

Lemma 3.2.4. *The Lipschitz property of $W_j^{(i)}(k/T)$ stated in Definition 3.1.1 implies that $|W_j^{(i)}(k/T) - W_j^{(i)}(t/T)| \leq T^{-1}(L_j|k - t|)$ for some Lipschitz constant L_j , and similarly $|\rho_j(k/T) - \rho_j(t/T)| \leq T^{-1}(R_j|k - t|)$. Furthermore, assume that there exists a positive constant, C , such that for all j , $|W_j(z)| \leq C2^{j/2}$. Then by the property of products of Lipschitz continuous variables, denoting $B_i = \max(L_i, R_i)$, we have*

$$|W_i^{(1)}(k/T)W_i^{(2)}(k/T)\rho_i(k/T) - W_i^{(1)}(t/T)W_i^{(2)}(t/T)\rho_i(t/T)| \leq T^{-1}CB_i|k - t|.$$

Using this result, we are now able to prove Proposition 3.2.3, on the convergence of the process cross-covariance.

Proof of Proposition 3.2.3.

$$\begin{aligned} c_T^{(1,2)}(z, \tau) &= \text{cov}(X_{t,T}^{(1)}, X_{t+\tau,T}^{(2)}), \\ &= E\left(\sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_j^{(1)}(k/T)\psi_{j,t-k}\xi_{j,k}^{(1)} \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_j^{(2)}(k/T)\psi_{j,t+\tau-k}\xi_{j,k}^{(2)}\right), \\ &= \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} C_j(k/T)\psi_{j,t-k}\psi_{j,t+\tau-k}. \end{aligned}$$

Using the result from Lemma 3.2.4 we obtain

$$\begin{aligned} |c_T^{(1,2)}(z, \tau) - c^{(1,2)}(z, \tau)| &= \left| \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} C_j(k/T)\psi_{j,t-k}\psi_{j,t+\tau-k} - \sum_{j=-\infty}^{-1} C_j(t/T)\Psi_j(\tau) \right|, \\ &\leq |T^{-1} \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} CB_j|k - t|\psi_{j,t-k}\psi_{j,t+\tau-k}|, \\ &= O(T^{-1}). \end{aligned}$$

Note also that the support of $\Psi_j(\tau)$ is bounded by $K2^{-j}$, and so the distance $|k - t|$ is bounded by this amount also. The Lipschitz constants B_j are uni-

formly bounded in j with $\sum_j B_j 2^{-j} < \infty$ (as stated in Definition 3.1.1) and $\Psi_j(\tau) = O(1)$.

□

The evolutionary wavelet cross-spectrum describes the contribution to the cross-covariance at a particular rescaled time, z , and scale, j . This interpretation becomes clear when considering the cross-covariance at zero lag. Since $\Psi_j(0) = 1$, substituting into equation (3.1) gives $c^{(1,2)}(z, 0) = \sum_{j=-\infty}^{-1} C_j(z)$. As the function $\Psi_j(\tau)$ is symmetric, this implies $c^{(1,2)}(z, \tau) = c^{(1,2)}(z, -\tau)$ and so the model only permits symmetric cross-covariances. The cross-covariance is maximised at $\tau = 0$ and so it is required that the two series are aligned so that there is no time delay between the signals. Note that the analysis of bivariate signals containing a time delay is considered in Chapter 6, using a wavelet lifting framework.

The relationship between the LSW cross-spectrum and the local cross-covariance is invertible, with the inverse relationship given by

$$C_j(z) = \sum_l A_{j,l}^{-1} \sum_{\tau} c^{(1,2)}(z, \tau) \Psi_l(\tau), \quad (3.2)$$

where $A_{i,j}$ is the autocorrelation wavelet inner product matrix.

The LSW cross-spectrum, $C_j(z)$, provides a measure of the dependence between the two series. However it is clear that deviations in $C_j(z)$ can be caused by fluctuations in the EWS of each of the processes as well as by changes in the dependence, and hence the wavelet cross-spectrum cannot be used alone as a measure of dependence between the series.

The LSW coherence, $\rho_j(z)$, can be represented in terms of the evolutionary wavelet cross-spectrum and the individual EWS of each process as shown in equation (3.3), providing a normalised measure of the relationship between two series

$$\rho_j(z) = \frac{C_j(z)}{\sqrt{S_j^{(1)}(z)S_j^{(2)}(z)}}. \quad (3.3)$$

In particular, the model permits a stationary or non-stationary coherence at each scale, $\rho_j(z)$, independently of the stationarity of the individual wavelet spectra $S_j^{(i)}(z)$ or the wavelet cross spectrum $C_j(z)$. The locally stationary wavelet coherence ranges from -1 , indicating complete negative correlation at scale j and rescaled time z , to $+1$ indicating complete correlation. A value of close to zero indicates a lack of correlation between the two series at the given scale and location. In the case where one of the spectra is zero, the coherence is undefined.

3.3 Estimation

The basic pre-estimator of the EWS is a quantity known as the wavelet periodogram (see Section 2.3.4). Similarly, the wavelet cross-spectrum is estimated from the *wavelet cross-periodogram*. Following Nason et al. (2000), the empirical nondecimated wavelet coefficients for the LSW process $X_{t,T}^{(i)}$, constructed using the wavelet system ψ , are given by $d_{j,t,T}^{(i)} = \sum_s X_{s,T}^{(i)} \psi_{j,s-t}$. The non decimated wavelet coefficients are used to construct the cross periodogram:

Definition 3.3.1. The wavelet cross-periodogram for the LSW processes $X_{t,T}^{(1)}, X_{t,T}^{(2)}$, is given by $I_{j,t,T}^{(1,2)} = d_{j,t,T}^{(1)} d_{j,t,T}^{(2)}$.

Proposition 3.3.2. Assume there exists a constant, C , such that for all $j \in \{-1, -2, \dots\}$, $|W_j^{(i)}(z)| \leq C2^{j/2}$. The expectation of the cross-periodogram, $I_{j,t,T}^{(1,2)}$, is given by

$$\mathbb{E}(I_{j,t,T}^{(1,2)}) = \sum_{i=-\infty}^{-1} W_i^{(1)}(t/T) W_i^{(2)}(t/T) \rho_i(t/T) A_{i,j} + 2^{-j} O(T^{-1}).$$

Also, the variance is given by

$$\begin{aligned} \text{Var}(I_{j,t,T}^{(1,2)}) &= \sum_{i=-\infty}^{-1} S_i^{(1)}(t/T) A_{i,j} \sum_{i=-\infty}^{-1} S_i^{(2)}(t/T) A_{i,j} \\ &+ \left(\sum_{i=-\infty}^{-1} W_i^{(1)}(t/T) W_i^{(2)}(t/T) \rho_i(t/T) A_{i,j} \right)^2 + 2^{-j} O(T^{-1}). \end{aligned}$$

Proof of Proposition 3.3.2. The expectation of the evolutionary wavelet cross-periodogram is given by

$$\begin{aligned} \mathbb{E}(I_{j,t,T}^{(1,2)}) &= \mathbb{E} \left[\sum_s \sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_i^{(1)}(k/T) \psi_{i,s-k} \xi_{i,k}^{(1)} \psi_{j,s-t} \sum_s \sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_i^{(2)}(k/T) \psi_{i,s-k} \xi_{i,k}^{(2)} \psi_{j,s-t} \right], \\ &= \mathbb{E} \left[\sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_i^{(1)}(k/T) \Psi_{i,j}(k-t) \xi_{i,k}^{(1)} \sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_i^{(2)}(k/T) \Psi_{i,j}(k-t) \xi_{i,k}^{(2)} \right], \\ &= \sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_i^{(1)}(k/T) W_i^{(2)}(k/T) \rho_i(k/T) \Psi_{i,j}^2(k-t), \\ &= \sum_{i=-\infty}^{-1} W_i^{(1)}(t/T) W_i^{(2)}(t/T) \rho_i(t/T) A_{i,j} + \text{Rest}_T, \end{aligned}$$

since $\text{cov}(\xi_{j,k}^{(1)}, \xi_{j',k'}^{(2)}) = \delta_{j,j'} \delta_{k,k'} \rho_j(k/T)$. The remainder can be found using the result from Lemma 3.2.4, and that the length of support of $\Psi_{i,j}^2(k-t)$ is bounded by $K(2^{-i} + 2^{-j})$, to be

$$\begin{aligned} \text{Rest}_T &\leq T^{-1} \sum_{i=-\infty}^{-1} CK B_i (2^{-i} + 2^{-j}) \sum_{k=-\infty}^{\infty} \Psi_{i,j}^2(k-t), \\ &\leq T^{-1} 2^{-j} CK \sum_{i=-\infty}^{-1} B_i 2^{-i} \sum_{k=-\infty}^{-1} 2^k A_{i,k} + T^{-1} 2^{-j} CK \sum_{i=-\infty}^{-1} B_i 2^{-i} \sum_{k=-\infty}^{-1} 2^k A_{k,j}, \\ &= 2^{-j} O(T^{-1}). \end{aligned}$$

Note also that $\sum_{k=-\infty}^{-1} 2^k A_{i,k} = 1$ for all i (Fryzlewicz et al., 2003) and the Lipschitz constants B_i are uniformly bounded in i with $\sum_i B_i 2^{-i} \leq \infty$. For the variance, we decompose the product using Isserlis's Theorem for zero mean

Gaussian random variables (Isserlis, 1918) to give

$$\text{Var}(I_{j,k,T}^{(1,2)}) = E(I_{j,t,T}^{(1)})E(I_{j,t,T}^{(2)}) + E(I_{j,t,T}^{(1,2)})^2,$$

which on substitution of the relevant terms provides the stated result. \square

We can see from Proposition 3.3.2 that the expectation of the wavelet cross-periodogram is composed of the sum of wavelet cross-spectra, $C_j(z)$. The wavelet cross-periodogram is therefore a natural estimator of the wavelet cross-spectrum, but we first need to correct for the bias incurred by the matrix A_J . Also, since the wavelet cross-periodogram has non-vanishing variance, it needs to be smoothed to obtain consistency. For this we use simple moving average smoothing. Other more advanced smoothing techniques such as wavelet shrinkage may also be viable.

The estimator of the wavelet cross-spectrum is therefore constructed by bias-correcting the periodogram to give

$$\tilde{I}_l(t/T) = \sum_{j=-J^*}^{-1} I_{j,t,T}^{(1,2)} A_{l,j}^{-1}. \quad (3.4)$$

Then smoothing over a window of $\pm M$ in time to give

$$\hat{C}_l(t/T) = \frac{1}{2M+1} \sum_{m=-M}^M \tilde{I}_{l,t+m,T}^{(1,2)}. \quad (3.5)$$

The maximum number of estimable scales, J^* , is determined by the length of the series. As shown in the proof of Proposition 3.3.4 below, it is necessary to choose $J^* < \log_2(T)$ to ensure the consistency of $\tilde{I}_l(z)$. The choice of J^* is discussed in more detail in Section 4.1.2. These correction and smoothing steps can be performed in any order.

In order to find the expectation and MSE of the estimator $\hat{C}_l(z)$, we first provide the intermediate result for the smoothed cross-periodogram.

Proposition 3.3.3. *The expectation of the smoothed cross-periodogram is given by*

$$\mathbb{E}(\tilde{I}_{j,t,T}^{(1,2)}) = \sum_{i=-\infty}^{-1} W_i^{(1)}(t/T) W_i^{(2)}(t/T) \rho_i(t/T) A_{i,j} + (2^{-j} + M)O(T^{-1}),$$

and the mean squared error is

$$MSE(\tilde{I}_{j,t,T}^{(1,2)}) = 2^{-2j} M^{-1} O(1).$$

Proof of Proposition 3.3.3.

$$\begin{aligned} \mathbb{E}(\tilde{I}_{j,t,T}^{(1,2)}) &= \frac{1}{2M+1} \sum_{m=-M}^M \mathbb{E}(I_{j,t+m,T}^{(1,2)}), \\ &= \frac{1}{2M+1} \sum_{m=-M}^M \sum_{i=-\infty}^{-1} W_i^{(1)}\left(\frac{t+m}{T}\right) W_i^{(2)}\left(\frac{t+m}{T}\right) \rho_i\left(\frac{t+m}{T}\right) A_{i,j} + 2^{-j} O(T^{-1}), \\ &= \sum_{i=-\infty}^{-1} W_i^{(1)}(t/T) W_i^{(2)}(t/T) \rho_i(t/T) A_{i,j} + Rest_T. \end{aligned}$$

Using Lemma 3.2.4, the remainder can be found as

$$\begin{aligned} Rest_T &\leq \frac{1}{2M+1} \sum_{m=-M}^M \sum_{i=-\infty}^{-1} T^{-1} B_i |m| A_{i,j} + 2^{-j} O(T^{-1}), \\ &\leq \frac{1}{(2M+1)T} \sum_{m=-M}^M |m| \sum_{i=-\infty}^{-1} B_i 2^{-i} \sum_{k=-\infty}^{-1} 2^k A_{k,j} + 2^{-j} O(T^{-1}), \\ &= \frac{M(M+1)}{(2M+1)T} \sum_{i=-\infty}^{-1} B_i 2^{-i} \sum_{k=-\infty}^{-1} 2^k A_{k,j} + 2^{-j} O(T^{-1}), \\ &= MO(T^{-1}) + 2^{-j} O(T^{-1}), \\ &= (M + 2^{-j}) O(T^{-1}). \end{aligned}$$

The variance of the smoothed cross-periodogram is given by

$$\begin{aligned} Var(\tilde{I}_{j,t,T}^{(1,2)}) &= \frac{1}{(2M+1)^2} \sum_{m=-M}^M \sum_{m'=-M}^M Cov\left(I_{j,t+m,T}^{(1,2)}, I_{j,t+m',T}^{(1,2)}\right), \\ &= \frac{1}{(2M+1)^2} \sum_{m=-M}^M \sum_{\tau} Cov\left(I_{j,t+m,T}^{(1,2)}, I_{j,t+m+\tau,T}^{(1,2)}\right), \end{aligned}$$

where $\tau = m' - m$. Decomposing the covariance using Isserlis's Theorem leads to the following expression, in terms of the periodogram ordinates of each process:

$$\begin{aligned} \text{Cov}\left(I_{j,t+m,T}^{(1,2)}, I_{j,t+m+\tau,T}^{(1,2)}\right) &= \mathbb{E}\left(d_{j,t+m,T}^{(1)}, d_{j,t+m+\tau,T}^{(1)}\right)\mathbb{E}\left(d_{j,t+m,T}^{(2)}, d_{j,t+m+\tau,T}^{(2)}\right), \\ &+ \mathbb{E}\left(d_{j,t+m,T}^{(1)}, d_{j,t+m+\tau,T}^{(2)}\right)\mathbb{E}\left(d_{j,t,T}^{(2)}, d_{j,t+m+\tau,T}^{(1)}\right). \end{aligned}$$

Here the first product is of the autocovariances of the wavelet periodogram ordinates and the last product consists of the cross-covariances, both for a fixed scale, j . These expectations can be shown using similar techniques as in proof of Proposition 3.3.2 to take the values:

$$\begin{aligned} \mathbb{E}(d_{j,t,T}, d_{j,t+m+\tau,T}) &= \sum_{i=-\infty}^{-1} S_i(t/T)A_{i,j}^\tau + (|m| + 2^{-j})O(T^{-1}), \\ \mathbb{E}(d_{j,t,T}^{(1)}, d_{j,t+m+\tau,T}^{(2)}) &= \sum_{i=-\infty}^{-1} W_i^{(1)}(t/T)W_i^{(2)}(t/T)\rho_i(t/T)A_{i,j}^\tau + (|m| + 2^{-j})O(T^{-1}), \end{aligned}$$

Where $A_{i,j}^\tau$ is a lagged version of the nondecimated wavelet inner product matrix, given by $A_{i,j}^\tau = \sum_n \Psi_{i,j}(n)\Psi_{i,j}(n + \tau)$. Substituting the relevant terms and summing over m , the variance is given by

$$\begin{aligned} \text{Var}(\tilde{I}_{j,t,T}^{(1,2)}) &= \frac{1}{(2M+1)^2} \sum_{m=-M}^M \sum_{\tau} \left[\sum_{i=-\infty}^{-1} S_i^{(1)}\left(\frac{t}{T}\right)A_{i,j}^\tau \sum_{i=-\infty}^{-1} S_i^{(2)}\left(\frac{t}{T}\right)A_{i,j}^\tau, \right. \\ &+ \left. \left(\sum_{i=-\infty}^{-1} W_i^{(1)}\left(\frac{t}{T}\right)W_i^{(2)}\left(\frac{t}{T}\right)\rho_i\left(\frac{t}{T}\right)A_{i,j}^\tau \right)^2 + (2^{-j} + |m|)O(T^{-1}) \right], \\ &= \frac{1}{(2M+1)^2} \sum_{m=-M}^M \left[I + II + (2^{-j} + |m|)O(T^{-1}) \right]. \end{aligned}$$

To find a bound for the variance we look at these terms separately. The first term can be bounded by

$$\begin{aligned}
 I &\leq \sum_{\tau} \left| \sum_{i=-\infty}^{-1} S_i^{(1)}\left(\frac{t}{T}\right) A_{i,j}^{\tau} \right| \sum_{\tau} \left| \sum_{i=-\infty}^{-1} S_i^{(2)}\left(\frac{t}{T}\right) A_{i,j}^{\tau} \right|, \\
 &= \sum_{\tau} \left| \sum_{i=-\infty}^{-1} S_i^{(1)}(t/T) \sum_n \Psi_i(n) \Psi_j(n+\tau) \right| \sum_{\tau} \left| \sum_{i=-\infty}^{-1} S_i^{(2)}(t/T) \sum_n \Psi_i(n) \Psi_j(n+\tau) \right|, \\
 &\leq \sum_n |c^{(1)}(t, n)| \sum_{\tau} |\Psi_j(n+\tau)| \cdot \sum_n |c^{(2)}(t, n)| \sum_{\tau} |\Psi_j(n+\tau)|, \\
 &= 2^{-2j} O(1),
 \end{aligned}$$

assuming $\sup_{z \in [0,1]} \sum_{\tau} |c(z, \tau)| < \infty$, and using that $\sum_{\tau} |\Psi_j(\tau)| = 2^{-j} O(1)$ (Nason et al., 2000, Proof of Proposition 3). A bound for the second term can be found in a similar fashion, but replacing the autospectrum of each process with the cross spectrum. This provide the result $II \leq 2^{-2j} O(1)$. Hence the variance is of order

$$\begin{aligned}
 \text{Var}(\tilde{I}_{t,T}^{(j)}) &= \frac{1}{(2M+1)^2} \sum_{m=-M}^M \left[2^{-2j} O(1) + 2^{-2j} O(1) + (2^{-j} + |m|) O(T^{-1}) \right], \\
 &= M^{-1} 2^{-2j} O(1) + 2^{-j} O(T^{-1}), \\
 &= M^{-1} 2^{-2j} O(1),
 \end{aligned}$$

and the mean squared error is given by

$$\text{MSE}(\tilde{I}_{j,t,T}^{(1,2)}) = 2^{-2j} M^{-1} O(1) + ((2^{-j} + M) O(T^{-1}))^2 = 2^{-2j} M^{-1} O(1).$$

□

Using this result, we are now in the position to prove the consistency of our estimator, $\hat{C}_l(z)$.

Proposition 3.3.4. *Suppose that the assumptions from Proposition 1 and 2 hold, and let $J^* = \alpha \log_2(T)$ where $\alpha \in (0, 1)$. Then for Haar and Shannon wavelets the estimator $\hat{C}_l(t/T)$ converges in probability to $W_l^{(1)}(t/T) W_l^{(2)}(t/T) \rho_l(t/T)$ for each fixed scale l , provided that $M^{-1} T^{\alpha} \rightarrow 0$ as $T \rightarrow \infty$ and $M \rightarrow \infty$.*

Proof of Proposition 3.3.4.

$$\begin{aligned}\mathbb{E}\hat{C}_l(t/T) &= \sum_{j=-J^*}^{-1} \mathbb{E}\tilde{I}_{j,t,T}^{(1,2)} A_{l,j}^{-1}, \\ &= W_l^{(1)}(t/T)W_l^{(2)}(t/T)\rho_l(t/T) + Rest_T,\end{aligned}$$

as the residual is given by

$$\begin{aligned}Rest_T &\leq \sum_{j=-J^*}^{-1} (M + 2^{-j})O(T^{-1})A_{l,j}^{-1}, \\ &= MO(T^{-1}) \sum_{j=-J^*}^{-1} A_{l,j}^{-1} + O(T^{-1}) \sum_{j=-J^*}^{-1} 2^{-j} A_{l,j}^{-1} = MO(T^{-1}),\end{aligned}$$

since $A_{l,j}^{-1}$ is bounded and $\sum_{j=-J^*}^{-1} 2^{-j} A_{l,j}^{-1} < \infty$. For the mean squared error, we use a similar result to that of Fryzlewicz and Nason (2006, Theorem 4.1).

The mean squared error can be written in the form

$$\begin{aligned}MSE(\hat{C}_l(t/T)) &= \mathbb{E}(\hat{C}_l(t/T) - W_l^{(1)}(t/T)W_l^{(2)}(t/T)\rho_l(t/T))^2, \\ &= \mathbb{E}\left(\sum_{j=-J^*}^{-1} \tilde{I}_{j,t,T}^{(1,2)} A_{l,j}^{-1} - \sum_{j=-\infty}^{-1} \beta_j(t/T)A_{l,j}^{-1}\right)^2,\end{aligned}$$

where $\beta_j(z) = \sum_{i=-\infty}^{-1} W_i^{(1)}(z)W_i^{(2)}(z)\rho_i(z)A_{i,j}$. Splitting this into two terms gives

$$\begin{aligned}MSE(\hat{C}_l(t/T)) &\leq 2\mathbb{E}\left(\sum_{j=-J^*}^{-1} (\tilde{I}_{j,t,T}^{(1,2)} - \beta_j(t/T))A_{l,j}^{-1}\right)^2 + 2\left(\sum_{j=-\infty}^{-J^*-1} \beta_j(t/T)A_{l,j}^{-1}\right)^2, \\ &= I + II.\end{aligned}$$

From Theorem 2 of Nason et al. (2000), we have that for Haar wavelets $A_{l,j}^{-1} \leq C2^{l/2}2^{j/2}$. Using this result we find

$$\begin{aligned}
I &\leq 2 \sum_{j=J^*}^{-1} \mathbb{E}(\tilde{I}_{j,t,T}^{(1,2)} - \beta_j(t/T))^2 (A_{l,j}^{-1})^2 \\
&\leq 2 \sum_{j=-J^*}^{-1} C 2^l 2^j 2^{-2j} M^{-1} O(1), \\
&= M^{-1} O(1) \sum_{j=-J^*}^{-1} 2^{-j}, \\
&= M^{-1} O(T^\alpha)
\end{aligned}$$

as $\mathbb{E}(\tilde{I}_{t,T}^{(1,2)} - \beta_j(t/T))^2$ is the MSE of the smoothed cross-periodogram, given by $MSE(\tilde{I}_{j,t,T}^{(1,2)}) = 2^{-2j} M^{-1} O(1)$. To consider the case of wavelets other than Haar, we first consider the case of Shannon wavelets. For the Shannon wavelet we have $A_{l,j}^{-1} \leq 2^j$ which results in a bound of $M^{-1} O(1) \leq M^{-1} O(T^\alpha)$. The bound is therefore valid for Shannon's wavelet as well as Haar, and so we conjecture that the proof is valid for all other Daubechies compactly supported wavelets.

For the second term we further assume $\beta_j(t/T) < C_1$ (as in Fryzlewicz and Nason (2006)), to give

$$II \leq 2 \left(\sum_{j=-\infty}^{-J^*-1} C_1 2^{l/2} 2^{j/2} \right)^2 \leq 2 \cdot 2^l \left(\sum_{j=-\infty}^{-J^*-1} C_1 2^{j/2} \right)^2 = 2^l O(1).$$

Combining these results gives $MSE(\hat{C}_l(t/T)) = M^{-1} O(T^\alpha)$. The estimator $\hat{C}_l(t/T)$, then converges in probability to $W_l^{(1)}(t/T) W_l^{(2)}(t/T) \rho_l(t/T)$ provided that $MT^\alpha \rightarrow 0$ as $T \rightarrow \infty$ and $M \rightarrow \infty$ for each fixed scale l . \square

The wavelet periodograms, $I_{j,t,T}^{(i)}$ for $i = 1, 2$ are smoothed and corrected in the same manner to give $\hat{S}_l^1(t/T)$ and $\hat{S}_l^2(t/T)$, and a similar result holds. From Proposition 3.3.4, we see that as M increases, our estimators of the cross and auto spectra converge in probability to the expected quantities.

Given estimates of the wavelet cross-spectrum, $\widehat{C}_l(t/T)$, and EWS, $\widehat{S}_l^{(i)}(t/T)$, of each process and provided that $S_l^{(1)}(t/T) > 0$ and $S_l^{(2)}(t/T) > 0$, the estimator of the locally stationary wavelet coherence given by

$$\widehat{\rho}_l(t/T) = \frac{\widehat{C}_l(t/T)}{(\widehat{S}_l^{(1)}(t/T)\widehat{S}_l^{(2)}(t/T))^{1/2}}, \quad (3.6)$$

converges in probability to $\rho_l(t/T)$ by Slutsky's theorem (Slutsky, 1925).

3.4 Relationship between the LSW and Fourier coherence

In order to assess the performance of the LSW coherence estimator, it is interesting to compare this method with existing methods for quantifying dependence between two time series. This section considers the link between LSW coherence and classical Fourier coherence.

To understand the relationship between the Fourier and LSW coherence measures, we express both quantities in terms of the covariance function. Using the definition of the LSW cross-spectra from equation (3.3), and the corresponding representation of the wavelet auto-spectra, results in the following expression for the LSW coherence:

$$\rho_j(z) = \frac{\sum_{\tau} c^{(1,2)}(z, \tau) \sum_l A_{jl}^{-1} \Psi_l(\tau)}{\sqrt{\sum_{\tau} c^{(1)}(z, \tau) \sum_l A_{jl}^{-1} \Psi_l(\tau) \sum_{\tau} c^{(2)}(z, \tau) \sum_l A_{jl}^{-1} \Psi_l(\tau)}}. \quad (3.7)$$

The equivalent representation for the Fourier coherence is:

$$K(\omega) = \frac{|\sum_{\tau} c_X^{(1,2)}(\tau) e^{-i\omega\tau}|}{\sqrt{\sum_{\tau} c_X^{(1)}(\tau) \cos(\omega\tau) \sum_{\tau} c_X^{(2)}(\tau) \cos(\omega\tau)}}. \quad (3.8)$$

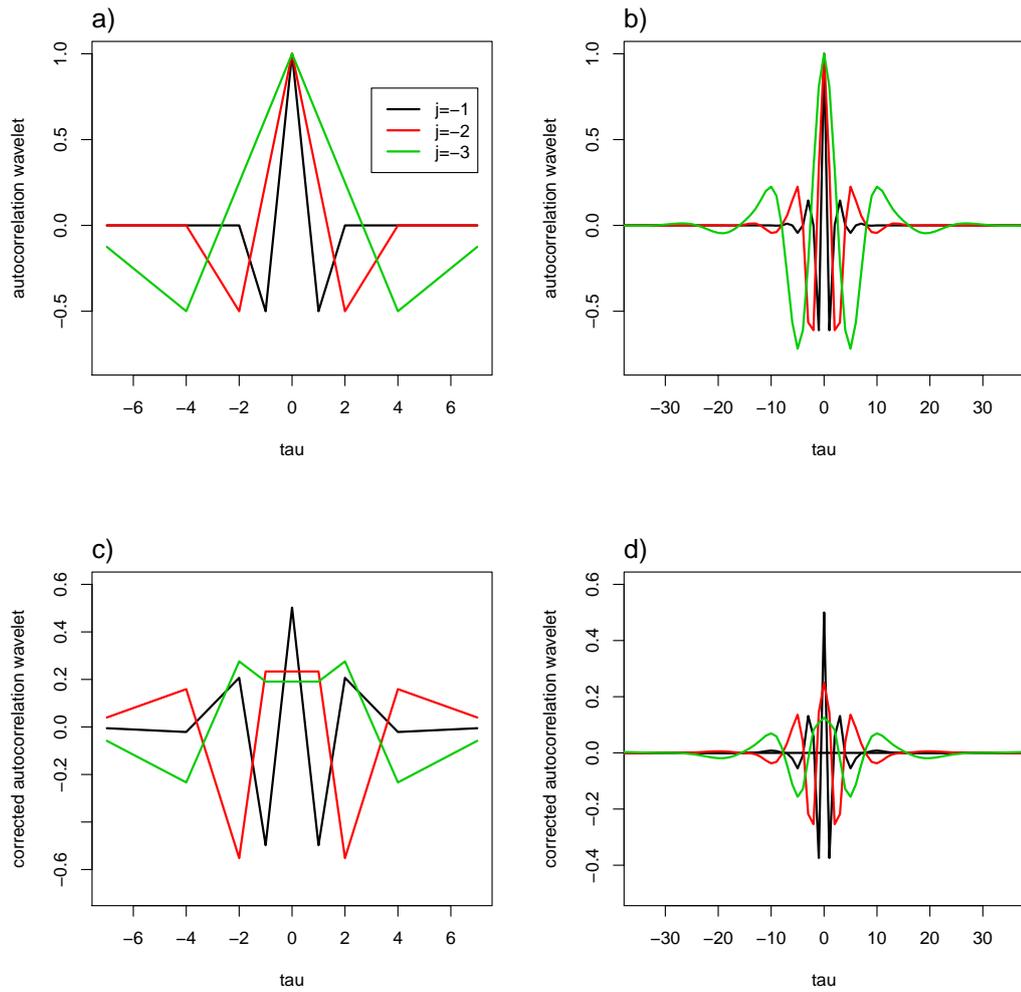


Figure 3.2: Autocorrelation wavelets and corrected autocorrelation wavelets for Haar (panels a and c) and Daubechies least asymmetric $N = 6$ (panels b and d).

Since in the LSW setting we have a measure of the localised covariance, the LSW coherence is defined over rescaled time, $z = k/T$, giving a localised measure of the coherence. In the case where the series is stationary, there is no dependence on z in equation (3.7) and the two representations become more comparable.

Comparing equations (3.7) and (3.8) in the stationary setting, we see that the functions $\sum_l A_{jl}^{-1} \Psi_l(\tau)$ in the LSW representation perform the same task as the Fourier exponentials in equation (3.8). These functions are a modified version of the autocorrelation wavelets, $\Psi_j(\tau)$, and unlike the Fourier exponentials which have a constant amplitude over the entire length of T , the amplitude of these functions decay as $|\tau| \rightarrow \infty$. The rate of decay depends on the scale, j , with the functions decaying at a much faster rate for fine scales. The autocorrelation wavelets and the corrected quantity $\sum_l A_{jl}^{-1} \Psi_l(\tau)$ are plotted in Figure 3.2 for two choices of wavelet.

The following result from von Sachs et al. (1997) provides a useful relation between the wavelet periodogram and the classical Fourier spectrum.

Theorem 3.4.1 (von Sachs et al. (1997)). *If $X_{t,T}$ is second order stationary, with spectral density $f(\omega)$, then*

$$\lim_{T \rightarrow \infty} \mathbb{E}(I_{j,k,T}) = \int_{-\pi}^{\pi} f(\omega) \widehat{\Psi}_j(\omega) d\omega,$$

where $\widehat{\Psi}_j(\omega) = |\widehat{\psi}_{j,k}(\omega)|^2$ and

$$\widehat{\psi}_{j,k}(\omega) = \sum_{s=-\infty}^{\infty} \psi_{j,k}(s) e^{i\omega s}.$$

The quantity $\widehat{\psi}_{j,k}(\omega)$ is not independent of k but is a phase shifted version of $\widehat{\psi}_{j_0}(\omega)$. However taking absolute values cancels the phase shift, and hence the expectation in Theorem 3.4.1 is independent of k . $\widehat{\Psi}_j(\omega)$ can be thought of as the (squared) Fourier domain expression of the inverse DWT operator.

For the Shannon wavelet, Nason et al. (2000) derive

$$\widehat{\Psi}_j(\omega) = 2^{-j} \chi_{C_j}(\omega), \quad (3.9)$$

where $\chi_A(\omega)$ is the indicator function of the set A and

$$C_j = \left[-\frac{\pi}{2^{-j-1}}, -\frac{\pi}{2^{-j}} \right] \cup \left[\frac{\pi}{2^{-j}}, \frac{\pi}{2^{-j-1}} \right]. \quad (3.10)$$

Using Theorem 3.4.1, equation (3.9) and assuming that $\omega > 0$, the expectation of the wavelet periodogram of a stationary process using Shannon wavelets, can be expressed in terms of the Fourier spectrum as

$$\mathbb{E}(I_{k,T}^j) = \int_{-\pi}^{\pi} f(\omega) 2^{-j} \chi_{C_j} d\omega = 2^{-j} \int_{\pi/2^{-j}}^{\pi/2^{-j-1}} f(\omega) d\omega. \quad (3.11)$$

For the Shannon wavelet, $A_{j,l=0}$ for $j \neq l$ since the supports of different $\widehat{\Psi}_j(\omega)$ do not overlap, and $A_{j,j} = 2^{-j}$ for $j < 0$ (Nason et al., 2000). Therefore the expectation of the corrected periodogram is given by

$$\mathbb{E}\left(\sum_l A_{jl}^{-1} I_{k,T}^l\right) = A_{jj}^{-1} \mathbb{E}(I_{k,T}^j) = \int_{\pi/2^{-j}}^{\pi/2^{-j-1}} f(\omega) d\omega \quad (3.12)$$

So that for stationary processes, the EWS, $S_j(z)$ can be thought of as the integral of the Fourier spectrum over dyadic intervals: $S_j(z) = \int_{\pi/2^{-j}}^{\pi/2^{-j-1}} f(\omega) d\omega$. This result extends naturally to the bivariate setting.

Proposition 3.4.2. *If $X_{t,T}^{(1)}$ and $X_{t,T}^{(2)}$ are second order stationary, with cross-spectral density $f^{(1,2)}(\omega)$, then*

$$\lim_{T \rightarrow \infty} \mathbb{E}(I_{j,k,T}^{(1,2)}) = \int_{-\pi}^{\pi} f^{(1,2)}(\omega) \widehat{\Psi}_j(\omega) d\omega, \quad (3.13)$$

where $\widehat{\Psi}_j(\omega)$ is as defined in Theorem 3.4.1.

Proof of Proposition 3.4.2. Using the spectral representation of $X_{t,T}^{(1)}$ and $X_{t,T}^{(2)}$:

$$\begin{aligned} X_t^{(1)} &= \int_{-\pi}^{\pi} \exp(i\omega t) d\zeta^{(1)}(\omega), \\ X_t^{(2)} &= \int_{-\pi}^{\pi} \exp(i\omega t) d\zeta^{(2)}(\omega), \end{aligned}$$

where $d\zeta^{(1)}(\omega)$ and $d\zeta^{(2)}(\omega)$ are orthogonal increment processes with $\mathbb{E}|d\zeta^{(1)}(\omega)|^2 = f^{(1)}(\omega)$, $\mathbb{E}|d\zeta^{(2)}(\omega)|^2 = f^{(2)}(\omega)$ and $\mathbb{E}(d\zeta^{(1)}(\omega)\overline{d\zeta^{(2)}(\omega)}) = f^{(1,2)}(\omega)$. Following von Sachs et al. (1997)

$$\begin{aligned} \lim_{T \rightarrow \infty} \mathbb{E}(I_{j,k,T}^{(1,2)}) &= \lim_{T \rightarrow \infty} \mathbb{E} \left(\overline{\sum_{t=1}^T X_t^{(1)} \psi_{j,k}(t) \sum_{t=1}^T X_t^{(2)} \psi_{j,k}(t)} \right), \\ &= \lim_{T \rightarrow \infty} \mathbb{E} \left(\overline{\sum_{t=1}^T \int \exp(i\omega t) d\zeta^{(1)}(\omega) \psi_{j,k}(t) \sum_{t=1}^T \int \exp(i\omega t) d\zeta^{(2)}(\omega) \psi_{j,k}(t)} \right), \\ &= \lim_{T \rightarrow \infty} \mathbb{E} \int \left(\sum_{t=1}^T \psi_{j,k}(t) \exp(i\omega t) \right) d\zeta^{(1)}(\omega) \overline{\int \left(\sum_{t=1}^T \psi_{j,k}(t) \exp(i\omega t) \right) d\zeta^{(2)}(\omega)}, \\ &= \lim_{T \rightarrow \infty} \mathbb{E} \int_{\omega} \int_{\omega'} (\omega) \left(\sum_{t=1}^T \psi_{j,k}(t) \exp(i\omega t) \right) d\zeta^{(1)} \left(\sum_{t=1}^T \psi_{j,k}(t) \exp(i\omega t) \right) d\zeta^{(2)}(\omega'), \\ &= \lim_{T \rightarrow \infty} \int \left| \sum_{t=1}^T \psi_{j,k}(t) \exp(i\omega t) \right|^2 f^{(1,2)}(\omega) d\omega, \\ &= \int f^{(1,2)}(\omega) \widehat{\Psi}_j(\omega) d\omega. \end{aligned}$$

□

It follows that for stationary processes the wavelet cross-spectrum, $C_j(z)$, can be thought of as the integral of the Fourier cross-spectrum over dyadic intervals; $C_j(z) = \int_{\pi/2^{-j}}^{\pi/2^{-j-1}} f^{(1,2)}(\omega) d\omega$. Furthermore, the LSW coherence can therefore be thought of as the integral of the Fourier coherence over dyadic intervals:

$$\rho_j(z) = \int_{\pi/2^{-j}}^{\pi/2^{-j-1}} K^{(1,2)}(\omega) d\omega. \quad (3.14)$$

Note that in the above working we have shown the results for Shannon wavelets. The expectation of the wavelet periodogram (and cross-periodogram) is dependent on the quantity $\hat{\Psi}_j(\omega)$, the squared Fourier domain expression of the inverse DWT operator. Different wavelets cover slightly different frequency ranges (Nason and von Sachs, 1999). For the Shannon wavelet the support of the dyadic-rescales does not overlap in the frequency domain. For other wavelets this is not the case, however multiplication by the inverse inner product matrix corrects for this leakage.

3.5 Phase estimation

Estimation of the phase between the series is another important consideration and this subject is discussed in further detail in Chapter 6, in the context of irregularly sampled observations. Phase estimation is possible in the special case of equispaced time series through the use of complex-valued wavelets (see e.g. Lawton (1993), Lina and Mayrand (1995), Kingsbury (2001)). However as noted in Section 2.3.4 the LSW cross-covariance is maximised at $\tau = 0$, and so it is required in the model that there is no time delay between the signals. Phase estimation is therefore not appropriate in the bivariate LSW framework and the model requires further modification to provide a suitable estimator of phase.

Although the bivariate LSW framework is not appropriate for phase estimation, it is still possible to produce phase estimates using the raw wavelet periodograms, without correcting for the bias implied by the LSW formulation. While we do not consider this subject in detail, this section provides an introduction to the subject, and a preview to the work of Chapter 6. The empirical nondecimated wavelet coefficients are constructed as before (see Section 3.3), but using complex wavelets. Since the periodogram ordinates are now complex-valued this requires a slightly different definition of the wavelet periodogram and cross-periodogram. Following the classical Fourier approach

described in Section 2.3.2 the wavelet periodogram and cross-periodogram, derived using discrete complex wavelets, are defined as follows.

Definition 3.5.1. The wavelet periodogram for the LSW process $X_{t,T}^{(i)}$, constructed using a complex wavelet system, is given by $I_{j,t,T}^{(i)} = |d_{j,t,T}^{(i)}|^2$.

Definition 3.5.2. The wavelet cross-periodogram for the LSW processes $X_{t,T}^{(1)}$, $X_{t,T}^{(2)}$, constructed using a complex wavelet system, is given by $I_{j,t,T}^{(1,2)} = d_{j,t,T}^{(1)} \overline{d_{j,t,T}^{(2)}}$.

The wavelet periodograms and cross-periodogram can then be used to estimate phase and coherence following the classical Fourier approach described in Section 2.3.1. The quantities are first smoothed over time to give $\tilde{I}_j^{(i)}(t/T)$ for $i = 1, 2$ and $\tilde{I}_j^{(1,2)}(t/T)$. The cross periodogram is now a complex quantity and can be written in terms of its real and imaginary parts:

$$\tilde{I}_j^{(1,2)}(t/T) = c_j(t/T) - iq_j(t/T).$$

Together these quantities can be used to estimate the complex wavelet coherence and phase:

$$\rho_j(t/T) = \frac{\sqrt{c_j(t/T)^2 + q_j(t/T)^2}}{\sqrt{\tilde{I}_j^{(1)}(t/T)\tilde{I}_j^{(2)}(t/T)}}, \quad \phi_j(t/T) = \tan^{-1} \left(\frac{-q_j(t/T)}{c_j(t/T)} \right).$$

The complex wavelet coherence provides a measure of the dependence between the series and satisfies $0 \leq \rho_j(t/T) \leq 1$, while the phase provides an indication of any time lag between the series. Measures of coherence and phase are also considered in Chapter 6, where we look at the dependence between two signals observed on irregular sampling grids.

Chapter 4

Locally stationary wavelet coherence: application

In the previous chapter it was seen that our proposed estimator of the locally stationary wavelet coherence provides an asymptotically unbiased and consistent measure of the dependence between two locally stationary time series. This chapter considers the practical application of the proposed coherence estimator. Several issues that arise in the estimation procedure are considered, and the methodology is demonstrated by application to a simulated example and experimental data from neuroscience.

4.1 Practical considerations

4.1.1 Stability of the estimator

When estimating the LSW coherence using equation (3.6), values of $\widehat{S}_i^{(1)}(t/T)$ and $\widehat{S}_i^{(2)}(t/T)$ close to or below zero can lead to instabilities in the estimator. By definition the wavelet spectra are strictly positive quantities as they are the squared amplitudes. However because of the correction step in the estimation procedure it is possible that the estimates of the evolutionary wavelet auto-spectra may take values below zero. In this section we discuss several

variants on the correction procedure, and the resulting effects on the coherence estimate. The following techniques for estimating the LSW coherence are considered:

1. Not correcting. Under some circumstances it may not be necessary to correct at all and missing out the correction step of the estimation procedure avoids introducing stability issues. The uncorrected wavelet coherence is given by

$$\rho_j^{UC}(z) = \frac{\sum_{i=-\infty}^{-1} W_i^{(1)}(t/T)W_i^{(2)}(t/T)\rho_i(t/T)A_{i,j}}{\sqrt{\sum_{i=-\infty}^{-1} S_i^{(1)}(t/T)A_{i,j}}\sqrt{\sum_{i=-\infty}^{-1} S_i^{(2)}(t/T)A_{i,j}}}.$$

There are two occasions when this estimator will be unbiased; i) when the matrix A_J is diagonal (as is the case with the Shannon wavelet) and ii) if the evolutionary wavelet spectra and cross spectrum are not dependent on scale. Generally A_J will not be diagonal, however for wavelets with long length of support the bias may be small and so the uncorrected measure may give a good estimate. Also, we cannot generally assume that the spectra and coherence will be constant over scale. Although these two assumptions may be unrealistic, it can be seen that under certain circumstances the uncorrected measure may give a good approximation to the true coherence.

2. Correcting. An unbiased estimate of the LSW coherence is provided by forming the measure using the corrected estimates of the cross spectrum and EWS as in equation (3.6). However as previously noted, this estimator may become unstable.

3. Regularisation. One way to ensure stability in the corrected estimator is to bound the spectra so that they can not take negative values. This bound can be applied to the individual spectra by taking

$$\hat{S}_j^R(z) = \max(\hat{S}_j(z), \epsilon_R) \quad (4.1)$$

or alternatively we could bound the product $\hat{S}_j^{(1)}(z)\hat{S}_j^{(2)}(z)$. The major issue here is with the choice of bound. The bound, ϵ_R , should be large enough that the resulting coherence estimate is less than one, but as small as possible to reduce the bias.

An improvement to this method is to estimate the autospectra using either non negative least squares (Lawson and Hanson, 1974) or by formulation as a local complimentary problem (Murty, 1988). Using these methods it is possible to correct the estimates of the spectra with the constraints that they must be greater than zero (or our chosen bound). This provides a better fit than the simple bounding method but the same problem is still encountered, in that it is necessary to set a lower limit for the values that the spectra can take.

Estimating the spectra using regularisation gives an asymptotically biased estimate:

$$\hat{S}_j^R(z) = \hat{S}_j(z) + R$$

where $R = 0$ if $\hat{S}_j(z) \geq \epsilon_R$ and $R = \epsilon_R - \hat{S}_j(z)$ otherwise.

4. Truncation of the correction matrix Another possible solution that has been considered is to truncate the correction matrix, A_J^{-1} , at zero to give $(A_J^{-1})^+$, thus ensuring that the resulting estimates will be positive. Instead of estimating via $\hat{S}_l^{(i)}(t/T)$, we instead use $\tilde{S}_l^{(i)}(t/T) = \sum_{j=-J^*}^{-1} \tilde{I}_{j,t,T}^{(i)} (A_{l,j}^{-1})^+$. The expectation of $\tilde{S}_l(t/T)$ is given by

$$\begin{aligned}
 \mathbb{E}\tilde{S}_l(t/T) &= \sum_{j=-J^*}^{-1} \mathbb{E}\tilde{I}_{j,t,T}^{(i)}(A_{l,j}^{-1})^+ \\
 &= \sum_{j=-J^*}^{-1} \left(\sum_{i=-\infty}^{-1} S_i(t/T)A_{ij} + (2^{-j} + M)O(T^{-1}) \right) (A_{l,j}^{-1})^+ \\
 &= \sum_{j=-J^*}^{-1} \sum_{i=-\infty}^{-1} S_i(t/T)A_{ij}(A_{l,j}^{-1})^+ + (2^{-j} + M)O(T^{-1}) \\
 &= \sum_{j=-J^*}^{-1} (A_{l,j}^{-1} + P_{l,j}) \sum_{i=-\infty}^{-1} S_i(t/T)A_{ij} + (2^{-j} + M)O(T^{-1}) \\
 &= \sum_{j=-J^*}^{-1} A_{l,j}^{-1} \sum_{i=-\infty}^{-1} S_i(t/T)A_{ij} + \sum_{j=-J^*}^{-1} P_{l,j} \sum_{i=-\infty}^{-1} S_i(t/T)A_{ij} + (2^{-j} + M)O(T^{-1}) \\
 &= S_l(t/T) + \sum_{j=-J^*}^{-1} P_{l,j} \sum_{i=-\infty}^{-1} S_i(t/T)A_{ij} + (2^{-j} + M)O(T^{-1})
 \end{aligned}$$

by rewriting $(A_{i,j}^{-1})^+$ as $(A_{i,j}^{-1})^+ = A_{i,j}^{-1} + P_{l,j}$. The bias incurred from correcting by $(A_J^{-1})^+$ instead of A_J^{-1} is therefore given by $\sum_{j=-J^*}^{-1} P_{l,j} \sum_{i=-\infty}^{-1} S_i A_{ij}$ which is a strictly positive quantity. This method is therefore similar to regularisation in that we are adding a correction factor to the original spectral estimate. The difference is that we do not need to choose the bound as the value of $P_{l,j}$, and hence the bias, is determined by the choice of wavelet. This method is therefore more straightforward to implement than method 3, but the major drawback is that it tends to result in a larger bias. This is because in many cases the resulting (unmodified) spectral estimates will be strictly positive, leading to an unbiased and stable estimate of the LSW coherence, yet we are still applying this modified correction procedure and introducing a needless source of bias. This is not the case with the regularisation procedure, which modifies the resulting spectral estimates only if they fall below the set bound.

For wavelets such as Daubechies with long lengths of support, the number of negative values is low and so the bias incurred by this alteration is minimal. However, particularly for wavelets with short length of support (such as Haar),

the resulting estimate substantially over-estimates the spectrum. The method may be useful when using wavelets with longer length of support, however in this case it may be just as effective to estimate the coherence using the uncorrected quantity.

5. Smoothing over scale. It is also possible to stabilise the estimator by smoothing over scale. Performing this extra smoothing reduces the variance of the estimator further and, by suitable choice of smoothing parameters, can ensure that the spectra do not take negative values. Though this method introduces bias, the bias is minimal, particularly at fine scales where not much smoothing is necessary to ensure stability. It is necessary to choose the amount of smoothing but this is less of an issue than the problem with choice of bound as it cannot force the coherence to take certain values.

We smooth over scale allowing a different vector of weights for each scale. This produces a $J^* \times J^*$ matrix of weights, denoted D , where the element $D_{l,j}$ is the contribution of scale j to the smoothed estimate of scale l , and the rows thus sum to 1. Correcting for the bias by multiplying by A_j^{-1} has the effect of applying a scale factor of about 2^j to scale j , so that fine scales are scaled up in comparison to coarse scales.

Since we choose to smooth over scale after correcting, we smooth using the following scheme: a) first the estimates are re-scaled by 2^{-j} , b) the estimates are smoothed over scale, c) the scaling factor of 2^j is re-introduced. The corrected, scale smoothed estimate is therefore given by:

$$\tilde{I}_{l,t,T}^{(1,2)} = \left(\sum_{j=-\infty}^{-1} 2^{-j} D_{l,j} \tilde{I}_{j,t,T}^{(1,2)} \right) 2^l.$$

In the examples below we smooth over the neighbouring 5 scales. Different choices of smoothing weights are allowed for the estimates of each scale. For fine scales it is often not necessary to smooth over scale at all so that at fine scales of l we may have $D_{l,j} = \delta_{j,l}$.

The matrix, D , is described by choosing the weights on the leading diagonal, and then the off-diagonal elements are chosen as decaying proportions of $1 - D_{l,l}$ with modifications for end values as follows:

$$\mathbf{D} = \begin{pmatrix} D_{1,1} & \frac{2(1-D_{2,2})}{6} & \frac{1(1-D_{3,3})}{6} & 0 & \dots \\ \frac{4(1-D_{1,1})}{6} & D_{2,2} & \frac{2(1-D_{3,3})}{6} & \frac{1(1-D_{4,4})}{6} & \\ \frac{2(1-D_{1,1})}{6} & \frac{2(1-D_{2,2})}{6} & D_{3,3} & \frac{2(1-D_{4,4})}{6} & \\ 0 & \frac{1(1-D_{2,2})}{6} & \frac{2(1-D_{3,3})}{6} & D_{4,4} & \\ \vdots & & & & \ddots \end{pmatrix}.$$

In order to compare the different approaches to stabilising the estimator, we apply the methods to time series simulated according to a known Fourier coherence. Since the series are stationary, we smooth over the full length of the series and hence obtain one value at every wavelet scale. The Fourier coherence is assumed to be piecewise constant (see Figure 4.1). This is so that the coherence varies over frequency but is approximately constant within each wavelet scale. The Fourier autospectra are assumed to be white noise processes i.e. the spectra are constant over frequency. The associated wavelet coherence has an erratic form which makes this coherence structure a good test data set for the different estimators, since any leakage across the scales will be clearly visible.

The corresponding LSW coherence for the different estimators, based on 100 simulations, is plotted in Figure 4.2. For these estimates we have used Daubechies extremal phase wavelets with $N = 6$. In each plot the grey lines show the individual sample paths, while the coloured lines represent the mean values over the 100 simulations. The ‘true’ coherence structure, as estimated from the Fourier coherence using the approximation $\int_{\pi/2^{-j}}^{\pi/2^{-j-1}} K_{12}(\omega)$, is plotted with the dashed line to act as a guide for the accuracy of the estimators.

Although stable, the uncorrected LSW coherence estimator (method 1) clearly suffers from leakage across scales and the resulting coherence appears

smoothed. The corrected measure (method 2) remedies this problem but it becomes unstable at coarse scales. The regularisation and truncation of the correction matrix methods (3 and 4) both ensure that the resulting corrected estimator is stable, but produce biased estimates of the coherence. For the scale-smoothed measure (method 5) the estimates are smoothed over scale using weights of $D_j = (0.9, 0.9, 0.85, 0.85, 0.80, 0.80, 0.75, 0.75, 0.70, 0.60, 0.55)$ on the leading diagonal. The amount of smoothing was chosen to ensure stable estimates at all scales with as little bias as possible. The scale-smoothed measure is both accurate and stable.

Therefore, we recommend overcoming the problems with stability of the estimator by smoothing the cross-periodogram over scale as well as time. Smoothing over scale reduces the variance of the estimator further, producing stable estimates of the LSW coherence. Though this additional smoothing introduces extra bias into the estimator, the bias can be minimised by suitable choice of the smoothing parameters. The weights on the leading diagonal are chosen to be as close to 1 as possible, while maintaining stability.

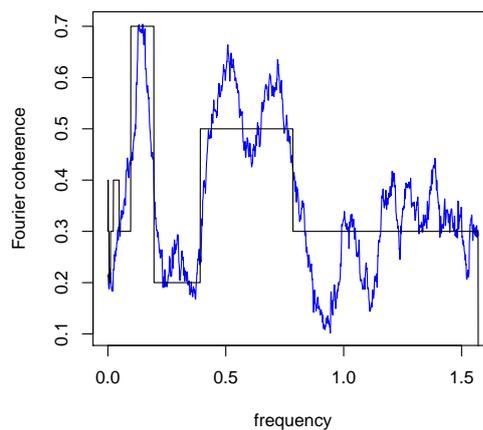


Figure 4.1: Piecewise constant Fourier coherence. The true coherence structure is shown in black and the estimated coherence for one realisation is shown in blue.

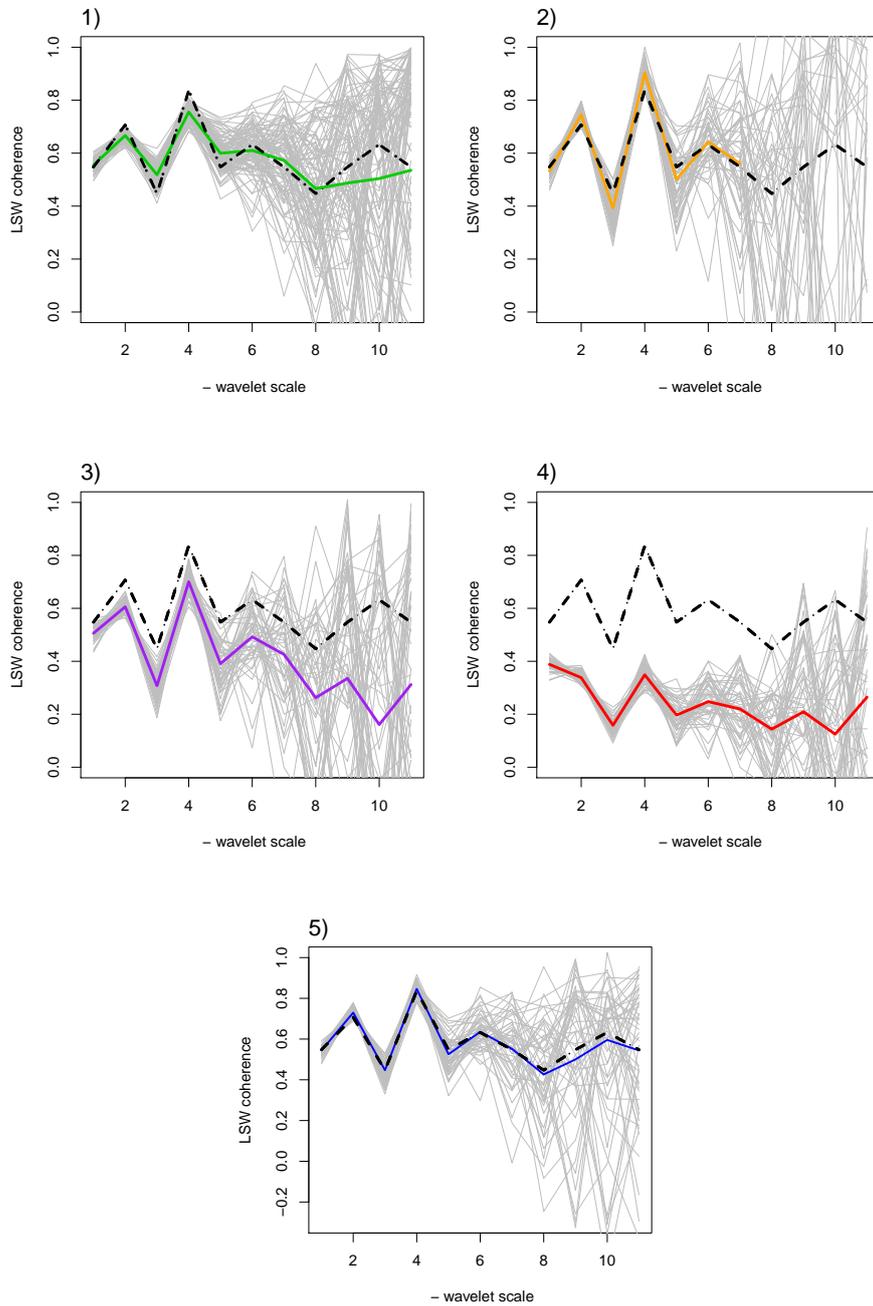


Figure 4.2: Comparison of stabilisation methods: 1) uncorrected 2) corrected 3) corrected with regularisation 4) corrected with truncated correction matrix 5) corrected and scale-smoothed. The true coherence structure is shown by the dashed line.

4.1.2 The effect of J^*

It is also necessary to select a value of $J^* = \alpha \log_2(T)$, the number of scales over which the periodogram is corrected. In order to ensure consistency of the estimator, α is chosen to be in the interval $(0, 1)$. A value of α close to 1 ensures a decomposition over the largest possible number of scales; however in practice not all of these scales will be informative and the proportion of informative scales will depend on the data and the choice of wavelet. As a default, we propose taking the proportion $\alpha = 0.7$ which provides accurate estimates across different simulated examples. To illustrate this, we consider an estimation experiment on simulated data. The coherence is estimated for various values of J^* , demonstrating the effect on the observed coherence. Ideally J^* should be as large as possible to give a decomposition over as many scales as is feasible, but making sure that this results in informative estimates of the coherence.

We simulate bivariate stationary time series of length 2^{10} with a known Fourier coherence. For each pair of time series the LSW coherence is estimated by correcting over $J^* = (10, 9, 8, 7, 6)$ scales, i.e. $\alpha = (1, 0.9, 0.8, 0.7, 0.6)$. Since the time series are stationary, the series are smoothed over time by averaging over the entire length of the series and hence the results are not affected by the choice of time smoothing parameter. Smoothing over scale is conducted with the weights $D_{l,l} = (0.9, 0.85, 0.85, 0.8, 0.75, 0.7, 0.6, 0.55, 0.45, 0.45)$ on the leading scales. This choice of scale-smoothing parameters was selected to ensure the stability of the estimates, while introducing a minimal amount of bias. The estimated LSW coherence for each choice of J^* is shown in Figure 4.3. Here the length of the estimate indicates the choice of J^* . The ‘true’ LSW coherence, corresponding to the integral of the underlying Fourier coherence over dyadic intervals, is plotted with the dashed line. From this we estimate that informative estimates of the coherence are produced for choices of J^* up to 7 (giving $\alpha = 0.7$).

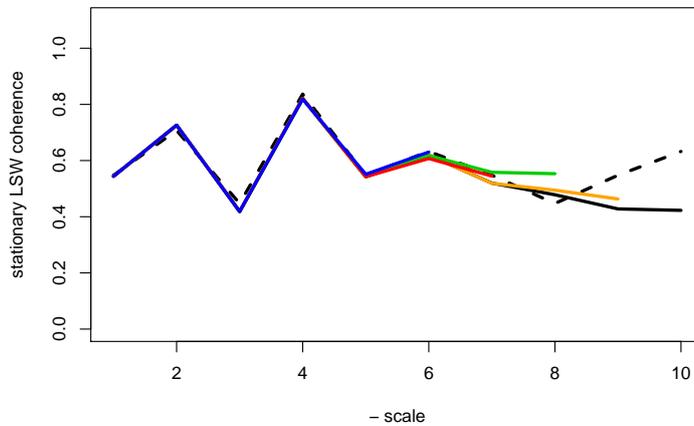


Figure 4.3: Estimated stationary LSW coherence for different choices of J^* . Here we plot the mean over 200 simulations and the choice of J^* is indicated by the length of the corresponding estimate. The underlying coherence structure is indicated by the dashed line.

4.1.3 The effect of smoothing over time

Smoothing over time is conducted using a simple moving average kernel smoother, described in equation (3.5), and it is necessary to choose the width of the smoothing window, M . We choose to increase the width of the smoothing window as the scale becomes coarser and the width of the smoothing window at scale j is denoted M_j . This is natural since the wavelet coefficients at coarse scales display a longer correlation length.

To justify this choice, we consider the case of smoothing the auto-periodogram of a white noise process. In this case, the variance of the periodogram is $\text{Var}(I_{j,k,T}) \approx 2$, using the result from Proposition 2.3.7 and since for white noise $\mathbb{E}(S_j(z)) = 2^j$. Figure 4.4 shows the effect of smoothing on the variance of the wavelet auto-periodogram of a white noise process. The plot is based on 1000 simulations of a white noise process. The variance of the raw periodogram is approximately 2 at all scales as expected, and the variance then reduces as the length of the smoothing window is increased. The plotted lines represent window lengths of $M = 0, 10, 25, 100$ and 200, with the final line

representing the variance after smoothing over the whole length of the series. For a fixed smoothing window over all scales, the variance of the smoothed periodogram estimates would increase as the scale becomes coarser. This is because at coarser scales the length of support of the wavelet is larger, and averaging over a fixed window length involves less independent information. It therefore makes sense that when smoothing over time, the width of the smoothing window should increase as the scale becomes coarser.

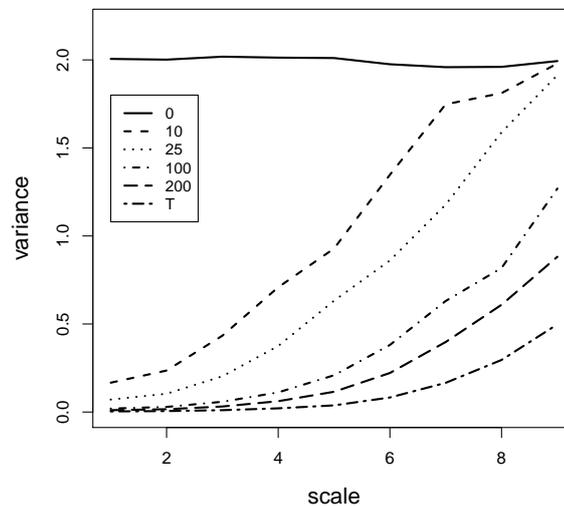


Figure 4.4: Effect of the width of the smoothing window on the variance of the periodogram of a white noise process. Choices of $M = 0, 10, 25, 100, 200$ and T are considered.

4.1.4 Other issues

Another issue that will affect the estimates is the choice of wavelet. The wavelet can be chosen according to the properties of the time series that it is representing. For example if the series contains sharp jumps then the Haar wavelet might be preferable, whereas for series with smoother features other wavelets, such as Daubechies least asymmetric (Daubechies, 1992), might be preferable.

4.2 Applications

4.2.1 Simulated example

The first example applies our method to data simulated from a bivariate locally stationary wavelet process with a known, nonstationary coherence structure that varies between scales. For even scales we take $\rho_j(k/T) = 0.5$, and for odd scales we assume a non-stationary structure that forms an ‘inverted v’ between 0.2 and 0.8. The EWS of each process are taken to be white noise so that $S_j^{(i)}(k/T) = 2^j$. For this example the series is of length $T = 2^{13}$, and we correct over $J^* = 9$ ($\alpha = 9/13$) scales.

The results from one such simulation, using Daubechies least asymmetric wavelet with $N = 5$ vanishing moments, are shown in Figure 4.6. The estimates are smoothed over scale using weights on the leading scales of $D_{l,l} = (0.95, 0.95, 0.95, 0.9, 0.9, 0.9)$ for $l = 1$ to 6. As previously, this choice of scale-smoothing parameters was selected to ensure stability of the estimates, while introducing a minimal amount of bias. For smoothing over time, we use a bandwidth of $M_j/T = (0.025, 0.05, 0.075, 0.1, 0.125, 0.15)$, so that the bandwidth increases as the scale becomes coarser. The estimated coherence structure from one realisation of the process is shown in Figure 4.5. The estimated coherence follows the true coherence closely.

4.2.2 Application to neuroscience data

The method is now demonstrated with application to experimental neuroscience data, previously described by Jones and Wilson (2005). We consider the coherence between local field potentials (LFP) in two functionally and anatomically connected areas of a rat’s brain: the hippocampus and the prefrontal cortex. The LFP provide a measure of averaged activity over local neuronal populations, and the estimated LSW coherence presents an indication of the extent to which activities in the two areas are coordinated, decomposed

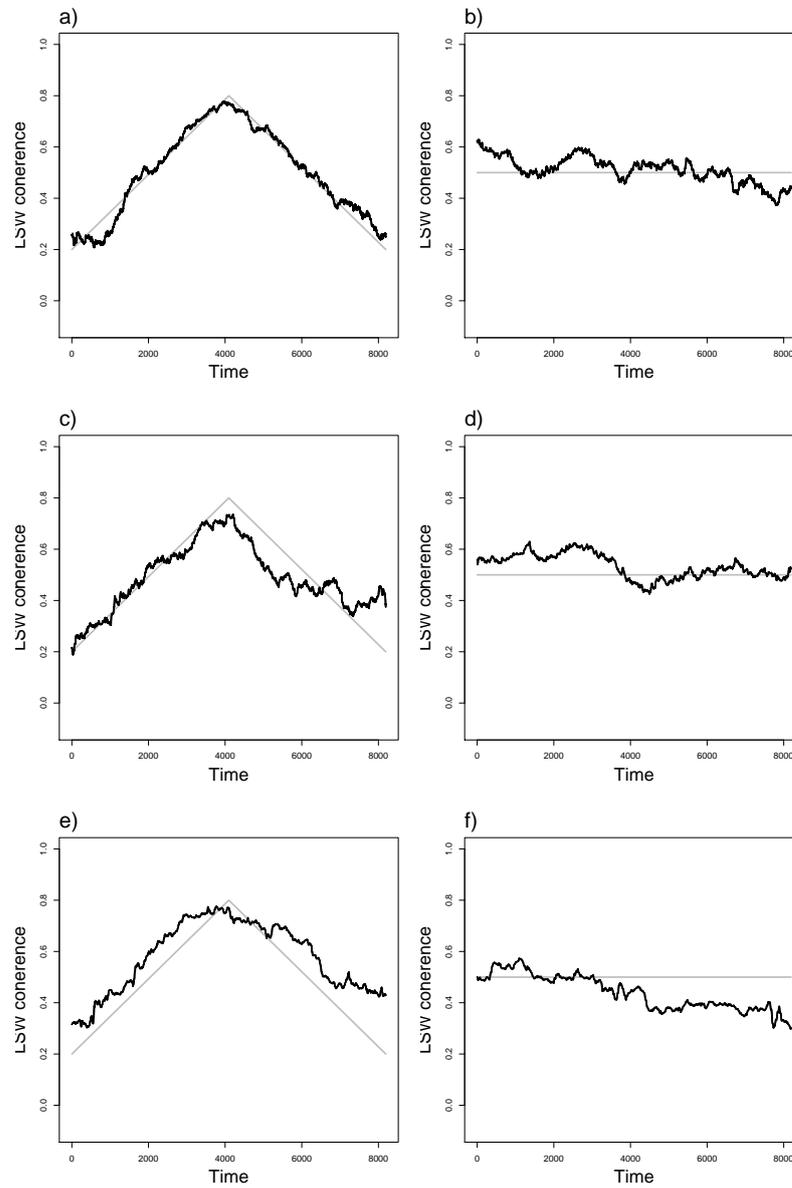


Figure 4.5: Results showing the true coherence in grey, and estimated coherence in black from the simulation example. Subplots a) to f) correspond to scales -1 to -6 respectively.

over scale and over time.

The measurements are taken while the rat is performing a maze-based task, designed to invoke spatial working memory and decision-making. The complete experimental procedure can be decomposed into smaller sections of 13 trials, with each trial comprising a ‘forced turn’ epoch and a ‘choice’ epoch. Each of these epochs are 6s in duration, corresponding to the section of the task when the rat is moving along the central arm of the maze, immediately prior to reaching a T-junction turning point. During forced turn runs the direction at the T-junction is pre-determined by a movable barrier, whereas in choice epochs the rat is free to choose either direction. It is expected that hippocampal-prefrontal interactions are selectively recruited during active decision-making on choice runs, and that these interactions will be reflected by increased coherence relative to forced turn epochs, presenting an indication of the extent to which activities in these connected brain regions are coordinated. In order to eliminate other sources of variation, trials in which the rat made the incorrect decision at the turning point were not included in the analysis.

During neurophysiological recordings, the local field potential data were recorded at a sampling frequency of 625Hz. As part of the preparatory data analysis the data was downsampled by taking every third measurement, giving a sampling frequency sufficient to estimate coherence at less than 100 Hz. In order to provide a comparison between our results and Fourier approaches, we use the approximation given in Section 3.4 that a wavelet scale j corresponds to Fourier frequencies in the interval $(2^{j-1}(\Delta t)^{-1}, 2^j(\Delta t)^{-1})$. In Jones and Wilson (2005), significant coherence was found in the theta frequency range of 4–12Hz. Using this approximation, the downsampling ensures that the theta frequency band is contained within wavelet scales -4 and -5 . We smooth over all 3 subsamples to ensure that information is not lost. Other amounts of downsampling were also investigated to ensure that features were not missed, but this extra analysis offered no additional insight.

Since choice runs invoke spatial working memory, we would expect to see

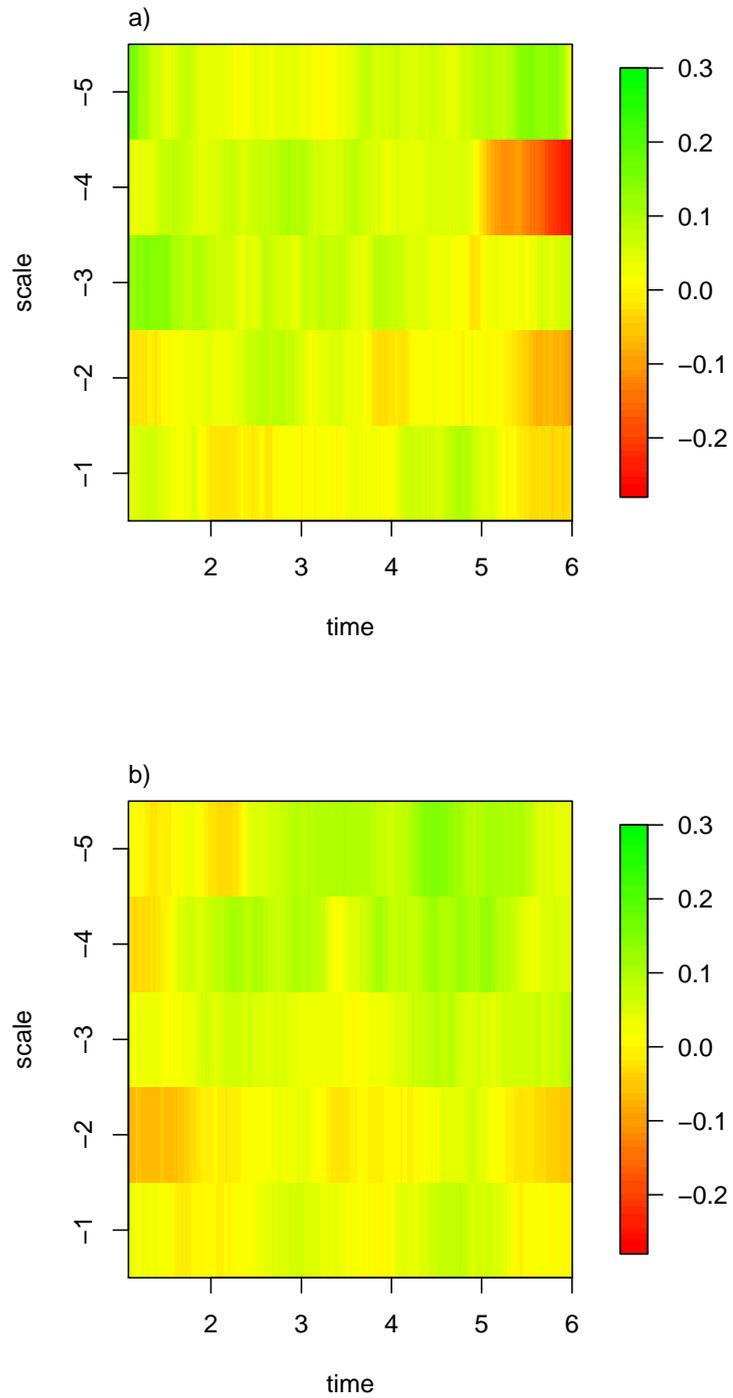


Figure 4.6: Trial averages for choice runs (a) choice forced-turn runs (b) scales -1 to -5 using Daubechies extremal phase wavelet with $N = 8$ vanishing moments.

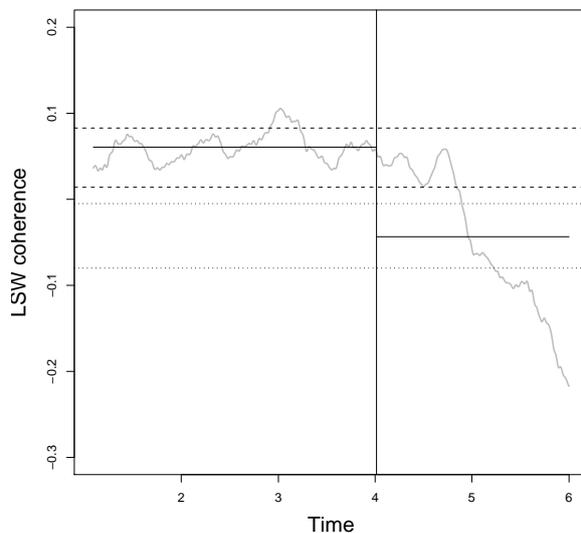


Figure 4.7: LSW coherence for scale -4 with means before and after the reference point plotted in bold. The associated confidence intervals are plotted with dashed lines.

a change in correlation structure when the rat approaches the turning point. For forced runs we would not expect to observe this change. The locally stationary wavelet coherence averaged over all 13 trials is shown in Figure 4.6. We emphasise that the locally stationary wavelet coherence is not only applicable in repeated measurements situations but works in the context of single measurements too, as illustrated in our simulation study of Section 5.1. In the present case study, the variability between trials is high and so the trial average provides a clearer comparison between groups. The presented results have been produced using Daubechies extremal phase wavelet, $N=8$, scale-smoothing using $D_{l,t} = (0.95, 0.9, 0.9, 0.9, 0.85)$ on the leading terms and smoothing over time using a bandwidth of $M_j/T = (0.05, 0.1, 0.15, 0.2, 0.25)$. Figure 4.6 shows an area of increased coherence towards the end of the series at scale -4 for the choice run data. This feature is not observed in the forced run data.

After 4s of the trial, the rat has travelled approximately $3/4$ of the way down the central arm of the maze. This is a good reference point as the rat's

movement at this stage of the task is consistent between trials. The actual time point at which the rat turns at the junction varies between trials, but occurred at 5.4s on average. From Figure 4.7, we see that at approximately 5s, after the reference point but before the estimated turning point, there is a marked change in the coherence structure; we judge the significance of this change by looking at the mean coherence before and after the 4s reference point. Confidence intervals are constructed by simulation: given the estimated wavelet coherence and autospectra, we simulate n new series. For each series we take the same estimator, the mean before and after the reference point, and then take the $n\alpha$ and $n(1 - \alpha)$ ordered values to give the lower and upper confidence limits. The estimate and resulting 90% confidence intervals for scale -4 of the choice data is shown in Figure 4.7. The mean wavelet coherence is significantly different between the two sections of the data, suggesting that there is a significant change in interaction between the hippocampus and prefrontal cortex as the rat approaches the turning point. This effect was only observed at scale -4 on the choice run data sets.

Figure 4.8 illustrates the original findings of Jones and Wilson (2005) using the Fourier transform. Note that these results represent the coherence for the full set of trials (17 in total), including those in which the rat made the wrong decision at the turning point. The plot shows the trial-averaged coherence for the entire length of the central-arm of the task, split into forced-turn (grey) and choice (red) trials. The coherence shown in red can therefore be compared to LSW coherence results shown in Figure 4.6a, while the coherence plotted in grey can be compared to Figure 4.6b. Similarly to the findings presented in this thesis, the Fourier transform method shows significant coherence only in the theta-frequency range, and only during choice epochs. However unlike the LSW coherence method, the results using the Fourier transform provide no information on the time-dependence of the results. Figure 4.9 shows the nonstationary coherence (for the 13 trials in which the rat made the correct decision at the turning point) as estimated by the short time Fourier transform.

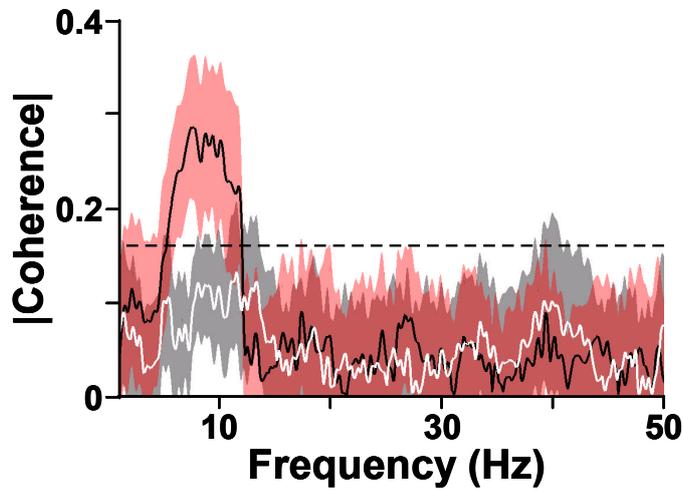


Figure 4.8: Trial-averaged central-arm coherence, reproduced from Jones and Wilson (2005, Figure 7b). Central-arm coherence is subdivided into forced-turn (grey) and choice (red) directions. Dashed line marks 95 % confidence level, with shaded band thickness corresponding to jackknife error bars (estimated over trials and nine tapers).

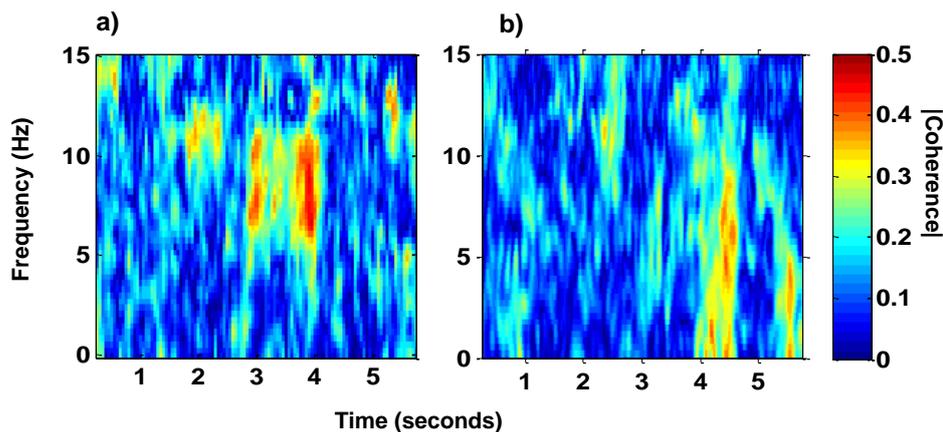


Figure 4.9: Trial-averaged coherence for the choice trials (a) and forced-turn trial (b) using the short time Fourier transform with a window width of 0.5 seconds, run at 0.005 second increments.

An area of increased coherence can be seen in the theta frequency range at a time of around three to four seconds, indicating a slight difference in the exact timing of the coherence change than that provided by the LSW coherence estimate.

4.3 Summary and further work

In Chapter 3 we introduced the bivariate LSW process model, providing an extension to the univariate time series model of Nason et al. (2000). Through this formulation, we proposed a novel measure of linear dependence, termed LSW coherence. An important difference between the LSW coherence and previous wavelet coherence measures lies in the particular bias correction implied by the LSW model. Use of our methodology was demonstrated with respect to a simulated example, as well as experimental data from neuroscience.

Our analysis of the neuroscience example corroborates the previous results from Jones and Wilson (2005) using the Fourier transform, indicating that the LSW coherence methodology provides a suitable tool for the analysis of these data sets. The real advantage of this new method over the global Fourier transform stems from the natural time localisation of wavelets, providing extra information on the time dependence of the coherence. The short time Fourier transform can be used to extend the Fourier methods to nonstationary settings by applying the Fourier transform to a localised time window that slides along the time axis; however the window width is constant over all frequencies. It is desirable to select a small time window for accurate localisation at high frequencies, but a large time window is needed to capture low frequency information. This trade-off becomes particularly hard to meet in the case of long time series that cover a large frequency range. In the example of Section 4.2.2 the data has been segmented into short trials, but the original data recordings are much longer, in some cases spanning as long as 24 hours. Application of the short time Fourier transform, and in particular choice of the window width,

would be problematic in these scenarios.

LSW coherence provides a representation of the dependence structure with less redundancy than similar methods that use the continuous wavelet transform. Methods using complex continuous wavelets do, however, have the advantage that this allows for an estimation of the phase between the signals. Extension of our method to allow for phase estimation would be very interesting. This is possible in the model through the use of discrete complex wavelets, as described in Section 3.5.

Another important consideration is the application to multivariate data. Existing methods of assessing multivariate dependence include those of Dahlhaus (2000), who extended univariate locally stationary processes to a multivariate setting and Ombao et al. (2005) which is an extension to the univariate application of smooth localized complex exponentials. The application of our method to multivariate cases is possible using the current methodology by taking pair-wise combinations of multivariate series. Generalisation of the bivariate LSW model to a fully multivariate model is also possible. In extending the current formulation to a multivariate model, inference is aided by the fact that the individual spectrum of each series is dependent only on the properties of the given series, and not on the other signals of the multivariate process. This property is illustrated by the fact the individual EWS of each process are estimated similarly to the univariate case of Nason et al. (2000). This is not the case for the bivariate Cramér representation; through the transfer function formulation, the individual spectra of each series are dependent on the other signals of the process (see Ombao et al. (2005, Section 3)). Therefore the complexity of the multivariate Cramér model increases quickly as the number of series increases.

Chapter 5

Periodogram estimation for irregular situations using wavelet lifting

As seen in the previous chapters, wavelets provide a useful tool for the representation and visualisation of nonstationary time series. Through the use of the NDWT, a sampled signal was decomposed into a set of wavelet coefficients which were then used to estimate the evolutionary periodogram and spectrum. Although classical wavelet techniques have proved useful in many settings, there are also limitations to these methods. For example the classical DWT and NDWT rely on the observations having a regular sampling rate. Time series with irregular spacing can occur in a variety of different settings. It may be that a time series is regularly spaced, but contains missing data due to, for example, machine failure or censoring. This type of irregularly spaced data is often termed ‘gappy’ (Mondal and Percival, 2008). Alternatively, the sampling grid may be intrinsically irregular. In climate research, for example, irregular spacing of observations is a common problem. As discussed further in Section 5.7, when analysing ice-core data, equidistant sampling along the depth of the core leads to a coarsening of time resolution with increasing age

(Witt and Schumann, 2005). The irregular nature of palaeoclimate time series motivates the work of this chapter, and we present our results on palaeoclimate signals in Section 5.7.

In the majority of applications, irregularly spaced time series are processed prior to analysis, restoring the data to a regular grid which is then suitable for analysis by standard spectral methods. A commonly used method for this purpose is interpolation, often using splines (Yiou et al., 1996).

Resampling techniques such as ‘sample and hold’ and ‘nearest neighbour’ are also popular (Broerson, 2008). Using resampling methods, the data is transformed to a regular grid using the irregularly sampled measurements close to the imposed grid structure. With sample and hold the underlying function is assumed to have a piecewise constant structure that is defined by the irregularly spaced observations. Nearest neighbour resampling simply replaces the unobserved regularly sampled points with the nearest neighbouring observation. With both of these resampling methods, the multiple use of single observations is a common problem. This can be avoided by using the slotting principle; an observation is only used if it is within a certain distance from the resampling grid point. Using slotting, the irregularly sampled data is therefore transformed to a regular grid with missing observations. Although it is convenient to work within a regularly spaced time series setting, a typical result of interpolation will be smoothing of the signal, leading to the loss of information at high frequencies (Frick et al., 1998).

Ideally we would like to avoid restoring the data to a regular grid to ensure that the analysis is not dependent on the chosen pre-processing method. There are several methods available that make direct use of the irregularly spaced observations. A popular method for stationary series is the Lomb-Scargle periodogram (see Lomb (1976) and Scargle (1982)) which is equivalent to least-square fitting of sinusoids to the data. If the data consists of pure Gaussian noise then the obtained Lomb-Scargle periodogram follows an exponential distribution, a result which can be used to test the significance of

detected events. The Lomb-Scargle periodogram and other methods for spectral analysis of irregularly sampled stationary time series are reviewed and compared in Stoica and Sandgren (2006). Mondal and Percival (2008) propose unbiased estimators of the wavelet variance for gappy data, but this relies on the data irregularities being of the missing data type. It is also assumed that the series has stationary d 'th order differences.

Mondal and Percival (2008) state that their method could also be extended to nonstationary situations by using similar techniques to those of Nason et al. (2000). Other methods for nonstationary series include the weighted wavelet Z-transform (WWZ) of Foster (1996) which has been applied to the analysis of palaeoclimate records (see e.g. Witt and Schumann (2005)). The application of wavelet analysis to irregular time series has also been considered by Frick et al. (1998) who introduce the use of 'gapped wavelets'. Another relatively new technique is that of Qi et al. (2002). The data is modelled using a basis of sines and cosines with time varying amplitudes and the coefficients are estimated using a Bayesian framework.

This chapter considers the use of spectral analysis based on one coefficient at a time wavelet lifting (Knight et al., 2010) to analyse irregularly sampled palaeoclimate signals. Sections 5.1 and 5.2 review the work of Knight et al. (2010) in constructing the nondecimated lifting periodogram. Section 5.3 explores this method further, examining the relationship between wavelet lifting scale and Fourier frequency. This is crucial when considering the application to palaeoclimate series which are more commonly described in terms of Fourier periods of oscillation rather than wavelet scale. We then introduce the empirical mode decomposition approach as an interesting comparison to the wavelet lifting techniques. Both methods are demonstrated on palaeoclimate time series in Section 5.7. Having highlighted the advantages and disadvantages of each method, the different approaches are compared in Section 5.8.

We begin by describing the nondecimated lifting transform before detailing its use in forming the nondecimated lifting periodogram for irregular data.

5.1 The nondecimated lifting transform

As explained in Section 2.2.2, the one coefficient at a time (OCAAT) lifting scheme can be used to decompose a signal into a lower resolution signal and a set of wavelet coefficients. Recall that the NDWT produces wavelet coefficients, $d_{j,k}$, at each dyadic scale, $j \in \{-1, \dots, -J\}$, for each time location, $k \in \{0, \dots, T - 1\}$. In contrast, the one coefficient at a time lifting scheme produces one wavelet coefficient, d_{j_r} , at each observation point, x_{j_r} , associated with a particular scale, α_{j_r} i.e. $x_{j_r} \rightarrow (d_{j_r}, \alpha_{j_r})$. Using the matrix representation of the OCAAT lifting transform, the coefficients have the form

$$\begin{pmatrix} d_{x_1} \\ \vdots \\ d_{x_n} \end{pmatrix} = R \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}, \quad (5.1)$$

where R is an $n \times n$ matrix built by the OCAAT lifting scheme. From this it follows that $d_{x_i} = \sum_{j=1}^n r_{i,j} f_j$ and each detail is a linear combination of the observed data.

To be able to construct a full spectral representation of a process, where for each time point we have an estimate at each (artificial) scale, Knight and Nason (2008) introduce a *nondecimated lifting transform* (NLT). The key to this method is in noting that the OCAAT lifting scheme produces a decomposition which is highly dependent on the removal order of points. Instead of considering just one order for removal of points, the nondecimated lifting transform considers a set of P randomly chosen removal orders. In this way, each x_{j_r} becomes associated with a set of tuples for (d_{j_r}, α_{j_r}) , one tuple for each removal order. Following Knight and Nason (2008) we now briefly summarise the NLT.

We start with a *trajectory*, $T = (x_{o_1}, \dots, x_{o_n})$ where (o_1, \dots, o_n) is a permutation of the set of point indices $\{1, \dots, n\}$. Thus the trajectory, T , gives the observation points in their order of removal. At each stage of the algorithm,

the data is lifted one coefficient at a time according to Jansen et al. (2001), but with the modification that the point to be removed is selected following the predefined removal trajectory, T . The first point to be lifted is x_{o_1} , the second x_{o_2} and so on. This is different to the standard OCAAT lifting scheme, which suggests that the selection of the point to be removed is based on a chosen criterion, such as the length of integral of the scaling function (Jansen et al., 2001). If the primary resolution level is set to be L , the points to be removed are x_{o_1} to $x_{o_{n-L}}$, with the last L entries of T denoting the scaling coefficients.

The nondecimated lifting transform of Knight and Nason (2008) consists of repeated applications of this modified algorithm, using different random trajectories, T_p for $p \in 1, \dots, P$. This provides P sets of wavelet coefficients, each described by a different matrix R^1, \dots, R^P . For each removal location, x_k , where $k = 1, \dots, T$, the NLT provides a set of details $\{d_{x_k}^p\}_{p=1, \dots, P}$ where $d_{x_k}^p$ is the coefficient at location x_k , obtained from trajectory T_p . For a given point, x_k , its removal rank will vary throughout the set of trajectories. If a point is removed early on in the algorithm, the resulting wavelet coefficient is likely to be associated with a small integral value and therefore a fine scale. Conversely, if a point is removed later in the algorithm it is likely to be associated with a larger integral value and hence a coarser scale. This is demonstrated in Figure 5.1. The wavelet coefficients at each time location will therefore be associated with different scales, where scale is now a continuous measure as defined in Section 2.2.2. This is different to the classical NDWT which produces one wavelet coefficient at each location and dyadic scale. If the series is of length n , then the number of possible trajectories is given by $n!$. In practice just a sub-sample of these trajectories is used for implementation. The number of trajectories should be ‘large enough’ to ensure that an ample number of coefficients is produced at all scales and locations, subject to computational constraints.

The NLT has been successfully applied to problems in nonparametric re-

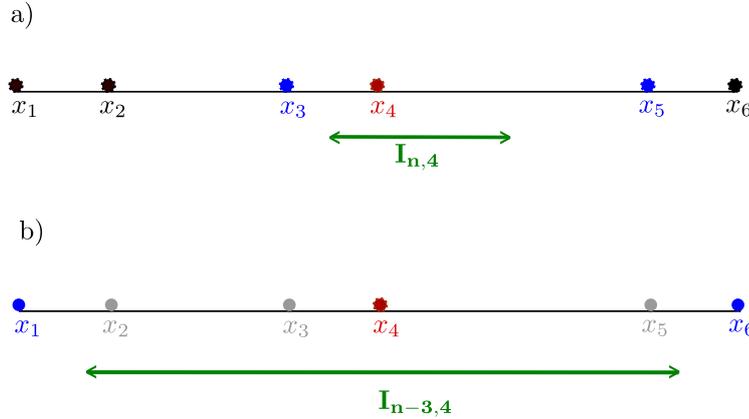


Figure 5.1: Relationship between wavelet lifting scale and removal order for the example illustrated in Figure 2.5. When the point x_4 is removed at the first stage of the algorithm it is associated with a small integral as shown in subplot a. When the point is removed after points x_2, x_3 and x_5 , it is associated with a larger integral, as shown in subplot b.

gression (Knight and Nason, 2008) and can also be extended for use in spectral estimation. Following Knight et al. (2010) we now describe how the NLT can be used to define a periodogram for irregularly spaced data.

5.2 The nondecimated lifting periodogram

Recall that the NDWT periodogram for a series of length T is a $T \times J$ array, containing the squared detail coefficients associated with locations $0, \dots, (T-1)$ at scales $-1, \dots, -J$ where $J = \log_2(T)$. Knight and Nason (2008) define a similar array for data with irregular observations. Since in a wavelet lifting setting the scale associated with each wavelet coefficient is a continuous quantity (see Section 2.2.2), the scale is first partitioned into a set of discrete intervals, or *artificial scales*, according to $\{l^i\}_{i \in \{1, 2, \dots, J^*\}}$, where J^* is chosen to provide the desired resolution level.

For each time location, x_k , and each artificial scale, l^i , the NDWT provides a set of squared wavelet detail coefficients $\{(d_{x_k}^p)^2 : \alpha_{x_k}^p \in l^i\}$. This is the set of all squared wavelet coefficients associated with time x_k that have integral length in the set l^i . In order to form the periodogram we would like to use

this set of wavelet coefficients to give a *unique* value. Knight et al. (2010) uses these sets of detail coefficients to estimate one single value, $\hat{f}_{x_k}(l^i)$, using a nonparametric regression approach. The squared details $(d_{x_k}^1)^2, \dots, (d_{x_k}^p)^2$ are modelled as a function of the corresponding scale $\alpha_{x_k}^1, \dots, \alpha_{x_k}^p$ according to

$$(d_{x_k}^p)^2 = f_{x_k}(\alpha_{x_k}^p) + \epsilon_p, \quad p \in 1, \dots, P. \quad (5.2)$$

For each time location x_k , f_{x_k} is the function of interest and in practice, we are interested in its values at the set of discrete intervals $\{l^i\}_{i \in \{1, 2, \dots, J^*\}}$. Using a linear smoother provides the estimate

$$\hat{f}_{x_k}(l^i) = \sum_{p=1}^P K_p(l^i)(d_{x_k}^p)^2, \quad \forall i \in \{1, 2, \dots, J^*\}, \quad (5.3)$$

where $K_p(l^i)$ are weight functions that are non-zero only for those p values such that $l_{x_k}^p$ is in the neighbourhood of l^i and the weights K_p are different for each x_k . The value $\hat{f}_{x_k}(l^i)$ provides an estimate of the squared details at time t_k and associated with interval length l^i . The matrix $\{\hat{f}_{x_k}(l^i)\}_{i=1 \dots J^*, k=1 \dots n}$ forms the NLT periodogram. Note that this method does not ensure that the periodogram estimates are positive, although it is possible to do so by modelling the log squared details and then exponentiating.

Alternatively, as we use in our implementation, it is also possible to estimate the periodogram ordinates by taking the average of the wavelet coefficients within each scale. Estimating the wavelet coefficients this way leads to the following definition of the NLT periodogram.

Definition 5.2.1. For each time point, x_k , $k \in 1, \dots, n$ and each artificial scale l^i for $i \in 1, \dots, J^*$, the wavelet lifting periodogram is defined as

$$I_{x_k}(l^i) = \text{mean}\{(d_{x_k}^p)^2 : \alpha_{x_k}^p \in l^i\}.$$

Taking the average of the coefficients rather than using a nonparamet-

ric regression approach ensures that the resulting periodogram estimates are not affected by the choice of regression method. The averaging method has further benefits over nonparametric regression techniques when, in later chapters, we consider the estimation of the coherence between two series. Therefore throughout the rest of this thesis we form the periodogram as stated in Definition 5.2.1. However, note that under different circumstances it may be preferable to form the periodogram using nonparametric regression.

Smoothing over time

Before plotting the NLT periodograms, we first smooth the estimates over time. In our implementation we use simple moving average smoothing, paralleling the regularly sampled case described in Section 3.3. The smoothed estimate is given by

$$\tilde{I}_{x_k}(l^i) = \frac{1}{\#(M_j)} \sum_{j \in M_j} I_{x_j}(l^i), \quad (5.4)$$

where $M_j = \{j : x_k - M < x_j \leq x_k + M\}$ and M denotes the width of the averaging window. Smoothing over time serves to reduce the variance, providing a clearer visualisation of the characteristics of the data. As in the regularly sampled case, it might also be preferable to modify this formulation so that the width of the time window is a function of scale. In the current methodology we use the same width of smoothing window for all scales but further research in this area would be advisable.

Missing values

Also note that, unlike the NDWT periodogram, it is possible to have missing values in the resulting NLT periodogram. Sometimes missing values can be filled by taking a larger sample of trajectories, but often the missing values occur as a natural result of the sampling structure. This is evident in the plots of Figure 5.3 by the missing values at small scales. The scale information

revealed in the periodogram depends on the sampling rate which in our chosen application is dependent on time. We will therefore observe higher frequency information at some points in time than at others. Although it is possible to use interpolation to provide periodogram estimates at the missing ordinates, we choose not to, using uncoloured areas to emphasise the information that cannot be provided by the series.

Example of periodogram construction

We now provide an example of the construction of the NLT periodogram for a simulated series. The series consists of 500 values, observed on an irregular sampling grid in which the sampling rate varies between 1 and 3. The signal consists of a noisy sine wave with period 25 and a sine wave with period 80 which is just present for the middle section of the data. The wavelet coefficients are calculated using a random sample of 5000 trajectories. Each coefficient is associated with a scale, and the range of observed scales is dependent on the sampling of the data and the length of the time series. Since the finest sampling difference observed in the data is 1, the lowest scale coefficients estimated from the data will have $\alpha_{x_k}^p = 1$. The largest scale coefficient that is estimated from the data depends on the randomly selected trajectories but will not exceed the length of the time series. In practice it is more convenient to work with $\log_2(\alpha_{x_k}^p)$ and the finest and coarsest scales may appear patchy so in this example we restrict the \log_2 -scale range to 2-7. Within this scale range we have discretised to form 20 artificial scales, $\{l^i\}_{i \in \{1, 2, \dots, 20\}}$.

The construction of the periodogram ordinates for one time point, x_{200} , is demonstrated in Figure 5.3 using the moving average method and in Figure 5.3b using spline smoothing. These plots correspond to a vertical slice through the full periodogram estimates. The full periodograms smoothed over time using a smoothing window of width 30 are plotted in 5.3c and 5.3d. The variations in the signal are identified clearly in the periodogram estimates. Estimation of the lifting periodograms was conducted using the Adlift package

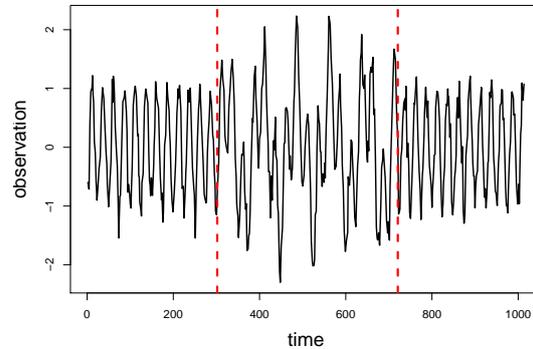


Figure 5.2: Simulated time series with irregular spacing. The mean sampling rate is 2 and the series contains oscillations with periods 25 and 80. The middle section between the vertical lines contains oscillations at both frequencies.

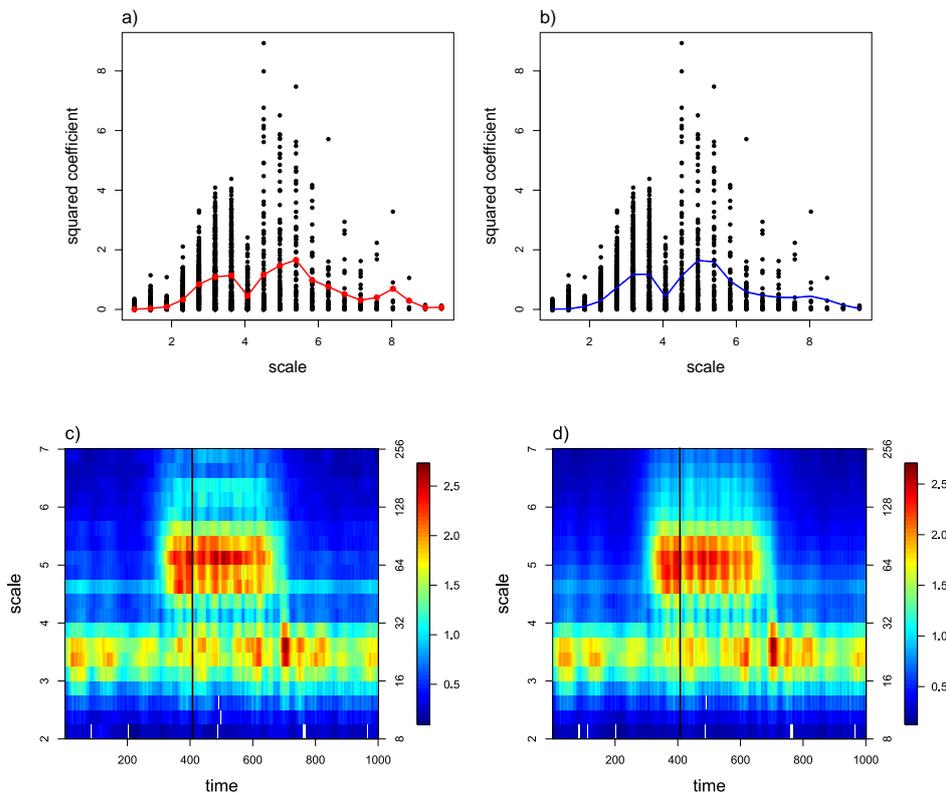


Figure 5.3: Estimating the periodogram ordinates for a given time point, x_{200} , using a) the average in each group b) nonparametric regression with splines. Estimation of the periodogram using c) moving average d) splines. All estimates are smoothed over time using a smoothing window of 30.

(Nunes and Knight, 2008) and the NLT package (Knight, 2008).

5.3 Relationship to Fourier frequency

Section 5.7 uses the lifting periodogram methodology to examine time series with physically meaningful variations in frequency. It is therefore important to understand the physical meaning of artificial scale as defined by wavelet lifting. This section contributes to the existing methodology by presenting new work on the relationship between wavelet lifting scale and Fourier frequency.

As previously described, the scale of each wavelet coefficient obtained from the one coefficient at a time lifting scheme is given by the integral of the scaling function at the last stage before the point is removed from the algorithm. At level n of the lifting transform, the integral of point j_n is given by $\int_{\mathbb{R}} \phi_{n,j_n}(x) dx$. This accounts for the length of the sampling interval between points. For example, at the first stage of the algorithm the scaling functions are given by $\phi_{n,k}(x_i) = \delta_{i,k}$ and the intervals associated with each grid point are constructed using the midpoints between successive grid points. The integral of point x_2 will therefore be $\frac{x_3 - x_1}{2}$.

The initial definition of the integrals of each coefficient remains unchanged no matter which prediction scheme is used in the lifting algorithm. However it is not necessarily the case that different lifting schemes will provide the same relationship between scale and frequency. For example, consider the case of linear prediction with either 1 or 2 neighbours either side. Removal of any given point using 2 neighbours either side will provide larger scale information than removing the point using one neighbour either side; this is because we are considering a larger overall interval of the data. The wavelet lifting scale therefore does not have a consistent relationship to frequency across different lifting schemes. To provide a detailed description of the relationship between wavelet scale and Fourier frequency, we start by considering the case of periodogram estimation using linear prediction with one neighbour either side.

Relationship between integral and scale: 1 neighbour

To determine the relationship between OCAAT lifting scale and Fourier frequency, we consider estimating the wavelet coefficients of a pure sinusoid using linear prediction with one neighbour either side. Assuming that the two neighbours are equidistant from the point that is being removed, the wavelet coefficient produced on removal of the point indexed j_n is given by

$$\begin{aligned} d_{j_n} &= c_{n,j_n} - \sum_{i \in J_n} a_i^n c_{n,i}, \\ &= \sin\left(\frac{2\pi x_{j_n}}{P}\right) - \frac{1}{2} \sin\left(\frac{2\pi(x_{j_n} + I_{n-1,j_n})}{P}\right) - \frac{1}{2} \sin\left(\frac{2\pi(x_{j_n} - I_{n-1,j_n})}{P}\right), \end{aligned}$$

where P is the period of the wave and the point j_n has coordinates (x_{j_n}, c_{n,j_n}) , and integral I_{n-1,j_n} . Note that since we assume that the two neighbours are both the same distance from the removal point, the prediction weights are given by $a_i^n = 1/2$ for $i \in J_n$, the set of neighbours.

For a signal of period P , the resulting detail coefficient, d_{j_n} , depends on the location, x_{j_n} , and the value of the integral, $I_{n-1,j_n} = \alpha_{j_n}$, and so we can write

$$|d_{j_n}| = d(x_{j_n}, \alpha_{j_n}) \quad \text{for} \quad j_n \in D_n. \quad (5.5)$$

where D_n is the set of indices of wavelet coefficients at scale n . In what follows we consider the general case, rather than specific removal locations x_{j_n} and so the subscripts in equation (5.5) are dropped, to give $d(x, \alpha)$. Taking the arithmetic average over x for each value of α we obtain

$$\bar{d}(\alpha) \quad \text{for} \quad \alpha \in \mathbb{R}^+. \quad (5.6)$$

The relationship between wavelet lifting scale and Fourier frequency can be described using the value of α that maximises this quantity. The value of α that maximises equation (5.6) corresponds to the scale of the periodogram at which we see a peak in the wavelet coefficients for a signal of period P ,

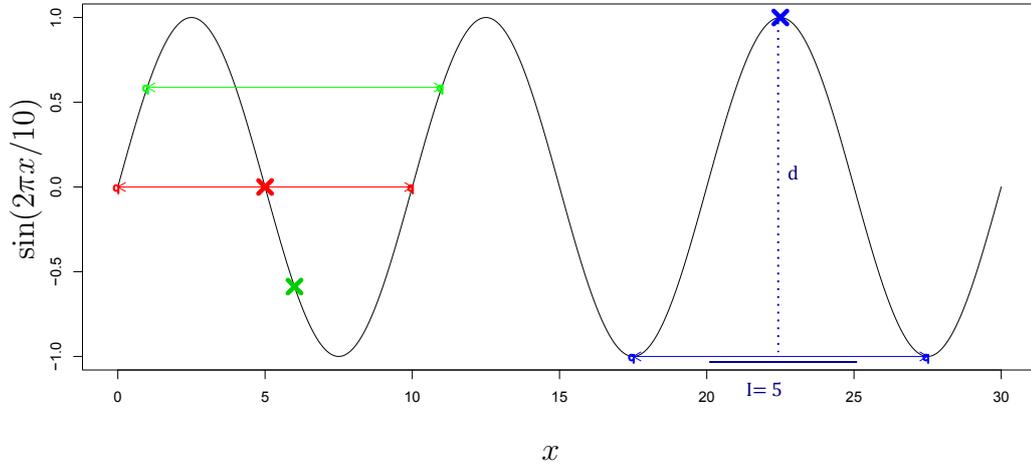


Figure 5.4: Relationship between the integral and associated detail coefficients using one neighbour either side linear prediction and a sine wave of period 10. Removing the point marked with the blue cross it is clear that the largest coefficient will be obtained by considering an integral of 5.

hence providing a relationship between wavelet scale and Fourier frequency. Predicting using linear prediction with one neighbour either side provides the relationship

$$\{\alpha : \bar{d}(\alpha) \geq \bar{d}(\beta), \beta \in \mathbb{R}^+\} = \frac{P}{2} \pm nP, \quad (5.7)$$

where the nP term occurs as a consequence of aliasing. This relationship can be verified intuitively by inspection of Figure 5.4, which clearly shows that the largest wavelet coefficients are obtained by considering an integral of $P/2$.

Effect of uneven sampling: 1 neighbour

We now consider relaxing the assumption that the two neighbours are equidistant from the removal point. Even if the test data set is regularly sampled, the equidistant sampling will not be preserved after the first point has been removed. Consider the case where the distances of each neighbour from the removal point, j_n , are given by d and $\beta_{n,j_n}d$. Here β_{n,j_n} is the scaling of the shortest to longest side so that $\beta_{n,j_n} \geq 1$. For schemes using linear prediction with one neighbour either side the quantities β_{n,j_n} and d are related to the

integral of the removal point through the relationship $d = 2I_{n,j_n}/(1 + \beta_{n,j_n})$. This measure is also related to the regression weights, $a_i^n, i \in J_n$, and hence the form of the resulting wavelet (see Figure 5.5). Note that for any given point the associated ‘measure of asymmetry’, β_{n,j_n} , will vary throughout the algorithm depending on the spacing of the remaining points, which in turn depends on the trajectory.

Figure 5.6 shows the average value of the wavelet coefficients against scale for different amounts of asymmetry, β . As the measure of asymmetry increases, the coefficients are more likely to peak at scales other than $P/2$. Therefore the effectiveness of α as a measure of scale is dependent on the sampling of the data. For values of $\beta \geq 2$ we often see the peak of coefficients at a scale of $\alpha = P$ rather than $\alpha = P/2$ as with equispaced points. This results in a ‘blurring over scale’ in the resulting spectrum.

In Figure 5.7 we demonstrate this effect using the simulated example from Section 5.2. Here we plot the complete lifting spectrum (Figure 5.7a) and the lifting spectrum decomposed according to the degree of asymmetry of the coefficients. The left hand axis of the plot indicates the (log) wavelet lifting scale and the right hand axis provides a guide for the equivalent Fourier period as implied by equation (5.7). For this plot the detail coefficients that are used to form the periodogram are split into three equal sized groups, according to their measure of asymmetry as defined by β . We therefore form three different periodograms which are constructed from detail coefficients with varying measures of asymmetry. Figure 5.7b contains the ‘least asymmetric’ third of coefficients, Figure 5.7d contains the ‘most asymmetric’ third of coefficients and 5.7c shows the remainder. The plot of all coefficients (a) can be found by combining plots b, c and d. Comparing plots 5.7b and 5.7d, we see that 5.7b provides a much clearer description of the series, eliminating the artifacts at higher scales. Overall we see that the detail coefficients that have asymmetrically placed neighbours often peak at a different scale and result in a blurring of the periodogram.

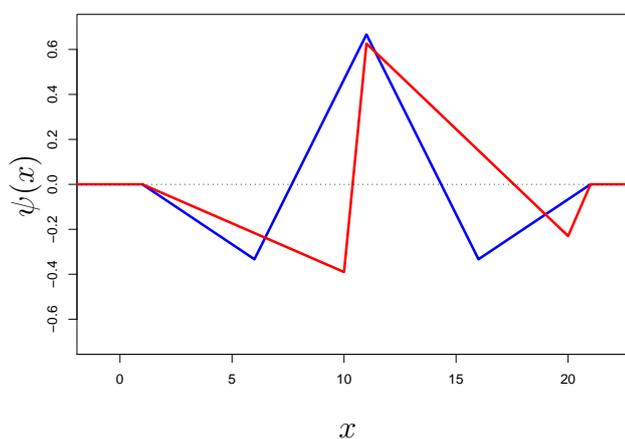


Figure 5.5: Effect of uneven sampling on the form of the resulting wavelet. Here demonstrated for the removal of a point with an integral of 5. The resulting wavelet for $\beta = 1$ is shown in blue and for $\beta = 10$ in red.

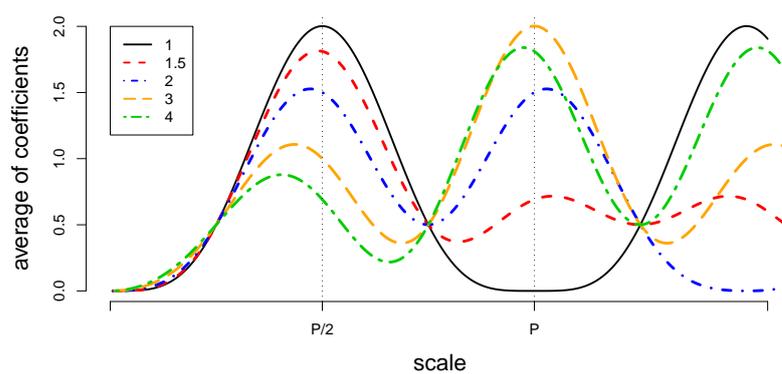


Figure 5.6: Average value of the squared detail coefficients, $\bar{d}(\alpha)^2$, against scale, α , for a sinusoid with period P . The different lines represent an increasing degree of asymmetry in the sampling: $\beta = 1, 1.5, 2, 3, 4$.

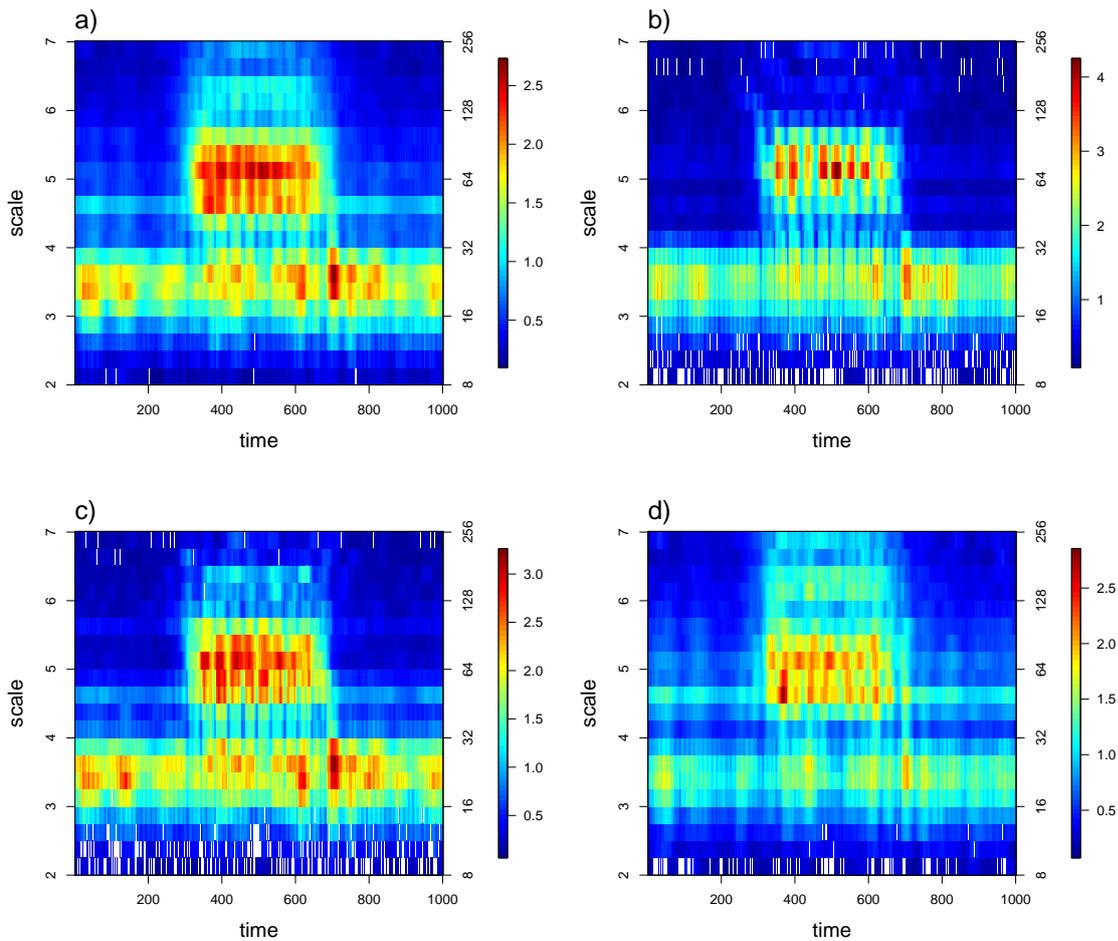


Figure 5.7: Lifting periodogram of the simulated time series illustrated in Figure 5.2, decomposed according to the degree of asymmetry of the neighbours (β) involved with production of each coefficient: a) all detail coefficients b) detail coefficients that satisfy $\beta \leq 1.45$ c) coefficients that satisfy $1.45 < \beta \leq 2.45$ and d) coefficients that satisfy $\beta > 2.4$.

Relationship between integral and scale: more than one neighbour

If more than one neighbour is used in the linear prediction scheme, the maximum absolute value of the coefficients is no longer achieved at a scale of $P/2$. Predicting using two neighbours either side of the removal points gives a relationship to Fourier frequency of $\{\alpha : \bar{d}_{N_2}(\alpha) \geq \bar{d}_{N_2}(\beta), \beta \in \mathbb{R}^+\} = 0.29P$ and for three neighbours either side we have $\{\alpha : \bar{d}_{N_3}(\alpha) \geq \bar{d}_{N_3}(\beta), \beta \in \mathbb{R}^+\} = 0.20P$. Figure 5.8 shows the average value of the squared coefficients according to integral for 1, 2 and 3 neighbours either side of the prediction point and assuming a regular spacing of points.

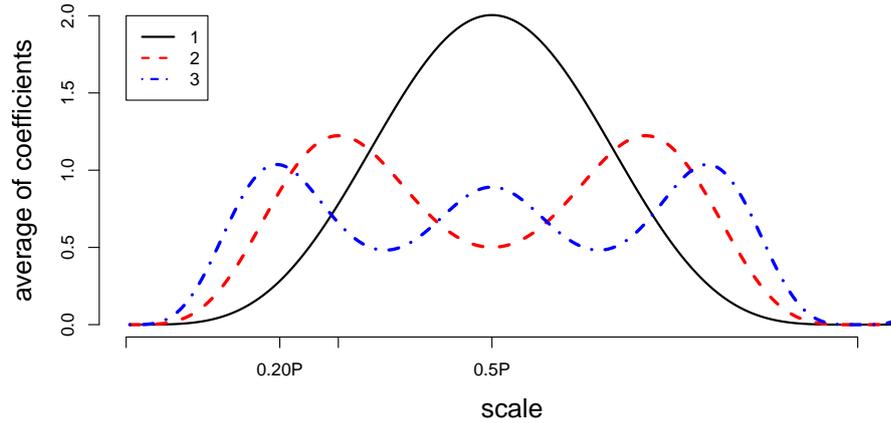


Figure 5.8: Average value of the squared detail coefficients, $\bar{d}(\alpha)^2$, against scale, α , obtained with a linear prediction scheme for different sizes of neighbourhood.

Choice of prediction scheme

It is possible to choose the prediction scheme according to the properties of the data or the aims of data analysis. For our analysis, we concentrate on one neighbour either side linear prediction. The use of linear prediction ensures that the prediction weights are positive. As described in Section 2.2.2 this ensures that scale is a monotonically increasing function of the index and hence an intuitive measure of scale. The inclusion of negative weights is problematic

for spectral estimation as it is possible to obtain negative scales. Choosing to use just one neighbour either side of the removal point ensures a good relationship to frequency with less blurring over scale than with a prediction scheme with more neighbours. This is illustrated in Figure 5.8 by the pronounced peak of the one-neighbour-either-side scheme. It is also feasible to use adaptive procedures as in Nunes et al. (2006) but since the choice of prediction scheme affects the relationship between scale and frequency it could be problematic to mix prediction schemes when constructing the periodogram.

5.4 Unblurring the periodogram

In order to produce clear visualisations of the data, we now consider methods of dealing with the ‘blurring over scale’ that is caused by the inclusion of asymmetric prediction values in the transform. As shown in Section 5.3, when the detail coefficient is obtained using neighbours which are not equidistant from the removal point, it is possible to see a peak in the periodogram at a different scale. Note that this is different to the bias of the regular NDWT periodogram that was introduced in Section 2.3.4. A bias equivalent to the regular NDWT periodogram bias is also present in the lifting periodogram, and is discussed further in Section 5.5. Three different methods for unblurring the periodogram are considered.

1. It is possible to remove the detail coefficients associated with asymmetric prediction points by setting a cut-off value for β_{n,j_n} denoted β_{cut} . Though this results in a loss of information, it ensures that only detail coefficients with a well defined relationship to frequency are included in the periodogram.
2. Only a small set, P , of the $T!$ possible trajectories, T_p , of observation

removal orders are actually utilised in the NLT lifting scheme. It is therefore possible to choose a subsample of these trajectories which minimises the number of detail coefficients with $\beta_{n,j_n} > \beta_{\text{cut}}$. To do this, we choose the elements of the trajectory T_p as the algorithm proceeds, rather than generating a random removal order a priori. At each stage of the algorithm, remaining points are split into two groups, according to whether the value of β_{n,j_n} is above or below the cut off point, giving $P_n^1 = \{x_i : \beta_{n,i} \leq \beta_{\text{cut}}\}$ and $P_n^2 = \{x_i : \beta_{n,i} > \beta_{\text{cut}}\}$. The point to be removed, x_{o_n} , is selected by randomly sampling from P_n^1 or if $P_n^1 = \emptyset$, then a point is randomly selected from P_n^2 . This way, points with $\beta_{n,j_n} > \beta_{\text{cut}}$ can still be removed at the end of the algorithm. Though there is still some loss of information, the number of detail coefficients cut in the final stage will be reduced. The points in P_n^1 and P_n^2 will change throughout the algorithm depending on the removal order of points.

3. The assignment of β_{n,j_n} to the point x_{j_n} is based on the nearest neighbours either side of x_{j_n} . To ensure that $\beta_{n,j_n} < \beta_{\text{cut}}$ it is also possible to search combinations of different neighbours of x_{j_n} to find a pair that produces a prediction with $\beta_{n,j_n} < \beta_{\text{cut}}$. The update step of this approach is not intuitive, as it involves updating points that are not next to the removal point, and so the value of the resulting updated integrals will not intuitively represent the sampling. To overcome this problem we discard the ‘in between’ observations. This is equivalent to performing the lifting algorithm on a subsampled set of the original points. To minimise the amount of discarded information, only variations of neighbours that will result in one point being discarded will be considered, and so in some situations it is not possible to find combinations of neighbours that produce coefficients with $\beta_{n,j_n} < \beta_{\text{cut}}$. In these cases, the point is lifted as normal, using the original neighbours, and the resulting coefficients are removed at the end of the algorithm.

The three different unblurring techniques are compared using a simulated example of an irregularly sampled sine wave. The signal has a period of 20 kyr, and the sampling is taken to be the same as that of the composite carbon dioxide series used in Section 5.7, between -100 kyr and -300 kyr BP, ensuring that the example provides a realistic trial for our chosen application.

The resulting periodograms using methods 1-3 are shown in Figure 5.9a-c. Here we have estimated the periodograms using 5000 trajectories, discretised using 20 scales between 10 and 16, and smoothed over time using a window of width 10 kyr. In order to compare the efficiency of the different methods, we also record the number of detail coefficients that are computed but not used in the final periodogram estimates. The proportion of used information (after using the $\beta_{n,j_n} \leq \beta_{\text{cut}}$ cut off at the end of the algorithm) for the three different methods is 0.59, 0.78, 0.55. For the third method this figure includes points that have been discarded due to the irregular sampling. Not including these points, the proportion of used information out of all the lifted (not discarded) points is 0.72.

The second method shows a clear improvement in the proportion of information that is being used in the final periodogram estimate; however we see that the resulting periodogram has missing observations at certain time points. From Figure 5.9d we see that the gaps in the periodogram correspond to areas of the series which contained a cluster of points with a high value of β_{n,j_n} . The method is effective for series which exhibit a uniform initial distribution of points with $\beta_{n,i} > \beta_{\text{cut}}$, but if this is not the case then there may be gaps in the estimated periodogram. Method 2 is therefore not suitable for our application, but provides an efficient solution in other situations. Both the first and third methods produce clear estimates of the periodogram but, since the proportion of used information is similar, we choose to use the simpler algorithm 1.

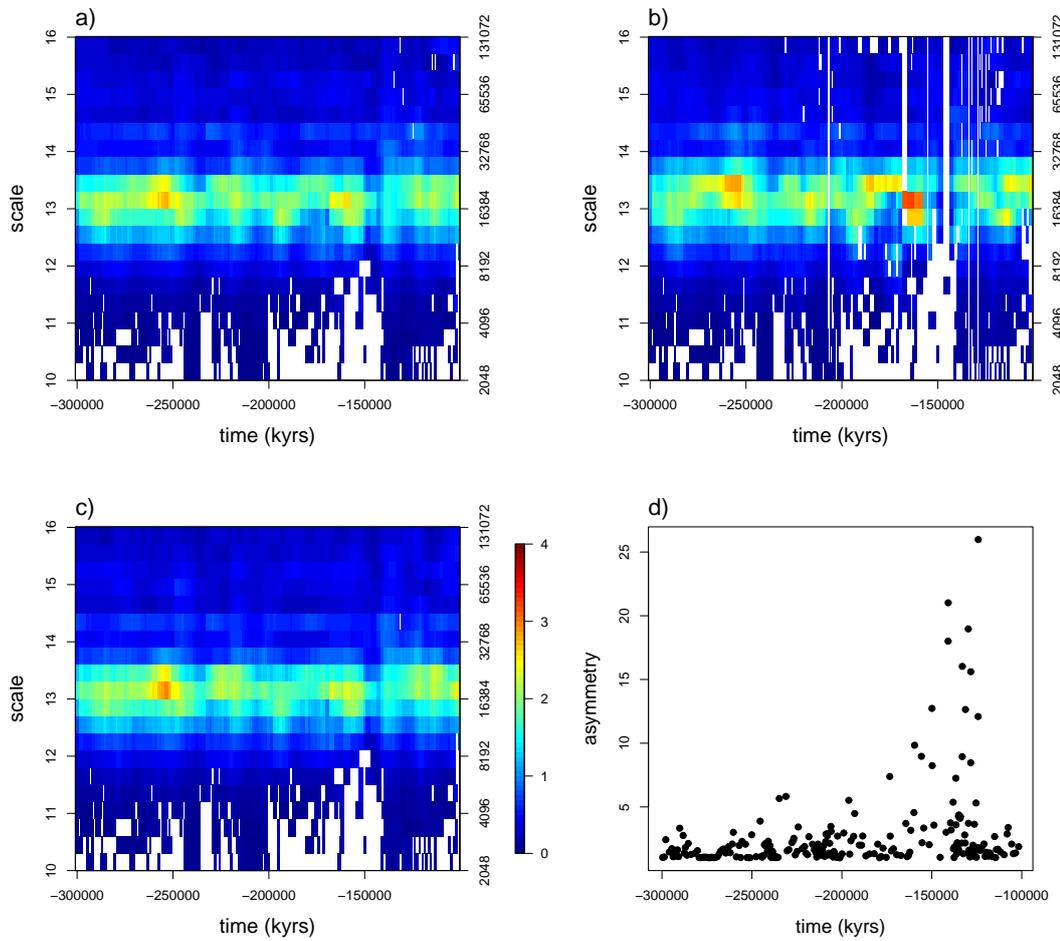


Figure 5.9: Different methods for unblurring the asymmetric periodogram: a) setting a cut-off value of $\beta_{\text{cut}} = 2.5$ b) favouring the removal of symmetric points c) searching for symmetric neighbours d) value of β for initial points.

5.5 Relationship between the NLT periodogram and the EWS

In Chapter 2 it was seen that the NDWT periodogram provides a biased estimate of the evolutionary wavelet spectrum. This bias causes a leakage across scales in the raw periodogram estimates but can be corrected by application of the inverse wavelet inner product matrix. Knight et al. (2010) propose an equivalent correction procedure for the NLT periodogram, based on the relationship between the NLT periodogram and the evolutionary wavelet spectrum.

In the formulation of Knight et al. it is assumed that the observed series can be modelled as LSW processes (as described in Definition 2.3.4) but with missing values. The appearance of each time point in the final data set is modelled as $I_t \sim \text{Bernoulli}(p)$ where $1-p$ is the probability that any given point is missing from the observed signal. Based on the LSW model formulation, Knight et al. find the expectation of the NLT periodogram in terms of the EWS:

Theorem 5.5.1 (Knight et al. (2010)). *For the wavelet periodogram estimators $\hat{f}_{x_k, T}$, and for $i \in 1, \dots, J^*$, $k \in 1, \dots, n$ we have*

$$\mathbb{E}(\hat{f}_{x_k, T}(l^i) | \text{fixed paths}) = \text{Trace}(A^{l^i, k} S^T) + O(T^{-1}),$$

where $S = (S_{l, j})_{l \leq -1, j \in 1 \dots n}$ with $S_{l, j} = S_l(x_j/T)$, $A^{l^i, k} = (a_{l \leq -1, j \in 1 \dots n}^{l^i, k})$ with $a^{l^i, k} = \sum_{j'=1}^n \{ \sum_{\alpha=1}^m K_\alpha(l^i) r_{k, j}^\alpha r_{k, j'}^\alpha \} \Psi_l(x_j - x_{j'})$, $r_{k, j}^\alpha$ are elements of R^α introduced in equation (5.1), and $\{K_\alpha(\cdot)\}_\alpha$ are as defined in equation (5.3). The expectation is conditioned on the removal trajectories being fixed.

This shows that the raw periodogram is not an unbiased estimator for the spectrum. Based on this formulation, Knight et al. also propose a correction procedure for removing the bias.

Although this methodology provides a promising framework for certain situations, the correction procedure is computationally intensive and application to the long palaeoclimate series considered in this chapter is therefore problematic. The method depends on the observed data being of the ‘regular with missing observations’ type, or at least requiring that the signal can be well approximated by this framework. An intrinsically irregular time series can be transferred to a regular grid with missing observations by imposing a regular grid structure with the sampling rate equal to the highest common denominator of the observed irregular sampling differences. However in many situations this imposed regular frame will be very sparse (in terms of the ratio of observed to missing values). Also the initial assumption that the appearance of missing values will follow a Bernoulli distribution will also not hold in many cases. In the case of most palaeoclimate series, the observed sampling rate is a function of time. The application of this correction method to the time series considered in this chapter would be an interesting advancement in future work, but at the moment we omit this extra step.

Another interesting point, is how the correction procedure of Knight et al. compares to our proposed method of removing coefficients associated with asymmetric prediction weights. The method we propose arises from the ability of the lifting integral to act as a measure of scale. In contrast, Knight et al. assume the model is constructed from nondecimated wavelets (from the LSW construction) with a known frequency relationship and so it does not consider this cause of leakage.

5.6 Empirical mode decomposition

In this section we introduce another function decomposition method: empirical mode decomposition (EMD). As we shall see in Section 5.7, EMD provides an interesting comparison to the wavelet lifting methodology emphasised in this chapter. Since we are interested in EMD from a comparative perspective, our

treatment of the technique consists of a review of the existing literature rather than contributing to any methodological advancements.

Empirical mode decomposition (EMD) was introduced by Huang et al. (1998) and provides a method of adaptive time series decomposition suitable for non-linear, non-stationary and irregularly sampled time series. EMD is used to decompose the original function into a collection of intrinsic mode functions (IMF). With the methods we have introduced so far (DFT, DWT, wavelet lifting) the signal is decomposed with respect to a set of pre-defined basis vectors. EMD differs from these methods in that the basis is dependent on the observed signal. We now introduce the EMD procedure and discuss some of the issues related to its application.

The decomposition procedure

We start by defining intrinsic mode functions (IMFs). These are extracted from the raw signal using a procedure known as *sifting* which is introduced subsequently.

Definition 5.6.1 (Huang et al. (1998)). An intrinsic mode function (IMF) is a function that satisfies two conditions:

1. In the whole data set the number of extrema and the number of zero crossings must either be equal or differ by at most one.
2. At any point the mean value of the envelope defined by the local maxima and the envelope defined by the local minima must be zero.

Each IMF represents a single oscillatory mode. An IMF is therefore comparable to a single Fourier harmonic but is more general as IMFs may contain frequency and amplitude modulations (Huang et al., 2003). Most observed time series are not IMFs, instead the data at any one time may involve more than one oscillatory mode. Motivated by this, Huang et al. (1998) introduce the EMD *sifting procedure* to decompose the observed data, $X(t)$, into a set of IMFs. The sifting process consists of the following steps:

1. The local extrema of $X(t)$ are identified. These are then used to define the upper, lower and mean envelopes as follows
 - a) The local maxima are connected to form the upper envelope $e_{\max}(t)$.
 - b) The local minima are connected to form the lower envelope $e_{\min}(t)$.
 - c) The mean of the envelopes is defined as $m_1(t)$.

2. The difference between the data and $m_1(t)$ is calculated. This forms the first component:

$$h_1(t) = X(t) - m_1(t).$$

These steps are illustrated in Figure 5.10 a and b. Ideally $h_1(t)$ would be an IMF, but in practice this is generally not the case and more than one sifting step is required to extract each mode. If $h_1(t)$ does not fulfill the IMF criteria of Definition 5.6.1 then the sifting process is repeated. Treating $h_1(t)$ as the data we obtain $h_{11}(t) = h_1(t) - m_{11}(t)$. The sifting process is repeated k times until the component $h_{1k}(t)$ given by $h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t)$ satisfies the IMF requirements. The first IMF component is therefore given by: $c_1(t) = h_{1k}(t)$. This should contain the finest scale (or shortest period) component of the signal and is separated from the original data using

$$r_1(t) = X(t) - c_1(t).$$

The residue still contains information about longer period components and so is then treated as the new data and subjected to the same sifting process. This procedure is then repeated on all subsequent residuals:

$$r_2(t) = r_1(t) - c_2(t), \dots, r_n(t) = r_{n-1}(t) - c_n(t),$$

until the predetermined stopping criteria is reached. At this stage the data has been decomposed into a set of IMFs and a residual term

$$X(t) = \sum_{i=1}^n c_i(t) + r_n(t).$$

where the residual can be interpreted as the mean trend of the data. The sifting process serves two purposes: Firstly to ensure that each IMF contains only one mode of oscillation, and secondly to smooth uneven amplitudes, making them more symmetric with respect to zero. This introduces the crucial question of when to stop sifting. Due to the second effect, over sifting can lead to pure frequency modulated signals with constant amplitude, obliterating the physically meaningful amplitude modulations. To preserve the natural amplitude variations of the IMF, the sifting procedure must be limited to as few steps as are mathematically permissible while still ensuring that each IMF satisfies the necessary conditions.

Stopping criteria for producing an IMF

The criterion for stopping the sifting process and designating the resulting component an IMF is of key importance. The number of sifting steps must be balanced between under-sifting and obtaining a false IMF, and over-sifting which renders the resulting decomposition less physically meaningful. In our implementation we use the criterion $|m_k(t)| < \text{tol}$ for all t , from Kim and Oh (2009), which directly relates to the criteria that the mean of the upper and lower envelopes must be zero. For other possible choices of stopping criteria see e.g. Huang et al. (1999).

Envelope estimation

The upper and lower envelopes are calculated as interpolated curves between extrema. It is generally recommended that cubic splines be used for this interpolation, as other methods tend to increase the required number of iterations and lead to over decomposition (Rilling et al., 2003). However this remains an area of active research, for example Pegram et al. (2008) propose a method

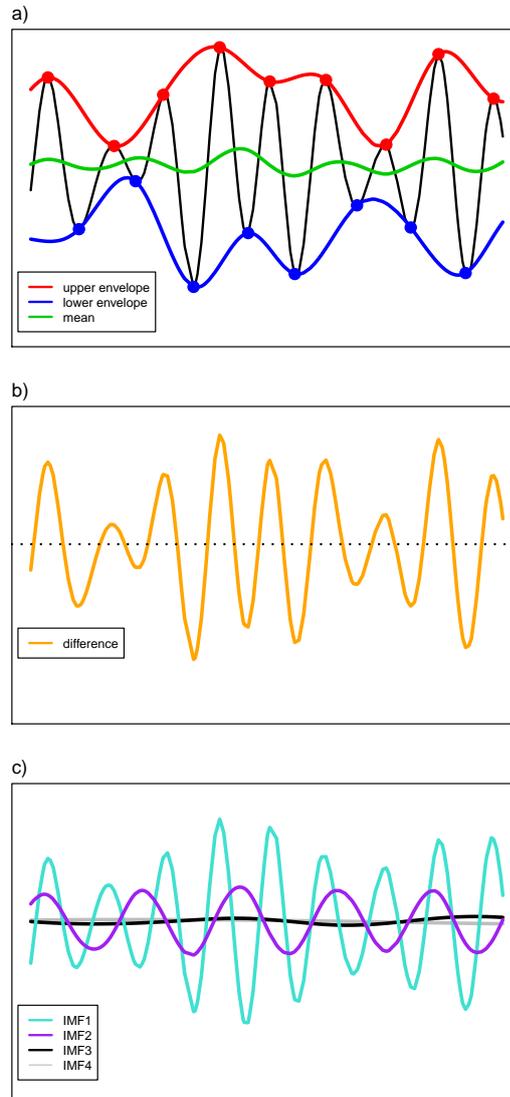


Figure 5.10: Example of EMD decomposition using a small section of the orbital forcing data introduced later in Section 5.7.2. a) Estimation of the upper and lower envelopes, b) Difference between the data and the mean of the envelopes. This is not an IMF so the sifting process is repeated. c) Final IMF decomposition.

based on rational splines and flexible treatment of the end conditions.

The issue of IMF selection

EMD is essentially defined by an algorithm and different IMF sets can be generated by using different parameters in the algorithm. As discussed in Huang et al. (1999), they all represent the truth approximately, while some might be closer to the truth than others. The problem of how to optimise the procedure to produce the best IMF set remains an open problem (Huang et al., 2003). The orthogonality index was introduced by Huang et al. (1998) and can be used to reject IMF's that are grossly non-orthogonal. The orthogonality index is given by

$$IO = \sum_{t=0}^T \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} \frac{c_j(t)c_k(t)}{X^2(t)}. \quad (5.8)$$

If the decomposition is orthogonal then the cross terms should be zero; hence an orthogonality index of close to zero indicates a decomposition that is close to orthogonal.

Intermittence

Problems in the EMD procedure can occur when the variations in the signal are of an intermittent nature. Intermittence in the observed signal commonly leads to 'mode mixing' in the EMD decomposition: a single IMF can consist of oscillations of widely different scales and hence cease to have physical meaning. Mode mixing is the main reason for instability in the EMD decomposition (Wu and Huang, 2009). In this context instability means that any small change can result in a new set of IMFs and hence the procedure suffers from a lack of uniqueness.

Huang et al. (2003) suggest a way to overcome the problem of mode mixing by restricting the frequency range of each IMF, implemented by limiting the distance between successive maxima. However as noted by Wu and Huang

(2009), this is based on a subjectively selected scale and with this intervention the EMD ceases to be totally adaptive. Also, the subjective selection of scales works only if there are clearly definable and separable timescales in the data. An alternative method to prevent mode mixing, ensemble empirical mode decomposition, was introduced by Wu and Huang (2009). The method is based on the observation that mode mixing causes instability, and an important issue is therefore whether the decomposition is sensitive to noise. This method is reviewed in further detail later.

Sampling

The EMD formulation assumes that the signal is described in continuous time, but in practice it is implemented on a discrete time signal. The sampling of the signal is therefore expected to affect the results of the decomposition in some way. EMD relies on the correct identification of signal extrema, and the extrema of a sampled signal are not necessarily the same as those of the underlying continuous time function (Rilling and Flandrin, 2009). For example Rilling et al. (2003) show that even for a pure tone, EMD may not extract the single mode correctly if the sampling period is insufficient.

Extensions to basic algorithm

There have been many suggestions on possible extensions to the basic EMD methodology, attempting to compensate for the described issues of mode mixing and sampling rate. We now briefly describe two interesting advancements to the original algorithm. Although these extensions are not considered in the results of the chapter, the ideas provide an interesting foundation for generating robust EMD decompositions of palaeoclimate signals.

Ensemble empirical mode decomposition (EEMD) was introduced by Wu and Huang (2009). Using EEMD the true IMF components are defined as the mean of an ensemble of trials, each consisting of the signal plus a finite amplitude white noise. This approach is motivated by the studies of Flandrin

et al. (2004) and Wu and Huang (2004), showing that EMD is effectively an adaptive dyadic filter bank when applied to white noise. When a signal is added to white noise, the components of the signal at different scales are automatically projected onto proper scales of reference established by the white noise in the background. Though each individual trial may produce noisy results, given enough trials (with different white noise realisations), the noise will be cancelled out in the ensemble mean. Wu and Huang (2009) show that EEMD can be used to eliminate the mode mixing problem, preserving the physical uniqueness of the decomposition.

Another variation on the classical EMD implementation is the Local EMD algorithm (Rilling et al., 2003). The standard EMD sifting procedure is applied to the full length of the signal and continues until the stopping criterion has been met. However Rilling et al. (2003) note that this can lead to over-sifting on the whole signal for the sake of a better local approximation in isolated regions. This can contaminate other parts of the signal by uniformising the amplitudes and over decomposing the signal by spreading out components over adjacent modes. To overcome this problem Rilling et al. (2003) introduce an intermediate step into the sifting process: Local zones where the error remains large are identified and isolated so that extra iterations are applied only to these regions.

EMD provides a data-driven decomposition method which is well suited to our application of nonstationary and irregularly sampled time series. As highlighted by this brief description there are issues with the application of EMD due to the non-uniqueness of the resulting decomposition and its sensitivity to sampling rate and noise. However under certain circumstances EMD can be used to provide extremely effective time series decompositions. The use of EMD, as well as its limitations, is demonstrated in Section 5.7. Here our emphasis is on comparison with wavelet lifting periodogram estimation. We demonstrate the advantages and disadvantages of both methods, and show how they can be used to complement each other.

5.7 Application to palaeoclimate time series

This section examines the use of wavelet lifting to estimate the periodogram of palaeoclimate data and compares the results to those obtained using EMD. We consider a composite 800 kiloyear (kyr) carbon dioxide series and also an orbital forcing time series which describes the effects of astronomical parameters. Before applying both methods to the data, we first provide a background to the subject area and briefly describe each data set.

A climate forcing is a change imposed on the Earth's energy balance that typically causes a change in global temperature (McGuffie and Henderson-Sellers, 2005). Climate forcings can be considered in two categories: External and internal climate forcings. Internal forcings are due to variations in the components of the climate system, e.g. changes in carbon dioxide levels or volcanic eruptions. External forcings, as considered in this section, are caused by variations in agents outside the system. We shall be considering the forcing effect caused by changes in astronomical parameters.

A decisive advance in the astronomical theory of palaeoclimates came from Milankovitch's theory of insolation (Milankovitch, 1969). The theory states that ice ages are a consequence of variations in the amount and distribution of solar insolation. Note that the distribution of insolation is likely to have a marked affect on climate due to the differing proportion of landmass to ocean in the Northern and Southern hemispheres. There is a general acceptance that the Milankovitch theory of insolation is at least partly correct, substantiated by the fact that geological records have been shown to exhibit spectral properties similar to those of astronomical phenomena (Roes and Allen, 1999). However the exact nature of the climate's response to the orbital parameters still leaves many unanswered questions. For example throughout the last 500 kyr the climate has been characterised by a strong 100 kyr cycle. This cycle shows an 80 kyr build up of ice followed by a 20 kyr termination and hence is often described as a 'saw tooth'. Looking back further than 1 myr BP, the dominant

period of variation was 41 kyr and the reason for this switch in the predominant period is unknown (Augustin et al., 2004). The detection of astronomical frequencies from geological records and comparison with the driving force from orbital parameters is crucial to establishing the mechanisms behind climate variations.

5.7.1 Characteristics of palaeoclimate data

There are many different sources that can be used to extract palaeoclimate time series; tree rings, marine sediments and ice cores are amongst the most popular. We shall be concentrating on the analysis of data from ice cores, but many of the characteristics of ice core records (e.g. uneven sampling) are also shared with other palaeoclimate series. A detailed summary of the characteristics and importance of ice core records is given by Wolff (2005).

Time series obtained from ice cores are characterised by an uneven sampling rate. This is because deeper down the ice core the snow/ice is under a stronger mass pressure, resulting in depletion, pinching and swelling of the layers (Witt and Schumann, 2005). Equidistant sampling along the depth of the core therefore leads to a coarsening of time resolution with increasing age. Isotope-based age estimates also cause uncertainties concerning the age axis of ice-core records, with the uncertainty increasing as the age of the samples increases. In general, though, ice-cores are well-dated (Wolff, 2005) and so for the purpose of this analysis we shall not attempt to make any modifications for the presence of dating uncertainty in the series. Our main focus is in providing a method that makes direct use of the irregularly sampled observations.

Ice-cores are a very important resource for the study of past climate change. One of their main advantages is that they contain information on many different environmental variables. The isotopic content of the water molecules can be used as a proxy for temperature. It is possible to use the oxygen isotopic ratio ($^{18}\text{O}/^{16}\text{O}$), or the hydrogen isotopic ratio ($^2\text{H}/^1\text{H}$) of deuterium to

normal hydrogen. These proxies are generally labelled $\delta^{18}\text{O}$ and δD . The two isotopic ratios represent different mechanisms related to temperature; $\delta^{18}\text{O}$ reflects changes in global ice volume and the hydrological cycle (Petit et al., 1999) and δD reflects local temperature via the amount of deuterium in precipitation. The overall result is that colder temperatures lead to water that is more depleted in the heavier isotopes (Wolff, 2005). The relationships are roughly linear over temperature, with ^{18}O decreasing by 0.7 ‰ (parts per thousand) and ^2H decreasing by 5.6 ‰ for each 1°C temperature decrease (McGuffie and Henderson-Sellers, 2005, Section 1.2.3). In Section 6.4 we consider the relationship between these two climate proxies.

Ice-cores also provide direct records of past changes in atmospheric trace-gas composition via air bubbles entrained in the ice. Unlike the isotope records, these measurements are not proxies. One issue with these records is that the bubbles form at a depth just below the surface, and the air takes just a few years to diffuse to that depth (Wolff, 2005). Direct records from air bubbles trapped in the ice therefore have a different depth-age relationship to measurements taken directly from the ice itself. At a given depth along the ice core, measurements taken from the ice (e.g. δD , $\delta^{18}\text{O}$) will relate to an older time period than those of the trace gases sampled at the same depth. This creates issues and uncertainties when looking at the phasing between climate proxies and trace gas changes.

In our analysis we consider a composite carbon dioxide series that dates back roughly 800 kyr, from 137 yrs BP to 797099 yrs BP where BP stands for before present (‘present’ is conventionally set to be 1950). The series consists of data from the EPICA Dome C and Vostock ice core projects (see Lüthi et al. (2008) and Petit et al. (1999)), and was obtained from the World Data Center for Paleoclimatology, NOAA Paleoclimatology Program, Boulder. The sampling of the data is highly irregular with sampling intervals ranging from 9 yrs to 6029 yrs.

5.7.2 Orbital forcing data

Astronomical theory is built on the hypothesis that climate is somehow determined by the seasonal and spatial distribution of insolation (Crucifix et al., 2006). There are three main components to the orbital forcing of climate: precession, obliquity and eccentricity. Berger (1978) provides trigonometrical expressions for these components, which together provide a time series of the astronomical effects on climate which we can then compare to the palaeoclimate records. Note that this series is derived directly from Newtonian mechanics of the solar system, rather than from observed records.

We now give a brief summary of the components of the orbital forcing signal. Further details on the calculation of these components can be found in the original paper (Berger, 1978) and the review of Crucifix et al. (2006). The astronomical effects on climate present in the orbital forcing time series are:

- *Climatic precession* is characterised by the movement of the perihelion (when the Earth is closest to the sun) with respect to the moving vernal equinox. It is due to two effects; the general precession (or wobble) of the Earth's axis of rotation, and the rotation of the Earth's elliptical figure with respect to the stars. Combined, the climatic precession has four dominant periods (23.7,22.4,18.9,19.1), with a mean period of roughly **21 kyr**.
- *Obliquity* is the angle of tilt of the Earth's axis and is responsible for the seasons (Crucifix, 2008). When the obliquity increases, the amplitude of the seasonal cycle in insolation increases, with summers in both hemispheres receiving more energy and the winters less energy. As a result, it is assumed that the winters become colder and summers warmer. Obliquity varies between 22° and 25° with a period of roughly **41kyr**.
- *Eccentricity* is the orbital deviation from circular. There are five dominant modes (404,95,124,9,131) with a **100 kyr** combination of periods.

Neither precession nor obliquity modifies the total amount of energy reaching the Earth in a year. Eccentricity does, but the effect is small (Crucifix, 2008). Since the effect of eccentricity is weak relative to the signals of obliquity and precession, if the response of the climate to the astronomical forcing parameters is linear then we would expect the 100-kyr signal in the data to be of negligible intensity (Mélise et al., 2001). However, as already noted, for the last 500 kyr the climate has been characterised by a strong 100 kyr signal. Nonlinear processes within the climate system have been proposed as the origin of the 100 kyr cycle (see Mélise et al. (2001), Crucifix and Rougier (2009)).

Given the quantification of these astronomical forcing effects by Berger (1978), we can use the generated time series to estimate the lifting periodograms. These can then be compared with the periodograms observed from the palaeological observations. Since the composite carbon dioxide record is observed on an irregular grid, we consider the orbital forcing data at the same irregular time points. Note that the orbital forcing time series is calculated from first principles and so, unlike the observed ice-core time series, does not contain observation and dating uncertainty. We also know the major frequencies of variation that are present in the data, allowing us to judge accurately the success of each method. For these reasons the orbital forcing time series acts as a good test data set.

The main aim of our analysis is to improve on the results of methods that rely on pre-processing the data to a regular grid. In particular, we are interested in noting if there is any high frequency information present in the series that has not previously been detected due to the loss of information at these scales. The detection of high frequency harmonics would be very interesting and could potentially lead to the verification/rejection of differing theories of glaciation. This is one of the issues being investigated in the ITOP (Integrating Theory and Observations over the Pleistocene) project led by Dr. Michel Crucifix.

5.7.3 Analysis

The lifting periodograms of each series were estimated using one neighbour either side linear prediction, forming the periodogram using 40 artificial scales, and smoothing the periodogram estimates over time using a window of 10 kyr. The orbital forcing signal, carbon dioxide record and corresponding wavelet lifting periodograms are shown in Figure 5.11. From the periodogram of the orbital forcing data the 21 kyr precession cycle is clearly visible. The signal shows a ‘beating’ over time due to the combination of tones. The lifting periodogram of the carbon dioxide data shows a 100 kyr signal which is most pronounced in the middle of the signal, when the amplitude of the orbital forcing signal is at its lowest.

The EMD decompositions of each signal were also considered, using an IMF stopping tolerance level of $|m_k(t)| < 0.175^2\sigma_X$ for all t , where σ_X is the standard deviation of the series. The envelopes were estimated using splines and the edge effects were dealt with by mirroring the end values. From the EMD decomposition of the orbital forcing data (Figure 5.12) we can clearly distinguish the different components of the signal: IMF 1 contains the precession component, IMF 2 contains the obliquity, and IMF 3 contains the eccentricity. The orthogonality index of the decomposition, as calculated using equation (5.8), is 0.097, indicating that the decomposition is close to orthogonal. The decomposition is also robust to changes in the options of the algorithm.

Two possible EMD decompositions of the carbon dioxide data are shown in Figure 5.13. The first decomposition is constructed using the same choices in the algorithm as with the orbital forcing example, while for the second decomposition the tolerance limit for the stopping criteria has been reduced to $|m_k(t)| < 0.125^2\sigma_X$. The corresponding lifting periodograms of the two decompositions are shown in Figures 5.14 and 5.15. These results appear difficult to interpret, in that we can not readily attribute known physical phenomena to the resulting IMFs. Unlike the orbital forcing data, the carbon dioxide series

is sensitive to changes in the algorithm components – small changes result in large differences in the resulting decomposition, particularly for the coarser scale IMFs. The orthogonality index for the first decomposition is very high at 0.228. Lowering the tolerance level to $0.125^2\sigma_X$ results in a decomposition with an improved orthogonality index of 0.019, but the decomposition has a larger number of IMFs and does not improve the overall interpretation of the decomposition. The results are still promising, for example it appears that IMF 4 is may be capturing part of the precession signal, but further work on a robust method is needed. It is also important to note that carbon dioxide responds indirectly to insolation and that many lags and feedbacks are involved.

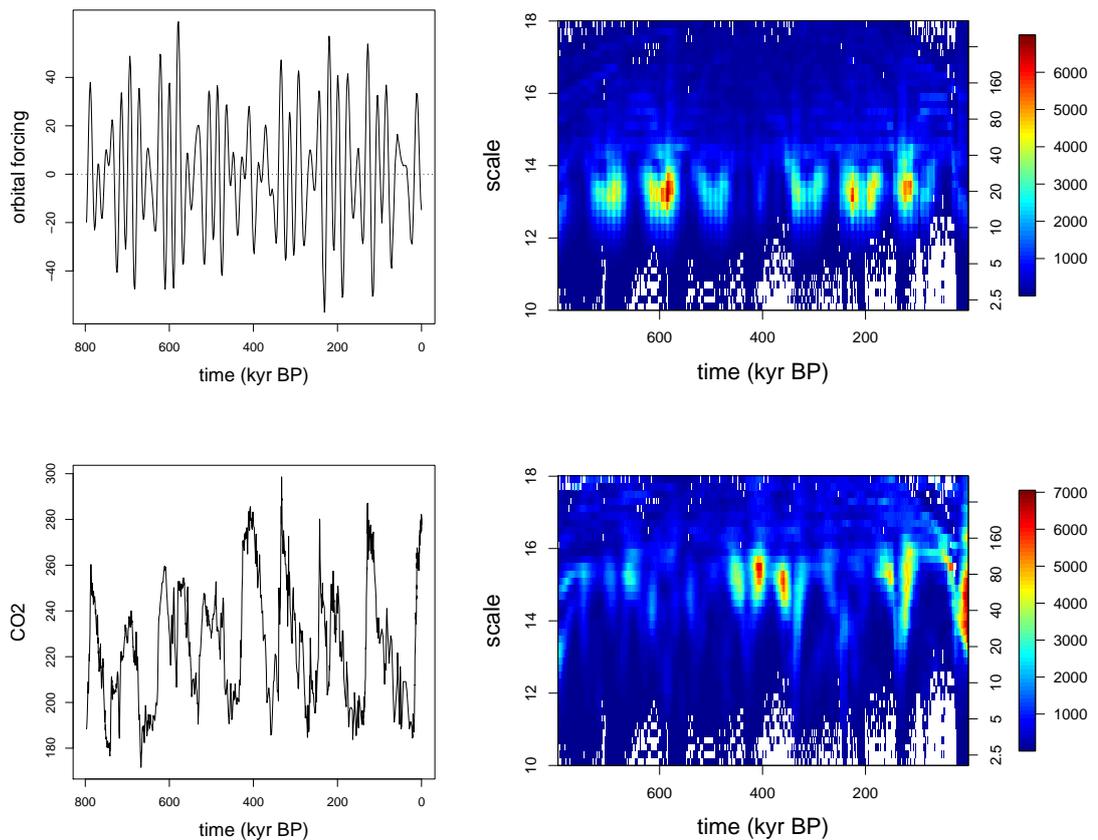


Figure 5.11: Time series (left) and wavelet periodograms (right) for the orbital forcing and carbon dioxide signals.

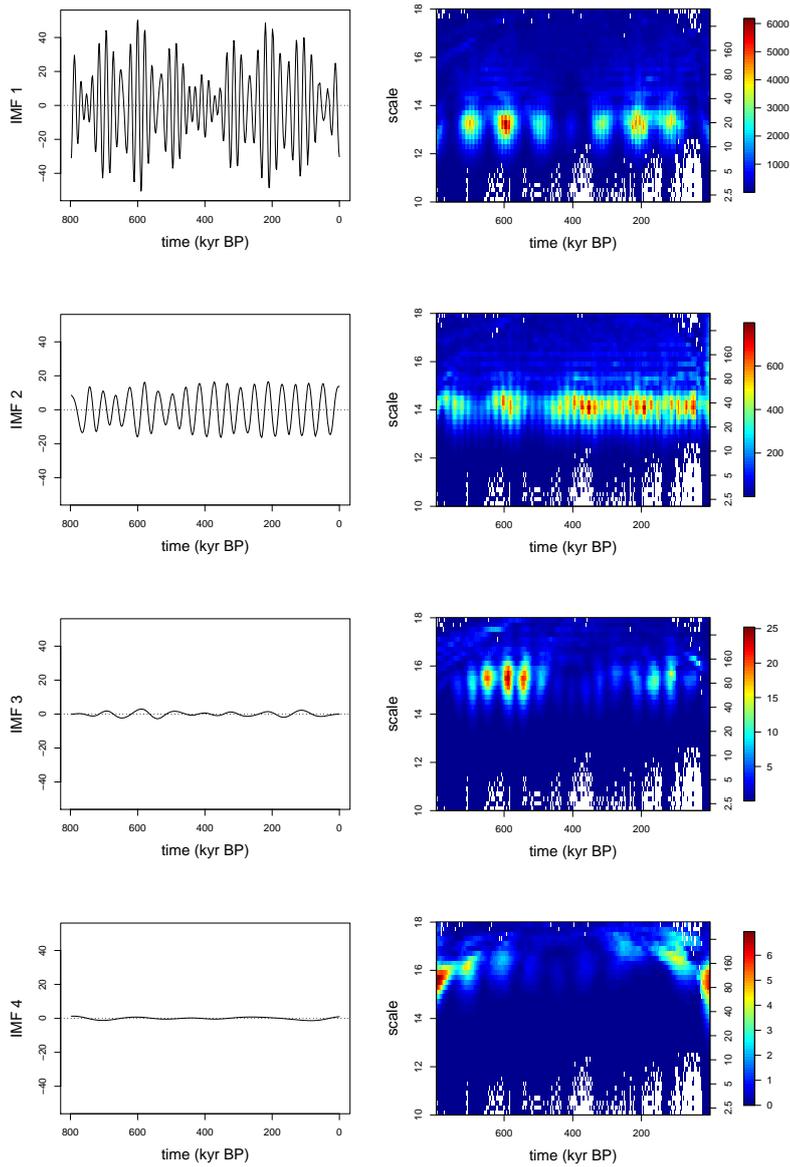


Figure 5.12: IMF decomposition (left) and associated lifting periodograms (right) for the orbital forcing data.

5.7. Application to palaeoclimate time series

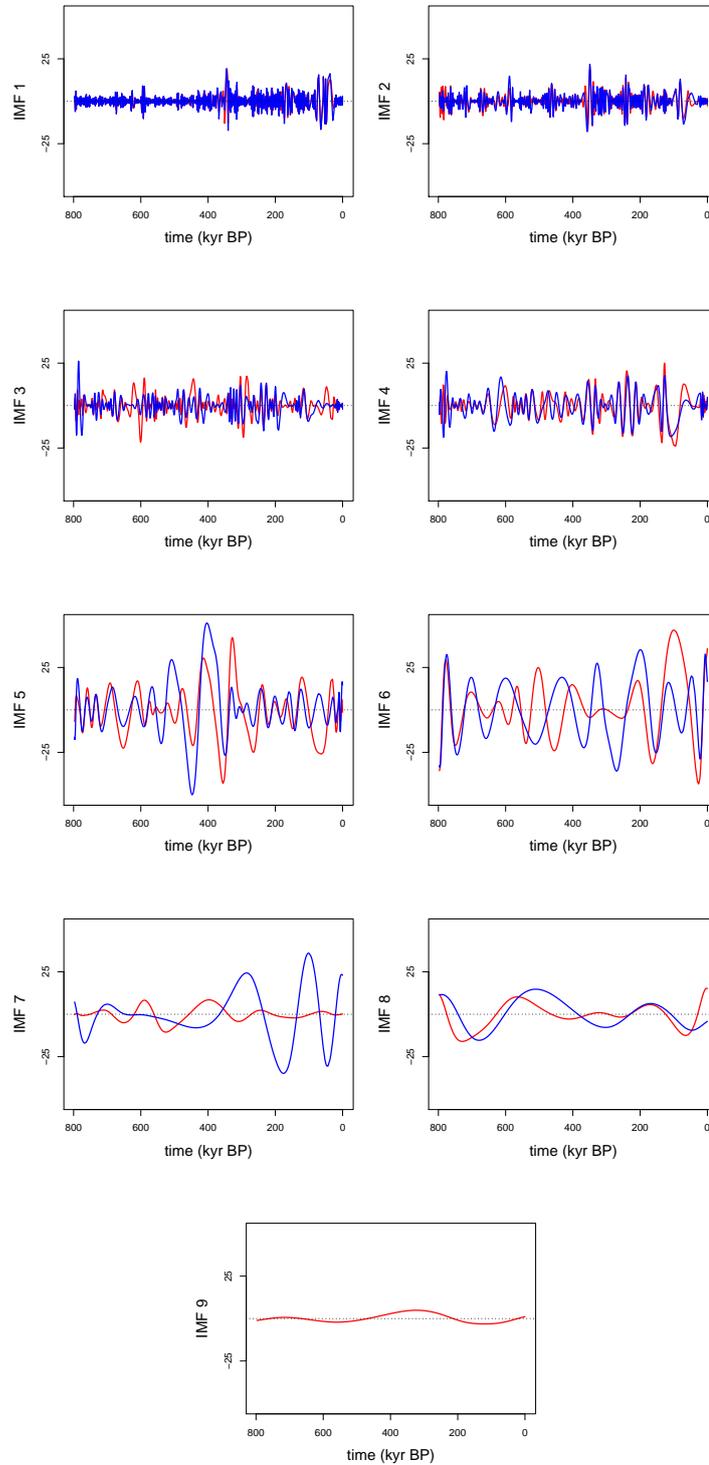


Figure 5.13: IMF decomposition for the Carbon Dioxide data using a tolerance level of $0.175^2\sigma_X$ (red) and a tolerance of $0.125^2\sigma_X$ (blue).

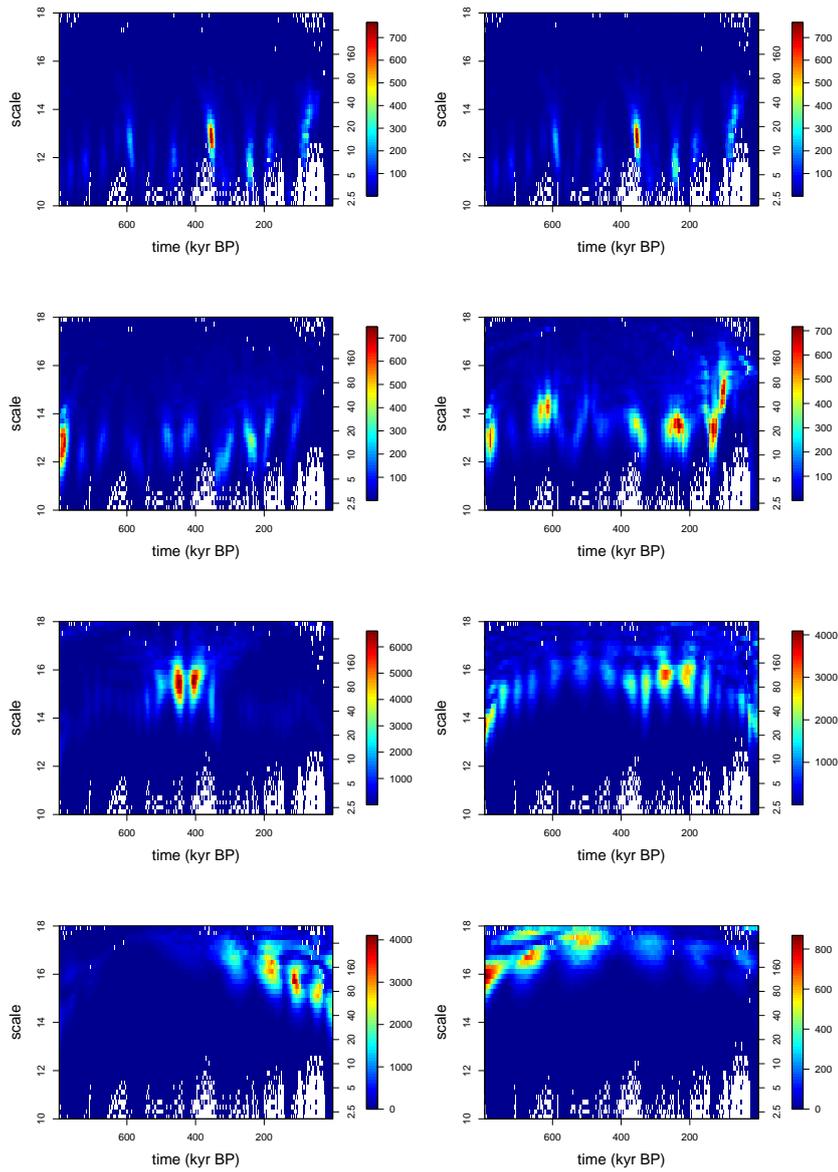


Figure 5.14: Lifting periodograms of the IMF components for the carbon dioxide data, decomposed using a tolerance level of $0.175^2\sigma_X$.

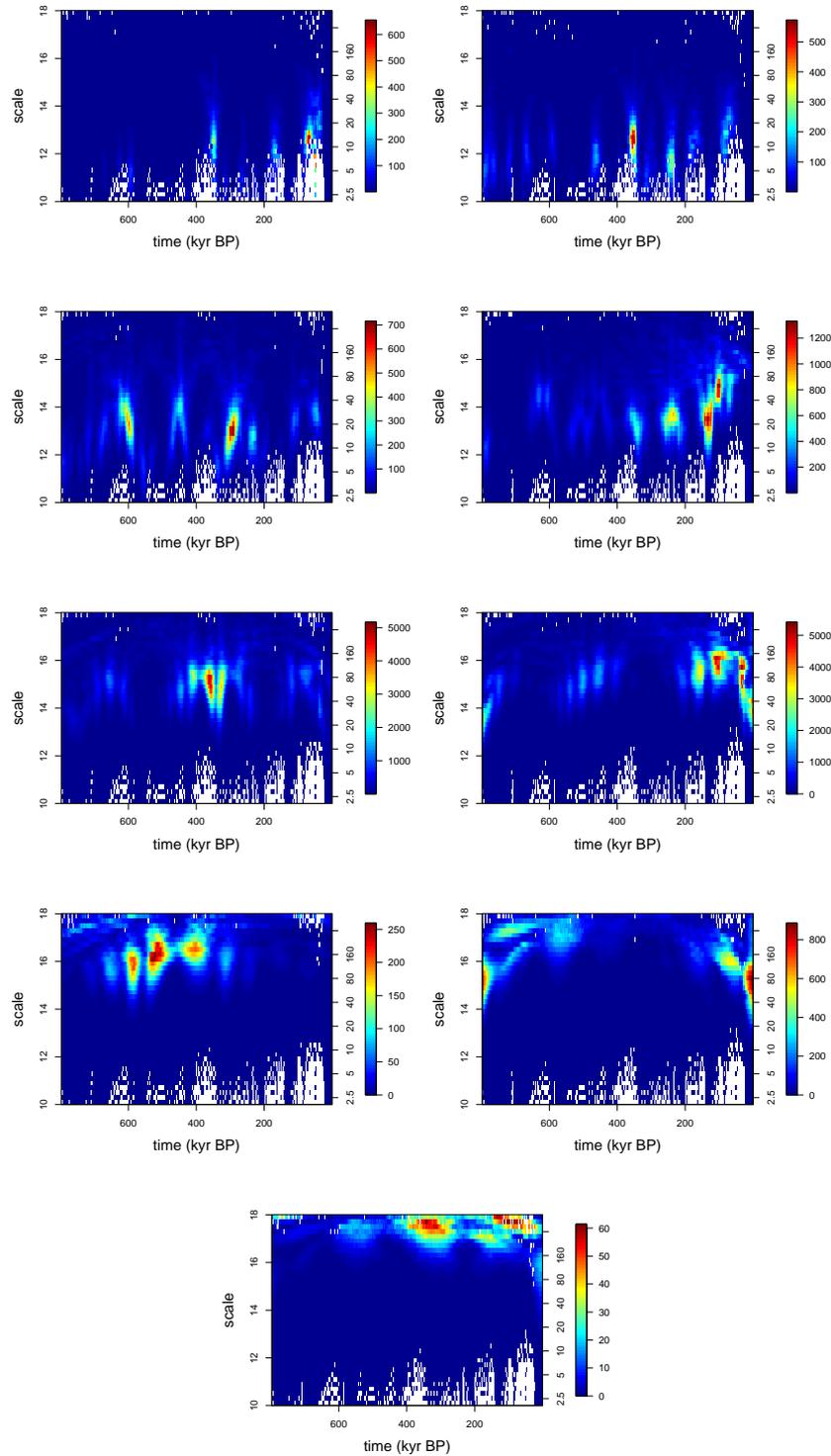


Figure 5.15: Lifting periodograms of the IMF components for the carbon dioxide data, decomposed using a tolerance level of $0.125^2\sigma_X$.

5.8 Method comparison

From the wavelet lifting periodogram of the orbital forcing data the 21 kyr precession cycle was clearly visible. Considering the EMD decomposition of the original signal we also observe other characteristics: the 41 kyr signal of obliquity (IMF 2) and the 100 kyr signal of eccentricity (IMF 3). These are clearly visible from the lifting periodograms of the individual IMF but the coefficients are of considerably lower magnitude than from IMF 1 which contains the precession component. Summing the periodograms of all the individual IMF (plus the residual term) provides the periodogram of the original series and so all this information is essentially contained within the lifting periodogram of the full series; however the coefficients of the obliquity and eccentricity signals are masked by those of the precession component. In this case, EMD provides a more informative decomposition of the series than obtained by wavelet lifting and highlights the limitations of the wavelet lifting periodogram methodology.

For the carbon dioxide time series the EMD decomposition is less convincing and a more complete picture of the overall characteristics of the data is observed from the lifting periodogram of the full series. The orbital forcing and carbon dioxide signals are observed on the same irregular time grid and the main difference in characteristics between them is the presence of noise in the carbon dioxide data. Since the basis of the EMD decomposition is derived from the data, signal extraction using EMD can be problematic as noise levels increase (Kijewski-Correa and Kareem, 2007). The basis is dependent on the observed signal and in the presence of noise this means that the basis is also affected by the noise.

Through the use of EMD we can recognise some limitations of the wavelet lifting periodogram methodology. The small amplitude periodic signals are lost in the decomposition due to the blurring over scale. EMD provides an orthogonal (or close to) decomposition and so does not suffer from this problem. It is possible to correct the periodogram estimates using the methodology of

Knight (2006), but at this time this is not feasible for our chosen application for the reasons discussed in Section 5.5. Note that this does not mean that EMD is superior to lifting for the analysis of palaeoclimate signals as demonstrated by the analysis of the carbon dioxide series.

The other main difference between the two methods is that EMD is adaptive. Adaptivity is preferable in representing nonlinear data as non-adaptive methods are generally characterised by the production of harmonics and a spreading of energy over scales (Huang et al., 1998). However we saw from the analysis of the carbon dioxide data that adaptivity in the EMD procedure can be problematic when the data contains noise. It is possible to use an adaptive wavelet lifting scheme, however methods for periodogram estimation under this framework are not well-developed.

5.9 Summary

In this chapter we have explored the ‘spectral estimation for irregular designs’ methodology of Knight et al. (2010). Our contribution to this area arises from our chosen application of palaeoclimatology. Time series from ice-cores, and other palaeoclimate sources, are inherently irregularly spaced with the sampling becoming coarser as we look further back in time.

Our application utilises a slightly different definition of the periodogram to that used by Knight et al. (2010). We estimate the ordinates at each time point and artificial scale using the average of the observed coefficients rather than using a non parametric smoothing approach. Though the latter approach may be beneficial in some situations, the simpler ‘averaging’ technique results in a more intuitive measure of coherence when estimating the dependence between two series in the next chapter.

Given our chosen application, it is also very important to define a relationship between wavelet scale, as defined by the lifting scheme, and Fourier frequency. This subject was discussed in Section 5.3 where we provided a mapping between scale and frequency for periodograms computed using linear prediction. We also showed that the relationship between scale and frequency is dependent on the symmetry of the neighbours used at the prediction stage, and proposed that coefficients generated using highly asymmetric neighbour pairs should not be used to construct the final periodogram.

Overall, we reinforce the work of Knight et al. (2010), highlighting that wavelet lifting is a promising new method for the analysis of irregularly sampled time series. The method allows us to visualise the periodicities present in palaeoclimate series, clearly capturing the major aspects of each signal. However there are limitations with the method, and we see from the application of EMD that smaller amplitude, but important, periodic signals were lost in the periodogram due to the ‘blurring over scale’.

Ideally we would like to be able to correct the periodogram to provide clearer estimates. Knight et al. (2010) propose a correction procedure for the raw lifting periodogram based on the formulation of the series as a LSW process with missing observations and this subject provides a promising topic for future work. Throughout our implementation we have concentrated on the use of linear prediction with one neighbour either side for the prediction step of the lifting scheme. The use of adaptive lifting for spectral estimation (Nunes et al., 2006) also provides an interesting direction for further research.

Our analysis did not provide any extra information regarding high frequency variation in the records. The detection of high frequency harmonics would provide an extremely interesting scientific result, potentially providing valuable insight into the mechanisms dominating climate change. This would also highlight the benefits of lifting over other competing methods that rely on pre-processing to a regular grid prior to analysis, and hence result in a loss of information at high frequencies. This remains an important consider-

ation for future work and advances are possible through application to other data sets, including different climatic variables and observations from different sites. Progress in this area is also possible through further advances in the lifting periodogram methodology.

Chapter 6

Bivariate spectral analysis for irregular situations

As in the regularly-sampled setting, it is often the case that we observe more than one time series and are interested not only in the individual spectral properties of these series but also in the dependence between the signals. Problems of this kind are common in palaeoclimatology since ice-cores provide measurements of several variables at each time point. For example, Section 6.4 considers the relationship between two different proxies for temperature. We have seen that in the regularly sampled setting the bivariate LSW model can be used to estimate the coherence between time series. However there is no well-defined method for estimating the coherence between two irregularly spaced time series. Potential approaches to this problem include that of Mondal and Percival (2008), who state that their method for univariate ‘gappy’ time series could be extended to bivariate situations. Also, Rilling et al. (2007) propose a bivariate extension to the EMD methodology suitable for data sets where the two components can be expressed as Cartesian coordinates of a point moving in a two-dimensional space. This formulation makes it possible to define extrema and signal envelopes for bivariate signals. While the standard methodology is designed to extract zero-mean oscillating components, the bivariate extension extracts zero-mean rotating components.

Chapter 5 demonstrated that lifting provides an intuitive way of adapting classical wavelet methodology to irregular designs. Based on the OCAAT lifting scheme, Knight et al. (2010) define the NLT periodogram, providing an analogue to the nondecimated wavelet periodogram that is suitable for irregularly sampled signals. This chapter extends the ideas of the NLT periodogram to the bivariate setting, providing estimates of the dependence between two (possibly irregularly sampled) time series. This work consists of two main advancements to the methodology. Firstly, in order to estimate the phase between the signals we define a *complex one coefficient at a time lifting scheme*, based on the OCAAT scheme but with a modification to produce complex-valued coefficients at each stage. We then introduce the *bivariate nondecimated lifting* framework. Lifting two series simultaneously enables the construction of a lifting cross-periodogram, providing information on the dependence between the series. We begin by detailing the proposed complex OCAAT lifting transform, then describe its use in a bivariate NLT setting. This provides a full spectral decomposition, including coherence and phase. The method is demonstrated using a simulated example as well as real palaeoclimate data.

6.1 The complex one coefficient at a time lifting scheme

The complex lifting scheme proposed in this section is motivated by the complex wavelet transform (CWT) for regular data introduced by Kingsbury (2001). The CWT is a dual transform and can be thought of in terms of two separate branches. The two branches consist of separate discrete wavelet transforms performed using two different (real) wavelets. The coefficients from one branch of the transform represent the real part of the wavelet coefficient while the coefficient from the second part of the transform represents the imaginary part. Therefore, using two different real wavelets, the CWT provides a com-

plex decomposition. The CWT can be compared to other discrete complex wavelet methods such as that of Belzer et al. (1995) and Lina and Mayrand (1995). These methods use complex wavelets which form orthogonal bases. In contrast, the CWT is a redundant transform in which the real and imaginary components of the wavelet individually form an orthogonal basis. This results in a transform that is nearly shift invariant but with less redundancy than the nondecimated DWT. Further details can be found in Kingsbury (1998), Kingsbury (2001) and Selesnick et al. (2005).

The proposed complex lifting transform follows the principles employed by the CWT, but in the context of wavelet lifting. Instead of using the DWT for each branch of the transform, our methodology is based on the one coefficient at a time lifting scheme of Jansen et al. (2009), ensuring that the method is well adapted for irregular design situations. Since the scheme combines the principles of the CWT with that of OCAAT lifting, we refer to the method as *complex one coefficient at a time* (complex OCAAT) lifting.

The complex OCAAT scheme follows the standard OCAAT scheme but at each prediction stage the algorithm proceeds following two different prediction methods, producing two detail coefficients, λ and μ , which represent the real and imaginary parts of the transform. The two prediction schemes are defined by their associated filters, \mathbf{L} and \mathbf{M} . At this stage the scheme is introduced using an abstract choice of the filters, but choice of the two prediction filters is discussed further in Section 6.1.1.

Having produced two coefficients using the filters \mathbf{L} and \mathbf{M} , the update stage then proceeds using just the first filter, \mathbf{L} . Therefore, unlike the CWT, the complex OCAAT scheme does not consist of two completely separate branches, as the same updating step is used for each branch. The proposed complex OCAAT lifting scheme is computed as follows:

- **Split:** Choose a point to be lifted according to some specification. The point to be lifted at stage n is denoted j_n .

- **Predict:** The real and complex detail coefficients are found by predicting using the filters **L** and **M**, according to

$$\lambda_{j_n} = c_{n,j_n} - \sum_{i \in J_n} l_i^m c_{n,i}, \quad (6.1)$$

$$\mu_{j_n} = m_{j_n}^n c_{n,j_n} - \sum_{i \in J_n} m_i^n c_{n,i}, \quad (6.2)$$

where J_n is the set of neighbours, $\{l_i^n\}_{i \in J_n \cup j_n}$ and $\{m_i^n\}_{i \in J_n \cup j_n}$ are the prediction weights associated with filters **L** and **M**, and $l_{j_n}^n = 1$. Note that while equation (6.1) is identical to the predict step of the standard OCAAT lifting scheme (equation (2.23)), the second predict step as given by equation (6.2) takes a slightly different form. The difference is that the formulation in this chapter does not require the prediction weights associated with the removal point $m_{j_n}^n$ to be 1. The detail coefficient obtained from scheme **M** is therefore not necessarily the difference between the observed value and the value predicted by the neighbours. This modified formulation provides a generalisation of the standard scheme which allows more flexibility in the choice of filter pairs. Since we update according to scheme **L**, this modification does not affect the update step. The full complex valued wavelet coefficient is found by combining the details from the two different schemes to give

$$d_{j_n} = \lambda_{j_n} + i\mu_{j_n}. \quad (6.3)$$

- **Update:** The scaling coefficients and integrals of the neighbouring points are then updated according to filter **L**:

$$\begin{aligned} c_{n-1,i} &= c_{n,i} + b_i^n d_{j_n}, \\ I_{n-1,i} &= I_{n,i} + l_i^n I_{n,j_n} \quad \forall i \in J_n, \end{aligned}$$

where b_i^n are the update weights associated with filter \mathbf{L} , as given by the standard OCAAT scheme.

The proposed complex OCAAT scheme is illustrated in Figure 6.1. For comparative purposes, Figure 6.2 is also included, illustrating an alternative scheme that consists of two separate branches. Choosing to update with just one filter (as in Figure 6.1) ensures that throughout the algorithm there is just one set of smooth coefficients, associated with one set of integrals. In the updating step the weights are dependent on the prediction method and so for the scheme depicted in Figure 6.2 the coefficients at each stage will be different, and associated with different integrals. Since the integrals are used as a measure of scale this means that coefficients at the same stage of the algorithm, but from different branches, could (depending on the choice of filters) be associated with different scales i.e. $\alpha_{j_n}^{\mathbf{L}} \neq \alpha_{j_n}^{\mathbf{M}}$. This is not an issue with the CWT procedure for regularly spaced data as the scale of the wavelets at a given stage is consistent between the two DWT schemes. The issue is bypassed in our formulation by updating according to just one of the filters as shown in Figure 6.1, ensuring that the wavelet coefficients produced by each scheme are associated with the same scale i.e. $x_{j_n} \rightarrow (\lambda_{j_n}, \mu_{j_n}, \alpha_{j_n})$.

Since at each stage the update step proceeds using prediction scheme \mathbf{L} , this also means that the ordering of the filters is important; interchanging \mathbf{L} and \mathbf{M} would mean updating with a different filter, leading to a different transform. Though the design of two completely separate branches is possible in concept, the design of such a scheme would be difficult due to the issues mentioned above. Using just one filter for the updating step also simplifies the filter design issue making it easier to ensure that the prediction weights of the filter \mathbf{L} are positive. This issue is discussed in further detail in the next section, where we introduce specific choices for the filters \mathbf{L} and \mathbf{M} .

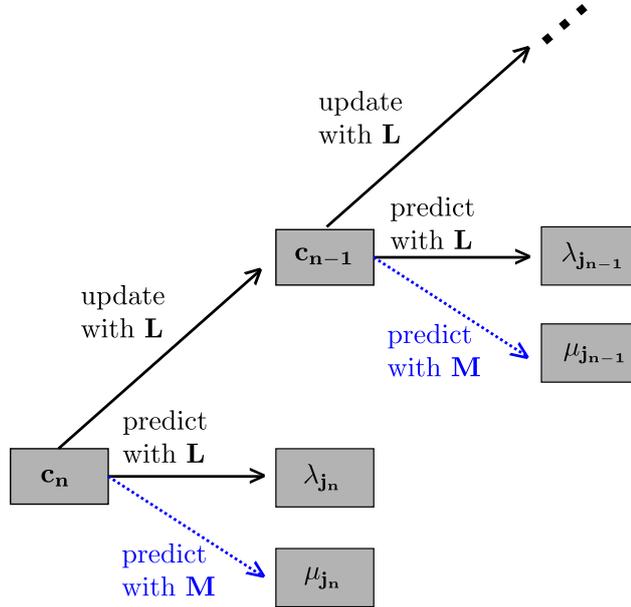


Figure 6.1: The complex OCAAT lifting scheme. The black lines correspond to the prediction and update steps of the standard OCAAT lifting scheme and the blue lines indicate the extra prediction step required for the complex scheme.

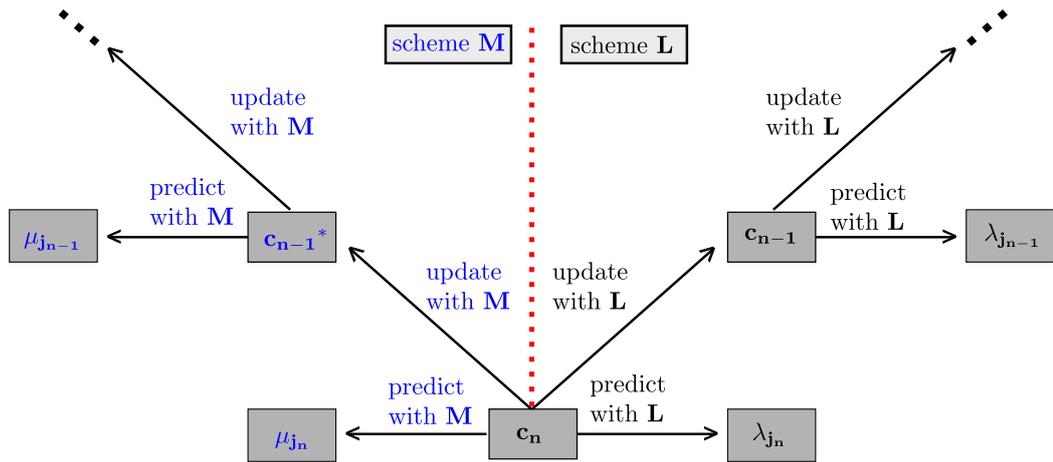


Figure 6.2: An alternative complex OCAAT lifting scheme, consisting of two separate branches with updating occurring in both. Starting with the same data vector c_n , the coefficients at later stages, c_{n-i} , c_{n-i}^* , may be associated with different scales.

6.1.1 Filters for the complex lifting scheme

Implementation of the complex OCAAT scheme relies on suitable choice of the filter pair \mathbf{L} and \mathbf{M} . As discussed in Section 5.3 certain schemes are favourable for spectral estimation. In particular, it is preferable to have positive weights on the neighbours in the prediction scheme. Negative prediction weights cause problems when updating the integrals as the integral is no longer a monotonically increasing value. Since we update using prediction scheme \mathbf{L} , this consideration applies only to the choice of filter \mathbf{L} .

In designing the filter pairs we therefore start by defining filter \mathbf{L} , choosing a filter which is well suited to spectral analysis. For this reason we concentrate on schemes which specify the use of linear prediction with symmetric neighbours for the leading prediction scheme (\mathbf{L}). The filter design problem is thus reduced to defining a second filter, \mathbf{M} . Since this filter is not used in the updating procedure, there is no restriction that the weights of the neighbours must be positive.

In choosing the second filter, we want to ensure that it provides different information about the local structure than that produced using the filter \mathbf{L} . In the regularly sampled CWT methodology, Kingsbury (2001) states that ideally the real and imaginary components of the wavelet should form a Hilbert transform pair, and hence be orthogonal to each other. In practice this is not possible with compactly supported wavelets and so the design of filters for the CWT involves a balance between meeting this criteria and ensuring that the wavelets are compactly supported. In our implementation we specify that the two filters should be orthogonal to each other. Note that this is not the same as specifying that the wavelets associated with each prediction scheme should be orthogonal to each other, but this choice still results in an accurate quantification of phase in the examples considered at the end of this Chapter. This specification is explored in further detail in Section 6.2.4.

Given the choice of filter for the leading prediction scheme, the remaining

task is in designing a suitable orthogonal filter for the second prediction scheme, \mathbf{M} . We now outline two possible filter pairs that can be used with the complex OCAAT lifting scheme.

Linear prediction with one neighbour either side

Consider the case where the real coefficients are obtained using a linear prediction scheme with one neighbour either side. The first filter is given by

$$\mathbf{L} = (l_1, 1, l_3),$$

where $l_1, l_3 > 0$, $l_1 + l_3 = 1$, and the exact values are determined by the sampling of the data. A second, orthogonal filter, is then constructed as

$$\mathbf{M} = (m_1, m_2, m_3),$$

such that $m_1 + m_3 = m_2$ to ensure annihilation of constants, $l_1 m_1 + m_2 + l_3 m_3 = 0$ to provide orthogonality and $l_1^2 + 1 + l_3^2 = m_1^2 + m_2^2 + m_3^2$ to provide equal weighting to the coefficients of each scheme. This can be solved by setting

$$m_1 = R m_3, \quad m_2 = (1 + R) m_3, \quad m_3,$$

where $R = \frac{l_1 - 2}{l_1 + 1}$ and $m_3 = \frac{l_1 + 1}{\sqrt{3}}$. If the time series is evenly spaced, then at the first stage of the algorithm we have $\mathbf{L} = (1/2, 1, 1/2)$ and the filter for the imaginary coefficients is given by $\mathbf{M} = (-m_3, 0, m_3) = (-\sqrt{3}/2, 0, \sqrt{3}/2)$. Therefore the second scheme is effectively looking at the difference between the neighbouring values.

For the removal of observation j_n the filter \mathbf{L} provides the real component of the detail coefficient, denoted λ_{j_n} , and the filter \mathbf{M} provides the complex-valued component of the detail coefficient, denoted μ_{j_n} . Using the notation introduced in Section 5.3, we can describe the relationship between scale and frequency for the coefficients in terms of the scale, α , at which we see a peak

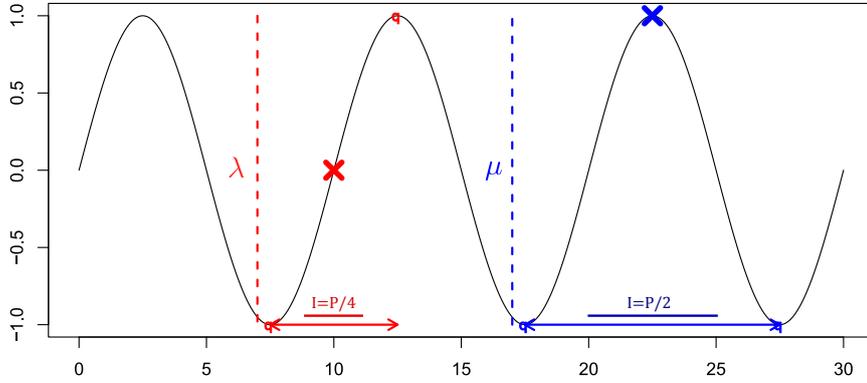


Figure 6.3: Relationship between the integral and maximum value of the associated detail coefficients for the prediction schemes defined by filters **L** (blue) and **M** (red), here shown for a sine wave of period 10. The integrals are shown by the double solid lines and the detail coefficients by the dashed lines. For scheme **L** the coefficients are maximised by considering an integral of $P/2$ and for scheme **M** the coefficients are maximised by considering an integral of $P/4$.

in the coefficients for a Fourier wave of period P . Writing the coefficients from each scheme as $|\lambda| = \lambda(\alpha, x)$, and $|\mu| = \mu(\alpha, x)$, provides the following relationships to frequency

$$\begin{aligned} \{\alpha : \bar{\lambda}(\alpha) \geq \bar{\lambda}(\beta), \beta \in \mathbb{R}^+\} &= P/2, \\ \{\alpha : \bar{\mu}(\alpha) \geq \bar{\mu}(\beta), \beta \in \mathbb{R}^+\} &= P/4, \end{aligned}$$

where $\bar{\lambda}(\alpha)$ and $\bar{\mu}(\alpha)$ are the arithmetic averages over x of $\lambda(\alpha, x)$ and $\mu(\alpha, x)$, for each value of α . This relationship is also illustrated in Figure 6.3, which for a sine wave with period 10, indicates for both schemes the integrals that provide the largest valued detail coefficients.

Linear prediction with two neighbours either side

Another example of a possible filter pair is the case where the real coefficients are obtained using a linear prediction scheme with two neighbours either side.

In this case the real valued filter is given by

$$\mathbf{L} = (l_1, l_2, 1, l_4, l_5),$$

where l_1, l_2, l_4, l_5 are defined by the standard OCAAT scheme and $l_1 + l_2 + l_4 + l_5 = 1$ to ensure annihilation of constants. The second filter can be defined by setting

$$\mathbf{M} = (m_1, m_2, m_3, m_4, m_5),$$

such that $l_1 m_1 + l_2 m_2 + m_3 + l_4 m_4 + l_5 m_5 = 0$ and $m_1 + m_2 + m_4 + m_5 = m_3$. Combining the two equations gives

$$m_1(l_1 + 1) + m_2(l_2 + 1) + m_4(l_4 + 1) + m_5(l_5 + 1) = 0.$$

One possible solution is to solve the two equalities

$$m_1(l_1 + 1) + m_5(l_5 + 1) = 0, \quad m_2(l_2 + 1) + m_4(l_4 + 1) = 0.$$

This provides the following set of weights for the second filter:

$$m_1, \quad m_2 = m_1, \quad m_3 = (R_1 + R_2 + 2)m_1, \quad m_4 = R_2 m_2, \quad m_5 = R_1 m_1,$$

where $R_1 = \frac{-(l_1+1)}{(l_5+1)}$, $R_2 = \frac{-(l_2+1)}{(l_4+1)}$, and we have further assumed that $m_2 = m_1$.

We can ensure that $m_1^2 + m_2^2 + m_3^2 + m_4^2 + m_5^2 = l_1^2 + l_2^2 + 1 + l_4^2 + l_5^2$ by setting

$$m_1 = \sqrt{\frac{l_1^2 + l_2^2 + l_4^2 + l_5^2 + 1}{2(R_1^2 + R_2^2 + R_1 R_2 + 2R_1 + 2R_2)}}.$$

If the data sampling is even, then at the first stage of the algorithm we have $M = (1/4, 1/4, 1, 1/4, 1/4)$, $R_1 = R_2 = -1$ and $M = (m_1, m_1, 0, -m_1, -m_1) = (\sqrt{5}/4, \sqrt{5}/4, 0, -\sqrt{5}/4, -\sqrt{5}/4)$. Similarly to the one neighbour either side scheme, this filter is essentially looking at the difference between the average of the left hand side neighbours and the average of the right hand side neighbours.

Using this scheme, the detail coefficients $|\lambda| = \lambda(\alpha, x)$, and $|\mu| = \mu(\alpha, x)$, obtained from filters \mathbf{L} and \mathbf{M} , have the following relationships to frequency

$$\begin{aligned} \{\alpha : \bar{\lambda}(\alpha) \geq \bar{\lambda}(\beta), \beta \in \mathbb{R}^+\} &= 0.29P, \\ \{\alpha : \bar{\mu}(\alpha) \geq \bar{\mu}(\beta), \beta \in \mathbb{R}^+\} &= 0.15P. \end{aligned}$$

6.1.2 The complex nondecimated lifting periodogram

Section 5.1 reviewed the use of the OCAAT lifting scheme to provide a non-decimated lifting transform (NLT), and Section 5.2 then reviewed the use of the NLT transform to define a periodogram. Similarly, the complex OCAAT lifting scheme can also be used to provide a nondecimated transform and to construct the periodogram. Since our method essentially combines the principles of the CWT with that of the NLT methodology, we call it the *complex nondecimated lifting transform* (CNLT) and refer to the resulting periodogram as the CNLT periodogram.

Recall that the NLT proceeds using repeated applications of the OCAAT lifting scheme with different random trajectories, T_p for $p = 1, \dots, P$. Thus for each removal location, x_k , where $k = 1, \dots, T$, the NLT provides a set of details $\{d_{x_k}^p\}_{p=1, \dots, P}$ where $d_{x_k}^p$ is the coefficient at location x_k , obtained from trajectory T_p . The CNLT is obtained by replacing the standard OCAAT scheme with the modified complex scheme. In the complex setting, a similar set of detail coefficients is obtained, but each detail is complex valued; $d_{x_k}^p = \lambda_{x_k}^p + \mu_{x_k}^p i$.

As with the real-valued lifting transform, the periodogram is found by discretising the wavelet coefficients according to the allocated set of artificial scales and then averaging the squared coefficients in each group.

Definition 6.1.1. For each time point, x_k , $k \in 1, \dots, n$ defined at scale l^i for $i \in 1, \dots, J^*$, the raw lifting periodogram as defined by the complex lifting

scheme is given by

$$\begin{aligned} {}_C I_{x_k}(l^i) &= \text{mean}\{|d_{x_k}^p|^2 : \alpha_{x_k}^p \in l^i\}, \\ &= \text{mean}\{(\lambda_{x_k}^p)^2 : \alpha_{x_k}^p \in l^i\} + \text{mean}\{(\mu_{x_k}^p)^2 : \alpha_{x_k}^p \in l^i\}. \end{aligned}$$

Note the presence of the C in this notation to distinguish the periodogram given by the complex lifting scheme from the standard lifting periodogram of Definition 5.2.1. The periodogram implied by the complex lifting scheme consists of the sum of the wavelet periodograms of the two individual lifting schemes as defined by filters \mathbf{L} and \mathbf{M} (updating according to \mathbf{L}). If the purpose of analysis is to provide a spectral estimate for a univariate time series, then it is preferable to form the periodogram using the real-valued NLT as this avoids mixing lifting schemes and so will generally result in a spectral estimate with less blurring over scale. This is discussed in further detail in Section 6.2.1. Our main interest in the complex lifting method arises in the multivariate setting which we now introduce.

6.2 The bivariate complex nondecimated lifting transform

This section extends the complex nondecimated lifting transform to the bivariate setting in order to estimate the dependence between two irregularly sampled time series. Although it is possible to define bivariate measures using the standard real-valued lifting transform, as demonstrated later, complex methods are preferable as this allows for the estimation of the phase between the series. The proposed bivariate nondecimated lifting transform follows the complex OCAAT scheme, but with two time series instead of one. The two series $X^{(1)}$, $X^{(2)}$ are assumed to contain observations on exactly the same (possibly irregular) time grid x_1, \dots, x_n . Series of this kind are common in palaeoclimatology as ice core samples provide measures of several different

variables.

In the bivariate complex nondecimated lifting transform (bivariate CNLT) the two series are lifted simultaneously, using the same set of trajectories $\{T_p\}_{p=1,\dots,P}$. This ensures that there is an exact correspondence between the coefficients of each series; i.e. for each coefficient of series $X^{(1)}$ there is a coefficient of series $X^{(2)}$ at exactly the same location and scale. The coefficients are lifted according to the complex OCAAT scheme, producing complex valued detail coefficients. The wavelet coefficients at a given stage, n , and trajectory, p , of the bivariate complex OCAAT scheme are denoted

$${}^1d_{j_n}^p = {}^1\lambda_{j_n}^p + {}^1\mu_{j_n}^p i \quad (6.4)$$

$${}^2d_{j_n}^p = {}^2\lambda_{j_n}^p + {}^2\mu_{j_n}^p i \quad (6.5)$$

where the λ are found from the first prediction scheme using filter \mathbf{L} and the μ are found using the second orthogonal prediction scheme with filter \mathbf{M} . For the removal of point j_n on trajectory p , we have $x_{j_n} \rightarrow ({}^1d_{j_n}^p, {}^2d_{j_n}^p, \alpha_{j_n}^p) = ({}^1\lambda_{j_n}^p, {}^1\mu_{j_n}^p, {}^2\lambda_{j_n}^p, {}^2\mu_{j_n}^p, \alpha_{j_n}^p)$.

After repeated application of the complex OCAAT scheme, using the same set of trajectories for both series, we obtain for each time point, x_k , two sets of detail coefficients; $\{{}^1d_{x_k}^p\}_{p=1,\dots,P}$ and $\{{}^2d_{x_k}^p\}_{p=1,\dots,P}$. These coefficients are used to estimate the lifting cross-periodogram, the cross-spectral analogue to the CNLT periodogram introduced in Definition 6.1.1.

Definition 6.2.1. For each time point, x_k , $k \in 1, \dots, n$ defined at scale l^i for $i \in 1, \dots, J^*$, the raw lifting cross-periodogram is given by

$$I_{x_k}^{(1,2)}(l^i) = \text{mean}\{{}^1d_{x_k}^p \overline{{}^2d_{x_k}^p} : \alpha_{x_k}^p \in l^i\},$$

where ${}^1d_{j_n}^p = {}^1\lambda_{j_n}^p + {}^1\mu_{j_n}^p i$ and ${}^2d_{j_n}^p = {}^2\lambda_{j_n}^p + {}^2\mu_{j_n}^p i$ are the complex valued wavelet coefficients obtained through the CNLT scheme.

The CNLT cross-periodogram consists of combinations of coefficients from

each series and provides information about the relationship between the signals. Unlike the CNLT periodogram, its bivariate counterpart is complex-valued. Similarly to the classical Fourier cross-spectrum methodology reviewed in Section 2.3.1, the CNLT cross-periodogram can be separated into its real and imaginary parts to define the CNLT co-periodogram and the CNLT quadrature periodogram. As demonstrated later, these quantities together with the individual lifting spectra of each process, can be used to produce an estimate of the coherence and phase between the two series. Following the methodology from classical Fourier analysis, the CNLT co- and quadrature periodograms are defined as follows:

Definition 6.2.2. For each time point, x_k , $k \in 1, \dots, n$ defined at scale l^i for $i \in 1, \dots, J^*$, the complex NLT co-periodogram and quadrature periodogram are given by

$$\begin{aligned} c_{x_k}(l^i) &= \text{mean}\{\lambda_{x_k}^p \lambda_{x_k}^p : \alpha_{x_k}^p \in l^i\} + \text{mean}\{\mu_{x_k}^p \mu_{x_k}^p : \alpha_{x_k}^p \in l^i\}, \\ q_{x_k}(l^i) &= \text{mean}\{\mu_{x_k}^p \lambda_{x_k}^p : \alpha_{x_k}^p \in l^i\} - \text{mean}\{\lambda_{x_k}^p \mu_{x_k}^p : \alpha_{x_k}^p \in l^i\}. \end{aligned}$$

These estimates, along with the individual periodograms of each series, ${}^1I_{x_k}(l^i)$ and ${}^2I_{x_k}(l^i)$, are smoothed over time using equation (5.4) to give ${}^1\tilde{I}_{x_k}(l^i)$, ${}^2\tilde{I}_{x_k}(l^i)$, $\tilde{c}_{x_k}(l^i)$, $\tilde{q}_{x_k}(l^i)$. The smoothed periodogram quantities are then used to calculate the bivariate measures of cross-periodogram, phase, and coherence:

$$\widehat{CS}_{x_k}(l^i) = \sqrt{\tilde{c}_{x_k}(l^i)^2 + \tilde{q}_{x_k}(l^i)^2}, \quad (6.6)$$

$$\hat{\phi}_{x_k}(l^i) = \tan^{-1} \left(\frac{-\tilde{q}_{x_k}(l^i)}{\tilde{c}_{x_k}(l^i)} \right), \quad (6.7)$$

$$\hat{\rho}_{x_k}(l^i) = \frac{\widehat{CS}_{x_k}(l^i)}{\sqrt{{}^1\tilde{I}_{x_k}(l^i){}^2\tilde{I}_{x_k}(l^i)}}. \quad (6.8)$$

In the regularly sampled setting the smoothing step was essential in form-

ing the coherence, as otherwise the resulting coherence would be one at all points. This is not the case in the bivariate lifting implementation as formation of the lifting periodogram (and related quantities) already incorporates some amount of smoothing due to the discretisation of the continuous scale measure. Choosing to discretise over a smaller number of scales implies a larger amount of smoothing since each artificial scale will contain a larger number of coefficients. Smoothing over time as well as scale serves to reduce the variance of the estimates, providing a clearer visualisation of the characteristics of the data.

The CNLT cross-periodogram provides a measure of the dependence between the series, but its magnitude is affected by the individual CNLT periodograms of the signals. As in the regularly sampled setting, it is therefore preferable to normalise this quantity. We call this normalised quantity the CNLT coherence. The estimated lifting coherence satisfies $0 \leq \hat{\rho}_{x_k}(l^i) \leq 1$. Note that this would not necessarily be the case if the CNLT periodogram and cross periodograms had been constructed using nonparametric regression as in Knight et al. (2010). Modifying the formulation of the periodogram simply to look at the average of the wavelet coefficients at each artificial scale, as in Definition 5.2.1, ensures that by the Cauchy-Schwarz inequality the resulting coherence is between 0 and 1. It is also possible to estimate a measure of phase between the series, providing an indication of any time lag between the signals. This interpretation is considered in further detail using simulated examples in Section 6.2.3.

In the regularly sampled LSW setting the periodogram estimates were corrected before forming the LSW coherence, providing an unbiased estimate of linear dependence. Note that we do not perform any form of correction in the irregularly sampled formulation and so the resulting estimates contain a bias, or ‘blurring effect’, due to the redundancy in the transform. Correction in the univariate framework has been considered by Knight et al. (2010) and provides an interesting possibility for future research.

6.2.1 Relationship to frequency and filter design

As previously, the integral at the last point before it is removed from the algorithm is used to provide a measure of scale for the wavelet coefficients. However, as noted in Section 5.3, different prediction schemes provide different mappings to frequency. Since the complex OCAAT scheme combines two different prediction schemes, this can lead to difficulties when interpreting the CNLT periodogram and cross periodogram.

Consider the case of a sinusoid of a known frequency. Scheme **L** provides the coefficients $|\lambda| = \lambda(\alpha, x)$, and scheme **M** provides the coefficients $|\mu| = \mu(\alpha, x)$, using the notation introduced in Section 5.3. The coefficients from each scheme have different relationships to frequency, given by

$$\begin{aligned} \{\alpha : \bar{\lambda}(\alpha) \geq \bar{\lambda}(\beta), \beta \in \mathbb{R}^+\} &= P_1 \\ \{\alpha : \bar{\mu}(\alpha) \geq \bar{\mu}(\beta), \beta \in \mathbb{R}^+\} &= P_2 \end{aligned}$$

where $\bar{\lambda}(\alpha)$ and $\bar{\mu}(\alpha)$ are the arithmetic averages over x of $\lambda(\alpha, x)$ and $\mu(\alpha, x)$, for each value of α . Here we use the location at which the coefficients peak (i.e. P_1, P_2) to describe the relationship to frequency. The CNLT periodograms and co-periodogram are composed of the sum of the wavelet coefficients from the two schemes, and the quadrature periodogram contains products of the coefficients. This gives a further two relationships to frequency:

$$\begin{aligned} \{\alpha : \overline{\lambda + \mu}(\alpha) \geq \overline{\lambda + \mu}(\beta), \beta \in \mathbb{R}^+\} &= P_3 \\ \{\alpha : \overline{\lambda\mu}(\alpha) \geq \overline{\lambda\mu}(\beta), \beta \in \mathbb{R}^+\} &= P_4 \end{aligned}$$

where $\overline{\lambda + \mu}(\alpha)$ and $\overline{\lambda\mu}(\alpha)$ are the averages over x of $\lambda(\alpha, x) + \mu(\alpha, x)$ and $\lambda(\alpha, x)\mu(\alpha, x)$ for each value of α . This information is summarised in Table 6.1, which considers the possible combinations of the real and imaginary coefficients.

To give a specific example, we consider the choice of prediction filters for

spectral component	coefficients	peak of coefficients
	$\overline{\lambda}(\alpha)$	P_1
	$\overline{\mu}(\alpha)$	P_2
${}_C I_{x_k}, c_{x_k}$	$\overline{\lambda + \mu}(\alpha)$	P_3
q_{x_k}	$\overline{\lambda\mu}(\alpha)$	P_4

Table 6.1: Relationship to frequency for the different bivariate CNLT periodogram components: periodogram (${}_C I_{x_k}$), co-periodogram (c_{x_k}) and quadrature periodogram (q_{x_k}). The cross periodogram, phase and coherence are formed using combinations of these elements. Here we use the value of α that maximises the wavelet coefficients to denote the relationship to frequency.

the one neighbour either side linear prediction scheme as proposed in Section 6.1.1. As described in Section 6.1.1, this scheme gives $P_1 = \frac{P}{2}$, $P_2 = \frac{P}{4}$. The individual periodograms and the co-periodogram are composed of the sum of the wavelet coefficients from the different schemes and show a peak at P_3 . The quadrature periodogram contains products of the coefficients from filters **L** and **M** and peaks at P_4 . This information is illustrated in Figure 6.4, where we see that for this particular scheme $P_3 = P_4 = \frac{P}{3}$. This means that the CNLT periodograms, co-periodogram and quadrature periodogram all show the same relationship between scale and frequency. As shown by equations (6.6)-(6.8), these three quantities are the only components needed to form the cross-periodogram, coherence and phase. Therefore, although for this scheme $P_1 \neq P_2$, the resulting measures of the cross-periodogram, coherence and phase still provide meaningful and interpretable estimates.

Note that this result is dependent on our choice of the weight m_3 in the prediction filter of **M**. Changing the value of m_3 , essentially moves the value of P_4 , as this alters the relative weightings of the real and imaginary coefficients. As illustrated in Figure 6.5, this means that the position of the peak of coefficients associated with the periodograms and co-periodogram will change. For $m_3 \neq \frac{4+1}{\sqrt{3}}$, it follows that $P_3 \neq P_4$, and interpretation of the cross periodogram, coherence and phase becomes unclear.

This example shows that in the bivariate CNLT scheme it is not essential that the real and imaginary coefficients exhibit the same relationship to fre-

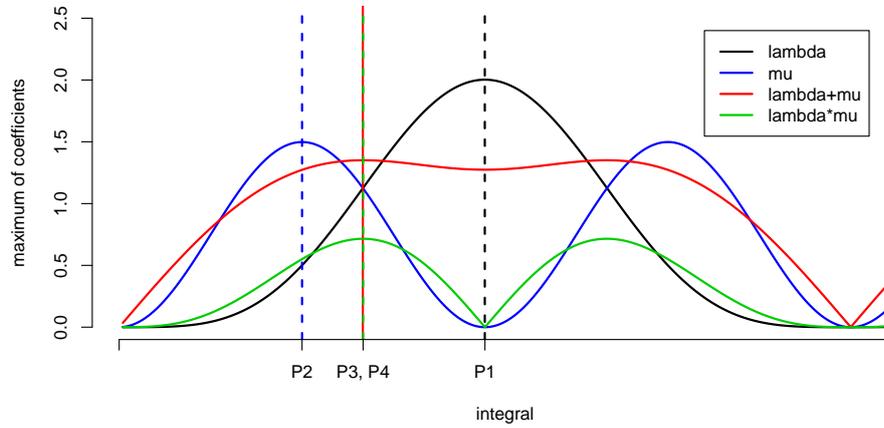


Figure 6.4: Relationship between detail coefficients and integral for $m_3 = \frac{l_1+1}{\sqrt{3}}$. This ensures that combining the coefficients by addition and multiplication results in the same relationship to frequency, as shown by the overlap of the red and green lines.

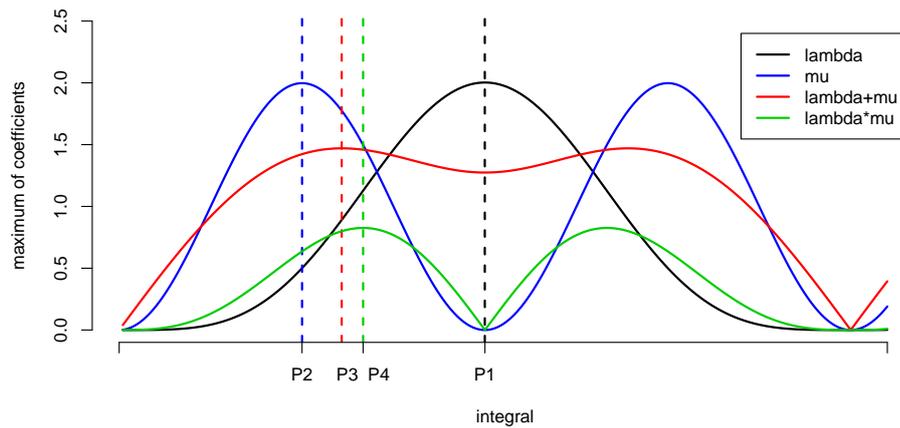


Figure 6.5: Relationship between detail coefficients and integral for $m_3 = 1$. Using this weighting the coefficients $\lambda_{x_k}^p$ and $\mu_{x_k}^p$ have the same maximum value. However combining the coefficients by addition or multiplication leads to different relationships to frequency, as shown by the different positions of the red and green lines in this example.

quency (i.e. it is not essential that $P_1 = P_2$). Instead it is sufficient to ensure that $P_3 = P_4$. This means that the CNLT periodograms, co-periodogram and quadrature periodogram all show the same relationship between scale and frequency and result in meaningful estimates of the cross-periodogram, coherence and phase. However, though it is not essential, the stronger requirement that $P_1 = P_2$ would also be beneficial. If the coefficients from the real and imaginary branches of the complex OCAAT scheme had the same relationship to frequency, this would result in clearer estimates of the CNLT periodogram. Ensuring the same scale-frequency relationship for the two prediction schemes is less practical to implement as it imposes further requirements on the filter design, but research into the design of prediction filters for the complex OCAAT scheme provides another interesting area for further research.

6.2.2 Example on simulated data

To demonstrate the complex bivariate NLT, we now provide an example using the one neighbour either side filters described in Section 6.1.1. The method is applied to a simulated bivariate time series as illustrated in Figure 6.6. The series are observed on an irregular grid with mean sampling rate 2. The first signal consists of a superposition of sine waves with periods 10, 30 and 70, while the second series consists of a sine wave with period 30 which has a time lag of 6 units introduced half way through the series, i.e.

$$\begin{aligned} X_t^{(1)} &= \sin\left(\frac{2\pi t}{10}\right) + \sin\left(\frac{2\pi t}{30}\right) + \sin\left(\frac{2\pi t}{70}\right) + Z_t^{(1)}, \\ X_t^{(2)} &= \sin\left(\frac{2\pi(t-\tau)}{30}\right) + Z_t^{(2)}, \end{aligned}$$

where $\text{diff}(t) \in \{n/10 : n = 10, 11, \dots, 30\}$, $\text{mean}(X_t) = 2$ and $Z_t^{(1)}, Z_t^{(2)}$ are i.i.d normal random variables with mean zero and variance 0.2. The time lag is given by $\tau = 0$ for $t < 200$ and $\tau = 6$ for $t \geq 200$.

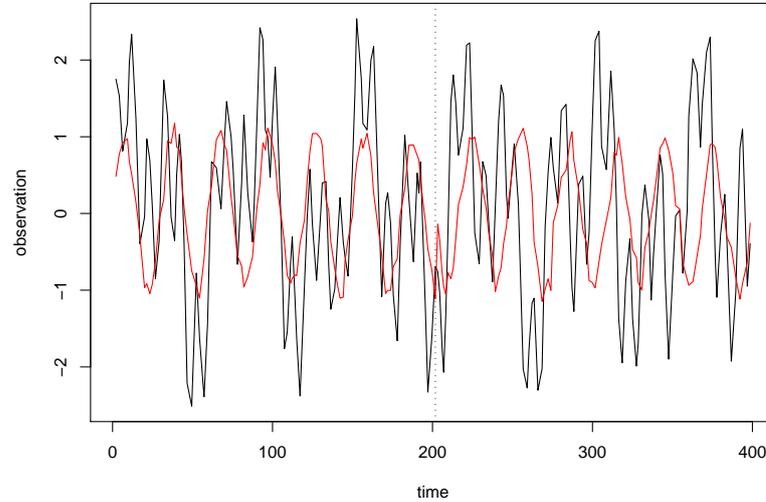


Figure 6.6: Simulated bivariate time series with irregular sampling. $X^{(1)}$ is shown in black and $X^{(2)}$ in red. The vertical line indicates the location where the time lag is introduced.

The coherence and associated quantities, as estimated by the complex bivariate NLT scheme, are shown in Figure 6.8. For these estimates we have used a sample of 2500 randomly sampled trajectories, formed the periodogram using $J^* = 25$ artificial scales, and smoothed over time using a smoothing window of width 60. The coherence provides a clear visualisation of the dependence between the two series, with a peak in the coherence occurring at a scale equivalent to a Fourier period of 30. The time lag that is introduced halfway through the second signal is also clearly captured by the estimate of phase between the series. The phase is approximately zero for the first half of the series, then shows a marked increase for the second half. The individual periodograms of each series are plotted in Figures 6.8a and b. Although the scheme provides an accurate description of the dependence between the series, the individual periodograms do not provide clear descriptions of the variation within each signal. From Figure 6.8a it is not possible to separate the three different periods of variation.

Figure 6.7 shows the coherence between the two series as estimated us-

ing the real-valued bivariate scheme. This is similar to the complex bivariate scheme but without the extra step in the OCAAT lifting procedure that produces the complex-valued coefficients. Although the dependence between the series is clearly demonstrated for the first half of the series, the coherence is not detected after the time delay is introduced. The real-valued scheme can therefore be used to estimate the dependence between two irregularly sampled series, but only if the series are in perfect alignment. Although the real-valued bivariate lifting transform is not ideal for estimating the dependence between series, the individual spectra of each series (subplots a and b) provide a much clearer representation of the series than is provided by the periodogram computed using the complex-valued scheme. This is because the complex NLT periodograms contain coefficients from two different prediction schemes. The prediction schemes have different scale-frequency relationships and so mixing the coefficients results in a blurring over scale.

Also note the difference in the right hand Fourier period axis between the two plots. The wavelet coefficients derived from the complex-valued scheme have a different scale-frequency relationship to those derived from the real-valued scheme.

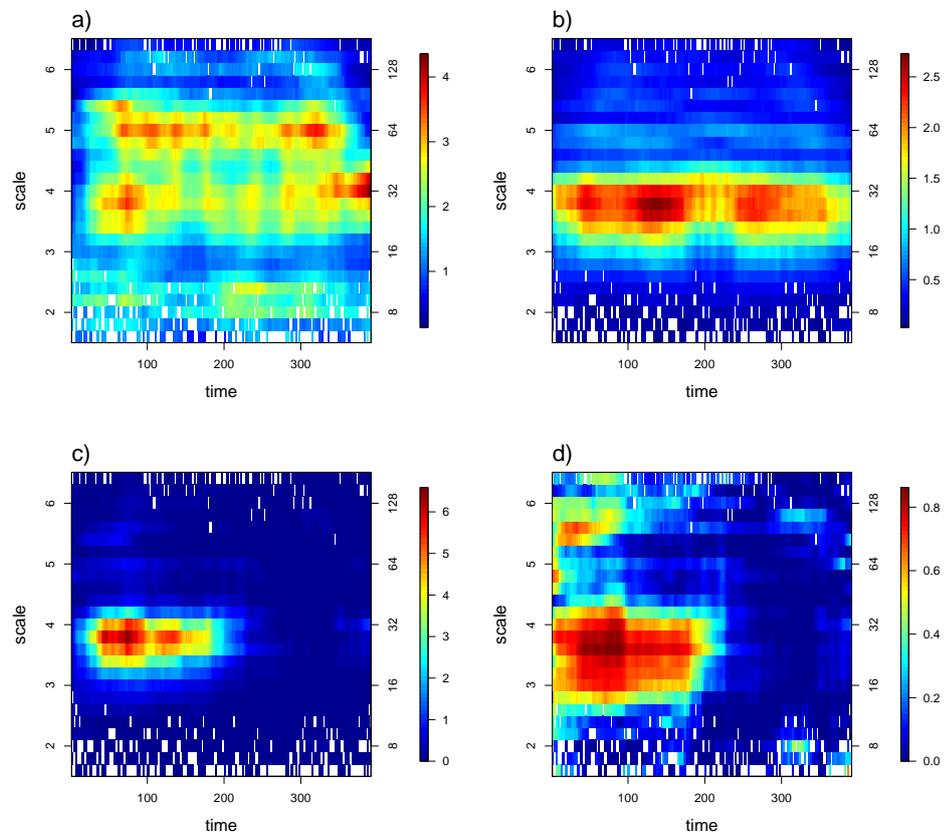


Figure 6.7: Estimation of coherence using the real-valued bivariate lifting scheme. a) periodogram of $X^{(1)}$ b) periodogram of $X^{(2)}$, c) cross periodogram and d) coherence.

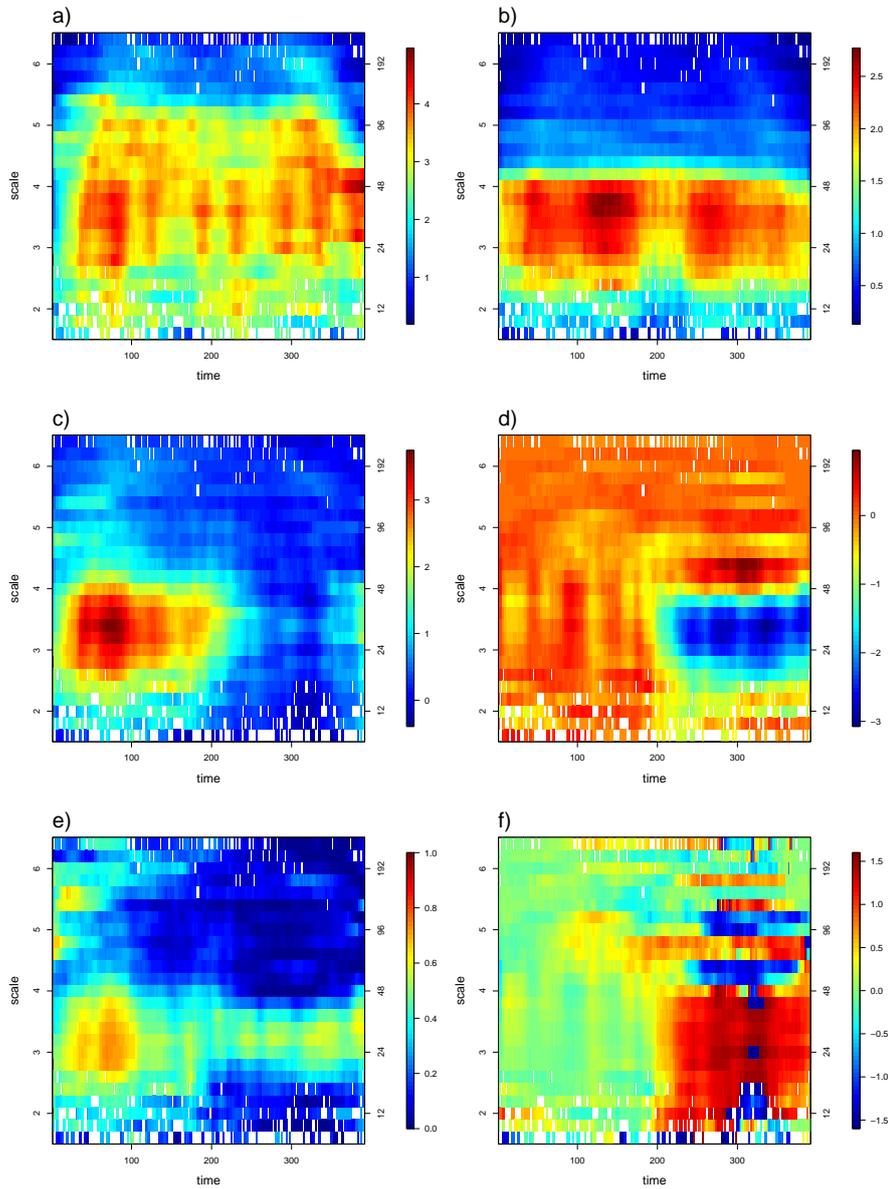


Figure 6.8: Estimation of coherence and phase using the complex lifting scheme. a) periodogram of $X^{(1)}$ b) periodogram of $X^{(2)}$, c) co-periodogram d) quadrature periodogram e) coherence and f) phase.

6.2.3 Physical meaning of phase

This section considers the interpretation of phase, as given by the one neighbour either side filter pairs described in Section 6.1.1. Two simulated examples are provided, illustrating the effect of introducing a time delay between the two observed signals.

Example 1: Purely random time lagged signal

The first example considers a bivariate time series where $X^{(1)}$ is a purely random process and $X^{(2)}$ is a time lagged version of $X^{(1)}$:

$$\begin{aligned} X_t^{(1)} &= Z_t, \\ X_t^{(2)} &= X_{t-\tau}^{(1)}, \end{aligned}$$

where $\text{diff}(t) = 1$ for all t and Z_t are i.i.d normal random variables with mean zero and variance 1.

The coherence and phase are estimated for for time lags of $\tau = 0, 1, 2$ and 5. Since the series are stationary, the estimates are averaged over the full time range and the resulting measures are dependent only on scale, l^i , and the magnitude of the time delay.

The estimated phase and coherence are shown in Figures 6.9 and 6.10. These estimates were produced assuming a time series of length 400, using a sample of 750 random trajectories, and discretising using 10 artificial scales between 0 and 6. For $\tau = 0$ the estimated coherence is 1 and the phase is zero at all scales as expected. After introducing a time delay between the series, the resulting phase shift becomes non-zero and the coherence is no longer 1. The magnitude of the phase shift for each value of τ is a function of scale, with the response varying as τ increases.

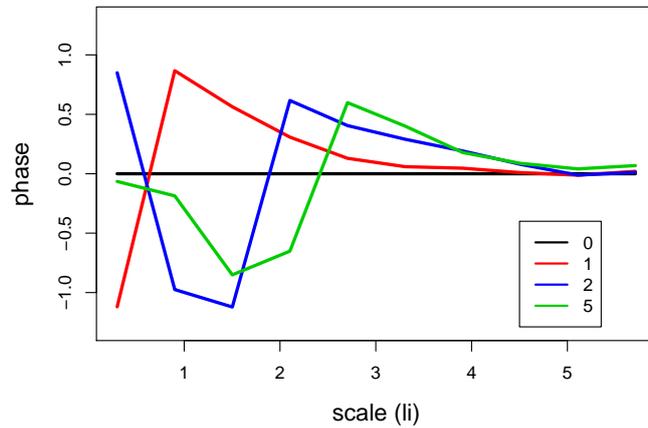


Figure 6.9: Phase shift between $X^{(1)}$ and $X^{(2)}$ from example 1 as a function of scale. The phase is estimated for $\tau = 0, 1, 2, 5$ and averaged over time.

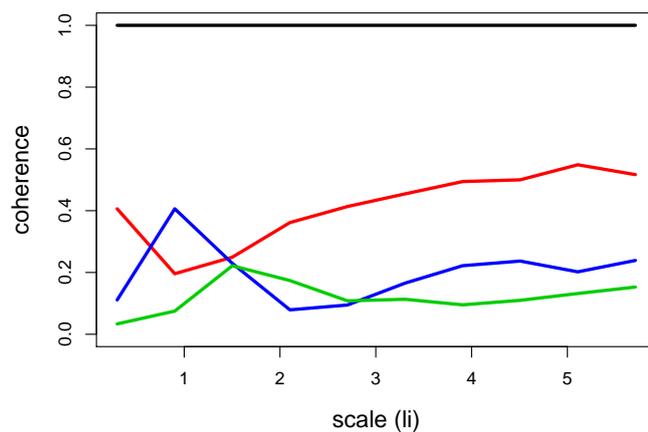


Figure 6.10: Coherence between $X^{(1)}$ and $X^{(2)}$ from example 1 as a function of scale. The coherence is estimated for $\tau = 0, 1, 2, 5$ and averaged over time.

Example 2: Sinusoid with time lag

The next example considers the effect of time lag on the estimation of coherence and phase between two series with periodic variations. The series are given by

$$\begin{aligned} X_t^{(1)} &= \sin\left(\frac{2\pi t}{30}\right), \\ X_t^{(2)} &= \sin\left(\frac{2\pi(t - \tau)}{30}\right), \end{aligned}$$

where $\text{diff}(t) \in \{n/10 : n = 10, 11, \dots, 30\}$, $\text{mean}(X_t) = 2$ and $\tau \in (0, 30)$. As with the previous example, the series are stationary and so we average the resulting estimates over the full time range.

The resulting coherence and phase for time shifts up to 15 are shown in Figures 6.11 and 6.12 respectively. For these results we used a bivariate time series of length 200, a sample of 750 trajectories, and discretised using 20 artificial scales between 1 and 5. For $\tau = 0$ the signals are identical and so the coherence is 1 at all scales, and the phase zero. An overall (summing over time lags) peak in coherence is observed at a scale of $\log_2(30/3)$, corresponding to the period of variation in the data. The magnitude of the coherence, and the response over scale, is affected by the magnitude of the time delay. The coherence is lowest at time delays around 7.5 ($P/4$), and at these shifts the peak at scale $\log_2(30/3)$ is also more pronounced. At a time shift of 15 ($P/2$) the signals are sign reversed versions of each other and, again, the observed coherence is 1 at all scales. The estimated phase also shows an overall peak at scale $\log_2(30/3)$. The phase response varies as a function of time delay and alternates between positive and negative values, with $|\phi(l^i)|$ maximised at $l^i = P/4$. This is also illustrated in Figure 6.13 which shows the phase at scale $\log_2(30/3)$ as a function of time delay.

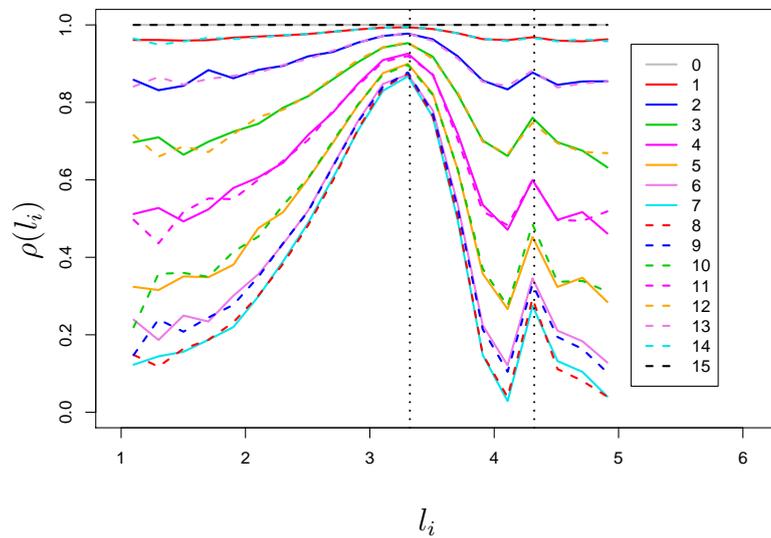


Figure 6.11: Coherence between $X^{(1)}$ and $X^{(2)}$ from example 2 as a function of scale, for various choices of τ . The vertical lines correspond to Fourier periods of 30 and 60.

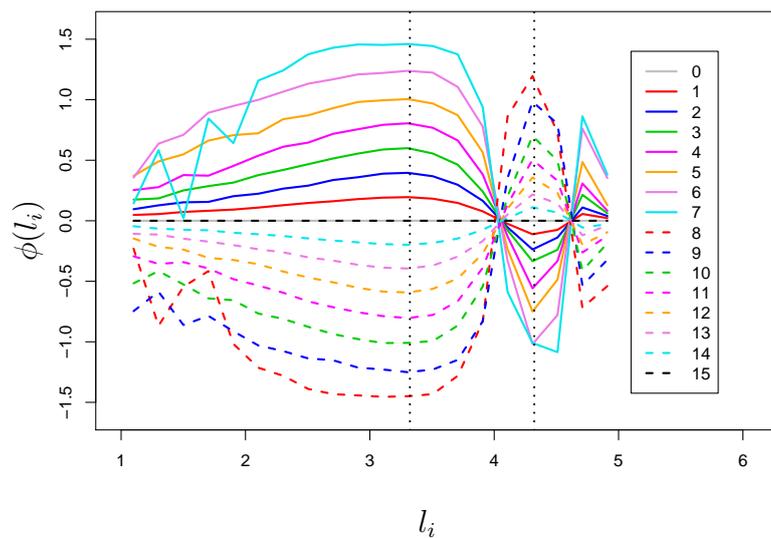


Figure 6.12: Phase between $X^{(1)}$ and $X^{(2)}$ from example 2 as a function of scale, for various choices of τ .

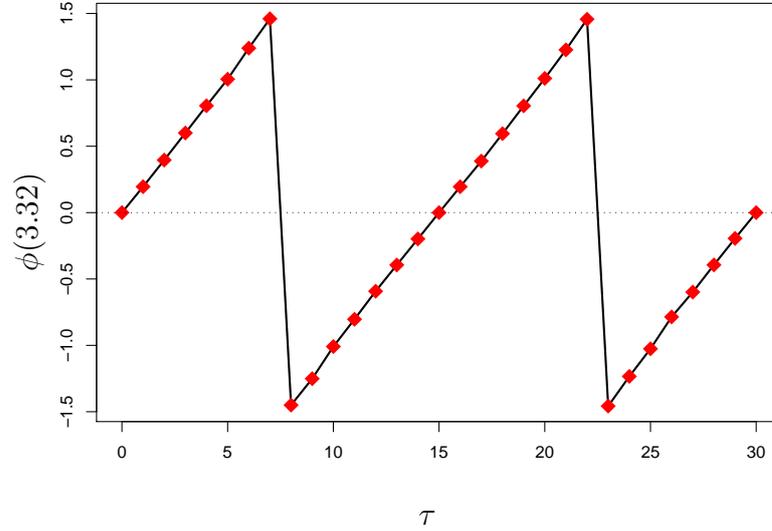


Figure 6.13: Phase at scale $\log_2(30/3)$, equivalent to a Fourier period of 30, as a function of the time delay between the series. The red points show the resulting estimates from the simulations and the black line indicates the observed trend.

6.2.4 Choice of filters

The filters were chosen by specifying the use of linear prediction for the first filter, \mathbf{L} , and then constructing the second filter, \mathbf{M} , to be orthogonal to the first. This ensures that the coefficients from the two prediction schemes provide different local information about the signal.

To justify this choice of filter pairs, we consider what happens as the dependence between the filters increases. Using the same time series as in the simulated example of Section 6.2.2, the coherence and phase between the series are estimated using the filters \mathbf{L} and \mathbf{N} where

$$\mathbf{N} = d_f \mathbf{L} + \sqrt{1 - d_f^2} \mathbf{M}.$$

Here d_f acts as a measure of dependence between the two filters. When $d_f = 0$ we have $\mathbf{N} = \mathbf{M}$ so that the filters are orthogonal, and when $d_f = 1$ we have

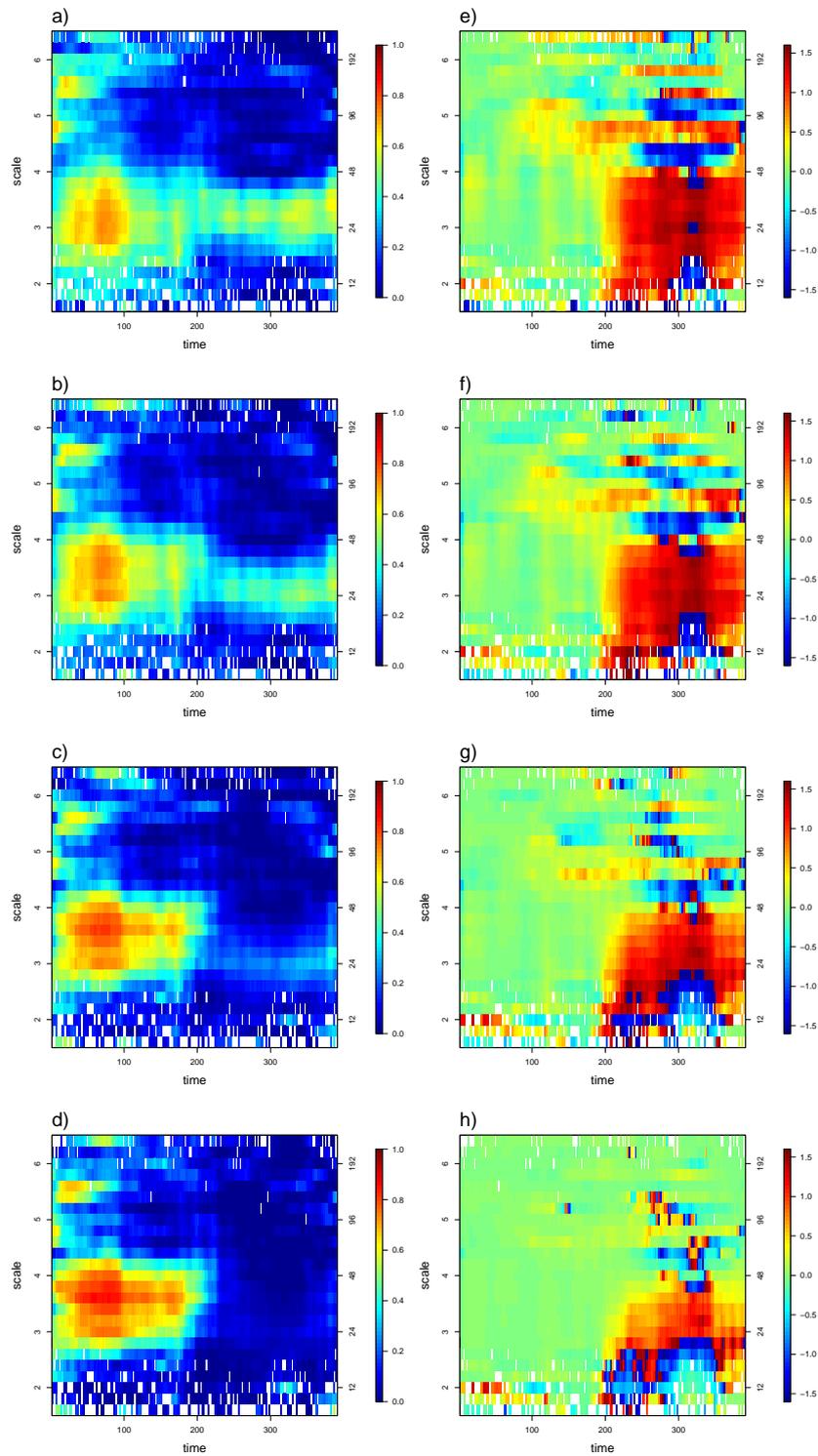


Figure 6.14: Coherence (plots a-d) and phase (plots e-h) for different degrees of dependence between the filters; $d_f = 0$ (plots a and e), $d_f = 0.5$ (plots b and f), $d_f = 0.75$ (plots c and g) and $d_f = 0.9$ (plots d and h).

$\mathbf{N} = \mathbf{L}$ and the second filter is exactly the same as the first. As before, the update stage proceeds using filter \mathbf{L} .

The effects of varying the amount of dependence between the filters on the resulting estimates of phase and coherence are demonstrated with respect to the simulated data example of Section 6.2.2. As with the original example, we have used a sample of 2500 randomly sampled trajectories, formed the periodogram using 25 artificial scales, and smoothed over time using a smoothing window of width 60. The estimates of coherence and phase for different choices of d_f are shown in Figure 6.14. This shows that as the filters become more dependent, the estimates of coherence and phase become less pronounced.

6.3 Application to regularly sampled time series

To compare the methods introduced in this chapter with those described in Chapters 3 and 4, we now consider the application of both methods to a simulated example with regular spacing of observations. The two time series have a nonstationary coherence structure and a time varying phase relationship. The signals are given by

$$\begin{aligned} X^{(1)} &= \sin\left(\frac{2\pi t}{6}\right)\mathbf{I}_{t \in (362, 662)} + \sin\left(\frac{2\pi t}{30}\right) + Z_t^{(1)}, \\ X^{(2)} &= 2\sin\left(\frac{2\pi t}{6}\right)\mathbf{I}_{t \in (362, 662)} + \sin\left(\frac{2\pi(t + \tau)}{30}\right) + Z_t^{(2)}, \end{aligned}$$

where $\tau = 0$ for $t \in (1, 256)$, $\tau = 3$ for $t \in (257, 512)$, $\tau = 6$ for $t \in (513, 768)$ and $\tau = 9$ for $t \in (769, 1024)$. Z_t are i.i.d. normal random variables with

mean zero and variance 1. Both series contain a short burst of variation with period 6, present in the middle section of the data, as well as a variation with period 30. The phase relationship between the two series is nonstationary since τ increases from 0 to 9 over the given time interval. The raw signals are plotted in Figure 6.15. To compare the different methods, Table 6.2 outlines the relationship between the dyadic scales as used in Chapters 3 and 4, Fourier frequency, and wavelet lifting scale. The variation with period 6 is contained within dyadic scale -1 and the variation with period 30 is contained within scale -4 .

The resulting LSW coherence estimate, using the methodology of Chapters 3 and 4, is illustrated in Figure 6.16a. Both the short burst at scale -2 , and the variation at scale -4 are clearly identified. However these results do not provide any phase information about the signals and the coherence at scale -4 decreases as the time delay between the signals increases.

The lifting coherence measure was estimated using a sample of 2000 trajectories and smoothing the resulting periodogram using a smoothing window of width 100. In Figure 6.16b the periodogram was constructed using a discretisation of $J^* = 6$ scales within the chosen scale range, to provide a clear comparison to the LSW coherence estimate. The lifting coherence measure identifies the main features of the data but there is more leakage over scale than shown by the unbiased LSW coherence estimate. However the lifting measure does have the advantage that the estimates are not restricted to dyadic scales. To illustrate this, 6.16c shows the results from the same set of simulations, discretised using 12 artificial scales instead of 6. The lifting methodology also provides an estimate of the phase between the series as shown in Figure 6.16d. The nonstationary phase at coarser scales is clearly captured by this estimate.

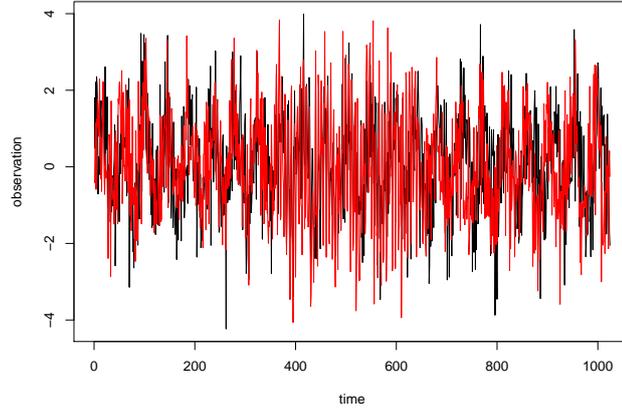


Figure 6.15: Data for simulated example: $X^{(1)}$ (black) and $X^{(2)}$ (red)

wavelet scale j	Fourier frequency $f = 1/P$	wavelet lifting scale	
		$\alpha = \log_2(P/2)$	$\alpha = \log_2(P/3)$
-1	0.250- 0.500	0-1	-0.6-0.4
-2	0.125 - 0.250	1-2	0.4-1.4
-3	0.063 - 0.125	2-3	1.4-2.4
-4	0.031 - 0.063	3-4	2.4-3.4
-5	0.016 - 0.031	4-5	3.4-4.4
-6	0.008 - 0.016	5-6	4.4-5.4
\vdots	\vdots	\vdots	\vdots

Table 6.2: Approximate relationship between wavelet scale, Fourier frequency and wavelet lifting scale. The scale-frequency relationship for the wavelet lifting scheme depends on the chosen prediction method and so we show the relationships for two different lifting schemes; The univariate lifting scheme using linear prediction with one neighbour either side ($\alpha = \log_2(P/2)$), and the bivariate scheme based on this filter ($\alpha = \log_2(P/3)$).

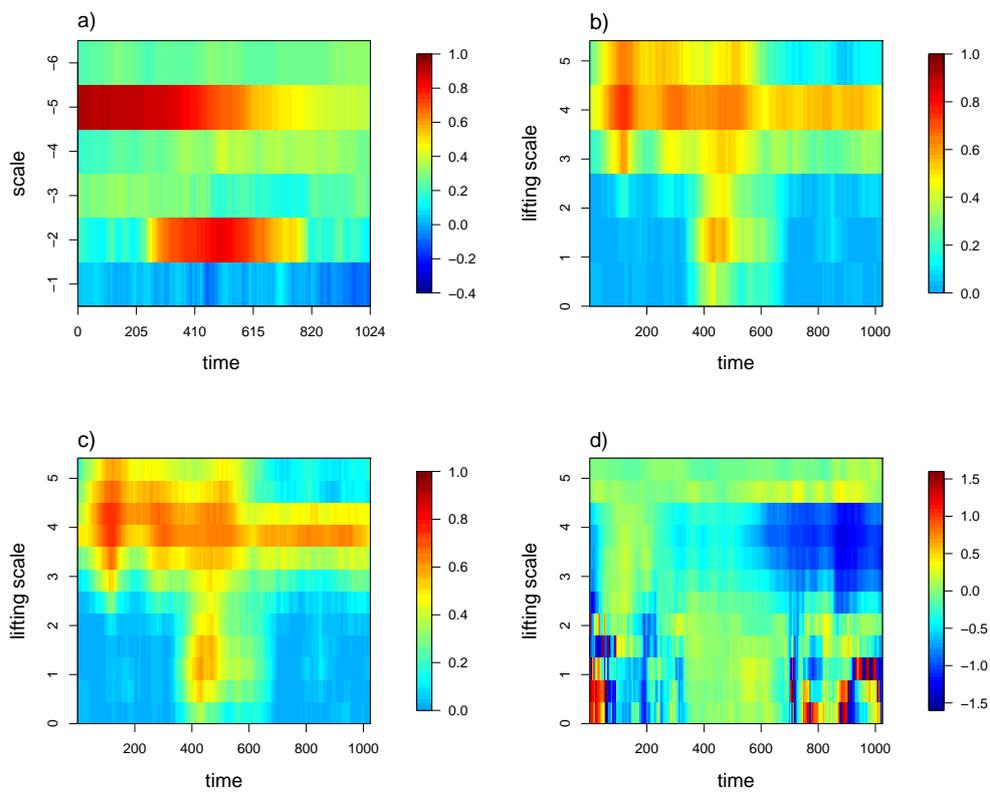


Figure 6.16: Estimated coherence for the data in Figure 6.15. a) LSW coherence b) wavelet lifting coherence using 6 artificial scales c) wavelet lifting coherence using 12 artificial scales d) wavelet lifting phase using 12 artificial scales.

6.4 Application to real data

The bivariate NLT methodology is now demonstrated with application to a stable isotope record from East Antarctica (Stenni et al., 2004). The series is over 40 kyrs long, with measurements from 42.5 yrs BP to 44821 yrs BP. There are 1417 observations in each series, with sampling differences ranging from 8.6 to 50.7 years. The record contains two different isotopic ratios which act as proxies for temperature: $\delta\mathbf{D}$ is a function of the ratio of deuterium (${}^2\mathbf{H}$) to normal hydrogen (${}^1\mathbf{H}$) and $\delta^{18}\mathbf{O}$ is a function of the ratio of ${}^{18}\mathbf{O}$ to ${}^{16}\mathbf{O}$. The two isotopic systems enable independent estimates of temperature, at least in East Antarctica (Stenni et al., 2004). Estimating the coherence and phase between these two temperature proxies provides interesting insights into their roles as climate proxies. Any phase delay between the two signals would be of particular interest.

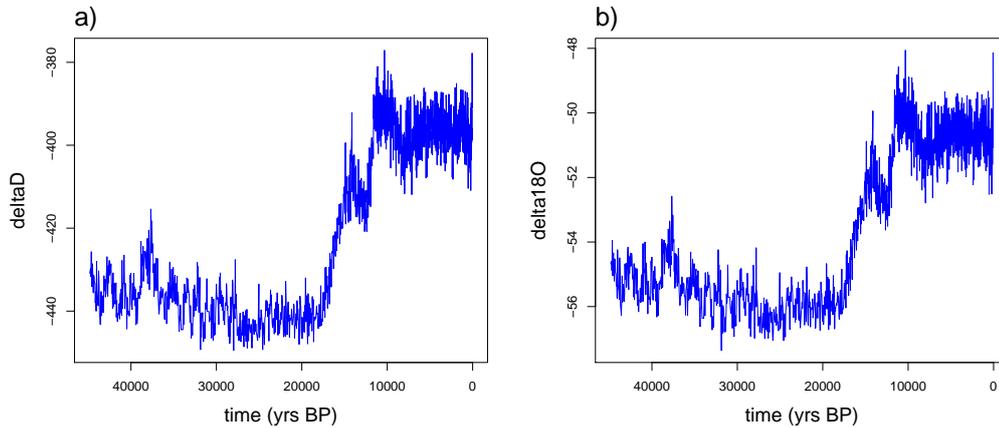


Figure 6.17: Time series of isotopic temperature proxies: a) deuterium ($\delta\mathbf{D}$) and b) oxygen ($\delta^{18}\mathbf{O}$).

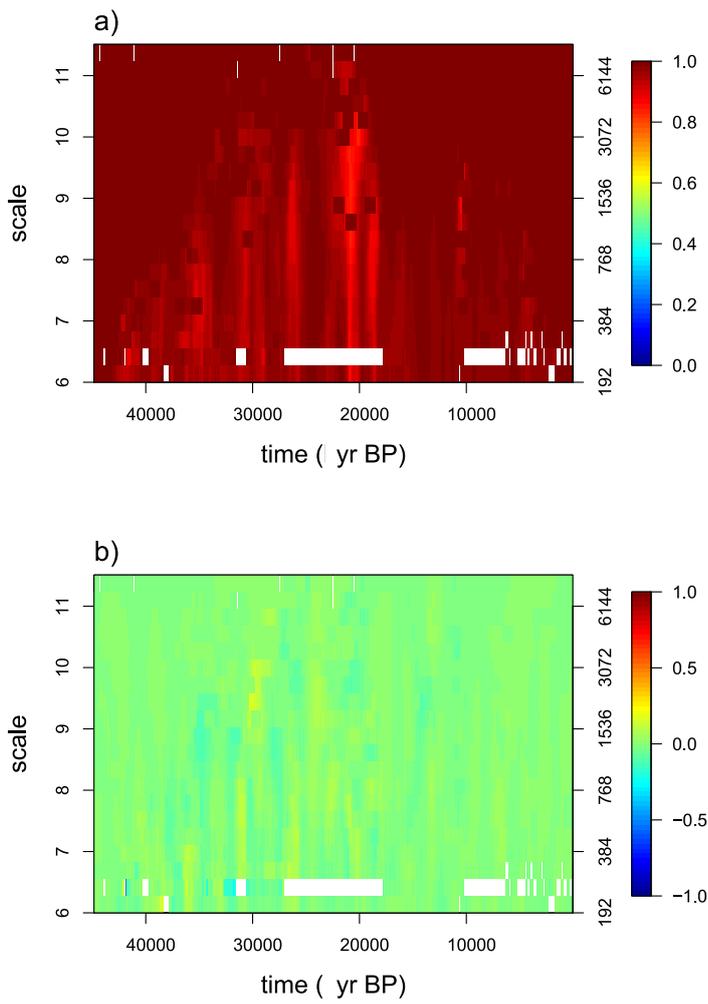


Figure 6.18: Coherence (a) and phase (b) between the $\delta\mathbf{D}$ and $\delta^{18}\mathbf{O}$ series.

The hydrogen and oxygen proxy records are shown in Figures 6.17a and 6.17b. The dependence between the two proxies is obvious from these plots, with both series showing the same overall trend and characteristics. The estimated coherence and phase between the two series is shown in Figure 6.18. For these estimates we have used a sample of 5000 trajectories, formed the periodogram using a discretisation of $J^* = 20$ artificial scales, and smoothed over time using a smoothing window of 1 kyr. The coherence between the series is close to one at all time points, with a mean of 0.99 and range of 0.83 to 1.00. The phase has a mean of 0.00 and a range of -0.21 to 0.376 . This suggests that there is no time delay in the observations of the two different temperature proxies.

No periodic variations were detected in the individual spectra of the series. Because their length is just 40 kyrs, the series do not capture the documented 100 kyr saw tooth trend exhibited by climate records. The longest period of variation captured by this data is shown by Figure 6.17 to be around 6 kyrs and so this series is unable to capture even the shortest of the orbital forcing signals, the precession cycle which has a period of 21 kyrs.

6.5 Summary and further work

In this chapter we extended the standard OCAAT lifting scheme to provide complex valued coefficients. The complex OCAAT scheme was used in the NLT framework to provide a complex-valued NLT transform. As with the standard NLT transform, the complex version can be used to construct a periodogram for data observed on irregular sampling grids. The complex NLT transform was also implemented in a bivariate lifting setting, allowing us to define the complex NLT cross-periodogram which provides a measure of the dependence between the two series. The NLT cross-periodogram was used to define two further measures of dependence: the coherence and phase. Note that this methodology is also applicable in a multivariate setting with more

than two time series. In this case the dependence can be assessed by taking pairwise combinations of the signals.

Our methodology was demonstrated with respect to simulated data as well as observed palaeoclimate records. With respect to the simulated data, we showed that it is also possible to estimate the coherence between the series using a real-valued lifting scheme. The real-valued bivariate lifting scheme provided informative estimates of the coherence between the series when there was no time delay, but all information on the dependence between the series was lost when the signals were not in perfect alignment. Here the complex valued scheme showed a clear improvement. The coherence between the series was estimated accurately both before and after the introduction of a time shift between the signals. Furthermore, the time shift was clearly captured by the estimated phase spectrum.

The work of this chapter provides promising results on estimating the dependence between series observed on an irregular grid; however there is still much scope for further research and improvement. One limitation of our methodology is highlighted by the problems encountered when estimating the univariate complex NLT periodograms. The periodograms contain coefficients from two different prediction schemes which have different scale-frequency relationships. This results in increased blurring over scale in the periodograms. It would be preferable to implement a dual lifting scheme in which both branches provide coefficients with the same scale-frequency relationship. The issue of designing suitable filter pairs for the complex scheme presents an important area for further research.

So far we have considered bivariate series that are irregularly sampled, but we have been assuming that the irregular sampling grid is exactly the same for each signal. Although this is a valuable tool in many situations, it would also be useful to compare series that are observed on *different* irregular grids. For example, it would be interesting to compare records from different sites. The sampling rate of an observed ice core is dependent on many different variables

(e.g. snow accumulation rate, extraction method) and, as such, the observation grids from two different cores will never be the same. Preliminary work in this area is positive, suggesting that it is possible to extend the methodology presented in this section to estimate the dependence between two series observed on separate irregular grids. Although it is beyond the scope of our current work, this provides another interesting area for future study.

Chapter 7

Conclusions

This thesis has considered the application of wavelet based methods for estimating the characteristics of nonstationary time series observed on both regular and irregular sampling grids and in the presence of bivariate observations. Three different situations are considered:

- Bivariate (Chapters 3 and 4)
- Irregularly spaced (Chapter 5)
- Bivariate and irregularly spaced (Chapter 6)

Chapters 3 and 4 introduce the locally stationary wavelet coherence which, based on the NDWT, provides a method for estimating the dependence between two regularly sampled time series. Chapter 5 demonstrates how, based on the NLT, we can estimate the spectral properties of irregularly sampled time series. Finally, Chapter 6 combines the methodology of the previous chapters, using the NLT in a bivariate setting to estimate the dependence between two time series observed on irregular sampling grids. Chapter 6 therefore provides a generalisation to the work of Chapters 3 and 4. To conclude, we now briefly summarise the main themes and findings from each chapter and discuss some directions for future work. Further discussions of each element of work are

provided by the individual summaries at the end of each chapter.

Chapter 3 addresses the problem of estimating the dependence between two nonstationary time series. This work was originally motivated by a problem in neuroscience in which recordings are taken from two different areas of the brain. In this setting the dependence between the series provides information on the interaction between the brain regions.

In order to quantify this interaction, we introduced the bivariate LSW process model. This parallels the univariate LSW model of Nason et al. (2000), but allows for the potential correlation structure between the two series, as given by the locally stationary wavelet coherence. The LSW coherence is a novel measure of the dependence between two time series and we proposed an estimation procedure for this quantity. In Chapter 4 it was seen that the bias correction implied by the LSW formulation can lead to stability problems with the estimator, and a robust method for ensuring the stability of the final measure was proposed. Our methodology was applied to the original problem from neuroscience, illustrating the practical use of the described techniques.

In its current form the bivariate LSW process model requires that the cross-covariance between the two processes is maximised at a shift of $\tau = 0$, and that the two signals are therefore aligned. This is not the case if there is a time shift between the signals, and so extension of the current formulation to allow for time delays between the signals is an important topic for future work.

Chapter 5 considers the analysis of palaeoclimate signals that, unlike the time series generated by the neuroscience experiment described in Chapters 3 and 4, have a naturally irregular sampling structure. Although we concentrate on the analysis of palaeoclimate signals, this is also an important problem in other applications, including experimental results that contain missing observations.

The first section of Chapter 5 reviews the work of Knight et al. (2010) in using the nondecimated lifting transform (based on the OCAAT lifting scheme

of Jansen et al. (2009)) to estimate the periodogram of irregularly sampled signals. The resulting NLT periodogram provides an analogue to the NDWT periodogram for regularly sampled data that was used in Chapters 3 and 4. We then explore the relationship between scale (as defined by the OCAAT lifting scheme) and Fourier frequency, which is important for the practical application of these techniques. It is also shown that the use of the lifting integral as a measure of scale can lead to a blurring in the resulting periodogram.

The use of empirical mode decomposition (EMD), another function decomposition method suitable for irregularly spaced time series, is also considered. Through the use of EMD we are able to draw attention to some of the limitations of the NLT periodogram method, in that small amplitude signals may be lost in the resulting periodogram due to the blurring over scale. This introduces an important area for further work, to produce periodogram estimates with less leakage over scale. Knight et al. (2010) propose a correction scheme for the raw periodogram estimates based on the formulation of the problem as an LSW process with missing observations; however this methodology is computationally intensive and so at this time we are unable to implement these techniques in our chosen application.

Chapter 6 extends the methodology introduced in Chapter 5 to estimate the dependence and phase between two signals observed on the same irregularly sampled grids. This is interesting in our chosen applications as palaeoclimate proxies typically contain information on more than one variable.

In order to provide an estimate of the phase between the signals, we first propose an extension to the OCAAT lifting scheme (Jansen et al., 2009) that was used in Chapter 5, to produce complex-valued wavelet coefficients at each stage. The complex OCAAT scheme is then implemented in a bivariate lifting framework to produce the complex NLT cross-periodogram. Analogous to the wavelet cross-periodogram introduced in Chapter 3, the NLT cross-periodogram provides information about the dependence between the series

but without relying on the regular spacing of observations. This measure, along with the individual periodograms of each series, was used to define a measure of coherence and phase between the signals. The bivariate lifting methodology was then demonstrated with respect to simulated examples as well as observed palaeoclimate records. We also compared the methods of this chapter with those introduced in Chapters 3 and 4 using a simulated example with regular sampling, highlighting the strengths and weaknesses of each technique. The LSW coherence estimate showed less blurring over scale than the bivariate lifting coherence estimate but suffered from the lack of phase information.

The work of this chapter provides promising results on estimating the dependence between series observed at irregular sampling intervals, but there is still scope for further research and improvement. One possible area for future research is the design of suitable filters for the prediction stage of the complex OCAAT scheme. In particular, the use of two prediction schemes with the same scale-frequency relationship would result in estimates with an improved localisation of characteristics. Also, we have considered bivariate series that are irregularly sampled, but we have been assuming that the irregular sampling grid is exactly the same for each signal. Although this is a valuable tool in many situations, it would also be useful to compare series that are observed on separate irregular grids.

Bibliography

- Abramovich, F., Bailey, T., and Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *Journal of the Royal Statistical Society. Series D*, 49(1):1–29.
- Antoniadis, A. (1999). Wavelets in statistics: a review. *Journal of the Italian Statistical Society*, pages 97–144.
- Antoniadis, A. (2007). Wavelet methods in statistics: some recent developments and their applications. *Statistics Surveys*, 1:16–55.
- Antoniadis, A., Gregoire, G., and Nason, G. (1999). Density and hazard rate estimation for Right-Censored data by using wavelet methods. *Journal of the Royal Statistical Society. Series B*, 61(1):63–84.
- Augustin, L., Barbante, C., Barnes, P., Barnola, J., Bigler, M., Castellano, E., Cattani, O., Chappellaz, J., Dahl-Jensen, D., Delmonte, B., et al. (2004). Eight glacial cycles from an Antarctic ice core. *Nature*, 429(6992):623–628.
- Barber, S. and Nason, G. (2004). Real nonparametric regression using complex wavelets. *Journal of the Royal Statistical Society. Series B*, 66(4):927–939.
- Belzer, B., Lina, J., and Villasenor, J. (1995). Complex, linear-phase filters for efficient image coding. *IEEE transactions on Signal Processing*, 43(10):2425–2427.
- Berger, A. (1978). Long-term variations of daily insolation and Quaternary climatic changes. *Journal of the Atmospheric Sciences*, 35(12):2362–2367.

- Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*. Holf, Rinehart and Winston.
- Brockwell, P. and Davis, R. (2006). *Time Series: Theory and Methods*. Springer.
- Broerson, P. M. T. (2008). Time series models for spectral analysis of irregular data far beyond the mean data rate. *Measurement Science and Technology*, 19:1–13.
- Chatfield, C. (2003). *The Analysis of Time Series*. Chapman and Hall.
- Cho, H. and Fryzlewicz, P. (2010). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *In submission*.
- Chui, C. K. (1997). *Wavelets: a Mathematical Tool for Signal Analysis*. Philadelphia: Society for Industrial and Applied Mathematics.
- Cohen, A., Daubechies, I., and Feauveau, J. C. (1992). Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(2):485–560.
- Coifman, R. R. and Donoho, D. (1995). Translation-invariant de-noising. Technical report, Statistics Department, Stanford University.
- Coifman, R. R. and Wickerhauser, M. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718.
- Crucifix, M. (2008). Global change: Climate’s astronomical sensors. *Nature*, 456:47–48.
- Crucifix, M., Loutre, M., and Berger, A. (2006). The climate reponse to the astronomical forcing. *Space science reviews*, 454:903–995.

- Crucifix, M. and Rougier, J. (2009). On the use of simple dynamical systems for climate predictions. *The European Physical Journal-Special Topics*, 174(1):11–31.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The annals of Statistics*, 25(1):1–37.
- Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Annals of Statistics*, pages 1762–1794.
- Dahlhaus, R. and Polonik, W. (2002). Empirical spectral processes and nonparametric maximum likelihood estimation for time series. *Empirical Process Techniques for Dependent Data*, pages 275–298.
- Dahlhaus, R. and Polonik, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *The Annals of Statistics*, 34(6):2790–2824.
- Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39.
- Daubechies, I. (1992). *Ten Lectures On Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics.
- Daubechies, I. and Sweldens, W. (1998). Factoring wavelet transforms into lifting steps. *Journal of Fourier Analysis and Applications*, 4(3):247–269.
- Davies, P. and Kovac, A. (2004). Densities, spectral densities and modality. *Annals of Statistics*, pages 1093–1136.
- Davis, R., Lee, T., and Rodriguez-Yam, G. (2006). Structural Break Estimation for Nonstationary Time Series Models. *Journal of the American Statistical Association*, 101:223–239.
- Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

- Donoho, D. L., Johnstone, I. M., and Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics*, 24(2):508–539.
- Eckley, I. A. and Nason, G. P. (2005). Efficient computation of the discrete autocorrelation wavelet inner product matrix. *Statistics and Computing*, 15(2):83–92.
- Flandrin, P., Rilling, G., and Goncalves, P. (2004). Empirical mode decomposition as a filter bank. *IEEE signal processing letters*, 11(2):112–114.
- Folland, G. B. (2009). *Fourier Analysis and Its Applications*. AMS Bookstore.
- Foster, G. (1996). Wavelets for period analysis of unevenly sampled time series. *The Astronomical Journal*, 112(4):1709–1729.
- Frick, P., Grossman, A., and Tchamitchian, P. (1998). Wavelet analysis for signals with gaps. *Journal of Mathematical Physics*, 39:4091–4107.
- Fryzlewicz, P., Nason, G., and von Sachs, R. (2008). A wavelet-Fisz approach to spectrum estimation. *Journal of Time Series Analysis*, 29(5):868–880.
- Fryzlewicz, P. and Nason, G. P. (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society: Series B*, 68:611–634.
- Fryzlewicz, P. and Ombao, H. (2009). Consistent Classification of Nonstationary Time Series Using Stochastic Wavelet Representations. *Journal of the American Statistical Association*, 104(485):299–312.
- Fryzlewicz, P., Van Bellegem, S., and von Sachs, R. (2003). Forecasting nonstationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, 55(4):737–764.
- Gabor, D. (1946). Theory of communication. *J. Inst. Elec. Eng.*, 93:429–457.

- Girardi, M. and Sweldens, W. (1997). A new class of unbalanced Haar wavelets that form an unconditional basis for L_p on general measure spaces. *Journal of Fourier Analysis and Applications*, 3(4):457–474.
- Grinsted, A., Moore, J., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Non-linear Processes in Geophysics*, 11(5/6):561–566.
- Haar, A. (1910). Zur theorie der orthogonalen funktionen-systeme. *Mathematische Annalen*, 69(4):331–371.
- Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Annals of statistics*, 23(3):905–928.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.
- Huang, N., Shen, Z., and Long, S. (1999). A new view of nonlinear water waves: the hilbert spectrum. *A. Rev. Fluid Mech*, 31:417–457.
- Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., and Tung, C. (1998). The empirical mode decomposition and hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc. London A*, 454:903–995.
- Huang, N., Wu, M., Long, S., Shen, S., Qu, W., Gloersen, P., and Fan, K. (2003). A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc. of Roy. Soc. London A*, 459(2037):2317–2345.
- Hubbard, B. B. (1996). *The World According to Wavelets*. A. K. Peters, Wellesley, Mass.
- Hudgins, L., Friehe, C., and Mayer, M. (1993). Wavelet transforms and atmospheric turbulence. *Physical Review Letters*, 71(20):3279–3282.

- Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12:134–139.
- Jansen, M., Nason, G., and Silverman, B. (2001). Scattered data smoothing by empirical bayesian shrinkage of second generation wavelet coefficients. In Unser, M. and Aldroubi, A., editors, *Wavelet Applications in Signal and Image Processing IX*, volume 4478, pages 87–97. SPIE.
- Jansen, M., Nason, G., and Silverman, B. (2004). Multidimensional nonparametric regression using lifting. Technical Report 04:17, Statistics Group, Department of Mathematics, University of Bristol, UK.
- Jansen, M., Nason, G. P., and Silverman, B. W. (2009). Multiscale methods for data on graphs and irregular multidimensional situations. *Journal Of The Royal Statistical Society Series B*, 71(1):97–125.
- Jones, M. W. and Wilson, M. A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biology*, 3:2187–2199.
- Kijewski-Correa, T. and Kareem, A. (2007). Performance of wavelet transform and empirical mode decomposition in extracting signals embedded in noise. *Journal of Engineering Mechanics*, 133:849.
- Kim, D. and Oh, H. (2009). EMD: A package for empirical mode decomposition and hilbert spectrum. *The R Journal*, 1:40–46.
- Kingsbury, N. (2001). Complex wavelets for shift invariant analysis and filtering of signals. *Applied and Computational Harmonic Analysis*, 10(3):234–253.
- Kingsbury, N. G. (1998). The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters. In *IEEE Digital Signal Processing Workshop, DSP*, pages 319–322.

- Knight, M. (2006). *A second generation wavelet construction and applications to regression and time series analysis*. PhD thesis, University of Bristol.
- Knight, M. (2008). *NLT: A nondecimated lifting transform for signal denoising*. R package version 2.0-1.
- Knight, M. I. and Nason, G. (2008). A ‘nondecimated’ lifting transform. *Statistics and Computing*, 19:1–16.
- Knight, M. I., Nunes, M., and Nason, G. (2010). Spectral estimation for locally stationary time series with missing observations. *In preparation*.
- Lachaux, J., Lutz, A., Rudrauf, D., Cosmelli, D., Le Van Quyen, M., Martinerie, J., and Varela, F. (2002). Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. *Neurophysiologie Clinique/Clinical Neurophysiology*, 32(3):157–174.
- Lawson, C. L. and Hanson, B. J. (1974). *Solving Least Squares Problems*. Prentice-Hall.
- Lawton, W. (1993). Applications of complex valued wavelet transforms to sub-band decomposition. *IEEE Transactions on Signal Processing*, 41(12):3566–3568.
- Lina, J. and Mayrand, M. (1995). Complex daubechies wavelets. *Applied and Computational Harmonic Analysis*, 2(3):219–229.
- Lomb, N. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39(2):447–462.
- Lüthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J.-M., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K., and Stocker, T. (2008). High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature*, 453:379–382.

- Mallat, S. et al. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693.
- Maraun, D. and Kurths, J. (2004). Cross wavelet analysis: significance testing and pitfalls. *Nonlinear Processes in Geophysics*, 11:505–514.
- McGuffie, K. and Henderson-Sellers, A. (2005). *A climate modelling primer*. Wiley.
- Mélice, J., Coron, A., and Berger, A. (2001). Amplitude and frequency modulations of the Earth’s obliquity for the last million years. *Journal of Climate*, 14(6):1043–1054.
- Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press.
- Milankovitch, M. (1969). *Canon of insolation and the ice-age problem*. Israel Program for Scientific Translations.
- Mondal, D. and Percival, D. (2008). Wavelet variance analysis for gappy time series. *Annals of the Institute of Statistical Mathematics*, pages 1–24.
- Murty, K. (1988). *Linear Complementarity, Linear and Nonlinear Programming*. Internet edition.
- Nason, G. (1996). Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 463–479.
- Nason, G. (2008). *Wavelet methods in statistics with R*. Springer.
- Nason, G., Kovac, A., and Maechler, M. (2008). *Wavethresh: Software to perform wavelet statistics and transforms*. R package version 4.2-1.
- Nason, G. and Silverman, B. (1994). The discrete wavelet transform in S. *Journal of Computational and Graphical Statistics*, 3(2):163–191.

- Nason, G. and Silverman, B. (1995). The stationary wavelet transform and some applications. In Antoniadis, A. and Oppeheim, G., editors, *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pages 281–300. New York: Springer-Verlag.
- Nason, G. and von Sachs, R. (1999). Wavelets in time-series analysis. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 357(1760):2511–2526.
- Nason, G., von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(2):271–292.
- Nunes, M. and Knight, M. (2008). *Adlift: An adaptive lifting scheme algorithm*. R package version 1.1-1.
- Nunes, M., Knight, M., and Nason, G. (2006). Adaptive lifting for nonparametric regression. *Statistics and Computing*, 16(2):143–159.
- Ombao, H., Raz, J., Sachs, R., and Malow, B. (2001). Automatic Statistical Analysis of Bivariate Nonstationary Time Series in Memory of Jonathan A. Raz. *Journal of the American Statistical Association*, 96(454).
- Ombao, H., Raz, J., von Sachs, R., and Guo, W. (2002). The SLEX model of a Non-Stationary random process. *Annals of the Institute of Statistical Mathematics*, 54(1):171–200.
- Ombao, H. and Van Bellegem, S. (2008). Evolutionary coherence of nonstationary signals. *IEEE Transactions on Signal Processing*, 56(6):2259–2266.
- Ombao, H., von Sachs, R., and Guo, W. (2005). SLEX Analysis of Multivariate Nonstationary Time Series. *Journal of the American Statistical Association*, 100(470):519–532.

- Pegram, G., Peel, M., and McMahon, T. (2008). Empirical mode decomposition using rational splines: an application to rainfall time series. *Proceedings of the Royal Society A*, 464(2094):1483.
- Percival, D. B. and Walden, A. T. (2000). *Spectral Analysis of Time Series. Wavelet Methods for Time Series Analysis* (Cambridge Series in Statistical and Probabilistic Mathematics).
- Petit, J., Jouzel, J., Raynaud, D., Barkov, N., Barnola, J.-M., Basile, I., Benders, M., Chappellaz, J., Davis, M., Delayque, G., Delmotte, M., Kotlyakov, V., Legrand, M., Lipenkov, V., Lorius, C., Ppin, L., Ritz, C., Saltzman, E., and Stievenard, M. (1999). Climate and atmospheric history of the past 420,000 years from the vostok ice core, Antarctica. *Nature*, 399:429–436.
- Priestley, M. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(2):204–237.
- Priestley, M. (1981). *Spectral Analysis of Time Series*. Academic Press.
- Priestley, M. B. (1988). *Nonlinear and Nonstationary Time Series*. Academic Press Inc.
- Priestley, M. B. and Tong, H. (1973). On the analysis of bivariate Non-Stationary processes. *Journal of the Royal Statistical Society. Series B*, 35(2):153–166. ArticleType: primary_article / Full publication date: 1973 / Copyright 1973 Royal Statistical Society.
- Qi, Y., Minka, T., and Picard, R. (2002). Bayesian spectrum estimation of unevenly sampled nonstationary data. In *IEEE international conference on acoustics speech and signal processing*, volume 2. Citeseer.
- Rilling, G. and Flandrin, P. (2009). Sampling effects on the empirical mode decomposition. *Advances in Adaptive Data Analysis*, 1(1):43–59.

- Rilling, G., Flandrin, P., and Gonçalves, P. (2003). On empirical mode decomposition and its algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing 3 NSIP-03*.
- Rilling, G., Flandrin, P., Gonçalves, P., and Lilly, J. (2007). Bivariate empirical mode decomposition. *IEEE Signal Processing Letters*, 14(12):936.
- Roes, G. and Allen, M. (1999). A comparison of competing explanations for the 100, 000-yr ice age cycle. *Geophysical research letters*, 26(15):2259–2262.
- Sanderson, J., Fryzlewicz, P., and Jones, M. W. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, doi: 10.1093/biomet/asq007.
- Scargle, J. (1982). Studies in astronomical time series analysis. II- Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853.
- Selesnick, I., Baraniuk, R., and Kingsbury, N. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151.
- Serroukh, A. and Walden, A. (2001). Wavelet scale analysis of bivariate time series I: motivation and estimation. *Journal of Nonparametric Statistics*, 13(1):1–36.
- Shumway, R. and Stoffer, D. (2000). *Time series analysis and its applications*. Springer Verlag.
- Slutsky, E. (1925). Über stochastische Asymptoter und Grenzwerte. *Metron*, 5:3–89.
- Stade, E. (2005). *Fourier analysis*. Wiley-Interscience Hoboken, NJ.
- Stenni, B., Jouzel, J., Masson-Delmotte, V., Röthlisberger, R., Castellano, E., Cattani, O., Falourd, S., Johnsen, S., Longinelli, A., Sachs, J., et al.

- (2004). A late-glacial high-resolution site and source temperature record derived from the EPICA Dome C isotope records (East Antarctica). *Earth and Planetary Science Letters*, 217(1-2):183–195.
- Stoica, P. and Sandgren, N. (2006). Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches. *Digital signal processing*, 16:712–734.
- Sweldens, W. (1996). The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis*, 3(2):186–200.
- Van Bellegem, S. and Dahlhaus, R. (2006). Semiparametric estimation by model selection for locally stationary processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(5):721–746.
- Van Bellegem, S. and von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Ann. Statist.*, 36(4):1879–1924.
- Varela, F., Lachaux, J., Rodriguez, E., and Martinerie, J. (2001). The brain-web: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4):229–239.
- Vidakovic, B. (1999). *Statistical modelling by wavelets*. Wiley: New York.
- von Sachs, R., Nason, G., and Kroisandt (1997). Adaptive estimation of the evolutionary wavelet spectrum. Technical Report 516, Dept. Statistics, Stanford Univ.
- Walter, G. G. and Shen, X. (2005). Wavelet like behavior of slepian functions and their use in density estimation. *Communications in statistics. Theory and methods*, 34(3):687–711.
- Wang, Y., Cavanaugh, J., and Song, C. (2001). Self-similarity index estimation

- via wavelets for locally self-similar processes. *Journal of statistical planning and inference*, 99(1):91–110.
- Whitcher, B., Craigmile, P., and Brown, P. (2005). Time-varying spectral analysis in neurophysiological time series using Hilbert wavelet pairs. *Signal Processing*, 85(11):2065–2081.
- Whitcher, B., Guttorp, P., and Percival, D. (2000). Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research*, 105(D11):941–962.
- Witt, A. and Schumann, A. (2005). Holocene climate variability on millennial scales recorded in Greenland ice cores. *Nonlinear Processes in Geophysics*, 12(3):345–352.
- Wolff, E. (2005). Understanding the past-climate history from Antarctica. *Antarctic Science*, 17(04):487–495.
- Wu, Z. and Huang, N. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London-A*, 460(2046):1597–1612.
- Wu, Z. and Huang, N. (2009). Ensemble Empirical Mode Decomposition: A Noise Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, 1:1–41.
- Yiou, P., Baert, E., and Loutre, M. (1996). Spectral analysis of climate data. *Surveys in Geophysics*, 17(6):619–663.