

# **A Statistical Study of Wavelet Coherence for Stationary and Nonstationary Processes**

A thesis presented for the degree of  
Doctor of Philosophy of Imperial College London  
April, 2011

**Edward Aidan Kingsley Cohen**

Department of Mathematics  
Imperial College London  
180 Queen's Gate  
London SW7 2BZ

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

E.A.K. Cohen

# Copyright

Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made **only** in accordance with instructions given by the Author and lodged in the doctorate thesis archive of the Imperial College London central library. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

The ownership of any intellectual property rights which may be described in this thesis is vested in Imperial College London, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement. Further information on the conditions under which disclosures and exploitation may take place is available from the Imperial College London registry.

# Abstract

The coherence function measures the correlation between a pair of random processes in the frequency domain. It is a well studied and understood concept, and the distributional properties of conventional coherence estimators for stationary processes have been derived and applied in a number of physical settings.

In recent years the wavelet coherence measure has been used to analyse correlations between a pair of processes in the time-scale domain, typically in hypothesis testing scenarios, but it has proven resistant to analytic study with resort to simulations for statistical properties. As part of the null hypothesis being tested, such simulations invariably assume joint stationarity of the series. In this thesis two methods of calculating wavelet coherence have been developed and distributional properties of the wavelet coherence estimators have been fully derived.

With the first method, in an analogous framework to multitapering, wavelet coherence is estimated using multiple orthogonal Morse wavelets. The second coherence estimator proposed uses time-domain smoothing and a single Morlet wavelet. Since both sets of wavelets are complex-valued, we consider the case of wavelet coherence calculated from discrete-time complex-valued and stationary time series. Under Gaussianity, the Goodman distribution is shown, for large samples, to be appropriate for wavelet coherence. The true wavelet coherence value is identified in terms of its frequency domain equivalent and degrees of freedom can be readily derived. The theoretical results are verified via simulations.

The notion of a spectral function is considered for the nonstationary case. Particular focus is given to Priestley's evolutionary process and a Wold-Cramér nonstationary representation where time-varying spectral functions can be clearly defined. Methods of estimating these spectra are discussed, including the continuous wavelet transform, which when performed with a Morlet wavelet and temporal smoothing is shown to bear close resemblance to Priestley's own estimation procedure.

The concept of coherence for bivariate evolutionary nonstationary processes is discussed in detail. In such situations it can be shown that the coherence function, as in the stationary case, is invariant of time. It is shown that for spectra that vary slowly in time the derived statistics of the temporally smoothed wavelet coherence estimator are appropriate. Further to this the similarities with Priestley's spectral estimator are exploited to derive distributional properties of the corresponding Priestley coherence estimator.

A well known class of the evolutionary and Wold-Cramér nonstationary processes are the modulated stationary processes. Using these it is shown that bivariate processes can be constructed that exhibit coherence variation with time, frequency, and time-and-frequency. The temporally smoothed Morlet wavelet coherence estimator is applied to these processes. It is shown that accurate coherence estimates can be achieved for each type of coherence, and that the distributional properties derived under stationarity are applicable.

# Acknowledgments

I would first and foremost like to thank my supervisor, Prof Andrew Walden, for his time, ideas and guidance. I am also grateful to the Engineering and Physical Sciences Research Council (EPSRC), whose funding has made this research possible.

I would like to thank members of Imperial College's Statistics Section for creating a stimulating and enjoyable environment in which to work.

I thank my family, and in particular my parents Clive and June for their continued love and support throughout my education. There is no doubt that without them I would not have been able to achieve this.

Finally I would like to thank Red for her kindness, love and unwavering support.

Ed

# Table of contents

<b>Abstract</b>	<b>4</b>
<b>1 Introduction</b>	<b>15</b>
<b>2 Complex-Valued Stationary Processes</b>	<b>22</b>
2.1 Random Complex Vectors . . . . .	23
2.2 Complex-Valued Processes . . . . .	25
2.3 Spectral Representation of Stationary Processes . . . . .	27
2.4 Gaussian Distributed Complex-Valued Vectors . . . . .	31
2.5 Complex Vector-Valued Gaussian Processes . . . . .	34
2.6 Coherence . . . . .	34
2.7 Spectral Density and Coherence Estimators . . . . .	37
2.7.1 Lag-window Spectral Estimators . . . . .	42
2.7.2 Segment Averaging . . . . .	44
2.7.3 Multitapering . . . . .	45
2.8 Multitaper Spectral Estimators and Coherence . . . . .	47
<b>3 Wavelet Coherence for Stationary Processes</b>	<b>54</b>
3.1 Representing the $L^2$ Function Space . . . . .	55
3.2 Time-Frequency Analysis . . . . .	56
3.3 The Continuous Wavelet Transform . . . . .	60
3.3.1 The Continuous Wavelet Transform in Discrete-Time . . . . .	61
3.3.2 Wraparound and the Maximum Analysis Scale . . . . .	63
3.3.3 Calculating the Minimum Analysis Scale . . . . .	64
3.4 The Wavelet Spectrum and Wavelet Coherence Measure . . . . .	65
<b>4 A Statistical Analysis of Morse Wavelet Coherence</b>	<b>70</b>
4.1 The Multiple Morse Wavelets . . . . .	71
4.2 The Morse Wavelet Coherence Estimator . . . . .	76
4.3 The Morse Wavelet Transform for Discrete-Time Processes . . . . .	77
4.3.1 Wraparound and the Maximum Analysis Scale . . . . .	78
4.3.2 Calculating the Minimum Analysis Scale . . . . .	78

---

4.3.3	Mapping Time-Scale to Time-Frequency . . . . .	78
4.4	The Statistics of the Morse Wavelet Coherence Estimator . . .	79
4.4.1	The Wavelet Spectral Matrix . . . . .	79
4.4.2	Orthogonality and Uncorrelatedness . . . . .	83
4.5	Key Results and Distributional Properties of WCOH . . . . .	84
4.6	Simulations and Results . . . . .	86
4.6.1	The Complex Process . . . . .	87
4.6.2	Specifics of the Morse Wavelet Parameters . . . . .	87
4.6.3	Results . . . . .	88
4.7	Concluding Remarks . . . . .	93
<b>5</b>	<b>Temporally Smoothed Wavelet Coherence</b>	<b>95</b>
5.1	Smoothing with the Morlet Wavelet . . . . .	96
5.2	WOSA for Continuous-Time Processes . . . . .	98
5.3	Discrete-Time Method . . . . .	101
5.3.1	Restrictions on $a$ and $b$ . . . . .	105
5.3.2	Review . . . . .	106
5.4	Statistical Properties of WOSA Estimators . . . . .	107
5.4.1	WOSA to Multitaper . . . . .	108
5.5	Statistical Analysis of TWCOH . . . . .	113
5.6	Example results . . . . .	117
5.7	Significance Testing Example . . . . .	118
5.8	Concluding Remarks . . . . .	121
<b>6</b>	<b>Nonstationary Processes</b>	<b>123</b>
6.1	Harmonizable Processes . . . . .	125
6.2	The Evolutionary Process . . . . .	127
6.2.1	Oscillatory Processes . . . . .	129
6.2.2	Semi-Stationarity . . . . .	131
6.3	Uniformly Modulated Processes . . . . .	134
6.3.1	Example . . . . .	135
6.4	Wold-Cramér Model of Nonstationarity . . . . .	137
6.5	Estimating Evolutionary Spectra . . . . .	139
6.5.1	Priestley's Method of Evolutionary Spectral Estimation	140
6.5.2	Smoothing Procedure . . . . .	142
6.5.3	Wavelets . . . . .	143
6.6	Spectral Estimation Examples . . . . .	147
6.6.1	Uniformly Modulated Process . . . . .	147
6.6.2	Wold-Cramér Nonstationary Process . . . . .	148
6.7	Other models for Nonstationarity . . . . .	152
6.7.1	Locally Stationary Processes . . . . .	152
6.7.2	Wavelet Processes . . . . .	154

---

<b>7</b>	<b>Coherence of Evolutionary Processes</b>	<b>156</b>
7.1	Multivariate Evolutionary Processes . . . . .	157
7.2	Coherence for Evolutionary Processes . . . . .	159
7.3	Multivariate Evolutionary Spectral Estimators . . . . .	162
7.4	Discrete-Time TWCOH Estimator . . . . .	164
7.5	Wavelet Coherence Examples . . . . .	167
7.5.1	Time-and-Frequency-Invariant Coherence . . . . .	167
7.5.2	Frequency-Dependent Coherence . . . . .	169
7.6	Priestley's Coherence Estimator . . . . .	172
7.6.1	Priestley's Estimator for Discrete-Time SOS Processes	173
7.6.2	Priestley's Estimator for Discrete-Time Semi-Stationary Processes . . . . .	175
7.7	Downburst Wind Example . . . . .	177
7.7.1	Coherence Estimator Details . . . . .	179
7.8	Concluding Remarks . . . . .	180
<b>8</b>	<b>Time-Varying-Coherence Models</b>	<b>182</b>
8.1	Multivariate Wold-Cramér Model . . . . .	182
8.1.1	Time-and-frequency-Dependent Coherence . . . . .	184
8.1.2	Time-Dependent Coherence . . . . .	185
8.1.3	Introducing Frequency-Dependence . . . . .	186
8.2	Sigma-Oscillatory Processes . . . . .	187
8.3	Simulations and Results . . . . .	188
8.3.1	Constant Coherence . . . . .	189
8.3.2	Time-Dependent Coherence . . . . .	189
8.3.3	Time-and-Frequency-Dependent Coherence . . . . .	193
8.4	Concluding Remarks . . . . .	194
<b>9</b>	<b>Conclusion</b>	<b>198</b>
<b>A</b>	<b>Proof of Theorem 4.4.1</b>	<b>201</b>
<b>B</b>	<b>Continuous-Time Version of Theorem 4.4.1</b>	<b>208</b>
<b>C</b>	<b>Proof of Theorem 5.5.1</b>	<b>212</b>
<b>D</b>	<b>Characteristic Width Derivation for the Downburst Wind Pro- cess</b>	<b>215</b>

# List of Figures

2.1	(a) Periodogram and (b) multitaper spectral estimator ( $K = 10$ ) for a SOS complex autoregressive process of order 1 — see Section 4.6.1. The spectral density function is also plotted. . .	47
2.2	Probability density functions $g_{\hat{\gamma}^2}(x; K, \gamma^2)$ for $K = 10$ and a true coherence value $\gamma^2$ of (a) 0, (b) 0.1, (c) 0.2, (d) 0.5, (e) 0.7, and (f) 0.9. The dotted line marks the true coherence and the dashed line is the expected value of the distribution. . . . .	51
2.3	Demonstration of calculating a confidence interval of true coherence from an individual coherence estimate. In this case we find the 95% confidence interval ( $\alpha = 0.05$ ) from a coherence estimate $\hat{\gamma}^2 = 0.7$ with 10 complex degrees of freedom. . . . .	52
2.4	PDF surface $g_{\hat{\gamma}^2}(x; K, \gamma^2)$ , $0 \leq x, \gamma^2 < 1$ for $K = 10$ . See text for an explanation. . . . .	53
3.1	Symbolic representation of the Gabor discretisation (left) and wavelet discretisation (right) of a signal as a paving of the time-frequency plane. . . . .	59
3.2	The real part of the Morlet wavelet in the time domain (left) and frequency domain (right) for scales (a),(b) $a = 1$ , (b),(c) $a = 2$ and (d),(e) $a = 5$ . . . . .	62
3.3	Symbolic representation of how the continuous wavelet transform can analyse an arbitrary point of the time-frequency plane, and the time-frequency trade-off that occurs. . . . .	63
4.1	Morse wavelet domain $\mathcal{D}_{C,\beta,\gamma}$ (solid) for $C = 10$ , $\beta = 8$ and $\gamma = 3$ . The disc of equal area is also plotted (dashed). . . . .	73
4.2	Each row of the plot shows the magnitude $ \psi_{k,40,15}^+ $ (left) and frequency response $\Psi_{k,40,15}^+$ (right). The rows are for (top to bottom) $k = 0, \dots, 4$ . . . . .	75

4.3	A Morse wavelet coherence estimate for a SOS CAR(1) process (see Section 4.6.1), with parameters $K = 10$ , $\beta = 40$ and $\gamma = 15$ , plotted as a function of time-scale (left) and time-frequency (right). Only the valid regions are shown. Areas of dark shading indicate high values of coherence. . . . .	88
4.4	Q-Q plots for 100 independent samples of $\hat{\gamma}_*^2(a, b)$ for (a) $N = 256$ and (b) $N = 512$ . The four curves on each plot correspond to four different $(a, b)$ pairs on the scale-time plane. $K = 10$ , $\beta = 40$ and $\gamma = 15$ . . . . .	89
4.5	Q-Q plots for 100 independent samples of $\hat{\gamma}_*^2(a, b)$ for (a) $K = 5$ and (b) $K = 10$ . The four curves on each plot correspond to four different $(a, b)$ pairs on the scale-time plane. $N = 1024$ , $\beta = 40$ and $\gamma = 15$ . . . . .	90
4.6	The theoretical mean values of the coherence estimator are shown by the triangles, and the sample mean of $\hat{\gamma}_*^2(a, b)$ over 1000 simulations, at four randomly chosen points $(a, b)$ on the scale-time plane, are shown by crosses. $N = 1024$ . . . . .	91
4.7	The four theoretical components of $\mathfrak{Y}(f)$ . In each plot the thin line marks the true value. The upper right and lower left plots show both real and imaginary parts of $R_{11}(f)$ and $R_{11}^*(f)$ , respectively. The crosses are the means of $\hat{\Omega}(a, b)$ , over 1000 simulations, for 100 different positive and negative scale values in the range $a_{\min} <  a  < a_{\max}$ , plotted at positions $f = f_0/a$ for a randomly-chosen time $b$ , $\beta = 40$ , $\gamma = 15$ and $N = 1024$ . . . . .	92
4.8	$\sum_{k=0}^{K-1} [\Psi_k^+(f)]^2$ for Morse parameters (a) $\beta = 40$ , $\gamma = 15$ and (b) $\beta = 30$ , $\gamma = 5$ . . . . .	93
4.9	As for Figure 4.7 but for Morse parameters $\beta = 30$ , $\gamma = 5$ and $N = 1024$ . . . . .	94
5.1	Illustration of the WOSA estimator equivalent to the TWCOH estimator for a translation point of interest $b = 400$ and scale $a = 10$ . The realisation of the process is shown on the bottom and the Gaussian taper component on the top. The solid vertical lines and bold portion represent the part used in calculating the estimate for $b = 400$ . The dashed lines shows an individual block of size $N_S$ , $N_B$ of which are averaged together. . . . .	104
5.2	Parameter relationships and statistical goodness-of-fit assessment when $d = 1$ , $f'(d) = 1.8$ , $N = 1024$ , $\kappa_0 = 12$ . (a) Block size $N_P$ versus $a_0$ . (b) Region of validity for $(a_0, b_0)$ given by the interior of the central triangle. (c) Degrees of freedom $\varphi$ versus $a_0$ . (d) Goodman Q-Q plots for four different $(a_0, b_0)$ locations. . . . .	107

5.3	Same style of plots as for Figure 5.2 but with $d = 2$ , $f'(d) = 1.4$ , $N = 1024$ , $\kappa_0 = 24$ . . . . .	108
5.4	Contours of constant complex degrees of freedom in the $N_S$ - $N_B$ plain. The right most line is for $\varphi = 2$ with lines increasing in steps of 2 from right to left. . . . .	113
5.5	Magnetic time series $\{Z_{1,t}\}$ and $\{Z_{2,t}\}$ . The real part is plotted with a solid line, and the imaginary part is dashed. . . . .	119
5.6	Wavelet coherence estimate as a function of scale $a$ and time $b$ of magnetic field data (left), and those values exceeding the 1% point of the null distribution (right). (Higher coherence corresponds to darker shading.) . . . . .	120
6.1	Simulated seismic signal using the UMP model. Figure copied from [22]. . . . .	135
6.2	Spectral analysis for the UMP. (a) shows the evolutionary spectral density of the UMP in time-frequency space (for the positive half of the spectrum only). (b) shows the ESDF in the Morlet wavelet's time-scale space (with the mapping $a = 1/f$ ). (c) is a single realisation of the non-smoothed wavelet spectrum in time-scale space. (d) is the mean non-smoothed wavelet spectrum. (e) and (g) are single realisations of the smoothed wavelet spectrum with smoothing parameters $\kappa_0 = 6$ and $\kappa_0 = 12$ respectively. (f) and (h) are the corresponding mean smoothed wavelet spectra. . . . .	149
6.3	Wold-Cramér spectra for the arterial blood flow process. (a) shows the Wold-Cramér spectrum in time-frequency space. (b) shows the Wold-Cramér spectrum in the Morlet wavelet's time-scale space (with the transform $a = 1/f$ ). (c) is a single realisation of the non-smoothed wavelet spectrum in time-scale space. (d) is the mean non-smoothed wavelet spectrum. (e) is a single realisation of the smoothed wavelet spectrum with smoothing parameter $\kappa_0 = 12$ . (f) is the corresponding mean smoothed wavelet spectra. . . . .	151
7.1	Real part of a realisation of the UMP $\{Z_t\}$ used in Section 7.5.1. The bars indicate the length $N_P$ of the signal portion used in analysis of a single point, for smoothing parameters (top) $\kappa_0 = 9.6$ and (bottom) $\kappa_0 = 24.6$ . . . . .	167

7.2	Probability density functions of the coherence estimates approximated from five thousand realisations of the bivariate UMP from Section 7.5.1 using the smoothing parameters $\kappa_0 =$ (a) 9.6, (b) 14.6, (c) 19.6 and (d) 24.6. Plotted with a dashed line is the appropriate Goodman distribution assumed under stationarity.	168
7.3	Q-Q plots constructed using five thousand estimates for the smoothing parameters $\kappa_0 =$ (a) 9.6, (b) 14.6, (c) 19.6 and (d) 24.6. . . . .	169
7.4	Real parts of a realisation of the bivariate UMP from Section 7.5.2. . . . .	170
7.5	Left column: Debiased mean coherence estimates from 5000 simulations for scales $a = 6, 12, 18, 24$ (top to bottom) using smoothing with (a) $\kappa_0 = 6$ (b) $\kappa_0 = 12$ . Dotted lines show true coherence values. Horizontal bars show the block size, $N_P$ , for $a = 6$ (top) and $a = 24$ (bottom). Right column: Corresponding Q-Q plots for same scales at a single time point $b = 512$ for smoothing with (b) $\kappa_0 = 6$ , (d) $\kappa_0 = 12$ . . . . .	171
7.6	Realisations of the bivariate downburst wind UMP $\{Z_1(t)\}$ (top) and $\{Z_2(t)\}$ (bottom). . . . .	178
7.7	Debiased mean coherence estimates (rings) for 100 simulations of the downburst wind process, plotted as a function of frequency at four different points in time. The true coherence is plotted with the solid line. . . . .	179
7.8	Q-Q plots at four separate frequencies. Each plot has four curves for four different time points. . . . .	181
8.1	Left: Realisation of the bivariate process. Right: Modulating functions $g_t$ (solid line), $\beta_t$ (dashed) and $d_t$ (dash-dot). . . . .	189
8.2	Constant coherence example. Left column: Debiased mean coherence estimates from 5000 simulations for scales (a) $a = 6$ (b) $a = 18$ using smoothing with $\kappa_0 = 12$ ( $\varphi = 10$ ). Dotted lines show true coherence values. Horizontal bars show the block size, $N_P$ , for the stated value of scale $a$ . Right column: Corresponding Q-Q plots at time points $b = 450$ and $550$ for scales (b) $a = 6$ , (d) $a = 18$ . . . . .	190
8.3	As for Figure 8.2, but for the first example of Section 8.3.2. . . . .	191
8.4	As for Figure 8.2, but for the second example of Section 8.3.2. . . . .	192
8.5	Left: Realisations of the processes producing the time-dependent coherence of Figure 8.6. Right: Modulating functions $g_t$ (solid line), $\beta_t$ (dashed) and $d_t$ (dash-dot). . . . .	193

8.6	Debiased mean coherence estimates from 5000 simulations for third example of time-dependent coherence for scale $a = 6$ with smoothing parameters of (a) $\kappa_0 = 12$ ( $\varphi = 10$ ) and (b) $\kappa_0 = 6$ ( $\varphi = 5$ ). . . . .	194
8.7	First example of time-and-frequency-dependent coherence plotted as a surface in time-frequency space. The white lines are for $f = 1/6$ and $1/18$ . . . . .	195
8.8	Left: Realisations of the processes producing the time-and-frequency-dependent coherence of Figure 8.7. Right: Modulating functions $g_t$ (solid line), $\beta_t$ (dashed) and $d_t$ (dash-dot). . . . .	195
8.9	(a) Debiased mean coherence estimates from 5000 simulations for time-and-frequency-dependent coherence of Figure 8.7 for scales $a = 18$ (top) and $a = 6$ (bottom) using smoothing with $\kappa_0 = 12$ ( $\varphi = 10$ ). Dotted lines show true coherence values. Horizontal bars show the block size, $N_P$ , for $a = 18$ (top) and $a = 6$ (bottom). (b) Corresponding Q-Q plots for both scales at the single time point $b = 512$ . . . . .	196
8.10	Second example of time-and-frequency-dependent coherence plotted as a surface in time-frequency space. The white lines are for $f = 1/6$ and $1/18$ . . . . .	196
8.11	(a) Debiased mean coherence estimates from 5000 simulations for time-and-frequency-dependent coherence of Figure 8.10 for scales $a = 18$ (top) and $a = 6$ (bottom) using smoothing with $\kappa_0 = 12$ ( $\varphi = 10$ ). Dotted lines show true coherence values. Horizontal bars show the block size, $N_P$ , for $a = 18$ (top) and $a = 6$ (bottom). (b) Corresponding Q-Q plots for both scales at the single time point $b = 512$ . . . . .	197

# Chapter 1

## Introduction

A random process  $\{Z_t\}$  (also called a random signal or time series) is a time-indexed sequence of random variables. Their abundance across all areas of engineering and the physical sciences has led to their study becoming a rich research discipline of great importance.

Random processes fall into two broad categories, stationary and nonstationary. A second order stationary process is defined as one where the mean is fixed for all time and the covariance of two points in the process separated by a time shift  $\tau$ ,  $\text{cov}\{Z_{t+\tau}, Z_t\}$ , is dependent only on  $\tau$ . This leads to convenient statistical properties, including a constant variance. Important examples of stationary processes include white noise, autoregressive processes and moving average processes. Their elegant statistical properties have resulted in stationary processes being the main focus of the time series literature, and stationarity is a common assumption placed on naturally occurring processes.

An appealing aspect of stationarity is the spectral interpretation that can be placed on it. Often it is the case that underlying properties and characteristics of a stationary process are obscured when observed in the time domain. It is convenient in many science and engineering disciplines to consider stationary processes as being driven by underlying random harmonic oscillations, and consequently it has been common practice to study these processes in the frequency domain. The spectral density function,  $S(f)$  say, of a stationary

process is a well defined object that represents to what extent oscillations at a particular frequency  $f$  are contributing to the variance (power) of the process. Naturally, it is often necessary to estimate the spectrum from a single finite length sample of the random process and there is extensive literature on the various methods of achieving this, e.g. [73].

When presented with a pair of random variables it is natural to investigate to what extent they are correlated. Similarly, understanding the nature of correlations that exists between a pair of random processes is of great importance [52]. Various measures of correlation have been presented in the literature e.g. [33]. With frequency often providing the most suitable domain in which to study the behaviour of a random process, it is natural to look here for correlations between two time series. For a pair of stationary random processes  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$ , each with their own spectral density functions  $S_{11}(f)$  and  $S_{22}(f)$ , it is also possible to define a cross-spectral density function  $S_{12}(f)$  which exposes frequencies that are common to both processes. The magnitude squared coherency, or ordinary coherence

$$\gamma^2(f) = \frac{|S_{12}(f)|^2}{S_{11}(f)S_{22}(f)}$$

gives a normalised measure on the interval  $[0, 1]$  of the correlation that exists between a pair of processes due to oscillations at frequency  $f$ . It has been studied in great detail [11] and applied extensively in the physical sciences [1], [29], [52], [89].

It happens to be the case that real-world processes often fail to meet the conditions of stationarity. This implies their spectral characteristics vary in time and it becomes necessary to perform a chronological spectral analysis of the process. Since their most noticeable developments by Morlet et al in the 1980s [65], [66] and Daubechies in the 1990s [23], wavelets have come to the fore as the tool of choice to perform such analysis. The continuous wavelet transform decomposes a signal into a time-scale space and its squared magnitude, the wavelet spectrum, gives a time-scale representation of the signal's

power, which intrinsically has a time-frequency interpretation. It is also possible to define the cross-wavelet spectrum that exposes common power between a pair of signals in time-scale space.

With the recent developments in wavelet theory there has been much interest in the idea of wavelet coherence as a means of measuring correlations that exist between a pair of processes in a joint time-scale space. The wavelet coherence, analogously to ordinary coherence, is the cross-wavelet spectrum normalised by the individual wavelet spectra. However it is recognised that smoothing of the individual terms is necessary to give meaningful wavelet coherence estimates, as is the case for ordinary coherence.

For the computation of sample coherence in the frequency domain smoothing is straightforward, but for wavelet coherence the question of how to smooth over time and scale has proven problematic [56], [98]. In fact the influential paper by Torrence and Compo [98, p. 77] called for further research on the properties of wavelet coherence. There are two common methods of smoothing found in the literature. The method suggested by Torrence and Webster [99] involves a smoothing over time followed by a smoothing over scale, and has been widely applied [3], [39], [77]. A simple smoothing over just time has also been applied [42], [53], [101].

The standard procedure for using wavelet coherence is to assume under the null hypothesis that the two processes are jointly stationary and Gaussian with a given second-order structure. For example, [39] look for intermittent coherent oscillations against background red Gaussian noise (low frequency) spectra. The background processes thus define the null hypothesis. If the computed wavelet coherence value at time-scale point exceeds the chosen simulated critical value, then the oscillation is declared present at that scale and time. For such tests it is necessary to know the statistical distribution of the wavelet coherence estimator under the null hypothesis, but the full statistical effects of such smoothing have not been derived in either case and statistical properties have only been determined by Monte Carlo simulation, e.g., [39], [60], [77].

We come to the main objective of this thesis; to provide key advances in the statistical theory of wavelet coherence. This is twofold.

- The first objective is to supply a fully parameterised distribution for wavelet coherence for jointly stationary Gaussian processes. This is achieved via a detailed analysis of two different smoothing procedures. The first uses the Morse wavelets [24], [68], a set of orthogonal wavelets with which a set of approximately uncorrelated wavelet spectra are computed from a single sample of the random processes. Averaging across these spectra gives the required smoothing to compute sensible coherence estimates. The second smoothing procedure uses the more traditional method of smoothing across a localised time interval (temporal smoothing) [42], [53], [101]. With both, the framework for wavelet coherence estimation is rigorously treated and fully parameterised distributions for the wavelet coherence estimators are derived. Special attention is given to the discrete-time case with an arbitrary sampling rate, as is necessary for this work to be applied in the physical sciences.
- The second objective of this thesis is to extend the results for stationarity to the realm of nonstationarity. By its very nature nonstationarity is a far more complex problem. Nonstationary processes in general fail to have a well defined statistical structure. Consequently, a meaningful and definable time-frequency spectrum, and hence coherence measure, is often unachievable. It becomes necessary to restrict analysis to nonstationary models which have a well-ordered statistical structure and a well defined time-frequency spectrum. Two such models are the evolutionary process presented by Priestley in [78], [79], [80], [83] and Wold-Cramér formulation [18]. A subclass of these models is the uniformly modulated process, a nonstationary process that has wide spread applications [22], [48], [41]. It will be shown that for the statistical results derived for the temporal smoothing method under the assumption of stationarity, can be extended to the coherence structures for these nonstationary models.

Recently there has been an increased interest in complex-valued random processes [90]. These present their own elegances, and difficulties. It is often the case that theoretical studies are made easier when propriety is enforced and this has traditionally formed the focus of the literature. Here we consider the most general class of complex-valued processes when the condition of propriety is relaxed. Real-valued processes can be considered as a subset of the complex-valued processes, and in this respect the approach taken here is most general.

The thesis progresses as follows. In Chapter 2 is presented a comprehensive introduction to the spectral theory of complex-valued multivariate stationary processes. This allows the introduction of the classical coherence function for stationary processes. Coherence necessarily needs to be estimated from a finite length sample of the processes of interest, and this comes about by forming estimators of the spectral functions. Some of the classical spectral estimators are discussed, including weighted overlapping segment averaging (WOSA) and multitaper spectral estimators. The key statistical results are discussed.

Wavelets as a means of time-scale/frequency analysis are introduced in Chapter 3. In particular the continuous wavelet transform, wavelet spectrum and wavelet coherence provide the focus for discussions. A review of the current literature on the wavelet coherence measure is given and the key existing results are presented.

The Morse wavelets are presented in Chapter 4. These set of orthogonal wavelets have previously been used for coherence estimation [7], however the authors had to resort to Monte Carlo simulations to determine statistical properties of the estimator. Here is presented a derivation of the fully parameterised distribution for the wavelet coherence estimator under the assumption of Gaussian stationarity. The theory is verified through simulations. This work can also be found in Cohen and Walden [14].

Chapter 5 gives a comprehensive and rigorous formulation of the temporally smoothed wavelet coherence estimator, specifically using a Morlet wavelet [9, p. 134]. It is recognised that this can be formulated as a WOSA spectral estimator, which itself can be cast into a multitaper form. This allows the

full parameterised distribution of the temporally smoothed wavelet coherence estimator under the assumption of Gaussian stationarity to be derived and these results are verified through simulations. A simple significance testing scenario demonstrates the theory in use. This work can also be found in Cohen and Walden [15].

Having presented a detailed statistical analysis of two wavelet coherence estimators under the assumption of stationarity, attention switches to nonstationary processes. Some of the key nonstationary models found in the literature are presented in Chapter 6. Special focus is given to Priestley's evolutionary processes and the Wold-Cramér nonstationary formulation. Both these processes are shown to have well defined time-varying spectral functions, and a discussion on estimating these via the temporally smoothed wavelet spectrum, and an existing spectral estimator (which will be called the *Priestley estimator*,) is presented.

For these two key models of nonstationarity it is possible to define a coherence measure. Evolutionary processes display coherence that is invariant of time, however more complex coherence structures that exhibit time-dependence are achievable by considering the Wold-Cramér formulation. Chapter 7 concentrates on the coherence for the evolutionary processes. A subclass of the evolutionary processes known as semi-stationary (which includes the uniformly modulated processes) can be considered to be approximately stationary on a closed interval of time. The stationary results for the temporally smoothed wavelet coherence estimator are shown to be applicable for the Priestley estimator and are extended to the semi-stationary case.

Chapter 8 concentrates on coherence for the Wold-Cramér nonstationary model. It will be shown that by considering a simple regression model, the full panoply of coherence structures can be achieved, namely frequency-dependence, time-dependence, and time-and-frequency-dependence. The temporally smoothed wavelet coherence estimator is shown to be effective for tracking coherence in each of these cases and the statistical results derived under the stationary assumption are applicable. This work can also be found

in Cohen and Walden [16].

## Chapter 2

# Complex-Valued Stationary Processes

Complex-valued random processes are inherent in science and engineering. They are key to disciplines as diverse as communications, radar and sonar; geophysics and oceanography; optics and electromagnetics; audio and acoustics; and medicine and bioengineering, to name a few. While the values of the random process being observed are not themselves complex-valued, it is common place in signal analysis to consider two channels of real signals. It has become the normal practice to make these two channels the real and imaginary components of a single complex-valued signal. Dealing with a single signal that exist in the complex plane brings into play the elegances and conveniences of complex analysis and provides insights that are often obscured through the use of a real-valued representations [90]. For example in oceanography, meteorology and electromagnetics where key insights into the underlying physics of a system are found through studying two-dimensional trajectories (e.g. ocean currents, wind trajectories, rotational polarisation of an electromagnetic signal), a complex representation simplifies the problem into the study of an ellipse in a complex plane.

The study of complex-valued random variables has traditionally been restricted to assumptions that the real and imaginary parts have equal variance

and are independent. These types of random variables are known as *proper* or *circular*. These assumptions are put in place for mathematical simplicity because the theory for the probability distributions for these complex variables is a natural extension of the real case (the Gaussian probability density function (PDF) for a real valued vector and a proper complex vector is unchanged under rotation, unlike improper complex vectors). For complex-valued random processes there exists analogous definitions and results for propriety and impropriety that will be discussed in this chapter.

For some time now it has been acknowledged amongst scientists and engineers that propriety is an insufficient assumption for a wide range of data sets recorded in nature. That is to say that the two channels of real data often have different variances and/or display correlation. The mathematical theory of improper complex-valued random variables and processes is a relatively new discipline with the major advancements being made from the 1990s onwards, e.g. [75], [76], and is only now starting to break through into mainstream literature [90].

In this chapter is presented a comprehensive and rigorous introduction to the spectral theory of improper complex-valued second order stationary processes. The theory will be presented for the univariate case and then extended to the general  $p \geq 1$  dimensional case. We begin with some key definitions and results for random complex-valued vectors.

## 2.1 Random Complex Vectors

Let  $\mathbf{Z} = [Z_1, \dots, Z_p]^T$  denote a  $p$ -dimensional random complex-valued column vector with mean zero. The covariance matrix of  $\mathbf{Z}$  is defined as

$$\boldsymbol{\Sigma}_{\mathbf{Z}} = E\{\mathbf{Z}\mathbf{Z}^H\},$$

where  $E$  is the expectation operator and  $^H$  denotes the Hermitian matrix operation (complex conjugate transpose).  $\boldsymbol{\Sigma}_{\mathbf{Z}}$  is a Hermitian and positive

semidefinite  $p \times p$  matrix. Additionally we can define the relation matrix [75]

$$\mathbf{R}_Z = E\{\mathbf{Z}\mathbf{Z}^T\}$$

which is complex-symmetric. The random complex vector  $\mathbf{Z}$  is said to be *proper* if its relation terms are zero,  $\mathbf{R}_Z = \mathbf{0}$ . If  $\mathbf{R}_Z$  has at least one non-zero entry then the vector  $\mathbf{Z}$  is said to be *improper*. By expressing the vector  $\mathbf{Z}$  as  $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$ , where  $\mathbf{X} = \text{Re}\{\mathbf{Z}\}$ ,  $\mathbf{Y} = \text{Im}\{\mathbf{Z}\}$  and  $i = \sqrt{-1}$ ,

$$\mathbf{R}_Z = E\{\mathbf{Z}\mathbf{Z}^T\} = E\{\mathbf{X}\mathbf{X}^T\} + E\{i\mathbf{X}\mathbf{Y}^T\} + E\{i\mathbf{X}^T\mathbf{Y}\} - E\{\mathbf{Y}\mathbf{Y}^T\},$$

and hence a complex vector is proper if the covariance of the real part is equal to the covariance of the imaginary part, and the real and imaginary parts are uncorrelated.

By forming the augmented vector

$$\check{\mathbf{Z}} = [\mathbf{Z}^T, \mathbf{Z}^H]^T = [Z_1, \dots, Z_p, Z_1^*, \dots, Z_p^*]^T, \quad (2.1)$$

where  $*$  represents complex conjugation, we can define the augmented covariance matrix  $\Sigma_{\check{\mathbf{Z}}} = E\{\check{\mathbf{Z}}\check{\mathbf{Z}}^H\}$  and obtain

$$\Sigma_{\check{\mathbf{Z}}} = \begin{bmatrix} \Sigma_Z & \mathbf{R}_Z \\ \mathbf{R}_Z^* & \Sigma_Z^* \end{bmatrix}$$

in which the conventional covariance and relation matrices for  $\mathbf{Z}$  both appear. This second-order structure and the representation (2.1) were first introduced in [75].

Using the representation  $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$ , we let  $\mathbf{\Lambda}$  denote the covariance matrix of the real-valued augmented random vector

$$\check{\mathbf{X}} = [\mathbf{X}^T, \mathbf{Y}^T]^T = [X_1, \dots, X_p, Y_1, \dots, Y_p]^T.$$

We can partition the covariance matrix  $\mathbf{\Lambda} = E\{\check{\mathbf{X}}\check{\mathbf{X}}^T\}$  into its four constituent

$p \times p$  matrices,

$$\Lambda = \begin{bmatrix} \Lambda_{\mathbf{X}\mathbf{X}} & \Lambda_{\mathbf{X}\mathbf{Y}} \\ \Lambda_{\mathbf{Y}\mathbf{X}} & \Lambda_{\mathbf{Y}\mathbf{Y}} \end{bmatrix},$$

where  $\Lambda_{\mathbf{X}\mathbf{Y}} = E\{\mathbf{X}\mathbf{Y}^T\}$  and the remaining three components are defined in an analogous way. We are able to express the covariance matrix for  $\mathbf{Z}$  as

$$\Sigma_{\mathbf{Z}} = E\{\mathbf{Z}\mathbf{Z}^H\} = (\Lambda_{\mathbf{X}\mathbf{X}} + \Lambda_{\mathbf{Y}\mathbf{Y}}) + i(\Lambda_{\mathbf{Y}\mathbf{X}} - \Lambda_{\mathbf{X}\mathbf{Y}})$$

and the relation matrix as

$$\mathbf{R}_{\mathbf{Z}} = E\{\mathbf{Z}\mathbf{Z}^T\} = (\Lambda_{\mathbf{X}\mathbf{X}} - \Lambda_{\mathbf{Y}\mathbf{Y}}) + i(\Lambda_{\mathbf{Y}\mathbf{X}} + \Lambda_{\mathbf{X}\mathbf{Y}}).$$

## 2.2 Complex-Valued Processes

Let  $\{Z_t, t \in \mathbb{Z}\}$  denote a set of complex-valued random variables indexed by a time variable  $t$  with  $E\{Z_t\} = \mu$  for all  $t \in \mathbb{Z}$ . We call  $\{Z_t\}$  a random process. Let us define the autocovariance at time  $t$  and lag  $\tau$  as  $\text{cov}\{Z_{t+\tau}, Z_t\} \equiv E\{(Z_{t+\tau} - \mu)(Z_t - \mu)^*\}$ . We further define the relation [76] (or complementary covariance [91]) at lag  $\tau$  as  $\text{rel}\{Z_{t+\tau}, Z_t\} \equiv E\{(Z_{t+\tau} - \mu)(Z_t - \mu)\} = \text{cov}\{Z_{t+\tau}, Z_t^*\}$ .  $\{Z_t\}$  is said to be *second order stationary* (SOS) if and only if  $\text{cov}\{Z_{t+\tau}, Z_t\}$  and  $\text{rel}\{Z_{t+\tau}, Z_t\}$  are functions of  $\tau$  only. For SOS processes we can define the autocovariance sequence  $\{s_\tau, \tau \in \mathbb{Z}\}$  as  $s_\tau \equiv \text{cov}\{Z_{t+\tau}, Z_t\}$  and the relation sequence  $\{r_\tau, \tau \in \mathbb{Z}\}$  as  $r_\tau \equiv \text{rel}\{Z_{t+\tau}, Z_t\}$ .

A complex-valued random process with non-zero mean,  $\{W_t\}$  say, can be turned into a zero-mean process  $Z_t = W_t - \mu_W$  by simply removing the mean  $\mu_W = E\{W_t\}$ . The second (and higher) order statistical properties of  $\{W_t\}$  are completely captured by the zero-mean process  $\{Z_t\}$ , specifically,  $\text{cov}\{Z_{t+\tau}, Z_t\} = \text{cov}\{W_{t+\tau}, W_t\}$  and  $\text{rel}\{Z_{t+\tau}, Z_t\} = \text{rel}\{W_{t+\tau}, W_t\}$ . Therefore without loss of generality we will from here onwards assume all SOS random processes to be of zero-mean.

Using the definitions of the autocovariance and relation sequences we can

divide the set of SOS processes into two distinct subsets. Proper SOS processes are defined as those whose relation sequence  $\{r_\tau, \tau \in \mathbb{Z}\}$  is identically zero for all lag times  $\tau$ . Improper SOS processes are those whose relation sequences take a non-zero value for at least one value of  $\tau$ . Noting that  $\mathbb{R} \subset \mathbb{C}$ , we can say the set of real valued random processes are a subset of the complex-valued random processes, further to this the autocovariance and relation sequences for real valued processes are equal and therefore within this framework they are considered to be improper.

If we construct the 2-dimensional process  $\{\check{\mathbf{Z}}_t = [Z_t, Z_t^*]^T, t \in \mathbb{Z}\}$  and the lag- $\tau$  covariance matrix is defined as

$$\Xi_\tau \equiv E\{\check{\mathbf{Z}}_{t+\tau}\check{\mathbf{Z}}_t^H\} = \begin{bmatrix} s_\tau & r_\tau \\ r_\tau^* & s_\tau^* \end{bmatrix},$$

then  $s_\tau$  and  $r_\tau$  take their usual meaning of the covariance and relation sequences respectively.  $\{Z_t\}$  is proper if and only if  $\Xi_\tau$  is a diagonal matrix for all  $\tau$ .

Let  $\{\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{p,t}]^T, t \in \mathbb{Z}\}$  denote a  $p$ -dimensional complex vector-valued random process. By constructing the vector process  $\{\check{\mathbf{Z}}_t = [\mathbf{Z}_t^T, \mathbf{Z}_t^H]^T = [Z_{1,t}, \dots, Z_{p,t}, Z_{1,t}^*, \dots, Z_{p,t}^*]^T, t \in \mathbb{Z}\}$ , it is said  $\{\mathbf{Z}_t\}$  is *jointly stationary* if all the elements of the lag- $\tau$  covariance matrix  $\Xi_\tau \equiv E\{\check{\mathbf{Z}}_{t+\tau}\check{\mathbf{Z}}_t^H\}$  are dependent on the lag time  $\tau$  only. In this case we have

$$\Xi_\tau = \begin{bmatrix} \mathbf{s}_\tau & \mathbf{r}_\tau \\ \mathbf{r}_\tau^* & \mathbf{s}_\tau^* \end{bmatrix}, \quad (2.2)$$

where  $\mathbf{s}_\tau$  and  $\mathbf{r}_\tau$  are  $p \times p$  matrices with  $(l, m)$ th elements  $s_{lm,\tau}$  and  $r_{lm,\tau}$  respectively. For the case  $l \neq m$ ,  $s_{lm,\tau} = E\{Z_{l,t+\tau}Z_{m,t}^*\}$  is the cross-covariance sequence and  $r_{lm,\tau} = E\{Z_{l,t+\tau}Z_{m,t}\}$  is the cross relation sequence for the processes  $\{Z_{l,t}\}$  and  $\{Z_{m,t}\}$ . For the case of  $l = m$  they are the standard autocovariance and relation sequences.

### 2.3 Spectral Representation of Stationary Processes

We now turn our attention to the frequency domain interpretation of SOS processes. The Fourier transform decomposes a deterministic signal into its constituent frequencies. The spectral representation theorem provides an analogous framework for stochastic signals, decomposing the variance of a SOS process into constituent frequencies. Formulated by Cramér in 1942 [17], here we present a version taken from [73, Section 4.1].

**Theorem 2.3.1.** *Let  $\{Z_t, t \in \mathbb{Z}\}$  be a SOS real or complex-valued process of zero-mean, then there exists a complex-valued stochastic process  $\{\zeta(f)\}$  such that with probability one,*

$$Z_t = \int_{-1/2}^{1/2} e^{i2\pi ft} d\zeta(f). \quad (2.3)$$

where  $\{\zeta(f)\}$  is an orthogonal increment process with properties

1.  $E\{d\zeta(f)\} = 0$  for all  $|f| \leq 1/2$ ;
2.  $E\{|d\zeta(f)|^2\} = dS^{(I)}(f)$ , say, for all  $|f| < 1/2$ , where the bounded non-decreasing function  $S^{(I)}(f)$  is called the integrated spectrum of  $\{Z_t\}$ ; and
3. for any two distinct frequencies  $f$  and  $f'$  contained in the interval  $[-1/2, 1/2)$

$$\text{cov}\{d\zeta(f'), d\zeta(f)\} = E\{d\zeta(f')d\zeta^*(f)\} = 0.$$

Equation (2.3) is called the *spectral representation* of  $\{Z_t, t \in \mathbb{Z}\}$  and states that all SOS processes are comprised of an infinite sum of complex exponentials  $\{e^{i2\pi ft}\}$  oscillating at frequencies  $f \in [-1/2, 1/2)$ , with random amplitudes  $|d\zeta(f)|$  and phases  $\arg\{d\zeta(f)\}$  (where  $\arg(z)$  is the phase angle of the complex number  $z$ ).

It can be shown using (2.3) and parts 2 and 3 of Theorem 2.3.1 that the

autocovariance  $\{s_\tau\}$  of process  $\{Z_t\}$  has the representation

$$s_\tau = \int_{-1/2}^{1/2} e^{i2\pi f\tau} dS^{(I)}(f).$$

If the integrated spectrum is differentiable everywhere with derivative  $S(\cdot)$  then  $E\{|d\zeta(f)|^2\} = dS^{(I)}(f) = S(f)df$ . The function  $S(\cdot)$  is called the spectral density function (SDF) and forms the following Fourier pair relationship with the autocovariance sequence

$$s_\tau = \int_{-1/2}^{1/2} S(f)e^{i2\pi f\tau} df \quad (2.4)$$

$$S(f) = \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f\tau}, \quad (2.5)$$

and more specifically

$$\text{var}\{Z_t\} = s_0 = \int_{-1/2}^{1/2} S(f)df.$$

In essence the SDF measures how the variance of a process is distributed across frequencies.  $S(f)df$  is the contribution to the variance in the stationary process due to oscillations at frequencies in a small interval about  $f$ .

At this point it is worth making a comment on the limits of the integral in (2.3). The reason for the range of integration  $[-1/2, 1/2)$  is because the value of  $\exp(i2\pi ft)$  is identical to  $\exp(i2\pi(f \pm k)t)$  for all  $t, k \in \mathbb{Z}$ . It is because of this periodicity that it is only necessary to consider frequencies in this range. When the random process is defined for the times  $t \in \mathbb{Z}$  we say it has a unit *sampling interval*,  $\Delta = 1$ .

For a continuous-time SOS processes  $\{Z(t), t \in \mathbb{R}\}$  ( $\text{cov}\{Z(t + \tau), Z(t)\}$  and  $\text{rel}\{Z(t + \tau), Z(t)\}$  depend only on  $\tau$ ) the representation simply becomes

$$Z(t) = \int_{-\infty}^{\infty} e^{i2\pi ft} d\zeta(f).$$

The orthogonal increment process possesses all the same properties as those given in Theorem 2.3.1, but now for  $|f| < \infty$ .

Suppose we wish to sample a continuous-time process at an arbitrary sampling interval  $\Delta > 0$ . Failure to sample at a fine enough rate results in components of the signal oscillating at frequencies  $|f| > 1/(2\Delta)$  to be indistinguishable from oscillations at frequencies  $f \pm N/(2\Delta)$ ,  $N \in \mathbb{Z}$ . This is known as aliasing. It is therefore necessary that  $\Delta$  is small enough such that  $S(f) = 0$  for  $|f| > 1/(2\Delta)$ . In this case the spectral representation generalises to

$$Z_t = \int_{-1/(2\Delta)}^{1/(2\Delta)} e^{i2\pi ft\Delta} d\zeta(f).$$

The frequency  $1/(2\Delta)$  is called the *Nyquist frequency* and will be represented by  $f_N$ .

We now approach the issue of spectral representation for a general  $p \geq 1$  dimensional jointly stationary process. The  $p$ -dimensional process  $\{\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{p,t}]^T, t \in \mathbb{Z}\}$  has the representation

$$\mathbf{Z}_t = \int_{-1/2}^{1/2} e^{i2\pi ft} d\zeta(f)$$

where  $\{\zeta(f)\}$  is a  $p$ -dimensional complex vector-valued orthogonal increment process with the following properties;

1.  $E\{d\zeta(f)\} = \mathbf{0}$  for all  $|f| \leq 1/2$ ;
2.  $E\{d\zeta(f)d\zeta^H(f)\} = d\mathbf{S}^{(I)}(f)$ , say, for all  $|f| < 1/2$ , where the matrix of bounded non-decreasing functions  $\mathbf{S}^{(I)}(f)$  is called the integrated spectral matrix of  $\{\mathbf{Z}_t\}$ ; and
3. for any two distinct frequencies  $f$  and  $f'$  contained in the interval  $[-1/2, 1/2)$

$$\text{cov}\{d\zeta(f'), d\zeta(f)\} = E\{d\zeta(f')d\zeta^H(f)\} = \mathbf{0}.$$

Furthermore, if the integrated spectral matrix  $\mathbf{S}^{(I)}(f)$  is differentiable for all  $f \in [-1/2, 1/2)$  with derivative  $\mathbf{S}(f)$ , then  $E\{d\zeta(f)d\zeta^H(f)\} = d\mathbf{S}^{(I)}(f) = \mathbf{S}(f)df$ .  $\mathbf{S}(f)$  is a  $p \times p$  matrix of the form

$$\mathbf{S}(f) = \begin{bmatrix} S_{11}(f) & \cdots & S_{1p}(f) \\ \vdots & \ddots & \vdots \\ S_{p1}(f) & \cdots & S_{pp}(f) \end{bmatrix}.$$

The diagonal element  $S_{kk}(f)$ ,  $k = 1, \dots, p$  is the SDF for the process  $\{Z_{k,t}\}$ ,  $k = 1, \dots, p$ . The off-diagonal element  $S_{jk}(f)$ ,  $j, k = 1, \dots, p$ , is the *cross spectral density function* (CSDF) for the processes  $\{Z_{j,t}\}$  and  $\{Z_{k,t}\}$ . The CSDF reveals the covariance between the increments at  $f$  of the orthogonal processes associated with the pair of processes and we have the relation

$$\mathbf{S}(f) = \sum_{\tau=-\infty}^{\infty} \mathbf{s}_{\tau} e^{-i2\pi f\tau}$$

where  $\mathbf{s}_{\tau}$  is the  $p \times p$  matrix of autocovariance and cross-covariance sequences at lag- $\tau$

$$\begin{bmatrix} s_{11,\tau} & \cdots & s_{1p,\tau} \\ \vdots & \ddots & \vdots \\ s_{p1,\tau} & \cdots & s_{pp,\tau} \end{bmatrix}.$$

In general a CSDF is complex-valued, whereas a SDF is always real and non-negative. The SDF of a real valued random process is symmetric around  $f = 0$ . The SDF of a complex-valued process is in general non-symmetric around the zero frequency. If we once again define the augmented  $2p$ -dimensional vector process  $\{\check{\mathbf{Z}}_t\}$  then we define the SDF matrix of  $\{\check{\mathbf{Z}}_t\}$  as

$$\mathbf{\Upsilon}(f) = \sum_{\tau=-\infty}^{\infty} \mathbf{\Xi}_{\tau} e^{-i2\pi f\tau},$$

where  $\mathbf{\Xi}_{\tau}$  is the lag- $\tau$  covariance matrix defined in (2.2). The SDF matrix of

$\{\check{\mathbf{Z}}_t\}$  takes the form

$$\begin{aligned} \Upsilon(f) &= \begin{bmatrix} \mathbf{S}(f) & \mathbf{R}(f) \\ \mathbf{R}^*(-f) & \mathbf{S}^*(-f) \end{bmatrix} \\ &= \begin{bmatrix} S_{11}(f) & \cdots & S_{1p}(f) & R_{11}(f) & \cdots & R_{1p}(f) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_{p1}(f) & \cdots & S_{pp}(f) & R_{p1}(f) & \cdots & R_{pp}(f) \\ R_{11}^*(-f) & \cdots & R_{1p}^*(-f) & S_{11}^*(-f) & \cdots & S_{1p}^*(-f) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ R_{p1}^*(-f) & \cdots & R_{pp}^*(-f) & S_{p1}^*(-f) & \cdots & S_{pp}^*(-f) \end{bmatrix}. \end{aligned} \quad (2.6)$$

We will also call  $\Upsilon(f)$  the augmented SDF of  $\{\mathbf{Z}_t\}$ . The diagonal elements of this  $2p \times 2p$  matrix are true spectral density functions for the individual processes. All off-diagonal terms are cross-spectra.  $S_{lm}(f)$ ,  $l \neq m$ , are simply conventional cross-spectral density functions.  $R_{ll}(f)$  is the relation spectral density function for the process  $\{Z_{l,t}\}$ , which can be thought of as being the CSDF of  $\{Z_{l,t}\}$  and  $\{Z_{l,t}^*\}$ .  $R_{lm}(f)$  are relation cross-spectral density functions, in other words  $R_{lm}(f)$  would be the CSDF for the processes  $\{Z_{l,t}\}$  and  $\{Z_{m,t}^*\}$ .

Often random processes appearing in nature are assumed for analysis purposes to be Gaussian distributed. We here define a Gaussian process by first defining a Gaussian distributed vector.

## 2.4 Gaussian Distributed Complex-Valued Vectors

The random zero-mean  $p$ -dimensional vector  $\mathbf{Z}$  is said to be a zero-mean complex Gaussian random vector if the real ( $\mathbf{X}$ ) and imaginary ( $\mathbf{Y}$ ) parts of  $\mathbf{Z}$  are jointly zero-mean Gaussian, i.e. the vector  $\check{\mathbf{X}} = [\mathbf{X}^T, \mathbf{Y}^T]^T$  with covariance  $\Lambda$  is distributed

$$\check{\mathbf{X}} \stackrel{d}{=} \mathcal{N}_{2p}(\mathbf{0}, \Lambda),$$

where the notation means that the  $\check{\mathbf{X}}$  has the same distribution as a real  $2p$ -dimensional vector-valued Gaussian random variable with mean  $\mathbf{0}$  and  $2p \times 2p$  positive semidefinite covariance matrix  $\mathbf{\Lambda}$ . Equivalently we will say

$$\mathbf{Z} \stackrel{d}{=} \mathcal{N}_p^C(\mathbf{0}, \mathbf{\Sigma}_Z),$$

which means that  $\mathbf{Z}$  has a complex Gaussian distribution with mean zero and covariance  $\mathbf{\Sigma}_Z$ . The joint distribution of  $\mathbf{X}$  and  $\mathbf{Y}$  is known to be [44]

$$f(\mathbf{x}, \mathbf{y}) = (2\pi)^{-p} [\det \mathbf{\Lambda}]^{-1/2} \exp\left\{-\frac{1}{2} \check{\mathbf{x}} \mathbf{\Lambda}^{-1} \check{\mathbf{x}}\right\},$$

where  $\mathbf{x}$ ,  $\mathbf{y}$  and  $\check{\mathbf{x}}$  are realisations of  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\check{\mathbf{X}}$  respectively.

It is well known that for a zero-mean real-valued Gaussian distributed variable  $X$  with unit variance, then

$$X^2 \stackrel{d}{=} \chi_1^2 \tag{2.7}$$

where  $\chi_\nu^2$  is the central chi-squared distribution with  $\nu$  degrees of freedom. For a zero-mean complex-valued Gaussian random variable  $Z = X + iY$ , where the real-valued variables  $X$  and  $Y$  are independent zero-mean Gaussian distributed with unit variance, we have

$$|Z|^2 = X^2 + Y^2 \stackrel{d}{=} \chi_2^2.$$

Wishart [107] extended (2.7) to a zero-mean real-valued  $p$ -dimensional Gaussian vector  $\mathbf{X}$  and derived the Wishart distribution for the  $p \times p$  matrix  $\mathbf{X}\mathbf{X}^T$ .

If

$$\mathbf{X} \stackrel{d}{=} \mathcal{N}_p(\mathbf{0}, \mathbf{\Lambda}),$$

then we write

$$\mathbf{X}\mathbf{X}^T \stackrel{d}{=} \mathcal{W}_p\{1, \mathbf{\Lambda}\}$$

where  $\mathcal{W}_p\{1, \mathbf{\Lambda}\}$  denotes a central Wishart distribution with a single degree

of freedom and mean  $\mathbf{\Lambda}$ .

Goodman [34] derived the complex Wishart distribution. Suppose we have a zero-mean complex-valued  $p$ -dimensional random vector  $\mathbf{Z}$  distributed as

$$\mathbf{Z} \stackrel{d}{=} \mathcal{N}_p^C(\mathbf{0}, \mathbf{\Sigma}),$$

if the vector  $\mathbf{Z}$  is proper (i.e.  $E\{\mathbf{Z}\mathbf{Z}^T\} = \mathbf{0}$ ) then we have

$$\mathbf{Z}\mathbf{Z}^H \stackrel{d}{=} \mathcal{W}_p^C\{1, \mathbf{\Sigma}\}$$

where  $\mathcal{W}_p^C\{1, \mathbf{\Sigma}\}$  denotes a  $p$ -dimensional complex central Wishart distribution with a single complex degree of freedom and mean  $\mathbf{\Sigma}$ . Goodman further showed [35]:

**Theorem 2.4.1.** *If  $\{\mathbf{Z}_k, k = 0, \dots, K - 1\}$  are a set of  $K$  independent identically distributed (IID)  $p$ -dimensional proper complex Gaussian vectors each of distribution*

$$\mathbf{Z}_k \stackrel{d}{=} \mathcal{N}_p^C(\mathbf{0}, \mathbf{\Sigma}),$$

*then the matrix  $\sum_{k=0}^{K-1} \mathbf{Z}_k \mathbf{Z}_k^H$  is a sum of  $K$  independent matrices each with a  $p$ -dimensional central Wishart distribution with a single degree of freedom, and hence*

$$\sum_{k=0}^{K-1} \mathbf{Z}_k \mathbf{Z}_k^H \stackrel{d}{=} \mathcal{W}_p^C\{K, \mathbf{\Sigma}\}$$

*where  $\mathcal{W}_p^C\{K, \mathbf{\Sigma}\}$  denotes the  $p$ -dimensional complex central Wishart distribution with  $K$  complex degrees of freedom and mean  $K\mathbf{\Sigma}$ .*

Interestingly, Wishart distributed matrices with the number of degrees of freedom less than or equal to the dimension of the matrix are always singular [35]. In this case the Wishart distribution  $\mathcal{W}_p^C\{K, \mathbf{\Sigma}\}$ ,  $K \leq p$ , is said to be a singular distribution. For the case  $K > p$  the Wishart distribution is non-singular. These results will be of use in Sections 2.8 and 4.5.

## 2.5 Complex Vector-Valued Gaussian Processes

Let  $\{\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{p,t}]^T, t \in \mathbb{Z}\}$  be a zero-mean  $p$ -dimensional vector-valued process. Now let  $\mathcal{Z}$  represent an  $N$  dimensional vector (finite  $N$ ) whose entries have been drawn from any  $N$  distinct points in the process, across time, or dimension, or both, i.e.

$$\mathcal{Z} = [Z_{\alpha_1, \beta_1}, Z_{\alpha_2, \beta_2}, \dots, Z_{\alpha_N, \beta_N}]$$

where the pairs  $\{(\alpha_j, \beta_j), j = 1, \dots, N\}$  are unique elements of  $\{1, \dots, p\} \times \mathbb{Z}$ .  $\{\mathbf{Z}_t\}$  is called a complex Gaussian vector-valued process if for any  $\mathcal{Z}$

$$\mathcal{Z} \stackrel{d}{=} \mathcal{N}_N^C(\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{Z}}).$$

## 2.6 Coherence

Given two or more random events it is common to ask; to what extent can we predict the outcome of one of the events given knowledge of the outcome of another event? Likewise, when presented with a pair of random processes we might wish to study whether we are able to predict properties of one process from knowledge of the second process. The degree of correlation for two zero-mean complex-valued random variables  $Z_1$  and  $Z_2$  is defined as

$$\rho = \frac{E\{Z_1 Z_2^*\}}{[E\{|Z_1|^2\}E\{|Z_2|^2\}]^{1/2}}. \quad (2.8)$$

In general this is complex-valued and it becomes useful to look at  $|\rho|^2$ , which on a scale of zero to one measures the correlation between a pair of complex-valued random variables.

For a pair of SOS processes  $\{Z_{1,t}, t \in \mathbb{Z}\}$  and  $\{Z_{2,t}, t \in \mathbb{Z}\}$  there are a number of time domain correlation measures that can be defined to analyse linear relations between the two processes, [33] provides an overview of the topic. We have shown that the frequency domain provides a natural environment in

which to study the statistical properties of SOS random processes and consequently it can be used to provide key insights into how a pair of processes are correlated. It has been shown with the spectral representation theorem that a SOS process  $\{Z_{1,t}, t \in \mathbb{Z}\}$  can be represented as an infinite sum of complex exponentials at frequencies  $f \in [-1/2, 1/2)$  with random amplitudes  $|d\zeta_1(f)|$  and random phase  $\arg(d\zeta_1(f))$ . Given a second SOS process  $\{Z_{2,t}, t \in \mathbb{Z}\}$  we may wish to measure if there is any linear relation between how the frequency components of  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  vary. Using (2.8) we can define the coherency at frequency  $f \in [-1/2, 1/2)$  for the random events  $\{d\zeta_1(f)\}$  and  $\{d\zeta_2(f)\}$  as

$$\gamma(f) = \frac{E\{d\zeta_1(f)d\zeta_2^*(f)\}}{[E\{|d\zeta_1(f)|^2\}E\{|d\zeta_2(f)|^2\}]^{1/2}}.$$

Using the relationships  $E\{|d\zeta_1(f)|^2\} = S_{11}(f)df$ ,  $E\{|d\zeta_2(f)|^2\} = S_{22}(f)df$  and  $E\{d\zeta_1(f)d\zeta_2^*(f)\} = S_{12}(f)df$  we define the ordinary coherence, or magnitude squared coherency (MSCOH), as

$$\gamma^2(f) = \frac{|S_{12}(f)|^2}{S_{11}(f)S_{22}(f)}.$$

Using the Cauchy-Schwarz inequality it is possible to show that  $0 \leq \gamma^2(f) \leq 1$  for all  $f$  [32, Section 13.2].  $\gamma^2(f)$  measures the linear dependency between a pair of processes due to oscillations at a frequency  $f$ . A value close to 0 indicates there is a weak dependency, a value close to 1 indicates a strong dependency.

For a general  $p$ -dimensional complex vector-valued process with augmented SDF matrix  $\Upsilon(f)$ , we notice that appropriate functions of the type

$$\frac{|\Upsilon_{lm}(f)|^2}{\Upsilon_{ll}(f)\Upsilon_{mm}(f)} \quad l, m = 1, \dots, 2p$$

(where  $\Upsilon_{lm}(f)$  is the  $(l, m)$ th element of  $\Upsilon(f)$ ) can define a coherence measure. These can be divided into the following three distinct classes:

1. the ordinary coherence  $\gamma_{lm}^2(f)$  for the pair of processes  $\{Z_{l,t}\}$  and  $\{Z_{m,t}\}$

is defined as

$$\gamma_{lm}^2(f) = \frac{|S_{lm}(f)|^2}{S_{ll}(f)S_{mm}(f)} = \frac{|\Upsilon_{lm}(f)|^2}{\Upsilon_{ll}(f)\Upsilon_{mm}(f)};$$

2. the conjugate coherence  $\gamma_{lm*}^2(f)$  for the pair of processes  $\{Z_{l,t}\}$  and  $\{Z_{m,t}^*\}$  is defined as

$$\gamma_{lm*}^2(f) = \frac{|R_{lm}(f)|^2}{S_{ll}(f)S_{mm}(-f)} = \frac{|\Upsilon_{l,m+p}(f)|^2}{\Upsilon_{ll}(f)\Upsilon_{m+p,m+p}(f)};$$

3. the conjugate coherence  $\gamma_{ll*}^2(f)$  for the pair of processes  $\{Z_{l,t}\}$  and  $\{Z_{l,t}^*\}$  is defined as

$$\gamma_{ll*}^2(f) = \frac{|R_{ll}(f)|^2}{S_{ll}(f)S_{ll}(-f)} = \frac{|\Upsilon_{l,l+p}(f)|^2}{\Upsilon_{ll}(f)\Upsilon_{l+p,l+p}(f)}.$$

We note that  $\gamma_{l*m*}(f) = \gamma_{lm}(-f)$ . The coherence measures  $\gamma_{ll}^2(f)$  and  $\gamma_{l*l*}(f)$  are trivial, being equal to one for all frequencies.

The ordinary coherence measure has been extensively studied and widely applied across a wide range of applications in engineering and the physical sciences (e.g. [1], [29], [89] and [95]). It most notably proved its worth in a seminal paper on global warming [52].

The conjugate coherence, between a process and its complex conjugate, has also proven to be of interest. Given a pair of stationary real-valued random processes  $\{U_t\}$  and  $\{V_t\}$ , where these are measurements of the components of a two dimensional system (for example they could represent the longitudinal and latitudinal components of ocean currents or wind vectors, or perpendicular components of an electromagnetic signal), it has become common practice to represent these two processes as a single complex random process  $\{Z_t\}$  where

$$Z_t = U_t + iV_t.$$

Using the spectral representation of a stationary process (2.3) it can be shown

that the contribution to the complex process from a single frequency  $f$  is a random ellipse in the complex plane given by

$$d\zeta_Z(f)e^{i2\pi ft} + d\zeta_Z(-f)e^{-i2\pi ft}.$$

It is the aspect ratio, rotational direction, and orientation of these elliptical components at a particular frequency that are of great interest in the sciences, e.g. [28], [47], [72] and [100]. It has been shown [88] that the conjugate coherence for the pair of conjugate processes  $\{Z_t\}$  and  $\{Z_t^*\}$  provides key statistical information on the properties of these random ellipses, thus estimating the coherence function allows inference on the rotational kinematics of the observed system.

## 2.7 Spectral Density and Coherence Estimators

In practice the spectral functions and coherence for a pair of processes needs to be estimated from finite length realisations. The coherence estimator is formed using estimators of the spectral functions. For example, we estimate the ordinary coherence  $\gamma_{12}^2(f)$  of a pair of processes  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  using the SDF estimators  $\hat{S}_{11}(f)$ ,  $\hat{S}_{22}(f)$  and CSDF estimator  $\hat{S}_{12}(f)$

$$\hat{\gamma}_{12}^2(f) = \frac{|\hat{S}_{12}(f)|^2}{\hat{S}_{11}(f)\hat{S}_{22}(f)}, \quad (2.9)$$

similarly conjugate coherence estimators are formed using estimators of the relation spectral functions.

The simplest and most natural spectral estimators are the periodogram type spectral estimators. Suppose we observe the points  $Z_{1,0}, \dots, Z_{1,N-1}$  of the zero-mean complex-valued SOS process  $\{Z_{1,t}\}$  at a sampling interval of  $\Delta$ , the simplest estimate of the autocovariance sequence is [73, section 6.2]

$$\hat{s}_\tau^{(p)} = \frac{1}{N} \sum_{n=0}^{N-|\tau|} Z_n Z_{n+|\tau|}^*.$$

Substituting into the relationship in (2.5), we define the periodogram spectral estimator  $\hat{S}^{(p)}(f)$  [73, section 6.3]. It can be shown that

$$\hat{S}_{11}^{(p)}(f) = \frac{\Delta}{N} \left| \sum_{t=0}^{N-1} Z_{1,n} e^{-i2\pi f n \Delta} \right|^2.$$

Note that to estimate the CSDF from an  $N$  point time series of  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  then we do so via

$$\hat{S}_{12}^{(p)}(f) = \frac{\Delta}{N} \left( \sum_{n=0}^{N-1} Z_{1,n} e^{-i2\pi f n \Delta} \right) \left( \sum_{n=0}^{N-1} Z_{2,n} e^{-i2\pi f n \Delta} \right)^*.$$

Percival and Walden [73, section 6.3] derive some important statistical properties of the periodogram. It can be shown that the expected value of the periodogram is the spectral density function convolved with Fejér's kernel

$$\mathcal{F}_N(f) = \frac{\Delta \sin^2(N\pi f \Delta)}{N \sin^2(\pi f \Delta)},$$

giving

$$E\{\hat{S}_{lm}^{(p)}(f)\} = \int_{-f_N}^{f_N} \mathcal{F}_N(f - f') S_{lm}(f') df', \quad l, m = 1, 2$$

and because

$$\lim_{N \rightarrow \infty} \mathcal{F}_N(\cdot) = \delta(\cdot)$$

(where  $\delta(\cdot)$  takes its usual meaning of representing the Dirac delta function) we see that asymptotically the periodogram is an unbiased estimator of the spectral density function

$$\lim_{N \rightarrow \infty} E\{\hat{S}_{jk}^{(p)}(f)\} = S_{jk}(f).$$

Even though the periodogram is asymptotically unbiased, Thomson [96] was of the opinion that in standard engineering applications the periodogram shows significant enough bias to render it untrustworthy. The reason for this considerable bias can be understood through the convolution the SDF undergoes.

In practical examples the number of data points is restricted. Due to the pronounced side-lobes and poor energy concentration in frequency displayed by Fejér's kernel, when a SDF is convolved with it, power is transferred from one frequency region to another. This phenomenon is known as *leakage*.

It is possible to reduce the leakage and hence the bias that occurs when forming a spectral estimator by a process known as tapering. It has been discussed that the bias in the periodogram is due to the convolution of the spectral density function with Fejér's kernel. Tapering is a technique that attempts to reduce the side-lobes of the convolving kernel and hence reduces associated leakages.

Let  $Z_0, \dots, Z_{N-1}$  be a portion of a zero-mean stationary process with spectral density function  $S(f)$ . A data taper is a sequence  $\{h_n, n = 0, \dots, N-1\}$  of length  $N$ ,  $\sum_{n=0}^{N-1} |h_n|^2 = 1$ , that is multiplied with the portion to form the sequence  $h_0 Z_0, h_1 Z_1, \dots, h_{N-1} Z_{N-1}$ . By performing a spectral estimate  $\hat{S}^{(d)}(f)$  on this new sequence in the same way that the periodogram was calculated then it can be shown that

$$E \left\{ \hat{S}^{(d)}(f) \right\} = \int_{-f_N}^{f_N} \mathcal{H}(f - f') S(f') df'$$

where

$$\mathcal{H}(f) \equiv \Delta \left| \sum_{n=0}^{N-1} h_n e^{-i2\pi f n \Delta} \right|^2.$$

CSDF estimators are defined in an analogous way. Here  $^{(d)}$  denotes a *direct spectral estimator*, the common terminology for tapered estimators. The key idea behind tapering is to select  $\{h_n\}$  so that  $\mathcal{H}(\cdot)$  has improved concentration in frequency over  $\mathcal{F}(\cdot)$ , thus reducing leakage and as a result bias in the spectral estimator.

While tapering reduces the bias properties of the spectral estimator, there is still the important issue of variance. It can be shown [73, section 6.6] that

the periodogram estimator  $\hat{S}^{(p)}(f)$  of the SDF  $S(f)$  is distributed

$$\hat{S}^{(p)}(f) \stackrel{d}{=} \begin{cases} S(f)\chi_2^2/2, & \text{for } 0 < f < f_N \\ S(f)\chi_1^2, & \text{for } f = 0 \text{ or } f_N \end{cases}$$

asymptotically as  $N \rightarrow \infty$ , and correspondingly

$$\text{var}\{\hat{S}^{(p)}(f)\} = \begin{cases} S^2(f), & \text{for } 0 < f < f_N \\ 2S^2(f), & \text{for } f = 0 \text{ or } f_N \end{cases}.$$

The same is true for direct spectral estimators given the autocovariance sequence satisfies a summability condition and data taper  $\{h_t\}$  is of a suitable form [6]. Such estimators for which the variance fails to vanish in the limit  $N \rightarrow \infty$  are known as inconsistent estimators.

It is desirable to detect structure in the spectral estimates, a large variance can hide important features. For this reason we want to smooth the direct spectral estimator in an attempt to reduce the variance. With reference to coherence estimation there is one other important need for smoothing as will now be demonstrated.

As has been said, to estimate the coherence for a pair of jointly SOS processes we need to use spectral estimators. However there is a fundamental problem with the use of a direct spectral estimator. Substituting the spectral estimator

$$\hat{S}_{12}^{(d)}(f) = \Delta \left( \sum_{n=0}^{N-1} h_n Z_{1,n} e^{-i2\pi f n \Delta} \right) \left( \sum_{n=0}^{N-1} h_n Z_{2,n} e^{-i2\pi f n \Delta} \right)^*$$

into the expression for the ordinary coherence estimator (2.9), the calculated value  $\hat{\gamma}_{12}^2(f)$  using direct estimators is identically unity at all frequencies  $f \in$

$[-f_N, f_N)$ , for any pair of jointly SOS processes  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$ :

$$\begin{aligned}\hat{\gamma}_{12}^2(f) &= \frac{|\hat{S}_{12}^{(d)}(f)|^2}{\hat{S}_{11}^{(d)}(f)\hat{S}_{22}^{(d)}(f)} \\ &= \frac{\left|\sum_{n=0}^{N-1} h_n Z_{1,n} e^{-i2\pi f n \Delta}\right|^2 \left|\sum_{n=0}^{N-1} h_n Z_{2,n} e^{-i2\pi f n \Delta}\right|^2}{\left|\sum_{n=0}^{N-1} h_n Z_{1,n} e^{-i2\pi f n \Delta}\right|^2 \left|\sum_{n=0}^{N-1} h_n Z_{2,n} e^{-i2\pi f n \Delta}\right|^2} \\ &= 1.\end{aligned}$$

Extending this to the  $p$ -dimensional complex-valued process  $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$  we can say the direct SDF matrix estimator for the augmented process  $\{\check{\mathbf{Z}}_t\}$  (defined in the usual way) is given as

$$\hat{\mathbf{Y}}^{(d)}(f) = \Delta \left( \sum_{n=0}^{N-1} h_n \check{\mathbf{Z}}_n e^{-i2\pi f n \Delta} \right) \left( \sum_{n=0}^{N-1} h_n \check{\mathbf{Z}}_n e^{-i2\pi f n \Delta} \right)^H.$$

Any function of the form

$$\frac{|\hat{\mathbf{Y}}_{lm}^{(d)}(f)|^2}{\hat{\mathbf{Y}}_{ll}^{(d)}(f)\hat{\mathbf{Y}}_{mm}^{(d)}(f)} \quad l, m = 1, \dots, 2p$$

takes a value of unity.

This demonstrates the fundamental problem with coherence estimation using direct spectral estimators. Statistically the answer of unity can be attributed to only a single complex degree of freedom in the calculation of the periodogram. To produce more meaningful coherence estimators from a single finite realisation of the processes it is necessary to introduce extra degrees of freedom into the estimator. This is typically done by performing an averaging or smoothing procedure on the spectral estimator.

Suppose we wish to estimate the coherence for a pair of processes at a given frequency, the coherence estimate will differ from the true value, consequently statistical procedures need to be employed to evaluate the estimator. For example we may wish to know; given the true ordinary coherence takes a

value of zero, what is the probability that the estimator is greater than a certain value? That is to say we may be interested in calculating  $P[\hat{\gamma}^2(f) > c]$  when  $\gamma^2(f) = 0$ . Additionally we may want to know what is the interval  $(a, b) \subset [0, 1]$  such that the true ordinary coherence lies within this with a  $(1 - \alpha)$  probability. To compute these it is necessary to know the distribution of the coherence estimator. This depends on the smoothing procedure adopted. With the ordinary coherence being calculated from three individual statistics in a non-linear way, determining such distributions is far from trivial.

In Sections 2.7.1 - 2.7.3 we introduce three common methods of spectral smoothing and briefly discuss some of the key statistical results.

### 2.7.1 Lag-window Spectral Estimators

One method of reducing variance in the spectral estimator is to smooth across the frequency domain. We justify this approach as follows [73, p. 235]; suppose the number of data points observed  $N$  is large enough such that we can assume the periodogram  $\hat{S}^{(p)}(f)$  to be an approximately unbiased estimator of the SDF  $S(f)$ . It can be shown that the periodogram estimator is pairwise uncorrelated at the *Fourier frequencies*  $\{f_k = k/(N\Delta) - f_N, k = 0, \dots, N - 1\}$  [73, p. 223]. If  $S(f)$  is slowly varying in the neighbourhood of, say,  $f_k$ , then

$$S(f_{k-M}) \approx \dots \approx S(f_k) \approx \dots \approx S(f_{k+M})$$

for some integer  $M > 0$ . Thus

$$\hat{S}^{(p)}(f_{k-M}), \dots, \hat{S}^{(p)}(f_k), \dots, \hat{S}^{(p)}(f_{k+M})$$

are a set of  $2M+1$  unbiased and uncorrelated estimators of  $S(f_k)$ . By averaging these, we define the spectral estimator

$$\bar{S}(f_k) = \frac{1}{2M+1} \sum_{j=-M}^M \hat{S}^{(p)}(f_{k-j}). \quad (2.10)$$

We now have

$$\text{var}\{\bar{S}(f_k)\} \approx \frac{\text{var}\{\hat{S}^{(p)}(f_k)\}}{2M+1}.$$

Using a weighting function  $\{g_j, j = -M, -M+1, \dots, M\}$  with  $\sum_{j=-M}^M g_j = 1$ , and a direct spectral estimator  $\hat{S}^{(d)}(f)$ , an estimator of the form (2.10) can be generalised to

$$\bar{S}(f_k) = \sum_{j=-M}^M g_j \hat{S}^{(d)}(f_{k-j}).$$

It is formed by a discrete convolution of  $\hat{S}^{(d)}(f)$  with  $\{g_j\}$  over a discrete set of frequencies. However because the direct spectral estimator  $\hat{S}^{(d)}(f)$  is defined for all frequencies in the range  $[-f_N, f_N]$ , it is also possible to smooth over a continuous set of frequencies using a continuous convolution. The lag window spectral estimator is defined as

$$\hat{S}^{(lw)}(f) = \int_{-f_N}^{f_N} W(f-\nu) \hat{S}^{(d)}(\nu) d\nu$$

$W(\cdot)$  is called the smoothing window. For a comprehensive discussion on the first and second order moments of the lag window spectral estimator, its asymptotic distribution, and how the smoothing window is chosen, the reader is referred to [73, sections 6.6 - 6.10], here we outline the key results.

The expected value and variance of the lag-window spectral estimator of a SDF  $S(f)$  is approximately given as

$$E\{\hat{S}^{(lw)}(f)\} \approx \int_{-f_N}^{f_N} W(f-\nu) S(\nu) d\nu \quad (2.11)$$

$$\text{var}\{\hat{S}^{(lw)}(f)\} \approx \frac{C_h S^2(f)}{B_W N \Delta} \quad (2.12)$$

where  $C_h$  is a constant that depends on the data taper and  $B_W$  is the bandwidth of the smoothing window  $W(\cdot)$  [73, p. 242]. We therefore see that in the limit of  $N \rightarrow \infty$  the variance becomes zero, thus the lag-window spectral estimator is a consistent estimator of the SDF. It is shown that  $\hat{S}^{(lw)}(f)$  is

approximately chi-squared distributed with approximately

$$2(E\{\hat{S}^{(lw)}(f)\}^2)/\text{var}\{\hat{S}^{(lw)}(f)\}$$

degrees of freedom.

### 2.7.2 Segment Averaging

It is possible to reduce the sampling variance of the periodogram by dividing the signal of interest into equal sized contiguous non-overlapping blocks, performing an individual periodogram on each block and then averaging the individual periodograms together to form a spectral estimator. The individual periodograms will be pairwise uncorrelated and hence averaging will reduce the variance of the estimator in a measurable way. To derive exact distributional results a further condition of pairwise independence is required for the individual estimators. This is not the case, however by matching moments (e.g. mean and variance) it is possible to conclude that the averaged estimator is approximately chi-squared distributed with  $2N_B$  degrees of freedom, where  $N_B$  is the number of individual signal blocks used. Welch [106] further developed the concept of block averaging with two important advancements. The first was to use a data taper to reduce leakage bias by having the spectral estimator for each individual block as a direct spectral estimator. The second advancement was that by allowing the blocks to overlap it is possible to form spectral estimators with better variance properties than using non-overlapping blocks [11]. This method is known in the literature as weighted overlapping segment averaging, or Welch's overlapping segment averaging (WOSA).

The statistical properties of coherence estimators formed from non-tapered, non-overlapping segment averaging have been researched in great detail. Using the work of Goodman [35], Carter et al [10], [11] derived distributions for the coherence estimator for Gaussian processes. We have said though that spectral estimates are more accurate using overlapping segments and hence is preferred for coherence estimation. Bortel and Sovka [4] recently derived

an approximate distribution for coherence estimates computed using WOSA spectral estimators, but their results are sub-optimal — see Chapter 5.

A method of spectral estimation which has been found to lend itself well to statistical analysis [102] are multitaper spectral estimators.

### 2.7.3 Multitapering

First introduced by Thomson [96], this method makes use of a set of multiple orthogonal tapers with which to create a set of approximately uncorrelated spectral estimators. Averaging across these individual spectral estimators provides the required smoothing operation to reduce variance and provide sufficient degrees of freedom. The multitaper spectral estimators will now be discussed in detail. It will be demonstrated how the methodology lends itself well to drawing conclusive analytical results on the statistical properties of coherence estimators.

Given an  $N$  point finite realisation  $\{Z_n, n = 0, \dots, N - 1\}$  of a random process  $\{Z_t, t \in \mathbb{Z}\}$  with sampling interval  $\Delta$ , the key to estimating the spectral density function  $S(f)$  is to use  $K$  orthogonal tapers  $\{u_{k,n}, n = 0, \dots, N - 1\}$ ,  $k = 0, \dots, K - 1$ , with which to form  $K$  approximately uncorrelated tapered series  $\{u_{k,n}Z_n, n = 0, \dots, N - 1\}$ ,  $k = 0, \dots, K - 1$ . Spectral estimates are performed by the standard direct Fourier method on each tapered series and then averaged. As has been previously stated in Section 2.7, tapers are used to suppress the side-lobes of Fejér's kernel in an effort to reduce leakage. As such, when deciding upon a set of tapers, as well as orthogonality a desirable property is that they are well concentrated in the frequency domain.

It was shown in [92] that a set of orthogonal tapers of length  $N$  that have optimal energy concentration in the frequency band  $[-W, W]$  ( $W < 1/2$ ) can be derived from the frequency domain functions that form the solutions to the following eigen-problem:

$$\int_{-W}^W \frac{\sin N\pi(f - f')}{\sin \pi(f - f')} U_k(N, W; f') df' = \lambda_k(N, W) \cdot U_k(N, W; f).$$

$U_k(N, W; f)$   $k = 0, 1, \dots, N-1$  are known as the discrete prolate spherical wave functions and are related to the discrete prolate spherical sequences (dps), or Slepian tapers  $u_{k,n}(N, W)$  via

$$u_{k,n}(N, W) = \frac{1}{\epsilon_k \lambda_k(N, W)} \int_{-W}^W U_k(N, W; f) e^{i2\pi f[n-(N-1)/2]} df$$

for  $k = 0, \dots, N-1$  and all  $n \in \mathbb{Z}^+$ .  $\epsilon_k$  is 1 for  $k$  even and  $i$  for  $k$  odd. The Slepian tapers are the real-valued sequences of length  $N$  that are most concentrated in the bandwidth  $[-W, W]$ .  $\lambda_k^2(N, W)$  is the proportion of the energy of the taper  $\{u_{k,n}, n = 0, \dots, N-1\}$  in this bandwidth and we order them  $\lambda_0^2(N, W) > \lambda_1^2(N, W) > \dots > \lambda_{N-1}^2(N, W)$ . Further to this, the tapers are pairwise orthonormal. That is to say  $\sum_{n=0}^{N-1} u_{k,n} u_{l,n} = \delta_{kl}$  where  $\delta_{kl}$  is the Kronecker delta.

From the  $N$  multiple tapers we choose the first  $K$  such that  $\lambda_k^2 \approx 1$ ,  $k = 0, \dots, K-1$ . The multitaper spectral estimator  $\hat{S}(f)$ ,  $f \in [-f_N, f_N]$  for the process  $\{Z_t\}$  is then given by

$$\hat{S}^{(mt)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k(f)$$

where

$$\hat{S}_k(f) = \Delta \left| \sum_{n=0}^{N-1} u_{n,k} Z_n e^{-i2\pi n f \Delta} \right|^2.$$

The Slepian tapers are the optimal set of tapers (in the sense of energy concentration) to use for multitapering, however in practice they are tricky to implement due to the computational requirements for their calculation. Alternative orthogonal tapers, such as the sine tapers [85], that are easier to implement have been derived. We now proceed by providing a comprehensive statistical analysis of multitaper ordinary coherence estimators for multivariate SOS complex Gaussian processes.

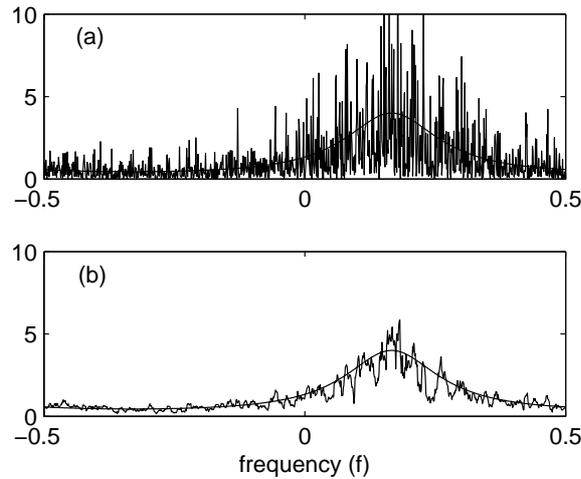


Figure 2.1: (a) Periodogram and (b) multitaper spectral estimator ( $K = 10$ ) for a SOS complex autoregressive process of order 1 — see Section 4.6.1. The spectral density function is also plotted.

## 2.8 Multitaper Spectral Estimators and Coherence

Using a matrix approach to spectral representation and Goodman's work on complex random vectors outlined in Section 2.4, we are now in a position to draw conclusions with regard to the statistical properties of coherence estimators for complex Gaussian vector-valued processes using the multitaper approach to spectral estimation.

Suppose we have a finite  $N$  point realisation  $\{\mathbf{Z}_n = [Z_{1,n}, \dots, Z_{p,n}]^T, n = 0, \dots, N - 1\}$  of a zero-mean complex Gaussian  $p$ -dimensional vector-valued process, sampled with interval  $\Delta$ . With a set of  $K$  orthogonal tapers we form the product  $\{u_{k,n}Z_{l,n}\}$  of the  $k$ th taper with the  $l$ th complex-valued process and compute its Fourier transform

$$J_{k,l}(f) = \Delta^{1/2} \sum_{n=0}^{N-1} u_{k,n} Z_{l,n} e^{-i2\pi f n \Delta}.$$

We now construct the vector

$$\begin{aligned}\check{\mathbf{J}}_k(f) &= \Delta^{1/2} \sum_{t=0}^{N-1} u_{j,n} \check{\mathbf{Z}}_n e^{-i2\pi f n \Delta} \\ &= [J_{k,1}(f), \dots, J_{k,p}(f), J_{k,1}^*(-f), \dots, J_{k,p}^*(-f)]^T\end{aligned}\quad (2.13)$$

which is the Fourier transform of the augmented vector

$$\check{\mathbf{Z}}_n = [Z_{1,n}, \dots, Z_{p,n}, Z_{1,n}^*, \dots, Z_{p,n}^*]^T.$$

In (2.13) we make use of the fact that if  $G(f)$  is the Fourier transform of a sequence  $g_n$ , then the Fourier transform of the sequence  $g_n^*$  is  $G^*(-f)$ .

For  $K$  orthogonal tapers the multitaper estimator  $\hat{\Upsilon}^{(mt)}(f)$  of the SDF matrix  $\Upsilon(f)$  is conveniently expressed as

$$\hat{\Upsilon}^{(mt)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \check{\mathbf{J}}_k(f) \check{\mathbf{J}}_k^H(f) = \begin{bmatrix} \hat{\mathbf{S}}^{(mt)}(f) & \hat{\mathbf{R}}^{(mt)}(f) \\ \hat{\mathbf{R}}^{(mt)*}(f) & \hat{\mathbf{S}}^{(mt)}(-f) \end{bmatrix}.$$

There are four important properties of the vectors  $\check{\mathbf{J}}_k(f)$ ,  $k = 0, \dots, K-1$  that should be noted.

The first property is that for all  $k = 0, 1, \dots, K-1$ , the vector  $\check{\mathbf{J}}_k(f)$  is a complex Gaussian zero-mean vector. This conclusion can be drawn from the fact that the Fourier transform of a realisation of a complex Gaussian process is a finite linear combination of Gaussian random variables and so is itself Gaussian-distributed.

Secondly, given the multiple tapers have good frequency concentration we can assume the covariance of  $\check{\mathbf{J}}_k(f)$  is equal to the augmented SDF matrix, i.e.  $E\{\check{\mathbf{J}}_k(f) \check{\mathbf{J}}_k^H(f)\} = \Upsilon(f)$  and thus

$$\check{\mathbf{J}}_k(f) \stackrel{d}{=} \mathcal{N}_{2p}^C(\mathbf{0}, \Upsilon(f)).\quad (2.14)$$

The third property, derived in [86, pp. 72-74], is that the set of  $2p$ -dimensional vectors  $\check{\mathbf{J}}_k(f)$ ,  $k = 0, \dots, K-1$  are asymptotically (with respect to  $N \rightarrow \infty$ )

pairwise uncorrelated. That is to say

$$E\{\check{\mathbf{J}}_l(f)\check{\mathbf{J}}_m^H(f)\} = \mathbf{0} \quad l \neq m, \quad l, m = 0, \dots, K-1.$$

Armed with the above two properties, we can conclude that the set of vectors  $\check{\mathbf{J}}_k(f)$ ,  $k = 0, \dots, K-1$  are asymptotically (with respect to  $N \rightarrow \infty$ ) IID with distribution (2.14).

The remaining fourth property is stated in the following theorem below;

**Theorem 2.8.1.** *For a  $p$ -dimensional vector-valued proper or improper Gaussian process*

$$\mathbf{Z}_t = \{[Z_{1,t}, \dots, Z_{p,t}]^T, \quad t \in \mathbb{Z}\},$$

the vector  $\check{\mathbf{J}}_k(f)$  is asymptotically proper (in the sense  $N \rightarrow \infty$ ) for all  $k = 0, \dots, K-1$ .

For the proof of Theorem 2.8.1 see [86, section 4.1.2].

Summarising, for a proper or improper zero-mean complex vector-valued Gaussian process  $\{\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{p,t}]^T, t \in \mathbb{Z}\}$ , the set of vectors  $\{\check{\mathbf{J}}_k(f), k = 0, \dots, K-1\}$  are complex Gaussian each with covariance  $\Upsilon(f)$ , asymptotically IID and asymptotically proper. Using Theorem 2.4.1 we can now conclude

$$\check{\mathbf{J}}_k(f)\check{\mathbf{J}}_k^H(f) \stackrel{d}{=} \mathcal{W}_{2p}^C\{1, \Upsilon(f)\},$$

and

$$\sum_{k=0}^{K-1} \check{\mathbf{J}}_k(f)\check{\mathbf{J}}_k^H(f) \stackrel{d}{=} \mathcal{W}_{2p}^C\{K, \Upsilon(f)\},$$

giving the estimator for the augmented SDF matrix  $\hat{\Upsilon}^{(mt)}(f)$  to be asymptotically  $2p$ -dimensional complex Wishart distributed with  $K$  degrees of freedom and mean  $\Upsilon(f)$ .

We now state the following theorem [35];

**Theorem 2.8.2.** *If the random  $p \times p$  matrix  $\hat{\Theta}$  is distributed*

$$\hat{\Theta} \stackrel{d}{=} \mathcal{W}_p^C\{K, \Theta\},$$

then for any  $l, m = 1, \dots, p$ ,  $l \neq m$ , the statistic  $\hat{\theta}_{lm} = |\hat{\Theta}_{lm}|^2 / [\hat{\Theta}_{ll}\hat{\Theta}_{mm}]$  has PDF

$$g_{\hat{\theta}_{lm}}(x; K, \theta_{lm}) = (K-1)(1-\theta_{lm}^2)^K(1-x)^{K-2} {}_2F_1(K, K; 1; \theta_{lm}x),$$

where  $\theta_{lm} = |\Theta_{lm}|^2 / [\Theta_{ll}\Theta_{mm}]$ .

${}_2F_1(\alpha_1, \alpha_2; \beta_1; z)$  is the hypergeometric function with 2 and 1 parameters,  $\alpha_1$ ,  $\alpha_2$  and  $\beta_1$ , and scalar argument  $z$ . This is the special case of the generalised hypergeometric series  ${}_pF_q(\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q; z)$  defined by [38, p. 1045]. The proof can be found in [35].

We therefore conclude the quantity

$$\hat{\gamma}^2(f) = \frac{|\hat{\Upsilon}_{jk}^{(mt)}(f)|^2}{\hat{\Upsilon}_{jj}^{(mt)}(f)\hat{\Upsilon}_{kk}^{(mt)}(f)}$$

that estimates a coherence  $\gamma^2(f)$  as described in Section 2.6 will asymptotically have the PDF

$$g_{\hat{\gamma}^2}(x; K, \gamma^2) = (K-1)(1-\gamma^2)^K(1-x)^{K-2} {}_2F_1(K, K; 1; \gamma^2x),$$

where the frequency  $f$  has been suppressed,  $\gamma^2$  is the true value of ordinary (or conjugate) coherence, and  $K$  is the number of tapers used in constructing the estimator. We say  $\hat{\gamma}^2$  is *Goodman distributed* with  $K$  complex degrees of freedom.

In Figure 2.2 is plotted the PDF of the Goodman distribution with 10 complex degrees of freedom for various values of the true coherence  $\gamma^2$ .

Using the work of [103] we now briefly demonstrate how the distribution can be used to interpret coherence estimates. Suppose we estimate the ordi-

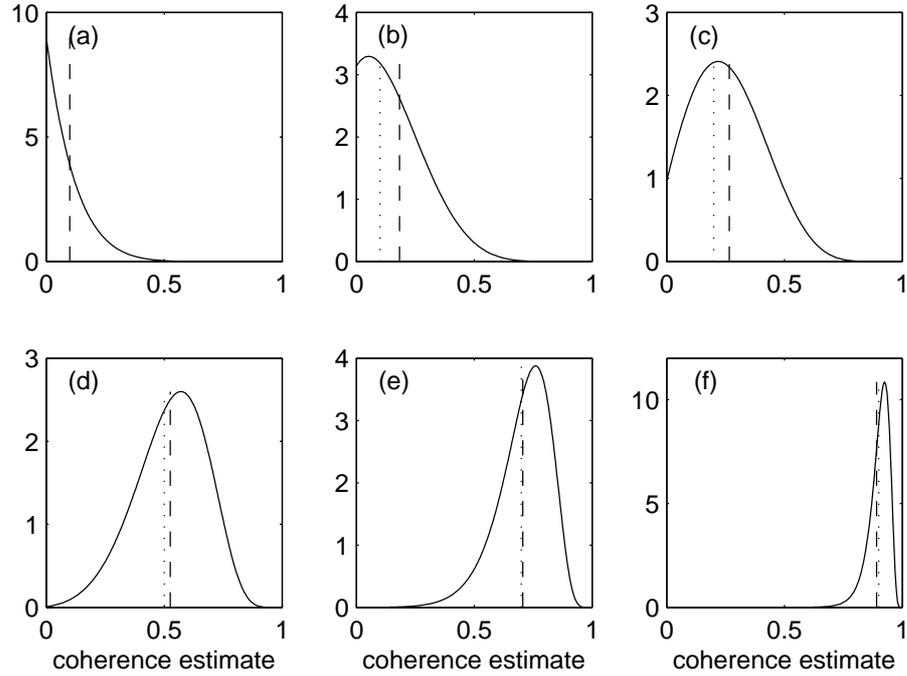


Figure 2.2: Probability density functions  $g_{\hat{\gamma}^2}(x; K, \gamma^2)$  for  $K = 10$  and a true coherence value  $\gamma^2$  of (a) 0, (b) 0.1, (c) 0.2, (d) 0.5, (e) 0.7, and (f) 0.9. The dotted line marks the true coherence and the dashed line is the expected value of the distribution.

nary coherence at a frequency  $f$  to be  $\hat{\gamma}^2(f)$ . Using the distribution we can apply an interval of confidence for the true coherence value  $\gamma^2(f)$ . We denote the cumulative probability distribution function of  $\hat{\gamma}^2(f)$  by  $G_{\hat{\gamma}^2}(x; K, \gamma^2) \equiv P[\hat{\gamma}^2 \leq x \mid K, \gamma^2(f)]$ . Given the true value  $\gamma^2$ , the  $(1 - \alpha)100\%$  confidence interval for  $\hat{\gamma}^2$  is given as  $[a_{\alpha/2}, a_{1-\alpha/2}]$  where

$$G_{\hat{\gamma}^2}(a_{\alpha/2}; K, \gamma^2) = \alpha/2, \quad G_{\hat{\gamma}^2}(a_{1-\alpha/2}; K, \gamma^2) = 1 - \alpha/2.$$

The 95% confidence limits as a function of  $\gamma^2$  are shown in Figure 2.3. The two lines are the plots  $\hat{\gamma}^2 = a_{\alpha/2}(\gamma^2)$  and  $\hat{\gamma}^2 = a_{1-\alpha/2}(\gamma^2)$ . Also shown is how to extract the  $(1 - \alpha)100\%$  confidence for the true ordinary (or conjugate) coherence  $\gamma^2$  from an estimate  $\hat{\gamma}^2 = x_0$ . This can be found by drawing a line across the plot at  $\hat{\gamma}^2 = x_0$ , the estimate's value, and finding the intersection

points with  $a_{\alpha/2}(\gamma^2)$  and  $a_{1-\alpha/2}(\gamma^2)$ .

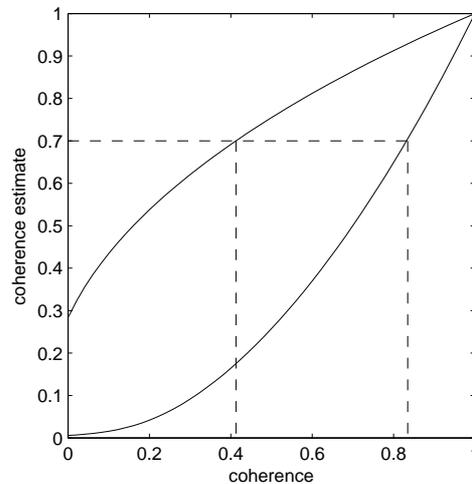


Figure 2.3: Demonstration of calculating a confidence interval of true coherence from an individual coherence estimate. In this case we find the 95% confidence interval ( $\alpha = 0.05$ ) from a coherence estimate  $\hat{\gamma}^2 = 0.7$  with 10 complex degrees of freedom.

It is also possible to use the distribution for coherence estimates to perform maximum likelihood investigations. Given an estimate  $\hat{\gamma}^2 = x_0$  of  $\gamma^2$ , we can improve our estimate of  $\gamma^2$  by locating the value  $\bar{\gamma}^2$ , say, such that  $g_{\hat{\gamma}^2}(x_0; K, \bar{\gamma}^2) = \max_{\gamma^2} g_{\hat{\gamma}^2}(x; K, \gamma^2)$ . Figure 2.4 shows the surface  $g_{\hat{\gamma}^2}(x; K, \gamma^2)$ ,  $0 \leq x, \gamma^2 < 1$ . The thick line from bottom right to top left marks the coherence estimate  $\hat{\gamma}^2 = x_0 = 0.7$ . The PDF which has a maximum at this  $x$  value is marked by the second thick line and corresponds to  $\gamma^2 = \bar{\gamma}^2 = 0.64$ . The estimate  $\bar{\gamma}^2$  is the maximum likelihood estimate

$$\left. \frac{dg_{\hat{\gamma}^2}(x_0; K, \gamma^2)}{d\gamma^2} \right|_{\gamma^2=\bar{\gamma}^2} = 0 \quad \text{with} \quad \left. \frac{d^2g_{\hat{\gamma}^2}(x_0; K, \gamma^2)}{d(\gamma^2)^2} \right|_{\gamma^2=\bar{\gamma}^2} < 0.$$

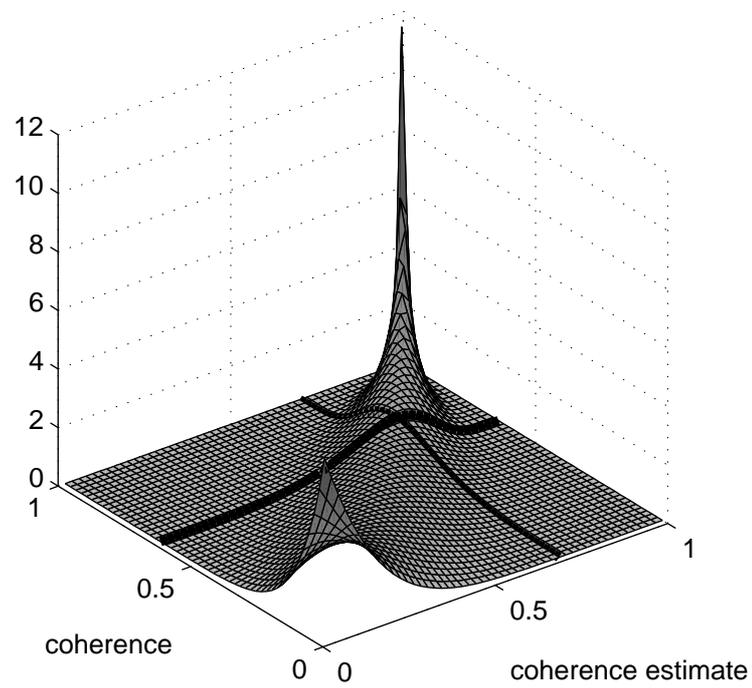


Figure 2.4: PDF surface  $g_{\gamma^2}(x; K, \gamma^2)$ ,  $0 \leq x, \gamma^2 < 1$  for  $K = 10$ . See text for an explanation.

## Chapter 3

# Wavelet Coherence for Stationary Processes

Fourier analysis has long been established as a method for decomposing a signal into its constituent frequencies. The Fourier transform of a signal represents the average contribution of a frequency across all time (or the time for which the observation occurs). For deterministic signals that behave with some regularity over time i.e. they have some underlying periodicity, then the Fourier transform is perfectly suitable for the task of frequency analysis. Similarly for stationary random processes whose second order properties remain unchanged in time and in Section 2.3 were shown to have well defined spectra, stochastic Fourier methods form a naturally suitable framework. However for deterministic signals whose behaviour does not show any regularity in time (e.g. speech), or likewise for random processes that fail to meet the conditions of second order stationarity (e.g. seismic signals), a Fourier transform acting on the signal is insufficient as it fails to preserve chronological information in its analysis.

It has been acknowledged for some time that stationary frequency analysis is insufficient. Flandrin notes [30, section 2.1] the concept of an instantaneous description of frequency first occurred in literature in 1890. In the last 30 years wavelets have come to the forefront as the method of choice in giving a

chronological (or spatial analysis) of a signal (or image). Wavelets are a tool by which a signal (or image) can be decomposed into *scales* while retaining the chronological (or spatial) information. They have been developed across a breadth of mathematical and scientific disciplines and through several separate trains of thought. Additionally, their use in interpreting data has a huge variety of applications. As a consequence, while a brief introduction to some of the key ideas will now be given, it is not possible to give the subject a proper treatment here. The reader is directed to [13], [59] and [74].

### 3.1 Representing the $L^2$ Function Space

Suppose a signal  $Z(t)$  is defined on the interval  $[-T, T] \subset \mathbb{R}$  and is square-integrable, that is to say  $Z \in L^2([-T, T])$ . The set of complex exponentials  $\{E_n(t) = e^{i2\pi nt/T}, n \in \mathbb{Z}\}$  form an orthonormal basis for  $L^2([-T, T])$  and as such  $Z(t)$  can be represented as a linear combination of these basis functions

$$Z(t) = \sum_{n \in \mathbb{Z}} \langle Z, E_n \rangle E_n(t)$$

where  $\langle \cdot, \cdot \rangle$  is the  $L^2([-T, T])$  inner-product

$$\langle f, g \rangle = \frac{1}{2T} \int_{-T}^T f(t)g^*(t)dt.$$

This is the Fourier series for the  $2T$  periodic signal  $Z(t)$ . It is possible to extend the interval  $[-T, T]$  to be of any finite size, and the set of functions  $\{e^{i2\pi nt/T}, n \in \mathbb{Z}\}$  will form an orthonormal basis. However in the infinite limit,  $\{e^{i2\pi ft}, f \in \mathbb{R}\}$  does not form an orthonormal basis for the  $L^2(\mathbb{R})$ . Elements of the space  $L^2(\mathbb{R})$  must decay to zero at  $\pm\infty$  and so in the same way that the complex exponential  $e^{it}$  generates the space  $L^2([-T, T])$ , it is desirable to have a wave-like function  $\psi$  that decays at  $\pm\infty$  and generates the space  $L^2(\mathbb{R})$ .

To have an analogous basis to the complex exponentials, these wave-like functions need to be able to oscillate at a different frequencies. For compu-

tational efficiency this is achieved by *binary dilations* (i.e. by a scaling factor of  $2^j$ ,  $j \in \mathbb{Z}$ ). Additionally, if  $\psi$  is decaying, then to cover the whole of the real line it needs to be able to shift along it. This is achieved by a *dyadic translation* of  $k/2^j$ ,  $k \in \mathbb{Z}$ .

It is possible to construct a function  $\psi$  such that the set of functions  $\{\psi_{jk}, j, k \in \mathbb{Z}\}$  defined as

$$\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k), \quad j, k \in \mathbb{Z}$$

form an orthonormal basis for  $L^2(\mathbb{R})$  (the scaling factor  $2^{j/2}$  ensures  $\psi_{jk}$  has unit  $L^2$  norm). That is to say, any function  $Z \in L^2(\mathbb{R})$  can be represented as

$$Z(t) = \sum_j \sum_k c_{jk} \psi_{jk}(t)$$

where

$$c_{jk} = \langle Z, \psi_{jk} \rangle = 2^{j/2} \int_{-\infty}^{\infty} Z(t) \psi^*(2^j t - k) dt.$$

(Note: suitable functions  $\psi$  considered here can be complex valued, hence the correctly placed conjugate.) Such functions are termed orthogonal or discrete wavelets (due to the discretised decomposition of the signal). Examples include the Haar wavelet [43] and Daubechies wavelet [23]. It is this concept of being able to represent functions using localised wave-like functions that have been scaled and translated in time that forms the basis of wavelet analysis.

## 3.2 Time-Frequency Analysis

As has been said, for signals that show a lack of regularity over time there is the need for a framework that provides a chronological frequency representation. One obvious solution is to use the short time Fourier transform (STFT) which essentially partitions the signal into smaller blocks of equal length, and then performs individual Fourier transforms on each block. One immediate problem we confront with this method is the problem of leakage. Suppose we have a

signal  $Z(t)$  which is observed on the interval  $[0, T]$ . We portion the signal into non-overlapping segments of equal length  $\tau$ . Then for an arbitrary segment  $Z_n(t)$  defined on interval  $[(n-1)\tau, n\tau)$ , the Fourier transform is defined as

$$\int_{-\infty}^{\infty} I_{[(n-1)\tau, n\tau)}(t) Z(t) e^{-i2\pi ft} dt,$$

where

$$I_A(t) = \begin{cases} 1 & t \in A \subset \mathbb{R} \\ 0 & \text{otherwise.} \end{cases} \quad (3.1)$$

From Fourier theory we know that

$$\begin{aligned} \mathcal{F}\{I_{[(n-1)\tau, n\tau)}(t) Z(t)\} &= \mathcal{F}\{I_{[(n-1)\tau, n\tau)}(t)\} * \mathcal{F}\{Z(t)\} \\ &= e^{i\pi n\tau} \frac{\sin(\pi f\tau)}{\pi f} * \mathcal{Z}(f) \end{aligned}$$

where  $*$  represents convolution,  $\mathcal{F}\{\cdot\}$  represents the Fourier transform and  $\mathcal{Z}(f) = \mathcal{F}\{Z(t)\}$ . In the frequency domain the signal undergoes a convolution with a sinc function. For small signal portions (small  $\tau$ ) the sinc function is less concentrated and as such the leakage due to convolution is increased.

As has been discussed in Section 2.7, it is possible to use tapering to reduce leakage. The Fourier transform on the signal portion would now become

$$\int_{-\infty}^{\infty} h(t - (2n-1)\tau/2) Z(t) e^{-i2\pi ft} dt$$

where  $h(t)$  is a function that vanishes outside the interval  $[-\tau/2, \tau/2]$ . If we rewrite the transform as

$$\int_{-\infty}^{\infty} Z(t) h(t - (2n-1)\tau/2) e^{-i2\pi ft} dt,$$

we see that we are correlating the signal with the time-localised waveform  $h(t)e^{i2\pi ft}$ .

It was Gabor in 1946 [31] who first proposed this method as a route to chronological frequency analysis of a signal. By recognising that a Fourier

transform is a linear transform that correlates the signal with a set of infinitely supported complex sinusoids, he proposed that by performing a linear time-frequency transform that correlates the signal with a family of waveforms that are well concentrated in time and in frequency it is possible to retain chronological information and build a time-frequency understanding of the signal. These localised waveforms are called time-frequency ‘atoms’ and form the basis for a number of time-frequency methods. [59] provides a good overview of the topic.

One immediate problem is Heisenberg’s uncertainty principles dictates it is impossible to have perfect resolution in both frequency and time. In fact the time resolution and frequency resolution are inversely proportional to one another. As such by choosing small signal segments we can achieve high time resolution, but as a consequence frequency resolution will be poor. Similarly increasing the block size will improve frequency resolution, but will incur a loss in time resolution.

A method that has come to the forefront in the last 30 years is wavelet analysis. First developed in the early 1980s by predominantly Morlet and Grossmann in a series of papers [37], [40], [65] and [66], wavelets are a development of Gabor’s work and have a natural framework for dealing with the uncertainty principle. For the analysis of high frequencies that typically operate over short time periods we have good time resolution, while at low frequencies that typically operate over long time periods, time resolution is sacrificed to provide good frequency resolution.

A wavelet  $\psi(\cdot)$  by definition is any function that integrates to zero and is square integrable with  $\|\psi\|_2 = 1$ . We however impose one more condition on  $\psi(\cdot)$ ,

$$C_\psi \equiv \int_0^\infty \frac{|\Psi(f)|^2}{f} df \text{ satisfies } 0 < C_\psi < \infty \quad (3.2)$$

where  $\Psi(\cdot)$  is the Fourier Transform of  $\psi$ . This is known as the admissibility condition and is necessary for a signal to be recreated from its continuous wavelet transform — see Section 3.3.

Figure 3.1 gives a pictorial representation of how Gabor time-frequency atoms and wavelets differ in their decomposition of a signal into time-frequency components and demonstrates the natural time-frequency resolution adjustment that wavelets can provide. It has been stated that through discrete dilations and translations of a wave-like function it is possible to construct an orthogonal basis for the  $L^2(\mathbb{R})$  function space. This decomposition, known as the *discrete wavelet transform* (DWT), projects the signal onto the discrete set of time-frequency tilings that are depicted in this figure. It is often useful to have a more fluid approach to analysing a signal on the time-frequency plane, so instead of being restricted by the discrete set of frequency octaves and the corresponding partitioning of the time domain, analysis of any frequency at any time can be achieved. The *continuous wavelet transform* provides such a method.

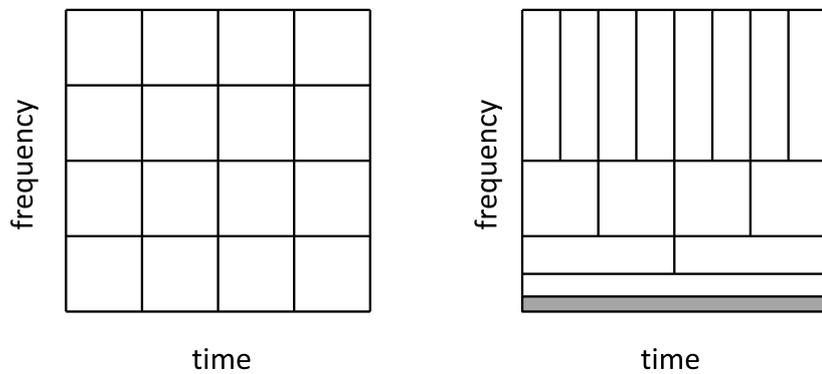


Figure 3.1: Symbolic representation of the Gabor discretisation (left) and wavelet discretisation (right) of a signal as a paving of the time-frequency plane.

### 3.3 The Continuous Wavelet Transform

For an analysing wavelet  $\psi(t)$  we define the continuous wavelet transform  $W(a, b; Z, \psi)$  of a continuous-time  $L^2(\mathbb{R})$  signal  $Z(t)$  at scale  $a \in \mathbb{R}$ ,  $|a| > 0$ , and time  $b \in \mathbb{R}$  as

$$W(a, b; Z, \psi) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} Z(t) \psi^* \left( \frac{t-b}{a} \right) dt. \quad (3.3)$$

In a vector space approach  $W(a, b; Z, \psi) = \langle Z, \psi_{a,b} \rangle$  where  $\langle \cdot, \cdot \rangle$  is the  $L^2(\mathbb{R})$  inner-product and

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi \left( \frac{t-b}{a} \right).$$

The continuous wavelet transform (CWT) is the projection of the signal  $Z(t)$  onto the function space  $\{\psi_{a,b}(t), |a| > 0, b \in \mathbb{R}\}$ .  $b$  acts as a translation along the time domain to the particular time of interest and  $a$  contracts or expands the wavelet to analyse the signal at different scales. It is these continuous domains for  $a$  and  $b$  that allow a continuous view point of the time-frequency plane. Furthermore the admissibility condition (3.2) is required for  $Z(t)$  to be reconstructed from its continuous wavelet transform [74, section 1.1].

To be able to make the connection between scale and frequency, we consider using wavelets with localised frequency support. In this instance, we can say that a wavelet  $\psi(\cdot)$  is approximately oscillating at a frequency  $f_0$  where

$$f_0 = \arg \max_f \{f [\Psi(f)]^2\} \quad (3.4)$$

is the central frequency of the wavelet. Here,  $\Psi(\cdot)$  is the Fourier transform of the wavelet  $\psi(\cdot)$ . We can therefore say that the scaled wavelet  $|a|^{-1/2} \psi(t/a)$  can be considered to be oscillating at the frequency  $f_0/a$  and as such we can interpret the wavelet transform at a scale  $a$  to be the contribution of oscillations at frequencies  $f_0/a$  to the composition of the signal at the time of interest  $b$  [59, p. 82], thus allowing a time-frequency wavelet analysis of a signals. It is often assumed that  $a$  is positive, however here we consider the more general case in

(3.3). This is more suited to complex signals whose spectra are non-symmetric about the zero-frequency axis.

In Figure 3.2 we plot the Morlet wavelet in the time domain (left) and the frequency domain (right), for varying scales. The Morlet wavelet defined as [9, p. 139]

$$\psi(t) = \pi^{-1/4} e^{-t^2/2} e^{i2\pi t}$$

is a complex exponential localised by a Gaussian function. We use this simple wavelet to demonstrate the time-frequency trade off. We can see that an increase in scale elongates the wavelet in the time domain but concentrates the wavelet in the frequency domain. Vice-versa, a decrease in scale spreads the wavelet in the frequency domain but concentrates the wavelet in the time domain. Figure 3.3 provides a visualisation of how the CWT can analyse arbitrary points of the time-frequency plane, and the time-frequency resolution trade-off that occurs.

### 3.3.1 The Continuous Wavelet Transform in Discrete-Time

Up until this point the theory of wavelets has been presented for continuous-time signals. Here we present a CWT that is suitable for discrete-time signals, achieved via the inverse discrete Fourier transform as discussed in [70].

For a continuous-time (real or complex-valued) process  $Z(t)$ , and analysing wavelet  $\psi(\cdot)$ , the CWT at a scale  $a$  and time  $b$  can be written in terms of the inverse Fourier transform

$$W(a, b; Z, \psi) = |a|^{1/2} \int_{-\infty}^{\infty} \mathcal{Z}(f) \Psi^*(af) e^{i2\pi fb} df \quad (3.5)$$

where  $\mathcal{Z}(f)$  is the Fourier transform of  $Z(t)$  at frequency  $f \in \mathbb{R}$ . Suppose we sample  $Z(t)$  at intervals of  $\Delta$ . We assume that the sampling interval is fine enough as to avoid aliasing (i.e.  $\mathcal{Z}(f) = 0$  for  $|f| > f_N$ ) and use the shorthand notation  $Z_t = Z(t\Delta)$ ,  $t = 0, \dots, N - 1$ , where  $N$  is assumed even. (3.5) can be

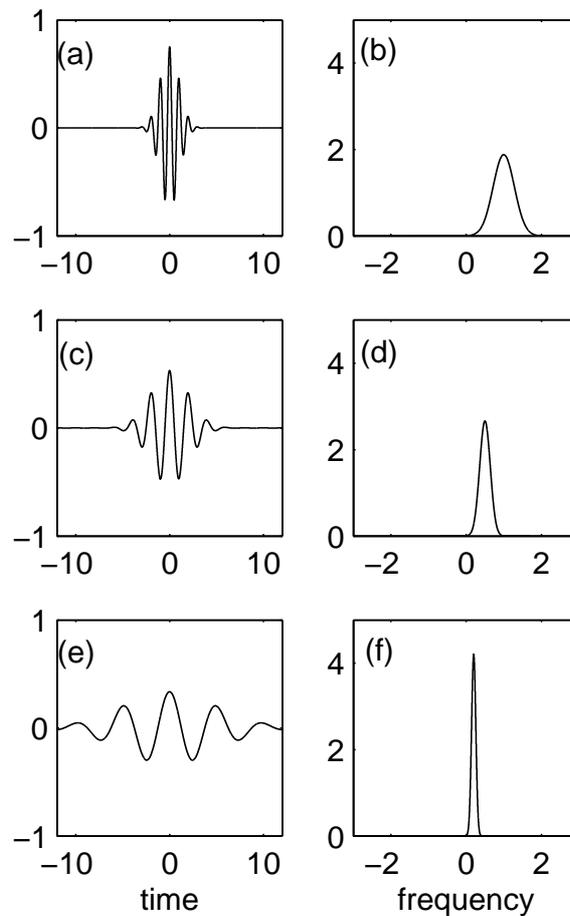


Figure 3.2: The real part of the Morlet wavelet in the time domain (left) and frequency domain (right) for scales (a),(b)  $a = 1$ , (b),(c)  $a = 2$  and (d),(e)  $a = 5$ .

discretised as [69]

$$W(a, b; Z, \psi) \approx \frac{|a|^{1/2}}{N\Delta} \sum_{l=-N/2}^{(N/2)-1} \mathcal{Z} \left( \frac{l}{N\Delta} \right) \Psi^* \left( \frac{al}{N\Delta} \right) e^{i2\pi lb/(N\Delta)}.$$

If we let  $a = a_0\Delta$  and  $b = b_0\Delta$  where  $a_0 \in \mathbb{Z} \setminus \{0\}$  and  $b_0 \in \{0, \dots, N-1\}$ , then this can be expressed as

$$W(a, b; Z, \psi) \approx \frac{|a_0|^{1/2}}{N\Delta^{1/2}} \sum_{l=0}^{N-1} \mathcal{Z}_l \Psi_l^*(a_0) e^{i2\pi lb_0/N} \quad (3.6)$$

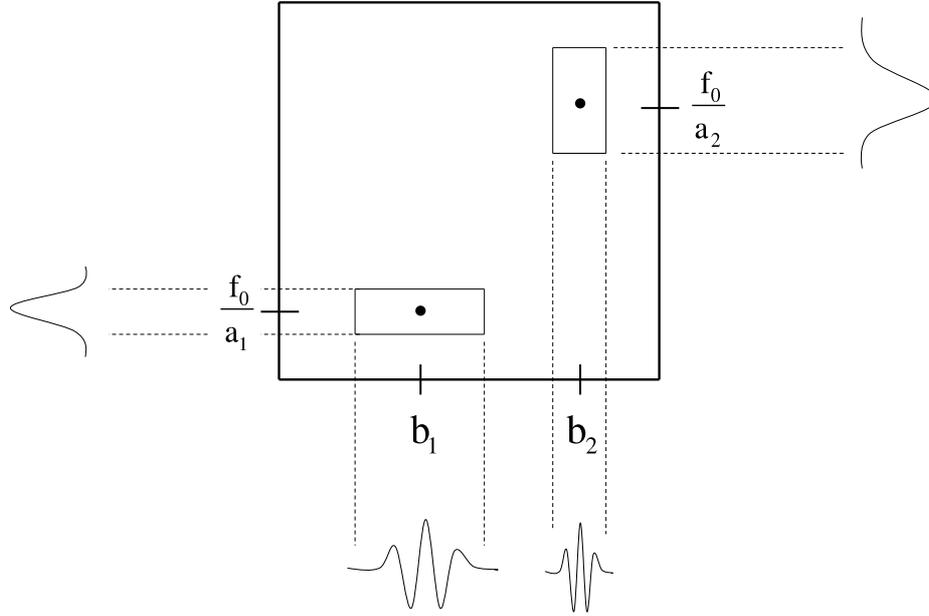


Figure 3.3: Symbolic representation of how the continuous wavelet transform can analyse an arbitrary point of the time-frequency plane, and the time-frequency trade-off that occurs.

where  $Z_l = \Delta \sum_{n=0}^{N-1} Z_n e^{-i2\pi nl/N}$ , for  $l = 0, \dots, N - 1$ , the discrete Fourier transform of  $\{Z_t\}$ , and

$$\Psi_l^*(a_0) = \begin{cases} \Psi^*(a_0 \frac{l}{N}), & l = 0, \dots, (N/2) - 1; \\ \Psi^*(a_0 \frac{l-N}{N}), & l = (N/2), \dots, N - 1. \end{cases}$$

### 3.3.2 Wraparound and the Maximum Analysis Scale

A consequence of employing the inverse discrete Fourier transform (DFT) implementation of the CWT is the cyclicity of the transform. While wavelets in general are not compactly supported, there will exist a  $t_{\min} < 0$  and  $t_{\max} > 0$  such that  $\psi(t) \approx 0$  for all  $t$  outside  $[t_{\min}, t_{\max}]$ . Hence for a scale  $|a| > 0$ ,  $\psi(t/|a|)$  will be approximately zero except on the interval  $|a|[t_{\max}, t_{\min}]$ . Since wavelets are centred at zero, at most half of this spread,  $\tau = |a|[t_{\max} - t_{\min}]/2$ ,

can be wrapped around due to cyclicity, and this will occur at the endpoints  $b = 0$  and  $b = (N - 1)\Delta$ . The highest scale at which everything is ruined by the cyclicity is when the wrapping on both sides meet; this occurs when  $|a| = a_{\max}$ , where  $2a_{\max} [t_{\max} - t_{\min}] / 2 = N\Delta$  or  $a_{\max} = N\Delta / [t_{\max} - t_{\min}]$ . This gives a corresponding minimum analysis frequency of  $f_0 / a_{\max}$ . While this defines the absolute upper bound on  $|a|$ , in practice we usually truncate plots at a scale  $|a|$  less than  $a_{\max}$ .

For each  $|a| < a_{\max}$  we also need to specify a valid range of  $b$ . We avoid wraparound provided  $\frac{a}{2}(t_{\max} - t_{\min}) < b < N\Delta - \frac{a}{2}(t_{\max} - t_{\min})$ .

### 3.3.3 Calculating the Minimum Analysis Scale

Olhede and Walden show in [71] that for  $a > 0$ , if  $a/(2\Delta)$  exceeds  $f'$  the frequency such that  $\Psi(f) \approx 0$  for  $f > f'$ , then

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} a^{1/2} \Psi(af) e^{i2\pi fb} df = \frac{1}{a^{1/2}} \bar{\psi}(b/a)$$

for  $a > 2\Delta f'$ , where  $\bar{\psi}(\cdot)$  denotes the wavelet obtained by inverse Fourier reconstruction over the truncated domain  $[-f_{\mathcal{N}}, f_{\mathcal{N}}] \subset (\infty, \infty)$ . Now consider the wavelet obtained,  $a^{-1/2} \tilde{\psi}(b/a)$  say, by inverse DFT reconstruction from

$$\Psi\left(a_0 \frac{l}{N}\right) = \Psi\left(a \frac{l}{N\Delta}\right).$$

We have, for  $N$  even

$$\frac{1}{a^{1/2}} \tilde{\psi}(b/a) \equiv \frac{a^{1/2}}{N\Delta} \sum_{t=0}^{(N/2)-1} \Psi\left(a \frac{l}{N\Delta}\right) e^{i2\pi lb_0/N}.$$

It can be shown that

$$\lim_{N \rightarrow \infty} \frac{1}{a^{1/2}} \tilde{\psi}(b/a) = \frac{1}{a^{1/2}} \bar{\psi}(b/a),$$

and hence as  $N \rightarrow \infty$ , we have

$$\frac{1}{a^{1/2}}\tilde{\psi}(b/a) \rightarrow \frac{1}{a^{1/2}}\psi(b/a),$$

provided  $a > 2\Delta f'$ ; this condition thus defines our minimum scale value by setting  $a_{\min} = 2\Delta f'$ . This gives us a maximum analysis frequency of  $f_{\max} = f_0/a_{\min} = f_0/2\Delta f'$ . An analogous argument holds for  $a < 0$ , leading us to require  $|a| > a_{\min}$ .

### 3.4 The Wavelet Spectrum and Wavelet Coherence Measure

The following theory was originally developed in the literature for real-valued processes. The extension to complex-valued processes is a natural one and considered here.

The wavelet transform  $W(a, b; Z, \psi)$  for a real or complex-valued process  $\{Z_t\}$  is in general complex-valued. In an analogy to Fourier analysis, we can readily define a wavelet power spectrum as  $|W(a, b; Z, \psi)|^2$ . Suppose we analyse a random process  $\{Z_t\}$  then we say  $|W(a, b; Z, \psi)|^2$  is an estimator for the true wavelet power spectrum which we define as  $E\{|W(a, b; Z, \psi)|^2\}$  [60]. In Fourier analysis the spectral density function gives the power of the stationary random process that can be attributed to variation at a particular frequency. The wavelet spectrum's connection to the variance of a stationary process is not so apparent. It has been shown in [98] that for a white noise process  $\{\epsilon_t\}$  with  $E\{\epsilon_{t+\tau}\epsilon_t^*\} = \sigma^2\delta_{\tau,0}$ , then the expected value of the wavelet spectrum is  $\sigma^2$ . Torrence and Compo [98] use this for comparison purposes. Suppose wavelet analysis is performed on a stationary process  $\{Z_t\}$  of known variance  $\sigma_Z^2$ , then  $|W(a, b; Z, \psi)|^2/\sigma_Z^2$  gives a measure of the power at scale  $a$  and time  $b$  relative to white noise. Further to this they show that for a Gaussian zero-mean SOS process  $\{Z_t\}$  with spectral density function  $S(f)$  then the normalised wavelet

spectrum is distributed as

$$\frac{|W(a, b; Z, \psi)|^2}{\sigma_Z^2} \stackrel{d}{=} \frac{1}{2} S(f_0/a) \chi_2^2 \quad (3.7)$$

where  $f_0$  is the reference frequency of the wavelet — see Section 3.3. For a real-valued wavelet and process the distribution becomes  $S(f_0/a) \chi_1^2$ . The confidence interval with confidence coefficient  $1 - \theta$  for the true wavelet spectrum,  $\mathcal{W}(a, b)$  say, can be shown to be given by

$$\frac{2}{\chi_2^2(\theta/2)} |W(a, b; Z, \psi)|^2 \leq \mathcal{W}(a, b) \leq \frac{2}{\chi_2^2(1 - \theta/2)} |W(a, b; Z, \psi)|^2.$$

Torrence and Compo demonstrate how the wavelet spectrum's distribution can be used for testing a null hypothesis that the random process  $\{Z_t\}$  is a stationary process with a certain stated background spectrum. For example [98] includes a study of the El Nino's NINO3 SST time series under the null hypothesis that states it is a first order autoregressive process. The wavelet spectrum provides the statistic with which to test this null hypothesis. Using the derived distribution (3.7) we reject the null hypothesis with a significance level  $\alpha$  if the peak of the wavelet spectrum is above the  $(1 - \alpha)$  confidence level of the background spectrum.

For a pair of signals  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$ , analogously to Fourier analysis, we can define the cross-wavelet spectrum as [49]

$$W(a, b; Z_1, \psi) W^*(a, b; Z_2, \psi).$$

In general the cross-wavelet spectrum is complex, and hence one can define the cross-wavelet power spectrum  $|W(a, b; Z_1, \psi) W^*(a, b; Z_2, \psi)|$ . We say the cross-wavelet power spectrum is an estimator for the true cross-wavelet power spectrum which is defined as [60]

$$E\{|W(a, b; Z_1, \psi) W^*(a, b; Z_2, \psi)|\}.$$

This exposes common power between a pair of signals in time-scale space [39] [60]. Suppose the zero-mean Gaussian SOS processes  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  have variances  $\sigma_{Z_1}^2$  and  $\sigma_{Z_2}^2$  respectively, Torrence and Compo give the distribution of the cross-wavelet spectrum as

$$\frac{|W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)|}{\sigma_{Z_1}\sigma_{Z_2}} \stackrel{d}{=} \frac{\Phi_\nu(\theta)}{\nu} \sqrt{S_{11}(f_0/a)S_{22}(f_0/a)}$$

where  $\Phi_\nu(\theta)$  is the confidence level associated with the probability  $\theta$  for a PDF defined by the square root of the product of two chi-squared distributions.  $S_{11}(f)$  and  $S_{22}(f)$  are the respective spectral density functions. If both the process and the wavelet are real-valued then  $\nu = 1$ , otherwise  $\nu = 2$ .

We have stated in Section 2.7 that the periodogram is itself chi-squared distributed with two degrees of freedom. The variance of the periodogram fails to converge to zero for an infinite time series and hence is an inconsistent estimator. To construct a consistent estimator smoothing needs to be applied to the periodogram. If we consider the wavelet power spectrum  $|W(a, b; Z, \psi)|^2$  to be an estimator of the true wavelet power spectrum, and the cross-wavelet power spectrum  $|W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)|$  to be an estimator of the true cross-wavelet power spectrum, then these are also inconsistent estimators [60]. To make the estimator consistent it is necessary to smooth the wavelet spectrum estimator. Smoothing introduces extra degrees of freedom into the estimator that reduces variance and narrows confidence intervals. Smoothing becomes particularly important in coherence estimation, which we discuss now.

The cross-wavelet power spectrum is an insufficient tool for analysing correlations between a pair of processes in time-scale space. Suppose one process has a flat spectrum, and the other's spectrum is strongly peaked, then the cross-wavelet spectrum may contain peaks that have no relation to the nature of correlation between the pair of processes. It is therefore necessary to normalise with respect to the individual wavelet power spectra. Liu [56]

introduced the quantity

$$\frac{|W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)|^2}{|W(a, b; Z_1, \psi)|^2|W(a, b; Z_2, \psi)|^2}$$

as a wavelet coherence estimator with which to measure the linear dependency of a pair of random processes  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  at a scale  $a$  and time  $b$ .

Importantly, failure to smooth the wavelet spectrum terms independently results in a wavelet coherence estimate of one for all scales  $a$  and times  $b$ . A sensible wavelet coherence estimator is therefore [39], [77]

$$\hat{\gamma}^2(a, b) = \frac{|\mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)\}|^2}{\mathcal{S}\{|W(a, b; Z_1, \psi)|^2\}\mathcal{S}\{|W(a, b; Z_2, \psi)|^2\}} \quad (3.8)$$

where  $\mathcal{S}\{\cdot\}$  represents a smoothing operation. The notation  $\gamma^2(\cdot, \cdot)$  represents wavelet coherence, whereas we recall  $\gamma^2(\cdot)$  represents the Fourier spectral coherence. The question of how to smooth the wavelet coherence estimator, and the resulting statistical properties, has proven problematic and the influential paper [98] calls for further research on the properties of wavelet coherence.

Wavelet coherence estimators have been used extensively for hypothesis testing purposes. For example in [7], [42] and [60] the null hypothesis states that the pair of processes are jointly stationary, independent white noise processes (i.e. they have a coherence of zero). Significant values of wavelet coherence are used to reject the null hypothesis and infer that there exists coherence in time-scale space. Similarly, [39], [77] and [98] look for intermittent coherent oscillations against a background coloured spectra with the background processes specifying the null hypothesis. Testing against a stationary background as opposed to a nonstationary background is preferred because a non-arbitrary test is required. There are difficulties in choosing a nonstationary structure with which to test against, mainly because of the vast array of non-stationary behaviour that could be considered — see Chapter 6. A stationary background provides a useful standard and forms a sound basis on which to decide upon further tests. It can also be the case that the stationary assumption holds

at certain regions in time and fails to hold at others. Testing against a stationary background with a time localised coherence estimator allows the test to be performed at all time points, and therefore zones where the stationary assumption holds and zones where it does not become immediately apparent.

For such tests it is necessary to know the distribution of the test statistic, namely the wavelet coherence, under the null hypothesis. This is dependent on the smoothing procedure adopted. Used in combination with a Morlet wavelet (introduced in Section 3.3), a smoothing method suggested in [98] involves smoothing over time  $b$ , followed by smoothing over scale  $a$ . This has been widely applied [3], [39], [77] and [99] for hypothesis testing purposes. A simpler smoothing operation over just time has been used in [53], [101].

Until now the statistical properties of smoothed wavelet coherence estimators have been resistant to analytical study. While we can derive the distribution of the wavelet and cross-wavelet spectra, the smoothing operation makes the situation more complicated. Consequently in the referenced studies, distributions of the wavelet coherence estimator were determined experimentally through Monte Carlo simulations of the processes used in specifying the null hypothesis.

A detailed analytical statistical study of wavelet coherence has proved difficult because adjacent values of the wavelet spectrum on the time-scale grid are correlated [60], and thus when averaging together the wavelet spectrum for a localised region of time-scale space, the resulting statistic is no longer chi-squared distributed. Maraun and Kurths go as far to say that [60]

“an analytical test statistic is highly non-trivial if not impossible”.

This problem is now rectified. A detailed statistical analysis for two different methods of smoothing the individual terms in the wavelet coherence estimator (3.8) is now provided. We look at using a set of orthogonal Morse wavelets in Chapter 4, and then at the more traditional method of smoothing in the time domain in Chapter 5. These results are also found in Cohen and Walden [14] and [15], respectively.

## Chapter 4

# A Statistical Analysis of Morse Wavelet Coherence

In Chapter 3 the concept of a wavelet coherence estimator, and the current problems in its formulation and derivation of statistical properties were highlighted, with the key point being that smoothing in the time domain, frequency domain, or both makes wavelet coherence estimators less amenable to statistical study. This was a similar problem for stationary Fourier based coherence estimators. In Section 2.8 it was shown that by averaging across a set  $\ell^2$  orthogonal tapers, it is possible to fully derive the distribution of the coherence estimator (the Goodman distribution) and that degrees of freedom can be calculated readily.

Suppose we have a set of  $K$  wavelets  $\{\psi_k, k = 0, \dots, K-1\}$  that are pairwise orthogonal in  $L^2(\mathbb{R})$ , then for a pair of processes  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  a wavelet coherence estimator taking the form

$$\hat{\gamma}^2(a, b) = \frac{\left| \sum_{k=0}^{K-1} W(a, b; Z_1, \psi_k) W^*(a, b; Z_2, \psi_k) \right|^2}{\left[ \sum_{k=0}^{K-1} |W(a, b; Z_1, \psi_k)|^2 \right] \left[ \sum_{k=0}^{K-1} |W(a, b; Z_2, \psi_k)|^2 \right]}$$

can provide the necessary averaging operation to achieve a non-trivial coherence estimate.

It was stated in Section 3.1 that when forming a direct spectral estimator of a stationary process it is desirable to choose a sequence to taper the observed time series that has good energy concentration in the frequency domain to reduce leakage effects. For a frequency band  $[-W, W]$ ,  $W < 1/2$ , the sequences of  $N$  points that have optimal concentration are the Slepian tapers. Furthermore the Slepian tapers are mutually orthogonal. A set of orthogonal functions that can be used as wavelets were first derived from the Slepian tapers [55]. The  $k$ th order complex Slepian pseudo-wavelet is defined as

$$\psi_{k,n}^S(N, W) = \frac{1}{\sqrt{2}} [u_{2k,n}(N, W) + iu_{2k+1,n}(N, W)]$$

where  $u_{k,n}(N, W)$  is the  $k$ th order Slepian taper of length  $N$  with respect to the frequency interval  $[-W, W]$ . While this function is defined discretely it can be used as a wavelet in the discrete-time form of the continuous wavelet transform — see Section 3.3.2.

The Slepian tapers only optimise energy concentration in a designated frequency band and have no consideration for the time domain. For wavelets it is desirable to consider the time and frequency domains simultaneously by maximising the energy in a bounded subset of  $\mathbb{R}^2$ . In an analogy to the derivation of the Slepian tapers, Daubechies and Paul [24], [25] introduced localisation operators which can be used in the time-frequency plane.

## 4.1 The Multiple Morse Wavelets

The discussion in this section is taken from [68]. In [24] the ‘coherence state’ associated with a point  $(s, f)$  in the time-frequency space is given as  $\phi_{s,f}(t) = e^{i2\pi ft} \phi(t-s)$ . The choice  $\phi(t) = \pi^{-1/4} e^{-t^2/2}$  (the Gabor wave function) attains the Heisenberg-Gabor inequality:  $\phi_{s,f}(t)$  has the best localisation around a point  $(s, f)$ , and is used here. The following resolution of identity holds for  $Z \in L^2(\mathbb{R})$

$$Z(t) = \int \int_{(s,f) \in \mathbb{R}^2} \phi_{s,f}(t) \langle \phi_{s,f}, Z \rangle ds df. \quad (4.1)$$

By restricting the signal to a domain  $\mathcal{D}$  of time-frequency space we can define the operator  $Z(t) \rightarrow (\mathcal{P}_{\mathcal{D}}Z)(t)$ , where

$$(\mathcal{P}_{\mathcal{D}}Z)(t) = \int \int_{(s,f) \in \mathcal{D}} \phi_{s,f}(t) \langle \phi_{s,f}, Z \rangle ds df.$$

The ratio of the energy of the signal limited to the domain  $\mathcal{D}$  to that of the original signal is a real-valued quantity,  $\mu(\mathcal{D})$  say, given by

$$\mu(\mathcal{D}) = \frac{\langle \mathcal{P}_{\mathcal{D}}Z, \mathcal{P}_{\mathcal{D}}Z \rangle}{\langle Z, Z \rangle}.$$

For bounded  $\mathcal{D}$  the operator  $\mathcal{P}_{\mathcal{D}}$  is a bounded, compact and self adjoint operator on a Hilbert space [24] and hence by the Hilbert-Schmidt Theorem [84, pp. 203, 209] there is a complete orthonormal basis  $\{h_k(t)\}$  for  $L^2(\mathbb{R})$  such that  $(\mathcal{P}_{\mathcal{D}}h_k)(t) = \lambda_k h_k(t)$ . As the spectrum of  $\mathcal{P}_{\mathcal{D}}$  is purely discrete, we can order the eigenvalues in size  $\lambda_0 \leq \lambda_1 \leq \dots$ , and further it can be shown the signal that maximises the energy concentration in  $\mathcal{D}$  is  $h_0(t)$ .

When  $\mathcal{D}$  corresponds to the disc of radius  $R$  about the origin in time-angular frequency space, i.e.  $\mathcal{D}_R = \{(s, f) : s^2 + (2\pi f)^2 \leq R^2\}$ , then it has been shown [24] that the orthogonal set of eigenfunctions take the form of scaled Hermite polynomials. Suppose instead we consider the domain

$$\mathcal{D}_{C,\beta,\gamma} = \left\{ (s, f) \in \mathbb{R}^2 : \left( \frac{s}{C_2} \right)^2 \left( \frac{C_1}{|2\pi f|} \right)^{2\gamma-2} + \left( \frac{C_1}{|2\pi f|} \right)^{2\gamma} + 1 \leq 2C \left( \frac{C_1}{|2\pi f|} \right)^{\gamma} \right\}.$$

$C_1$  and  $C_2$  are given as

$$C_1 = \frac{2^{-1/\gamma} \Gamma\left(r + \frac{1}{\gamma}\right)}{\Gamma(r)}; \quad C_2 = \frac{\beta \gamma^{-1} 2^{1/\gamma} \Gamma\left(r - \frac{1}{\gamma}\right)}{\Gamma(r)}$$

where  $\Gamma(\cdot)$  is the gamma function — see Figure 4.1. For the Hermitian operator  $\mathcal{P}_{\mathcal{D}_{C,\beta,\gamma}}$  we label the eigenvalues  $\lambda_{k;\beta,\gamma}^2(C)$   $k = 0, 1, \dots$  where  $\lambda_{0;\beta,\gamma}^2(C) \leq$

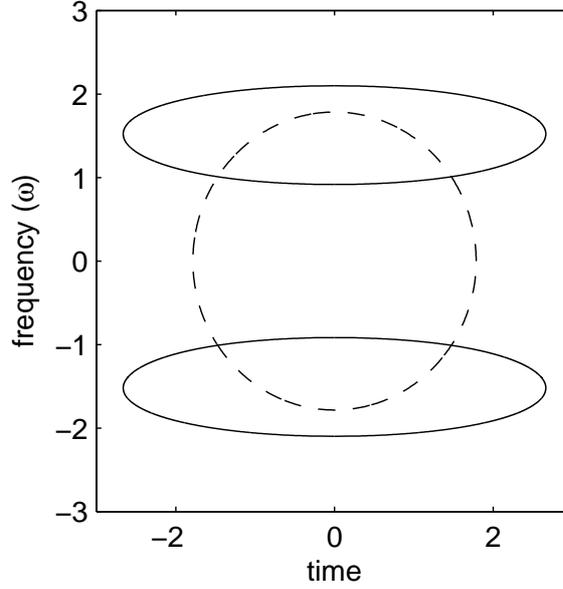


Figure 4.1: Morse wavelet domain  $\mathcal{D}_{C,\beta,\gamma}$  (solid) for  $C = 10$ ,  $\beta = 8$  and  $\gamma = 3$ . The disc of equal area is also plotted (dashed).

$\lambda_{1;\beta,\gamma}^2(C) \leq \dots$ . Olhede and Walden [68] showed that for particular values of  $\beta$  and  $\gamma$  the energy concentration of the eigenfunctions of the operator  $\mathcal{P}_{\mathcal{D}_{C,\beta,\gamma}}$ , given by the square of the corresponding eigenvalues  $\lambda_{0;\beta,\gamma}^2(C) \leq \lambda_{1;\beta,\gamma}^2(C) \leq \dots$ , outperform those of the operator  $\mathcal{P}_{\mathcal{D}_R}$  when the area of  $\mathcal{D}_{C,\beta,\gamma}$  equals that of  $\mathcal{D}_R$ . The orthogonal set of eigenfunctions for this set of domains are known as the *generalised Morse wavelets*.

The generalised Morse wavelets possess an added property which makes them attractive for the analysis of complex signals. For each eigenvalue in the spectrum of  $\mathcal{P}_{\mathcal{D}_{C,\beta,\gamma}}$  there exists two corresponding eigen-solutions, an analytic and anti-analytic form. The analytic solution for eigenvalue  $\lambda_{k;C,\beta,\gamma}$  is denoted  $\psi_{k;\beta,\gamma}^+(\cdot)$  and has the Fourier transform

$$\Psi_{k;\beta,\gamma}^+(f) = \sqrt{2}A_{k;\beta,\gamma}(2\pi f)^\beta e^{-(2\pi f)^\gamma} L_k^c(2[2\pi f]^\gamma)$$

for  $f > 0$  and zero otherwise. Here,  $c = r - 1$ ,

$$A_{k;\beta,\gamma} = \sqrt{[\pi\gamma 2^r \Gamma(k+1)/\Gamma(k+r)]},$$

and  $L_k^c(\cdot)$  denotes the generalised Laguerre polynomial

$$L_k^c \sum_{m=0}^k (-1)^m \frac{\Gamma(k+c+1)}{\Gamma(c+m+1)\Gamma(k-m+1)} \frac{x^m}{m!}.$$

It is given the term *analytic* because its frequency support is solely on the positive half of the real line;

$$\Psi_{k;\beta,\gamma}^+(f) = 0 \text{ for all } f < 0. \quad (4.2)$$

The anti-analytic solution is denoted  $\psi_{k;\beta,\gamma}^-(\cdot)$  with Fourier transform

$$\Psi_{k;\beta,\gamma}^-(f) = \sqrt{2} A_{k;\beta,\gamma} (2\pi|f|)^\beta e^{-(2\pi|f|)^\gamma} L_k^c(2[2\pi|f|]^\gamma)$$

for  $f < 0$  and zero otherwise. It is given the term *anti-analytic* because its frequency support is solely on the negative half of the real line;

$$\Psi_{k;\beta,\gamma}^-(f) = 0 \text{ for all } f > 0. \quad (4.3)$$

Some important properties of the Morse wavelets are given below [68]:

$$\Psi_{k;\beta,\gamma}^+(0) = \Psi_{k;\beta,\gamma}^-(0) = 0; \quad (4.4)$$

$$\psi_{k;\beta,\gamma}^+(t) = \psi_{k;\beta,\gamma}^{-*}(t); \quad (4.5)$$

$$\Psi_{k;\beta,\gamma}^+(f) = \Psi_{k;\beta,\gamma}^-(f); \quad (4.6)$$

and the orthogonality condition

$$\int_{-\infty}^{\infty} \psi_{k;\beta,\gamma}^+(t) \psi_{l;\beta,\gamma}^{+*}(t) dt = \delta_{lk};$$

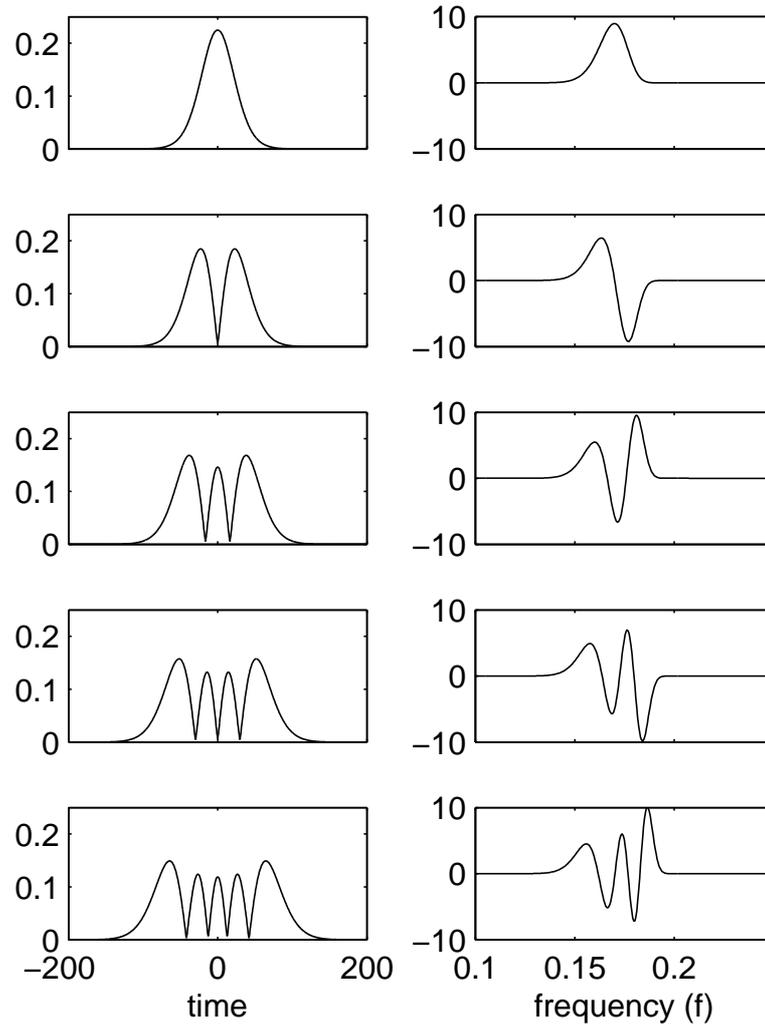


Figure 4.2: Each row of the plot shows the magnitude  $|\psi_{k,40,15}^+|$  (left) and frequency response  $\Psi_{k,40,15}^+$  (right). The rows are for (top to bottom)  $k = 0, \dots, 4$ .

and [69]

$$\psi_{k;\beta,\gamma}^+(-t) = \psi_{k;\beta,\gamma}^-(t). \quad (4.7)$$

Combining (4.5) and (4.7) we get

$$\psi_{k;\beta,\gamma}^+\left(\frac{t-b}{-a}\right) = \psi_{k;\beta,\gamma}^{+*}\left(\frac{t-b}{a}\right),$$

so that, with  $Z(t)$  denoting a complex-valued signal,

$$\begin{aligned}
W(a, b; Z^*, \psi_{k;\beta,\gamma}^+) &= \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} Z^*(t) \psi_{k;\beta,\gamma}^{+*} \left( \frac{t-b}{a} \right) dt \\
&= \left[ \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} Z(t) \psi_{k;\beta,\gamma}^{+*} \left( \frac{t-b}{-a} \right) dt \right]^* \\
&= W^*(-a, b; Z, \psi_{k;\beta,\gamma}^+). \tag{4.8}
\end{aligned}$$

Interestingly this is an analogous result to the Fourier transform of a conjugated signal.

When using the Morse wavelets, the parameters  $\beta$  and  $\gamma$  are chosen such that the wavelet functions have strong concentration properties. The first  $K$  Morse wavelets are selected such that  $\lambda_{k;\beta,\gamma}^2(C) \approx 1$  for  $k = 0, \dots, K-1$ . Throughout the rest of this chapter it will be common to fix the  $\beta$  and  $\gamma$  wavelet parameters and refer to  $\psi_{k;\beta,\gamma}^{\pm}$  as simply  $\psi_k^{\pm}$ . Further to this, to condense notation we denote the  $k$ th analytic wavelet transform of the signal  $Z(t)$  at scale  $|a| > 0$  and translation  $b \in \mathbb{R}$  as

$$W_k(a, b; Z) \equiv W(a, b; Z, \psi_{k;\beta,\gamma}^+). \tag{4.9}$$

## 4.2 The Morse Wavelet Coherence Estimator

We will demonstrate in Section 4.4.2 that the pairwise orthogonality of a set of  $K$  Morse wavelets gives rise to  $K$  approximately uncorrelated wavelet spectra and cross-spectra and that in the white noise case these are exactly uncorrelated. These can therefore be averaged to introduce extra degrees of freedom into the wavelet coherence (WCOH) calculation. The Morse WCOH estimator for the pair of complex-valued process  $\{Z_1(t)\}$  and  $\{Z_2(t)\}$  takes the form

$$\gamma^2(a, b) = \frac{\left| \sum_{k=0}^{K-1} W_k(a, b, Z_1) W_k^*(a, b, Z_2) \right|^2}{\left[ \sum_{k=0}^{K-1} |W_k(a, b, Z_1)|^2 \right] \left[ \sum_{k=0}^{K-1} |W_k(a, b, Z_2)|^2 \right]}. \tag{4.10}$$

It will now be shown that by averaging together uncorrelated wavelet spectra, this method of coherence estimation will be more amenable to statistical analysis and distributional results will naturally follow. For this work to be applicable in the physical sciences it is necessary to consider the discrete-time case and arbitrary sampling rate  $\Delta$ . We therefore now take special care in formulating the Morse wavelet transform for discrete-time processes.

### 4.3 The Morse Wavelet Transform for Discrete-Time Processes

In Section 3.3 was presented a discrete-time formulation of the continuous wavelet transform. By replacing the wavelet in (3.6) with a general  $k$ th order analytic Morse wavelet, for a scale  $a = a_0\Delta$  and  $b = b_0\Delta$ ,  $a \in \mathbb{Z} \setminus \{0\}$  and  $b \in \{0, \dots, N - 1\}$ , we have

$$W_k(a, b; Z, \psi^+) = \frac{|a_0|^{1/2}}{N(\Delta t)^{1/2}} \sum_{l=0}^{N-1} Z_l \Psi_{k,l}^+(a_0) e^{i2\pi l b_0/N} \quad (4.11)$$

where

$$\Psi_{k,l}^+(a_0) = \begin{cases} \Psi_k^+(a_0 \frac{l}{N}), & l = 1, \dots, \frac{N}{2} - 1 \\ \Psi_k^+(a_0 \frac{l-N}{N}), & l = \frac{N}{2}, \dots, N - 1. \end{cases} \quad (4.12)$$

This form of the wavelet transform allows computation through the use of the inverse fast Fourier transform algorithm, which has computational expense in the order of  $N \log_2 N$ . While this is slower than the DWT pyramid algorithm (which is of order  $N$  [74, p. 68]), it is comparable and provides an efficient method of calculation. There are some important considerations when using this implementation of the CWT. These were discussed in Section 3.3 and here are tailored to the multiple wavelet case.

### 4.3.1 Wraparound and the Maximum Analysis Scale

A discussion on the maximum analysis scale for a finite discrete-time process is given in Section 3.3.2. While the Morse wavelets are not compactly supported, there exists  $t_{\min}(K) < 0$  and  $t_{\max}(K) > 0$  such that  $\psi_k^+(t) \approx 0$ ,  $k = 0, \dots, K - 1$ , for all  $t$  outside  $[t_{\min}(K), t_{\max}(K)]$ . Time support of the wavelet  $\psi_k^+(t)$  increases with  $k$  — see Figure 4.2 — and consequently it is the time support of  $\psi_{K-1}^+(t)$  that needs to be considered. The maximum scale is given as  $a_{\max}(K) = N\Delta / [t_{\max}(K) - t_{\min}(K)]$ .

### 4.3.2 Calculating the Minimum Analysis Scale

The minimum analysis scale for the single wavelet was discussed in Section 3.3.3. If the cut-off frequency of a  $k$ th analytic Morse wavelet is given as  $f'_k$ , then (see Figure 4.2)  $f'_{K-1} > f'_{K-2} > \dots > f'_0$  and consequently for all  $k = 0, \dots, K - 1$ ,  $\Psi_k^+(f) \approx 0$  for  $f > f'_{K-1}$ . The minimum scale value is therefore given as

$$a_{\min}(K) = \frac{f'_{K-1}}{f_N} = 2\Delta f'_{K-1}. \quad (4.13)$$

An analogous argument holds for  $a < 0$ , leading us to require  $a_{\min}(K) < |a| < a_{\max}(K)$ .

### 4.3.3 Mapping Time-Scale to Time-Frequency

The mapping of time-scale to time-frequency for a single wavelet was given in Section 3.3, and so for the  $k$ th Morse wavelet the reference frequency as determined by (3.4) will be denoted  $f_0^{(k)}$ . To relate scale and frequency for multi-wavelet calculations we use the relationship  $f(a) = f_0/a$  where the reference frequency now takes the form [71]

$$f_0 = \arg \max_f \left\{ f \sum_{k=0}^{K-1} [\Psi_k(f)]^2 \right\}, \quad (4.14)$$

where  $\Psi_k(f)$  is the Fourier transform of the Morse wavelet  $\psi_k(t)$ .

## 4.4 The Statistics of the Morse Wavelet Coherence Estimator

In Section 2.8 the statistical properties of multitaper coherence estimators were derived. Central to the derivation was the use of matrices in representing the spectral estimators of a complex vector-valued process. It will now be shown that by taking the same approach with the CWT we are able to derive some important statistical result.

### 4.4.1 The Wavelet Spectral Matrix

We remind ourselves that for the observed portion  $\{\mathbf{Z}_n = [Z_{1,n}, \dots, Z_{p,n}]^T, n = 0, \dots, N - 1\}$  of a  $p$ -dimensional complex-valued process, the vector  $\check{\mathbf{J}}_k(f)$  was defined as

$$\check{\mathbf{J}}_k(f) = \Delta^{1/2} \sum_{n=0}^{N-1} u_{k,n} \check{\mathbf{Z}}_n e^{-i2\pi f n \Delta}$$

where  $\{u_{k,n}, n = 0, \dots, N - 1\}$  is the  $k$ th of the  $K$  orthogonal tapers and  $\{\check{\mathbf{Z}}_n, n = 0, \dots, N - 1\}$  is the augmented vector  $\{[\mathbf{Z}_n^T, \mathbf{Z}_n^H]^T, n = 0, \dots, N - 1\}$ . The matrix

$$\hat{\Upsilon}^{(mt)}(f) = (1/K) \sum_{k=0}^{K-1} \check{\mathbf{J}}_k(f) \check{\mathbf{J}}_k^H(f)$$

is the multitaper estimator of the spectral matrix  $\Upsilon(f)$  defined in (2.6). Coherence estimators are of the form

$$\frac{|\hat{\Upsilon}_{lm}^{(mt)}(f)|^2}{\hat{\Upsilon}_{ll}^{(mt)}(f) \hat{\Upsilon}_{mm}^{(mt)}(f)}, \quad l, m = 1, \dots, 2p.$$

We recognise that this vector and matrix approach can also be used with the multiple Morse wavelets. For a particular choice of wavelet parameters  $\beta$  and  $\gamma$  we use the  $K$  most concentrated analytic wavelets  $\{\psi_{k;\beta,\gamma}^+(t), k = 0, \dots, K - 1\}$ . Using the previously defined augmented vector-valued process  $\{\check{\mathbf{Z}}_t\}$ , for each  $k = 0, \dots, K - 1$ , in an analogous way to how  $\check{\mathbf{J}}_k(f)$  was defined, and by using

the notation outlined in (4.9) we can construct the wavelet transform vector

$$\begin{aligned}\check{\mathbf{W}}_k(a, b) &= W_k(a, b; \check{\mathbf{Z}}) \\ &= [W_k(a, b; Z_1), \dots, W_k(a, b; Z_p), W_k(a, b; Z_1^*), \dots, W_k(a, b; Z_p^*)]^T.\end{aligned}$$

We condense notation further by saying

$$W_{k,l}(a, b) = W(a, b; Z_l, \psi_k^+(t)).$$

Using (4.8) we can express  $\check{\mathbf{W}}_k(a, b)$  as

$$[W_{k,1}(a, b), \dots, W_{k,p}(a, b), W_{k,1}^*(-a, b), \dots, W_{k,p}^*(-a, b)]^T.$$

We now define the matrix

$$\hat{\mathbf{\Omega}}(a, b) = (1/K) \sum_{k=0}^{K-1} \check{\mathbf{W}}_k(a, b) \check{\mathbf{W}}_k^H(a, b). \quad (4.15)$$

Let  $\hat{\Omega}_{l,m}(a, b)$  be the  $(l, m)$ th element of  $\hat{\mathbf{\Omega}}(a, b)$ . For  $l \neq m$ , any function of the form

$$\frac{|\hat{\Omega}_{l,m}(a, b)|^2}{\hat{\Omega}_{l,l}(a, b)\hat{\Omega}_{m,m}(a, b)} \quad l, m = 1, \dots, 2p \quad (4.16)$$

is a WCOH estimator of type (4.10) where  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  are replaced by two component processes of the augmented process  $\{\check{\mathbf{Z}}_t\}$ . With analogous definitions to those provided for classical spectral coherence in Section 2.4, (4.16) can define three types of WCOH estimator:

1. the ordinary WCOH estimator  $\hat{\gamma}_{lm}^2(a, b)$  for the pair of processes  $\{Z_{l,t}\}$

and  $\{Z_{m,t}\}$  is defined as

$$\begin{aligned}\hat{\gamma}_{lm}^2(a, b) &= \frac{|\sum_{k=0}^{K-1} W_{k,l}(a, b)W_{k,m}^*(a, b)|^2}{\sum_{k=0}^{K-1} |W_{k,l}(a, b)|^2 \sum_{k=0}^{K-1} |W_{k,m}(a, b)|^2} \\ &= \frac{|\hat{\Omega}_{l,m}(a, b)|^2}{\hat{\Omega}_{l,l}(a, b)\hat{\Omega}_{m,m}(a, b)};\end{aligned}\quad (4.17)$$

2. the conjugate WCOH estimator  $\hat{\gamma}_{lm_*}^2(a, b)$  for the pair of processes  $\{Z_{l,t}\}$  and  $\{Z_{m,t}^*\}$  is defined as

$$\begin{aligned}\hat{\gamma}_{lm_*}^2(a, b) &= \frac{|\sum_{k=0}^{K-1} W_{k,l}(a, b)W_{k,m}^*(-a, b)|^2}{\sum_{k=0}^{K-1} |W_{k,l}(a, b)|^2 \sum_{k=0}^{K-1} |W_{k,m}(-a, b)|^2} \\ &= \frac{|\hat{\Omega}_{l,m+p}(a, b)|^2}{\hat{\Omega}_{l,l}(a, b)\hat{\Omega}_{m+p,m+p}(a, b)};\end{aligned}\quad (4.18)$$

3. the conjugate WCOH estimator  $\hat{\gamma}_{ll_*}^2(a, b)$  for the pair of processes  $\{Z_{l,t}\}$  and  $\{Z_{l,t}^*\}$  is defined as

$$\begin{aligned}\hat{\gamma}_{ll_*}^2(a, b) &= \frac{|\sum_{k=0}^{K-1} W_{k,l}(a, b)W_{k,l}^*(-a, b)|^2}{\sum_{k=0}^{K-1} |W_{k,l}(a, b)|^2 \sum_{k=0}^{K-1} |W_{k,l}(-a, b)|^2} \\ &= \frac{|\hat{\Omega}_{l,l+p}(a, b)|^2}{\hat{\Omega}_{l,l}(a, b)\hat{\Omega}_{l+p,l+p}(a, b)}.\end{aligned}\quad (4.19)$$

We now state the following theorem for the vector  $\check{\mathbf{W}}_k(a, b)$ ;

**Theorem 4.4.1.** *Let  $\{[Z_{1,0}, \dots, Z_{p,0}]^T, \dots, [Z_{1,N-1}, \dots, Z_{p,N-1}]^T\}$  be a realisation of length  $N$  of a bivariate zero-mean complex-valued Gaussian (proper or improper) SOS process  $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$  with bounded and continuous spectra. Then asymptotically, as  $N \rightarrow \infty$ ,*

(i)  $\check{\mathbf{W}}_k(a, b)$  is a proper zero-mean complex-valued Gaussian vector.

(ii) The covariance matrix of  $\check{\mathbf{W}}_k(a, b)$  is given by

$$\begin{aligned} \mathbf{\Omega}_k(a, b) &\equiv E\{\check{\mathbf{W}}_k(a, b)\check{\mathbf{W}}_k^H(a, b)\} \\ &\approx \frac{|a|}{N\Delta} \sum_{q=0}^{(N/2)-1} \left[ \Psi_k^+ \left( |a| \frac{q}{N\Delta} \right) \right]^2 \begin{cases} \mathbf{\Upsilon} \left( \frac{q}{N\Delta} \right), & a > 0; \\ \mathbf{\Upsilon} \left( -\frac{q}{N\Delta} \right), & a < 0. \end{cases} \end{aligned} \quad (4.20)$$

*Proof.* This is found in Appendix A.  $\square$

We now interpret equation (4.20) from Theorem 4.4.1. In Section 4.3.3 we define the reference frequency  $f_0^{(k)}$  for the  $k$ th analytic Morse wavelet to be the frequency that maximises  $f [\Psi_k^+(f)]^2$ . We can therefore say that  $|a| [\Psi_k^+(|a|f)]^2$  attains its maximum at  $f = f_0^{(k)}/|a|$ . Provided  $\mathbf{\Upsilon}(f)$  is sufficiently smooth, and the frequency domain support of  $[\Psi_k^+(|a|f)]^2$  sufficiently limited, then (4.20) becomes

$$\mathbf{\Omega}_k(a, b) \approx \mathbf{\Upsilon} \left( \frac{f_0^{(k)}}{a} \right) \frac{|a|}{N\Delta} \sum_{q=0}^{(N/2)-1} \left[ \Psi_k^+ \left( |a| \frac{q}{N\Delta} \right) \right]^2. \quad (4.21)$$

From (4.2) and using Rayleigh's theorem [5, p. 112] we know

$$1 = \int_0^\infty [\Psi_k^+(f)]^2 df = |a| \int_0^\infty [\Psi_k^+(|a|f)]^2 df,$$

hence for large  $N$

$$\frac{|a|}{N\Delta} \sum_{q=0}^{(N/2)-1} \left[ \Psi_k^+ \left( |a| \frac{q}{N\Delta} \right) \right]^2 \approx 1 \quad (4.22)$$

since for  $|a| > a_{\min}$  we know from (4.13) that  $\Psi_k^+(|a|f) = 0$  for  $f \geq f_N$ . Combining (4.21) and (4.22), we obtain

$$\mathbf{\Omega}_k(a, b) \approx \mathbf{\Upsilon} \left( \frac{f_0^{(k)}}{a} \right).$$

Instead of a single Morse wavelet, consider now the multiple orthogonal wavelets average

$$\hat{\Omega}(a, b) \equiv (1/K) \sum_{k=0}^{K-1} \check{\mathbf{W}}_k(a, b) \check{\mathbf{W}}_k^H(a, b).$$

Referring back to Section 4.3.3 and from [71] we can say the function

$$|a| \sum_{k=0}^{K-1} [\Psi_k^+(|a|f)]^2$$

peaks at  $f_0/a$  where  $f_0 > 0$  is the multiple wavelet reference frequency defined in (4.14). If  $\Upsilon(f)$  is sufficiently smooth and the frequency domain support of  $\sum_{k=0}^{K-1} [\Psi_k^+(|a|f)]^2$  sufficiently limited then we have the result

$$E\{\hat{\Omega}(a, b)\} \approx \Upsilon\left(\frac{f_0}{a}\right). \quad (4.23)$$

#### 4.4.2 Orthogonality and Uncorrelatedness

It has been said that the use of an orthogonal set of multiple wavelets is to provide us with a set of approximately uncorrelated estimators. To gain a deeper understanding of this concept we consider  $E\{\check{\mathbf{W}}_l(a, b) \check{\mathbf{W}}_m^H(a, b)\}$  for  $l \neq m$ ,  $l, m = 0, \dots, K - 1$ . We use the results in the proof of part (ii) of Theorem 4.4.1 to infer that  $E\{\check{\mathbf{W}}_l(a, b) \check{\mathbf{W}}_m^H(a, b)\}$  is equal to

$$\frac{|a|}{N\Delta} \sum_{q=0}^{(N/2)-1} \Psi_l^+\left(|a|\frac{q}{N\Delta}\right) \Psi_m^+\left(|a|\frac{q}{N\Delta}\right) \begin{cases} \Upsilon\left(\frac{q}{N\Delta}\right), & a > 0; \\ \Upsilon\left(-\frac{q}{N\Delta}\right), & a < 0. \end{cases} \quad (4.24)$$

The orthogonality of a pair of  $L^2(\mathbb{R})$  wavelets  $\psi_l(t)$  and  $\psi_m(t)$   $l \neq m$  is defined as  $\int_{-\infty}^{\infty} \psi_l(t) \psi_m^*(t) dt = 0$ . Through a change of variables it can be shown that for  $|a| > 0$  the functions  $|a|^{-1/2} \psi_l(t/a)$  and  $|a|^{-1/2} \psi_m(t/a)$   $l \neq m$  are also orthogonal. With use of the power theorem [5, p. 113]  $|a| \int_{-\infty}^{\infty} \Psi_l^+(f) \Psi_m^{+*}(f) df =$

0, and thus for large  $N$

$$\frac{|a|}{N\Delta} \sum_{q=0}^{(N/2)-1} \Psi_l^+ \left( |a| \frac{q}{N\Delta} \right) \Psi_m^+ \left( |a| \frac{q}{N\Delta} \right) \approx 0 \quad l \neq m.$$

We can now conclude that for a multivariate process whose augmented SDF matrix  $\Upsilon(f)$  is sufficiently smooth on the support of  $\Psi_l^+(f)\Psi_m^{+*}(f)$ , then (4.24) is equal to zero and the vectors  $\{\check{\mathbf{W}}_k(a, b), k = 0, \dots, K - 1\}$  will be approximately asymptotically uncorrelated. In the white noise case (flat spectra) they are exactly asymptotically uncorrelated.

Uncorrelated Gaussian variables of the same distribution implies IID, we therefore say that for processes with smooth spectra,  $\{\check{\mathbf{W}}_k(a, b), k = 0, \dots, K - 1\}$  will be a set of approximately IID zero-mean complex-valued Gaussian vectors.

## 4.5 Key Results and Distributional Properties of WCOH

We summarise the key results of this chapter. For a set of  $K$  orthogonal Morse wavelets  $\{\psi_k^+, k = 0, \dots, K - 1\}$  and  $N$  point realisation

$$\{[Z_{1,0}, \dots, Z_{p,0}]^T, \dots, [Z_{1,N-1}, \dots, Z_{p,N-1}]^T\}$$

of a  $p$ -dimension zero-mean complex-valued (proper or improper) Gaussian SOS process  $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$  with bounded and continuous spectra:

- (i) The vectors  $\{\check{\mathbf{W}}_k(a, b), k = 0, \dots, K - 1\}$  are zero-mean complex-valued Gaussian  $2p$ -dimensional vectors. For processes with smooth spectra they are approximately IID.
- (ii)  $\check{\mathbf{W}}_k(a, b), k = 0, \dots, K - 1$  is asymptotically proper.
- (iii) Provided  $\Upsilon(f)$  is smooth over the support of  $[\Psi_k^+(|a|f)]^2$ , the covariance  $\Omega_k(a, b)$  of  $\check{\mathbf{W}}_k(a, b)$  is approximately equal to  $\Upsilon \left( \frac{f_0^{(k)}}{a} \right)$ .

(iv) Provided  $\Upsilon(f)$  is smooth over the support of  $\sum_{k=0}^{K-1} [\Psi_k^+(|a|f)]^2$ , the average

$$\hat{\Omega}(a, b) \equiv (1/K) \sum_{k=0}^{K-1} \check{\mathbf{W}}_k(a, b) \check{\mathbf{W}}_k^H(a, b)$$

has expected value approximately equal to  $\Upsilon(f_0/a)$ .

From (i) and (ii), and with the condition in (iii) holding, with reference to the work of Goodman [35] outlined in Theorem 2.4.1 we can conclude  $\Omega_k(a, b)$  is, for large  $N$ , approximately  $2p$ -dimensional complex central Wishart distributed with a single complex degree of freedom, and mean  $\Upsilon(f_0^{(k)}/a)$ , i.e.

$$\Omega_k(a, b) \stackrel{d}{=} \mathcal{W}_{2p}^C \left\{ 1, \Upsilon(f_0^{(k)}/a) \right\}.$$

By considering (iv), provided the smoothness condition holds we can further say that  $K\hat{\Omega}(a, b)$  is, for large  $N$ , approximately  $2p$ -dimensional complex central Wishart distributed with  $K$  complex degrees of freedom and mean  $K\Upsilon(f_0/a)$ , i.e.

$$\hat{\Omega}(a, b) \stackrel{d}{=} (1/K) \mathcal{W}_{2p}^C \left\{ K, \Upsilon(f_0/a) \right\}. \quad (4.25)$$

If  $K < 2p$ , the matrix  $\hat{\Omega}(a, b)$  will be singular. If we wish to obtain just a single ordinary or conjugate coherence we could delete the other rows and columns of  $\hat{\Omega}(a, b)$  e.g. to obtain the wavelet coherence estimator  $\hat{\gamma}_{12}^2$  of  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  we only need retain the first and second columns of  $\hat{\Omega}(a, b)$  and can delete all others. In this case it would be sufficient to have  $K \geq 2$  for non-singularity. We further conclude from [35], outlined in Section 2.8 and Theorem 2.8.2, that all WCOH estimators of types (4.17), (4.18) and (4.19) will be Goodman distributed. To parameterise the Goodman distribution we need the true value parameter. If we look at the example of the conjugate WCOH estimator for  $\{Z_{1,t}\}$  and  $\{Z_{1,t}^*\}$ , deleting the appropriate rows and columns of  $\Upsilon(f)$ , the

spectral matrix for  $\{[Z_{1,t}, Z_{1,t}^*]^T\}$  is

$$\mathbf{\Upsilon}(f) = \begin{bmatrix} S_{11}(f) & R_{11}(f) \\ R_{11}^*(f) & S_{11}(-f) \end{bmatrix}.$$

(4.25) requires us to evaluate this at  $f = f_0/a$ , giving

$$\mathbf{\Upsilon}(f_0/a) = \begin{bmatrix} S_{11}(f_0/a) & R_{11}(f_0/a) \\ R_{11}^*(f_0/a) & S_{11}(-f_0/a) \end{bmatrix}.$$

So the estimator  $\hat{\gamma}_{11*}^2(a, b)$  has the Goodman distribution with  $K$  degrees of freedom and true parameter value

$$\gamma_{11*}^2(a, b) = \frac{|R_{11}(f_0/a)|^2}{S_{11}(f_0/a)S_{11}(-f_0/a)} = \gamma_{11*}^2(f_0/a).$$

The PDF for  $0 \leq \hat{\gamma}_{11*}^2(a, b) \leq 1$ , is

$$g_{\hat{\gamma}_{11*}^2}(x; K, \gamma_{11*}^2) = (K-1)(1-\gamma_{11*}^2)^K(1-x)^{K-2} {}_2F_1(K, K; 1; \gamma_{11*}^2 x) \quad (4.26)$$

where  $(a, b)$  has been suppressed.

## 4.6 Simulations and Results

To demonstrate the theory of WCOH estimators as derived in this chapter, we look at one form of the possible coherences derived in Section 4.4, namely that between a process and its own complex conjugate. If  $\{Z_t, t \in \mathbb{Z}\}$  is a complex process then we concentrate on the bivariate process  $\{\check{\mathbf{Z}}_t = [Z_t, Z_t^*]^T\}$ .

### 4.6.1 The Complex Process

We use the complex SOS autoregressive process  $\{Z_t\}$  of order 1 (CAR(1) process) which is of the form

$$Z_t = \phi Z_{t-1} + \varsigma_t$$

where  $\phi$  is a complex-valued parameter such that the value of  $1/\phi$  lies outside the unit circle, and  $\{\varsigma_t\}$  is doubly white noise [76]. A zero-mean process  $\{\varsigma_t\}$  is said to be doubly white if the autocovariance sequence is of the form  $s_{\varsigma,\tau} = \sigma_{\varsigma}^2 \delta_{\tau,0}$  where  $\sigma_{\varsigma}^2 = E\{|\varsigma_t|^2\}$ , and the relation sequence is of the form  $r_{\varsigma,\tau} = r_{\varsigma} \delta_{\tau,0}$  where  $r_{\varsigma} = E\{\varsigma_t^2\}$ . The specific doubly white noise sequence used here is as used in [87]

$$\varsigma_t = \sqrt{0.1} \epsilon_t + \frac{0.3}{\sqrt{0.1}} e^{i\pi/6} \epsilon_t^*$$

with proper Gaussian white noise  $\epsilon_t = \epsilon_{X,t} + i\epsilon_{Y,t}$  where  $\{\epsilon_{X,t}\}$  and  $\{\epsilon_{Y,t}\}$  are zero-mean uncorrelated real-valued Gaussian white noise sequences. The true value of the conjugate coherence for the processes  $\{Z_t\}$  and  $\{Z_t^*\}$  is  $\gamma_*^2(f) = 0.36$  for all  $f \in [-1/2, 1/2)$ .

### 4.6.2 Specifics of the Morse Wavelet Parameters

When choosing the Morse wavelet parameters  $\beta$  and  $\gamma$  there are two issues to consider. Firstly we require that the frequency domain support of  $\sum_{k=0}^{K-1} [\Psi_k^+(|a|f)]^2$  be suitably narrow as to validate the approximation in equation (4.20) of Theorem 4.4.1. Secondly, to have a valid range of scales  $a_{\min}(K) < |a| < a_{\max}(K)$  we require that  $a_{\min}(K) < a_{\max}(K)$ . After some analysis in maximising energy concentrations, wavelet parameters of  $\beta = 40$ ,  $\gamma = 15$  and  $K = 10$  were used — see Figure 4.2. For this choice  $t_{\min}(10) = -200$ ,  $t_{\max}(10) = 200$ ,  $f'(10) = 0.2$ ,  $a_{\min}(10) = 0.4$  and  $a_{\max}(10) = 0.64, 1.28$  and  $2.56$  for  $N = 256, 512$  and  $1024$ , respectively. (The choice  $N = 128$  gives

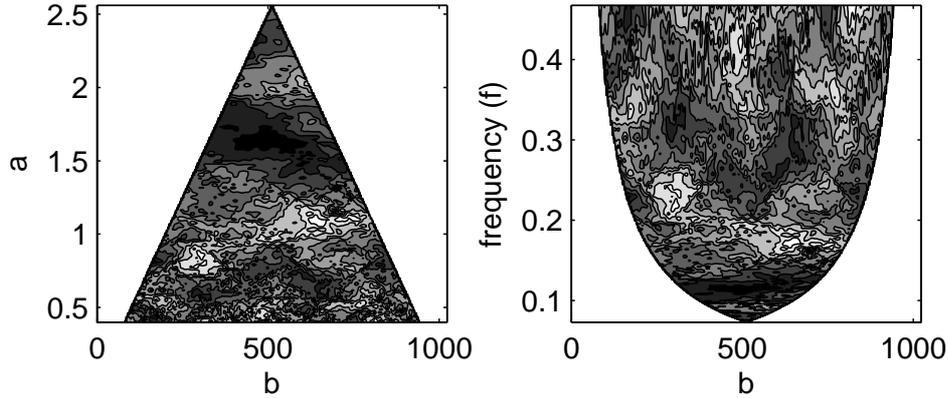


Figure 4.3: A Morse wavelet coherence estimate for a SOS CAR(1) process (see Section 4.6.1), with parameters  $K = 10$ ,  $\beta = 40$  and  $\gamma = 15$ , plotted as a function of time-scale (left) and time-frequency (right). Only the valid regions are shown. Areas of dark shading indicate high values of coherence.

$a_{\max}(10) < a_{\min}(10)$  and so is unsuitable.) Additionally,  $f_0 = 0.18$ . We have stated that for non-singularity of the matrix  $\hat{\Omega}(a, b)$  it is required that  $K \geq 2$ . The choice  $K = 10$ , a typical number for complex degrees of freedom for an estimator, is therefore appropriate.

Figure 4.3 shows a wavelet coherence estimate for a single 1024 point sample of the complex autoregressive process defined in Section 4.6.1 using the wavelet parameters specified above. Only the valid region of the time-scale and time-frequency plane is considered.

### 4.6.3 Results

For our simulations, the true value of conjugate coherence is  $\gamma_*^2(f) = 0.36$  for all  $|f| \leq 1/2$ , thus the true value parameter  $\gamma_*^2(a, b)$  used for the parameterisation of the Goodman distribution in equation (4.26) takes the value of 0.36 for all  $(a, b)$  pairs in the time-scale domain depicted in Figure 4.3. The theory developed in this chapter therefore predicts that the distribution of  $\hat{\gamma}_*^2(a, b)$  is invariant to  $(a, b)$  for the autoregressive model outlined above.

The distribution of  $\hat{\gamma}_*^2(a, b)$  is examined in Figure 4.4 for  $N = 256$  and

$N = 512$  at four randomly chosen points on the time-scale plane with the use of quantile-quantile (Q-Q) plots. Each curve is produced by ordering 100 independently-simulated estimates  $\hat{\gamma}_*^2(a, b)$  into ascending order of size. These ordered estimates are plotted against the theoretical quantiles of the Goodman distribution. The probability of a value less than the  $j$ th ordered estimate is approximately  $p_j = j/101$ ,  $j = 1, \dots, 100$ . The corresponding theoretical quantile of the Goodman distribution is the value  $y_j$  that satisfies

$$p_j = \int_0^{y_j} g_{\hat{\gamma}_*^2}(x; K, \gamma_*^2) dx = G(y_j)$$

where  $G(y_j)$  is the cumulative distribution function of the Goodman distribution which can be calculated using the algorithm in [54]. The values  $y_1, \dots, y_{100}$  are plotted on the y-axis against the ordered estimates. A good fit to the Goodman distribution corresponds to a curve that is close to the dashed line. For both values of  $N$  and for all four  $(a, b)$  pairs there is a good fit.

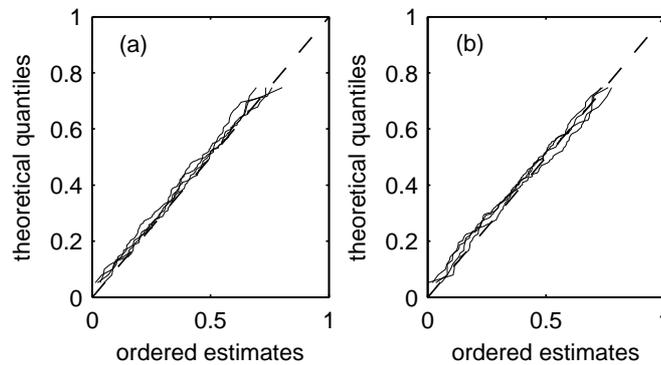


Figure 4.4: Q-Q plots for 100 independent samples of  $\hat{\gamma}_*^2(a, b)$  for (a)  $N = 256$  and (b)  $N = 512$ . The four curves on each plot correspond to four different  $(a, b)$  pairs on the scale-time plane.  $K = 10$ ,  $\beta = 40$  and  $\gamma = 15$ .

Figure 4.5 are Q-Q plots for two different values of  $K = 5$  and  $K = 10$ . Once again curves for four different  $(a, b)$  pairs are plotted for a process portion of length  $N = 1024$ . We can see that  $K = 10$  gives a better fit to the Goodman distribution than  $K = 5$ . By referring to Section 4.4.2, we have said that the

wavelet vectors  $\{\tilde{\mathbf{W}}_k(a, b), k = 0, \dots, K - 1\}$  are only exactly uncorrelated for processes with flat spectra. For the process used in our simulations the spectra are far from flat so averaging over a greater number of wavelets may help cancel out errors that occur from any correlation between the vectors  $\{\tilde{\mathbf{W}}_k(a, b), k = 0, \dots, K - 1\}$ .

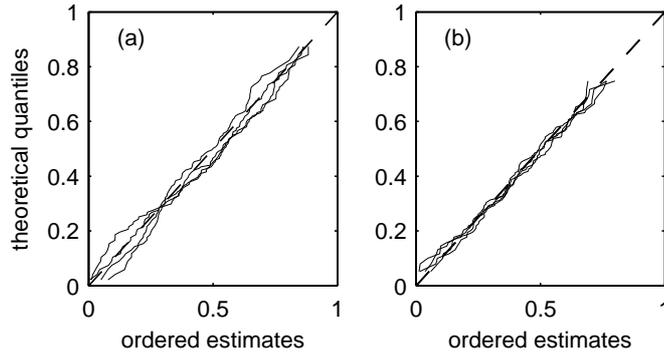


Figure 4.5: Q-Q plots for 100 independent samples of  $\hat{\gamma}_*^2(a, b)$  for (a)  $K = 5$  and (b)  $K = 10$ . The four curves on each plot correspond to four different  $(a, b)$  pairs on the scale-time plane.  $N = 1024$ ,  $\beta = 40$  and  $\gamma = 15$ .

Figure 4.6 shows, for each  $K = 1$  to 10, the theoretical mean value of the WCOH estimators (triangles) and the sample mean of  $\hat{\gamma}_*^2(a, b)$  over 1000 simulations (crosses) at four chosen points  $(a, b)$ . As can be seen, the four crosses are almost indistinguishable for all values of  $K$  and they match the theoretical value very closely. We see for  $K = 1$  that the mean of  $\hat{\gamma}_*^2(a, b)$  is exactly unity as expected. For  $K \geq 2$ , the theoretical mean of the coherence estimator is given by [103]

$$E\{\hat{\gamma}_*^2(a, b)\} = \frac{1}{K} + \gamma_*^2(a, b) \left[ \frac{K-1}{K+1} \right] {}_2F_1(1, 1; K+2; \gamma_*^2(a, b)).$$

A key result of Section 4.4 is the specification of  $E\{\hat{\mathbf{\Omega}}(a, b)\} \approx \mathbf{\Upsilon}(f_0/a)$  which allows full parameterisation of the Wishart distribution in (4.25) and hence the Goodman distribution in (4.26). We examine this in Figure 4.7 for

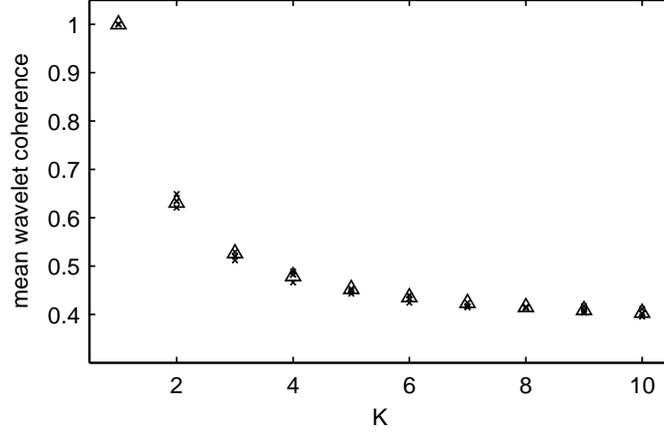


Figure 4.6: The theoretical mean values of the coherence estimator are shown by the triangles, and the sample mean of  $\hat{\gamma}_*^2(a, b)$  over 1000 simulations, at four randomly chosen points  $(a, b)$  on the scale-time plane, are shown by crosses.  $N = 1024$ .

the parameter choices of  $\beta = 40$ ,  $\gamma = 15$  and  $K = 10$ , giving  $f_0 = 0.18$ . The four thin lines are the four components of

$$\Upsilon(f) = \begin{bmatrix} S(f) & R(f) \\ R^*(f) & S(-f) \end{bmatrix}$$

for  $-1/2 \leq f < 1/2$ . The upper right and lower left plots show the real and imaginary components of the relation spectra separately. The means of  $\hat{\Omega}(a, b)$  are plotted as crosses (over 1000 simulations), for 100 different positive and negative scale values  $a$  in the range  $a_{\min} < |a| < a_{\max}$ , plotted at positions  $f = f_0/a$  for a single randomly chosen time  $b$ . In general, there is a very good agreement, except for the high positive frequency end of  $S_{11}(f)$  (Figure 4.7, top left), and its reflection, (Figure 4.7, bottom right), where there is a slight discrepancy.

Let us consider  $\sum_{k=0}^{K-1} [\Psi_k^+(f)]^2$ , shown in Figure 4.8(a). We can see that  $\sum_{k=0}^{K-1} [\Psi_k^+(f)]^2 \not\approx 0$  for  $f \in [0.14, 0.2]$ . For illustration we look at the extremes of the range of  $a$  values and concentrate on  $S_{11}(f)$ , (Figure 4.7, top left). For  $a = a_{\min} = 0.4$ , the matrix  $\Upsilon(f)$  will be smoothed over  $[0.14/|a|, 0.2/|a|] =$

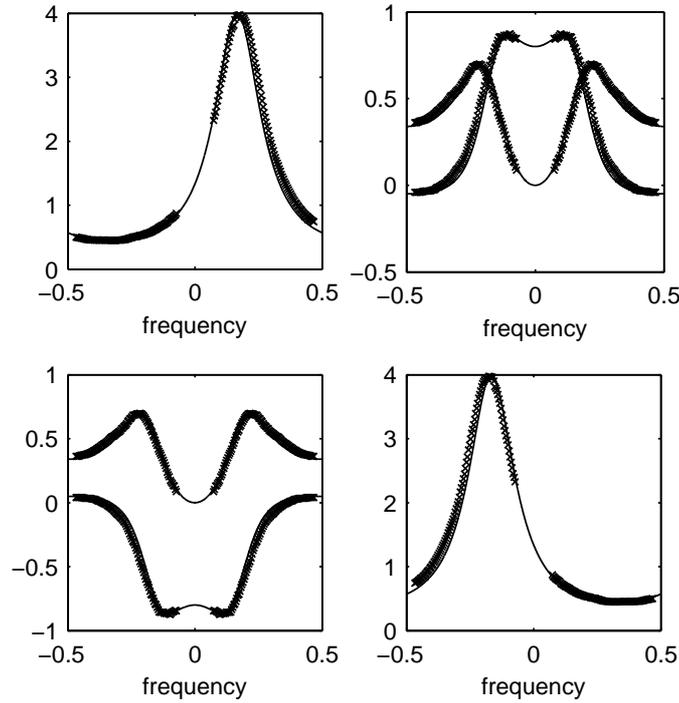


Figure 4.7: The four theoretical components of  $\boldsymbol{\Upsilon}(f)$ . In each plot the thin line marks the true value. The upper right and lower left plots show both real and imaginary parts of  $R_{11}(f)$  and  $R_{11}^*(f)$ , respectively. The crosses are the means of  $\hat{\boldsymbol{\Omega}}(a, b)$ , over 1000 simulations, for 100 different positive and negative scale values in the range  $a_{\min} < |a| < a_{\max}$ , plotted at positions  $f = f_0/a$  for a randomly-chosen time  $b$ ,  $\beta = 40$ ,  $\gamma = 15$  and  $N = 1024$ .

$[0.35, 0.5]$ . Figure 4.7 shows  $S_{11}(f)$  to vary quite rapidly over this interval. As a result, when the mean of  $\hat{\boldsymbol{\Omega}}_{11}(a, b)$  is plotted at  $f = f_0/a_{\min} = 0.18/0.4 = 0.45$  there is a slight error. In other words, for this  $a$  value, there is significant variation in  $S_{11}(f)$  over the support of  $\sum_{k=0}^{K-1} [\Psi_k^+(|a|f)]^2$ . When  $N = 1024$ ,  $a = a_{\max} = 2.56$  then  $[0.14/|a|, 0.2/|a|] = [0.055, 0.078]$ , a narrow interval over which there is little variation in  $S_{11}(f)$ . The mean of  $\hat{\boldsymbol{\Omega}}_{11}(a, b)$  is plotted at  $f = f_0/a = 0.18/2.56 = 0.07$ , and is in good agreement with  $S_{11}(0.07)$ .

To further demonstrate the requirement for a tight frequency support of  $\sum_{k=0}^{K-1} [\Psi_k^+(f)]^2$  in providing a good approximation in (4.23), we compare Fig-

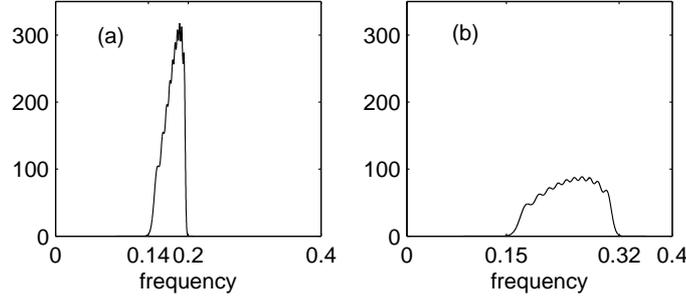


Figure 4.8:  $\sum_{k=0}^{K-1} [\Psi_k^+(f)]^2$  for Morse parameters (a)  $\beta = 40$ ,  $\gamma = 15$  and (b)  $\beta = 30$ ,  $\gamma = 5$ .

ures 4.7 and 4.9. Figure 4.9 is constructed from Morse wavelets with parameters  $\beta = 30$  and  $\gamma = 5$ . Figure 4.8(b) shows that for these parameters  $\sum_{k=0}^{K-1} [\Psi_k^+(|a|f)]^2$  has a significantly wider support than for the parameters  $\beta = 40$  and  $\gamma = 15$ . It is clear to see from Figure 4.9 that the approximation in (4.23) in this case is less valid.

## 4.7 Concluding Remarks

The long-standing problem of finding the statistical distribution of a WCOH estimator applied to SOS processes has been addressed, and the Goodman distribution has been shown to be applicable for the case of estimation via multiple Morse wavelets. Using simulations of a complex autoregressive process the theoretical results have been illustrated.

In order to make the theory presented here accessible for practical purposes attention has been paid to the discrete-time implementation. If for Theorem 4.4.1 we work instead with continuous-time processes  $\{[Z_1(t), Z_2(t)]^T, t \in \mathbb{R}\}$  and the continuous-time CWT in (3.3), results are exact:  $\check{\mathbf{W}}_k(a, b)$  is a zero-mean proper complex-valued vector and the covariance is precisely

$$E \{ \check{\mathbf{W}}_k(a, b) \check{\mathbf{W}}_k^H(a, b) \} = |a| \int_{-\infty}^{\infty} [\Psi_k^+(af)]^2 \Upsilon(f) df.$$

The proof is given in Appendix B.

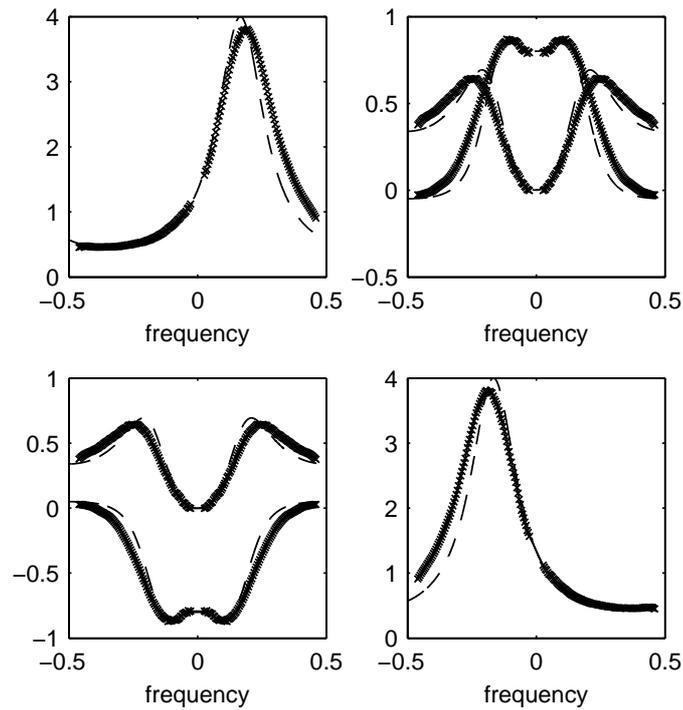


Figure 4.9: As for Figure 4.7 but for Morse parameters  $\beta = 30$ ,  $\gamma = 5$  and  $N = 1024$ .

By knowing the distribution of the WCOH estimator it is possible to test a hypothesis on the nature of correlations between a pair of random processes as outlined in Section 3.4. The reader is referred to Section 2.8 where a discussion is given on how the Goodman distribution can be used for such analysis. In Section 5.5 an example will be provided using actual data.

Until now the method of choice in the literature for smoothing WCOH estimators has been with the use of a single Morlet wavelet and smoothing in time and scale, or just in time. It is the second of these approaches that we now focus on.

## Chapter 5

# Temporally Smoothed Wavelet Coherence

We now examine the statistical properties of the WCOH estimator for a pair of signals  $Z_1(t)$  and  $Z_2(t)$  using a single Morlet wavelet and temporal smoothing over a small localised time interval centred at the translation of interest  $b$ . This smoothing method has been used in [42], [53] and [101]. Consider the wavelet coherence estimator

$$\hat{\gamma}^2(a, b) = \frac{|\mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)\}|^2}{\mathcal{S}\{|W(a, b; Z_1, \psi)|^2\}\mathcal{S}\{|W(a, b; Z_2, \psi)|^2\}}, \quad (5.1)$$

with this approach to smoothing the numerator of (5.1) is

$$\mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)\} = \frac{1}{\delta} \int_{b-\delta/2}^{b+\delta/2} W(a, \tau; Z_1, \psi)W^*(a, \tau; Z_2, \psi)d\tau \quad (5.2)$$

and the denominator terms are

$$\mathcal{S}\{|W(a, b; Z_1, \psi)|^2\} = \frac{1}{\delta} \int_{b-\delta/2}^{b+\delta/2} |W(a, \tau; Z_1, \psi)|^2 d\tau \quad (5.3)$$

$$\mathcal{S}\{|W(a, b; Z_2, \psi)|^2\} = \frac{1}{\delta} \int_{b-\delta/2}^{b+\delta/2} |W(a, \tau; Z_2, \psi)|^2 d\tau. \quad (5.4)$$

Lachaux et al [53] recognised that when using a Morlet wavelet the terms (5.2), (5.3) and (5.4) can be interpreted as continuous-time WOSA spectral estimators of the spectral density and cross-spectral density functions of  $Z_1(t)$  and  $Z_2(t)$  — see Section 2.7.2.

The first step in the approach for deriving distributional results for the temporally smoothed wavelet coherence estimator (TWCOH) will be to use this continuous-time interpretation to formulate a discrete-time equivalent. The WOSA estimator will then be cast into a multitaper formulation that allows highly accurate degrees of freedom to be derived. This will remove approximations made in Bortel and Sovka [4] in their derivation of ordinary coherence using segment averaging. Analytical results will then be derived for the distribution of the TWCOH estimator for complex-valued, jointly stationary and Gaussian time series. The work in this chapter can also be found in Cohen and Walden [15].

## 5.1 Smoothing with the Morlet Wavelet

The Morlet wavelet was first discussed in Section 3.3. A function of the general form

$$\psi(t) = Ce^{-t^2/(2d)}e^{i\omega_0 t}$$

defined on the real line with  $d > 0$  is a complex sinusoid localised/tapered by a Gaussian envelope [9, p. 139]. In order that this function can be treated like a wavelet it must integrate to zero (or approximately zero). With a choice of  $\omega_0 = 2\pi$ , the integral of the function is less than  $10^{-4.4}$  provided  $0.8 \leq d \leq 10$ . A wavelet must also have unit  $L^2$  norm so additionally it is required that

$$1 = \int_{-\infty}^{\infty} |\psi(t)|^2 dt = C^2 \int_{-\infty}^{\infty} e^{-(t/d)^2} dt = C^2 d\pi^{1/2}$$

giving  $C = d^{-1/2}\pi^{-1/4}$ . So

$$\psi(t) = d^{-1/2}\pi^{-1/4}e^{i2\pi t}e^{-\frac{1}{2}\left(\frac{t}{d}\right)^2}. \quad (5.5)$$

The Fourier transform of  $\psi(t)$  (also of unit norm) is given as

$$\Psi(f) = \int_{-\infty}^{\infty} \psi(t) e^{-i2\pi ft} dt = \pi^{1/4} (2d)^{1/2} e^{-2[d\pi(f-1)]^2},$$

which, like the Morse wavelets, is real-valued.  $\Psi(f)$  is a Gaussian function centred around 1, i.e. it has a maximum value  $\Psi(1)$ . From (5.5) the scaled ( $|a| > 0$ ) and translated ( $b \in \mathbb{R}$ ) form of the Morlet wavelet is given as

$$\begin{aligned} \psi_{a,b}(t) &= \pi^{-1/4} (|a|d)^{-1/2} e^{-\frac{1}{2} \left(\frac{t-b}{ad}\right)^2} e^{i2\pi \left(\frac{t-b}{a}\right)} \\ &= \lambda(a, b, d; t) e^{i2\pi \left(\frac{t-b}{a}\right)} \end{aligned} \quad (5.6)$$

where  $\lambda(a, b, d; t) \equiv \pi^{-1/4} (|a|d)^{-1/2} e^{-\frac{1}{2} \left(\frac{t-b}{ad}\right)^2}$  is the Gaussian component of  $\psi_{a,b}(t)$ .  $\lambda(a, b, d; t)$  is also of unit  $L^2$  norm. Using (5.6), the CWT (3.3) can be expressed as

$$W(a, b; Z, \psi) = e^{i2\pi b/a} \int_{-\infty}^{\infty} Z(t) \lambda(a, b, d; t) e^{-i2\pi t/a} dt.$$

This simple form follows from the choice of wavelet and is key to formulating this approach to statistical analysis. We define

$$T_{Z;a,b}(t) = Z(t) \lambda(a, b, d; t).$$

This is the product of the process  $Z(t)$  with a normalised Gaussian function centred at time  $b$ . We can therefore represent the CWT as  $W(a, b; Z, \psi) = e^{i2\pi b/a} \mathcal{T}_{Z;a,b}(1/a)$  where  $\mathcal{T}_{Z;a,b}(f)$  is the Fourier transform of the tapered process  $T_{Z;a,b}(t)$ , i.e.

$$\mathcal{T}_{Z;a,b}(f) = \int_{-\infty}^{\infty} Z(t) \lambda(a, b, d; t) e^{-i2\pi ft} dt, \quad (5.7)$$

which is evaluated at  $f = 1/a$ . Consequently, (5.2) becomes

$$\mathcal{S}\{W(a, b; Z_1, \psi) W^*(a, b; Z_2, \psi)\} = \frac{1}{\delta} \int_{b-\delta/2}^{b+\delta/2} \mathcal{T}_{Z_1;a,\tau}(1/a) \mathcal{T}_{Z_2;a,\tau}^*(1/a) d\tau \quad (5.8)$$

with analogous expressions for the denominator terms (5.3) and (5.4).

It was recognised in [53] that (5.8) is a WOSA-type estimator, however very few details were given. To understand why this connection can be made we need to consider the continuous-time formulation of the WOSA spectral estimator.

## 5.2 WOSA for Continuous-Time Processes

For a signal  $Z(t)$  we split a time interval  $[b - T/2, b + T/2]$  of length  $T$  centred on  $b$  into  $N_B$  overlapping intervals of equal measure. Let  $\theta$  denote the length of each subinterval. Each subinterval is offset from the previous one by  $\Lambda = (T - \theta)/(N_B - 1)$ . Successive subintervals will have  $100(1 - \Lambda/\theta)\%$  overlap. The centre of the  $j$ th interval is given as  $t_j = b - [(T - \theta)/2] + j\Lambda$ . We therefore express the  $j$ th segment of the signal as  $Z(t)I_{A_j}$  where  $A_j = [b - (T/2) + j\Lambda, b - (T/2) + \theta + j\Lambda]$ , with  $I_A(t)$  being the characteristic function for  $A \subset \mathbb{R}$  defined in (3.1).

Let  $h(t)$  be a continuous-time data taper of unit  $L^2$  norm.  $h(t)$  is designed to be compactly supported on the interval  $[-1/2, 1/2]$ . The scaled function  $h_\theta(t) = \theta^{-1/2}h(t/\theta)$  will be unit norm and compactly supported on the interval  $[-\theta/2, \theta/2]$ . We can now define the  $j$ th weighted Fourier transform

$$J_{Z;j}(f) = \int_{-\infty}^{\infty} Z(t)I_{A_j}h_\theta(t - t_j)e^{-i2\pi ft}dt.$$

This is the Fourier transform of a signal portion  $Z(t)I_{A_j}$  tapered by  $h_\theta(t)$ . As described in Section 2.7.2, a WOSA spectral estimator is formed from averaging a set of individual direct spectral estimators on overlapping segments. The WOSA cross-spectral estimator for the continuous-time processes  $Z_1(t)$  and  $Z_2(t)$  averaging over the  $N_B$  individual segment estimates is given as

$$\hat{S}_{12,b}^{(W)}(f) = (1/N_B) \sum_{j=0}^{N_B-1} J_{Z_1;j}(f)J_{Z_2;j}^*(f).$$

The index  $b$  indicates that this is the estimator for the signal portion  $[b - T/2, b + T/2]$ . We will just consider the CSDF estimators, the most general form. SDF estimators follow by letting both processes be the same.

If we now let the offset between the overlapping segments  $\Lambda \rightarrow 0$ , while letting the number of segments  $N_B \rightarrow \infty$  such that  $N_B \Lambda = (T - \theta)$ , then

$$\begin{aligned} \hat{S}_{12,b}^{(W)}(f) &\rightarrow \lim_{\Lambda \rightarrow 0} \frac{\Lambda}{T - \theta} \sum_{j=0}^{N_B-1} J_{Z_1;j}(f) J_{Z_2;j}^*(f) \\ &= \frac{1}{T - \theta} \int_{b - \frac{T-\theta}{2}}^{b + \frac{T-\theta}{2}} J_{Z_1;t'}(f) J_{Z_2;t'}^*(f) dt' \end{aligned} \quad (5.9)$$

where

$$J_{Z;t'}(f) = \int_{-\infty}^{\infty} Z(t) I_{[t' - \frac{\theta}{2}, t' + \frac{\theta}{2}]}(t) h_{\theta}(t - t') e^{-i2\pi f t} dt. \quad (5.10)$$

This limiting form of WOSA estimator provides effectively a 100% overlap. By setting  $\delta = T - \theta$ , (5.9) becomes

$$\hat{S}_{12,b}^{(W)}(f) = \frac{1}{\delta} \int_{b - \frac{\delta}{2}}^{b + \frac{\delta}{2}} J_{Z_1;t'}(f) J_{Z_2;t'}^*(f) dt'$$

and the similarity with (5.8) becomes obvious. To be more precise we make the approximation

$$e^{-\frac{1}{2} \left( \frac{t-\tau}{ad} \right)^2} \approx e^{-\frac{1}{2} \left( \frac{t-\tau}{ad} \right)^2} I_{[\tau-3|a|d, \tau+3|a|d]}, \quad (5.11)$$

which uses the fact that even though a Gaussian function of standard form  $e^{-t^2/(2\sigma^2)}$  lacks compact support, it can be approximated as existing on the interval  $[-3\sigma, 3\sigma]$ . Using this approximation, from (5.7)  $\mathcal{T}_{Z;a,\tau}(1/a)$  can be expressed as

$$\int_{-\infty}^{\infty} Z(t) I_{[\tau-3|a|d, \tau+3|a|d]} \lambda(a, \tau, d; t) e^{-i2\pi t/a} dt, \quad (5.12)$$

which is in the form of (5.10) evaluated at frequency  $f = 1/a$ , where  $h(t)$  is a unit  $L^2$  norm Gaussian-like function,  $h(t) \equiv 6^{1/2} \pi^{-1/4} e^{-(1/2)(6t)^2}$  and  $\theta \equiv 6|a|d$ . By combining (5.8) and (5.12) we make the following conclusion: the

term  $\mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)\}$  is equivalent to a limiting 100%-overlap continuous-time WOSA cross-spectral estimator for the portions  $Z_1(t)I_{[\alpha, \beta]}$  and  $Z_2(t)I_{[\alpha, \beta]}$  evaluated at the frequency  $1/a$ , with  $\alpha = b - [(\delta + \theta)/2] = b - (\delta/2) - 3|a|d$  and  $\beta = b + [(\delta + \theta)/2] = b + (\delta/2) + 3|a|d$ .

In conclusion, by separating the Morlet wavelet into its components we can see that the Gaussian part acts as a data taper, and the complex exponential acts with the integration operation to perform a Fourier transform on the tapered process. This is therefore equivalent to a direct spectral estimator on a localised portion of the signal. By averaging these direct spectral estimators across a time interval centred at  $b$ , the smoothed Morlet wavelet spectrum is equivalent to a WOSA estimator on a localised region of the signal, with maximal overlapping.

We are left with the problem of choosing  $\delta$ , the width of the smoothing/integration window. When performing temporal smoothing on a wavelet spectrum using the Morlet wavelet, not only does the period of the sinusoid increase/decrease as  $|a|$  increases/decreases but the width of the Gaussian-type component that acts as a data taper also increases/decreases with  $|a|$ . Consequently Lachaux et al [53] noticed that  $\delta$  must become a function of scale, so that for translation  $b$  the smoothing occurs over the interval  $[b - (\delta(|a|))/2, b + (\delta(|a|))/2]$ .

Given a specific choice of the parameter  $d$  the Morlet wavelet exhibits  $\omega_d$  sinusoidal periods within the Gaussian envelope. We can say that  $\omega_d \approx 6|a|d/|a| = 6d$  and hence is independent of  $|a|$ . The number of sinusoid oscillations that could be contained in an interval of length  $\delta$  is  $\kappa_\delta = \delta/|a|$ . It is the value  $\kappa_\delta$  that we wish to keep constant for all values  $|a|$  and as such we let  $\delta = |a|\kappa$ , for a constant  $\kappa > 0$ , giving

$$\mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)\} = \frac{1}{|a|\kappa} \int_{b - \frac{|a|\kappa}{2}}^{b + \frac{|a|\kappa}{2}} \mathcal{T}_{Z_1; a, \tau}(1/a) \mathcal{T}_{Z_2; a, \tau}^*(1/a) d\tau.$$

While continuous-time provides a nice framework in which to explain and develop the theory, a discrete-time formulation is required for application pur-

poses, which is now discussed. It will be shown that by casting the problem into a discrete-time framework, asymptotic statistical properties of the TWCOH estimator can be derived with distributions and exact degrees of freedom.

### 5.3 Discrete-Time Method

We recall the discrete-time CWT discussed in Section 3.1. Let  $Z_n \equiv Z(n\Delta)$ ,  $n = 0, \dots, N-1$ , be an observed portion of a complex-valued SOS process  $\{Z_t\}$  sampled at an interval  $\Delta$  small enough to avoid aliasing, we can express the CWT for  $a = a_0\Delta$  and  $b = b_0\Delta$  as

$$W(a, b; Z, \psi) \approx \frac{|a_0|^{1/2}}{N\Delta^{1/2}} \sum_{q=0}^{N-1} \mathcal{Z}_q \Psi_q^*(a_0) e^{i2\pi b_0/N}$$

where  $\mathcal{Z}_q = \Delta \sum_{n=0}^{N-1} Z_n e^{-i2\pi nq/N}$ , for  $q = 0, \dots, N-1$ , is the discrete Fourier transform of  $\{Z_t\}$ , and

$$\Psi_q^*(a_0) = \begin{cases} \Psi^*(a_0 \frac{q}{N}), & q = 0, \dots, (N/2) - 1; \\ \Psi^*(a_0 \frac{q-N}{N}), & q = (N/2), \dots, N-1. \end{cases} \quad (5.13)$$

Hence,

$$W(a, b; Z, \psi) \approx \frac{|a|^{1/2}}{N} \sum_{n=0}^{N-1} Z_n \sum_{q=0}^{N-1} \Psi_q^*(a_0) e^{i2\pi q(b_0-n)/N}.$$

By making use of the form of (5.13) this may be rewritten as

$$\begin{aligned} W(a, b; Z, \psi) &\approx \frac{|a|^{1/2}}{N} \sum_{n=0}^{N-1} Z_n \sum_{q=-(N/2)}^{(N/2)-1} \Psi^*(a_0 \frac{q}{N}) e^{i2\pi q(b_0-n)/N} \\ &\approx \Delta \sum_{n=0}^{N-1} Z_n |a|^{1/2} \int_{-f_N}^{f_N} \Psi^*(af) e^{i2\pi f(b-n\Delta)} df. \end{aligned} \quad (5.14)$$

Provided

$$|a|f_N > f'(d) \quad \text{where} \quad f'(d) \equiv \max\{|f_1|, |f_2|\}, \quad (5.15)$$

and the effective support of  $\Psi(f)$  is  $[f_1, f_2]$ , then

$$\begin{aligned} |a|^{1/2} \int_{-f_N}^{f_N} \Psi^*(af) e^{i2\pi f(b-n\Delta)} df &= |a|^{1/2} \int_{-\infty}^{\infty} \Psi^*(af) e^{i2\pi f(b-n\Delta)} df \\ &= \frac{1}{|a|^{1/2}} \psi^*([n\Delta - b]/a) \\ &= \psi_{a,b}^*(n\Delta). \end{aligned} \quad (5.16)$$

From (5.14) and (5.16), we take the discretised form of the CWT for our analysis to be defined by  $W(a, b; Z, \psi) \equiv \Delta \sum_{n=0}^{N-1} Z_n \psi_{a,b}^*(n\Delta)$ . The form of  $\psi_{a,b}(\cdot)$  used is given by (5.6). So  $W(a, b; Z, \psi) = \Delta \sum_{n=0}^{N-1} Z_n \lambda_{a,b,d;n} e^{-i2\pi(\frac{n\Delta-b}{a})}$ , where  $\lambda_{a,b,d;n} \equiv \lambda(a, b, d; n\Delta)$  is a taper for which  $\Delta \sum_{n=-\infty}^{\infty} \lambda_{a,b,d;n}^2 \approx 1$ . Let  $T_{Z;a,b,n} = Z_n \lambda_{a,b,d;n}$ . Then,

$$W(a, b; Z, \psi) = e^{i2\pi b/a} \Delta \sum_{n=0}^{N-1} T_{Z;a,b,n} e^{-i2\pi n\Delta/a}. \quad (5.17)$$

Hence,  $W(a, b; Z, \psi) = e^{i2\pi b/a} \mathcal{T}_{Z;a,b}(1/a)$ , where  $\mathcal{T}_{Z;a,b}(f)$  is the Fourier transform of  $\{T_{Z;a,b,n}, n = 0, \dots, N-1\}$ , i.e.,  $\mathcal{T}_{Z;a,b}(f) \equiv \Delta \sum_{n=0}^{N-1} T_{Z;a,b,n} e^{-i2\pi f n\Delta}$ .

In the continuous-time case the smoothing operation requires averaging by integrating the wavelet spectra and cross-spectrum across a localised time interval. For the discrete-time case this becomes an average over a set of  $N_B = 2M + 1$  discrete-time indices in the neighbourhood of the translation point of interest  $b = b_0\Delta$ . Given jointly stationary series  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$

$$\begin{aligned} \mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)\} &= \frac{1}{N_B} \sum_{l=b_0-M}^{b_0+M} W(a, l\Delta; Z_1, \psi)W^*(a, l\Delta; Z_2, \psi) \\ &= \frac{1}{N_B} \sum_{l=b_0-M}^{b_0+M} \mathcal{T}_{Z_1;a,l\Delta}(1/a)\mathcal{T}_{Z_2;a,l\Delta}^*(1/a). \end{aligned} \quad (5.18)$$

Given that  $\mathcal{T}_{Z_1;a,l\Delta}(\cdot)$  is the frequency domain function calculated on the segment of  $\{Z_{1,n}\}$  centred at time index  $l$  and tapered by the Gaussian taper

$\lambda_{a,l,d;n}$ , then (5.18) is a WOSA cross-spectral estimator involving the signal portions  $\{\tilde{Z}_{1,n}, n = 0, \dots, N_P - 1\}$  and  $\{\tilde{Z}_{2,n}, n = 0, \dots, N_P - 1\}$  evaluated at frequency  $1/a$  where

$$\tilde{Z}_{m,n} \equiv Z_{m,(b_0 - \frac{N_S - 1}{2} - M + n)} \quad (5.19)$$

for  $m = 1, 2$ , and  $N_P = N_B + N_S - 1$ , with  $N_S$  assumed odd and equal to the number of terms included by the Gaussian taper.  $N_P$  will be called the *block size*. Using the truncated approximation to the Gaussian function in (5.11), the support of the dimensionless taper

$$\lambda_n^0 \equiv \pi^{-1/4} (|a_0 d|)^{-1/2} \exp(-[(n - b_0)/(a_0 d)]^2 / 2)$$

is  $n \in [b_0 - \lceil 3|a_0 d| \rceil, b_0 + \lceil 3|a_0 d| \rceil]$ , where  $\lceil x \rceil$  denotes the smallest integer exceeding  $x$ . We condense notation by letting  $\nu \equiv \lceil 3|a_0 d| \rceil$  and  $N_S = 2\nu + 1$ , and we can force  $\sum_{b_0 - \nu}^{b_0 + \nu} (\lambda_n^0)^2 = 1$  by renormalisation.

A useful shifted version of the taper is denoted

$$g_n \equiv \lambda_{n - \nu + b_0}^0, \quad n = 0, \dots, N_S - 1. \quad (5.20)$$

In the WOSA interpretation to the temporal smoothing, the averaging occurs over blocks of data that are offset by a single sample point and as such we say there is a percentage overlap of  $100(1 - 1/N_S)\%$ . As in the continuous-time case  $M$  is chosen to be proportional to  $|a|$ , with constant of proportionality (smoothing parameter)  $\kappa_0$ .

Table 5.1 gives a summary of the various parameters that have been defined for the continuous and discrete-time schemes. We note that the block size  $N_P$  is a function of the scale  $|a_0|$ , the wavelet parameter  $d$  and the smoothing parameter  $\kappa_0$ .

Figure 5.1 illustrates the various parameters for the WOSA estimator that is equivalent to the TWCOH estimator. On the bottom is a realisation of a SOS random process made up of 1024 points. On the top is the Gaussian taper component of the Morlet wavelet. In this example  $a = 10$  and  $\kappa_0 = 12$ .

	Cont.	Disc.
tapering	$\theta = 6 a d$	$N_S = 2\nu + 1; \nu = \lceil 3 a_0 d \rceil$
smoothing	$\delta =  a \kappa$	$N_B = 2M + 1; M = \lceil  a_0 \kappa_0 \rceil$
block size	$T = \delta + \theta$	$N_P = N_B + N_S - 1$

Table 5.1: Comparison of parameters in continuous-time and discrete-time cases.

The bold section of the signal represents the portion  $\{\tilde{Z}_n, n = 0, \dots, N_P - 1\}$  that is used to determine the WOSA spectral estimator for translation time  $b = 400$  (the dotted line). The number of points between the two solid lines is  $N_P = 301$ . The number of points between two dashed lines is  $N_S = 61$ .  $N_B$  of these sub-blocks (each overlapping with the next by a single point) are averaged together to give the WOSA estimator. In this example  $N_B = 241$ .

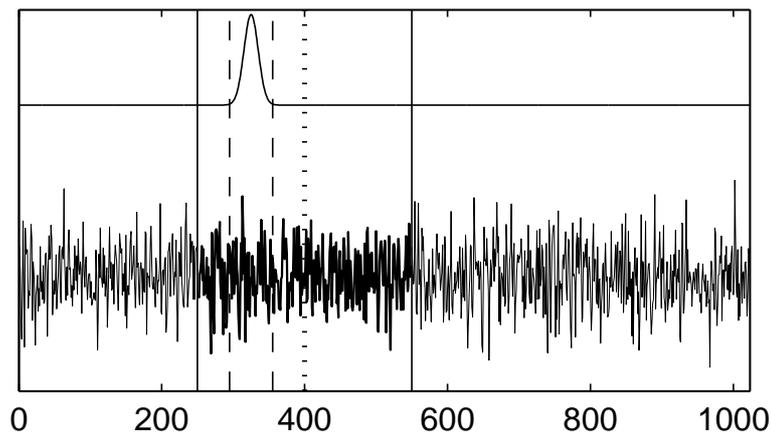


Figure 5.1: Illustration of the WOSA estimator equivalent to the TWCOH estimator for a translation point of interest  $b = 400$  and scale  $a = 10$ . The realisation of the process is shown on the bottom and the Gaussian taper component on the top. The solid vertical lines and bold portion represent the part used in calculating the estimate for  $b = 400$ . The dashed lines shows an individual block of size  $N_S$ ,  $N_B$  of which are averaged together.

### 5.3.1 Restrictions on $a$ and $b$

We recall from Section 3.3 that in the implementation of the discrete-time CWT there are restrictions on the scale values  $a$  and translation points  $b$  that need to be considered for validity. The minimum usable absolute scale is specified as

$$a_{\min}(d) = f'(d)/f_{\mathcal{N}} \quad (5.21)$$

where the cut-off frequency of the Morlet wavelet  $f'(d)$  is defined in this case as the lowest frequency such that  $\Psi(f) < 10^{-6}\Psi(1)$  for all  $f > f'(d)$ .

Consider a specific positive  $a > a_{\min}$ . For any  $b_0$  the Gaussian taper will extend from  $b_0 - \nu$  to  $b_0 + \nu$  — see Table 5.1. The data extends from 0 to  $N - 1$  so to avoid wraparound effects with the CWT we only consider time points  $\nu \leq b_0 \leq N - 1 - \nu$ . However, we also need to consider the smoothing operation in (5.18) which restricts the choice of  $b_0$  further to  $\nu + M(a) \leq b_0(a) \leq N - 1 - \nu - M(a)$  giving

$$b_{\min}(a) = \nu + M(a); \quad b_{\max}(a) = N - 1 - \nu - M(a). \quad (5.22)$$

We note that for  $b_{\min}(a)$ , the signal portion, namely  $\{\tilde{Z}_{m,n}, n = 0, \dots, N_P - 1\}$ , is  $\{Z_{m,0}, \dots, Z_{m,N_P - 1}\}$ , and for  $b_{\max}(a)$  the signal portion is  $\{Z_{m,N - N_P}, \dots, Z_{m,N - 1}\}$ , for  $m = 1, 2$ , meaning that as expected there are no end-effects. Obviously we must require that  $b_{\min}(a) \leq b_{\max}(a)$ , which gives the condition that

$$\begin{aligned} \nu + M(a) &\leq N - 1 - \nu - M(a) \\ N_S + 2M(a) &\leq N \end{aligned} \quad (5.23)$$

i.e.,  $N_S + N_B - 1 = N_P \leq N$ , which states that the block size  $N_P$  used in calculating a single point in the WOSA-type estimator can not be larger than the total number of observed points  $N$ . To determine the maximum absolute

analysis scale  $a_{\max}$ , we consider (5.23) which can be written

$$\begin{aligned} (N_S - 1)/2 + M(a) &\leq (N - 1)/2 \\ \lceil 3|a_0|d \rceil + \lceil |a_0|\kappa_0 \rceil &\leq (N - 1)/2. \end{aligned} \quad (5.24)$$

For an arbitrary pair of positive reals  $x$  and  $y$ ,  $\lceil xy \rceil = xy + \xi$ ,  $0 \leq \xi \leq 1$  and as such (5.24) becomes

$$3|a_0|d + |a_0|\kappa_0 \leq [(N - 1)/2] - \xi_0, \quad \text{for } 0 \leq \xi_0 \leq 2.$$

Hence,  $a_{\max}$  is given by

$$a_{\max} = \Delta \frac{[(N - 1)/2] - 2}{3d + \kappa_0} = \Delta \frac{N - 5}{6d + 2\kappa_0}.$$

Figures 5.2(a) and 5.3(a) give the variation of the block size  $N_P$  for positive  $a_0$  with  $a_{\min}/\Delta \leq a_0 \leq a_{\max}/\Delta$  for parameters  $d = 1$ ,  $f'(d) = 1.8$ ,  $N = 1024$ ,  $\kappa_0 = 12$  and  $d = 2$ ,  $f'(d) = 1.4$ ,  $N = 1024$ ,  $\kappa_0 = 24$  respectively. The corresponding valid regions of  $(a_0, b_0)$  is given by the interior of the central triangle in Figures 5.2(b) and 5.3(b).

### 5.3.2 Review

It has been shown that by using a Morlet wavelet, temporal smoothing can be represented mathematically as a WOSA estimation based on a subsection of the processes of block size  $N_P$  centred at the translation point of interest  $b$ , with analysis at scale  $a$  corresponding to frequency  $f = 1/a$ . In the continuous-time case, the overlap can be considered to be 100%, for the discrete-time case it is  $100(1 - 1/N_S)\%$ . We conclude that if we know the distributional properties of such WOSA spectral estimators, and the coherence estimators formed from them then the distributional properties of TWCOH estimators will follow.

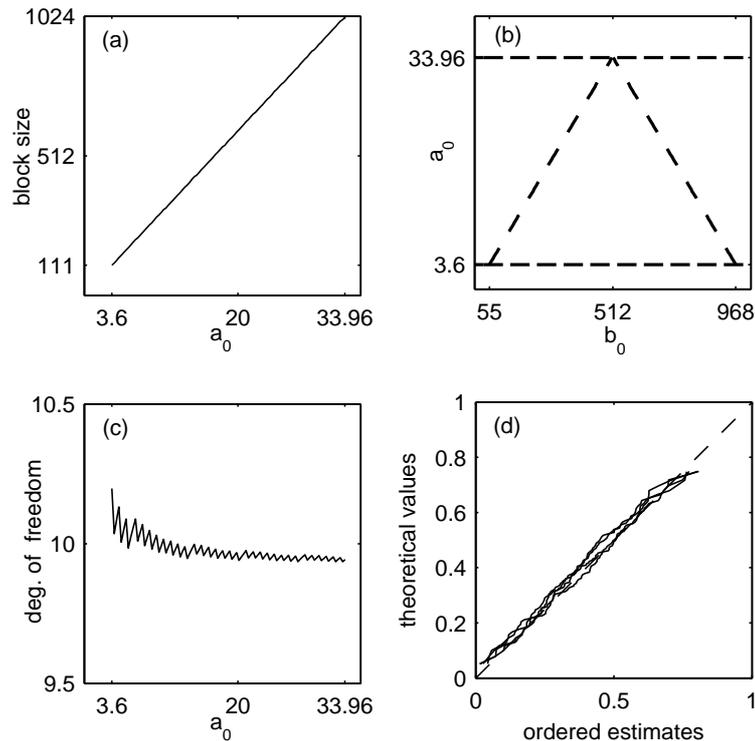


Figure 5.2: Parameter relationships and statistical goodness-of-fit assessment when  $d = 1$ ,  $f'(d) = 1.8$ ,  $N = 1024$ ,  $\kappa_0 = 12$ . (a) Block size  $N_P$  versus  $a_0$ . (b) Region of validity for  $(a_0, b_0)$  given by the interior of the central triangle. (c) Degrees of freedom  $\varphi$  versus  $a_0$ . (d) Goodman Q-Q plots for four different  $(a_0, b_0)$  locations.

## 5.4 Statistical Properties of WOSA Estimators

In [102] it is shown that the cross and standard WOSA estimators can mathematically be cast into a multitaper form. The distributional properties of these multitaper spectral estimators and subsequent coherence estimators are known and consequently are appropriate for the WOSA spectral and coherence estimators. Using this route it will be shown that the distributional properties of TWCOH estimators naturally follow.

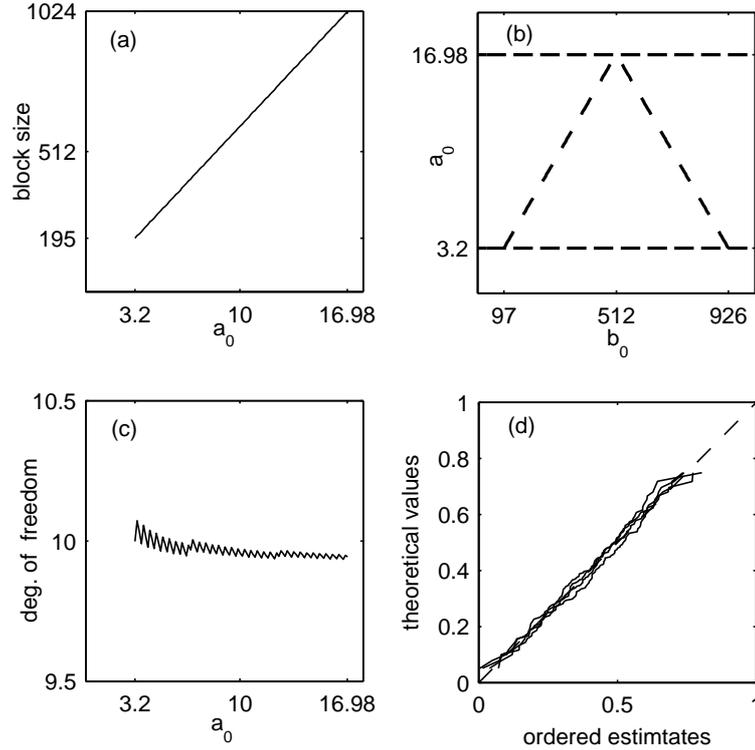


Figure 5.3: Same style of plots as for Figure 5.2 but with  $d = 2$ ,  $f'(d) = 1.4$ ,  $N = 1024$ ,  $\kappa_0 = 24$ .

#### 5.4.1 WOSA to Multitaper

For the estimator (5.18) we define

$$\hat{S}_{12,b}(1/a) \equiv \mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)\}. \quad (5.25)$$

$\hat{S}_{12,b}(1/a)$  is the cross spectral estimator for the signal portions  $\{\tilde{Z}_{1,n}, n = 0, \dots, N_P - 1\}$  and  $\{\tilde{Z}_{2,n}, n = 0, \dots, N_P - 1\}$  centred at  $b$  (defined in (5.19)) evaluated at the frequency  $1/a$ . WOSA spectral density function estimators are defined in an analogous manner. Offset between portions is by a single point and  $N_P = N_B + N_S - 1$ ,  $N_B = 2M + 1$  and  $N_S = 2\nu + 1$ .

In the WOSA formulation of the TWCOH estimator, the block size of  $N_P$  is divided into  $N_B$  sub-blocks, each of size  $N_S$ . A shift of unity is applied

between each contiguous block and using the data taper (5.20)  $g_0, \dots, g_{N_S-1}$ , scaled so that  $\sum_{n=0}^{N_S-1} g_n^2 = 1$ , the WOSA cross-spectrum estimator is defined as

$$\hat{S}_{12,b}(f) \equiv \frac{1}{N_B} \sum_{j=0}^{N_B-1} \hat{S}_{12,b;j}(f) \quad (5.26)$$

where the estimator for sub-block  $j$ ,  $0 \leq j \leq N_B - 1$ , is  $\hat{S}_{12,b;j}(f)$  given by

$$\Delta \left[ \sum_{l=0}^{N_S-1} g_l \tilde{Z}_{1,l+j} e^{-i2\pi fl\Delta} \right] \left[ \sum_{n=0}^{N_S-1} g_n \tilde{Z}_{2,n+j}^* e^{i2\pi fn\Delta} \right].$$

To convert this to the multitaper estimator we begin by rewriting this as a sum over the whole signal portion and padding the taper with zeros to make it of length  $N_P$  i.e. [8]

$$\Delta \left[ \sum_{l=0}^{N_P-1} g'_{j,l} \tilde{Z}_{1,l+j} e^{-i2\pi fl\Delta} \right] \left[ \sum_{n=0}^{N_P-1} g'_{j,n} \tilde{Z}_{2,n+j}^* e^{i2\pi fn\Delta} \right].$$

where for  $n = 0, \dots, N_P - 1$

$$g'_{j,n} = \begin{cases} g_{n-j}, & \text{if } n = j, \dots, j + N_S - 1; \\ 0, & \text{otherwise.} \end{cases} \quad (5.27)$$

Using this representation (5.26) can be rewritten as

$$\hat{S}_{12,b}(f) = \frac{\Delta}{N_B} \sum_{j=0}^{N_B-1} \left[ \sum_{l=0}^{N_P-1} g'_{j,l} \tilde{Z}_{1,l+j} e^{-i2\pi fl\Delta} \right] \left[ \sum_{n=0}^{N_P-1} g'_{j,n} \tilde{Z}_{2,n+j}^* e^{i2\pi fn\Delta} \right]$$

There exists a matrix representation of this equation [102]. We let

$$D_{m,l} = \tilde{Z}_{m,l}^* \exp(i2\pi fl\Delta),$$

for  $m = 1, 2$ , and  $\mathbf{D}_m = [D_{m,0}, \dots, D_{m,N_P-1}]^T$ . We also let  $\mathbf{B}$  be the  $N_P \times N_B$

matrix with  $j$ th column given by  $g'_{j,n}/\sqrt{N_B}$ , i.e.

$$\mathbf{B} = \frac{1}{\sqrt{N_B}} \begin{bmatrix} g_0 & 0 & 0 & \cdots & 0 & 0 \\ g_1 & g_0 & 0 & \cdots & 0 & 0 \\ g_2 & g_1 & g_0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g_{N_S-1} & g_{N_S-2} \\ 0 & 0 & 0 & \cdots & 0 & g_{N_S-1} \end{bmatrix}. \quad (5.28)$$

Then

$$\hat{S}_{12,b}(f) = \Delta \mathbf{D}_1^H \mathbf{B} \mathbf{B}^T \mathbf{D}_2. \quad (5.29)$$

The outer product  $\mathbf{B} \mathbf{B}^T$  is a symmetric positive semidefinite  $N_P \times N_P$  matrix and as such will have real-valued eigenvectors and non-negative eigenvalues [94, p. 296].

By choosing the taper  $\{g_n\}$  to have all non-zero elements, the columns of  $\mathbf{B}$  are linearly independent, i.e.  $\mathbf{B}$  is a full rank matrix, namely  $\text{rank}\{\mathbf{B}\} = N_B$ . Further to this, because  $\{g_n\}$  is real-valued we have  $\text{rank}\{\mathbf{B}\} = \text{rank}\{\mathbf{B}^T \mathbf{B}\} = \text{rank}\{\mathbf{B} \mathbf{B}^T\}$  giving  $\mathbf{B} \mathbf{B}^T$  to be itself of rank  $N_B$ . From this we determine that the matrix  $\mathbf{B} \mathbf{B}^T$  has  $N_B$  positive eigenvalues which we shall index such that  $\sigma_0 \geq \cdots \geq \sigma_{N_B-1}$  and the  $(N_P - N_B)$  remaining eigenvalues will be zero, i.e.  $\sigma_{N_B}, \dots, \sigma_{N_P-1} = 0$ . Associated with the positive eigenvalues are the set of corresponding orthonormal eigenvectors  $\mathbf{u}_0, \dots, \mathbf{u}_{N_B-1} \in \mathbb{R}^{N_P}$ .

Using singular value decomposition it is possible to rewrite  $\mathbf{B} \mathbf{B}^T$  in terms of the eigenvalues and eigenvectors,

$$\mathbf{B} \mathbf{B}^T = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^T = \sum_{j=0}^{N_B-1} \sigma_j \mathbf{u}_j \mathbf{u}_j^T$$

where  $\mathbf{U}$  is an  $N_P \times N_B$  matrix with  $j$ th column  $\mathbf{u}_j$ , and  $\mathbf{\Sigma} = \text{diag}\{\sigma_0, \dots, \sigma_{N_B-1}\}$ .

It is possible to write  $\mathbf{U} \mathbf{\Sigma} \mathbf{U}^T = \mathbf{A} \mathbf{A}^T$  where  $\mathbf{A} = [\mathbf{u}_0 \sqrt{\sigma_0}, \dots, \mathbf{u}_{N_B-1} \sqrt{\sigma_{N_B-1}}]$

is an  $N_P \times N_B$  matrix with orthogonal columns, meaning we can write the matrix representation of the WOSA estimator in (5.29) instead as

$$\hat{S}_{12,b}(f) = \Delta \mathbf{D}_1^H \mathbf{A} \mathbf{A}^T \mathbf{D}_2.$$

Taking this out of matrix notation this can be rewritten as [102]

$$\hat{S}_{12,b}(f) = \frac{\Delta}{N_B} \sum_{j=0}^{N_B-1} \gamma_j \left[ \sum_{l=0}^{N_P-1} u_{j,l} \tilde{Z}_{1,l} e^{i2\pi f l \Delta} \right] \left[ \sum_{n=0}^{N_P-1} u_{j,n} \tilde{Z}_{2,n} e^{i2\pi f n \Delta} \right]^* \quad (5.30)$$

where  $u_{j,l}$  is the  $l$ th element of the vector  $\mathbf{u}_j$  and  $\gamma_j$  is termed the  $j$ th *weighting factor* related to the  $j$ th eigenvalue via  $\gamma_j = N_B \sigma_j$ . It can be shown that due to the standardisation of  $\{g_n\}$  we have  $\sum_{j=0}^{N_B-1} \gamma_j = N_B$ , [102].

We have said the set of  $N_B$  eigenvectors  $\mathbf{u}_0, \dots, \mathbf{u}_{N_B-1}$  are orthonormal, i.e. for all  $l, m = 0, \dots, N_B - 1$  we have  $\langle \mathbf{u}_l, \mathbf{u}_m \rangle = \delta_{lm}$ . We therefore recognise the representation of  $\hat{S}_{12,b}(f)$  in (5.30) as a weighted multitaper cross-spectral estimator that uses a set of  $N_B$  orthonormal tapers to create a set of individual spectral estimators across which averaging can be carried out.

A multitaper formulation of the spectrum

$$\hat{S}_{11,b}(1/a) \equiv \mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_1, \psi)\},$$

and likewise  $\hat{S}_{22}(1/a)$ , follow by simply setting both time series to be the same. For reasons already outlined we might also want to consider relation terms

$$\hat{R}_{12,b}(1/a) \equiv \mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2^*, \psi)\}, \quad (5.31)$$

i.e. where the second process is conjugated. This takes the form

$$\hat{R}_{12,b}(f) = \frac{\Delta}{N_B} \sum_{j=0}^{N_B-1} \gamma_j \left[ \sum_{l=0}^{N_P-1} u_{j,l} \tilde{Z}_{1,l} e^{i2\pi f l \Delta} \right] \left[ \sum_{n=0}^{N_P-1} u_{j,n} \tilde{Z}_{2,n}^* e^{i2\pi f n \Delta} \right]^* \quad (5.32)$$

and again  $\hat{R}_{11}(1/a)$  and  $\hat{R}_{22}(1/a)$  analogously follow by setting the two time

series to be complex conjugates of each other.

It is shown in [102] that the estimator (5.30) consisting of  $N_B$  orthonormal tapers with corresponding weighting factors  $\gamma_0, \dots, \gamma_{N_B-1}$  has

$$\varphi = \frac{N_B^2}{\sum_{j=0}^{N_B-1} \gamma_j^2} = \frac{1}{\sum_{j=0}^{N_B-1} \sigma_j^2} \quad (5.33)$$

complex degrees of freedom. The quantity in (5.33) is critical in determining the distribution of the TWCOH estimator. The degrees of freedom are a function of the eigenvalues of the matrix  $\mathbf{B}\mathbf{B}^T$ , which in turn are a function of taper  $\{g_n\}$ , which in turn depends solely on the choice of  $N_S$ , together with the amount of averaging  $N_B$ .

An analytical expression for the degrees of freedom can be achieved by using a consequence of the singular valued decomposition, that is that the sum of the squares of the eigenvalues of a real-valued symmetric matrix is equal to the sum of the squares of the elements. For the  $N_B$  ranked  $N_P \times N_P$  matrix  $\mathbf{B}\mathbf{B}^T$  this gives

$$\sum_{j=0}^{N_B-1} \sigma_j^2 = \sum_{j=1}^{N_P} \sum_{k=1}^{N_P} [(\mathbf{B}\mathbf{B}^T)_{jk}]^2$$

and consequently using (5.33) we have

$$\varphi = \frac{1}{\sum_{j=1}^{N_P} \sum_{k=1}^{N_P} [(\mathbf{B}\mathbf{B}^T)_{jk}]^2}.$$

By definition  $(\mathbf{B}\mathbf{B}^T)_{jk} = \sum_{l=1}^{N_P} B_{jl}B_{kl}$ , where  $B_{jl}$  is the  $(j, l)$ th element of  $\mathbf{B}$ , hence using (5.27),  $(\mathbf{B}\mathbf{B}^T)_{jk} = (1/N_B) \sum_{l=1}^{N_P} g'_{j,l}g'_{k,l}$  giving

$$\varphi = \frac{N_B^2}{\sum_{j=1}^{N_P} \sum_{k=1}^{N_P} \left[ \sum_{l=1}^{N_P} g'_{j,l}g'_{k,l} \right]^2}. \quad (5.34)$$

While an expression for  $\varphi$  solely in terms of  $N_S$  and  $N_B$  has proven elusive, in Figure 5.4 are shown the contours of constant degrees of freedom  $\varphi$  as a

function of  $N_S$  and  $N_B$ . It is clear to see from this plot that complex degrees of freedom  $\varphi$  are constant for a constant gradient or ratio  $N_B/N_S$ . For example, for  $\varphi = 10$  complex degrees of freedom, the ratio  $N_B/N_S = 200/50 = 4$ . From Table 5.1 we can see that for the sort of values used in practice  $N_B/N_S \approx \kappa_0/(3d)$  and thus the choices  $\kappa_0 = 12$ ,  $d = 1$  or  $\kappa_0 = 24$ ,  $d = 2$ , are both associated with approximately ten complex degrees of freedom.

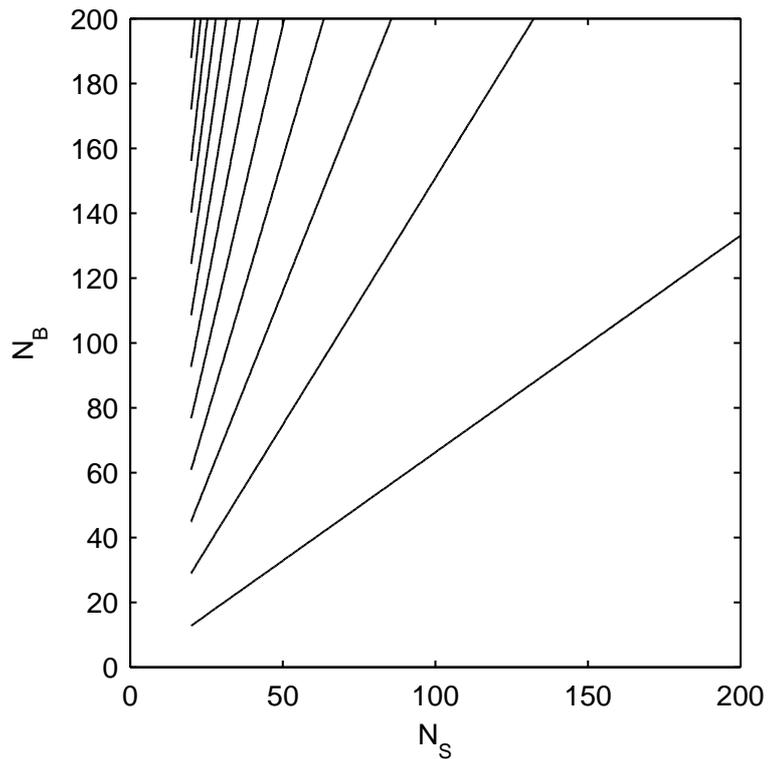


Figure 5.4: Contours of constant complex degrees of freedom in the  $N_S$ - $N_B$  plain. The right most line is for  $\varphi = 2$  with lines increasing in steps of 2 from right to left.

## 5.5 Statistical Analysis of TWCOH

Using the multitaper formulation of the WOSA estimator we are now in a position to derive the distribution of the TWCOH estimator developed by Cohen

and Walden [15]. For notation purposes we consider the  $p = 2$  dimensional case. Extending to higher dimensions is straight forward.

Let  $\{\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}]^T, t \in \mathbb{Z}\}$  be a zero-mean complex-valued bivariate random process. The two processes are assumed to be jointly SOS. As described in Section 2.3, using the augmented process  $\{\check{\mathbf{Z}}_t = [\mathbf{Z}_t^T, \mathbf{Z}_t^H]^T, t \in \mathbb{Z}\}$  we can define the augmented SDF matrix  $\Upsilon(f)$  as in (2.6).

For ease of notation we will henceforth write  $\mathcal{S}\{Z_1, Z_2\}$  in place of the term  $\mathcal{S}\{W(a, b; Z_1, \psi)W^*(a, b; Z_2, \psi)\}$ . Dependence on  $a, b$  and  $\psi$  is implicit. Consider the matrix

$$\hat{\Omega}(a, b) = \begin{bmatrix} \mathcal{S}\{Z_1, Z_1\} & \mathcal{S}\{Z_1, Z_2\} & \mathcal{S}\{Z_1, Z_1^*\} & \mathcal{S}\{Z_1, Z_2^*\} \\ \mathcal{S}\{Z_2, Z_1\} & \mathcal{S}\{Z_2, Z_2\} & \mathcal{S}\{Z_2, Z_1^*\} & \mathcal{S}\{Z_2, Z_2^*\} \\ \mathcal{S}\{Z_1^*, Z_1\} & \mathcal{S}\{Z_1^*, Z_2\} & \mathcal{S}\{Z_1^*, Z_1^*\} & \mathcal{S}\{Z_1^*, Z_2^*\} \\ \mathcal{S}\{Z_2^*, Z_1\} & \mathcal{S}\{Z_2^*, Z_2\} & \mathcal{S}\{Z_2^*, Z_1^*\} & \mathcal{S}\{Z_2^*, Z_2^*\} \end{bmatrix}.$$

Using the definitions in (5.25) and (5.31) this matrix can equivalently be written as  $\hat{\Upsilon}_b(1/a)$  given by

$$\begin{bmatrix} \hat{S}_{11,b}(1/a) & \hat{S}_{12,b}(1/a) & \hat{R}_{11,b}(1/a) & \hat{R}_{12,b}(1/a) \\ \hat{S}_{21,b}(1/a) & \hat{S}_{22,b}(1/a) & \hat{R}_{21,b}(1/a) & \hat{R}_{22,b}(1/a) \\ \hat{R}_{11,b}^*(1/a) & \hat{R}_{12,b}^*(1/a) & \hat{S}_{11,b}(-1/a) & \hat{S}_{12,b}(-1/a) \\ \hat{R}_{21,b}^*(1/a) & \hat{R}_{22,b}^*(1/a) & \hat{S}_{21,b}(-1/a) & \hat{S}_{22,b}(-1/a) \end{bmatrix},$$

with  $\hat{\Omega}(a, b) = \hat{\Upsilon}_b(1/a)$ . Equations (5.30) and (5.32) give the terms of this matrix.  $\hat{\Upsilon}_b(1/a)$  is an estimator for  $\Upsilon(1/a)$  for different values of  $b$ ; the  $b$  merely defines which blocks of size  $N_P$  of the time series are being used.

It has been shown in Section 2.8 that when performing a multitaper spectral estimator on a realisation of an augmented complex-Gaussian process  $\{\check{\mathbf{Z}}_n, n = 0, \dots, N - 1\}$  using  $K$  orthogonal tapers  $\{u_{k,n}, n = 0, \dots, N - 1\}$ ,

$k = 0, \dots, K - 1$ , the set of vectors  $\{\check{\mathbf{J}}_k(f), k = 0, \dots, K - 1\}$ , where

$$\check{\mathbf{J}}_k(f) = \Delta^{1/2} \sum_{n=0}^{N-1} u_{k,n} \check{\mathbf{Z}}_{l,n} e^{-i2\pi fn\Delta}.$$

are all proper and complex-Gaussian IID. Consequently we can invoke the argument of [35] and state that the augmented SDF matrix estimator

$$\hat{\mathbf{Y}}^{(mt)}(f) = (1/K) \sum_{k=0}^{K-1} \check{\mathbf{J}}_k(f) \check{\mathbf{J}}_k^H(f)$$

is central complex-Wishart distributed with integer  $K$  degrees of freedom. In the multitaper formulation of the WOSA spectral estimator provided in this chapter we have to consider a weighted multitaper estimator of the form

$$\hat{\mathbf{Y}}_b(1/a) = (1/N_B) \sum_{k=0}^{N_B-1} \gamma_k \check{\mathbf{J}}_k(f) \check{\mathbf{J}}_k^H(f)$$

where in this instance  $\check{\mathbf{J}}_k(f) = \Delta^{1/2} \sum_{n=0}^{N_P-1} u_{k,n} [\tilde{\mathbf{Z}}_n^T, \tilde{\mathbf{Z}}_n^H]^T e^{-i2\pi fn\Delta}$ . We can write this

$$\hat{\mathbf{Y}}_b(1/a) = (1/N_B) \sum_{k=0}^{N_B-1} \tilde{\mathbf{J}}_k(f) \tilde{\mathbf{J}}_k^H(f)$$

where  $\tilde{\mathbf{J}}_k(f) = \sqrt{\gamma_k} \check{\mathbf{J}}_k(f)$ . Due to the weighting factors  $\{\gamma_k, k = 0, \dots, N_B - 1\}$ , the vectors  $\tilde{\mathbf{J}}_k(f)$ , while still complex-Gaussian and proper, are no longer identically distributed and consequently a different approach is needed to deriving the distribution properties of the SDF matrix estimator  $\hat{\mathbf{Y}}_b(1/a)$ .

**Theorem 5.5.1.** *Let  $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$  be a zero-mean complex-valued bivariate Gaussian SOS random process. Then, asymptotically*

$$\hat{\mathbf{Y}}_b(1/a) \stackrel{d}{=} (1/\varphi) \mathcal{W}_4^C\{\varphi, \mathbf{Y}(1/a)\}$$

where  $\mathcal{W}_4^C\{\varphi, \mathbf{Y}(1/a)\}$  denotes the four-dimensional complex central Wishart distribution with  $\varphi$  complex degrees of freedom and mean  $\varphi \mathbf{Y}(1/a)$ .

*Proof.* This is found in Appendix C.  $\square$

We therefore conclude that asymptotically as  $N_P \rightarrow \infty$ ,

$$\hat{\Omega}(a, b) \stackrel{d}{=} (1/\varphi)\mathcal{W}_4^C\{\varphi, \Upsilon(1/a)\}.$$

Referring back to Section 4.4.1 where we define the three different classes of wavelet coherence for a multivariate complex-valued SOS process using Morse wavelets, for temporal smoothing the three classes of wavelet coherence are now:

1. the ordinary TWCOH estimator  $\hat{\gamma}_{lm}^2(a, b)$  for the pair of processes  $\{Z_{l,t}\}$  and  $\{Z_{m,t}\}$ , defined as

$$\begin{aligned} \hat{\gamma}_{lm}^2(a, b) &= \frac{|\mathcal{S}\{Z_l Z_m\}|^2}{\mathcal{S}\{Z_l Z_l\}\mathcal{S}\{Z_m Z_m\}} \\ &= \frac{|\hat{\Omega}_{l,m}(a, b)|^2}{\hat{\Omega}_{l,l}(a, b)\hat{\Omega}_{m,m}(a, b)}; \end{aligned} \quad (5.35)$$

2. the conjugate TWCOH estimator  $\hat{\gamma}_{lm_*}^2(a, b)$  for the pair of processes  $\{Z_{l,t}\}$  and  $\{Z_{m,t}^*\}$ , defined as

$$\begin{aligned} \hat{\gamma}_{lm_*}^2(a, b) &= \frac{|\mathcal{S}\{Z_l Z_m^*\}|^2}{\mathcal{S}\{Z_l Z_l\}\mathcal{S}\{Z_m^* Z_m^*\}} \\ &= \frac{|\hat{\Omega}_{l,m+p}(a, b)|^2}{\hat{\Omega}_{l,l}(a, b)\hat{\Omega}_{m+p,m+p}(a, b)}; \end{aligned} \quad (5.36)$$

3. the conjugate TWCOH estimator  $\hat{\gamma}_{ll_*}^2(a, b)$  for the pair of processes  $\{Z_{l,t}\}$  and  $\{Z_{l,t}^*\}$ , defined as

$$\begin{aligned} \hat{\gamma}_{ll_*}^2(a, b) &= \frac{|\mathcal{S}\{Z_l Z_l^*\}|^2}{\mathcal{S}\{Z_l Z_l\}\mathcal{S}\{Z_l^* Z_l^*\}} \\ &= \frac{|\hat{\Omega}_{l,l+p}(a, b)|^2}{\hat{\Omega}_{l,l}(a, b)\hat{\Omega}_{l+p,l+p}(a, b)}. \end{aligned} \quad (5.37)$$

Invoking Theorem 2.8.2, an ordinary or conjugate TWCOH estimator  $\hat{\gamma}^2(a, b)$  of type (5.35), (5.36) or (5.37) will asymptotically be Goodman distributed with PDF

$$g_{\hat{\gamma}^2}(x; \varphi, \gamma^2) = (\varphi - 1) (1 - \gamma^2)^\varphi (1 - x)^{\varphi-2} {}_2F_1(\varphi, \varphi; 1; \gamma^2 x). \quad (5.38)$$

where  $\gamma^2$  is the corresponding true value of ordinary or conjugate coherence — see Section 2.6 — evaluated at the frequency  $1/a$ .

## 5.6 Example results

To illustrate the derived theory we consider the simulations of the same improper CAR(1) process  $\{Z_{1,t}\}$  used in Section 4.6, using a sampling rate of  $\Delta = 1$ . We examine the conjugate coherence between  $\{Z_{1,t}\}$  and  $\{Z_{1,t}^*\}$  that has a true value of  $\gamma_{11}^2(f) = 0.36$  for all  $|f| < 1/2$ .

First, we consider analysis with a Morlet wavelet of form (5.5) with  $d = 1$ , a parameter choice which provides a “reasonable compromise between frequency and time resolution” [101]. The corresponding cut-off frequency is  $f' = 1.8$ . As described in Section 5.4, a choice of  $\kappa_0 = 12$  gives approximately 10 complex degrees of freedom, a sensible practical choice. In more precise terms, every valid scale value  $a_{\min} \leq a_0 \leq a_{\max}$  will combine with values of  $d$  and  $\kappa_0$  to produce a corresponding value of  $N_S$ ,  $N_B$  and  $N_P$  — see table 5.1. With use of the appropriate Gaussian taper the complex degrees of freedom  $\varphi$  can be obtained through either forming the matrix  $\mathbf{B}$  and calculating the eigenvalues of  $\mathbf{B}\mathbf{B}^T$  and using (5.33), or directly from the taper using (5.34). These degrees of freedom are shown as a function of scale in Figure 5.2(c). It is seen that  $\varphi \approx 10$  for all valid  $a_0$  values, as predicted in Section 5.4.

Realisations of the CAR(1) process used were of length  $N = 1024$ . For each simulated sequence at each  $(a, b)$  point

$$\begin{bmatrix} \mathcal{S}\{Z_1, Z_1\} & \mathcal{S}\{Z_1, Z_1^*\} \\ \mathcal{S}\{Z_1^*, Z_1\} & \mathcal{S}\{Z_1^*, Z_1^*\} \end{bmatrix} \equiv \begin{bmatrix} \hat{S}_{11,b}(1/a) & \hat{R}_{11,b}(1/a) \\ \hat{R}_{11,b}^*(1/a) & \hat{S}_{11,b}(-1/a) \end{bmatrix}$$

was calculated using (5.18). Following this the TWCOH estimator

$$\hat{\gamma}_{11*}^2(a, b) = \frac{|\hat{R}_{11,b}(1/a)|^2}{\hat{S}_{11,b}(1/a)\hat{S}_{11,b}(-1/a)}$$

was computed. This was repeated 100 times and the ordered values of  $\hat{\gamma}_{11*}^2(a, b)$  were plotted against the theoretical quantiles of the Goodman distribution with  $\gamma_{11*}^2(a, b) = 0.36$ . The resulting Q-Q plots are shown in Figure 5.2(d) where the individual curves are for four different  $(a, b)$  locations within the permitted ranges. At all locations we note that  $N_P > 128$ , which provides a suitably large signal portion to render excellent agreement with the Goodman distribution.

For a second example, we use the choice of  $d = 2$  and  $\kappa_0 = 24$ . The corresponding cut-off frequency is  $f' = 1.4$ . The corresponding degrees of freedom are shown in Figure 5.3(c) and as expected  $\varphi \approx 10$  for all  $a_0$  values. With these parameter values, the resulting Q-Q plots are shown in Figure 5.3(d) and again indicate an excellent agreement with the Goodman distribution.

## 5.7 Significance Testing Example

We now look at actual data recorded by a pair of spacecraft involved in the solar physics ‘cluster’ experiment [62]. We consider the two complex-valued magnetic field time series  $\{Z_{1,n}\}$  and  $\{Z_{2,n}\}$ , plotted in Figure 5.5. The component measured parallel to the ecliptic plane forms the real part of the series, and the perpendicular component forms the imaginary part. The series is sampled at  $\Delta = 2$  s and is of length 1000 s (i.e. 500 data points).

We examine the ordinary coherence between the process  $\{Z_{1,n}\}$  and  $\{Z_{2,n}\}$ . The left plot of Figure 5.6 shows the TWCOH estimate, while the right plot shows only values exceeding the 1% point of the Goodman’s distribution under the null hypothesis  $\gamma^2(a, b) = 0$ , i.e., values  $\hat{\gamma}^2 > 1 - \alpha^{1/(\varphi-1)}$  for  $\alpha = 0.01$ ,  $\varphi = 10$ .

There is an important issue that needs to be considered for such analysis,

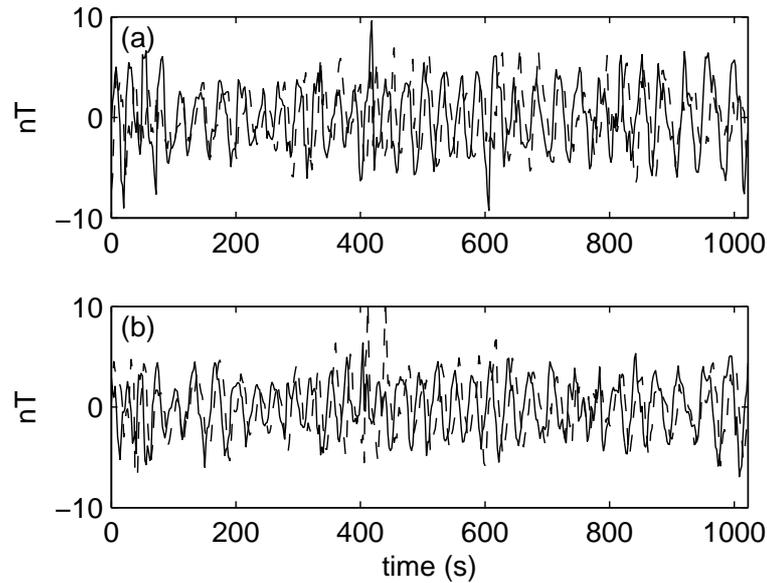


Figure 5.5: Magnetic time series  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$ . The real part is plotted with a solid line, and the imaginary part is dashed.

and that is the one of multiple testing. When performing multiple independent tests of the hypothesis each at a  $(1 - \alpha)$  level, the experiment wide error rate, the probability of incorrectly rejecting the null hypothesis, increases with the more tests that are carried out. This is known as the multiple testing problem. Testing for wavelet coherence is also an example of multiple testing. While the experiment is not being repeated, the multiple points on the time-scale plane that are separately being analysed can be considered as a set of multiple tests. If all points on the time-scale plane were uncorrelated from each other then we would expect the significant points to be randomly distributed across time-scale plane. However it is shown in [60, Appendix A] that adjacent values of WCOH on the time-scale plane are correlated due to the sum across data points in the CWT, and the smoothing process. This means that false positive coherence values appear in patches, giving the illusion that they represent a significant area of WCOH. To overcome this problem a scheme is needed to decide whether a patch is significant or a result of multiple testing.

Maraun and Kurths [60] [61] design a method for testing whether an area

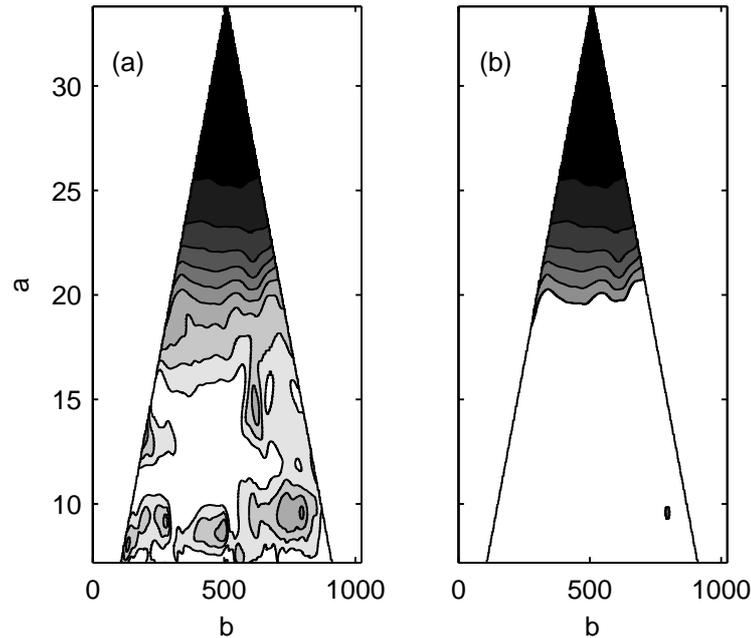


Figure 5.6: Wavelet coherence estimate as a function of scale  $a$  and time  $b$  of magnetic field data (left), and those values exceeding the 1% point of the null distribution (right). (Higher coherence corresponds to darker shading.)

of WCOH is significant or not by studying the correlations that exist between adjacent points on the time-scale plane. Given a single point is significant, Maraun and Kurths determine the size and geometry of the area surrounding that point that will also show significance, termed the *critical patch*. A point on the time-scale plane is then deemed *area-wise significant* if every critical patch containing the point of interest lies completely within a patch of pointwise significant points.

An adaptation of the scheme would be needed to take into account the new findings on TWCOH distributions presented here, this is beyond the scope of this thesis. However in this example the size of the patch for scales exceeding 20 is clearly larger than 1% of the overall area, so with a high degree of certainty it can be said that this represents nonnull coherence. Using the mapping between scale and frequency  $f = 1/a$ , this equates to coherent waves in the magnetic field for periods exceeding 20s. We notice there is a much

smaller patch at the scale  $a = 10$ ; perhaps this is a spurious patch that results from multiple testing.

## 5.8 Concluding Remarks

In this chapter explicit expressions have been provided that show the temporally smoothed wavelet spectral estimators can be written in terms of WOSA-type spectral estimators. These in turn can be expressed in terms of multitaper estimators utilising orthonormal tapers. The consequence of formulating the TWCOH in this way is the number of complex degrees of freedom are immediately revealed. From this a fully specified Goodman distribution results for the TWCOH estimator using a Morlet wavelet, applied to SOS complex-valued Gaussian processes.

Due to the form of the Morlet wavelet we have used a Gaussian envelope to taper the complex sinusoid, or equivalently taper the signal. Other sensible choices of taper would fit into the general approach presented here and the statistical analysis would proceed in an identical way, meaning this approach to be quite general.

The key advantage of this approach is the ease with which we can control the size of the data taper and the smoothing window. In the discrete-time formulation a finite portion of the signal (of length  $N_P$ ) is used to estimate either the wavelet spectrum or TWCOH at a particular time of interest  $b$ . Consequently for the derived distribution of the matrix  $\hat{\Upsilon}_b(1/a) \equiv \hat{\Omega}(a, b)$  or TWCOH estimator  $\hat{\gamma}^2(a, b)$  to be valid we only require stationarity across the signal portion of length  $N_P$ , rather than the whole of the observed time series. This method should therefore lend itself to analysing signals that while approximately stationary on a closed interval of time, may in fact globally display nonstationary behaviour.

In the remainder of this thesis we turn our attention to the concept of coherence in nonstationary process and consider the temporal smoothing approach presented in this chapter as a method for analysing the spectral properties of

nonstationary process. We first begin with a review of nonstationarity.

## Chapter 6

# Nonstationary Processes

Up until this point discussions have centred solely on second order stationary processes. Stationarity provides a convenient mathematical framework in which to work with a clearly defined frequency representation, and consequently is commonly assumed to model many random processes occurring in nature. The reality is that true stationarity is a rare luxury and most naturally occurring signals exhibit some form of nonstationary behaviour. There has therefore been much thought given to modelling nonstationary processes, and in particular the concept of a nonstationary spectrum.

It has been shown in Chapter 2 that the stationarity of the second order properties of a random process allows the autocovariance to be represented as a linear combination of infinitely supported complex exponentials providing a spectral representation of the process that is invariant with time, with the autocovariance sequence  $s_\tau$  and SDF  $S(f)$  forming a Fourier pair. For nonstationary processes the covariance depends explicitly on both time variables rather than just the difference of them as in the stationary case. Consequently for a spectral function to be defined it is necessary to be able to represent the autocovariance as localised oscillations in a joint time-frequency space with a spectral function depending on both time and frequency,  $S(t, f)$  say.

The class of autocovariance functions explode when the stationarity condition is lifted and it is often the case that while a ‘time-frequency’ representation

of a process' second order properties may be achievable, frequency is just a name and has no physical interpretation. [58] offers a useful discussion on the properties we should expect from a time-varying spectral function, and the description of the frequency content of a time-localised region of the covariance function is considered the key property. Other properties that are desirable include  $S(t, f)$  reducing to the standard SDF in the stationary case, along with it being estimable.

Armed with the statistics of a time-scale coherence estimator in the stationary case, we look to extend this work to coherence in a nonstationary setting. However because we need to take care in how a nonstationary spectrum, and ultimately coherence, is defined we first need to take care in choosing the classes of processes that will be considered. In this chapter we have a look at some of the nonstationary models that have appeared in the literature and discuss their time-frequency interpretation. We begin with a look at the harmonizable processes which cover a wide range of possible covariance structures. It will be shown that although mathematically elegant, the derivation of a time-frequency spectral density in this class often lacks physical meaning. Particular focus will be given to the evolutionary processes which are a subclass of the harmonizable processes and one of the few nonstationary models to be accompanied by a meaningful spectral density function and estimation procedure that has been successfully applied in the physical sciences [45]. It will become apparent that the best route to defining nonstationary spectral density functions is by considering natural extensions to SOS representations. We therefore also take a look at the Wold-Cramér nonstationary representation that offers a moving average view of nonstationarity.

It is here that we introduce a change in notation. Previously  $\{Z_t\}$  has been used to represent a complex SOS process which has real part  $\{X_t\}$ . From now on  $\{X_t\}$  will be used to denote a SOS process, and  $\{Z_t\}$  will represent a nonstationary process. All processes will be assumed to be complex-valued from here. We also note that the literature generally considers continuous-time processes, although discrete-time representations are often a straightforward

case. To aid the introduction of ideas here we will look at continuous-time representations. In this chapter and in Chapter 7 discrete-time formulations will also be considered.

## 6.1 Harmonizable Processes

For a complex-valued zero-mean random process  $\{Z(t), t \in \mathbb{R}\}$  the covariance structure at times  $t_0$  and  $t_1$  is defined as

$$\Gamma(t_0, t_1) \equiv \text{cov}\{Z(t_0)Z(t_1)\} = E\{Z(t_0)Z^*(t_1)\}.$$

By definition complex-valued SOS processes have a covariance structure that depends only on the difference  $t_1 - t_0$ . Consider the zero-mean complex-valued process with representation

$$Z(t) = \int_{-\infty}^{\infty} e^{i2\pi ft} d\zeta_Z(f) \quad (6.1)$$

where now, in contrast to SOS processes, the complex-valued stochastic process  $\{\zeta_Z(f)\}$  has the following key property

$$E\{d\zeta_Z(f')d\zeta_Z^*(f)\} = d^2\lambda(f', f) \quad f', f \in \mathbb{R}.$$

Processes that take this more general form of the spectral representation (2.3) are called harmonizable [57]. The function  $\lambda(f', f) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{C}$  is the complex-valued Loève distribution of two frequency variables. The representation (6.1) is ensured by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |d^2\lambda(f', f)| < \infty. \quad (6.2)$$

The covariance structure for  $\{Z(t)\}$  at times  $t_0$  and  $t_1$  is given as

$$\Gamma(t_0, t_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i2\pi(f't_0 - ft_1)} d^2\lambda(f, f') \quad (6.3)$$

and in general is a function of both  $t_0$  and  $t_1$ , rather than just the difference between them as in the stationary case.

Stationary processes are a subclass of the harmonizable processes in the special case where  $\{\zeta_Z(f)\}$  is an orthogonal increment process leading to the Loève distribution function taking the form  $\lambda(f', f) = \mu(f)\delta(f - f')$ , where  $\mu(f)$  takes the role of the stationary process' integrated spectrum.

Assuming the two-dimensional distribution function  $\lambda(f', f)$  is differentiable across both domains we can define the Loève spectral density function of a harmonizable process as  $\Lambda(f, f')df'df = d^2\lambda(f', f)$ . Hanssen et al [46] give a more intuitive approach to the spectral analysis of harmonizable processes that relates to the formulation of the spectral theory for stationary processes presented in Section 2.3.

Instead of two global time variables  $t_0$  and  $t_1$ , we consider the autocovariance sequence at time  $t$  and lag  $\tau$  which for a zero-mean process is defined as

$$s(t, \tau) \equiv \text{cov}\{Z(t), Z(t + \tau)\} = E\{Z(t + \tau)Z^*(t)\}.$$

Using (6.1) we can write

$$\begin{aligned} s(t, \tau) &= E \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i2\pi f't} e^{i2\pi f(t+\tau)} d\zeta_Z^*(f') d\zeta_Z(f) \right\} \\ &= \int_{-\infty}^{\infty} e^{i2\pi\nu t} e^{i2\pi f\tau} \mathcal{S}(\nu, f) d\nu df \end{aligned} \quad (6.4)$$

where

$$\mathcal{S}(\nu, f) d\nu df = E\{d\zeta^*(f - \nu)d\zeta(f)\}.$$

Hanssen et al call  $\mathcal{S}(\nu, f)$  the dual-frequency spectral density function. In (6.4) the frequency variable  $\nu = f - f'$  is interpreted as the frequency offset or *local* frequency relative to the *global* frequency  $f$ . The autocovariance sequence  $s(t, \tau)$  and dual-frequency spectral density function form a 2-

dimensional Fourier pair

$$\mathcal{S}(\nu, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(t, \tau) e^{-i2\pi\nu t} e^{-i2\pi f\tau} d\nu df,$$

i.e. frequency  $f$  is associated with the lag  $\tau$  as in the stationary case, and frequency  $\nu$  is associated with the time coordinate  $t$ . We interpret  $\nu$  as the nonstationary frequency coordinate [46], and  $f$  the stationary frequency coordinate. Any dependence on  $\nu$  is a sign of nonstationarity. A time-frequency spectral density function for the harmonizable process can be derived by invoking the inverse Fourier transform with respect to local frequency  $\nu$  and time  $t$

$$S(t, f)df = df \int_{-\infty}^{\infty} e^{i2\pi\nu t} \mathcal{S}(\nu, f) d\nu. \quad (6.5)$$

It is possible to say [109] that any random process possessing a covariance structure that has the representation (6.3), where (6.2) holds, is harmonizable. This encompasses an extensive class of covariance structures. It is however noticed in [46] that in general  $S(t, f)$  cannot be interpreted as representing the local oscillating behaviour of the covariance structure. We therefore look at a model of nonstationarity that presents a more rigorous treatment of the concept of a time-frequency spectral density.

## 6.2 The Evolutionary Process

This alternative model of nonstationarity was presented in a series of papers by Priestley [78], [79], [80], [83]. Let  $\{X(t), t \in \mathbb{R}\}$  be a complex-valued continuous-time SOS process with zero-mean, that is to say  $\text{cov}\{X(t_0), X(t_1)\}$  is a function of  $(t_0 - t_1)$  only. Referring back to Section 2.3 the covariance structure can be represented as

$$\text{cov}\{X(t_0), X(t_1)\} = \int_{-\infty}^{\infty} e^{i2\pi f(t_0 - t_1)} dS^{(I)}(f)$$

with respect to a measure  $S^{(I)}(f)$ , known as the integrated spectrum. We remind ourselves that the process admits a spectral representation

$$X_t = \int_{-\infty}^{\infty} e^{i2\pi ft} d\zeta_X(f)$$

where  $\{\zeta_X(f)\}$  is a stochastic process with orthogonal increments, and

$$E\{|d\zeta_X(f)|^2\} = dS^{(I)}(f).$$

If  $S^{(I)}(f)$  is differentiable for all  $f$ , the SDF of the process is defined via  $S(f)df = dS^{(I)}(f)$ .

Priestley introduced a logical extension for a nonstationary process. He considered the class of processes for which there exists a family  $\mathcal{F}$  of functions  $\{\phi_t(f)\}$  defined on the  $f$ -axis and indexed by parameter  $t$ , together with a measure  $\mu(f)$  such that for each  $t_0$  and  $t_1$ , the covariance of any such zero-mean process,  $\{Z(t), t \in \mathbb{Z}\}$  say, admits a representation of the form

$$\text{cov}\{Z(t_0), Z(t_1)\} = \int_{-\infty}^{\infty} \phi_{t_0}(f)\phi_{t_1}^*(f)d\mu(f). \quad (6.6)$$

Provided for each  $t \in \mathbb{R}$ ,  $\phi_t(f)$  is square integrable with respect to  $\mu(f)$ , then  $\text{var}\{Z(t)\}$  is finite for all  $t$ . Given (6.6)  $\{Z(t)\}$  will have an autocovariance sequence that is evolving in global time  $t$ , in contrast to stationary covariance sequences that are solely a function of the lag  $\tau$ . This process will admit a representation of the form

$$Z(t) = \int_{-\infty}^{\infty} \phi_t(f)d\zeta(f), \quad t \in \mathbb{R} \quad (6.7)$$

where  $\{\zeta(f)\}$  is an orthogonal increment process and  $E\{|d\zeta(f)|^2\} = d\mu(f)$ . The measure  $\mu(f)$  plays the role of the integrated spectrum  $S^{(I)}(f)$  in the stationary case.

It was shown in Section 2.3 that the SDF,  $S(f)$  say, for a SOS process  $\{X(t)\}$  measures how the variance of the process is distributed across fre-

quency, with the identity

$$\text{var}\{X(t)\} = \int_{-\infty}^{\infty} S(f)df.$$

For the nonstationary representation (6.7) we would want to describe a time-varying spectral density function by the same principle. Assuming the differentiability of  $\mu(f)$ , the fact  $\{\zeta(f)\}$  is an orthogonal increment process means

$$\begin{aligned} \text{var}\{Z(t)\} &= \text{cov}\{Z(t), Z(t)\} \\ &= \int_{-\infty}^{\infty} |\phi_t(f)|^2 d\mu(f) \\ &= \int_{-\infty}^{\infty} S_t(f)df \end{aligned}$$

where

$$S_t(f)df = |\phi_t(f)|^2 d\mu(f). \quad (6.8)$$

This seems to indicate that the power of a process at time  $t$  in a small interval of width  $df$  about  $f$  is given by  $S_t(f)df$ . However we still have no justification for treating  $f$  as something we can physically interpret as frequency. To represent the second (and higher) order structure of a random process in terms of local frequency contributions we need  $\phi_t(f)$  to be oscillating at frequency  $f$ .

### 6.2.1 Oscillatory Processes

Suppose that the function  $\phi_t(f)$ , now considered a function of  $t$ , possesses a Fourier transform for each fixed  $f$  whose modulus has an absolute maximum at frequency  $\nu(f)$ , say, then we may regard  $\phi_t(f)$  as an amplitude-modulated complex sinusoid with frequency  $\nu(f)$ , and hence write  $\phi_t(f)$  in the form

$$\phi_t(f) = A_t(f)e^{i2\pi\nu(f)t}.$$

For  $\phi_t(f)$  to be an oscillating function, the modulating function  $A_t(f)$  must admit the generalised Fourier representation

$$A_t(f) = \int_{-\infty}^{\infty} e^{i2\pi\theta t} dK(\theta, f) \quad (6.9)$$

(where  $dK(\theta, f)$  is the differential with respect to  $\theta$ ) with  $|dK(\theta, f)|$  having an absolute maximum at  $\theta = 0$  for any fixed  $f$ , [109, p. 457]. Priestley [78] defines  $\phi_t(f)$  to be an *oscillatory function* if uniformly in  $f$

$$\arg \max_{\theta} |dK(\theta, f)| = 0.$$

Further to this if  $\{\phi_t(f)\}$  is such that  $\nu(f)$  is a single-valued function of  $f$ , then a suitable variable change in the representation (6.6) can be made, along with a redefining of  $A_t(f)$  and the measure  $\mu(f)$ , to give the representation

$$Z(t) = \int_{-\infty}^{\infty} A_t(f) e^{i2\pi ft} d\zeta(f) \quad t \in \mathbb{R} \quad (6.10)$$

(where  $E\{|d\zeta(f)|^2\} = d\mu(f)$ ). A family  $\mathcal{F}$  of oscillatory functions  $\{A_t(f)e^{i2\pi ft}\}$  is called an *oscillatory family*, and a process  $\{Z(t)\}$  that can be represented in terms of an oscillatory family, as in (6.10), is called an *oscillatory process*.

Using (6.8), the evolutionary spectral density function (ESDF) of the oscillatory process  $\{Z(t)\}$  with respect to the family  $\{A_t(f)e^{i2\pi ft}\}$  is

$$S_t(f) = |A_t(f)|^2 S(f) \quad (6.11)$$

where  $S(f)df = d\mu(f)$ , (assuming  $\mu(f)$  is differentiable).

Yaglom [109] recognises that a zero-mean oscillatory process with representation (6.10) will, using (6.9), have covariance structure

$$\Gamma(t_0, t_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i2\pi[t_0(f+\theta) - t_1(f+\theta')]} dK(\theta, f) dK^*(\theta', f) d\mu(f).$$

Noticing this as being of the form (6.3), provided condition (6.2) is met, i.e.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |dK(\theta, f)| |dK^*(\theta', f)| d\mu(f) < \infty,$$

then the process admits a harmonic representation. We can therefore say that the oscillatory processes are a subclass of the harmonizable processes in which the time-frequency spectral density (6.5) as defined by [46] has a physically meaningful interpretation.

### 6.2.2 Semi-Stationarity

In general the possible choice for the family  $\mathcal{F}$  with which to represent an oscillatory process is not unique, and thus it is useful to have a standardised way of choosing which family provides the best representation for the oscillatory process. Priestley proposed that  $\mathcal{F}$  be chosen such that the amplitude modulation  $A_t(f)$  is least variable. To do so it is first necessary to define a measure for the variability across a family of functions. In his analysis, Priestley uses angular frequency  $\omega = 2\pi f$ , and  $A_t(f)$  admits the generalised Fourier representation

$$A_t(f) = \int_{-\infty}^{\infty} e^{i\omega t} dH(\omega, f)$$

(where  $dH(\omega, f)$  is the differential with respect to  $\omega$ ) with  $|dH(\omega, f)|$  having an absolute maximum at  $\omega = 0$  for any fixed  $f$ .

For each family  $\mathcal{F}$ , Priestley defined the function

$$B_{\mathcal{F}}(\omega) = \int_{-\infty}^{\infty} |\omega| |dH(\omega, f)|.$$

Here we opt to use physical frequency  $\theta$ , and it can be shown that

$$\int_{-\infty}^{\infty} |\omega| |dH(\omega, f)| = 2\pi \int_{-\infty}^{\infty} |\theta| |dK(\theta, f)|.$$

The function  $B_{\mathcal{F}}(f) = 2\pi \int_{-\infty}^{\infty} |\theta| |dK(\theta, f)|$  is a measure of the width, or con-

centration of  $dK(\theta, f)$  about the origin in the  $\theta$  domain (i.e. the Fourier domain with respect to  $t$ ). It is therefore a measure of the rate at which  $A_t(f)$  is changing in time, and consequently is a measure of nonstationarity. If  $A_t(f)$  is slowly-varying,  $B_{\mathcal{F}}(f)$  will be ‘small’, whereas if  $A_t(f)$  is changing quickly  $B_{\mathcal{F}}(f)$  will be ‘large’ [80]. Stationary processes are represented by the family  $\mathcal{F} = \{1 \cdot e^{i2\pi ft}\}$ , and  $B_{\mathcal{F}}(f) = 0$ .

Yaglom [110] states that this train of thought makes more physical sense if the function  $A_t(f)$  is rescaled such that  $A_0(f) = 1$ . Suppose a zero-mean complex-valued oscillatory process  $\{Z(t)\}$  has representation

$$Z(t) = \int_{-\infty}^{\infty} A'_t(f) e^{i2\pi ft} d\zeta'(f),$$

provided  $A'_0(f) \neq 0$ ,  $\{Z(t)\}$  has the equivalent representation

$$Z(t) = \int_{-\infty}^{\infty} A_t(f) e^{i2\pi ft} d\zeta(f),$$

where  $A_t(f) = A'_t(f)/A'_0(f)$ , and  $\{d\zeta(f)\} \equiv \{A'_0(f)d\zeta'(f)\}$  remains an orthogonal increment process. With this rescaling of the modulating function, and using (6.9), we now have the identity

$$A_0(f) = \int_{-\infty}^{\infty} dK(\theta, f) = 1.$$

This means  $dK(\theta, f)$  can to some extent be interpreted, analogously to a PDF, as a spectral density function for the change  $A_t(f)$  undergoes in time.

From [5, p. 159] we have

$$\left| \frac{dA_t(f)}{dt} \right| \leq 2\pi \int_{-\infty}^{\infty} |\theta| |dK(\theta, f)| = B_{\mathcal{F}}(f),$$

and consequently  $1/B_{\mathcal{F}}(f)$  can be very roughly interpreted as the interval over which  $A_t(f)$  is approximately constant in time for fixed  $f$ .

Priestley defines a family of functions  $\mathcal{F}$  to be *semi-stationary* if  $B_{\mathcal{F}}(f)$  is bounded for all  $f$ . For a semi-stationary family  $\mathcal{F}$ , its *characteristic width* is

given by

$$B_{\mathcal{F}} = [\sup_f B_{\mathcal{F}}(f)]^{-1}. \quad (6.12)$$

Thus,  $1/B_{\mathcal{F}}$  measures the rate of change of the fastest varying modulating function and  $B_{\mathcal{F}}$  is an approximate measure of the maximum time interval over which  $A_t(f)$  can be treated as constant and hence  $\mathcal{F}$  can be approximated as the stationary family  $\{e^{i2\pi ft}\}$ .

We are now in a position to define a semi-stationary process. It is said [81, p. 150] that  $\{Z(t)\}$  is a *semi-stationary process* if there exists a semi-stationary family  $\mathcal{F}$  in terms of which  $\{Z(t)\}$  has a representation of the form (6.10). For a particular semi-stationary process  $\{Z(t)\}$ , consider the class  $\mathcal{C}$  of semi-stationary families  $\mathcal{F}$ , for each of which  $\{Z(t)\}$  admits a spectral representation (6.10). We can now define the *characteristic width*  $B_Z$  of the process  $\{Z(t)\}$  by

$$B_Z = \sup_{\mathcal{F} \in \mathcal{C}} \{B_{\mathcal{F}}\}. \quad (6.13)$$

$B_Z$  is the approximate maximum time interval over which a semi-stationary process  $\{Z(t)\}$  can be considered to be stationary [78], [80]. Yaglom [110] notes that this is a very crude approximation to the interval of stationarity, but it will be shown in Section 6.5 that it does play an important role in spectral estimation.

While the mathematical formulation of evolutionary spectra is elegant, it is also rather abstract. The implication of Priestley's work is that  $\mathcal{F}$  should be chosen such that it has maximum characteristic width. Such a representation is termed *natural* [63]. However as Walker notes in his discussions attached to [45], there is no obvious method of calculating such a family.

Mélard considers this problem in [63]. Using very simplistic, finite length nonstationary random processes the complete theory is applied, including a derivation for the natural representation. However even for these over simplistic models the theory is far from trivial and Mélard concedes that a general approach to determining natural representations is probably impossible.

One type of widely studied and applied nonstationary process does however

allow a more straightforward analysis. In an attempt to make some of the ideas here more accessible we take a look at this process, known in the literature as the uniformly modulated process.

### 6.3 Uniformly Modulated Processes

Consider a zero-mean SOS process  $\{X(t), t \in \mathbb{R}\}$  and a purely deterministic real and finite-valued modulating function  $h(t)$ . The process  $\{Z_X(t) \equiv h(t)X(t)\}$  will have autocovariance

$$\text{cov}\{Z_X(t_0), Z_X(t_1)\} = s_X(t_1 - t_0)h(t_0)h(t_1)$$

where  $s_X(\tau)$  is the autocovariance sequence of  $\{X(t)\}$ . Therefore, provided  $h(t)$  is not a constant for all  $t$ ,  $\{Z_X(t)\}$  is a nonstationary process. Such a process is known as a modulated stationary process and this model has been used in such instances as distinguishing earthquake and nuclear blast seismic time series [22], climatology [48], and in structural engineering [93]. Using the spectral representation of SOS process  $\{X(t)\}$ , we have

$$Z_X(t) = \int_{-\infty}^{\infty} h(t)e^{i2\pi ft} d\zeta_X(f)$$

which takes the form of an oscillatory process (6.10). Recalling (6.9) we know that this process will be oscillatory if  $h(t)$  has a generalised Fourier transform  $dK(\theta)$  defined via

$$h(t) = \int_{-\infty}^{\infty} e^{i2\pi t\theta} dK(\theta) \quad (6.14)$$

with the absolute maximum of  $|dK(\theta)|$  occurring at zero. For this to be satisfied it is sufficient that  $h(t)$  is a nonnegative function whose Fourier transform exists [109, p. 459].

Assuming the existence of the SDF of  $\{X(t)\}$  we see from (6.11) that the ESDF of  $\{Z_X(t)\}$  is

$$S_t(f) = h^2(t)S_X(f) \quad (6.15)$$

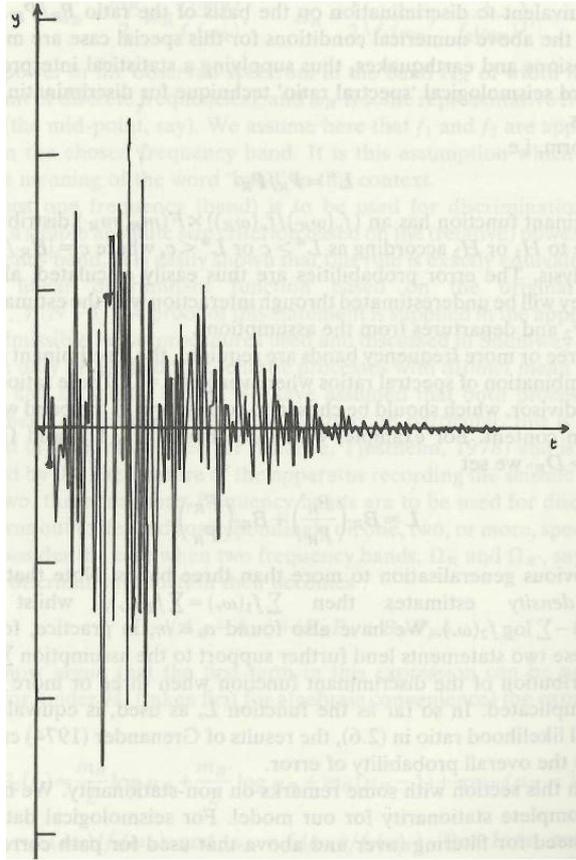


Figure 6.1: Simulated seismic signal using the UMP model. Figure copied from [22].

where  $S_X(f)df = E\{|d\zeta_X(f)|^2\}$ . The power of different frequency components of  $\{X(t)\}$  are all modified in the same way; an oscillatory process of this form is called a uniformly modulated process (UMP) by Priestley [83]. Further to being oscillatory, Priestley notes in [78] that a UMP is always semi-stationary.

### 6.3.1 Example

We now take a simple example of a UMP and demonstrate some of the definitions and terms introduced thus far in this chapter. We look at the example of [79, p. 238]. Consider the UMP  $\{Z(t)\}$  generated with SOS process  $\{X(t)\}$

and modulating function

$$h(t) = e^{-t^2/(2\alpha^2)}.$$

With respect to family  $\mathcal{F}_0 = \{e^{-t^2/(2\alpha^2)}e^{i2\pi ft}\}$ ,  $\{Z(t)\}$  has the ESDF

$$S_t(f) = e^{-t^2/\alpha^2} S_X(f)$$

where  $S_X(f)$  is the SDF of  $\{X(t)\}$ . The modulating function  $h(t)$  is dependent only on time so the generalised Fourier transform  $dK(\theta, f)$  is independent of frequency  $f$  and assuming differentiability we have  $dK(\theta, f) = dK(\theta) = k(\theta)d\theta$  where

$$k(\theta) = \int_{-\infty}^{\infty} e^{-t^2/(2\alpha^2)} e^{-i2\pi\theta t} dt = \alpha\sqrt{2\pi} e^{-2(\pi\alpha\theta)^2}.$$

This also means the function  $B_{\mathcal{F}_0}(f)$  is independent of  $f$ , hence using (6.12) the characteristic width of the family  $\mathcal{F}_0$  is

$$\begin{aligned} B_{\mathcal{F}_0} &= \left[ 2\pi \int_{-\infty}^{\infty} |\theta| |k(\theta)| d\theta \right]^{-1} \\ &= \left[ \alpha(2\pi)^{3/2} \int_{-\infty}^{\infty} |\theta| e^{-2(\pi\alpha\theta)^2} \right]^{-1} \\ &= \alpha\sqrt{\pi/2}. \end{aligned}$$

It is often unclear as to whether there exists other families that represent the nonstationary process, meaning that while an exact value for  $B_Z$ , the characteristic width of the process, is unobtainable, (6.13) tells us that  $B_{\mathcal{F}_0}$  will provide us with a lower bound to  $B_Z$  and hence

$$B_Z \geq \alpha\sqrt{\pi/2}.$$

To have a clearer understanding of the role of  $B_Z$  we need to look at methods for estimating the ESDF. This will be covered in Section 6.5. Here we continue our discussion on nonstationary models.

## 6.4 Wold-Cramér Model of Nonstationarity

We now look at an alternative way of extending a stationary process representation to the nonstationary case. The theory is typically presented in the discrete-time case and so will be considered here. The Wold Decomposition Theorem for stationary processes [108] states that any SOS zero-mean process  $\{X_t\}$  has the representation

$$X_t = \sum_{\tau=0}^{\infty} h_{\tau} \xi_{t-\tau}$$

where  $h_0 = 1$ ,  $\sum_{\tau} |h_{\tau}|^2 < \infty$  and  $\{\xi_t\}$  is a white noise process satisfying  $E\{\xi_j \xi_k^*\} = \sigma_{\xi}^2 \delta_{jk}$  with  $\text{var}\{X_t\} = \sigma_{\xi}^2 \sum_{\tau} |h_{\tau}|^2$ . We now apply the same principle, but with a time-varying filter.

Let  $\{Z_t, t \in \mathbb{Z}\}$  be a purely non-deterministic nonstationary process with zero-mean. It can be shown [18] that there exists a one-sided infinite moving-average representation, or Wold-Cramér decomposition of the process

$$Z_t = \sum_{\tau=0}^{\infty} \tilde{h}_{t,\tau} \tilde{\xi}_{t-\tau}$$

where convergence in the mean is guaranteed. It is further possible, as demonstrated in [64], to normalise the innovation  $\tilde{\xi}_t$  and rescale the time-dependent filter  $\tilde{h}_{t,\tau}$  to give

$$Z_t = \sum_{\tau=0}^{\infty} h_{t,\tau} \xi_{t-\tau} \tag{6.16}$$

where  $\{\xi_t\}$  is now a unit variance white noise process.

There are strong comparisons to be drawn between this representation and the oscillatory processes. With  $\{\xi_t\}$  being itself a SOS process, it is possible to represent it in the usual way as

$$\xi_t = \int_{-1/2}^{1/2} e^{i2\pi ft} d\zeta_{\xi}(f) \tag{6.17}$$

where the process  $\{\zeta_\xi(f)\}$  is a random orthogonal increment process with

$$E\{d\zeta_\xi(f)d\zeta_\xi^*(f')\} = \begin{cases} 1 df & f = f' \\ 0 & f \neq f'. \end{cases}$$

Using (6.16) and (6.17), we can now go on to say that

$$\begin{aligned} Z_t &= \sum_{\tau=0}^{\infty} h_{t,\tau} \xi_{t-\tau} \\ &= \sum_{\tau=0}^{\infty} h_{t,\tau} \int_{-1/2}^{1/2} e^{i2\pi f(t-\tau)} d\zeta_\xi(f) \\ &= \int_{-1/2}^{1/2} \sum_{\tau=0}^{\infty} h_{t,\tau} e^{i2\pi f(t-\tau)} d\zeta_\xi(f) \\ &= \int_{-1/2}^{1/2} H_t(f) e^{i2\pi ft} d\zeta_\xi(f) \end{aligned} \quad (6.18)$$

where  $H_t(f) = \sum_{\tau=0}^{\infty} h_{t,\tau} e^{-i2\pi f\tau}$ . The expression in (6.18) takes the same form as representation (6.10). Due to the innovation process being uncorrelated, and its SDF being one for all frequencies, the Wold-Cramér spectrum  $S_t(f)$  for the process  $\{Z_t\}$  is defined as [64]

$$S_t(f) = |H_t(f)|^2.$$

The advantage of this approach is it leads to an easily accessible, and unique spectrum. However we are unable to assume that  $H_t(f)$  has a generalised Fourier transform peaked at zero and hence in general the family  $\{H_t(f)e^{i2\pi ft}\}$  is not oscillatory. We are therefore unable, without further examination of  $h_{t,\tau}$  and  $H_t(f)$ , to treat the function  $S_t(f)$  as a meaningful frequency representation of the process, although there are obviously representations of the form (6.16) where  $h_{t,\tau}$  and  $H_t(f)$  will satisfy the conditions required for  $\{Z_t\}$  to be oscillatory. For one such example we return to the UMP and look at how it can be cast in a Wold-Cramér formulation.

Consider the discrete-time UMP  $Z_t = g_t X_t$  where  $\{X_t\}$  is a discrete-time

SOS process and  $g_t$  is a deterministic, non-negative and finite-valued modulating function.  $\{X_t\}$  will admit the Wold decomposition

$$X_t = \sum_{\tau=0}^{\infty} a_{\tau} \epsilon_{t-\tau}.$$

with  $a_0 = 1$  and  $\sum_{\tau} |a_{\tau}|^2 < \infty$ . The process  $\{Z_t\}$  is therefore representable in the form

$$Z_t = g_t \sum_{\tau=0}^{\infty} a_{\tau} \epsilon_{t-\tau} = \sum_{\tau=0}^{\infty} g_t a_{\tau} \epsilon_{t-\tau}$$

and hence  $\{Z_t\}$  is in the form of the Wold-Cramér nonstationary decomposition (6.16) where  $h_{t,\tau} = g_t a_{\tau}$ . We already know a process represented in this way has a Wold-Cramér spectrum  $S_t(f) = |H_t(f)|^2$  where  $H_t(f)$  is the Fourier transform with respect to  $\tau$  of the filter  $h_{t,\tau}$ . With  $h_{t,\tau} = g_t a_{\tau}$  we have  $H_t(f) = g_t A(f)$ , where  $A(f) = \sum_{\tau=0}^{\infty} a_{\tau} e^{-i2\pi f\tau}$ . Therefore in the case of the UMP  $H_t(f) = g_t A(f)$  has the representation

$$H_t(f) = A(f)g_t = \int_{-1/2}^{1/2} A(f)G(\theta)e^{i2\pi\theta t}d\theta = \int_{-1/2}^{1/2} dK(\theta, f)$$

where  $dK(\theta, f) = A(f)G(\theta)d\theta$ .  $|dK(\theta, f)|$  is ensured to have an absolute maximum at  $\theta = 0$  because  $|G(\theta)|$  itself has an absolute maximum at  $\theta = 0$  due to  $g_t$  being a non-negative finite valued function. We therefore arrive at the required condition for an oscillatory process via the Wold-Cramér formulation.

## 6.5 Estimating Evolutionary Spectra

We now return to the oscillatory processes, the main focus of this chapter. It has been shown that this class of processes have traditional Fourier based spectral representations that exist on the time-frequency domain and have an intuitive physical meaning. A further key advantage of the oscillatory processes is that Priestley, along with formulating their theoretical framework, also provided a scheme for estimating their spectra. The theory presented is

taken from [78] where it was formulated in continuous-time. A discrete-time formulation will be considered in Chapter 7.

### 6.5.1 Priestley's Method of Evolutionary Spectral Estimation

Consider the zero-mean continuous-time semi-stationary process  $\{Z(t), t \in \mathbb{R}\}$  with representation

$$Z(t) = \int_{-\infty}^{\infty} A_t(f) e^{i2\pi ft} d\zeta(f)$$

with respect to the natural family  $\{A_t(f) e^{i2\pi ft}\}$ , and we are able to define the ESDF  $S_t(f) = |A_t(f)|^2 S(f)$  where  $S(f) df = E\{|d\zeta(f)|^2\}$ . Let  $\{g(u)\}$  be a filter which we assume to have the standard attribute of being unit  $L^2$  norm. We define the process,  $\{Y_t(f_0)\}$  say, for a fixed frequency  $f_0$  as

$$Y_t(f_0) = \int_{-\infty}^{\infty} g(u) Z(t-u) e^{-i2\pi f_0(t-u)} du.$$

This can be re-written as

$$Y_t(f_0) = \int_{-\infty}^{\infty} \Gamma_{t, f+f_0}(f) A_t(f+f_0) e^{i2\pi ft} d\zeta^*(f+f_0)$$

where  $\Gamma_{t, \lambda}(\theta)$  is the *generalised transfer function* of filter  $g(u)$  with respect to the family  $\mathcal{F} = \{A_t(f) e^{i2\pi ft}\}$  defined as

$$\Gamma_{t, \lambda}(\theta) = \int_{-\infty}^{\infty} g(u) \{A_{t-u}(\lambda) / A_t(\lambda)\} e^{-i2\pi \theta u} du.$$

Due to the orthogonal increments of  $\{\zeta(f)\}$ , it follows that

$$E\{|Y_t(f_0)|^2\} = \int_{-\infty}^{\infty} |\Gamma_{t, f+f_0}(f)|^2 |A_t(f+f_0)|^2 d\mu(f+f_0). \quad (6.19)$$

We now define  $B_g$  the characteristic width of the filter  $g(u)$  as

$$B_g = \int_{-\infty}^{\infty} |u| |g(u)| du. \quad (6.20)$$

Now suppose  $g(u)$  is chosen such that  $B_g < \epsilon B_Z$  for some fixed  $\epsilon > 0$ , then

$$\Gamma_{t,f+f_0}(f) = \Gamma(f) + r(t, f_0, f)$$

where  $|r(t, f_0, f)| < \epsilon |A_t(f + f_0)|$  [78, p. 214]. It is shown that

$$E\{|Y_t(f_0)|^2\} = \int_{-\infty}^{\infty} |\Gamma(f)|^2 |A_t(f + f_0)|^2 d\mu(f + f_0) + \mathcal{O}(\epsilon)$$

where  $\mathcal{O}(\epsilon)$  denotes a term that can be made arbitrarily small by choosing  $B_g$  to be sufficiently small relative to  $B_Z$ . The exact bounds on this term are evaluated at [78, p. 215]. Priestley therefore concludes that to  $\mathcal{O}(\epsilon)$

$$E\{|Y_t(f_0)|^2\} = \int_{-\infty}^{\infty} |\Gamma(f)|^2 S_t(f + f_0) df \quad (6.21)$$

remembering that the exact value is given by

$$E\{|Y_t(f_0)|^2\} = \int_{-\infty}^{\infty} |\Gamma_{t,f+f_0}(f)|^2 S_t(f + f_0) df.$$

From (6.21)

$$E\{|Y_t(f_0)|^2\} = S_{\Gamma,t}(f_0) + \mathcal{O}(\epsilon)$$

where

$$S_{\Gamma,t}(f_0) = \int_{-\infty}^{\infty} |\Gamma(f)|^2 S_t(f + f_0) df.$$

Given a realisation of  $\{Z(t)\}$  observed for  $0 \leq t \leq T$ , Priestley proposed the spectral estimator  $|U_t(f_0)|^2$  at time  $t$  and frequency  $f_0$  where

$$U_t(f_0) = \int_{t-T}^t g(u) Z(t-u) e^{-i2\pi f_0(t-u)} du.$$

By designing  $\{g(u)\}$  such that  $B_g \ll B_Z$ , and with the assumption that  $B_Z \ll T$ , the limits in the integral can be replaced with  $(-\infty, \infty)$  and we obtain

$$E\{|U_t(f_0)|^2\} = \int_{-\infty}^{\infty} |\Gamma(f)|^2 S_t(f + f_0) df + \mathcal{O}(B_g/B_Z).$$

For  $|U_t(f_0)|^2$  to be an unbiased estimator of  $S_t(f_0)$  we require two conditions to be satisfied. Firstly,  $B_g \ll B_Z$ , and secondly we require  $|\Gamma(f)|^2$  to be a pseudo-delta function with respect to  $S_t(f)$  for all frequencies and time. (A function  $v(\cdot)$  is *pseudo-delta* with respect to  $w(\cdot)$  if  $\int_{-\infty}^{\infty} v(u)w(u+t)du \approx w(t) \int_{-\infty}^{\infty} v(u)du$ ). Assuming these two conditions are satisfied

$$\int_{-\infty}^{\infty} |\Gamma(f)|^2 S_t(f+f_0)df = S_t(f_0) \int_{-\infty}^{\infty} |\Gamma(f)|^2 df = S_t(f_0). \quad (6.22)$$

We therefore see that there is a natural trade off that needs to be confronted. The resolution in time is determined by the ratio  $B_g/B_Z$ . By letting  $\{g(u)\}$  tend to a delta-function we obtain perfect time resolution but with the consequence that the bandwidth of the function  $|\Gamma(f)|^2$  becomes infinite, losing all frequency resolution. However if we attempt to retain perfect frequency resolution by having  $|\Gamma(f)|^2$  delta-like then the term  $\mathcal{O}(B_g/B_Z)$  becomes significant and time-resolution is lost. A full analysis of how to quantify each of these errors and assess the suitability of a particular filter with respect to an oscillatory function is given in [79].

Provided the filter  $\{g(u)\}$  is chosen such that its characteristic width is small with respect to  $B_Z$ , and we have  $S_t(f)$  being smooth compared to  $|\Gamma(f)|^2$ , then we may write

$$E\{|U_t(f_0)|^2\} \approx S_t(f_0).$$

### 6.5.2 Smoothing Procedure

It has been shown via the work of Priestley that for a suitable  $\{g(u)\}$  such that  $B_g \ll B_Z$ , and with the assumption that the bandwidth of  $S_t(f)$  at every time  $t$  is substantially larger than  $|\Gamma(f)|^2$ , then  $|U_t(f_0)|^2$  is an approximately unbiased estimator of  $S_t(f_0)$ . However to reduce sample variance and increase the accuracy for the estimator Priestley proposed that one should smooth with a weighted average in the time domain at the sacrifice of a loss of time resolution. Therefore with a suitable weighting function (positive, square integrable,

integrates to one, and vanishes outside an interval) that we denote as  $W_{T'}(t)$  with width parameter  $T'$ , Priestley defines

$$V_t(f_0) = \int_{-\infty}^{\infty} W_{T'}(u) |U_{t-u}(f_0)|^2 du.$$

We now have

$$E\{V_t(f_0)\} = \int_{-\infty}^{\infty} \bar{S}_t(f + f_0) |\Gamma(f)|^2 df$$

where

$$\bar{S}_t(f + f_0) = \int_{-\infty}^{\infty} W_{T'}(u) S_{t-u}(f + f_0) du,$$

and as such if  $\bar{S}_t(f)$  is smooth in comparison to  $|\Gamma(f)|^2$  then

$$E\{V_t(f_0)\} \approx \bar{S}_t(f).$$

[78] gives an approximate expression for the variance of the smoothed estimator demonstrating the improvement it makes over the non-smoothed estimator.

### 6.5.3 Wavelets

The wavelet transform of signal  $X(t)$  at a scale  $|a| > 0$  and time  $b \in \mathbb{R}$  is given by

$$W(a, b; X, \psi) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} X(u) \psi^* \left( \frac{u - b}{a} \right) du.$$

$|W(a, b; X, \psi)|^2$ , known as the wavelet spectrum, represents the energy of a signal that can be attributed to variations at a particular spatial scale, for a given time  $b$ . In [82] and [93], utilising the continuous wavelet transform as a method for estimating the ESDF is considered for the discrete set of binary dilations  $\{a_j = 2^j, j \in \mathbb{Z}\}$ . While both sets of authors state that the wavelet spectrum provides an estimator for the spectral density, there is little consideration given to the assumptions required to provide validity to the statement. Here we use Priestley's non-wavelet work on spectral estimators to present a more natural link between the ESDF and the wavelet spectrum,

and we present a rigorous evaluation of the assumptions required.

Consider the Morlet wavelet [9, p. 139] of general form

$$\psi(t) = g(t)e^{i2\pi\hat{f}t}$$

where  $g(t)$  is a unit norm Gaussian function centred at zero with scale parameter (or standard deviation)  $d$  (that will be suppressed in the notation), and  $\hat{f}$  is the centre frequency of the wavelet. The CWT at scale  $|a| > 0$  and translation  $b \in \mathbb{R}$  for a zero-mean semi-stationary process  $\{Z(t)\}$ , represented by semi-stationary family  $\mathcal{F} = \{A_t(f)e^{i2\pi ft}\}$ , can be expressed as

$$W(a, b; Z, \psi) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} Z(u)g((u-b)/a)e^{-i2\pi\hat{f}(u-b)/a} du.$$

By making a simple substitution this becomes

$$W(a, b; Z, \psi) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} Z(u+b)g(u/a)e^{-i2\pi\hat{f}u/a} du,$$

and by recognising that  $g(u)$  is symmetric, a reverse of the time variable  $u$  gives

$$W(a, b; Z, \psi) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} Z(b-u)g(u/a)e^{i2\pi\hat{f}u/a} du.$$

If we let  $g_a(u) = |a|^{-1/2}g(u/a)$  and  $f_a = \hat{f}/a$ , further manipulation gives

$$W(a, b; Z, \psi) = e^{i2\pi f_a b} \int_{-\infty}^{\infty} g_a(u)Z(b-u)e^{i2\pi f_a u} du$$

which aside from the phase shifting pre-multiplying factor, takes the exact same form as Priestley's continuous-time formulation of the filtered process  $Y_t(f_0)$ , where  $f_a$  takes the role of  $f_0$ . Comparing with (6.19) we can make the following conclusions

$$E\{|W(a, b; Z, \psi)|^2\} = \int_{-\infty}^{\infty} |G_{a,b,f+f_a}(f)|^2 |A_b(f+f_0)|^2 d\mu(f+f_a)$$

where  $G_{a,t,f+f_a}(f)$  is the generalised transfer function of  $g_a(u)$  with respect to the family  $\mathcal{F}$ . Now continuing on from Priestley's work we can say provided  $B_{g_a} \leq \epsilon B_Z$  we have

$$E\{|W(a, b; Z, \psi)|^2\} = \int_{-\infty}^{\infty} |G_a(f)|^2 S_b(f + f_a) df + \mathcal{O}(\epsilon) \quad (6.23)$$

where  $G_a(f)$  is the transfer function (Fourier transform) of  $g_a(u)$ .

In Chapter 5 we considered the Morlet wavelet  $g(u) \exp(i2\pi t)$

$$g(u) = d^{-1/2} \pi^{-1/4} e^{-t^2/(2d^2)},$$

with  $1 \leq d \leq 10$ . With regards to the above formulation we can make the following calculations. Firstly

$$\begin{aligned} B_{g_a} &= \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} |u| |g(u/a)| du \\ &= |a|^{1/2} \int_{-\infty}^{\infty} |a| |u| |g(u)| du \\ &= |a|^{3/2} B_g. \end{aligned}$$

Using the following result

$$\begin{aligned} B_g &= \int_{-\infty}^{\infty} |u| |g(u)| du \\ &= 2 \int_0^{\infty} u g(u) du \\ &= -2d^{3/2} \pi^{-1/4} \int_0^{\infty} \frac{d}{du} e^{-u^2/(2d^2)} du \\ &= 2d^{3/2} \pi^{-1/4}, \end{aligned}$$

gives  $B_{g_a} = 2(|a|d)^{3/2} \pi^{-1/4}$ . Further to this, via the properties of the Fourier transform, we have  $G_a(f) = \sqrt{|a|} G(af)$ , where  $G(\cdot)$  is the Fourier transform

of  $g(\cdot)$ . We therefore conclude that

$$E\{|W(a, b; Z, \psi)|^2\} \approx \int_{-\infty}^{\infty} |a||G(af)|^2 S_b\left(f + \frac{1}{a}\right) df \quad (6.24)$$

provided  $2(|a|d)^{3/2}\pi^{-1/4} \ll B_Z$ . Assuming  $|a||G(af)|^2$  is a pseudo-delta function with respect to  $S_t(f)$ ;

$$E\{|W(a, b; Z, \psi)|^2\} \approx S_b(1/a). \quad (6.25)$$

With Priestley's estimator the time and frequency resolution remain constant across the time-frequency plane, the wavelet method differs in the fact that the time and frequency resolutions are changing with the analysis scale  $a$ . A trade-off occurs as we alter  $a$  such that for large  $|a|$  the width  $B_{g_a}$  increases and hence so do the errors in the approximation (6.24). However as we increase  $|a|$  the function  $|a||G(af)|^2$  tightens and becomes more delta-like meaning the approximation in (6.25) becomes more just.

### Smoothing procedure

As stated, Priestley used a temporal smoothing procedure to reduce the sample variance in this evolutionary spectral estimator. It is the same reason that the wavelet spectrum  $|W(a, b; Z, \psi)|^2$  was smoothed in time — as presented in Chapter 5. Specifically, a simple non-weighted smoothing window is used where the width of the smoothing window is proportional to  $|a|$  to ensure degrees of freedom in the estimator remain constant. Putting this into the Priestley formulation we have

$$\mathcal{S}\{|W(a, b; Z, \psi)|^2\} = \int_{-\infty}^{\infty} W_{T'(a)}(u) |W(a, u - b, Z, \psi)|^2 du$$

where

$$W_{T'(a)}(u) = \begin{cases} 1/T'(a) & |u| < T'(a)/2 \\ 0 & \text{otherwise,} \end{cases}$$

and as such for  $B_{g_a} \ll B_Z$  we have

$$E\{\mathcal{S}\{|W(a, b; Z, \psi)|^2\}\} = \int_{-\infty}^{\infty} \bar{S}_b\left(f + \frac{1}{a}\right) |a| |G(af)|^2 df$$

where

$$\bar{S}_b(f) = (1/T'(a)) \int_{-T'(a)/2}^{T'(a)/2} S_{b-u}(f) du.$$

Making the assumption  $\bar{S}_t(f)$  is approximately flat compared to  $|a| |G(af)|^2$  (which becomes more appropriate as  $a$  increases) then we conclude

$$E\{\mathcal{S}\{|W(a, b; Z, \psi)|^2\}\} \approx \bar{S}_t(f_a).$$

## 6.6 Spectral Estimation Examples

We now look at applying the discrete-time temporally smoothed Morlet wavelet spectrum, as presented in Chapter 5, to some example nonstationary processes. Thus far only the continuous-time formulation of the temporally smoothed Morlet wavelet spectrum has been examined for the spectral estimation of a continuous-time evolutionary processes. A discrete-time formulation will be considered in detail in Chapter 7 in conjunction with a multivariate framework. The following examples are for motivational purposes.

### 6.6.1 Uniformly Modulated Process

We first look at a UMP with a Gaussian modulating function as discussed in Section 6.3. Specifically, consider the process  $\{Z_t = g_t X_t, t \in \mathbb{Z}\}$  where  $\{X_t\}$  is the CAR(1) process described in Section 4.6.1 and used in Chapters 4 and 5 for simulation purposes, and deterministic function  $\{g_t, t \in \mathbb{Z}\}$  takes the form

$$g_t = e^{-(t-512)^2/(2\alpha^2)}$$

where  $\alpha = 1024/6$  (i.e. for a 1024 point realisation of the signal starting at  $t = 0$ , the signal effectively grows from, and decays to zero). For the error

terms in (6.23) to be ignored we require  $2(|a|d)^{3/2}\pi^{-1/4} \ll B_Z$ . From the analysis of Gaussian modulating functions in Section 6.3.1,  $B_Z \geq 214$ , giving the condition  $|a| \ll 27$  for the wavelet spectral estimation procedure to be valid.

In Figure 6.2 are eight plots that demonstrate the use of the non-smoothed and smoothed Morlet wavelet spectrum in estimating the ESDF of the UMP. By comparing Figure 6.2(b) and 6.2(d) it seems that for the range of scales  $a$ , while the non-smoothed wavelet spectrum  $|W(a, b; \psi, Z)|^2$  is a respectable estimator for the ESDF, there is some discrepancy between the mean (taken over 1000 simulations) and the true spectrum. As Figure 6.2(c) also demonstrates, without a smoothing operation the spectral estimator is poor with large variance across the time-scale plane. Smoothing the wavelet spectrum, as in Figures 6.2(e) and 6.2(g) improves the individual spectral estimates by reducing variance, however we now see that the mean spectral estimators in Figures 6.2(f) and 6.2(h) are distorted representations of the true ESDF.

### 6.6.2 Wold-Cramér Nonstationary Process

We now progress from the UMP to a more complex form of nonstationary process. We first introduce the continuous-time version of the Wold-Cramér representation (6.16);

$$Z(t) = \int_{-\infty}^{\infty} h(t, \tau)\nu(t - \tau)d\tau \quad (6.26)$$

where  $h(t, \tau)$  is a time-varying filter and  $\{\nu(t)\}$  is a zero-mean continuous-time white noise process with unit variance. This has the equivalent representation (analogously to (6.18))

$$Z(t) = \int_{-\infty}^{\infty} H_t(f)e^{i2\pi ft}d\zeta_\nu(f).$$

To apply the wavelet spectrum to the Wold-Cramér evolutionary model we look at a practical engineering example. Wang and Fish [104] [105] use the

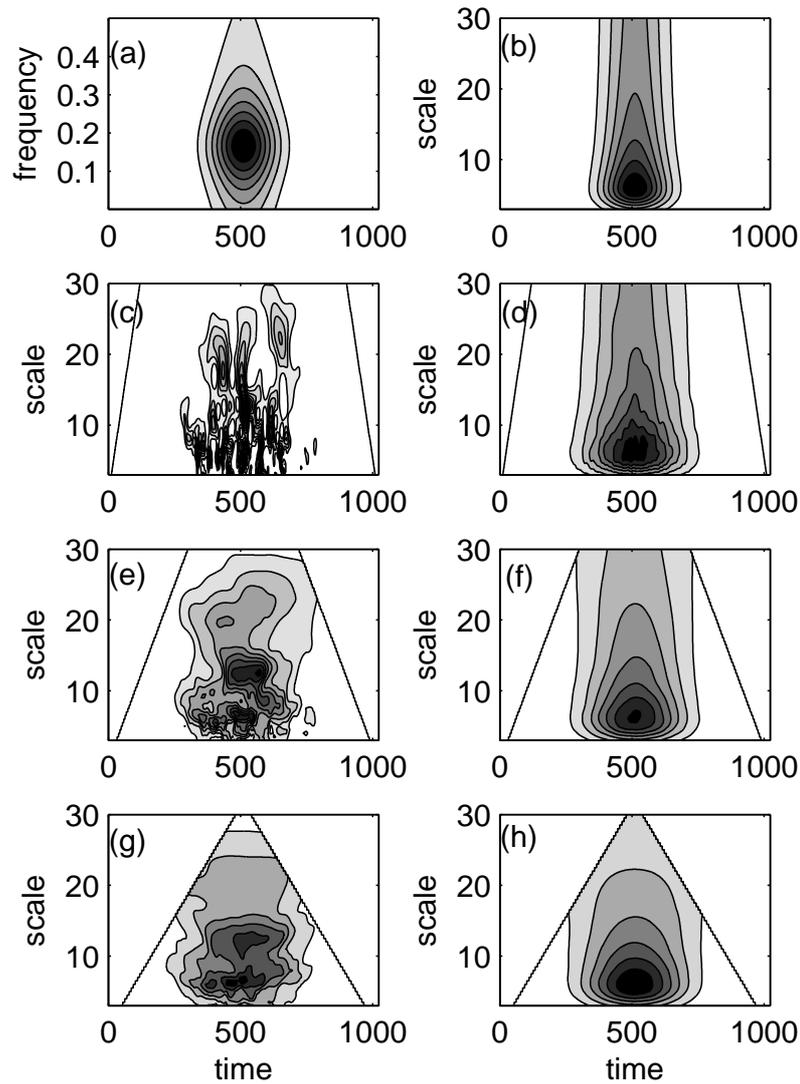


Figure 6.2: Spectral analysis for the UMP. (a) shows the evolutionary spectral density of the UMP in time-frequency space (for the positive half of the spectrum only). (b) shows the ESDF in the Morlet wavelet's time-scale space (with the mapping  $a = 1/f$ ). (c) is a single realisation of the non-smoothed wavelet spectrum in time-scale space. (d) is the mean non-smoothed wavelet spectrum. (e) and (g) are single realisations of the smoothed wavelet spectrum with smoothing parameters  $\kappa_0 = 6$  and  $\kappa_0 = 12$  respectively. (f) and (h) are the corresponding mean smoothed wavelet spectra.

continuous-time Wold-Cramér nonstationary framework to model arterial blood flow. The time-varying filter in (6.26) takes the form

$$h(\tau, t) = A \sqrt{\frac{p(t)}{\sigma_t(t)}} \exp\left\{-\frac{\tau^2}{2\sigma_t^2(t)}\right\} \exp[2\pi i f_m(t)\tau].$$

Using the discrete-time theory we can define the continuous-time Wold-Cramér spectrum in the usual way as the magnitude squared of the Fourier transform of  $h(\tau, t)$  (with respect to  $\tau$ ), giving [69]

$$S_t(f) = \frac{p(t)}{[2\pi]^{1/2}\sigma_f(t)} \exp\left\{-\frac{[f - f_m(t)]^2}{2\sigma_f^2(t)}\right\} \quad (6.27)$$

where  $\sigma_f(t) = [2^{3/2}\sigma_t(t)]^{-1}$ . In these expressions  $p(t)$ ,  $f_m(t)$  and  $\sigma_f(t)$  describe how the mean frequency, bandwidth and Doppler signal power, respectively, vary with time. In this specific case we have that

$$\sigma_f(t) = \alpha + \zeta \exp[-(t - \xi)^2/(2\delta^2)] \quad (6.28)$$

where  $\alpha = 100\text{Hz}$ ,  $\zeta = 200\text{Hz}$ ,  $\xi = 0.1325\text{s}$  and  $p(t) = \sigma_f(t)$ . We analyse 4330 point realisations of this process simulated with a sampling interval of 0.15 ms. For these conditions, the time-dependent evolutionary spectra as defined by (6.27) is plotted in Figure 6.3(a). Using the scale-frequency mapping  $a = 1/f$  it is possible to represent the Wold-Cramér spectrum in a Morlet wavelet time-scale space as in Figure 6.3(b). Figure 6.3(c) shows a single realisation of the non-smoothed wavelet spectrum, with Figure 6.3(d) showing the mean non-smoothed wavelet spectrum over 1000 simulations. It is clear to see from the plot that the non-smoothed wavelet spectrum has a mean that matches the true Wold-Cramér spectrum, however variance across the time-scale plane is high. By smoothing the wavelet spectrum (with parameter  $\kappa_0 = 12$ ), we are able to produce an estimator of the Wold-Cramér spectrum that has much less variation across the time-scale plane — Figure 6.3(e). Smoothing gives an improved estimator from a single realisation, although the mean smoothed

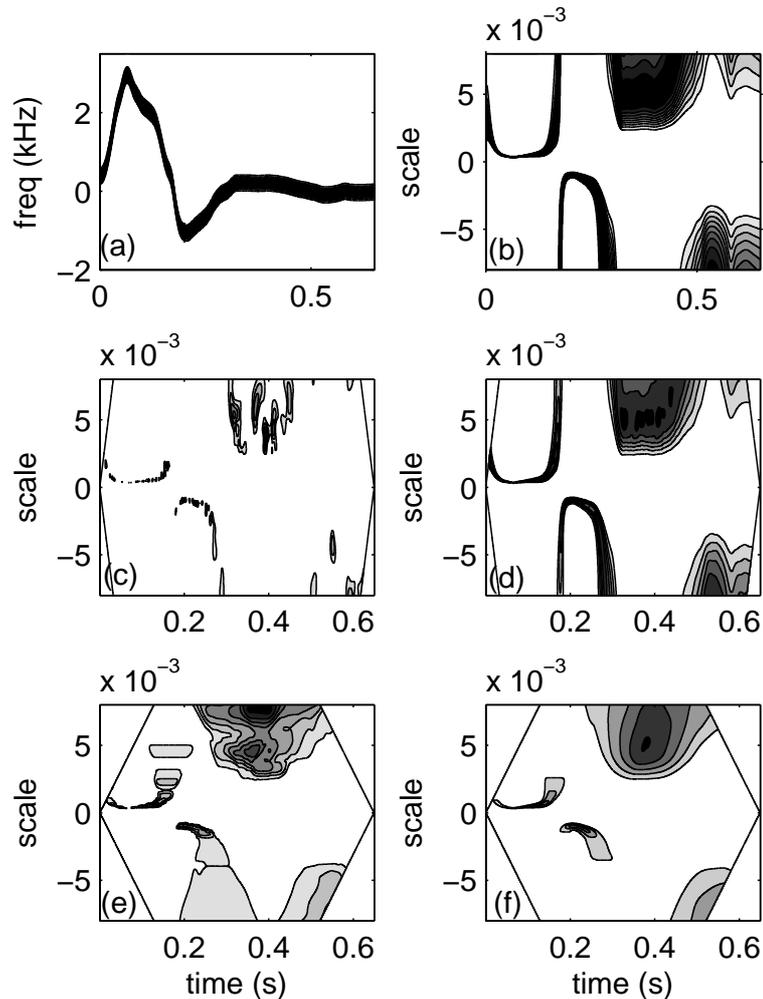


Figure 6.3: Wold-Cramér spectra for the arterial blood flow process. (a) shows the Wold-Cramér spectrum in time-frequency space. (b) shows the Wold-Cramér spectrum in the Morlet wavelet's time-scale space (with the transform  $a = 1/f$ ). (c) is a single realisation of the non-smoothed wavelet spectrum in time-scale space. (d) is the mean non-smoothed wavelet spectrum. (e) is a single realisation of the smoothed wavelet spectrum with smoothing parameter  $\kappa_0 = 12$ . (f) is the corresponding mean smoothed wavelet spectra.

wavelet spectrum in Figure 6.3(f) is noticeably different to the true Wold-Cramér spectrum (i.e. smoothing bias has been introduced).

The fact that the expected value of the wavelet spectrum is such a close match to the true Wold-Cramér spectrum indicates two important things.

Firstly we can say, without further analysis on the time-varying filter  $h(\tau, t)$ , that this process appears to be oscillatory. The wavelet spectrum gives us a direct link to true frequency and with this close match it appears as though frequency can be interpreted as a meaningful quantity. Secondly the characteristic width of this nonstationary process is significantly larger than the characteristic width of the Morlet wavelet.

## 6.7 Other models for Nonstationarity

We now briefly outline some further models of nonstationarity that have been developed in the literature and discuss their suitability in modelling naturally occurring nonstationary time series.

### 6.7.1 Locally Stationary Processes

For parametric and nonparametric analysis of stationary random processes it is often necessary to take an asymptotic approach because estimating the statistical characteristics of a time series from a finite set of observations has obvious short-comings. For instance, suppose the SOS process  $\{X_t\}$  is observed at times  $0, \dots, T - 1$ , say, statistical properties, including estimators for the spectral density, are best assessed in the limit of  $T \rightarrow \infty$ . In the nonstationary case such an asymptotic approach is often meaningless as extending time into the future provides no information on the statistics of the present. For example with Priestley's evolutionary model, to estimate the spectral density at a particular time point,  $t_0$  say, we require sampled data points to be located in close proximity to  $t_0$ , dictated by the characteristic width of the process. However this restriction on the number of data points leads to the recurring problem of leakage, and extending the number of data points is counterproductive to analysing a localised portion of the process. For an asymptotic theory of nonstationary processes we turn to the work of Dahlhaus [19], [20], [21].

Dahlhaus begins with the motivating example of the nonstationary process

$$Z_t = a(t)Z_{t-1} + \epsilon_t \quad (6.29)$$

where  $\{\epsilon_t\}$  is a Gaussian white noise process. By observing  $\{Z_t\}$  at times  $t = 0, \dots, T-1$  we have information about  $a(t)$  on the *grid*  $[0, \dots, T-1]$ . We now restrict ourselves to inference on the time interval  $[0, T-1]$ . Dahlhaus proposed that the asymptotic analysis considers an increase in  $T$  to be an observation of  $a(t)$  on a finer grid, but on the same interval, as supposed to extending time into the future. That is, we observe the process

$$Z_{t,T} = a\left(\frac{t}{T}\right) Z_{t-1,T} + \epsilon_t, \quad \text{for } t = 0, \dots, T-1$$

where  $a$  is now rescaled to the time interval  $[0, 1]$ . An increase in  $T$  increases the number of observations we make of  $a : [0, 1] \rightarrow \mathbb{R}$ . Dahlhaus extended this to the evolutionary model by considering a rescaling of the amplitude modulating function with processes of the form

$$Z_{t,T} = \int_{-1/2}^{1/2} A\left(\frac{t}{T}, f\right) e^{i2\pi ft} d\zeta(f).$$

The motivating example (6.29) fails to exactly have this representation, so Dahlhaus gave the more general definition that a sequence of zero-mean random variables  $\{Z_{t,T}, t = 0, \dots, T-1\}$  is *locally stationary* with transfer function  $A^0$  if there exists a representation

$$Z_{t,T} = \int_{-1/2}^{1/2} A_{t,T}^0(f) e^{i2\pi ft} d\zeta(f)$$

where  $\{\zeta(f)\}$  is an orthogonal increment process and there exists a periodic function (unit period)  $A : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}$  such that for some  $K > 0$

$$\sup_{t,f} \left| A_{t,T}^0(f) - A\left(\frac{t}{T}, f\right) \right| \leq KT^{-1}$$

for all  $T$ ;  $A(t, f)$  is assumed continuous in  $t$ .

In contrast to Priestley's formulation, the parameter  $T$  now plays a key role in defining the process, as indicated by its presence in the notation. An increase in  $T$  should not be thought of as an observation of a continuous-time process on an increasingly fine grid, but instead as an abstract setting by which we increase the amount of data we collect to study the local structure of the nonstationary process.

We will not go into anymore detail on the local stationary process here. They have been presented to raise awareness of their existence. Dahlhaus himself notes in [19] that it is meant as a theoretical exercise in defining a meaningful asymptotic theory to nonstationarity and applications to the physical sciences are extremely limited. He states that Priestley's formulation with accompanying estimation procedure is designed for physically meaningful representations and is much more suited to applications in the physical sciences.

### 6.7.2 Wavelet Processes

The concept of the wavelet nonstationary process was developed by Nason et al [67]. The Cramér representation of a stationary process (2.3) and Priestley's evolutionary representation (6.10) use the complex sinusoids,  $\{e^{i2\pi ft}\}$ , as their basis of decomposition. The wavelet processes instead use a set of locally supported wavelets  $\{\psi_{jk}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$  for decomposition where

$$\psi_{jk}(t) = 2^{j/2}\psi(2^j(t - k))$$

for a sufficiently well concentrated mother wavelet  $\psi(t)$ . A zero-mean nonstationary random process can be constructed as

$$Z_t = \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{\infty} \eta_{jk}\psi_{jk}(t)$$

where  $\eta_{jk}$  is a random coefficient with  $E\{\eta_{jk}\} = 0$  and  $\text{cov}\{\eta_{jk}\eta_{lm}\} = W_{jk}\delta_{jl}\delta_{km}$ . This allows for a genuine and analytical time-scale spectrum  $W_{jk}$  to be defined. The time-scale wavelet spectrum is the power of the process associated with oscillations over a scale  $2^{-j}$  and time index  $k$ . Further to this, an asymptotic framework akin to the work of Dahlhaus was also constructed by Nason et al. These processes, called *locally stationary wavelet processes*, use the same double indexing  $t, T$  to construct an abstract setting for asymptotic statistical analysis. The locally stationary wavelet processes have been extended to a multivariate setting, and the concept of wavelet coherence, which is now well defined in time-scale space, has been analysed in [50]. This includes an asymptotic statistical treatment of the statistics of a wavelet coherence estimator.

The wavelet process model provides a logical framework for defining nonstationarity and offers a straightforward route to simulating nonstationary processes that have a well defined wavelet spectrum. This can again be seen however as a theoretical nicety. The use of wavelet processes in modelling naturally occurring time series is troublesome and examples are rare. The complex exponentials together with the concept of frequency form a more natural and physically meaningful basis in which to work and consequently have more immediate and obvious applications for modelling naturally occurring nonstationary processes, for example the uniformly modulated processes.

## Chapter 7

# Coherence of Evolutionary Processes

In the previous chapter we introduced the evolutionary processes as formulated by Priestley and showed that through careful considerations it is possible to derive a meaningful time-varying spectral density function, the ESDF, that is estimable with both Priestley's spectral estimator and the temporally smoothed Morlet wavelet spectrum. We have stated that coherence for a pair of stationary processes is a measure of correlation in frequency space. The ordinary coherence for a pair of jointly SOS processes  $\{X_{1,t}\}$  and  $\{X_{2,t}\}$  is defined as

$$\gamma^2(f) = \frac{|S_{12}(f)|^2}{S_{11}(f)S_{22}(f)}$$

where  $S_{11}(f)$  and  $S_{22}(f)$  are the spectral density functions and  $S_{12}(f)$  is the cross-spectral density function. In the final two chapters we explore the concept of coherence for nonstationary processes. Provided that for a pair of nonstationary processes  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  with meaningful spectral density functions  $S_{11}(t, f)$  and  $S_{22}(t, f)$ , we can also define a meaningful cross-spectral density function  $S_{12}(t, f)$ , then it seems logical that we can define the coher-

ence in an analogous way

$$\gamma^2(t, f) = \frac{|S_{12}(t, f)|^2}{S_{11}(t, f)S_{22}(t, f)} \quad (7.1)$$

that measures the oscillatory correlations at a particular point in time. In this chapter we turn our attention to the notion of coherence in Priestley's oscillatory processes, and specifically to estimating coherence for these processes with Priestley spectral estimation approach and the TWCOH estimator.

## 7.1 Multivariate Evolutionary Processes

Priestley and Tong presented the bivariate framework of the oscillatory processes in [83]. Extension to the  $p > 2$  case is straightforward and consequently not considered here.

Consider the continuous bivariate process  $\{\mathbf{Z}_t = [Z_1(t), Z_2(t)]^T, t \in \mathbb{R}\}$  for which we can write the component processes as

$$Z_l(t) = \int_{-\infty}^{\infty} A_{l,t}(f) e^{i2\pi ft} d\zeta_l(f) \quad l = 1, 2$$

where the orthogonal increment processes  $\{\zeta_l(f)\}$  follow the standard rules for  $l = 1, 2$

$$\begin{aligned} E\{d\zeta_l(f) d\zeta_l^*(f')\} &= 0 \quad f \neq f', \\ E\{|d\zeta_l(f)|^2\} &= d\mu_l(f), \end{aligned}$$

and for  $l, k = 1, 2, l \neq k$

$$\begin{aligned} E\{d\zeta_l(f) d\zeta_k^*(f')\} &= 0 \quad f \neq f', \\ E\{d\zeta_l(f) d\zeta_k^*(f)\} &= d\mu_{lk}(f). \end{aligned}$$

We now let  $\{\mathcal{F}_1, \mathcal{F}_2\}$  denote the vector family of oscillatory functions  $\{\phi_{1,t}(f) = A_{1,t}(f)e^{i2\pi ft}, \phi_{2,t}(f) = A_{2,t}(f)e^{i2\pi ft}\}$  with respect to which we can

represent the bivariate process  $\{[Z_1(t), Z_2(t)]^T, t \in \mathbb{R}\}$ . [83] defines the evolutionary cross-spectral density function (ECSDF). Provided the measure  $\mu_{12}(f)$  is differentiable, then with respect to the families  $\mathcal{F}_1$  and  $\mathcal{F}_2$  the ECSDF  $S_{12,t}(f)$  of the bivariate oscillatory process  $\{\mathbf{Z}(t)\}$  is given as

$$S_{12,t}(f)df = A_{1,t}(f)A_{2,t}^*(f)d\mu_{12}(f),$$

or alternatively is defined as

$$S_{12,t}(f) = A_{1,t}(f)A_{2,t}^*(f)S_{12}(f)$$

where  $S_{12}(f)$  is the CSDF for the stationary processes defined in the usual way (2.3) by the orthogonal increment processes  $\{d\zeta_1(f)\}$  and  $\{d\zeta_2(f)\}$ .

At this point it becomes convenient to use matrices. We can equally say that the bivariate process  $\{\mathbf{Z}_t\}$  can be expressed as

$$\mathbf{Z}(t) = \int_{-\infty}^{\infty} \mathbf{A}_t(f)e^{i2\pi ft}d\boldsymbol{\zeta}(f) \quad (7.2)$$

where

$$\mathbf{A}_t(f) = \begin{bmatrix} A_{1,t}(f) & 0 \\ 0 & A_{2,t}(f) \end{bmatrix}, \quad d\boldsymbol{\zeta}(f) = [d\zeta_1(f), d\zeta_2(f)]^T. \quad (7.3)$$

In this situation matrices are being used purely as a method of simplifying notation. We are dealing with two individual processes represented with their own modulating function and orthogonal process. As such, the off-diagonal terms in  $\mathbf{A}_t(f)$  are zero. The consequence of non-zero off-diagonal terms is looked at in Section 8.2.

Using this notation it is convenient to express the ESDF matrix as [83]

$$\mathbf{S}_t(f) = \mathbf{A}_t(f)\mathbf{S}(f)\mathbf{A}_t^H(f) \quad (7.4)$$

where

$$\mathbf{S}(f)df = E\{d\zeta(f)d\zeta^H(f)\} = \begin{bmatrix} S_{11}(f) & S_{12}(f) \\ S_{21}(f) & S_{22}(f) \end{bmatrix} df.$$

The evolutionary spectral properties of the bivariate process are represented by the matrix

$$\mathbf{S}_t(f) = \begin{bmatrix} S_{11,t}(f) & S_{12,t}(f) \\ S_{21,t}(f) & S_{22,t}(f) \end{bmatrix}.$$

## 7.2 Coherence for Evolutionary Processes

Now that we have a clearly defined concept of the cross-spectrum for a bivariate oscillatory process it is possible to proceed with a definition of coherence in the usual manner. The ordinary coherence  $\gamma_t^2(f)$ , now a function of both  $f$  and time index  $t$ , is defined as [83]

$$\gamma_t^2(f) = \frac{|S_{12,t}(f)|^2}{S_{11,t}(f)S_{22,t}(f)}.$$

Using (7.4) we have

$$\gamma_t^2(f) = \frac{|A_{1,t}(f)A_{2,t}^*(f)S_{12}(f)|^2}{|A_{1,t}(f)|^2S_{11}(f)|A_{2,t}(f)|^2S_{22}(f)}.$$

This in fact reduces down to the ordinary coherence for the stationary processes defined in the normal way by  $\{\zeta_1(f)\}$  and  $\{\zeta_2(f)\}$ , giving

$$\gamma_t^2(f) = \frac{|S_{12}(f)|^2}{S_{11}(f)S_{22}(f)}. \quad (7.5)$$

Thus the resulting coherence for an oscillatory process is itself time-invariant. We retain the index  $t$  to distinguish it from stationary coherence.

As in the stationary case, for complex-valued processes we can extend the framework to deal with relation terms. This was not considered by Priestley, however the extension is natural and presented here. First, let us consider a univariate zero-mean complex-valued oscillatory process  $\{Z(t)\}$  with represen-

tation (6.10) and ESDF  $S_t(f)$ . The conjugated process will have representation

$$\begin{aligned} Z^*(t) &= \int_{-\infty}^{\infty} A_t^*(f) e^{-i2\pi ft} d\zeta^*(f) \\ &= \int_{-\infty}^{\infty} A_t^*(-f) e^{i2\pi ft} d\zeta^*(-f). \end{aligned} \quad (7.6)$$

We conclude from the properties of  $A_t(f)$  that the Fourier transform of  $A_t^*(-f)$  will too be peaked at zero and hence  $\{Z^*(t)\}$  is oscillatory with ESDF  $S_t(-f)$ . Furthermore it follows that the characteristic width of  $\{Z^*(t)\}$  is equal to the characteristic width of  $\{Z(t)\}$ .

For a bivariate process  $\{\mathbf{Z}(t)\}$ , let  $\{\check{\mathbf{Z}}(t) = [\mathbf{Z}^T(t), \mathbf{Z}^H(t)]^T, t \in \mathbb{R}\}$  be the augmented vector process. Using (7.2) and (7.6) this can be represented as

$$\check{\mathbf{Z}}(t) = \int_{-\infty}^{\infty} \check{\mathbf{A}}_t(f) e^{i2\pi ft} d\check{\zeta}(f)$$

where

$$\check{\mathbf{A}}_t(f) = \begin{bmatrix} \mathbf{A}_t(f) & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_t^*(-f) \end{bmatrix}, \quad \check{\zeta}(f) = [\zeta^T(f), \zeta^H(-f)]^T.$$

The augmented ESDF matrix is defined as

$$\Upsilon_t(f) = \check{\mathbf{A}}_t \Upsilon(f) \check{\mathbf{A}}_t^H$$

where  $\Upsilon(f)df = E\{d\check{\zeta}(f)d\check{\zeta}(f)^H\}$  is the augmented SDF (see Section 2.3 and equation (2.6)) for the complex-valued SOS process with representation

$$\mathbf{X}(t) = \int_{-\infty}^{\infty} e^{i2\pi ft} d\zeta(f).$$

Analogously, the augmented ESDF matrix takes the form

$$\begin{aligned} \mathbf{\Upsilon}_t(f) &= \begin{bmatrix} \mathbf{S}_t(f) & \mathbf{R}_t(f) \\ \mathbf{R}_t^*(-f) & \mathbf{S}_t^*(-f) \end{bmatrix} \\ &= \begin{bmatrix} S_{11,t}(f) & S_{12,t}(f) & R_{11,t}(f) & R_{12,t}(f) \\ S_{21,t}(f) & S_{22,t}(f) & R_{21,t}(f) & R_{22,t}(f) \\ R_{11,t}^*(-f) & R_{12,t}^*(-f) & S_{11,t}^*(-f) & S_{12,t}^*(-f) \\ R_{21,t}^*(-f) & R_{22,t}^*(-f) & S_{21,t}^*(-f) & S_{22,t}^*(-f) \end{bmatrix}. \end{aligned} \quad (7.7)$$

The diagonal elements of this matrix are true evolutionary spectral density functions for the individual processes. All off-diagonal terms are cross-spectra.  $S_{lm,t}(f)$ ,  $l \neq m$ , are simply conventional evolutionary cross-spectral density functions.  $R_{ll,t}(f)$  is the evolutionary relation spectral density function for the process  $\{Z_l(t)\}$ , which can be thought of as being the ECSDF of  $\{Z_l(t)\}$  and  $\{Z_l^*(t)\}$ .  $R_{lm,t}(f)$  are evolutionary relation cross-spectral density functions, in other words  $R_{lm,t}(f)$  would be the ECSDF for the processes  $\{Z_l(t)\}$  and  $\{Z_m^*(t)\}$ . As in the stationary case, we can extend the notion of coherence to the three following types:

1. the ordinary coherence  $\gamma_{lm,t}^2(f)$  for the pair of processes  $\{Z_l(t)\}$  and  $\{Z_m(t)\}$  is defined as

$$\gamma_{lm,t}^2(f) = \frac{|S_{lm,t}(f)|^2}{S_{ll,t}(f)S_{mm,t}(f)}; \quad (7.8)$$

2. the conjugate coherence  $\gamma_{lm^*,t}^2(f)$  for the pair of processes  $\{Z_l(t)\}$  and  $\{Z_m^*(t)\}$  is defined as

$$\gamma_{lm^*,t}^2(f) = \frac{|R_{lm,t}(f)|^2}{S_{ll,t}(f)S_{mm,t}(-f)}; \quad (7.9)$$

3. the conjugate coherence  $\gamma_{l^*l,t}^2(f)$  for the pair of processes  $\{Z_l(t)\}$  and

$\{Z_i^*(t)\}$  is defined as

$$\gamma_{u_*,t}^2(f) = \frac{|R_{u,t}(f)|^2}{S_{u,t}(f)S_{u,t}(-f)}. \quad (7.10)$$

As in (7.5) these reduce down to the time-invariant stationary equivalents.

To estimate the ordinary or conjugate coherence measures defined above we need to construct an estimator  $\hat{\Upsilon}_t(f)$  of the augmented ESDF matrix  $\Upsilon_t(f)$ . In Section 6.5 was presented two estimation procedures for the ESDF of a univariate process. The Priestley method of estimation and the temporally smoothed wavelet spectrum. Rigorous treatment on the conditions for forming an unbiased estimator were discussed and the wavelet method was demonstrated on some simple nonstationary processes. We now extend these spectral estimators to the multivariate case and demonstrate how distributional properties of the TWCOH estimator lend themselves to the coherence estimation in the evolutionary case.

### 7.3 Multivariate Evolutionary Spectral Estimators

The following multivariate framework was presented in [83]. Here we extend it to augmented complex vector-valued processes. Given a finite length- $T$  realisation of a bivariate zero-mean complex-valued process  $\{\mathbf{Z}(t) = [Z_1(t), Z_2(t)]^T\}$ , for a unit norm filter  $\{g(u)\}$  (with Fourier transform  $\Gamma(f)$ ) the complex vector-valued filtered process  $\mathbf{U}_t(f_0)$  is defined as

$$\mathbf{U}_t(f_0) = \int_{-\infty}^{\infty} g(u)\check{\mathbf{Z}}(t-u)e^{-i2\pi f_0(t-u)}du,$$

where  $\{\check{\mathbf{Z}}(t) = [\mathbf{Z}^T(t), \mathbf{Z}^H(t)]^T\}$ . Analogously to the univariate case, defining the characteristic width of  $\{\mathbf{Z}(t)\}$  to be  $B_{\mathbf{Z}} = \min\{B_{Z_1}, B_{Z_2}\}$  [83], provided  $B_g \ll B_{\mathbf{Z}} \ll T$ , then the results in Section 6.5.1 apply to all the elements of  $\Upsilon_t(f)$ , giving

$$E\{\mathbf{U}_t(f_0)\mathbf{U}_t^H(f_0)\} \approx \Upsilon_t(f_0)$$

where  $\Upsilon_t(f)$  is the augmented ESDF matrix for the process  $\{\mathbf{Z}(t)\}$ , assumed to be smooth with respect to  $|\Gamma(f)|^2$ . Using a weighting function  $W_{T'}(u)$  with width parameter  $T'$ , the weighted temporal average is defined as

$$\mathbf{V}_t(f_0) = \int_{-\infty}^{\infty} W_{T'}(u) \mathbf{U}_{t-u}(f_0) \mathbf{U}_{t-u}^H(f_0) du,$$

with

$$E\{\mathbf{V}_t(f_0)\} = \int_{-\infty}^{\infty} \bar{\Upsilon}_t(f + f_0) |\Gamma(f)|^2 df,$$

where

$$\bar{\Upsilon}_t(f + f_0) = \int_{-\infty}^{\infty} W_{T'}(u) \Upsilon_{t-u}(f + f_0) du.$$

As such if all entries in  $\bar{\Upsilon}_t(f)$  are smooth in comparison to  $|\Gamma(f)|^2$  then

$$E\{\mathbf{V}_t(f_0)\} \approx \bar{\Upsilon}_t(f_0).$$

It was shown in Section 6.5.3 that manipulation of the continuous Morlet wavelet transform allowed direct comparisons to be drawn between Priestley's spectral estimator and the temporally smoothed Morlet wavelet spectrum. Following an analogous argument, but replacing the univariate process with the multivariate augmented process  $\{\check{\mathbf{Z}}(t)\}$  we conclude that for the Morlet wavelet

$$\psi(t) = d^{-1/2} \pi^{-1/4} e^{i2\pi t} e^{-\frac{1}{2}(\frac{t}{d})^2},$$

at scales  $|a| > 0$  and times  $b \in \mathbb{R}$ , provided  $B_{g_a} = 2(|a|d)^{3/2} \pi^{-1/4} \ll B_{\mathbf{Z}}$  and the entries of the augmented ESDF are smooth and slowly-varying in  $f$ , then

$$E\{W(a, b; \check{\mathbf{Z}}, \psi) W^H(a, b; \check{\mathbf{Z}}, \psi)\} \approx \Upsilon_b(1/a). \quad (7.11)$$

Further to this we have that the temporally smoothed wavelet spectrum

$$\begin{aligned} \mathcal{S}\{W(a, b; \check{\mathbf{Z}}, \psi)W^H(a, b; \check{\mathbf{Z}}, \psi)\} = \\ \int_{b-\frac{T'(a)}{2}}^{b+\frac{T'(a)}{2}} W(a, u-b; \check{\mathbf{Z}}, \psi)W^H(a, u-b; \check{\mathbf{Z}}, \psi)du \approx \bar{\Upsilon}_b(1/a) \end{aligned}$$

where

$$\bar{\Upsilon}_b(1/a) = (1/T'(a)) \int_{-T'(a)/2}^{T'(a)/2} \Upsilon_{b-u}(1/a)du$$

with  $T'(a)$  as the length of the rectangular smoothing window, proportional to scale  $a$ .

## 7.4 Discrete-Time TWCOH Estimator

We now move to a discrete-time formulation. Suppose we sample  $N$  points  $\{\mathbf{Z}_n, n = 0, \dots, N-1\}$  of the continuous-time semi-stationary process  $\{\mathbf{Z}(t)\}$  with a sampling interval  $\Delta$ , provided  $\Upsilon_t(f) \approx 0$  for all  $|f| > 1/(2\Delta)$  and at every  $t$  in the observation window, we can use the discrete-time formulation of the CWT from Section 5.3 to write

$$W(a, b; \check{\mathbf{Z}}, \psi) = \Delta \sum_{n=0}^{N-1} \check{\mathbf{Z}}_n \lambda_{a,b,d;n} e^{-i2\pi\left(\frac{n\Delta-b}{a}\right)} \quad (7.12)$$

where  $\lambda_{a,b,d;n} = \pi^{-1/4}(|a|d)^{-1/2}e^{-\frac{1}{2}\left(\frac{n\Delta-b}{ad}\right)^2}$ .

To derive distributional results for TWCOH in the semi-stationary case we first recall the stationary results of Section 5.5. With consideration of the valid values of scales  $a$  and times  $b$  — see Section 5.3.1 — for a  $N$  point sampled realisation  $\{\mathbf{X}_n, n = 0, \dots, N-1\}$  of a jointly SOS bivariate process  $\{\mathbf{X}(t) = [X_1(t), X_2(t)]^T, t \in \mathbb{R}\}$ , the matrix of temporally smoothed wavelet

spectra is defined as

$$\hat{\Omega}_{\mathbf{X}}(a, b) \equiv \begin{bmatrix} \mathcal{S}\{X_1, X_1\} & \mathcal{S}\{X_1, X_2\} & \mathcal{S}\{X_1, X_1^*\} & \mathcal{S}\{X_1, X_2^*\} \\ \mathcal{S}\{X_2, X_1\} & \mathcal{S}\{X_2, X_2\} & \mathcal{S}\{X_2, X_1^*\} & \mathcal{S}\{X_2, X_2^*\} \\ \mathcal{S}\{X_1^*, X_1\} & \mathcal{S}\{X_1^*, X_2\} & \mathcal{S}\{X_1^*, X_1^*\} & \mathcal{S}\{X_1^*, X_2^*\} \\ \mathcal{S}\{X_2^*, X_1\} & \mathcal{S}\{X_2^*, X_2\} & \mathcal{S}\{X_2^*, X_1^*\} & \mathcal{S}\{X_2^*, X_2^*\} \end{bmatrix},$$

where we use the short hand notation

$$\begin{aligned} \mathcal{S}\{X_j, X_k\} &\equiv \mathcal{S}\{W(a, b; X_j, \psi)W^*(a, b; X_k, \psi)\} \\ &= \frac{1}{2M+1} \sum_{l=b_0-M}^{b_0+M} W(a, l\Delta; X_j, \psi)W^*(a, l\Delta; X_k, \psi). \end{aligned}$$

Table 5.1 defines the smoothing parameter  $M$ . It has been shown that  $\hat{\Omega}_{\mathbf{X}}(a, b)$  is equivalent to  $\hat{\Upsilon}_b(1/a)$ , a maximally overlapped WOSA spectral estimator of the augmented SDF matrix  $\Upsilon(f)$  of process  $\{\mathbf{X}_t\}$ , utilising a sub-portion of the observed process of length  $N_P$  — see (5.19) — centered around time  $b$ . For a pair of jointly stationary Gaussian processes it was shown that asymptotically as  $N_P \rightarrow \infty$

$$\hat{\Omega}_{\mathbf{X}}(a, b) \stackrel{d}{=} (1/\varphi)W_4^C\{\varphi, \Upsilon(1/a)\}, \quad (7.13)$$

where the complex degrees of freedom of the estimator are  $\varphi$ , calculated from the multitaper formulation — see (5.33).

We now turn our attention to the semi-stationary processes. As has been stated, the result in (7.13) is an asymptotic distribution for large  $N_P$ . In the stationary case asymptotic considerations are natural due to the time-invariant statistical properties of an SOS process. In the nonstationary case more care is needed.

A discussion on the asymptotic considerations needed for nonstationary processes was presented in Section 6.7. For the spectral estimation of semi-stationary processes, restrictions on the size of  $N_P$  are enforced by the rate at which the ESDF changes in time. For a zero-mean semi-stationary process

$\{\mathbf{Z}_t\}$  with augmented EDSF matrix  $\Upsilon_t(f)$  and characteristic width  $B_{\mathbf{Z}}$ , not only does the condition  $B_{g_a} = 2(|a|d)^{3/2}\pi^{-1/4} \ll B_{\mathbf{Z}}$  need to be satisfied for (7.11) to hold, but for the distributional results for stationary processes to be valid in the semi-stationary case it is also necessary that stationarity can be assumed across the time interval  $\Delta N_P$ . This is ensured by the stricter condition of  $\Delta N_P < B_{\mathbf{Z}}$ . Provided this condition is satisfied the signal portion upon which the analysis is performed,  $\{\tilde{\mathbf{Z}}_n = \mathbf{Z}_{(b_0 - \frac{N_S-1}{2} - M+n)}, n = 0, \dots, N_P - 1\}$ , can be assumed to be a portion of a bivariate SOS process and the EDSF matrix will be approximately constant over time  $\Delta N_P$ . Therefore the matrix

$$\hat{\Omega}_{\mathbf{Z}}(a, b) \equiv \mathcal{S}\{W(a, b; \tilde{\mathbf{Z}}, \psi)W^H(a, b; \tilde{\mathbf{Z}}, \psi)\}$$

will have expected value  $\Upsilon_b(1/a)$ . With this restriction on the size of  $N_P$  the asymptotic result of (7.13) now needs to be treated with some caution.

By considering the multitaper formulation of temporally smoothed Morlet wavelet spectra (5.30), Walden [102] in fact shows that for finite signal lengths, up to second order moments, the Wishart distribution is still appropriate provided the analysis frequency, in this case  $f = 1/a$ , lies within a frequency band  $W < |f| < f_N - W$ , where  $W \rightarrow 0$  as the signal portion and taper lengths tend to infinity. ( $2W$  is the bandwidth of the overall spectral window of the multitaper estimator — see [102]). It has been shown in Section 5.6 that (7.13) appears valid for typical scale values and smoothing parameters, including those that give relatively small values of  $N_P$  (in the order of 128), and so we can say that  $\hat{\Omega}_{\mathbf{Z}}(a, b)$  will be approximately distributed as

$$\hat{\Omega}_{\mathbf{Z}}(a, b) \stackrel{d}{=} \mathcal{W}_4^C\{\varphi, \Upsilon_b(1/a)\}.$$

Invoking Theorem 2.8.2, we conclude that the three types of TWCOH estimator (5.35), (5.36) and (5.37) will be approximately Goodman distributed with true coherence values (7.8), (7.9) and (7.10), respectively, and  $\varphi$  complex degrees of freedom — see (5.33).

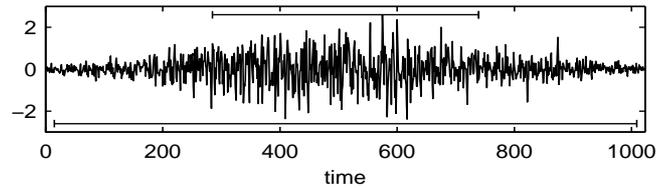


Figure 7.1: Real part of a realisation of the UMP  $\{Z_t\}$  used in Section 7.5.1. The bars indicate the length  $N_P$  of the signal portion used in analysis of a single point, for smoothing parameters (top)  $\kappa_0 = 9.6$  and (bottom)  $\kappa_0 = 24.6$ .

This result will now be verified with some examples.

## 7.5 Wavelet Coherence Examples

### 7.5.1 Time-and-Frequency-Invariant Coherence

Firstly we consider a special case where the coherence is not only invariant to time, but also invariant to frequency. Once again we consider the improper SOS CAR(1) process which we will now label  $\{X_t\}$ , and form the jointly stationary bivariate process  $\{[X_t, X_t^*]^T\}$ . It is known that the conjugate coherence, which we label  $\gamma_*^2(f)$ , is equal to 0.36 for all frequencies. If we now let  $\{Z_t\}$  be a UMP with  $Z_t = h_t X_t$  where  $h_t$  is a finite-valued deterministic modulating function, Priestley's work on the coherence of evolutionary processes tells us that the conjugate coherence  $\gamma_{*,t}^2(f)$  for the bivariate process  $\{[Z_t, Z_t^*]^T\}$  will be equal to 0.36 for all frequencies and all time.

We simulate five thousand realisations of the CAR(1) SOS process and modulate with the function  $h_t = \exp(-(t - 512)^2/2\sigma^2)$ , where  $\sigma = 1024/6$ . Using the analysis of Section 6.3.1,  $B_Z \geq 214$ . Four different smoothing factors of  $\kappa_0 = 9.6, 14.6, 19.6$  and  $24.6$  (resulting in complex degrees of freedom equal to 8, 12, 16 and 20 respectively) are applied in calculating the wavelet coherence measures at the time-scale point  $b = 512$  and  $a = 18$ . The five thousand wavelet coherence samples for the four different levels of smoothing were collated in a histogram format, and with the use of kernel smoothing the

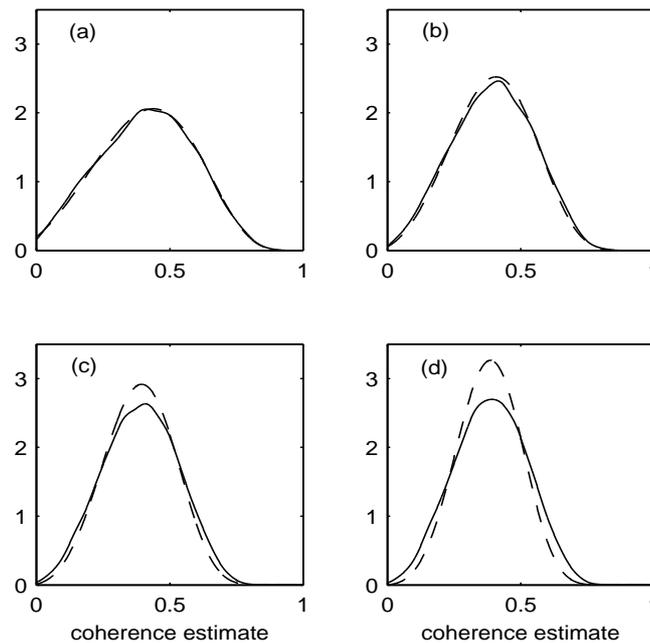


Figure 7.2: Probability density functions of the coherence estimates approximated from five thousand realisations of the bivariate UMP from Section 7.5.1 using the smoothing parameters  $\kappa_0 =$  (a) 9.6, (b) 14.6, (c) 19.6 and (d) 24.6. Plotted with a dashed line is the appropriate Goodman distribution assumed under stationarity.

estimated PDF is plotted in Figure 7.2. The dotted line is the true PDF of the Goodman distribution for the appropriate complex degrees of freedom. In Figure 7.3 we also draw the corresponding Q-Q plots for the four smoothing parameters.

For the smallest smoothing window (where  $N_P = 455$ ) the distribution of the wavelet coherence estimates match the appropriate Goodman distribution very closely. As the smoothing window increases the degrees of freedom also increase but the match with the appropriate Goodman distribution weakens. We attempt to explain this observation with the idea that as the smoothing window increases,  $N_P$  becomes significantly larger than  $B_Z$  and the assumption that the process is stationary across the signal portion fails to hold. Hence the statistics start to deviate from the those derived using this assumption.

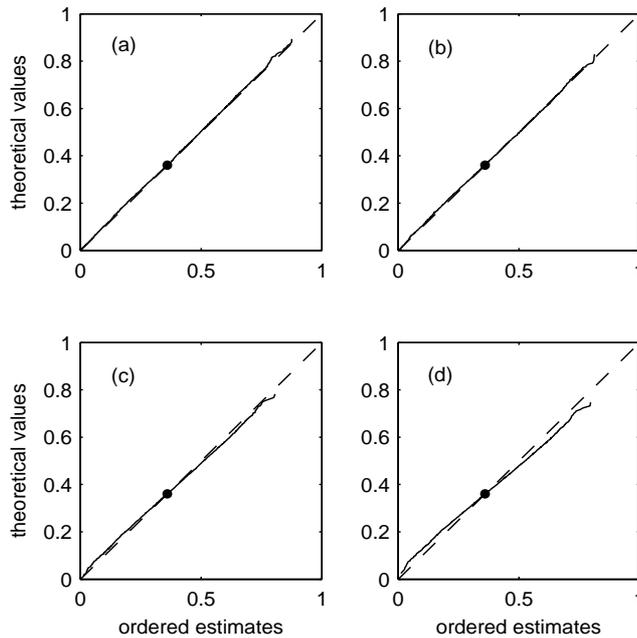


Figure 7.3: Q-Q plots constructed using five thousand estimates for the smoothing parameters  $\kappa_0 =$  (a) 9.6, (b) 14.6, (c) 19.6 and (d) 24.6.

The variance in the coherence estimates increase, which manifests itself as a lengthening of the distribution's tails. This is also recognised as being the cause of the skew away from unit gradient in the Q-Q plot — see Figure 7.3. On these Q-Q plots is marked 0.36, the true value of the MSCOH, through which the curve runs.

### 7.5.2 Frequency-Dependent Coherence

We now consider a bivariate UMP in which we introduce frequency variation to the coherence as follows. Consider a SOS process  $\{W_t\}$  with spectral density  $S_{WW}(f)$ . From  $\{W_t\}$  we define the pair of stationary processes  $X_{1,t} = W_t + \epsilon_t$  and  $X_{2,t} = W_t + \nu_t$  where  $\{\epsilon_t\}$  and  $\{\nu_t\}$  are zero-mean independent white noise processes with variances  $\sigma_\epsilon^2$  and  $\sigma_\nu^2$  respectively.  $\{X_{1,t}\}$  will have SDF  $S_{WW}(f) + \sigma_\epsilon^2$  and  $\{X_{2,t}\}$  has SDF  $S_{WW}(f) + \sigma_\nu^2$ . The CSDF is equal to  $S_{WW}(f)$ . The ordinary coherence for the pair of jointly stationary processes  $\{X_{1,t}\}$  and

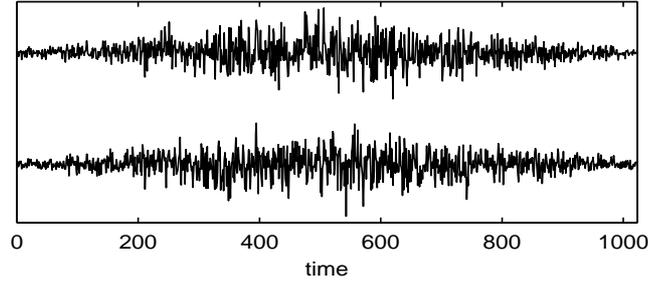


Figure 7.4: Real parts of a realisation of the bivariate UMP from Section 7.5.2.

$\{X_{2,t}\}$  takes the form

$$\gamma^2(f) = \frac{S_{WW}^2(f)}{[S_{WW}(f) + \sigma_\epsilon^2][S_{WW}(f) + \sigma_\nu^2]}. \quad (7.14)$$

With a finite-valued deterministic modulating function  $g_t$  we define the bivariate UMP  $\{[g_t X_{1,t}, g_t X_{2,t}]^T\}$ . The coherence  $\gamma_t^2(f)$  is time invariant and equal to the expression in (7.14).

For this specific example we let  $\{W_t\}$  be the CAR(1) process. The white noise processes are Gaussian with  $\sigma_\epsilon^2 = \sigma_\nu^2 = 1$ , and we modulate  $\{X_{1,t}\}$  and  $\{X_{2,t}\}$  with the function  $g_t = \exp(-(t - t_0)/2\sigma^2)$ , with  $t_0 = 512$  and  $\sigma = 1024/6$ . In Figure 7.4 we plot the real part of a single realisation of the nonstationary signals  $\{g_t X_{1,t}\}$  (top) and  $\{g_t X_{2,t}\}$  (bottom). Five thousand realisations of the bivariate UMP were generated. Under stationarity, the Goodman distribution for coherence dictates that if  $\hat{\gamma}^2(a, b)$  is the raw TWCOH estimator with  $\varphi$  complex degrees of freedom, then an (approximately) unbiased estimator is given by  $\tilde{\gamma}^2(a, b) = [\hat{\gamma}^2(a, b) - (1/\varphi)]/[1 - (1/\varphi)]$ . The debiased mean coherences across time for the scales  $a = 6, 12, 18$  and  $24$  are plotted in Figures 7.5(a) and (c) for  $\kappa_0 = 6$  ( $\varphi = 5$  complex degrees of freedom) and  $\kappa_0 = 12$  ( $\varphi = 10$ ), respectively. Dotted lines show true coherence values,  $\gamma_b^2(1/a)$ . To provide an indication of how the block size  $N_P$  — the length of the portion of signal that is used in calculating the coherence — varies with  $a$ , the horizontal bars show the block size for  $a = 6$  (top) and

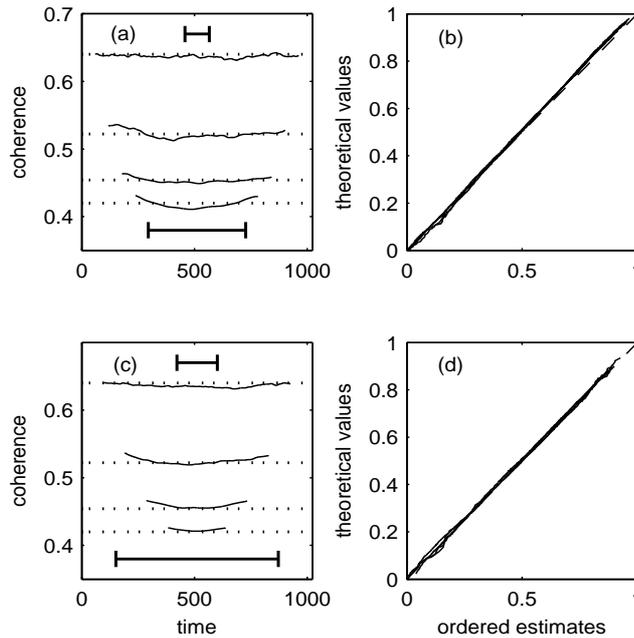


Figure 7.5: Left column: Debiased mean coherence estimates from 5000 simulations for scales  $a = 6, 12, 18, 24$  (top to bottom) using smoothing with (a)  $\kappa_0 = 6$  (b)  $\kappa_0 = 12$ . Dotted lines show true coherence values. Horizontal bars show the block size,  $N_P$ , for  $a = 6$  (top) and  $a = 24$  (bottom). Right column: Corresponding Q-Q plots for same scales at a single time point  $b = 512$  for smoothing with (b)  $\kappa_0 = 6$ , (d)  $\kappa_0 = 12$ .

$a = 24$  (bottom). Figures 7.5(b) and (d) give the corresponding Q-Q plots for the four scales at the time point  $b = 512$ . (The results are representative for general  $b$  values.) Estimated debiased mean coherence remains approximately constant across time as the evolutionary model predicts and there is good agreement with the expected value under the assumed Goodman distribution. The Q-Q plots show very close agreement between ordered sample values (over the 5000 simulations) and quantiles of the Goodman distribution having  $\varphi$  complex degrees of freedom.

## 7.6 Priestley's Coherence Estimator

With respect to Priestley's estimation scheme, Kiehm and Mélard apply the spectral estimator experimentally in [51] to examples of process that give time dependent coherences [63]. Specifically they look at some bivariate examples of the one-sided infinite moving-average representation whose univariate formulation was presented in Section 6.4. A multivariate formulation, along with a study of coherence for this model will be looked at in Chapter 8. By simulating these processes Kiehm and Mélard were able to gain some experimental insight into the bias of coherence estimators derived using this spectral estimation procedure, however mathematical expressions are not apparent in the literature.

In [36] the authors offer a spectral estimation procedure for evolutionary processes. Like Priestley's estimator, it is also based on the filtering of a complex demodulated signal. However to reduce bias, instead of smoothing in the time domain the spectral estimates are averaged over a discrete set frequencies  $\{f_m, m = -M, -M + 1, \dots, M\}$ , say, that are evenly spaced around the frequency of interest  $f_0$ . Relying on the filter being sufficiently concentrated as to render the individual estimates at each  $f_m, m = -M, \dots, M$  uncorrelated (a consequence of the property of the orthogonal increment process), the coherence estimator for a pair of processes formed from the averaged spectral estimators is shown to be Goodman distributed, under the null hypothesis of Gaussian joint stationarity.

The methodology used in deriving the distribution of the TWCOH in the stationary case is not dependent on the tapering function being Gaussian shaped, or indeed that the smoothing window is rectangular. With this in mind we can therefore relate this statistical analysis to the Priestley evolutionary spectral estimators to derive statistical results for the ESDF matrix estimators and time-frequency coherence. We first derive the results for Gaussian jointly stationary processes. This will then be extended to a general semi-stationary bivariate process.

## 7.6.1 Priestley's Estimator for Discrete-Time SOS Processes

Priestley proposes a discrete-time formulation to his continuous-time ESDF estimator in [81]. Here we extend it to consider complex vector-valued processes. Let  $\{\mathbf{X}_t\}$  be a zero-mean bivariate complex-valued jointly SOS process sampled with unit time interval ( $\Delta = 1$ ). For the augmented process  $\{\check{\mathbf{X}}_t = [\mathbf{X}_t^T, \mathbf{X}_t^H]^T\}$ , let

$$\mathbf{U}_t(f_0) = \sum_{u=-\infty}^{\infty} g_u \check{\mathbf{X}}_{t-u} e^{i2\pi f_0(t-u)}$$

where  $\{g_u\}$  is a discretised filter with  $\sum_{u=-\infty}^{\infty} g_u^2 = 1$  and Fourier transform  $\Gamma(f) = \sum_{u=-\infty}^{\infty} g_u e^{-i2\pi f u}$ . The logical extension to a signal sampled at an arbitrary sampling interval  $\Delta$  is

$$\mathbf{U}_t(f_0) = \Delta \sum_{u=-\infty}^{\infty} g(u\Delta) \check{\mathbf{X}}(t - u\Delta) e^{-i2\pi f_0(t-u\Delta)}, \quad |f_0| < f_N.$$

The filter  $\{g(\cdot)\}$  is chosen to be symmetric around zero (i.e.  $g(u\Delta) = g(-u\Delta)$ ), and as such by reversing the time variable we can write

$$\mathbf{U}_t(f_0) = \Delta \sum_{u=-\infty}^{\infty} g(u\Delta) \check{\mathbf{X}}(t + u\Delta) e^{-i2\pi f_0(t+u\Delta)}.$$

We now make the substitution  $n\Delta = t + u\Delta$  to give

$$\begin{aligned} \mathbf{U}_t(f_0) &= \Delta \sum_{n=-\infty}^{\infty} g(n\Delta - t) \check{\mathbf{X}}(n\Delta) e^{-i2\pi f_0(n\Delta)} \\ &= e^{-i2\pi f_0 t} \Delta \sum_{n=-\infty}^{\infty} \check{\mathbf{X}}(n\Delta) g(n\Delta - t) e^{-i2\pi f_0(n\Delta - t)}. \end{aligned}$$

Say we observe a finite length realisation of the random process at the times  $0, \Delta, 2\Delta, \dots, (N-1)\Delta$ , the sum can be reduced to

$$\mathbf{U}_t(f_0) = e^{-i2\pi f_0 t} \Delta \sum_{n=0}^{N-1} \tilde{\mathbf{X}}(n\Delta) g(n\Delta - t) e^{-i2\pi f_0 (n\Delta - t)},$$

which besides from the complex phase factor outside of the sum, is equivalent to the discrete-time formulation of the wavelet transform (7.12). In this discrete formulation, for the smoothed estimator  $\mathbf{V}_t(f_0)$  becomes

$$\mathbf{V}_t(f_0) = \sum_{u=-M}^M W_u \mathbf{U}_{t-u}(f_0) \mathbf{U}_{t-u}^H(f_0)$$

where  $\{W_{-M}, W_{-M+1}, \dots, W_M\}$  is a set of weights that sum to one, and the time period  $2M\Delta$  corresponds to the time interval  $T'$  from Priestley's continuous-time formulation. For an arbitrary  $(l, m)$ th element  $V_{lm,t}(f_0)$  of the  $4 \times 4$  matrix  $\mathbf{V}_t(f_0)$ , assuming  $\{g(u)\}$  vanishes outside an interval that is much less than  $(N-1)\Delta$ , and thus can be represented as a finite sequence  $g_0, \dots, g_{N_S-1}$  with  $N_S - 1 \ll N$ , in an identical way to (5.29)

$$V_{lm,t}(f_0) = \Delta \mathbf{D}_l^H \mathbf{B} \mathbf{B}^T \mathbf{D}_m.$$

Here  $\mathbf{D}_l = [D_{l,0}, \dots, D_{l,N_P-1}]^T$  where  $D_{l,n} = \tilde{X}_{l,n}^* \exp\{i2\pi f_0 l \Delta\}$  and the signal portion  $\{\tilde{X}_{l,n} \equiv \tilde{X}_{n,(t-\frac{N_S-1}{2}-M+n)}, n = 0, \dots, N_P-1\}$  is the  $N_P = 2M + N_S$  contiguous points of the process used in calculating  $\mathbf{V}_t(f_0)$ .  $\mathbf{B}$  is a matrix of size  $2M+1$  by  $N_P$  and the  $j$ th column will be the vector  $\sqrt{W_j} g'_{j,n}$ ,  $n = 0, 1, \dots, N_P$  with

$$g'_{j,n} = \begin{cases} g_{n-j}, & \text{if } n = j, \dots, j + N_S - 1; \\ 0, & \text{otherwise.} \end{cases}$$

The matrix  $\mathbf{B}\mathbf{B}^T$  will have  $N_B = 2M + 1$  non-zero positive eigenvalues  $\sigma_0 \geq \dots \geq \sigma_{N_B-1}$  and as such the SDF estimator  $V_{lm,t}(f_0)$  of  $\Upsilon_{lm}(f_0)$  will have

$$\varphi = \frac{1}{\sum_{j=0}^{N_B-1} \sigma_j^2} \quad (7.15)$$

complex degrees of freedom.  $\{\mathbf{X}_t\}$  is jointly stationary with augmented SDF matrix  $\Upsilon(f)$ . Under the evolutionary framework it can be considered as a oscillatory process represented by family  $\mathcal{F} = \{1 \cdot e^{i2\pi ft}\}$  of infinite characteristic width with (the time-invariant) augmented ESDF matrix  $\Upsilon_t(f) = \Upsilon(f)$ . Assuming  $\Upsilon(f)$  is smooth with respect to the support of  $|\Gamma(f)|^2$ , we have

$$E\{\mathbf{V}_t(f)\} \approx \Upsilon(f) \quad |f| < f_N,$$

and asymptotically as  $N_P \rightarrow \infty$

$$\mathbf{V}_t(f) \stackrel{d}{=} (1/\varphi)\mathcal{W}_4^C\{\varphi, \Upsilon(f)\}. \quad (7.16)$$

### 7.6.2 Priestley's Estimator for Discrete-Time Semi-Stationary Processes

We now make the progression from stationarity to semi-stationarity using a similar argument to that presented for the wavelet case in Section 7.4. We replace the stationary process in Section 7.6.1 with a bivariate zero-mean semi-stationary process  $\{\mathbf{Z}_t\}$  with augmented ESDF matrix  $\Upsilon_t(f)$  and characteristic width  $B_{\mathbf{Z}}$ , and given that we can assume that not only  $B_g \ll B_{\mathbf{Z}}$ , but also the stronger condition  $\Delta N_P < B_{\mathbf{Z}}$ , then  $\Upsilon_t(f)$  will be approximately constant across the signal portion and

$$E\{\mathbf{V}_t(f)\} \approx \Upsilon_t(f), \quad |f| < f_N.$$

Once again, the restriction on the size  $N_P$  means a certain amount of caution needs to be taken with the asymptotic result (7.16), but once again referring

to the work of Walden [102] — see the discussion in Section 7.4 — we can say that given  $N_P$  is of a reasonable size, up to second order moments,

$$\mathbf{V}_t(f) \stackrel{d}{=} (1/\varphi)\mathcal{W}_p^C\{\varphi, \mathbf{\Upsilon}_t(f)\}.$$

We can now define the following three types of coherence estimator for  $j, k = 1, 2$ :

1. the estimator

$$\hat{\gamma}_{jk,t}^2(f) = \frac{|V_{j,k,t}(f)|^2}{V_{j,j,t}(f)V_{k,k,t}(f)} \quad (7.17)$$

for the ordinary coherence for the pair of processes  $\{Z_{j,t}\}$  and  $\{Z_{k,t}\}$  as defined in (7.8);

2. the estimator

$$\hat{\gamma}_{jk^*,t}^2(f) = \frac{|V_{j,k+2,t}(f)|^2}{V_{j,j,t}(f)V_{k+2,k+2,t}(f)} \quad (7.18)$$

for the conjugate coherence for the pair of processes  $\{Z_{j,t}\}$  and  $\{Z_{k,t}^*\}$  as defined in (7.9);

3. the estimator

$$\hat{\gamma}_{jj^*,t}^2(f) = \frac{|V_{j,j+2,t}(f)|^2}{V_{j,j,t}(f)V_{j+2,j+2,t}(f)} \quad (7.19)$$

for the conjugate coherence for the pair of processes  $\{Z_{j,t}\}$  and  $\{Z_{j,t}^*\}$  as defined in (7.10).

Invoking Theorem 2.8.2, we conclude that the three estimators (7.17), (7.18) and (7.19) will be approximately Goodman distributed with true coherence values (7.8), (7.9) and (7.10), respectively, and  $\varphi$  complex degrees of freedom — see (7.15).

We demonstrate these theoretical results on an example found in the engineering literature.

## 7.7 Downburst Wind Example

We now look at a pair of uniformly modulated processes that model wind fluctuations as described in [41]. Specifically we look at the unique uniformly modulating the stationary vector-valued process  $\{[X_1(t), X_2(t)]^T\}$  defined by the spectral density matrix

$$\begin{aligned} \mathbf{S}(f) &= \begin{bmatrix} S_{11}(f) & S_{12}(f) \\ S_{21}(f) & S_{22}(f) \end{bmatrix} \\ &= \begin{bmatrix} S(f) & S(f)\gamma(f) \\ S(f)\gamma(f) & S(f) \end{bmatrix} \end{aligned}$$

where

$$S(f) = \frac{4\sigma_u^2 L_u^z / U}{[1 + 70.8(fL_u^z / U)^2]^{5/6}}$$

and

$$\gamma(f) = \exp\left(-\frac{k_z |f| \Delta z}{U}\right).$$

$U$  is the mean wind speed which in this example is taken to be 40 m/s.  $\sigma_u$  represents the standard deviation in the wind fluctuation and assigned the value 6 m/s.  $L_u^z$  is a length scale with the value of 80 m,  $k_z$  is a decay factor that has been assigned 2,  $\Delta z$  represents the distance between two locations and is assigned  $\Delta z = 30$  m.

The nonstationary downburst wind process is the uniformly modulated bivariate process  $\{[Z_1(t), Z_2(t)]^T \equiv [A_1(t)X_1(t), A_2(t)X_2(t)]^T\}$ , where the modulating functions are of the form

$$A_j(t) = \alpha_j t^{\beta_j} e^{-\lambda_j t}$$

with  $\alpha_j = \lambda_j^{\beta_j} / (\beta_j^{\beta_j} e^{-\beta_j})$  and a unity maximum value at  $t_{\max,j} = \beta_j / \lambda_j$ . In this example we set  $\beta_1 = 3$  and  $\beta_2 = 4$ . The modulating functions' maximum values are set to occur at  $t_{\max,1} = 2500$  s and  $t_{\max,2} = 3000$  s. Parameters  $\alpha_j$  and  $\lambda_j$ ,  $j = 1, 2$ , follow from the above constants.

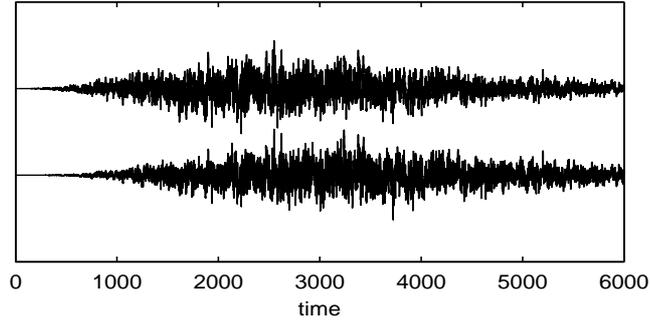


Figure 7.6: Realisations of the bivariate downburst wind UMP  $\{Z_1(t)\}$  (top) and  $\{Z_2(t)\}$  (bottom).

It is shown in Appendix D that the characteristic width of the family  $\mathcal{F}_j = \{A_j(t)e^{i2\pi ft}\}$ , with modulating function  $A_j(t) = \alpha_j t^{\beta_j} e^{-\lambda t}$ , takes the form

$$B_{\mathcal{F}} = \left[ \frac{1}{2\pi} \alpha_j (\beta_j!) \lambda_j^{1-\beta_j} \text{beta} \left( 1, \frac{\beta_j - 1}{2} \right) \right]^{-1}$$

where  $\text{beta}(\cdot, \cdot)$  is the standard beta function. This gives  $B_{\mathcal{F}_1} = 1173$  s, and  $B_{\mathcal{F}_2} = 1381$  s.

We know that the coherence for a pair of uniformly modulated processes is independent of time and equal to the coherence for the stationary processes that are being modulated. As such the coherence for  $\{Z_1(t)\}$  and  $\{Z_2(t)\}$  is equal to the coherence for  $\{X_1(t)\}$  and  $\{X_2(t)\}$ , i.e.

$$\gamma^2(f) = \exp \left( -2 \frac{k_z |f| \Delta z}{U} \right).$$

Realisations of the uniformly modulated processes come from generating realisations of the processes  $\{X_1(t)\}$  and  $\{X_2(t)\}$  by the algorithm outlined in [12]. The algorithm requires the auto and cross-covariance sequences for processes  $\{X_1(t)\}$  and  $\{X_2(t)\}$ . These were formulated by sampling the SDF matrix defined above on a frequency grid, and then performing an inverse Fourier transform. Realisations were generated at a sampling rate of  $\Delta = 0.5$  s for  $N = 12000$  points.

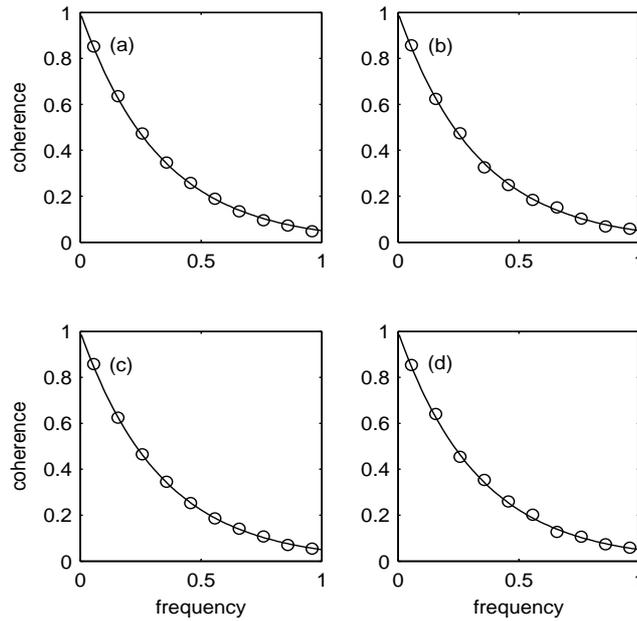


Figure 7.7: Debiased mean coherence estimates (rings) for 100 simulations of the downburst wind process, plotted as a function of frequency at four different points in time. The true coherence is plotted with the solid line.

### 7.7.1 Coherence Estimator Details

To obtain accurate estimates of the ESDF matrix of  $\{[Z_1(t), Z_2(t)]^T\}$  (and hence accurate estimates of the coherence), we require to satisfy the condition that for the choice of discrete-time filter  $\{g_u\}$ ,  $B_g \ll B_{\mathbf{Z}} = \min\{B_{Z_1}, B_{Z_2}\}$ . Further to this however is that we also require the magnitude squared of the Fourier transform of  $\{g_u\}$  be a pseudo-delta function with respect to the evolutionary spectra's frequency dependency. This gives us a typical trade of problem of how to keep the filter sufficiently concentrated in both time and frequency.

It is known — see Section 2.7.3 — that the sequence of finite length  $N_S$  that maximises energy concentration within a specified bandwidth  $[-W, W]$  is the zero'th order discrete prolate spheroidal sequence, denoted  $\{v_u^0(N_S, W), u = 0, \dots, N_S - 1\}$ . In this example we let  $g_u = v_u^0(N_S, W)$  with  $N_S = 2\nu + 1$ ,

$\nu = 50$ , and  $W = 3/101$ .  $B_g$  is calculated using a Riemann sum approximation to (6.20):

$$B_g = \Delta \sum_{u=0}^{N_S-1} |g_u| |(u - \nu)\Delta| = 24, \quad (7.20)$$

hence  $B_g/B_Z \leq 24/1173$ . In addition to this nearly 100% of  $\{g_u\}$  energy is contained in the frequency band  $[-0.0297/\Delta, 0.0297/\Delta] = [-0.0594, 0.0594]$  and as such we consider the Fourier transform  $\Gamma(f)$  to be a pseudo-delta function with respect to  $\mathcal{S}(f)$  that is defined on the frequency domain  $[-1, 1]$ . Using a rectangular smoothing widow of size  $N_B = 321$  gives approximately 8 complex degrees of freedom and we can be satisfied that  $\Delta N_P = \Delta(N_P + N_S - 1) = 0.5 \times 421$  is small enough in comparison to  $B_Z$  as to assume stationarity across the signal portion of length  $N_P$  used for analysis.

In Figure 7.7 is plotted the debiased mean (over 100 simulated processes) according to the predicted Goodman distribution as a function of frequency for four arbitrary points in time. The true coherence  $\gamma^2(f)$  is also plotted and these show excellent agreement. Figure 7.8 are Q-Q plots for the 100 coherence samples for the predicted Goodman distribution, at the same four arbitrary times. Each plot has four curves for four different frequencies. Once again the agreement is excellent.

## 7.8 Concluding Remarks

We have shown that by considering the mathematical similarities between the Priestley evolutionary spectral estimator and the temporally smoothed wavelet spectrum, it is possible to show that the distributional results for TWCOH, as derived under the stationary assumption in Chapter 5, are applicable for both methods of coherence estimation when applied to slowly-varying discrete-time semi-stationary processes.

While the Goodman distribution for TWCOH in the stationary case is only asymptotically appropriate, with the use of toy examples and a practical engineering example of the uniformly modulated processes it has been shown

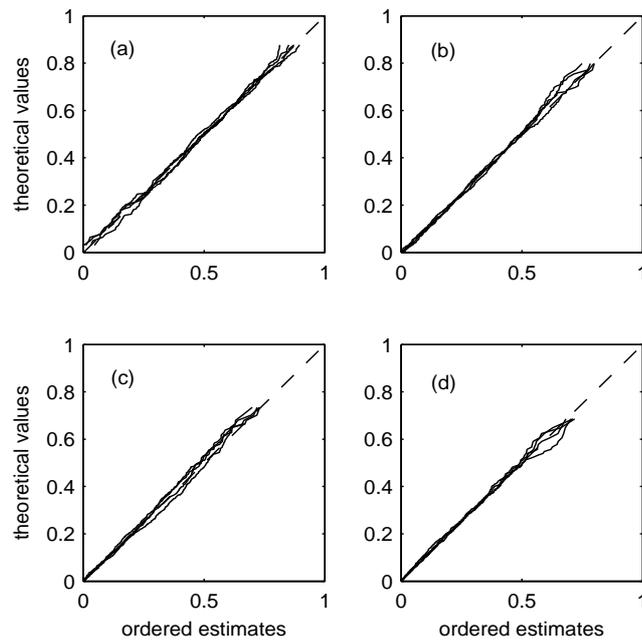


Figure 7.8: Q-Q plots at four separate frequencies. Each plot has four curves for four different time points.

that the approximations required in the nonstationary case are valid and that experimental results show an excellent match to the theory.

## Chapter 8

# Time-Varying-Coherence Models

In this chapter we turn our attention to the Wold-Cramér nonstationary model presented in Section 6.4. Via the work of Mélard and Herteleer-de Schutter [64], in which a multivariate framework was presented, it will be shown that it is possible to construct nonstationary bivariate models in which a coherence measure can be defined that displays dependence in time. By considering a simple regression-type model constructed from modulated stationary processes, a range of coherence structures can be achieved. In this chapter we will look at applying the TWCOH estimator to these models with the hope of showing the distributional results for TWCOH first presented for the stationary case in Chapter 5, and then extended to the semi-stationary case in Chapter 7, are again appropriate for these more complex coherence structures. The following work can be found in Cohen and Walden [16].

### 8.1 Multivariate Wold-Cramér Model

Our entry into the multivariate Wold-Cramér model for nonstationary processes is via the work of Mélard and Herteleer-de Schutter [64] and follows naturally from the univariate model set out in Section 6.4. Let  $\{\mathbf{Z}_t, t \in \mathbb{Z}\}$  be

a nonstationary  $p$ -variate purely non-deterministic random process. Without loss of generality we assume  $E\{\mathbf{Z}_t\} = \mathbf{0}$ . Then there exists a one-sided infinite moving-average representation or Wold-Cramér decomposition of the process given by

$$\mathbf{Z}_t = \sum_{\tau=0}^{\infty} \tilde{\mathbf{h}}_{t,\tau} \tilde{\boldsymbol{\xi}}_{t-\tau}, \quad (8.1)$$

where  $\boldsymbol{\Sigma}_t$ , the covariance matrix of the innovation  $\tilde{\boldsymbol{\xi}}_t$  has maximum rank for each  $t$ . The normalised innovations are given by  $\boldsymbol{\xi}_t = \boldsymbol{\Sigma}_t^{-1/2} \tilde{\boldsymbol{\xi}}_t$  so that alternatively

$$\mathbf{Z}_t = \sum_{\tau=0}^{\infty} \mathbf{h}_{t,\tau} \boldsymbol{\xi}_{t-\tau}, \quad (8.2)$$

where  $\mathbf{h}_{t,\tau} = \tilde{\mathbf{h}}_{t,\tau} \boldsymbol{\Sigma}_{t-\tau}^{1/2}$  and now  $\{\boldsymbol{\xi}_t\}$  is a stationary white noise process with identity covariance matrix. Further,  $\mathbf{h}_{t,\tau}$  is a matrix that can have non-zero off-diagonal elements and for which  $\sum_{\tau} |(\mathbf{h}_{t,\tau})_{lm}| < \infty$  for all  $t$  and  $l, m = 1, \dots, p$ .  $\mathbf{Z}_t$  also has the representation

$$\mathbf{Z}_t = \int_{-1/2}^{1/2} e^{i2\pi ft} \mathbf{H}_t(f) d\boldsymbol{\zeta}_{\boldsymbol{\xi}}(f),$$

where  $\mathbf{S}_{\boldsymbol{\xi}}(f)df = E\{d\boldsymbol{\zeta}_{\boldsymbol{\xi}}(f)d\boldsymbol{\zeta}_{\boldsymbol{\xi}}^H(f)\}$  and  $\mathbf{H}_t(f) = \sum_{\tau} \mathbf{h}_{t,\tau} \exp(i2\pi f\tau)$  is a matrix that can have non-zero off-diagonal elements. Since the innovation process in (8.2) is unit variance white noise it has SDF matrix  $\mathbf{S}_{\boldsymbol{\xi}}(f) = \mathbf{I}_p$  so the Wold-Cramér spectrum of  $\{\mathbf{Z}_t\}$  is, by analogy to (7.4), given by

$$\mathbf{S}_t(f) = \mathbf{H}_t(f) \mathbf{H}_t^H(f). \quad (8.3)$$

The coherence between two component processes of  $\{\mathbf{Z}_t\}$ ,  $Z_{l,t}$  and  $Z_{m,t}$  say, will be defined as

$$\gamma_{lm,t}^2(f) = \frac{|S_{lm,t}(f)|^2}{S_{ll,t}(f)S_{mm,t}(f)}, \quad (8.4)$$

where  $S_{lm,t}(f)$  is the  $(l, m)$ th element of  $\mathbf{S}_t(f)$ .

## 8.1.1 Time-and-frequency-Dependent Coherence

Consider the example of a simple regression model taken from [64]. Let

$$Z_{1,t} = \sum_{\tau=0}^{\infty} b_{t,\tau} \xi_{t-\tau}, \quad (8.5)$$

$$Z_{2,t} = \beta_t Z_{1,t} + d_t \epsilon_t. \quad (8.6)$$

Here  $\{Z_{1,t}\}$  is a purely non-deterministic process with normalised innovations  $\{\xi_t\}$ ; it has the Wold-Cramér decomposition (6.16), or (8.2) with  $p = 1$ .  $\{\epsilon_t\}$  is a sequence of uncorrelated variables with zero mean and unit variance. The assumption is also made that  $\{Z_{1,t}\}$  and  $\{\epsilon_t\}$  are independent processes, and that  $d_t$  and  $\beta_t$  are non-random, real-valued finite functions of time. (Particular forms for these functions will be introduced subsequently.) From (8.1) with  $p = 2$  the Wold-Cramér decomposition of  $\{\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}]^T\}$  may be written

$$\mathbf{Z}_t = \tilde{\boldsymbol{\xi}}_t + \sum_{\tau=1}^{\infty} \tilde{\mathbf{h}}_{t,\tau} \tilde{\boldsymbol{\xi}}_{t-\tau}$$

where [64]

$$\tilde{\mathbf{h}}_{t,\tau} = \frac{b_{t,\tau}}{b_{t-\tau,0}} \begin{bmatrix} 1 & 0 \\ \beta_t & 0 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_t^{1/2} = \nu_t \begin{bmatrix} b_{t,0}^2 + b_{t,0}d_t & \beta_t b_{t,0}^2 \\ \beta_t b_{t,0}^2 & \beta_t^2 b_{t,0}^2 + d_t^2 + b_{t,0}d_t \end{bmatrix},$$

with  $\nu_t = [\beta_t^2 b_{t,0}^2 + (b_{t,0} + d_t)^2]^{-1/2}$ . The calculation of  $\mathbf{h}_{t,\tau} = \tilde{\mathbf{h}}_{t,\tau} \boldsymbol{\Sigma}_{t-\tau}^{1/2}$ , followed by Fourier transformation, gives

$$\begin{aligned} \begin{bmatrix} H_{11,t}(f) \\ H_{12,t}(f) \end{bmatrix} &= \sum_{\tau=0}^{\infty} \nu_{t-\tau} b_{t,\tau} e^{-i2\pi f\tau} \begin{bmatrix} b_{t-\tau,0} + d_{t-\tau} \\ b_{t-\tau,0} \beta_{t-\tau} \end{bmatrix} \\ \begin{bmatrix} H_{21,t}(f) \\ H_{22,t}(f) \end{bmatrix} &= \beta_t \begin{bmatrix} H_{11,t}(f) \\ H_{12,t}(f) \end{bmatrix} + \nu_t d_t \begin{bmatrix} -b_{t,0} \beta_t \\ b_{t,0} + d_t \end{bmatrix}, \end{aligned} \quad (8.7)$$

where  $H_{lm,t}(f) = (\mathbf{H}_t(f))_{lm}$ . Putting  $\mathbf{H}_t(f)$  into (8.3) means that the time-frequency coherence (8.4) becomes

$$\gamma_t^2(f) = 1 - \frac{S_t^0(f)}{S_{11,t}(f)S_{22,t}(f)} \quad (8.8)$$

where

$$S_t^0(f) = \nu_t^2 d_t^2 |(b_{t,0} + d_t)H_{11,t}(f) + b_{t,0}\beta_t H_{12,t}(f)|^2 \quad (8.9)$$

$$S_{11,t}(f) = |H_{11,t}(f)|^2 + |H_{12,t}(f)|^2 \quad (8.10)$$

$$S_{22,t}(f) = \beta_t^2 S_{11,t}(f) + d_t^2 + 2\nu_t d_t \beta_t \times [(b_{t,0} + d_t)\operatorname{Re}\{H_{12,t}(f)\} - b_{t,0}\beta_t \operatorname{Re}\{H_{11,t}(f)\}]. \quad (8.11)$$

So the model (8.5) and (8.6) results in coherence that is, most generally, both time and frequency-dependent, in contrast to the oscillatory processes (7.5). This comes about through the introduction of off-diagonal terms in  $\mathbf{H}_t(f)$  in contrast to (7.3). The frequency variation of coherence for this model is determined entirely by  $H_{11,t}(f)$  and  $H_{12,t}(f)$ .

### 8.1.2 Time-Dependent Coherence

Mélard and Herteleer-de Schutter provided the general model presented above, however determining the entries of the matrix  $\mathbf{H}_t(f)$  requires an infinite sum and consequently we look to simplify the model further. Suppose the non-deterministic process  $\{Z_{1,t}\}$  is in fact just a modulation of the white noise process  $\{\xi_t\}$ , i.e.,

$$Z_{1,t} = g_t \xi_t, \quad (8.12)$$

$$Z_{2,t} = \beta_t Z_{1,t} + d_t \epsilon_t, \quad (8.13)$$

for some  $g_t$ . The nonstationary process  $\{Z_{1,t}\}$  has the Wold-Cramér representation (8.5) with  $b_{t,\tau} = g_t$  when  $\tau = 0$ , and zero otherwise. Then,

$$\begin{bmatrix} H_{11,t}(f) \\ H_{12,t}(f) \end{bmatrix} = \nu_t g_t \begin{bmatrix} g_t + d_t \\ g_t \beta_t \end{bmatrix}, \quad (8.14)$$

where  $\nu_t = [\beta_t^2 g_t^2 + (g_t + d_t)^2]^{-1/2}$ . Expressions for  $S_t^0(f)$ ,  $S_{11,t}(f)$  and  $S_{22,t}(f)$  can be found using (8.9), (8.10) and (8.11) respectively, along with (8.14). We find  $S_t^0(f) = d_t^2 g_t^2$ ,  $S_{11,t}(f) = g_t^2$  and  $S_{22,t}(f) = \beta_t^2 g_t^2 + d_t^2$ . From (8.8)

$$\gamma_t^2(f) = 1 - \frac{d_t^2}{\beta_t^2 g_t^2 + d_t^2}. \quad (8.15)$$

As expected this coherence varies in time but is frequency invariant. If  $d_t = 0$ , then  $Z_{1,t} = g_t \xi_t$  and  $Z_{2,t} = \beta_t g_t \xi_t$ , and the coherence takes a value of unity. Alternatively, by letting  $\beta_t = 0$  we have  $Z_{1,t} = g_t \xi_t$  and  $Z_{2,t} = d_t \epsilon_t$  (with  $\xi_t$  and  $\epsilon_t$  being independent white noise processes), and  $\gamma_t^2(f) = 0$ .

### 8.1.3 Introducing Frequency-Dependence

In Section 8.3.3 we look at examples of time-and-frequency-dependent coherence, with processes modelled as follows. We let  $\{Z_{1,t}\}$  be a uniform modulation of a moving average process of order  $q$ , or MA( $q$ ) process,

$$Z_{1,t} = g_t(\theta_0 \xi_t + \theta_1 \xi_{t-1} + \dots + \theta_q \xi_{t-q}).$$

In terms of (8.5),  $b_{t,\tau} = g_t \theta_\tau$  for  $\tau = 0, \dots, q$ , and zero otherwise. The frequency variation of coherence for this model is determined entirely by  $H_{11,t}(f)$  and  $H_{12,t}(f)$  given by (8.7). Some quite complex evolution of coherence in the time-frequency plane can be achieved, while still being straight-forward to compute. In the examples in Section 8.3.3 we take  $\theta_\tau = (1/5) \cdot 2^{(1-\tau)}$  for  $\tau = 0, \dots, 10$ , and zero otherwise.

## 8.2 Sigma-Oscillatory Processes

Suppose  $g_t$  is chosen so that  $Z_{1,t} \equiv g_t \xi_t$  is an UMP, then  $\{Z_{1,t}\}$  is by definition an oscillatory process. Is  $\{Z_{2,t}\}$  given by (8.6) also an oscillatory process? Suppose  $\beta_t$  and  $d_t$  are chosen such that  $\beta_t Z_{1,t} = \beta_t g_t \xi_t$  and  $d_t \epsilon_t$  are UMPs, so that  $\{Z_{2,t}\}$  is a sum of two independent oscillatory processes. It is known [97] that the sum of two independent oscillatory processes is not necessarily oscillatory (the class of oscillatory processes is not closed with respect to the sum of independent elements); a sufficient condition for the sum to again be oscillatory is given in [97] and in the present example this turns out to be that

$$d_t^2 = \beta_t^2 g_t^2, \quad (8.16)$$

which is the same as requiring that the time-varying variances of the two oscillatory components of  $\{Z_{2,t}\}$  be equal. Also, the insertion of (8.16) into (8.15) means that the coherence becomes simply  $\gamma_t^2(f) = 1/2$ , so that the coherence becomes time-invariant as for oscillatory processes in (7.5).

A process  $\{Z_{1,t}\}$  is called *sigma-oscillatory* [2] if it can be expressed as  $Z_{1,t} = \sum_{i=1}^{L_1} U_{i,t}$ , where each  $\{U_{i,t}\}$  is an oscillatory process with representation — see (6.10),

$$U_{i,t} = \int_{-1/2}^{1/2} A_{i,t}(f) e^{i2\pi ft} d\zeta_{U,i}(f),$$

and are pairwise independent.  $\{Z_{1,t}\}$  is not necessarily oscillatory. Consider a second sigma-oscillatory process  $Z_{2,t} = \sum_{j=1}^{L_2} V_{j,t}$ , where

$$V_{j,t} = \int_{-1/2}^{1/2} B_{j,t}(f) e^{i2\pi ft} d\zeta_{V,j}(f).$$

The pairwise independence of the oscillatory components and the orthogonality of the increments of  $\{\zeta_{U,i}(f)\}$  and  $\{\zeta_{V,j}(f)\}$  means that the time-varying SDF

of  $\{Z_{1,t}\}$  is given by

$$S_{11,t}(f) = \sum_{i=1}^{L_1} |A_{i,t}(f)|^2 S_{U,i}(f) \quad (8.17)$$

where  $S_{U,i}(f)df = E\{|d\zeta_{U,i}(f)|^2\}$ , and similarly for  $\{Z_{2,t}\}$ . The time-varying cross-spectrum for  $\{Z_{1,t}\}$  and  $\{Z_{2,t}\}$  is

$$S_{12,t}(f) = \sum_i \sum_j A_{i,t}(f) B_{j,t}^*(f) S_{UV,ij}(f)$$

where  $S_{UV,ij}(f)df = E\{d\zeta_{U,i}(f)d\zeta_{V,j}^*(f)\}$ . The coherence is then given by (7.1).

Consider the model (8.5) and (8.6), and suppose  $\{Z_{1,t}\}$  is oscillatory, or sigma-oscillatory with  $L_1 = 1$ ; its time-varying SDF is given by (8.17). For modulating functions  $\beta_t$  and  $d_t$  which are non-negative and finite valued,  $Z_{2,t} = \beta_t Z_{1,t} + d_t \epsilon_t$  is a sigma-oscillatory process with  $L_2 = 2$  and  $B_{1,t}(f) \equiv \beta_t A_{1,t}(f)$  and  $B_{2,t}(f) \equiv d_t$ . The time-frequency coherence reduces to

$$\begin{aligned} \gamma_t^2(f) &= \frac{\beta_t^2 |A_{1,t}(f)|^2 S_{U,1}(f)}{\{\beta_t^2 |A_{1,t}(f)|^2 S_{U,1}(f) + d_t^2\}} \\ &= 1 - \frac{d_t^2}{\beta_t^2 |A_{1,t}(f)|^2 S_{U,1}(f) + d_t^2}. \end{aligned} \quad (8.18)$$

When  $\{Z_{1,t}\}$  is a uniformly modulated uncorrelated processes,  $Z_{1,t} = g_t \xi_t$ , then  $A_{1,t}(f) = g_t$  and  $S_{U,1}(f) = 1$  and (8.18) reduces to (8.15), as it should.

### 8.3 Simulations and Results

We now apply the TWCOH estimator  $\hat{\gamma}^2(a, b)$  of Chapter 5 to a variety of examples of the regression model (8.5) and (8.6). We take  $\hat{\gamma}^2(a, b)$  as an estimator of  $\gamma_b^2(1/a)$ , the true (model) coherence (8.8) at discrete time  $b$  and continuous frequency  $1/a$ . We are also interested in testing the suitability of the Goodman distribution (5.38) for  $\hat{\gamma}^2(a, b)$  in this nonstationary setting. Clearly the success of such a scheme on any particular data will depend at least on our

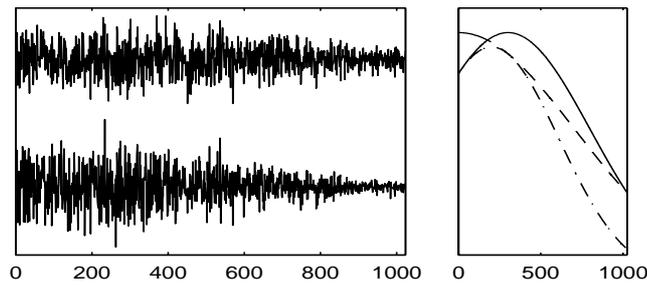


Figure 8.1: Left: Realisation of the bivariate process. Right: Modulating functions  $g_t$  (solid line),  $\beta_t$  (dashed) and  $d_t$  (dash-dot).

choice of  $N_P$ , which determines time-localization. All simulated white noise processes are Gaussian distributed.

### 8.3.1 Constant Coherence

For the model defined by (8.12) and (8.13) we showed in Section 8.1.1 that if  $g_t$  is chosen so that  $Z_{1,t} \equiv g_t \xi_t$  is an UMP, then  $\{Z_{1,t}\}$  is by definition an oscillatory process, and if (8.16) holds, then  $\{Z_{2,t}\}$  is also an oscillatory process and the coherence is constant, i.e.,  $\gamma_b^2(1/a) = 1/2$ . We used modulating functions  $g_t = G_{300,5 \cdot 10^5,t}$  and  $\beta_t = G_{0,10^6,t}$  and  $d_t = \beta_t g_t$ , using the parametric form  $G_{t_0,\lambda;t} = \exp[-(t - t_0)^2/\lambda]$ . Realisations of the bivariate process and modulating functions are shown in Figure 8.1.

Simulation results are shown in Figure 8.2 for scales  $a = 6$  and  $a = 18$ , with smoothing parameter  $\kappa_0 = 12$  ( $\varphi = 10$ ). Horizontal bars show the block size,  $N_P$ , for the stated value of scale  $a$ . Figure 8.2(b) gives Q-Q plots for  $a = 6$  for  $b = 450$  and  $550$ , and Figure 8.2(d) does likewise for  $a = 18$ . The estimation results are excellent despite the different modulations used.

### 8.3.2 Time-Dependent Coherence

Here we examine three models with time-dependent coherence. In all these cases  $\gamma_b^2(1/a)$  is defined by  $\gamma_t^2(f)$  given by (8.15).

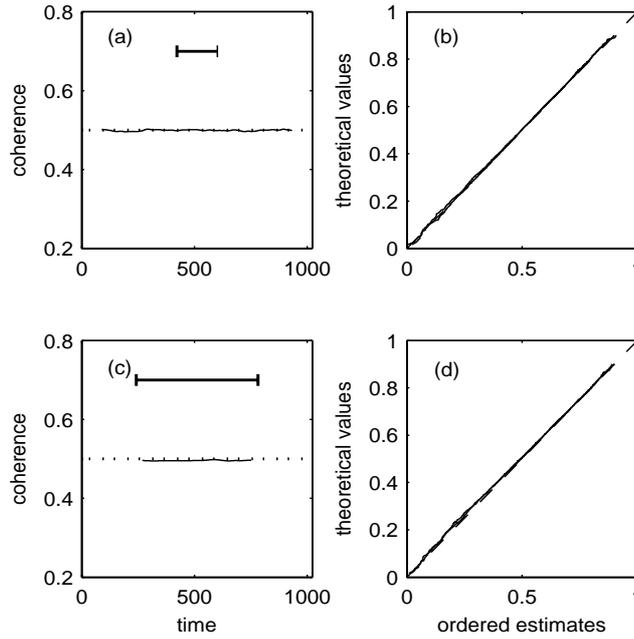


Figure 8.2: Constant coherence example. Left column: Debiased mean coherence estimates from 5000 simulations for scales (a)  $a = 6$  (b)  $a = 18$  using smoothing with  $\kappa_0 = 12$  ( $\varphi = 10$ ). Dotted lines show true coherence values. Horizontal bars show the block size,  $N_P$ , for the stated value of scale  $a$ . Right column: Corresponding Q-Q plots at time points  $b = 450$  and  $550$  for scales (b)  $a = 6$ , (d)  $a = 18$ .

To have coherence that varies with time we generalise  $d_t$  in (8.13) to be a finite non-negative function which need not satisfy (8.16).  $\{Z_{1,t}\}$  remains oscillatory, and while  $\{\beta_t Z_{1,t}\}$  and  $\{d_t \epsilon_t\}$  are both UMPs and thus individually oscillatory,  $\{Z_{2,t}\}$  will merely be sigma-oscillatory. In our first example we leave  $g_t$  and  $\beta_t$  as in Section 8.3.1, and let  $d_t = G_{900,5 \cdot 10^5;t}$ . Results are shown in Figure 8.3 (in same format as Figure 8.2). We see that TWCOH for  $a = 6$  successfully tracks the time-varying coherence even though  $\{Z_{2,t}\}$  is not now oscillatory. For  $a = 18$ , TWCOH and the Q-Q plot would be satisfactory for any practical purposes.

We can generalise further by releasing the restriction on  $\{Z_{1,t}\}$  and  $\{d_t \epsilon_t\}$  being UMPs, and allow them to be more general modulated stationary processes. Examples two and three which follow show that although neither  $\{Z_{1,t}\}$

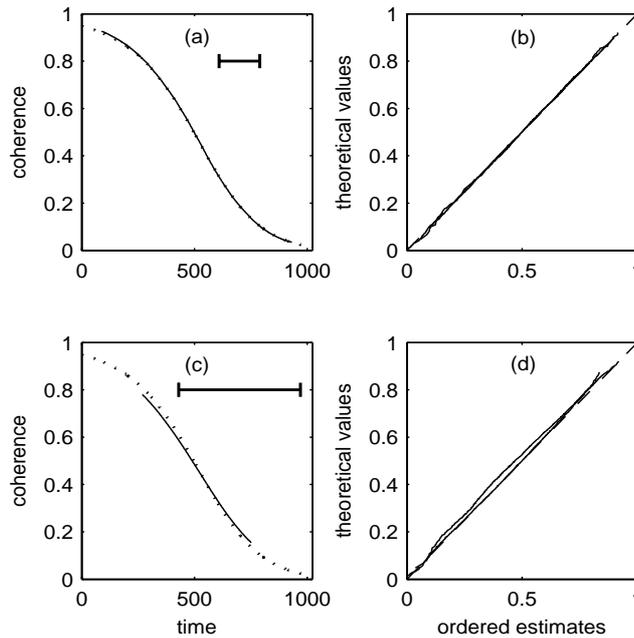


Figure 8.3: As for Figure 8.2, but for the first example of Section 8.3.2.

nor  $\{Z_{2,t}\}$  are oscillatory, TWCOH is still able to correctly track the time-evolving coherence. For the model defined by (8.12) and (8.13) we took  $g_t$  and  $d_t$  to be cosine functions oscillating with a periodicity of  $T_g$  and  $T_d$  respectively, while  $\beta_t = G_{0,10^6;t}$  as before. With these examples we also explore in more detail the effect of the temporal smoothing contained in TWCOH on the effectiveness of this estimator. It seems logical that a WCOH measure based on a temporal smoothing will lack resolution in the time domain, hence we expect the performance of the estimator to lessen the more rapidly the true coherence varies.

In example two we set  $T_g = 5000$ ,  $T_d = 6000$ . This model was purposefully designed to have slowly varying coherence relative to the observation period. Five thousand realisations of the bivariate process were generated. Results are presented in Figure 8.4 and again show a good match for the (debiased) WCOH estimator with the true coherence, and for ordered sample values and theoretical quantiles of the Goodman distribution, (with  $\varphi = 10$  complex

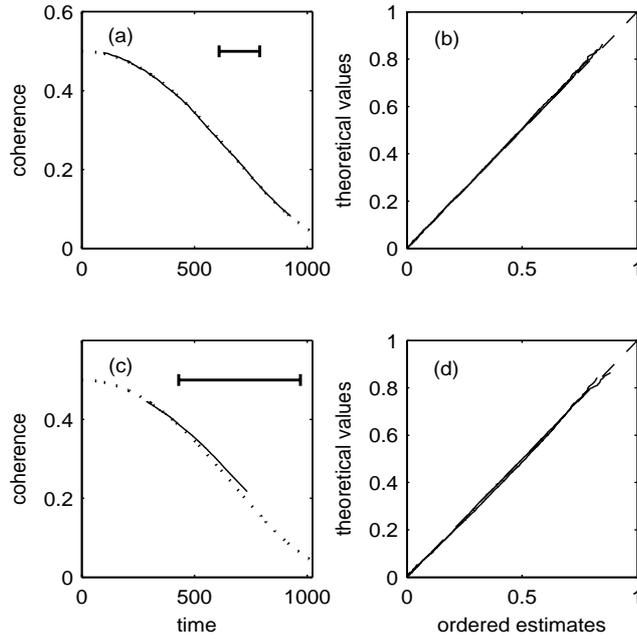


Figure 8.4: As for Figure 8.2, but for the second example of Section 8.3.2.

degrees of freedom), via the Q-Q plots.

In the third example for time-dependent coherence we utilise a model in which the coherence varies at a faster rate with respect to the observation period, and is a nonmonotonic function of time within the time window. We keep  $\beta_t$  unchanged, but set  $T_g = 400$  and  $T_d = 800$ . A realisation of the bivariate process is plotted in Figure 8.5. Debiased mean coherence results are shown for scale  $a = 6$  in Figure 8.6(a) when using a smoothing factor of  $\kappa_0 = 12$  which gives  $\varphi = 10$  complex degrees of freedom. At scale 6 the time resolution of TWCOH is inadequate for the variation in the coherence — compare the horizontal bar in Figure 8.6(a) and the curvature of the true coherence. In Figure 8.6(b) we reduce the smoothing factor to  $\kappa_0 = 6$ , and whilst the degrees of freedom are halved to  $\varphi = 5$ , the reduction in temporal smoothing allows the estimator to more effectively track the modulation of the signals and of the coherence.

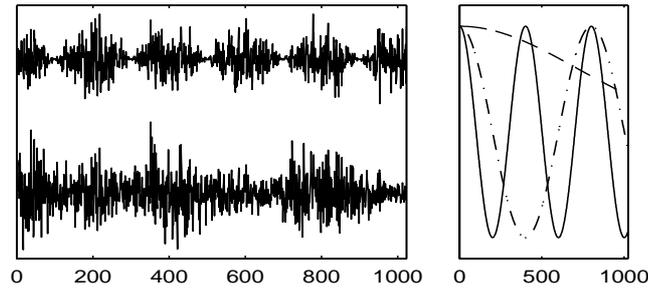


Figure 8.5: Left: Realisations of the processes producing the time-dependent coherence of Figure 8.6. Right: Modulating functions  $g_t$  (solid line),  $\beta_t$  (dashed) and  $d_t$  (dash-dot).

### 8.3.3 Time-and-Frequency-Dependent Coherence

Here we use the model of Section 8.1.3. Our first example of time-and-frequency-dependent coherence uses modulating functions  $g_t = G_{300,5 \cdot 10^5;t}$ ,  $\beta_t = G_{0,10^6;t}$  and  $d_t = G_{900,5 \cdot 10^5;t}$  so that  $\{Z_{1,t}\}$ ,  $\{\beta_t Z_{1,t}\}$  and  $\{d_t \epsilon_t\}$  are all oscillatory, and  $\{Z_{2,t}\}$  is sigma-oscillatory. The time-frequency surface specified by (8.8)-(8.11) is shown in Figure 8.7; frequencies  $f = 1/6$  and  $f = 1/18$ , corresponding to scales  $a = 6$  and  $18$ , are highlighted. A realisation of the bivariate process is plotted in Figure 8.8. For this example the coherence surface is quite rapidly varying with time (as in the first example of Section 8.3.2, Figure 8.3), but as can be seen in Figure 8.7 this is only slowly modified by frequency/scale. Results are presented in Figure 8.9 and show a good match for the (debiased) WCOH estimator with the true coherence, while the Q-Q plots are also satisfactory.

For our second example of time-and-frequency-dependent coherence we changed the modulating functions  $g_t$  and  $d_t$  to be cosines with periods  $T_g = 6000$  and  $T_d = 5000$ , as in the second example of Section 8.3.2. The time-frequency surface for this example is shown in Figure 8.10; the time variation of coherence is now modified rapidly as frequency/scale changes. Results are presented in Figure 8.11 and show a good match with the theoretical ideal when  $a = 6$ , but at  $a = 18$  the frequency-related change in coherence appears

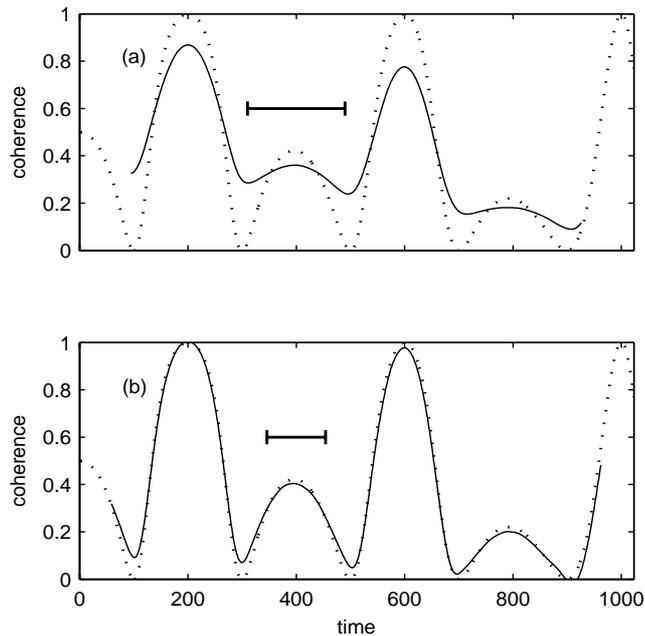


Figure 8.6: Debiased mean coherence estimates from 5000 simulations for third example of time-dependent coherence for scale  $a = 6$  with smoothing parameters of (a)  $\kappa_0 = 12$  ( $\varphi = 10$ ) and (b)  $\kappa_0 = 6$  ( $\varphi = 5$ ).

sufficient to cause coherence to be slightly overestimated for all times where coherence is estimable, (and a minor perturbation in the corresponding Q-Q plot). The coherence surface is more rapidly varying at  $a = 18$  than at  $a = 6$  and the slight error is probably due to the support of the Morlet wavelet in frequency/scale — see Figure 3.2 — becoming more apparent in such a case.

## 8.4 Concluding Remarks

Using the modulated stationary processes, a bivariate nonstationary model that can exhibit time-dependent and time-and-frequency-dependent coherence structures has been presented. The TWCOH estimator has been shown to be accurate in estimating coherence provided the time variation of the coherence is roughly linear over the extent of  $N_P$ . If there is notable curvature, then the smoothing parameter  $\kappa_0$  (and consequently degrees of freedom) must be re-

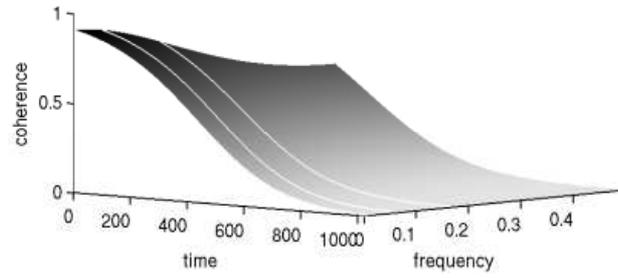


Figure 8.7: First example of time-and-frequency-dependent coherence plotted as a surface in time-frequency space. The white lines are for  $f = 1/6$  and  $1/18$ .

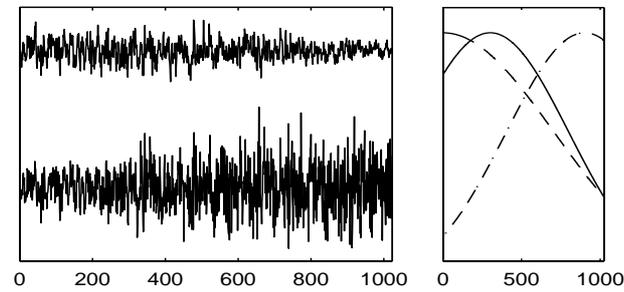


Figure 8.8: Left: Realisations of the processes producing the time-and-frequency-dependent coherence of Figure 8.7. Right: Modulating functions  $g_t$  (solid line),  $\beta_t$  (dashed) and  $d_t$  (dash-dot).

duced appropriately. Such iterative refinement is similar to the idea of ‘window closing’ in ordinary spectrum analysis [73, p. 276]. Rapid variation of coherence with frequency can also be problematic due to the support of the Morlet wavelet. Nevertheless, it has been demonstrated that the TWCOH estimator can perform extremely well for coherence from nonstationary processes arising from modulated time series, and moreover its statistical properties (e.g., Goodman distribution, bias) may be calculated exactly as in the stationary case.

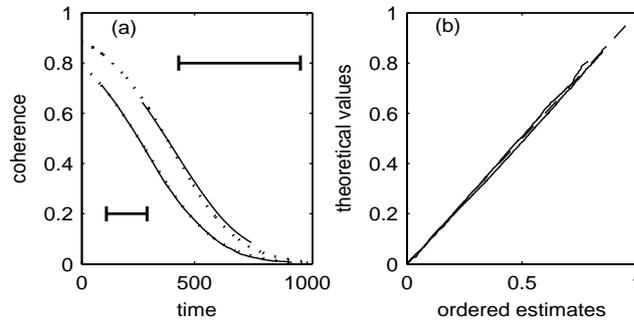


Figure 8.9: (a) Debiased mean coherence estimates from 5000 simulations for time-and-frequency-dependent coherence of Figure 8.7 for scales  $a = 18$  (top) and  $a = 6$  (bottom) using smoothing with  $\kappa_0 = 12$  ( $\varphi = 10$ ). Dotted lines show true coherence values. Horizontal bars show the block size,  $N_P$ , for  $a = 18$  (top) and  $a = 6$  (bottom). (b) Corresponding Q-Q plots for both scales at the single time point  $b = 512$ .

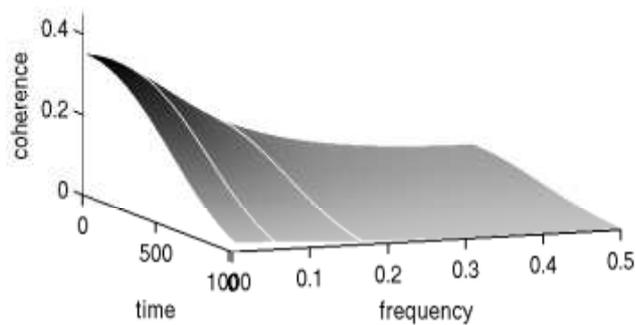


Figure 8.10: Second example of time-and-frequency-dependent coherence plotted as a surface in time-frequency space. The white lines are for  $f = 1/6$  and  $1/18$ .

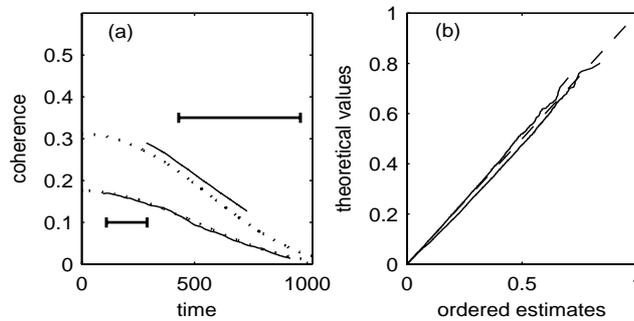


Figure 8.11: (a) Debiased mean coherence estimates from 5000 simulations for time-and-frequency-dependent coherence of Figure 8.10 for scales  $a = 18$  (top) and  $a = 6$  (bottom) using smoothing with  $\kappa_0 = 12$  ( $\varphi = 10$ ). Dotted lines show true coherence values. Horizontal bars show the block size,  $N_P$ , for  $a = 18$  (top) and  $a = 6$  (bottom). (b) Corresponding Q-Q plots for both scales at the single time point  $b = 512$ .

## Chapter 9

### Conclusion

The well recognised problem of finding an analytical statistical theory for wavelet coherence has been addressed. A statistical study for two methods of smoothing the estimator has been presented for the general class of improper (or proper) discrete-time complex-valued Gaussian SOS processes.

With the first method, in an analogous framework to multi-tapering, wavelet coherence was estimated using multiple orthogonal Morse wavelets. The second coherence estimator proposed used temporal smoothing together with a single Morlet wavelet. Under Gaussianity, the Goodman distribution has been shown, for large samples, to be appropriate for wavelet coherence in both cases. The true wavelet coherence value has been identified in terms of its frequency domain equivalent and degrees of freedom derived, providing a full parameterisation of the distribution. The theoretical results have been verified via simulations. An example of their use in testing for significant values of wavelet coherence has been demonstrated on a recorded data set.

For theoretical reasonings and application purposes, Gaussianity is by far the most common distribution assumed in the literature. It is therefore a reasonable but nonetheless restrictive assumption that has been placed on the random processes, and the statistical results presented here are reliant upon it. There is certainly scope for further work on the distribution of wavelet coherence for non-Gaussian processes. We note that the suitability of the Goodman

distribution in describing wavelet coherence is dependent on the CWT at a time-scale point being itself Gaussian. This follows naturally from the Gaussianity of the process, with the CWT (3.6) just being a sum of Gaussian variables. With the central limit theorem in mind, a Gaussian distributed CWT, and hence Goodman distributed coherence, may not necessarily be restricted to only Gaussian processes, although a Gaussian CWT for non-Gaussian random processes will only ever be an approximation. Further research will be needed to investigate when this approximation is appropriate.

Transporting the statistics in the stationary case to the nonstationary case required a careful consideration of the meaning and formulation of a time dependent spectral function. This meant particular focus was given to Priestley's evolutionary process and the Wold-Cramér nonstationary representation where time-varying spectral functions and coherence measures can be clearly defined. Although there are several ways of modelling nonstationary processes, these particular models include the important example of the UMP, a process with widespread applications.

Two methods of estimating these time-dependent spectra were discussed. The temporally smoothed wavelet spectrum has been shown to bear close resemblance to Priestley's own estimation procedure, allowing conditions to be constructed for its validity. It is clear that when these estimators are applied to a nonstationary process, careful consideration of the various estimator parameters (e.g. the filter width and smoothing parameter,) with respect to the characteristic width of the nonstationary process is necessary. It is recognised that the characteristic width is certainly a crude estimate for the interval of stationarity of a semi-stationary process, and a more rigorous formulation for an interval of stationarity would make for an interesting and useful field of research.

The concept of coherence for bivariate evolutionary nonstationary processes was discussed in detail. In such situations it has been shown that the coherence function, as in the stationary case, is invariant of frequency. For spectra that vary 'slowly' in time, termed semi-stationary, the derived statistics of the

temporally smoothed wavelet coherence estimator are appropriate. Further to this the similarities with Priestleys spectral estimator are exploited to derive distributional properties of the corresponding Priestley coherence estimator. The Goodman distribution was once again shown to be appropriate. The results were verified with examples of the uniformly modulated process.

Uniformly modulated processes, when used in a simple regression model, have been shown to exhibit coherence that varies with time, frequency, and time-and-frequency. The temporally smoothed Morlet wavelet coherence estimator has been applied to these processes and it has been shown that accurate coherence estimates can be achieved for each type of coherence with the Goodman distribution being appropriate.

The statistical analysis of nonstationary coherence given here is not one with asymptotic considerations. We have seen that such treatments in the nonstationary case require a more abstract setting. This work is concerned with a more practical approach to coherence estimations and while these results are only approximations in this setting, they show a remarkable robustness considering the relatively small sample sizes that are required for a time-localised coherence estimator. With a detailed discrete-time formulation the theoretical framework is easily applicable and is believed will provide a useful contribution to the ongoing work on nonstationary coherence.

# Appendix A

## Proof of Theorem 4.4.1

*Proof.* To aid notation and understanding we present the proof for the bivariate case  $\{\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}], t \in \mathbb{Z}\}$ . Extension to the general  $p$ -dimension case is straight forward. Processes are assumed to have bounded and continuous spectra, thus ruling out long-memory processes and processes with line spectra. Part 1 of the theorem concerns the large-sample approximation that  $\check{\mathbf{W}}_k$  is *proper*. To show this we need to show that all the relational values for the vector  $\check{\mathbf{W}}_k$  are equal to zero. This is equivalent to showing that  $E\{\check{\mathbf{W}}_k \check{\mathbf{W}}_k^T\} = \mathbf{0}$ .

We firstly assume  $a > 0$ . Using (4.8) to rewrite the last two components of  $\check{\mathbf{W}}_k$  in terms of the complex conjugates of the two processes, (4.2), (4.11) and (4.12) give

$$E\{\check{\mathbf{W}}_k \check{\mathbf{W}}_k^T\} = \frac{a_0}{N^2 \Delta} \sum_{q,s=0}^{(N/2)-1} E\{\boldsymbol{\chi}_{(q,s)}\} \Psi_{k,q}^+(a_0) \Psi_{k,s}^+(a_0) e^{i2\pi b_0(q+s)/N}, \quad (\text{A.1})$$

where

$$\boldsymbol{\chi}_{(q,s)} = \begin{bmatrix} X_{1,q}X_{1,s} & X_{1,q}X_{2,s} & X_{1,q}Y_{1,s} & X_{1,q}Y_{2,s} \\ X_{2,q}X_{1,s} & X_{2,q}X_{2,s} & X_{2,q}Y_{1,s} & X_{2,q}Y_{2,s} \\ Y_{1,q}X_{1,s} & Y_{1,q}X_{2,s} & Y_{1,q}Y_{1,s} & Y_{1,q}Y_{2,s} \\ Y_{2,q}X_{1,s} & Y_{2,q}X_{2,s} & Y_{2,q}Y_{1,s} & Y_{2,q}Y_{2,s} \end{bmatrix},$$

and

$$X_{l,q} = \Delta \sum_{n=0}^{N-1} Z_{l,n} e^{-i2\pi qn/N}, \quad Y_{l,q} = \Delta \sum_{n=0}^{N-1} Z_{l,n}^* e^{-i2\pi qn/N}.$$

Substituting in the expressions for  $X_{l,q}$  and  $Y_{l,q}$  we have

$$E \{ \chi_{(q,s)} \} = \Delta^2 \sum_{m,n=0}^{N-1} E \{ \boldsymbol{\eta}_{(m,n)} \} e^{-i2\pi(mq+ns)/N}, \quad (\text{A.2})$$

where

$$\boldsymbol{\eta}_{(m,n)} = \begin{bmatrix} Z_{1,m}Z_{1,n} & Z_{1,m}Z_{2,n} & Z_{1,m}Z_{1,n}^* & Z_{1,m}Z_{2,n}^* \\ Z_{2,m}Z_{1,n} & Z_{2,m}Z_{2,n} & Z_{2,m}Z_{1,n}^* & Z_{2,m}Z_{2,n}^* \\ Z_{1,m}^*Z_{1,n} & Z_{1,m}^*Z_{2,n} & Z_{1,m}^*Z_{1,n}^* & Z_{1,m}^*Z_{2,n}^* \\ Z_{2,m}^*Z_{1,n} & Z_{2,m}^*Z_{2,n} & Z_{2,m}^*Z_{1,n}^* & Z_{2,m}^*Z_{2,n}^* \end{bmatrix}.$$

By the spectral representation theorem — Theorem 2.3.1 — there exists an orthogonal increment process  $\{\zeta(f)\}$  defined on the interval  $[-f_N, f_N]$ , such that

$$Z_t = \int_{-f_N}^{f_N} e^{i2\pi ft} \Delta d\zeta(f). \quad (\text{A.3})$$

Suppose we associate  $\{\zeta_j(f)\}$  with the process  $\{Z_{j,t}\}$  and  $\{\zeta_l(f)\}$  with the process  $\{Z_{l,t}\}$  for  $j, l = 1, 2$ . For  $f, f' \in [-f_N, f_N]$

$$E \{ d\zeta_j(f) d\zeta_l^*(f') \} = \begin{cases} 0 & f \neq f' \\ S_{jl}(f) df & f = f' \end{cases} \quad (\text{A.4})$$

$$E \{ d\zeta_j(f) d\zeta_l(f') \} = \begin{cases} 0 & f \neq -f' \\ R_{jl}(f) df & f = -f' \end{cases}. \quad (\text{A.5})$$

We now use the spectral representation in (A.3) and the properties of the orthogonal increment processes in (A.4) and (A.5) to simplify our sum. We will demonstrate their use on two generic elements of the matrix  $\boldsymbol{\eta}_{(m,n)}$ , namely

$E \{Z_{j,m}Z_{l,n}^*\}$  and  $E \{Z_{j,m}Z_{l,n}\}$ . Using the spectral representation

$$E \{Z_{j,m}Z_{l,n}^*\} = \iint_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{i2\pi(fm-f'n)\Delta} E \{d\zeta_j(f)d\zeta_l^*(f')\}.$$

Integrating across  $f'$  and using (A.4),

$$E \{Z_{j,m}Z_{l,n}^*\} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{i2\pi f(m-n)\Delta} S_{jl}(f)df, \quad (\text{A.6})$$

and using a similar argument for  $E \{Z_{j,m}Z_{l,n}\}$  we obtain

$$E \{Z_{j,m}Z_{l,n}\} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{i2\pi f(m-n)\Delta} R_{jl}(f)df. \quad (\text{A.7})$$

So,

$$E \{\boldsymbol{\eta}_{(m,n)}\} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{i2\pi f(m-n)\Delta} \boldsymbol{\Phi}(f)df,$$

where

$$\begin{aligned} \boldsymbol{\Phi}(f) &= \begin{bmatrix} R_{11}(f) & R_{12}(f) & S_{11}(f) & S_{12}(f) \\ R_{21}(f) & R_{22}(f) & S_{21}(f) & S_{22}(f) \\ S_{11}^*(-f) & S_{12}^*(-f) & R_{11}^*(-f) & R_{12}^*(-f) \\ S_{21}^*(-f) & S_{22}^*(-f) & R_{21}^*(-f) & R_{22}^*(-f) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}(f) & \mathbf{S}(f) \\ \mathbf{S}^*(-f) & \mathbf{R}^*(-f) \end{bmatrix}. \end{aligned}$$

Substituting this into (A.2),  $E \{\boldsymbol{\chi}_{(q,s)}\}$  is given by

$$\Delta^2 \sum_{m,n=0}^{N-1} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{i2\pi[m(f-\frac{q}{N\Delta})-n(f+\frac{s}{N\Delta})]\Delta} \boldsymbol{\Phi}(f)df.$$

Taking the sums into the integral and using [73, p. 26]

$$\sum_{m=0}^{N-1} e^{i2\pi mf} = e^{i(N-1)\pi f} \frac{\sin[N\pi f]}{\sin[\pi f]} \quad (\text{A.8})$$

we see that

$$E \{ \chi_{(q,s)} \} = \Delta^2 e^{-i(N-1)\pi(\frac{q}{N} + \frac{s}{N})} \times \int_{-f_N}^{f_N} \frac{\sin[N\pi(f\Delta - \frac{q}{N})] \sin[N\pi(f\Delta + \frac{s}{N})]}{\sin[\pi(f\Delta - \frac{q}{N})] \sin[\pi(f\Delta + \frac{s}{N})]} \Phi(f) df. \quad (\text{A.9})$$

Assuming  $\Phi(f)$  is continuous with a continuous derivative, (like the sine ratios), we now replace the integral by a left-endpoint Riemann sum, which has an error magnitude bounded by a term of order  $1/N$ , [26, p16]. We assume  $N$  is large. The integrand is evaluated at  $l/(N\Delta)$ ,  $l = -(N/2), \dots, (N/2) - 1$ . Using (A.8) again,  $E \{ \chi_{(q,s)} \}$  is thus approximated as

$$\frac{\Delta}{N} \sum_{l=-(N/2)}^{(N/2)-1} \sum_{m=0}^{N-1} e^{i2\pi m(\frac{l-q}{N})} \sum_{n=0}^{N-1} e^{-i2\pi n(\frac{l+s}{N})} \Phi(\frac{l}{N\Delta}), \quad (\text{A.10})$$

(within an error of order  $1/N$ ). Since

$$\sum_{t=0}^{N-1} e^{\pm i2\pi tk/N} = \begin{cases} N & \text{if } k = mN \text{ for integer } m; \\ 0, & \text{otherwise,} \end{cases} \quad (\text{A.11})$$

it follows that the only non-zero case is for  $l = q = s = 0$  so that (A.10) becomes simply

$$E \{ \chi_{(q,s)} \} \approx \begin{cases} N\Delta \Phi(0), & \text{if } q = 0, s = 0; \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.12})$$

Substituting (A.12) into (A.1) we obtain

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^T \} \approx \frac{a_0}{N} [\Psi_{k,0}^+(a_0)]^2 \Phi(0) = \mathbf{0},$$

since  $\Psi_{k,0}^+(a_0) = \Psi_k^+(0) = 0$ . So,  $\check{\mathbf{W}}_k$  approximates a *proper* complex-valued random vector when  $a > 0$ .

Now consider the case  $a < 0$ . For  $a < 0$  we have from (4.6) that  $\Psi_k^+(af) =$

$\Psi_k^-(|a|f)$ . So from (4.11),

$$W_k(a, b; Z) = \frac{|a_0|^{1/2}}{N\Delta^{1/2}} \sum_{q=0}^{N-1} \mathcal{Z}_q \Psi_{k,q}^-(|a_0|) e^{i2\pi qb_0/N}.$$

We note however that due to cyclicity of the discrete Fourier transform  $\mathcal{Z}_0 = \mathcal{Z}_N$  and  $\Psi_{k,0}^-(|a_0|) = \Psi_{k,N}^-(|a_0|) = \Psi_k^-(0) = 0$ , so that this can be written

$$W_k(a, b; Z) = \frac{|a_0|^{1/2}}{N\Delta^{1/2}} \sum_{q=1}^N \mathcal{Z}_q \Psi_{k,q}^-(|a_0|) e^{i2\pi qb_0/N},$$

where, from (4.3),

$$\Psi_{k,q}^-(|a_0|) = \begin{cases} 0, & \text{for } q = 1, \dots, (N/2); \\ \Psi_k^-(|a_0|^{\frac{q-N}{N}}), & \text{for } q = (N/2) + 1, \dots, N. \end{cases}$$

Then

$$E\{\check{\mathbf{W}}_k \check{\mathbf{W}}_k^T\} = \frac{|a_0|}{N^2\Delta} \sum_{q,s=(N/2)+1}^N E\{\chi_{(q,s)}\} \Psi_{k,q}^-(|a_0|) \Psi_{k,s}^-(|a_0|) e^{i2\pi b_0(q+s)/N}. \quad (\text{A.13})$$

With reference to the form of  $E\{\chi_{(q,s)}\}$  in (A.10), and the identity (A.11), the only non-zero case is  $q = s = N$ ,  $l = 0$ . We therefore get

$$E\{\chi_{(q,s)}\} \approx \begin{cases} N\Delta\Phi(0) & \text{if } q = s = N; \\ 0 & \text{otherwise.} \end{cases}$$

Then (A.13) gives

$$\begin{aligned} E\{\check{\mathbf{W}}_k \check{\mathbf{W}}_k^T\} &\approx \frac{|a_0|}{N} [\Psi_{k,N}^-(|a_0|)]^2 \Phi(0) e^{i4\pi b_0} \\ &= \frac{|a_0|}{N} [\Psi_{k,0}^-(|a_0|)]^2 \Phi(0) e^{i4\pi b_0} = \mathbf{0}, \end{aligned}$$

since  $\Psi_{k,0}^-(|a_0|) = \Psi_k^-(0) = 0$ . So,  $\check{\mathbf{W}}_k$  approximates a *proper* complex-valued

random vector for  $a < 0$ , and the first part of the proof is complete.

We now turn our attention to part 2 of the theorem and the form of  $E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \}$ . Consider firstly  $a > 0$ . Then (4.2), (4.11) and (4.12) give

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \} = \frac{a_0}{N^2 \Delta} \sum_{q,s=0}^{(N/2)-1} E \{ \boldsymbol{\rho}_{(q,s)} \} \Psi_{k,q}^+(a_0) \Psi_{k,s}^+(a_0) e^{i2\pi b_0(q-s)/N}, \quad (\text{A.14})$$

where now

$$E \{ \boldsymbol{\rho}_{(q,s)} \} = \Delta^2 \sum_{m,n=0}^{N-1} E \{ \boldsymbol{\xi}_{(m,n)} \} e^{-i2\pi(mq-ns)/N}, \quad (\text{A.15})$$

with

$$\boldsymbol{\xi}_{(m,n)} = \begin{bmatrix} Z_{1,m} Z_{1,n}^* & Z_{1,m} Z_{2,n}^* & Z_{1,m} Z_{1,n} & Z_{1,m} Z_{2,n} \\ Z_{2,m} Z_{1,n}^* & Z_{2,m} Z_{2,n}^* & Z_{2,m} Z_{1,n} & Z_{2,m} Z_{2,n} \\ Z_{1,m}^* Z_{1,n}^* & Z_{1,m}^* Z_{2,n}^* & Z_{1,m}^* Z_{1,n} & Z_{1,m}^* Z_{2,n} \\ Z_{2,m}^* Z_{1,n}^* & Z_{2,m}^* Z_{2,n}^* & Z_{2,m}^* Z_{1,n} & Z_{2,m}^* Z_{2,n} \end{bmatrix}.$$

Using (A.6) and (A.7) again gives

$$E \{ \boldsymbol{\xi}_{(m,n)} \} = \int_{-f_N}^{f_N} e^{i2\pi f(m-n)\Delta} \boldsymbol{\Upsilon}(f) df$$

where  $\boldsymbol{\Upsilon}(f)$  is given in (2.6). Substituting this into (A.15), we see that  $E \{ \boldsymbol{\rho}_{(q,s)} \}$  is given by

$$\Delta^2 \sum_{m,n=0}^{N-1} \int_{-f_N}^{f_N} e^{i2\pi [m(f-\frac{q}{N\Delta}) - n(f-\frac{s}{N\Delta})] \Delta} \boldsymbol{\Upsilon}(f) df.$$

As in the first part of the proof we take the sums into the integral and then replace the integral by a left-endpoint Riemann sum, with  $N$  large, so  $E \{ \boldsymbol{\rho}_{(q,s)} \}$  is approximated as

$$\frac{\Delta}{N} \sum_{l=-(N/2)}^{(N/2)-1} \sum_{m=0}^{N-1} e^{i2\pi m(\frac{l-q}{N})} \sum_{n=0}^{N-1} e^{-i2\pi n(\frac{l-s}{N})} \boldsymbol{\Upsilon}(\frac{l}{N\Delta}). \quad (\text{A.16})$$

Using (A.11) and (A.16) we can see that the only non-zero case is for  $l = q = s$

so that (A.16) becomes simply

$$E \{ \boldsymbol{\rho}_{(q,s)} \} \approx \begin{cases} N\Delta \Upsilon\left(\frac{q}{N\Delta}\right), & \text{if } q = s; \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.17})$$

Substituting (A.17) into (A.14), and using the fact that  $a = a_0\Delta$ , we obtain

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \} \approx \frac{a}{N\Delta} \sum_{q=0}^{(N/2)-1} \Upsilon\left(\frac{q}{N\Delta}\right) [\Psi_k^+(a\frac{q}{N\Delta})]^2,$$

completing the proof for  $a > 0$ .

We now look at  $E \{ \check{\mathbf{W}} \check{\mathbf{W}}^H \}$  when  $a < 0$ . We can adjust our indexing as for (A.13) to give

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \} = \frac{|a_0|}{N^2\Delta} \sum_{q,s=(N/2)+1}^N E \{ \boldsymbol{\rho}_{(q,s)} \} \Psi_{k,q}^-(|a_0|) \Psi_{k,s}^-(|a_0|) e^{i2\pi b_0(q-s)/N}. \quad (\text{A.18})$$

We recall that  $E \{ \boldsymbol{\rho}_{(q,s)} \}$  is approximated by (A.16). Combining (A.11) and (A.16) we have contributions to the sum when  $q = s$  with  $l = q - N$ , giving

$$E \{ \boldsymbol{\rho}_{(q,s)} \} \approx \begin{cases} N\Delta \Upsilon\left(\frac{q-N}{N\Delta}\right) & \text{if } q = s; \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.19})$$

Substituting (A.19) into (A.18) and reindexing gives

$$\begin{aligned} E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \} &\approx \frac{|a|}{N\Delta} \sum_{q=-(N/2)+1}^0 \Upsilon\left(\frac{q}{N\Delta}\right) [\Psi_k^-(|a|\frac{q}{N\Delta})]^2 \\ &= \frac{|a|}{N\Delta} \sum_{q=-(N/2)+1}^0 \Upsilon\left(\frac{q}{N\Delta}\right) [\Psi_k^+(a\frac{q}{N\Delta})]^2 \\ &= \frac{|a|}{N\Delta} \sum_{q=0}^{(N/2)-1} \Upsilon\left(-\frac{q}{N\Delta}\right) [\Psi_k^+(|a|\frac{q}{N\Delta})]^2, \end{aligned}$$

completing the proof for  $a < 0$ . □

## Appendix B

# Continuous-Time Version of Theorem 4.4.1

**Theorem B.0.1.** *Let  $\{[Z_1(t), \dots, Z_p(t)]^T, t \in \mathbb{R}\}$  be a realisation of a  $p$ -dimensional zero-mean complex-valued Gaussian (proper or improper) SOS continuous-time process. Then,*

(i)  $\check{\mathbf{W}}_k(a, b)$  is a proper zero-mean complex-valued Gaussian vector.

(ii) The covariance matrix  $E\{\check{\mathbf{W}}_k(a, b)\check{\mathbf{W}}_k^H(a, b)\}$  is given by

$$|a| \int_{-\infty}^{\infty} [\Psi_k^+(af)]^2 \mathbf{\Upsilon}(f) df. \quad (\text{B.1})$$

*Proof.* To prove that  $\check{\mathbf{W}}_k$  is proper, it is necessary to show that all the relational values are zero. This is equivalent to showing  $E\{\check{\mathbf{W}}_k\check{\mathbf{W}}_k^T\} = 0$ .

Using the definition of the continuous wavelet transform for  $|a| > 0$  and  $b \in \mathbb{R}$

$$W(a, b; Z(t), \psi_k^+) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} Z(t) \psi_k^{+*} \left( \frac{t-b}{a} \right) dt,$$

we have

$$E\{\check{\mathbf{W}}_k\check{\mathbf{W}}_k^T\} = \frac{1}{|a|} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\{\boldsymbol{\eta}(t, t')\} \psi_k^{+*} \left( \frac{t-b}{a} \right) \psi_k^{+*} \left( \frac{t'-b}{a} \right) dt dt' \quad (\text{B.2})$$

where

$$\boldsymbol{\eta}(t, t') = \begin{bmatrix} Z_1(t)Z_1(t') & Z_1(t)Z_2(t') & Z_1(t)Z_1^*(t') & Z_1(t)Z_2^*(t') \\ Z_2(t)Z_1(t') & Z_2(t)Z_2(t') & Z_2(t)Z_1^*(t') & Z_2(t)Z_2^*(t') \\ Z_1^*(t)Z_1(t') & Z_1^*(t)Z_2(t') & Z_1^*(t)Z_1^*(t') & Z_1^*(t)Z_2^*(t') \\ Z_2^*(t)Z_1(t') & Z_2^*(t)Z_2(t') & Z_2^*(t)Z_1^*(t') & Z_2^*(t)Z_2^*(t') \end{bmatrix}.$$

Using the continuous-time version of the spectral representation theorem — Theorem 2.3.1, and the proof of Theorem 4.4.1, we conclude

$$E \{ \boldsymbol{\eta}(t, t') \} = \int_{-\infty}^{\infty} e^{i2\pi f(t-t')} \boldsymbol{\Phi}(f) df \quad (\text{B.3})$$

where

$$\boldsymbol{\Phi}(f) = \begin{bmatrix} R_{11}(f) & R_{12}(f) & S_{11}(f) & S_{12}(f) \\ R_{21}(f) & R_{22}(f) & S_{21}(f) & S_{22}(f) \\ S_{11}^*(-f) & S_{12}^*(-f) & R_{11}^*(f) & R_{12}^*(f) \\ S_{21}^*(-f) & S_{22}^*(-f) & R_{21}^*(f) & R_{22}^*(f) \end{bmatrix}.$$

Substituting (B.3) back into (B.2) gives

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^T \} = \frac{1}{|a|} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i2\pi f(t-t')} \boldsymbol{\Phi}(f) \psi_k^{+*} \left( \frac{t-b}{a} \right) \psi_k^{+*} \left( \frac{t'-b}{a} \right) dt dt' df. \quad (\text{B.4})$$

Using (4.5), we have  $\psi_k^{+*}(t) = \psi_k^-(t)$ . With the Fourier transform of a Morse wavelets being real-valued, the following Fourier identities follow:

$$\int_{-\infty}^{\infty} e^{i2\pi ft} \psi_k^{+*} \left( \frac{t-b}{a} \right) dt = a e^{i2\pi fb} \Psi_k^-( -af)$$

and

$$\int_{-\infty}^{\infty} e^{-i2\pi ft'} \psi_k^{+*} \left( \frac{t'-b}{a} \right) dt' = a e^{i2\pi fb} \Psi_k^-( af)$$

where  $\Psi_k^-(\cdot)$  is the Fourier transform of the  $k$ th anti-analytic generalised Morse

wavelet  $\psi_k^-(\cdot)$ . We can therefore reduce (B.4) to

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^T \} = |a| e^{i4\pi fb} \int_{-\infty}^{\infty} \Psi_k^-(af) \Psi_k^-(af) \Phi(f) df.$$

We now note that  $\Psi_k^-(\nu) = 0$  for all  $\nu \geq 0$  and so  $\Psi_k^-(af) \Psi_k^-(af) = 0$  for all  $f \in \mathbb{R}$  and thus

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^T \} = 0.$$

We now deal with  $E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \}$ . Using the definition of the continuous wavelet transform for  $|a| > 0$  and  $b \in \mathbb{R}$  we have

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \} = \frac{1}{|a|} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E \{ \boldsymbol{\xi}(t, t') \} \psi_k^{+*} \left( \frac{t-b}{a} \right) \psi_k^+ \left( \frac{t'-b}{a} \right) dt dt' \quad (\text{B.5})$$

where

$$\boldsymbol{\xi}(t, t') = \begin{bmatrix} Z_1(t)Z_1^*(t') & Z_1(t)Z_2^*(t') & Z_1(t)Z_1(t') & Z_1(t)Z_2(t') \\ Z_2(t)Z_1^*(t') & Z_2(t)Z_2^*(t') & Z_2(t)Z_1(t') & Z_2(t)Z_2(t') \\ Z_1^*(t)Z_1^*(t') & Z_1^*(t)Z_2^*(t') & Z_1^*(t)Z_1(t') & Z_1^*(t)Z_2(t') \\ Z_2^*(t)Z_1^*(t') & Z_2^*(t)Z_2^*(t') & Z_2^*(t)Z_1(t') & Z_2^*(t)Z_2(t') \end{bmatrix}.$$

Using the continuous-time version of the spectral representation theorem — Theorem 2.3.1, and the proof of Theorem 4.4.1, we conclude

$$E \{ \boldsymbol{\xi}(t, t') \} = \int_{-\infty}^{\infty} e^{i2\pi f(t-t')} \boldsymbol{\Upsilon}(f) df \quad (\text{B.6})$$

where

$$\boldsymbol{\Upsilon}(f) = \begin{bmatrix} S_{11}(f) & S_{12}(f) & R_{11}(f) & R_{12}(f) \\ S_{21}(f) & S_{22}(f) & R_{21}(f) & R_{22}(f) \\ R_{11}^*(f) & R_{12}^*(f) & S_{11}^*(-f) & S_{12}^*(-f) \\ R_{21}^*(f) & R_{22}^*(f) & S_{21}^*(-f) & S_{22}^*(-f) \end{bmatrix}.$$

Substituting (B.6) back into (B.5) gives

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \} = \frac{1}{|a|} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i2\pi f(t-t')} \mathbf{r}(f) \psi_k^{+*} \left( \frac{t-b}{a} \right) \psi_k^+ \left( \frac{t'-b}{a} \right) dt dt' df \quad (\text{B.7})$$

Now using the Fourier identities

$$\int_{\mathbb{R}} e^{i2\pi ft} \psi_k^{+*} \left( \frac{t-b}{a} \right) dt = a e^{i2\pi fb} \Psi_k^+(af)$$

and

$$\int_{\mathbb{R}} e^{-i2\pi ft'} \psi_k^+ \left( \frac{t'-b}{a} \right) dt' = a e^{-i2\pi fb} \Psi_k^+(af)$$

where  $\Psi_k^+(\cdot)$  is the Fourier transform of  $\psi_k^+(\cdot)$ , (B.7) becomes

$$E \{ \check{\mathbf{W}}_k \check{\mathbf{W}}_k^H \} = |a| \int_{\mathbb{R}} \mathbf{r}(f) [\Psi_k^+(af)]^2 df.$$

□

## Appendix C

### Proof of Theorem 5.5.1

*Proof.* Write  $Z_{l,n} = X_{l,n} + iY_{l,n}$ , with  $\{X_{l,n}\}$  and  $\{Y_{l,n}\}$  real-valued, for  $l = 1, 2$ . Now define the real-valued vector  $[X_{1,n}, X_{2,n}, Y_{1,n}, Y_{2,n}]^T = [\mathbf{X}_n^T, \mathbf{Y}_n^T]^T$ . Then if

$$\mathbf{T} = \begin{bmatrix} \mathbf{I}_2 & i\mathbf{I}_2 \\ \mathbf{I}_2 & -i\mathbf{I}_2 \end{bmatrix},$$

we see that

$$\mathbf{T} \begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_n + i\mathbf{Y}_n \\ \mathbf{X}_n - i\mathbf{Y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_n \\ \mathbf{Z}_n^* \end{bmatrix} = \check{\mathbf{Z}}_n. \quad (\text{C.1})$$

Form the product  $\{u_{j,n}\check{\mathbf{Z}}_n\}$  of the  $j$ th real-valued taper with  $\check{\mathbf{Z}}_n$ , and then compute its (scaled) Fourier transform

$$\check{\mathbf{J}}_j(f) = \Delta^{1/2} \sum_{n=0}^{N-1} u_{j,n} \begin{bmatrix} \mathbf{Z}_n \\ \mathbf{Z}_n^* \end{bmatrix} e^{-i2\pi fn\Delta} = \begin{bmatrix} \mathbf{J}_j(f) \\ \mathbf{J}_j^*(-f) \end{bmatrix},$$

where  $\mathbf{J}_j(f) \equiv \Delta^{1/2} \sum_{n=0}^{N-1} u_{j,n} \mathbf{Z}_n e^{-i2\pi fn\Delta}$ . From (C.1),

$$\check{\mathbf{J}}_j(f) = \mathbf{T} \Delta^{1/2} \sum_{n=0}^{N-1} u_{j,n} \begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} e^{-i2\pi fn\Delta} = \mathbf{T} \begin{bmatrix} \mathbf{J}_{X,j}(f) \\ \mathbf{J}_{Y,j}(f) \end{bmatrix},$$

where  $\mathbf{J}_{X,k}(f) \equiv \Delta^{1/2} \sum_{n=0}^{N-1} u_{j,n} \mathbf{X}_n e^{-i2\pi f n \Delta}$  and similarly for  $\mathbf{J}_{Y,j}(f)$ . Hence,

$$\begin{bmatrix} \mathbf{J}_j(f) \\ \mathbf{J}_j^*(-f) \end{bmatrix} = \mathbf{T} \begin{bmatrix} \mathbf{J}_{X,j}(f) \\ \mathbf{J}_{Y,j}(f) \end{bmatrix}, \quad (\text{C.2})$$

and as an estimator of  $\mathbf{\Upsilon}(f)$  we take the multitaper multivariate estimator  $\hat{\mathbf{\Upsilon}}(f) = (1/N_B) \sum_{j=0}^{N_B-1} \gamma_j \check{\mathbf{J}}_j(f) \check{\mathbf{J}}_j^H(f)$ , which can be written as

$$\frac{1}{N_B} \sum_{j=0}^{N_B-1} \gamma_j \begin{bmatrix} \mathbf{J}_j(f) \mathbf{J}_j^H(f) & \mathbf{J}_j(f) \mathbf{J}_j^T(-f) \\ \{\mathbf{J}_j(-f) \mathbf{J}_j^T(f)\}^* & \{\mathbf{J}_j(-f) \mathbf{J}_j^H(-f)\}^* \end{bmatrix}.$$

(C.2) means that  $\hat{\mathbf{\Upsilon}}(f)$  can also be written as

$$\frac{1}{N_B} \sum_{j=0}^{N_B-1} \gamma_j \mathbf{T} \begin{bmatrix} \mathbf{J}_{X,j}(f) \mathbf{J}_{X,j}^H(f) & \mathbf{J}_{X,j}(f) \mathbf{J}_{Y,j}^H(f) \\ \mathbf{J}_{Y,j}(f) \mathbf{J}_{X,j}^H(f) & \mathbf{J}_{Y,j}(f) \mathbf{J}_{Y,j}^H(f) \end{bmatrix} \mathbf{T}^H,$$

so that

$$\hat{\mathbf{\Upsilon}}(f) = \begin{bmatrix} \hat{\mathbf{S}}(f) & \hat{\mathbf{R}}(f) \\ \hat{\mathbf{R}}^H(f) & \hat{\mathbf{S}}^T(-f) \end{bmatrix} = \mathbf{T} \hat{\mathbf{\Sigma}}(f) \mathbf{T}^H, \quad (\text{C.3})$$

where,

$$\hat{\mathbf{\Sigma}}(f) \equiv \begin{bmatrix} \hat{\mathbf{S}}_{X,X}(f) & \hat{\mathbf{S}}_{X,Y}(f) \\ \hat{\mathbf{S}}_{Y,X}(f) & \hat{\mathbf{S}}_{Y,Y}(f) \end{bmatrix},$$

and, e.g.,  $\hat{\mathbf{S}}_{X,Y}(f) \equiv (1/N_B) \sum_{j=0}^{N_B-1} \gamma_j \mathbf{J}_{X,j}(f) \mathbf{J}_{Y,j}^H(f)$ . Since  $[\mathbf{X}_n^T, \mathbf{Y}_n^T]^T$  is real-valued and Gaussian, it follows immediately [102, p. 776] that asymptotically we may take the distribution of  $\hat{\mathbf{\Sigma}}(f)$  to be

$$\hat{\mathbf{\Sigma}}(f) \stackrel{d}{=} (1/\varphi) \mathcal{W}_4^C\{\varphi, \mathbf{\Sigma}(f)\}, \quad 0 < |f| < f_N. \quad (\text{C.4})$$

$\mathbf{\Sigma}(f)$  is the true spectral matrix for the process  $\{[\mathbf{X}_n^T, \mathbf{Y}_n^T]^T\}$ .

We want the distribution of  $\hat{\mathbf{\Upsilon}}(f) = \mathbf{T} \hat{\mathbf{\Sigma}}(f) \mathbf{T}^H$ . Now  $\mathbf{\Upsilon}(f) = \mathbf{T} \mathbf{\Sigma}(f) \mathbf{T}^H$ . Combining this with (C.4) and the result of the Lemma given below we see

that asymptotically,

$$\hat{\Upsilon}(f) \stackrel{d}{=} (1/\varphi)\mathcal{W}_4^C\{\varphi, \Upsilon(f)\}, \quad 0 < |f| < f_{\mathcal{N}}. \quad (\text{C.5})$$

□

**Lemma C.0.2.** *If  $\mathbf{W}$  has the  $\mathcal{W}_q^C\{\ell, \Sigma\}$  distribution and  $\mathbf{M}$  is a complex-valued  $q \times q$  matrix, then  $\mathbf{M}\mathbf{W}\mathbf{M}^H$  has the  $\mathcal{W}_q^C\{\ell, \mathbf{M}\Sigma\mathbf{M}^H\}$  distribution.*

*Proof.* The characteristic function of  $\mathbf{W}$  is [35, p. 163]

$$\phi(\Theta_0) = E\{\exp[i \cdot \text{tr}(\mathbf{W}\Theta_0)]\} = [\det(\mathbf{I}_q - i\Theta_0\Sigma)]^{-\ell}. \quad (\text{C.6})$$

The characteristic function,  $\phi(\Theta)$ , of  $\mathbf{M}\mathbf{W}\mathbf{M}^H$  is therefore

$$\begin{aligned} E\{\exp[i \cdot \text{tr}(\mathbf{M}\mathbf{W}\mathbf{M}^H\Theta)]\} &= E\{\exp[i \cdot \text{tr}(\mathbf{W}\mathbf{M}^H\Theta\mathbf{M})]\} \\ &= [\det(\mathbf{I}_q - i\mathbf{M}^H\Theta\mathbf{M}\Sigma)]^{-\ell} \\ &= [\det(\mathbf{I}_q - i\Theta\mathbf{M}\Sigma\mathbf{M}^H)]^{-\ell}, \end{aligned} \quad (\text{C.7})$$

where  $\text{tr}$  and  $\det$  denote trace and determinant respectively, and we have used the fact that if  $\mathbf{A}$  is an  $j \times k$  complex-valued matrix and  $\mathbf{B}$  is an  $k \times j$  complex-valued matrix, then  $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$  and  $\det(\mathbf{I}_j + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_k + \mathbf{B}\mathbf{A})$ , [27]. Comparing (C.6) and (C.7) it follows that  $\mathbf{M}\mathbf{W}\mathbf{M}^H$  has the  $\mathcal{W}_q^C\{\ell, \mathbf{M}\Sigma\mathbf{M}^H\}$  distribution. □

## Appendix D

# Characteristic Width Derivation for the Downburst Wind Process

The characteristic width of the UMP is calculated in a similar way to the example given in Section 6.3.1. We first need to calculate  $k(\theta)$ , the Fourier transform of the modulating function  $A(t) = \alpha t^\beta e^{-\lambda t} I_{[0, \infty)}$ . This is defined as

$$\begin{aligned} k(\theta) &= \int_0^\infty \alpha t^\beta e^{-\lambda t} e^{-i2\pi\theta t} dt \\ &= \int_0^\infty \alpha t^\beta e^{-t(\lambda + i2\pi\theta)} dt \end{aligned}$$

To evaluate this, for integer valued  $n$  we begin by defining the integral

$$I_n(\theta) = \int_0^\infty t^n e^{-t(\lambda + i2\pi\theta)} dt.$$

Using integration by parts we have

$$\begin{aligned}
 I_n(\theta) &= \left[ \frac{t^n e^{-t(\lambda+i2\pi\theta)}}{-(\lambda+i2\pi\theta)} \right]_0^\infty - \frac{n}{-(\lambda+i2\pi\theta)} \int_0^\infty t^{n-1} e^{-t(\lambda+i2\pi\theta)} dt \\
 &= \frac{n}{\lambda+i2\pi\theta} \int_0^\infty t^{n-1} e^{-t(\lambda+i2\pi\theta)} dt \\
 &= \frac{n}{\lambda+i2\pi\theta} I_{n-1}(\theta).
 \end{aligned}$$

Iterating, this becomes

$$I_n(\theta) = \frac{n!}{(\lambda+i2\pi\theta)^n} I_0 \tag{D.1}$$

where

$$I_0(\theta) = \int_0^\infty e^{-t(\lambda+i2\pi\theta)} dt = \frac{1}{\lambda+i2\pi\theta}. \tag{D.2}$$

Combining (D.1) and (D.2) gives

$$I_n(\theta) = \frac{n!}{(\lambda+i2\pi\theta)^{n+1}}.$$

The examples considered in [41] and Section 7.7 have  $\beta$  integer valued, and so we can say

$$k(\theta) = \frac{\alpha(\beta!)}{(\lambda+i2\pi\theta)^{(\beta+1)}}$$

The characteristic width  $B_{\mathcal{F}}$  of the family  $\mathcal{F} = \{\alpha t^\beta e^{-\lambda t} I_{[0,\infty)} e^{i2\pi f t}\}$  is given by (6.12), and in this example becomes

$$B_{\mathcal{F}} = \left[ 2\pi \int_{-\infty}^\infty |\theta| \left| \frac{\alpha(\beta!)}{(\lambda+i2\pi\theta)^{(\beta+1)}} \right| d\theta \right]^{-1}.$$

This can be rewritten as

$$\begin{aligned}
 B_{\mathcal{F}} &= \left[ 2\pi\alpha(\beta!) \int_{-\infty}^\infty \frac{|\theta|}{(\lambda^2 + (2\pi\theta)^2)^{(\beta+1)/2}} d\theta \right]^{-1} \\
 &= \left[ 2\pi\alpha(\beta!) \int_{-\infty}^\infty \frac{|\theta|}{\lambda^{(\beta+1)} \left(1 + \frac{(2\pi\theta)^2}{\lambda^2}\right)^{(\beta+1)/2}} d\theta \right]^{-1}. \tag{D.3}
 \end{aligned}$$

By recognising that the integrand in (D.3) is even and with the appropriate substitution  $\nu = (2\pi\theta)/\lambda$  we have

$$B_{\mathcal{F}} = \left[ \frac{2\alpha(\beta!)}{2\pi\lambda^{\beta-1}} \int_0^\infty \frac{\nu}{(1+\nu^2)^{(\beta+1)/2}} d\nu \right]^{-1}.$$

For for  $|\arg(\alpha)| < \pi$ ,  $p > 0$  and  $0 < \operatorname{Re}\{\mu\} < p\operatorname{Re}\{\eta\}$ , the following identity holds [38, p. 325]

$$\int_0^\infty \nu^{\mu-1} (1 + \alpha\nu^p)^{-\eta} d\nu = \frac{1}{p} \alpha^{-\mu/p} \operatorname{beta} \left( \frac{\mu}{p}, \eta - \frac{\mu}{p} \right)$$

where  $\operatorname{beta}(\cdot, \cdot)$  is the beta function. We conclude from (6.13) that the characteristic width  $B_Z$  of a UMP  $\{Z(t)\}$  defined via the family  $\mathcal{F}$  will be bounded by

$$B_Z \geq B_{\mathcal{F}} = \left[ \frac{1}{2\pi} \alpha(\beta!) \lambda^{1-\beta} \operatorname{beta} \left( 1, \frac{\beta-1}{2} \right) \right]^{-1}.$$

## References

- [1] M. A. Abramowicz, G. Bao, S. Larsson, and P. J. Wiita. On the variability coherence observed in black hole candidates at different X-ray energies. *The Astrophysical Journal*, **489**, 819–821, 1997.
- [2] F. Battaglia. Some extensions in the evolutionary spectral analysis of a stochastic process. *della Unione Matematica Italiana*, Serie V, **XVI-B**, 1154–1166, 1979.
- [3] D. S. Bloomfield, R. T. J. McAteer, B. W. Lites, P. G. Judge, M. Mathioudakis, and F. P. Keenan. Wavelet phase coherence analysis: Application to a quiet-sun magnetic element. *The Astrophysical Journal*, **617**, 623–632, 2004.
- [4] R. Bortel and P. Sovka. Approximation of statistical distribution of magnitude squared coherence estimated with segment overlapping. *Signal Processing*, **87**, 1100–1117, 2007.
- [5] R. N. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill International Editions, 2nd Edition, 1986.
- [6] D. R. Brillinger. *Time Series: Data Analysis and Theory (Expanded Edition)*. San Francisco: Holden-Day, 1981.
- [7] J.-S. Brittain, D. M. Halliday, B. A. Conway, and J. B. Nielsen. Single-trial multiwavelet coherence in application to neurophysiological time series. *IEEE Transactions on Biomedical Engineering*, **54**, 854–862, 2007.

- 
- [8] T. P. Bronez. On the performance advantage of multitaper spectral analysis. *IEEE Transactions on Biomedical Engineering*, **40**, 2941–2946, 1992.
- [9] R. Carmona, W. L. Hwang, and B. Torresani. *Practical Time-Frequency Analysis*. 1st Edition, San Diego, CA: Academic, 1998.
- [10] G. C. Carter. Coherence and time delay estimation. *Proceedings of the IEEE*, **75**, 236–255, 1987.
- [11] G. C. Carter, C. H. Knapp, and A. H. Nuttall. Statistics of the estimate of the magnitude-coherence function. *IEEE Transactions on Audio and Electroacoustics*, **AU-21**, 388–389, 1973.
- [12] G. Chan and A. T. A. Wood. Simulation of stationary Gaussian vector fields. *Statistics and Computing*, **9**, 265–268, 1999.
- [13] C. K. Chui. *Wavelet Analysis and its applications, Vol. 1: An Introduction to Wavelets*. Academic Press, 1992.
- [14] E. A. K. Cohen and A. T. Walden. A statistical analysis of Morse wavelet coherence. *IEEE Transactions on Signal Processing*, **58**, 980–989, 2010.
- [15] E. A. K. Cohen and A. T. Walden. A statistical study of temporally smoothed wavelet coherence. *IEEE Transactions on Signal Processing*, **58**, 2964–2973, 2010.
- [16] E. A. K. Cohen and A. T. Walden. Wavelet coherence for certain nonstationary bivariate processes. *IEEE Transactions on Signal Processing*, **59**, Due for publication July 2011.
- [17] H. Cramér. On harmonic analysis in certain functional spaces. *Arkiv Math. Astr. Fysik*, **7**, 1–7, 1942.
- [18] H. Cramér. On some classes of nonstationary stochastic processes. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, **2**, 57–58. Los Angeles, CA: University of California Press, 1961.

- 
- [19] R. Dahlhaus. Asymptotic statistical inference for nonstationary processes with evolutionary spectra. *Athens Conference on Applied Probability and Time Series Analysis, II: Time Series Analysis*, 146–159, 1996.
- [20] R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications*, **62**, 139–168, 1996.
- [21] R. Dahlhaus. Time series models to nonstationary processes. *The Annals of Statistics*, **25**, 1–37, 1997.
- [22] G. R. Dargahi-Noubary and P. J. Laycock. Spectral ratio discriminants and information theory. *Journal of Time Series Analysis*, **2**, 71–86, 1981.
- [23] I. Daubechies. *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.
- [24] I. Daubechies. Time-frequency localization operators: A geometric phase space approach. *IEEE Transactions on Signal Processing*, **34**, 605–612, 1988.
- [25] I. Daubechies and T. Paul. Time-frequency localization operators: A geometric phase space approach: II. the use of dialations. *Inverse Problems* **4**, 661–680, 1988.
- [26] P. Davis and P. Rabinowitz. *Numerical Integration*. London, U.K.: Blaisdell, 1967.
- [27] M. L. Eaton. *Multivariate Statistics*. New York Wiley, 1983.
- [28] W. J. Emery and R. E. Thomson. *Data Analysis Methods in Physical Oceanography*. New York: Pergamon, 1998.
- [29] S. A. Santos Filho, C. J. Tierra-Criollo, A. P. Souza, M. A. Silva Pinto, M. L. Cunha Lima, and G. M. Manzano. Magnitude squared of coherence to detect imaginary movement. *EURASIP Journal on Advances in Signal Processing*, **2009**, 1–13 2009.

- 
- [30] P. Flandrin. Maximum signal energy concentration in a time-frequency domain. *1988 International Conference on Acoustics, Speech, and Signal Processing*, **4**, 2176–2179, 1988.
- [31] D. Gabor. Theory of communication. *IEE (London)*, **93-III**, 429–441, 1946.
- [32] W. A. Gardner. *Introduction to Random Processes with Applications to Signals and Systems*. Macmillan, New York, 1st Edition, 1985.
- [33] W. A. Gardner. A unifying view of coherence in signal processing. *Signal Processing* **29**, 113–140, 1992.
- [34] N. R. Goodman. On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two-dimensional stationary Gaussian process. *Engineering Statistics Laboratory – Scientific Paper 10 (AD 134 919)*, New York University, New York, 1957.
- [35] N. R. Goodman. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Annals of Mathematical Statistics*, **34**, 152–177, 1963.
- [36] N. R. Goodman and M. R. Dubman. Theory of time-varying spectral analysis and complex Wishart matrix processes. *Multivariate Analysis (ed. P. R. Krishnaiah)*, **2**. New York Academic Press, 351–366, 1969.
- [37] P. Goupillaud, A. Grossmann, and J. Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, **23**, 85–102, 1984.
- [38] I. S. Gradshteyn and I. M. Ryzhik. *Tables of Integrals, Series, and Products*. New York: Academic Press, 7th Edition, 1980.
- [39] A. Grinsted, J. C. Moore, and S. Jevrejeva. Applications of the cross wavelet transform and the wavelet coherence to geophysical time series. *Non-linear Processes in Geophysics*, **11**, 561–566, 2004.

- [40] A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal of Mathematical Analysis*, **15**, 723–736, 1984.
- [41] X. Chen and G. Huang. Wavelets-based estimation of multivariate evolutionary spectra and its application to nonstationary downburst winds. *Engineering Structures*, **31**, 976–989, 2009.
- [42] K. Gurley, T. Kijewski, and A. Kareem. First- and higher-order correlation detection using wavelet transforms. *Journal of Engineering Mechanics*, **129**, 188–201, 2003.
- [43] A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, **69**, 331–371, 1910.
- [44] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- [45] J. K. Hammond. Evolutionary spectra in random vibrations. *Journal of the Royal Statistical Society Series B*, **35**, 167–188, 1973.
- [46] A. Hanssen, Y. Larsen, and L. L. Scharf. Complex time-frequency and dual-frequency spectra of harmonizable processes. *Proceedings EUSIPCO, Vienna, Austria*, 2004.
- [47] Y. Hayashi. Space-time spectral analysis of rotary vector series. *Journal of Atmospheric Sciences*, **36**, 757–766, 1979.
- [48] L. J. Herbst. Periodogram analysis and variance fluctuations. *Journal of the Royal Statistical Society Series B*, **25**, 442–450, 1963.
- [49] L. Hudgins, C. A. Friehe, and M. E. Mayer. Wavelet transforms and atmospheric turbulence. *Physics Review Letters*, **71**, 3279–3282, 1993.
- [50] J. Sanderson, P. Fryzlewicz, and M. W. Jones. Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, **97**, 435–446, 2010.

- [51] J. L. Kiehm and G. Mélard. Evolutive cospectral analysis of time-dependent ARMA processes. *Time Series Analysis*, eds. O.D. Anderson and M.R. Perryman, North-Holland Publishing Company, 1981.
- [52] C. Kuo, C. R. Lindberg, and D. J. Thomson. Coherence established between atmospheric carbon dioxide and global temperature. *Nature*, **343**, 709–714, 1990.
- [53] J.-P. Lachaux, A. Lutz, D. Rudrauf, D. Cosmelli, M. Le Van Quyen, J. Martinerie, and F. Varela. Estimating the time-course of coherence between single-trial brain signals: An introduction to wavelet coherence. *Clinical Neurophysiology*, **32**, 157–174, 2003.
- [54] P. F. Lee. An algorithm for computing the cumulative distribution function for magnitude squared coherence estimation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **29**, 117–119, 1981.
- [55] J. M. Lilly and J. Park. Multiwavelet spectral and polarization analyses of seismic records. *Geophysics Journal International*, **122**, 1001–1021, 1995.
- [56] P. C. Liu. Wavelet spectrum analysis and ocean wind waves. *Wavelet Analysis and its Applications, Vol 4: Wavelets in Geophysics*, eds. P. Kumar, E. Foufoula-Georgiou, San Diego CA: Academic, 151–156, 1994.
- [57] M. Loève. *Probability Theory*. D. Van Nostrand Company Inc., 3rd Edition, 1963.
- [58] R. M. Loynes. On the concept of the spectrum for non-stationary processes. *Journal of the Royal Statistical Society Series B*, **30**, 1–30, 1968.
- [59] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic, 2nd Edition, 1999.
- [60] D. Maraun and J. Kurths. Cross wavelet analysis: Significance testing and pitfalls. *Non-linear Processes in Geophysics* **11**, 504–514, 2004.

- 
- [61] D. Maraun and J. Kurths. Nonstationary Gaussian processes in wavelet domain: Synthesis, estimation, and significant testing. *Physical Review E*, **75**, 016707, 2007.
- [62] T. Medkour and A. T. Walden. A variance equality test for two correlated complex Gaussian variables with applications to spectral power comparisons. *IEEE Transactions on Signal Processing*, **55**, 881–888, 2007.
- [63] G. Mélard. Examples of the evolutionary spectrum theory. *Journal of Time Series Analysis*, **6**, 81–90, 1985.
- [64] G. Mélard and A. Herteleer-De Schutter. Contributions to evolutionary spectral theory. *Journal of Time Series Analysis*, **10**, 41–63, 1989.
- [65] J. Morlet, G. Arens, and E. Fourgeau. Wave-propagation and sampling theory 1: Complex signal and scattering in multilayered media. *Geophysics*, **47**, 203–221, 1982.
- [66] J. Morlet, G. Arens, and E. Fourgeau. Wave-propagation and sampling theory 2: Sampling theory and complex waves. *Geophysics*, **47**, 222–236, 1982.
- [67] G. Nason, R. von Sachs, and G. Kroisandt. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistics Society Series B*, **62**, 271–292, 1998.
- [68] S. C. Olhede and A. T. Walden. Generalized Morse wavelets. *IEEE Transactions on Signal Processing*, **50**, 2661–2670, 2002.
- [69] S. C. Olhede and A. T. Walden. Noise reduction in directional signals using multiple Morse wavelets illustrated on quadrature Doppler ultrasound. *IEEE Transactions on Biomedical Engineering*, **50**, 51–57, 2003.
- [70] S. C. Olhede and A. T. Walden. Polarization phase relationships via a multiple Morse wavelets. I. Fundamentals. *Proceedings: Mathematical, Physical and Engineering Sciences*, **459**, 413–444, 2003.

- [71] S. C. Olhede and A. T. Walden. Polarization phase relationships via a multiple Morse wavelets. II. Data analysis. *Proceedings: Mathematical, Physical and Engineering Sciences*, **459**, 641–657, 2003.
- [72] M. Orlic, B. Penzar, and I. Penzar. Adriatic sea and land breezes: Clockwise versus anticlockwise rotation. *Journal of Applied Meteorology*, **27**, 675-679, 1988.
- [73] D. B. Percival and A. T. Walden. *Spectral Analysis for Physical Applications*. Cambridge University Press, 1993.
- [74] D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 2000.
- [75] B. Picinbono and P. Bondon. Second-order complex random vectors and normal distributions. *IEEE Transactions on Signal Processing*, **44**, 2637–2640, 1996.
- [76] B. Picinbono and P. Bondon. Second-order statistics of complex signals. *IEEE Transactions on Signal Processing*, **45**, 411-420, 1997.
- [77] M. I. Plett. Transient detection with cross wavelet transforms and wavelet coherence. *IEEE Transactions on Signal Processing*, **55**, 1605-1611, 2007.
- [78] M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society Series B*, **27**, 204–237, 1965.
- [79] M. B. Priestley. Design relations for non-stationary processes. *Journal of the Royal Statistical Society Series B*, **28**, 228–240, 1966.
- [80] M. B. Priestley. Power spectral analysis of non-stationary random processes. *Journal of Sound and Vibration*, **6**, 86–97, 1971.
- [81] M. B. Priestley. *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press, 1988.

- [82] M. B. Priestley. Wavelets and time-dependent spectral analysis. *Journal of Time Series Analysis*, **17**, 85–103, 1996.
- [83] M. B. Priestley and H. Tong. On the analysis of bivariate non-stationary processes. *Journal of the Royal Statistical Society Series B*, **35**, 153–166, 1973.
- [84] M. Reed and B. Simon. *Methods of Modern Mathematical Physics: 1 Functional Analysis*. New York: Academic, 1st Edition, 1972.
- [85] K. S. Riedel and A. Sidorenko. Minimum bias multiple taper spectral estimation. *IEEE Transactions on Signal Processing*, **43**, 188–195, 1995.
- [86] P. Rubin-Delanchy. *Some New Results in the Analysis of Complex-Valued Time Series*. Ph.D. Thesis, Imperial College London, 2008.
- [87] P. Rubin-Delanchy and A. T. Walden. Simulation of improper complex-valued sequences. *IEEE Transactions on Signal Processing*, **55**, 5517–5521, 2007.
- [88] P. Rubin-Delanchy and A. T. Walden. Kinematics of complex-valued time series. *IEEE Transactions on Signal Processing*, **56**, 4189–4198, 2008.
- [89] J. M. Schoffelen, R. Oostenveld, and P. Fries. Neuronal coherence as a mechanism of effective corticospinal interaction. *Science*, **308**, 111–113, 2005.
- [90] P. J. Schreier and L. L. Scharf. *Statistical Signal Processing of Complex-Valued Data*. Cambridge University Press, 2010.
- [91] P. J. Schreier and L. L. Scharf. Second-order analysis of improper complex random vectors and processes. *IEEE Transactions on Signal Processing*, **51**, 714–725, 2003.
- [92] D. Slepian. Prolate spheroidal wave-functions, Fourier-analysis, and uncertainty. 5: Discrete case. *Bell Systems Technical Journal*, **57**, 1371–1430, 1978.

- [93] P. D. Spanos and G. Failla. Evolutionary spectra estimation using wavelets. *Journal of Engineering Mechanics*, **130**, 952–960, 2004.
- [94] G. Strang. *Linear Algebra and its Applications*. 3rd Edition: Orlando, FL: Harcourt Brace Jovanovich, 1988.
- [95] D. J. Thomson. Coherence between SOHO MDI velocity and ulysses magnetic fields. *Proceedings of the SOHO 10/GONG 2000 Workshop: Helio- and asteroseismology at the dawn of the millennium*, 551–554, 2000.
- [96] D. J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, **70**, 1055–1096, 1982.
- [97] H. Tong. Some comments on spectral representation of nonstationary stochastic processes. *Journal of Applied Probability*, **10**, 881–885, 1973.
- [98] C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bulletins of the American Meteorological Society*, **79**, 61–78, 1998.
- [99] C. Torrence and P. Webster. Interdecadal changes in the ENSO-monsoon system. *Journal of Climate*, **12**, 2679–2690, 1999.
- [100] H. van Haren. On the polarization of oscillatory currents in the Bay of Biscay. *Journal of Geophysical Research*, **108**, 6-1–6-7, 2003.
- [101] B. Ph. van Milligen, E. Sánchez, T. Estrada, C. Hidalgo, B. Brañas, B. Carreras, and L. García. Wavelet bicoherence: A new turbulence analysis model. *Physics of Plasmas*, **2**, 3017–3032, 1995.
- [102] A. T. Walden. A unified view of multitaper multivariate spectral estimation. *Biometrika*, **87**, 767–788, 2000.
- [103] A. T. Walden and D. B. Percival. *Spectral Analysis of Multiple Time Series*. Manuscript in publication, 2011.
- [104] Y. Wang and P. J. Fish. Arterial Doppler signal simulation by time domain processing. *European Journal on Ultrasound*, **3**, 71–81, 1996.

- 
- [105] Y. Wang and P. J. Fish. Simulating nonstationary in-phase and quadrature Doppler signals using a time-varying impulse response filter. *Ultrasound in Medicine and Biology*, **22**, 529–531, 1996.
- [106] P. D. Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, **Au-15**, 70–73, 1967.
- [107] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A**, 32–52, 1928.
- [108] H. Wold. *A study in the analysis of stationary time series*. Almqvist and Wiksell, 2nd Edition, 1954.
- [109] A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions Volume I: Basic Results*. Springer, 1987.
- [110] A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions Volume II: Supplementary Notes and References*. Springer, 1987.