THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

LSE Research **Online**

Jean Sanderson, Piotr Fryzlewicz and M. W. Jones

Estimating linear dependence between nonstationary time series using the locally stationary wavelet model

Article (Accepted version)
(Unrefereed)

This version available at: http://eprints.lse.ac.uk/29141/
Available in LSE Research Online: November 2011

http://eprints.lse.ac.uk

Measuring dependence between non-stationary time series using the locally stationary wavelet model

Jean Sanderson, Piotr Fryzlewicz and Matthew Jones

November 28, 2008

Abstract

Large volumes of neuroscience data comprise multiple, non-stationary electrophysiological or neuroimaging time series recorded from different brain regions. Estimating the dependence between such neural time series accurately is critical, since changes in the dependence structure are presumed to reflect functional interactions between neuronal populations. We propose a new method of wavelet coherence, derived from the new bivariate locally stationary wavelet (LSW) time series model. Since wavelets are localised in both time and scale, this approach leads to a natural, local and multiscale estimate of nonstationary dependence. Our methodology is illustrated by application to a simulated example, and to electrophysiological data relating to interactions between the rat hippocampus and prefrontal cortex during working memory and decision-making. We thereby demonstrate that this novel LSW model can be of use in systems neuroscience applications.

1 Introduction

It is often of interest to study the interdependence between two simultaneously recorded time series. However, in the majority of applications we cannot assume that the underlying data are stationary. For example, we later consider a problem from neuroscience in which population local field potential (LFP) recordings were made from two areas of a rat's brain as it performed a maze-based task designed to invoke spatial working memory and decision-making. Estimating the time-varying dependence structure between these signals can provide direct insight into the timing and fundamental nature of interactions between brain regions (Varela et al., 2001); in our example, the dynamic modulation of hippocampal and cortical activities is likely to reflect behaviour-dependent interactions between the two structures (Jones and Wilson, 2005). However, since the rat's behaviour, sensory environment and the excitability and rhythmicity of its neuronal networks can vary over sub-second timescales, it is likely that the resulting time series and the dependence between them will also be non-stationary. Dynamic methods are therefore necessary to capture the subtleties of neuronal network coordination.

We begin by introducing some relevant concepts from stationary time series analysis. Given two stationary time series, $\{X_t^{(1)}\}$ and $\{X_t^{(2)}\}$, their cross-covariance function has

a Fourier representation in terms of the cross-spectral density function, $f^{(1,2)}(\omega)$:

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

The cross-covariance function provides a natural estimate of the relationship between the two series in the time domain, while the cross-spectral density function can be used similarly in the spectral domain (Brillinger, 1975). The coherence function is derived by normalising the cross-spectrum by the individual spectra and, roughly speaking, measures the correlation between the signals as a function of frequency.

Although the coherence function provides a good option for stationary series, other methods are necessary when the data is non-stationary. An extension to Fourier analysis, to allow for non-stationarity, is the windowed, or short time Fourier transform (Gabor, 1946) which estimates the spectrum over time as well as frequency by applying the Fourier transform to a localised time window that slides along the time axis; the data within each time window is presumed stationary. One drawback to this approach is that the window width is constant over all frequencies. It is desirable for the time window to change so that we use a larger window to observe more information for lower frequencies, and a smaller window for more precise time resolution at high frequencies. Complex demodulation was introduced by Tukey (1961) and allows the frequency components of a time series to be described as a function of time. This method can also be used for bivariate situations, but unlike conventional spectral analysis which estimates components at multiple frequencies simultaneously, complex demodulation estimates the contribution of one specific reference frequency. Bivariate ARMA models with time dependent parameters have been used to estimate coherence (Schack and Krause, 1995). Complications with this procedure arise as it is necessary to select the order of the ARMA model which may change over time. Dahlhaus (2000) introduces a method for modelling multivariate processes based on the locally stationary process model (Dahlhaus, 1997). Other methods for the analysis of non-stationary bivariate time series include the Auto-SLEX method (Ombao et al., 2001). Using this method, the data is automatically segmented into approximately stationary dyadic blocks. SLEX basis waveforms are used instead of the Fourier exponentials of the classical representation, allowing estimation of the time-varying spectra and coherence.

Due to the natural localisation in both time and scale, wavelets are a popular tool for modelling the dependence between two non-stationary series; a thorough introduction to wavelets can be found in Vidakovic (1999). Unlike time resolved Fourier coherence, the wavelet transform uses shorter windows for higher frequencies, which leads to more 'natural' localisation. The concept of the wavelet cross spectrum, in terms of the continuous wavelet transform (CWT), was introduced by Hudgins et al. (1993). Wavelet cross spectral analysis using the CWT has been applied to applications such as climatology (Maraun and Kurths (2004), Grinsted et al. (2004)) and neuroscience (Lachaux et al., 2002). Wavelet cross-covariance and correlation has also been defined based on the maximal overlap discrete wavelet transform (MODWT) (see Whitcher et al. (2000) and Serroukh and Walden (2000)). Their formulation assumes that the d'th order backwards

differences of the series can be modelled as a stationary process. A similar concept to that of cross-wavelet analysis is that of Hilbert wavelet pairs (Whitcher et al., 2005). Using this approach, the coherence and phase between the signals can be estimated with less redundancy than with the CWT.

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

In section 2 we introduce the joint LSW process model and introduce several related quantities, such as the evolutionary wavelet cross spectrum and the local cross-covariance, and also recall briefly the univariate equivalents to these quantities. Section 3 details the estimation procedure based on the LSW periodograms and cross-periodogram. In Section 4 we then illustrate the methodology with a simulated example and application to experimental neurophysiological data.

2 Wavelet coherence using the LSW model

2.1 Definition of the LSW model for a bivariate time series

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

Definition 2.1. The joint LSW process $(X_{t,T}^{(1)}, X_{t,T}^{(2)})_{t=0,\ldots,T-1}$, for $T = 2^j \geq 1$ is a triangular stochastic array with representation

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

where $\{\psi_{k,t}\}$ are discrete, real valued, compactly supported, non-decimated wavelet vectors with scale and location parameters $j \in \{-1, -2, \ldots\}$ and $k \in Z$, respectively. For each $j < -1$, the functions $W_j^{(i)}(k/T)$ are assumed to be Lipschitz continuous with Lipschitz constants, L_j, fulfilling $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$. Similarly, $\rho_j(k/T)$ are assumed to

be Lipschitz continuous with Lipschitz constants, R_j, fulfilling $\sum_{j=-\infty}^{-1} 2^{-j} R_j < \infty$. The functions are defined on *rescaled time* $z = k/T \in [0,1]$ which enables asymptotic estimation. Also, $\xi_{j,k}^{(i)}$ are zero mean orthonormal identically distributed random variables with the following properties

- $cov(\xi_{j,k}^{(i)}, \xi_{j',k'}^{(i)}) = \delta_{j,j'} \delta_{k,k'}$

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

where $\delta_{i,j}$ is the Kronecker delta function, giving $\delta_{i,j} = 1$ for $i = j$ and 0 otherwise.

In Definition 2.1, the parameters $W_j^{(i)}(k/T)$ can be thought of as time and scale dependent transfer functions and the non-decimated wavelet vectors, ψ_j, can be thought of as building blocks analogous to Fourier exponentials in the spectral domain representation. Each process is composed of the sum of components over wavelet scale, j, and thus provides a multiscale representation of the process. The quantity $\rho_j(k/T)$ is a direct measure of the dependence between the innovation sequences of each process at scale j. Therefore, unlike other measures of dependence such as the cross-correlation function, $\rho_j(k/T)$ does not depend on the parameters of the model. This dependence measure is our main quantity of interest within the model.

Haar wavelets are perhaps the simplest example of a wavelet system that can be used in this formulation and they are defined by:

$$\psi_{j,k} = 2^{j/2} I_{0,\dots,2^{-j-1}-1}(k) - 2^{j/2} I_{2^{-j-1},\dots,2^{-j}-1}(k)$$

Here the notation $j = -1$ denotes the finest scale wavelet, $j = -2$ the next finest scale and so forth. Other real-valued Daubechies' compactly supported wavelets (Daubechies, 1992) can also be used. Throughout the paper we assume that the errors are normally distributed. In principle other distributions could be used, however this assumption is made to ensure theoretical tractability.

We do not generally deal with the parameters $W_j^{(i)}(k/T)$ directly, instead they are described by a quantity termed the evolutionary wavelet spectrum. The evolutionary wavelet spectrum of the process $X_{t,T}^{(i)}$ is defined by $S_j^{(i)}(z) = |W_j^{(i)}(z)|^2$. Analogous to the spectrum, $f(\omega)$, in the classical stationary representation, the evolutionary wavelet spectrum quantifies the contribution to variance within a LSW process over scale, j, and rescaled time, $z = k/T$. The EWS is the main quantity of interest within the univariate LSW framework, however in the bivariate setting we are interested not only in these quantities, but also in the dependence between the series as quantified by the LSW coherence, $\rho_j(k/T)$. The LSW coherence is analogous to the classical coherence, $K(\omega)$, in spectral domain setting and quantifies the dependence between the series over scale and time. To demonstrate how these quantities provide a description of the joint process, we provide a brief example.
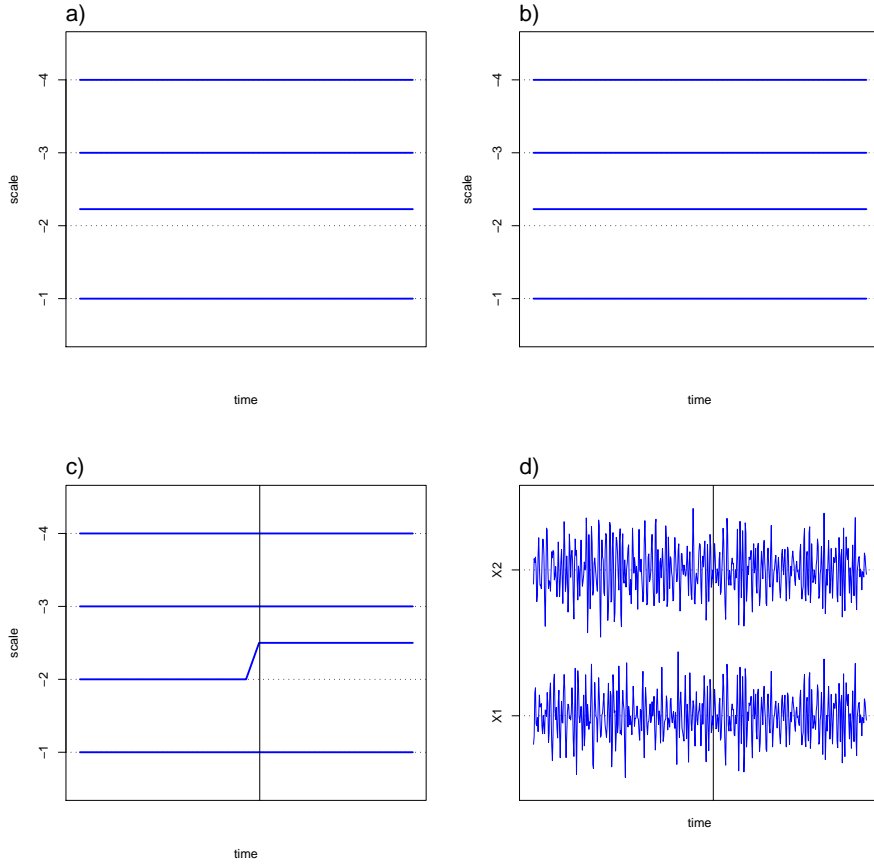
Figure 1: Example of the construction of an LSW process. a) EWS of $X^{(1)}$, b) EWS of $X^{(2)}$, c) LSW coherence between series, d) simulation of $X^{(1)}$ and $X^{(2)}$.

We start by specifying the EWS of two processes, and the dependence between them as measured by $\rho_j(k/T)$. Suppose that for both series the spectrum is zero at all scales apart from $j = -2$. At this scale, both spectra are constant at $S^{(1)}_{-2}(z) = S^{(2)}_{-2}(z) = 5$. The processes are assumed to have zero coherence at all scales apart from $j = -2$, and at this scale the LSW coherence is 0 for the first half of the series, then increases to 1 for the latter half of the series. The EWS and coherence are plotted in Figure 2, along with one simulation of time series with these properties. The resulting series are independent until time $T/2$, and then exactly the same for the latter half of the series since the coherence is specified to be 1 at the only non-zero scale.

2.2 The evolutionary wavelet cross-spectrum and coherence

The evolutionary wavelet cross-spectrum provides a measure of the dependence between two time series. It provides a natural stage in estimating the LSW coherence and is defined as follows.

Definition 2.2. The *evolutionary wavelet cross-spectrum* of the LSW processes $X_{t,T}^{(1)}$ and $X_{t,T}^{(2)}$, is given by

$$C_j(z) = W_j^{(1)}(z)W_j^{(2)}(z)\rho_j(z)$$

A measure of local cross-covariance may be associated with the evolutionary wavelet cross-spectrum. As shown by Proposition 2.4, the LSW process cross-covariance given by $c_T^{(1,2)}(z,\tau) = cov(X_{t,T}^{(1)}, X_{t+\tau,T}^{(2)})$, asymptotically tends to the local cross-covariance.

Definition 2.3. The local cross-covariance of the LSW processes $X_{t,T}^{(1)}$ and $X_{t,T}^{(2)}$, is given by

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

where $\Psi_j(\tau) = \sum_k \psi_{jk}(0)\psi_{jk}(\tau)$ is the autocorrelation wavelet.

Proposition 2.4. *Assuming there exists a constant, C, such that for all j, $|W_j^{(i)}(z)| \leq C2^{j/2}$. As $T \to \infty$, then for $\tau \in Z$ and $z \in (0,1)$, $|c_T^{(1,2)}(z,\tau) - c^{(1,2)}(z,\tau)| = O(T^{-1})$.*

The assumption of Proposition 2.4 is satisfied if, for example, $X_{t,T}$ is a white noise process with spectrum given by $S_j(z) = |W_j(z)|^2 = 2^j$. The evolutionary wavelet cross-spectrum describes the contribution to the cross-covariance at a particular time, z, and scale, j. This interpretation becomes clear when considering the cross-covariance at zero lag. Since $\Psi_j(0) = 1$, substituting into equation (1) we have $c^{(1,2)}(z,0) = \sum_{j=-\infty}^{-1} C_j(z)$. The relationship between the cross-spectrum and the local cross-covariance is invertible, with the inverse relationship given by

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

where $A_{i,j}$ is the autocorrelation wavelet inner product matrix $A_{i,j} = \sum_\tau \Psi_i(\tau)\Psi_j(\tau)$. The evolutionary wavelet cross-spectrum, $C_j(z)$, provides a measure of the dependence between the two series. However it is clear that deviations in $C_j(z)$ can be caused by fluctuations in the EWS of each of the processes as well as by changes in the dependence, and hence the cross-spectrum cannot be used alone as a measure of dependence between the series.

The locally stationary wavelet coherence, $\rho_j(z)$, can be represented in terms of the evolutionary wavelet cross-spectrum and the individual EWS of each process as shown in equation (3), providing a normalised measure of the relationship between two series.

$$\rho_j(z) = \frac{C_j(z)}{\sqrt{S_j^{(1)}(z)S_j^{(2)}(z)}} \tag{3}$$

The locally stationary wavelet coherence ranges from -1, indicating complete negative correlation, to $+1$ indicating complete correlation. A value of close to zero indicates a lack of correlation between the two series at the given scale and location.

2.3 Relationship between LSW and Fourier coherence

In order to understand the relationship between the Fourier and LSW coherence, we express both quantities in terms of the covariance function. Using the definition of the LSW cross spectra from equation (3), and the corresponding representation of the auto spectra, we obtain the following expression for the LSW coherence:

$$\rho_j(z) = \frac{\sum_\tau c^{(1,2)}(z,\tau) \sum_l A_{jl}^{-1} \Psi_l(\tau)}{\sqrt{\sum_\tau c^{(1)}(z,\tau) \sum_l A_{jl}^{-1} \Psi_l(\tau) \sum_\tau c^{(2)}(z,\tau) \sum_l A_{jl}^{-1} \Psi_l(\tau)}} \tag{4}$$

The equivalent representation for the Fourier coherence is:

$$K(\omega) = \frac{|\sum_\tau c_X^{(1,2)}(\tau) e^{-i\omega\tau}|}{\sqrt{\sum_\tau c_X^{(1)}(\tau) cos(\omega\tau)} \sqrt{\sum_\tau c_X^{(2)}(\tau) cos(\omega\tau)}} \tag{5}$$

Since in the LSW setting we have a measure of the localised covariance, the LSW coherence is defined over rescaled time, $z = k/T$, giving a localised measure of the coherence. In the case where the series is stationary, there is no dependence on z in equation (4) and the two representations become more comparable.

Comparing equations (4) and (5) in the stationary setting, we see that the functions $\sum_l A_{jl}^{-1} \Psi_l(\tau)$ in the LSW representation perform the same task as the Fourier exponentials in (4). These functions are a modified version of the autocorrelation wavelets, $\Psi_j(\tau)$ and unlike the Fourier exponentials which have a constant amplitude over the the entire length of T, the amplitude of these functions decay as $|\tau| \to \infty$. The rate of decay depends on the scale, j, with the functions decaying at a much faster rate for fine scales. Putting more weight on small lags of τ makes sense in the locally-stationary setting as our estimators of the spectra of each process are defined locally, and so in this setting the associated dependencies are are only significant locally.

3 Estimation

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

Definition 3.1. The wavelet periodograms for the LSW processes $X_{t,T}^{(i)}$, for $i = 1, 2$, are given by $I_{j,t,T}^{(i)} = |d_{j,t,T}^{(i)}|^2$. The wavelet cross-periodogram is given by $I_{j,t,T}^{(1,2)} = d_{j,t,T}^{(1)} d_{j,t,T}^{(2)}$.

Proposition 3.2. *Assume there exists a constant, C, such that for all j, $|W_j^{(i)}(z)| \le C2^{j/2}$. The expectation of the cross-periodogram, $I_{j,t,T}^{(1,2)}$, is given by*

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

Also, the variance is given by

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

We can see from Proposition 3.2 that the expectation of the wavelet cross-periodogram is composed of the sum of wavelet cross-spectra, $C_j(z)$. The cross-periodogram is therefore a natural estimator of the wavelet cross-spectrum, but we first need to correct for the bias incurred by the matrix $A_{i,j}$. Also, since the cross-periodogram has non-vanishing variance, it needs to be smoothed to obtain consistency. For this we use simple moving average smoothing. Other more advanced smoothing techniques such as DWT shrinkage are also potentially viable.

The estimator of the cross-spectrum is therefore constructed by firstly correcting the periodogram to give $\tilde{I}_l(t/T) = \sum_{j=-J*}^{-1} I_{j,t,T}^{(1,2)} A_{l,j}^{-1}$ for some $J^* < \log_2(T)$, chosen to ensure the consistency of $\tilde{I}_l(z)$, then smoothing over time to give $\hat{C}_l(t/T) = \frac{1}{2M+1} \sum_{m=-M}^{M} \tilde{I}_{l,t+m,T}^{(1,2)}$

Proposition 3.3. *Suppose that the assumptions from Proposition 1 and 2 hold, and let $J^* = \alpha \log_2(T)$ where $\alpha \in (0,1)$. The estimator $\hat{C}_l(t/T)$ converges in probability to $W_l^{(1)}(t/T)W_l^{(2)}(t/T)\rho_l(t/T)$ for each fixed scale l, provided that $MT^{\alpha-1} \to 0$ as $T \to \infty$ and $M \to \infty$.*

The wavelet periodograms, $I_{j,t,T}^{(i)}$ for $i = 1,2$ are smoothed and corrected similarly to give $\hat{S}_l^1(t/T)$ and $\hat{S}_l^2(t/T)$.

Proposition 3.4. *Suppose that the assumptions from Proposition 1 and 2 hold, and let $J^* = \alpha \log_2(T)$ where $\alpha \in (0,1)$. Then $\hat{S}_l^{(i)}(t/T)$ converges in probability to $S_l^{(i)}(t/T)$ provided that $MT^{\alpha-1} \to 0$ as $T \to \infty$ and $M \to \infty$ for each fixed scale l.*

From Propositions 3.3 and 3.4, we see that as the length of the series increases, our estimators of the cross and auto spectra converge in probability to the expected quantities. Given a sufficiently large length of series, this is valid for any choice of smoothing parameter, M.

Given estimates of the cross-spectrum, $\hat{C}_l(t/T)$, and individual spectra, $\hat{S}_l^{(i)}(t/T)$, of each process and provided that $S_l^{(1)}(t/T) > 0$ and $S_l^{(2)}(t/T) > 0$, the estimator of the

locally stationary wavelet coherence given by

$$\widehat{\rho}_l(t/T) = \frac{\widehat{C_l}(t/T)}{\sqrt{\widehat{S}_l^{(1)}(t/T)\widehat{S}_l^{(2)}(t/T)}}$$

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

3.1 Practical considerations

The corrected and time-smoothed periodograms provide a consistent estimator of the LSW cross-spectrum. In practice, however, the estimator can become unstable. This is largely because after the correction step in the estimation procedure it is possible that the estimator may take values close to or below zero. There are several options to overcoming this problem. For example we could bound the denominator to be above zero, or correct the spectra using a non-negative constraint such as with non-negative least squares (Lawson and Hanson, 1974).

We recommend overcoming the problems with stability of the estimator by smoothing the cross-periodogram over scale as well as time. Though this introduces extra bias into the estimator, we can minimise this bias by suitable choice of the smoothing parameters. Smoothing over scale reduces the variance of the estimator further and enables us to form an estimator of the cross-spectrum that is stable upon normalisation.

We smooth over scale allowing a different vector of weights for each scale. This produces a $J \times J$ matrix of weights, denoted D, where the element $D_{l,j}$ is the contribution of scale j, to the smoothed estimate of scale l, and the columns thus sum to 1. Correcting for bias using the bias correction matrix A^{-1} has the effect of applying a scale factor of about 2^j to scale j, so that fine scales are scaled up in comparison to coarse scales. Since we choose to smooth over scale after correcting, we smooth using the following scheme: a) Firstly the estimates are re-scaled by 2^{-j}, b) the estimates are smoothed over scale, c) the scaling factor of 2^{-j} is re-introduced. The corrected, scale smoothed parameter is therefore given by:

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

In practice, we generally smooth over the neighbouring 3 or 5 scales. Different choices of smoothing weights are allowed for the estimates of each scale. For fine scales it is often not necessary to smooth over scale at all so that at fine scales of l we may have $D_{l,l} = 1$. Details of our choice of smoothing parameters are provided in the examples.

There are also other issues that can affect the quality of the estimate, namely the choice of wavelet and the choice of time-smoothing parameters, M_j. The wavelet can be chosen according to the properties of the time series that it is representing. For example if the series contains sharp jumps then the Haar wavelet might be preferable, whereas for series with smoother features other wavelets, such as Daubechies' least asymmetric (Daubechies, 1992), might be preferable. For smoothing over time we choose to increase

the width of the smoothing window as the scale becomes coarser. This is natural since the wavelet coefficients at coarse scales display a stronger degree of auto-correlation.

4 Applications

4.1 Simulated example

As an example, we apply our method to data simulated from a bivariate LSW process with a known, non-stationary coherence structure that varies between scales. For even scales we take $\rho_j(k/T) = 0.2$, and for odd scales we assume a non-stationary structure that forms an 'inverted v' between 0.2 and 0.8. The autospectra of each process are taken to be $S_j^{(1)}(z) = S_j^{(2)}(z) = 2^j$ (i.e. that of white noise). In this case we take $T = 2^{13}$, giving a decomposition of 13 scales. The results from one such simulation, using Daubechies least asymmetric wavelet with support $N = 5$, are shown in Figure 2 below. The estimates are smoothed over scale using weights on the leading scales of $D_{l,l} = (0.95, 0.95, 0.95, 0.9, 0.9, 0.9)$ for $l = 1$ to 6. For each estimate we smooth over the neighboring 5 scales, using larger weightings for scales immediately next to the scale of interest. So for scale -4, for example, we take the off-diagonal weights to be $D_{,4} = (0, \frac{1-D_{4,4}}{6}, \frac{1-D_{4,4}}{3}, D_{4,4}, \frac{1-D_{4,4}}{3}, \frac{1-D_{4,4}}{6}, 0, ..., 0)$. For smoothing over time, we use a bandwidth of $M_j/T = (0.025, 0.05, 0.075, 0.1, 0.125, 0.15)$. The estimated coherence structure from one realisation of the process is shown in Figure **??**. The estimated coherence follows the true coherence closely.

4.2 Application to neuroscience data

We now illustrate our method with application to experimental neuroscience data, previously described by Jones and Wilson (2005). We consider the coherence between local field potentials (LFP) in two functionally and anatomically connected areas of a rat's brain: the hippocampus and the prefrontal cortex. The LFP provide a measure of averaged activity over local neuronal populations, and the estimated LSW coherence presents an indication of the extent to which activities in the two areas are coordinated, decomposed over scale and over time. The data consist of 13 trials, with each trial comprising a 'forced turn' epoch and a 'choice' epoch. Each of these epochs are 6s in duration, corresponding to the section of the task when the rat is moving along the central arm of the maze, immediately prior to reaching a T-junction decision point. During forced turn runs the direction at the T-junction is pre-determined by a movable barrier, whereas in choice epochs the rat is free to choose either direction and so presumably employs cognitive processes including spatial working memory and decision-making. It is expected that hippocampal-prefrontal interactions are thereby selectively recruited during forced-turn runs. Trials in which the rat made the incorrect decision at the turning point were not included in the analysis.

The LFP data are sampled at 625Hz, however in Jones and Wilson (2005), significant coherence was found in the theta frequency range ($4 - 12Hz$) so we choose to downsample

the data by taking every third measurement. Using the approximation that a wavelet scale j, corresponds to Fourier frequencies in the interval $(\frac{2^{j-1}}{\Delta t}, \frac{2^j}{\Delta t})$ (Percival and Walden, 2000, Chapter 4.), this downsampling ensures that the theta frequency band is contained within scales -4 and -5. We smooth over all 3 subsamples to ensure that information is not lost. Other amounts of downsampling were also investigated to ensure that features were not missed, but this extra analysis offered no additional insight.

Since choice runs invoke spatial working memory, we would expect to see a change in correlation structure when the rat reaches the decision point. For forced runs we would not expect to observe this change. The LSW coherence averaged over all 13 trials is shown in Figure 3. These estimates have been produced using Daubechies extremal phase wavelet (N=8), scale-smoothing with a weighting of $D_{l,l} = (0.95, 0.9, 0.9, 0.9, 0.85, 0.75)$ on the leading terms and off diagonal elements as with the previous example. For smoothing over scale we used a bandwidth of $Mj/T = (0.05, 0.1, 0.15, 0.2, 0.25, 1)$ so that we are smoothing over an increased number of observations as the scale becomes coarser. Figure 3 shows an area of increased coherence towards the end of the series at scale -4 for the choice run data. This feature is not observed in the forced run data.

Since we are interested in whether the dependence structure changes over the course of the experiment (when the rat is forced to make a decision), we judge the significance of our results by looking at the mean coherence before and after the decision point. Confidence intervals are constructed by simulation: given the estimated coherence structure $\{\hat{\rho}_j(z)\}_{j=-\infty}^{-1}$ and autospectra $\{\hat{S}_j^{(1)}(z)\}_{j=-\infty}^{-1}$, $\{\hat{S}_j^{(2)}(z)\}_{j=-\infty}^{-1}$, we simulate n new series. For each series we take the same estimator, the mean before and after the decision point, and then take the $n\alpha$ and $n(1-\alpha)$ ordered values to give the lower and upper confidence limits. The estimate and resulting 90% confidence intervals for scale -4 of the choice data is shown in Figure 4. The mean coherence is significantly different between the two sections of the data, suggesting that there is a significant change in interaction between the hippocampus and prefrontal cortex before and after the decision point. This effect was only observed at scale -4 on the choice run data sets.
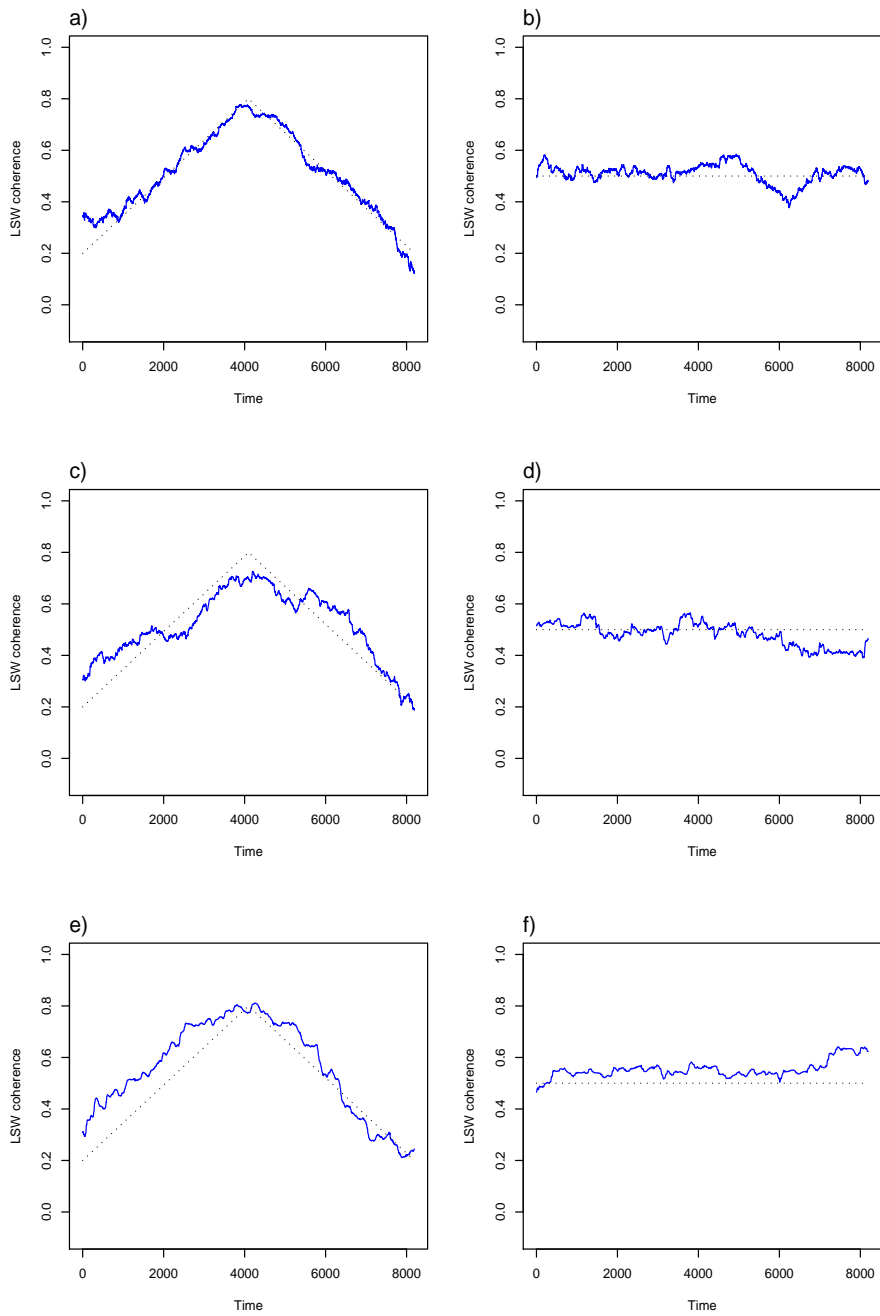
Figure 2: Results showing the true coherence (dashed) and estimated coherence (solid) from the simulation example. Subplots a) to f) correspond to scales -1 to -6 respectively.
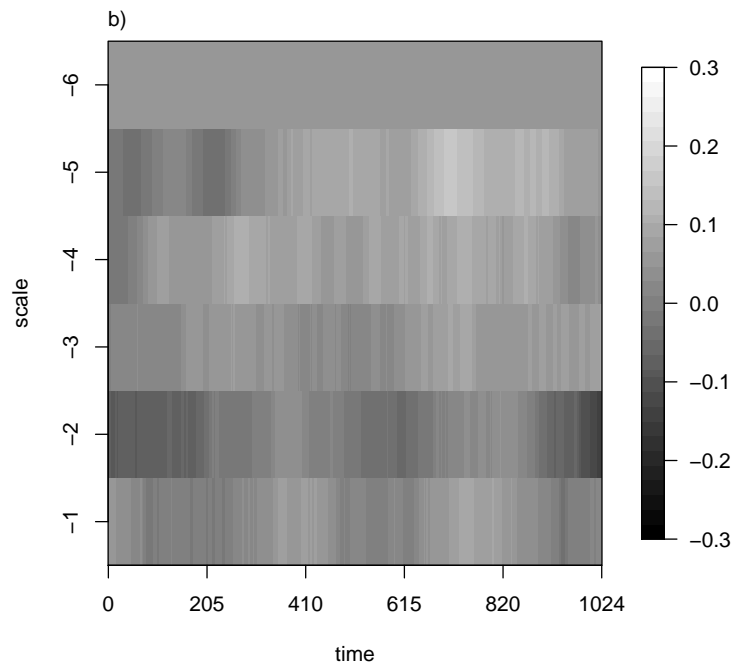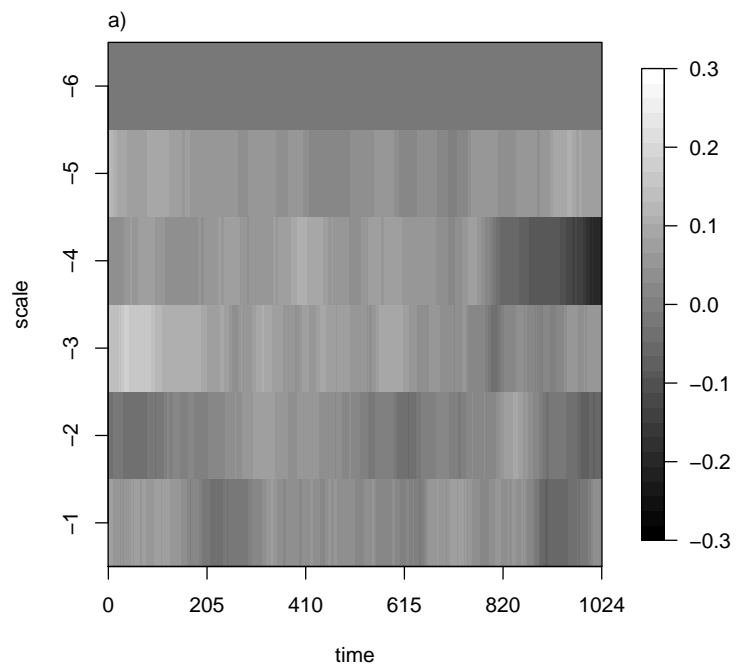
Figure 3: Trial averages for the choice runs (a) and forced runs (b), scales −1 to −6 using Daubechies extremal phase wavelet, N=8.
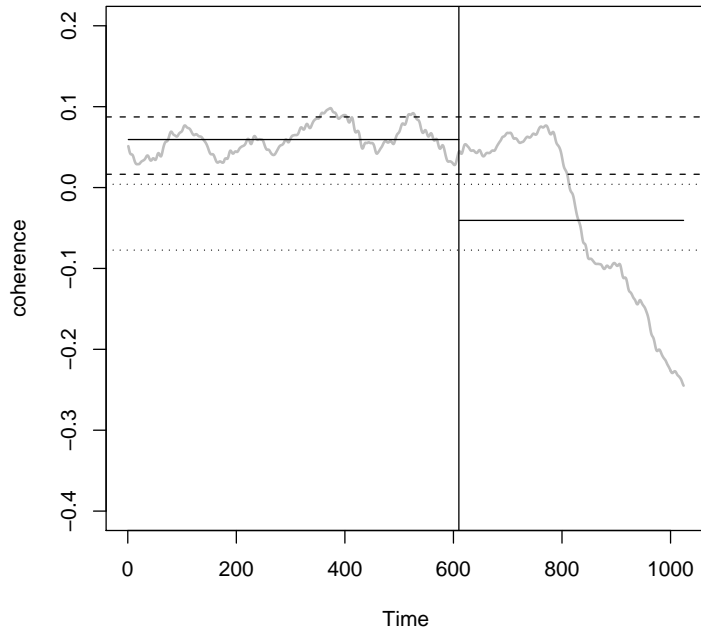
Figure 4: LSW coherence for scale −4. The decision point is marked with a vertical line, means before and after this point are plotted in bold and the associated confidence intervals are plotted with dashed lines.

5   Discussion

We have shown that the LSW theory can be usefully extended to the bivariate setting, providing a measure of dependence between non-stationary time series. In the described neuroscience application the method is applied to trials of 6s in duration and the findings are in agreement with the previous results from Jones and Wilson (2005) using the short time Fourier transform. With the short time Fourier transform it is necessary to take a narrow window width in order to assume stationarity. This means that the frequency resolution will be poor in comparison to results from wavelet analysis where the width of the analysing wavelet is naturally adjusted to scale. Our method of LSW coherence can also be easily applied to much longer series without the problem of choosing a fixed window length that is hard to tune over a broad range of frequencies. Also of note, our LSW model demonstrates significant variation in hippocampal-prefrontal coherence averaged over a limited dataset of only 13 trials; this is a valuable facet when analyses are applied to experimental data.

LSW coherence provides a representation of the dependence structure with less redundancy than similar methods using the CWT. Methods using complex continuous wavelets do however have the advantage that this allows for an estimation of the phase between the signals. Estimates of phase offsets can be useful in neuroscience, since they allow

14

us to make inferences regarding the connectivity of neuronal networks (for example, the synaptic time delay between connected structures). The sign of our estimate of $\rho_j(z)$ can provide some information; in particular changes in the sign of the estimated coherence would imply a shift of phase relationships. The extension of our method to allow for phase estimation would be very interesting and provide a more complete estimate of the relationship between the signals. This is possible in the model, perhaps through the use of complex wavelets.

Another important consideration is the application to multivariate data. Existing methods of assessing multivariate dependence include those of Dahlhaus (2000), who extended univariate locally stationary processes to a multivariate setting, Ombao et al. (2005) which is an extension to the SLEX framework and Guo and Dai (2006) who use a Cholesky decomposition approach. The application of our method to multivariate cases is possible using the current methodology, by taking pairwise combinations of multivariate series. Generalisation to a fully multivariate model is also possible and will be a consideration in future work that allows application to multi-site recordings of neuronal activity from disparate brain structures.

A   Appendix

For use in the following proofs, we first justify the Lipschitz property of the product $W_i^{(1)}(k/T)W_i^{(2)}(k/T)\rho_i(k/T)$.

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

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

Proof of Proposition 2.4

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

15

using Lemma A.1. The support of $\Psi_j(\tau)$ is bounded by $K2^{-j}$, and so the distance $|k - t|$ is bounded by this amount also. Since $\sum_j B_j 2^{-j} < \infty$ and $\Psi_j(\tau) = O(1)$, this gives

$$|c_T^{(1,2)}(z,\tau) - c^{(1,2)}(z,\tau)| = |T^{-1}CK \sum_{j=-\infty}^{-1} B_j 2^{-j} \Psi_j(\tau)| = O(T^{-1})$$

Proof of Proposition 3.2 The process cross-covariance is given by

$$
\begin{aligned}
E(I_{j,t,T}^{(1,2)}) &= E\Big[\sum_s \sum_{i=-\infty}^{-1} \sum_{k=\infty}^{\infty} W_i^{(1)}(k/T)\psi_{i,s-k}\xi_{i,k}^{(1)}\psi_{j,s-t} \sum_s \sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_i^{(2)}(k/T)\psi_{i,s-k}\xi_{i,k}^{(2)}\psi_{j,s-t}\Big] \\
&= E\Big[\sum_{i=-\infty}^{-1} \sum_{k=\infty}^{\infty} W_i^{(1)}(k/T)\Psi_{i,j}(k-t)\xi_{i,k}^{(1)} \sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_i^{(2)}(k/T)\Psi_{i,j}(k-t)\xi_{i,k}^{(2)}\Big] \\
&= \sum_{i=-\infty}^{-1} \sum_{k=\infty}^{\infty} W_i^{(1)}(k/T)W_i^{(2)}(k/T)\rho_i(k/T)\Psi_{i,j}^2(k-t) \\
&= \sum_{i=-\infty}^{-1} \Big(W_i^{(1)}(t/T)W_i^{(2)}(t/T)\rho_i(t/T) + O(T^{-1}B_i(2^{-i}+2^{-j}))\Big) \sum_{k=-\infty}^{\infty} \Psi_{i,j}^2(k-t) \\
&= \sum_{i=-\infty}^{-1} W_i^{(1)}(t/T)W_i^{(2)}(t/T)\rho_i(t/T)A_{ij} + Rest_T
\end{aligned}
$$

since the length of support of $\Psi_{i,j}^2(k-t)$ is bounded by $K(2^{-i}+2^{-j})$ and $cov(\xi_{j,k}^{(1)},\xi_{j',k'}^{(2)}) = \delta_{j,j'}\delta_{k,k'}\rho_j(k/T)$. The remainder is given by

$$
\begin{aligned}
Rest_T &\leq \sum_{i=-\infty}^{-1} T^{-1}CKb_i(2^{-i}+2^{-j})A_{ij} \\
&\leq T^{-1}2^{-j}CK \sum_{i=-\infty}^{-1} B_i 2^{-i} \sum_{k=-\infty}^{-1} 2^k A_{ik} + T^{-1}2^{-j}CK \sum_{i=-\infty}^{-1} B_i 2^{-i} \sum_{k=-\infty}^{-1} 2^k A_{kj} \\
&= O(T^{-1}2^{-j})
\end{aligned}
$$

as $\sum_{k=-\infty}^{-1} 2^k A_{ik} = 1$ and $\sum_{i=-\infty}^{-1} B_i 2^{-i} \leq \infty$. For the variance, we decompose the product using Isserlis' Theorem for zero mean Gaussian random variables (Isserlis, 1918) to give

$$
\begin{aligned}
var(I_{j,k,T}^{(1,2)}) &= E(I_{j,t,T}^{(1)})E(I_{j,t,T}^{(2)}) + E(I_{j,t,T}^{(1,2)})^2 \\
&= \sum_{i=-\infty}^{-1} S_i^{(1)}(t/T)A_{i,j} \sum_{i=-\infty}^{-1} S_i^{(2)}(t/T)A_{i,j} + \Big(\sum_{i=-\infty}^{-1} C_i(t/T)A_{ij}\Big)^2 + O(2^{-j}T^{-1})
\end{aligned}
$$

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

16

Lemma A.2. *The expectation of the smoothed cross-periodogram is given by*

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

Also, the variance is given by

$$\begin{aligned}
var(\widetilde{I}_{j,t,T}^{(1,2)}) &= \frac{1}{(2M+1)^2}\sum_{m=-M}^{M}\sum_{\tau}\Big[\sum_{i=-\infty}^{-1}S_i^{(1)}(t/T)\Big)A_{i,j}^{\tau}\sum_{i=-\infty}^{-1}S_i^{(2)}(t/T)\Big)A_{i,j}^{\tau} \\
&+ \Big(\sum_{i=-\infty}^{-1}W_i^{(1)}(t/T)W_i^{(2)}(t/T)\rho_i(t/T)A_{i,j}^{\tau}\Big)^2 + O(T^{-1}2^{-j}|m|)\Big]
\end{aligned}$$

and the MSE by

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

Proof of Lemma A.2 The expectation can be proved using very similar techniques to that of the expectation in Proposition 1. The variance is given by

$$var(\widetilde{I}_{t,T}^{(j)}) = \frac{1}{(2M+1)^2}\sum_{m=-M}^{M}\sum_{\tau}Cov\Big(I_{j,t+m,T}^{(1,2)}, I_{j,t+m+\tau,T}^{(1,2)}\Big) \qquad (6)$$

where $\tau = m' - m$. The covariance can be decomposed using Isserlis' Theorem, to give

$$\begin{aligned}
cov\Big(I_{j,t+m,T}^{(1,2)}, I_{j,t+m+\tau,T}^{(1,2)}\Big) &= E\Big(d_{j,t+m,T}^{(1)}, d_{j,t+m+\tau,T}^{(1)}\Big)E\Big(d_{j,t+m,T}^{(2)}, d_{j,t+m+\tau,T}^{(2)}\Big) \\
&+ E\Big(d_{j,t+m,T}^{(1)}, d_{j,t+m+\tau,T}^{(2)}\Big)E\Big(d_{j,t,T}^{(2)}, d_{j,t+m+\tau,T}^{(1)}\Big)
\end{aligned}$$

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

$$cov(d_{j,t+m,T}, d_{j,t+M+\tau,T}) = \sum_{i=-\infty}^{-1}S_i(t/T)A_{i,j}^{\tau} + O(T^{-1}2^{-j}|m|)$$

$$cov(d_{j,t+m,T}^{(1)}, d_{j,t+\tau+m,T}^{(2)}) = \sum_{i=-\infty}^{-1}W_i^{(1)}(t/T)W_i^{(2)}(t/T)\rho_i(t/T)A_{i,j}^{\tau} + O(T^{-1}2^{-j}|m|)$$

where $A_{i,j}^{\tau}$ is defined as $\sum_n \Psi_{i,j}(n)\Psi_{i,j}(n+\tau)$, and has the property $|\sum_j A_{i,j}^{\tau}2^j| \le 1$. Substituting the relevant terms into equation (6) and summing over m, the variance is given by

$$\begin{aligned}
var(\widetilde{I}_{j,t,T}^{(1,2)}) &= \frac{1}{(2M+1)^2}\sum_{m=-M}^{M}\sum_{\tau}\Big[\sum_{i=-\infty}^{-1}S_i^{(1)}\Big(\frac{t}{T}\Big)A_{i,j}^{\tau}\sum_{i=-\infty}^{-1}S_i^{(2)}\Big(\frac{t}{T}\Big)A_{i,j}^{\tau} \\
&+ \Big(\sum_{i=-\infty}^{-1}C_i\Big(\frac{t}{T}\Big)\Big(\frac{t}{T}\Big)A_{i,j}^{\tau}\Big)^2 + O(T^{-1}2^{-j}|m|)\Big] = \frac{1}{(2M+1)^2}\sum_{m=-M}^{M}\Big[I+II+O(T^{-1}2^{-j}|m|)\Big]
\end{aligned}$$

17

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

$$
\begin{aligned}
I & \leq \sum_{\tau} |\sum_{i=-\infty}^{-1} S_i^1\left(\frac{t}{T}\right) A_{i,j}^{\tau}| . \sum_{\tau} |\sum_{i=-\infty}^{-1} S_i^2\left(\frac{t}{T}\right) A_{i,j}^{\tau}| \\
& = \sum_{\tau} |\sum_{i=-\infty}^{-1} S_i^{(1)}(t/T) \sum_n \Psi_i(n) \Psi_j(n+\tau)| . \sum_{\tau} |\sum_{i=-\infty}^{-1} S_i^{(2)}(t/T) \sum_n \Psi_i(n) \Psi_j(n+\tau)| \\
& \leq \sum_n |c^{(1)}(t,n)| \sum_{\tau} |\Psi_j(n+\tau)| . \sum_n |c^{(2)}(t,n)| \sum_{\tau} |\Psi_j(n+\tau)| \\
& \leq K_1 2^{-j} \sum_n c^{(1)}(t,n) . K_2 2^{-j} \sum_n c^{(2)}(t,n) = O(2^{-j}) O(2^{-j}) = O(2^{-2j})
\end{aligned}
$$

assuming $\sup_{z \in [0,1]} \sum_{\tau} |c(t,\tau)| < \infty$, and using that $\sum_{\tau} |\Psi_j(\tau)| = O(2^{-j})$. The second term can be bounded similarly to $II \leq O(2^{-2j})$. Hence substituting the relevant terms, the variance is of order

$$
var(\widetilde{I}_{t,T}^{(j)}) = \frac{1}{(2M+1)^2} \sum_{m=-M}^{M} \left[O(2^{-2j}) + O(2^{-2j}) + \sum_{\tau} O(T^{-1} 2^{-j} |m|) \right] = O(T^{-1} 2^{-2j} M)
$$

and the mean squared error is given by

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

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

Proof of Proposition 3.3 The expectation of the smoothed and corrected cross-periodogram is given by

$$
\begin{aligned}
E(\hat{C}_l(t/T)) & = \sum_{j=-J*}^{-1} E(\widetilde{I}_{j,t,T}^{(1,2)}) A_{l,j}^{-1} \\
& = \sum_{j=-J*}^{-1} (\sum_{i=-\infty}^{-1} W_i^{(1)}(t/T) W_i^{(2)}(t/T) \rho_i(t/T) A_{ij} + O(T^{-1} 2^{-j} M)) A_{l,j}^{-1} \\
& = W_l^{(1)}(t/T) W_l^{(2)}(t/T) \rho_l(t/T) + O(T^{-1} M)
\end{aligned}
$$

The MSE is given by

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

where $\beta_j(z) = \sum_{i=-\infty}^{-1} W_i^{(1)}(z) W_i^{(2)}(z) \rho_i(z) A_{i,j}$. This can be split into two terms:

$$MSE(\hat{C}_l(t/T)) \le 2E\Big(\sum_{j=-J*}^{-1} (\tilde{I}_{j,t,T}^{(1,2)} - \beta_j(t/T)) A_{l,j}^{-1} \Big)^2 + 2\Big(\sum_{j=-\infty}^{-J*-1} \beta_j(t/T) A_{l,j}^{-1} \Big)^2 = I + II$$

The first term is given by

$$
\begin{aligned}
I &\le 2 \sum_{j=J*}^{-1} E(\tilde{I}_{j,t,T}^{(1,2)} - \beta_j(t/T))^2 (A_{l,j}^{-1})^2 \le 2 \sum_{j=-J*}^{-1} C2^l 2^j O(2^{-2j} T^{-1} M^2) \\
&= O(2^l T^{-1} M^2) \sum_{j=-J*}^{-1} 2^{-j} = O(2^l T^{-1} M^2).O(T^\alpha) = O(2^l T^{\alpha-1} M^2)
\end{aligned}
$$

as $E(\tilde{I}_{j,t,T}^{(1,2)} - \beta_j(t/T))^2$ is the MSE of $\tilde{I}_{j,t,T}^{(1,2)}$, found previously, and we know that $A_{l,j}^{-1} \le C2^{l/2} 2^{j/2}$. For the second term we further assume $\beta_j(t/T) < C_1$ (as in Fryzlewicz and Nason (2006)), to give

$$II \le 2\Big(\sum_{j=-\infty}^{-J*-1} C_1 2^{l/2} 2^{j/2} \Big)^2 \le 2.2^l \Big(\sum_{j=-\infty}^{-J*-1} C_1 2^{j/2} \Big)^2 = O(2^l)$$

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

References

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* Holf, Rinehart and Winston.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Stat.*, 25:1–37.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Stat.*, 28:1762–1794.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Fryzlewicz, P. and Nason, G. P. (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *J. Roy. Statist. Soc. Ser. B*, 68:611–634.

Gabor, D. (1946). Theory of communication. *J. Inst. Elec. Eng.*, 93:429–457.

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlin. Process. Geophys.*, 11:561–566.

Guo, W. and Dai, M. (2006). A likelihood approximation for locally stationary processes. *Statist. Sinica*, 16:825–845.

Hudgins, L., Friehe, A., and Mayer, M. (1993). Wavelet transforms and atmospheric turbulence. *Phys. Rev. Lett.*, 71:3279–3282.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12:134–139.

Jones, M. and Wilson, M. A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biol.*, 3:2187–2199.

Lachaux, J., Lutz, A., Rudrauf, D., Cosmelli, D., Le Van Quyen, M., Martinerie, J., and Varela, F. (2002). Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. *Neurophysiol. Clin.*, 32:157–174.

Lawson, C. L. and Hanson, B. J. (1974). *Solving Least Squares Problems.* Prentice-Hall.

Maraun, D. and Kurths, J. (2004). Cross wavelet analysis: significance testing and pitfalls. *Nonlin. Process. Geophys.*, 11:505–514.

Nason, G., von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc. B*, 62:271–292.

Ombao, H. C., Raz, J. A., von Sachs R., and Malow, B. (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Am. Stat. Assoc.*, 96:543–560.

Ombao, H. C., von Sachs R., and Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Am. Stat. Assoc.*, 100:519–531.

Percival, D. B. and Walden, A. T. (2000). *Spectral Analysis of Time Series.* Wavelet Methods for Time Series Analysis (Cambridge Series in Statistical and Probabilistic Mathematics).

Schack, B. and Krause, W. (1995). Dynamic power and coherence analysis of ultra short-term cognitive processes - a methodological study. *Brain Topogr.*, 8:127–136.

Serroukh, A. and Walden, A. T. (2000). Wavelet scale analysis of bivariate time series i: Motivation and estimation. *J. Nonparametr. Stat.*, 13:1–36.

Slutsky, E. (1925). Über stochastische Asymptoter und Grenzwerte. *Metron*, 5:3–89.

Tukey, J. (1961). Discussion emphasising the connection between analysis of variance and spectrum analysis. *Technometrics*, 3:1–29.

Varela, F., Lachaux, L., Rodriguez, E., and Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nat. Rev. Neurosci.*, 2:229–39.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* Wiley.

Whitcher, B., Craigmile, P., and Brow, P. (2005). Time-varying spectral analysis in neurophysiological time series using hilbert wavelet pairs. *Signal Process.*, 85:2065–2081.

Whitcher, B., Guttorp, P., and Percival, D. (2000). Wavelet analysis of covariance with application to atmospheric time series. *J. Geophys. Res.*, 105:14941 – 14962.