Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes

Timothy Park, Idris A. Eckley and Hernando C. Ombao

The following paper is published as:

Park, T., Eckley, I.A., Ombao, H.C., Estimating Time-Evolving Partial Coherence Between Signals via Multivariate Locally Stationary Wavelet Processes, *Signal Processing, IEEE Transactions on* , vol.62, no.20, pp.5240-5250, Oct.15, 2014.

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6868283&isnumber=6889149

Abstract

We consider the problem of estimating time-localized cross-dependence in a collection of non-stationary signals. To this end we develop the multivariate locally stationary wavelet framework which provides a time-scale decomposition of the signals and thus naturally captures the time-evolving scale-specific cross-dependence between components of the signals. Under the proposed model, we rigorously define and estimate two forms of cross-dependence measures: wavelet coherence and wavelet partial coherence. These dependence measures differ in a subtle but important way. The former is a broad measure of dependence which may include indirect associations, i.e. dependence between a pair of signals that is driven by another signal. Conversely, wavelet partial coherence measures direct linear association between a pair of signals, i.e. it removes the linear effect of other observed signals. Our time-scale wavelet partial coherence estimation scheme thus provides a mechanism for identifying hidden dynamic relationships within a network of non-stationary signals, as we demonstrate on electroencephalograms recorded in a visual-motor experiment.

1 Introduction

Historically much of the literature on non-stationary signals is focused on the univariate setting. For reviews of this area see Cohen (1989); Dahlhaus (2012); Daubechies (1990); Kayhan et al. (1994); Kumar and Fuhrmann (1992); Priestley (1988) and references therein. However with advanced data collection devices such as those used in the medical and mobile sectors, there is a need for rigorous approaches to assess and confirm time-localized direct vs. indirect dependence (or lack thereof) between signals. It is often difficult to infer dynamic cross-dependence between components of multivariate signals such as the multi-channel EEG collected during a visual-motor task (see Figure 1) which we will revisit later.

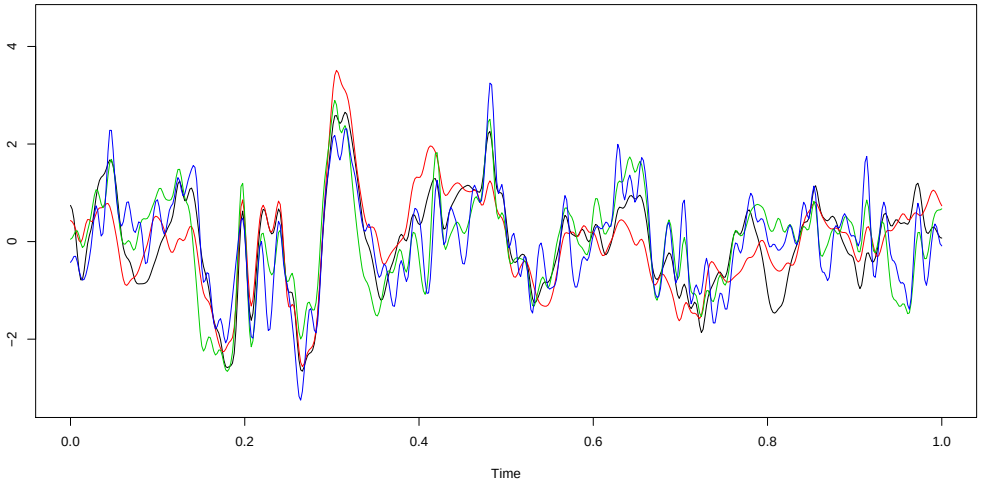


Figure 1: Plot of a 4-channel EEG.

We consider precisely this challenge, developing a novel approach for characterizing and estimating cross-dependence between non-stationary signals having dynamic and complex cross-dependence structures. In doing so, we highlight two specific forms of dependence which can be estimated between pairs of signals within a multivariate collection. The simplest form is that of the (time-dependent) coherence between two signals. This describes the linear relationship between two signals - more precisely it is a time-evolving squared cross-correlation between filtered signals, Ombao and Van Bellegem (2008). However, in so doing we may also include indirect associations driven by

another observed signal in the collection. The alternative is partial coherence. This provides a measure of the direct linear relationship between two signals over time, thus removing the (linear) effects of other observed signals. The difference between direct vs indirect associations is illustrated in Figure 2. This measure has broad potential scientific impact, for example the the neuroscience and genomic communities are keenly interested in such associations.

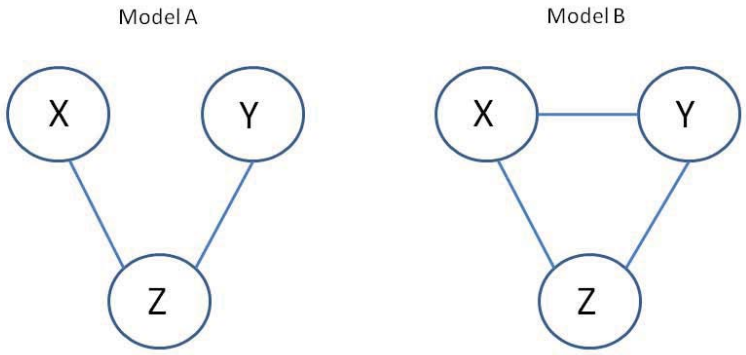


Figure 2: Indirect vs. Direct Associations Between Signals. Left: X and Y are indirectly linked through Z. Right: X and Y are directly linked. Coherence between X and Y is non-zero for both networks. Partial coherence is non-zero for the network on the right (with direct link) but zero for the left network because the link between X and Y is indirect.

Previous Work In recent years, several papers have appeared trying to address the non-stationary modelling challenge associated with such large and complex signals. In Dahlhaus (2000a), Dahlhaus presents a Fourier based model for multivariate locally stationary signals with time-varying spectral structure. A similar approach was also developed by Walden and Cohen (2012). Under the Dahlhaus framework, Ombao and Van Bellegem Ombao and Van Bellegem (2008) demonstrate that the time-varying coherence is equivalent to the modulus-squared cross-correlation between filtered segmented signals. Segment sizes are obtained data-adaptively by iteratively increasing segment lengths as long as the stationarity assumption within each segment is not violated. Such a data-adaptive windowing approach, however, is computationally demanding. An alternative Fourier based approach

to model multivariate non-stationary series is the smooth localized complex exponential (SLEX) model of Ombao et al. Ombao et al. (2005). Here the best representation of the signal is selected from the SLEX library using a complexity-penalized Kullback-Leibler criterion. Although capable of handling massive signals, the SLEX method is restricted to choosing representations obtained from temporally-dyadic segmentation. Moreover we note that both Ombao and Van Bellegem (2008) and Ombao et al. (2005) only develop methods for the estimation of coherence which, as we shall show later, can mask understanding of the direct relationships between pairs of signal components.

Cohen and Walden Cohen and Walden (2010) overcome the limitations of dyadic temporal splits within SLEX by using a wavelet basis to adapt to nonstationarity in the spectra of each channel for the case of jointly stationary processes. The assumption of jointly stationary processes is not present in Cohen and Walden (2011) and Sanderson et al. (2010) who both use wavelet based models to quantify non-stationary linear dependence between components of a *bivariate* non-stationary signals. More recently, within the more restricted context of changepoint detection of piecewise stationary signals, Cho and Fryzlewicz (2014) has extended the approach of Sanderson et al. (2010) to a p-variate setting. However none of these contributions directly address the issues that are germane to truly multivariate non-stationary signals (with three or more components). Specifically, as Koopmans (1964) identified in the stationary context, one major practical issue is to identify whether the (time-dependent) connection or cross-dependence between two channels is either (a.) direct or (b.) indirect (i.e., driven by another channel or common set of channels). It is this challenge which lies at the heart of this article.

Our Work: The modelling framework which we propose in this paper is an alternative formulation of the model form proposed by Sanderson et al. (2010). The model proposed by Sanderson et al. (2010) decomposes the spectral and cross-spectral structure into two different components: the within-channel structure being encapsulated within the transfer functions whilst the cross-channel structure is contained within the process innovations. Instead we propose a more parsimonious form, whereby both spectral components are described within a matrix of transfer functions. Specifically, to extract cross-dependence structures, we introduce the multivariate locally stationary wavelet

framework (MvLSW) - which is a stochastic representation that is ideally suited for non-stationary signals. This framework permits the direct estimation of both the coherence *and* partial coherence in a computationally efficient manner. In addition the framework also permits direct simulation of processes with a specific time-scale partial coherence form, including processes with abrupt changes in partial coherence. This direct simulation is necessary to perform resampling-based inference.

The format of the rest of the paper is as follows. Our main contributions are developed in Sections 2 and 3. Specifically, in Section 2.1 we develop the multivariate locally stationary wavelet framework for modelling multivariate signals. We then introduce the local wavelet spectral matrix as a representation of the properties of the signals in Section 2.2. In Section 2.3 we use the MvLSW model to develop our two key cross-dependence quantities: wavelet coherence and partial coherence. Section 3 gives detail of the estimator for the local wavelet spectral matrix as well as establishing its asymptotic properties. Finally Section 4 provides an example of how our approach can be used to identify direct time-dependent relationships between components of a signal which we demonstrate on multi-channel electroencephalograms (EEGs) recorded during a visual-motor experiment, as well as on simulated data.

2 Locally Stationary Wavelet Processes

This section describes the multivariate LSW (MvLSW) modelling framework, together with various time-scale measures which we introduce to describe the spectral and cross-spectral behaviour of such non-stationary signals. For completeness we start by briefly reminding the reader of key aspects associated with *univariate* LSW theory as introduced by Nason et al. (2000), their building blocks (discrete wavelets) and the associated evolutionary wavelet spectrum (EWS).

The key building blocks in constructing LSW processes, discrete wavelets, are founded on $\{h_k\}$ and $\{g_k\}$, the usual low and high-pass quadrature mirror filters associated with the construction of Daubechies' compactly supported continuous-time wavelets. The associated **discrete wavelets**, $\psi_j = \{\psi_{j,0}, \psi_{j,1}, \ldots, \psi_{j,N_j-1}\}$ are vectors of length N_j for scales $j \in \mathbb{N}$ which can be calculated using the following: $\psi_{1,n} = \sum_k g_{n-2k}\delta_{0,k} = g_n$ for $n = 0, \ldots, N_1 - 1$ and $\psi_{j+1,n} = \sum_k h_{n-2k}\psi_{j,k}$, for $n =$

$0, \ldots, N_{j+1} - 1$. Here $\delta_{0,k}$ is the usual Kronecker-delta function, and $N_j = (2^j - 1)(N_h - 1) + 1$ where N_h is the number of non-zero elements within the filter $\{h_k\}$. The discrete wavelets form the corner-stone of the (univariate) LSW time series model. Specifically, assume that $T = 2^J$ for some $J \in \mathbb{Z}$. Then the LSW process, $X_{t;T}$, is defined to be a sequence of (doubly-indexed) stochastic processes having the following representation in the mean-square sense:

$$X_{t;T} = \sum_{j=1}^{\infty} \sum_{k} W_j(k/T) \psi_{j,t-k} \xi_{j,k}. \tag{1}$$

As described in Nason et al. (2000), the representation consists of the discrete wavelets; $\{W_j(u)\}_{u \in (0,1)}$, a smoothly varying transfer function and $\{\xi_{j,k}\}$, a collection of zero-mean, unit-variance uncorrelated random variables. A number of smoothness assumptions are also required on the $\{W_j(\cdot)\}$ to ensure that the transfer function can be estimated (see Nason et al. (2000) for details).

The transfer function, $W_j(k/T)$, provides a measure of the time-varying contribution to the variance at a particular scale, j. Consequently, to describe the power contained at a given scale and location, Nason et al. (2000) introduce the evolutionary wavelet spectrum (EWS), $S_j(u) = |W_j(u)|^2$, for $j \in \mathbb{N}$. This can be estimated using the wavelet periodogram for a one-dimensional non-stationary signal, see Nason et al. (2000) for details.

2.1 The Multivariate LSW model

We now introduce our multivariate generalization of the LSW framework. In what follows we will refer to each (univariate) component signal as a channel. Our main goal is to develop a framework for modeling multivariate non-stationary signals under which we rigorously define the time-varying second order properties, and in particular the locally stationary cross-dependence between the different channels. In our framework we allow individual channels to experience their own uniquely localized non-stationary behaviour. More importantly we explicitly describe the potentially locally stationary correlation between channels. Under our model this correlation will be broken down into contributions from different scales. This is known as the coherence structure. It is important to be able to represent this structure adequately as it will reveal how the channels relate to each other and how this can change over time.

We start by considering a P-dimensional vector, $\mathbf{X}_{t;T} = [X^{(1)}_{t;T}, X^{(2)}_{t;T}, \ldots, X^{(P)}_{t;T}]'$, each element of which is an individual channel of the signal. To represent this signal under a multivariate model we replace the transfer function, $W_j(k/T)$, from the (univariate) LSW model with a $P \times P$ matrix of functions, $\mathbf{V}_j(k/T)$, known as the *transfer function matrix*. The innovations, $\{\xi_{jk}\}$, are also replaced by a set of random vectors, $\{\mathbf{z}_{j,k}\} = \{[z^{(1)}_{j,k}, \ldots, z^{(P)}_{j,k}]'\}$. The definition of the *multivariate LSW model* is then given as follows.

Definition 2.1.1 *The P-variate locally stationary wavelet process $\{\mathbf{X}_{t;T}\}_{\{t=0,\ldots,T-1\}}$, $T = 2^J$, $J \in \mathbb{N}$ is represented by,*

$$\mathbf{X}_{t;T} = \sum_{j=1}^{\infty} \sum_{k} \mathbf{V}_j(k/T) \psi_{j,t-k} \mathbf{z}_{j,k}, \tag{2}$$

where $\{\psi_{j,t-k}\}_{jk}$ is a set of discrete non-decimated wavelets; $\mathbf{V}_j(k/T)$ is the transfer function matrix, which is defined to have a lower-triangular form. We assume that each element of the transfer function matrix is a Lipschitz continuous function with Lipschitz constants L_j satisfying $\sum_{j=1}^{\infty} 2^j L_j^{(p,q)} < \infty$; $\mathbf{z}_{j,k}$ are uncorrelated random vectors with mean vector $\mathbf{0}$ and variance-covariance matrix equal to the $P \times P$ identity matrix.

We will henceforth drop the explicit dependence of the process on T, although naturally it will still be assumed.

Remark. The distributional property of the random elements in Definition 2.1.1 means that the elements have the following covariance property: $\operatorname{cov}\left(z^{(i)}_{j,k}, z^{(i')}_{j',k'}\right) = \delta_{i,i'}\delta_{j,j'}\delta_{k,k'}$. In other words the $\{z^{(i)}_{j,k}\}$ are random orthonormal increment sequences, which are themselves uncorrelated. Dependence between channels is encapsulated *only* in the transfer function matrix which also controls the contribution to the variance made by each channel at a particular time within each scale. This differs from the approach in Sanderson et al. (2010) where the dependence structure is encapsulated within the innovations $\mathbf{z}$.

Remark. The primary difference between our approach and that of Sanderson et al. (2010), or indeed the more recent contribution of Cho and Fryzlewicz (2014), is that in our framework we encapsulate the spectral structure (including cross-channel dependence) entirely within the transfer

function matrix. This is in contrast to the Sanderson *et al.* framework, where the spectral structure is encapsulated both within (i) the transfer functions (spectrum) and (ii) process innovations (cross-channel dependence). As such our framework permits one to estimate the *partial* coherence in a straightforward manner, since this structure is entirely embedded within the transfer function matrix. Computationally there are also benefits to this particular formulation: for example, this approach can be implemented via matrix operations, whilst in the formulation of [14] one would conduct the estimation scheme on each channel individually. More importantly, perhaps, it is possible to simulate multivariate time series with a given partial coherence form directly within this framework. The ability to perform such simulations means that resampling based inference can be performed in this setting.

Many different forms of transfer function matrix could be chosen, however for ease of interpretation we choose for it to have a lower triangular form. The lower triangular form of $\mathbf{V}_j(u)$ makes it very easy to generalize to multiple dimensions. It is also easy to see how linear dependencies between the channels are produced. If the off diagonal terms are non-zero then there will be (time-varying) dependence between the series, however if $\mathbf{V}_j(u)$ is diagonal then the channels will be uncorrelated with each other. Here, we do not estimate $\mathbf{V}_j(u)$ but estimate the spectral quantities which we discuss in the next subsection. Moreover the lower triangular form can represent a general spectral structure even if the channel order is permuted. This is explained further in Proposition 2.2.5.

2.2 Local Wavelet Spectral and Covariance Matrices of Non-Stationary signals

We next introduce the local wavelet spectral matrix which describes the time-scale decomposition of power in our multivariate time series. Recall that in the univariate LSW context the concept of an evolutionary wavelet spectrum describes a time-scale decomposition of power. Since we are dealing with multivariate signals, and have replaced the transfer function with a transfer function matrix, we will introduce its multivariate analog – the *local wavelet spectral matrix*.

Definition 2.2.1 *Let $\mathbf{X}_t$ be a MvLSW signal with associated time-dependent transfer function ma-*

trix $\mathbf{V}_j(u)$*. Then the local wavelet spectral (LWS) matrix at scale* j *and rescaled time* u *is defined to be,*

$$\mathbf{S}_j(u) = \mathbf{V}_j(u)\mathbf{V}_j'(u), \tag{3}$$

where $\mathbf{V}_j^{'}(u)$ *denotes the transpose of* $\mathbf{V}_j(u)$*.*

Remark. The LWS matrix provides a measure of the local contribution to both the variance of the channels and cross-covariance between channels made at a particular time, u, and scale, j. By the construction of Definition 2.2.1 it is clear that for any given transfer function matrix the LWS matrix is symmetric and positive semi-definite for every fixed time-scale combination. The diagonal elements of the LWS matrix are the spectra of the individual channels of the signals and are denoted $S_j^{(p,p)}(u)$. The off diagonal terms, $S_j^{(p,q)}(u)$, describe the cross-spectra between the series. It is also natural to consider whether a connection can be established between the LWS matrix and the local auto and cross-covariance. We start to explore this connection in the following definition. However prior to doing so we introduce the discrete autocorrelation wavelet, $\Psi_j(\tau)$, which is defined by $\Psi_j(\tau) \equiv \sum_k \psi_{j,k}\psi_{j,k-\tau}$ for $j \in \mathbb{N}$ and $\tau \in \mathbb{Z}$ (see Eckley and Nason (2005) for further details).

Definition 2.2.2 *Let* $c^{(p,p)}(u,\tau)$ *denote the local autocovariance of channel* p *at lag* τ *and* $c^{(p,q)}(u,\tau)$ *be the local cross-covariance between channels* p *and* q*. We can define these function in terms of the elements of the LWS matrix and the discrete autocorrelation wavelets,*

$$\begin{aligned} c^{(p,p)}(u,\tau) &= \sum_{j=1}^{\infty} S_j^{(p,p)}(u)\Psi_j(\tau), \\ c^{(p,q)}(u,\tau) &= \sum_{j=1}^{\infty} S_j^{(p,q)}(u)\Psi_j(\tau). \end{aligned} \tag{4}$$

The following proposition establishes that, up to choice of wavelet, the LWS matrix is unique for a specified MvLSW model form.

Proposition 2.2.3 *Given the corresponding MvLSW process, the LWS matrix is uniquely defined.*

Proof: See Appendix 6.

We also consider if under this definition the local auto- and cross-covariance functions exactly represent the covariance between elements of the signals.

Proposition 2.2.4 *Let $c^{(p,q)}(u,\tau)$ denote the local cross covariance stated in Definition 2.2.2. This function can also be represented, approximately, in terms of the covariance between elements of the signal because*

$$\left| c^{(p,q)}(u,\tau) \;-\; \mathrm{cov}\left(X^{(p)}_{[uT]}, X^{(q)}_{[uT]+\tau} \right) \right| = \mathcal{O}(T^{-1}).$$

Proof: See Appendix 7.

Remark. Given the lower triangular form of the transfer function matrix, $\mathbf{V}_j(u)$, it is natural to ask if the representation is reliant on a certain ordering of the channels of $\mathbf{X}_t$. It is possible to show that under any permutation of this ordering $\mathbf{X}_t$ will have a MvLSW representation and the spectral properties will be unchanged.

Proposition 2.2.5 *Let $\mathbf{X}_t$ be a MvLSW process with LWS matrix, $\mathbf{S}_j(u)$. Also let $\mathbf{X}^*_t$ be a permutation of $\mathbf{X}_t$ such that $\mathbf{X}^*_t = \mathbf{M}\mathbf{X}_t$ for some permutation matrix $\mathbf{M}$. Then the LWS matrix of $\mathbf{X}^*_t$, $\mathbf{S}^*_j(u)$ has the form $\mathbf{S}^*_j(u) = \mathbf{M}\mathbf{S}_j(u)\mathbf{M}'$.*

Proof: See Appendix 8.

2.3 Coherence and Partial Coherence within the MvLSW setting

We now introduce a measure of cross-dependence between different channels at a particular scale. We can quantify this dependence by defining the wavelet coherence between channels. For our multivariate series we will define the coherence in terms of the wavelet *coherence matrix*.

Definition 2.3.1 *For scale, j, rescaled time point, $u \in (0,1)$, the wavelet coherence matrix, $\boldsymbol{\rho}_j(u)$ is defined as,*

$$\boldsymbol{\rho}_j(u) = \mathbf{D}_j(u)\mathbf{S}_j(u)\mathbf{D}_j(u). \tag{5}$$

Here $\mathbf{S}_j(u)$ is the LWS matrix defined previously. We also define $\mathbf{D}_j(u)$ to be a diagonal matrix whose elements are $S^{(p,p)}_j(u)^{(-1/2)}$.

The (p, q) element of the wavelet coherence matrix, $\rho_j^{(p,q)}(u)$, is the coherence between channels p and q of the series. This individual element can also be expressed as,

$$\rho_j^{(p,q)}(u) = \frac{S_j^{(p,q)}(u)}{\sqrt{S_j^{(p,p)}(u) S_j^{(q,q)}(u)}}. \tag{6}$$

Remark. Given this expression it is clear that the coherence between channels will take a value between -1 and 1 at any given point in time. A value close to ± 1 indicates a strong positive/negative linear dependence between channels at that time and scale. A value close to 0 shows there is little or no linear dependence between channels. Setting $p = q$ in Equation (6) demonstrates that the diagonal elements of $\boldsymbol{\rho}_j(u)$ are equal to 1. In Fourier analysis a quantity with these properties would generally be referred to as coherency however we will follow the terminology of Sanderson et al. (2010) and refer to it as coherence.

When analyzing the coherence structure of a multivariate signal it may, superficially, appear that two channels are linked as there is significant coherence between them. However, it may in fact be the case that there is not a direct link between them but they are both linked via a third series (see Figure 2). To this end we conclude our modelling framework by introducing the wavelet partial coherence. This provides a measure of the coherence between two channels after removing the effects of all other channels. Partial coherence can again be defined in matrix form using the LWS matrix. The definition of wavelet partial coherence below is analogous to the Fourier domain definition developed in Dahlhaus (2000b).

Definition 2.3.2 *We define the matrix* $\mathbf{G}_j(u) = \mathbf{S}_j(u)^{-1}$ *and the diagonal matrix* $\mathbf{H}_j(u)$ *with elements* $G_j^{(p,p)}(u)^{-(1/2)}$. *The wavelet partial coherence matrix at scale,* j, *and rescaled time,* u, *is defined to be*

$$\boldsymbol{\Gamma}_j(u) = -\mathbf{H}_j(u)\mathbf{G}_j(u)\mathbf{H}_j(u). \tag{7}$$

The off diagonal terms of this matrix are the partial coherences between channels. That is the coherence between the channels after the linear effects of all other channels have been removed.

3 Estimation of the MvLSW Spectral Dependence Quantities

In this section we turn our attention to estimating the spectral quantities of a MvLSW signal. Specifically we first consider the estimation of the LWS matrix before turning to the estimation of the wavelet coherence and partial coherence which were introduced in Section 2.

First, we define the *empirical wavelet coefficient vector*, $\mathbf{d}_{j,k} = [d_{j,k}^{(1)} \ldots, d_{j,k}^{(P)}]'$ whose elements are the empirical wavelet coefficients for each signal channel

$$\mathbf{d}_{j,k} = \sum_{t=0}^{T-1} \mathbf{X}_t \psi_{jk}(t). \tag{8}$$

We use the empirical wavelet coefficient vector to produce the raw *wavelet periodogram matrix*, $\mathbf{I}_{j,k}$:

$$\mathbf{I}_{j,k} = \mathbf{d}_{j,k}\mathbf{d}'_{j,k}. \tag{9}$$

Moreover, we denote $I_{j,k}^{(p,q)}$ to be the (p,q)-th entry of the periodogram matrix where $p, q \in \{1, \ldots, P\}$. The raw wavelet periodogram matrix is the starting point for estimating the LWS matrix. In order to achieve a final estimator with the correct properties we explore the asymptotic properties of the raw periodogram matrix as an estimator for this quantity. In particular, given the results in the one-dimensional setting, it is natural to enquire whether the raw wavelet periodogram is biased.

Proposition 3.0.3 *Let $\{\mathbf{X}_t\}$ be a MvLSW signal with underlying LWS matrix, $\mathbf{S}_j(u)$, and empirical wavelet coefficients, $\{\mathbf{d}_{j,k}\}$. Then*

$$\begin{aligned} E\left[\mathbf{I}_{j,k}\right] &= \sum_{l=1}^{J} A_{jl}\mathbf{S}_l(k/T) + \mathcal{O}(T^{-1}) \quad \textit{and} \\ \mathrm{Var}\left\{I_{j,k}^{(p,q)}\right\} &= \sum_{l=1}^{J} A_{jl} S_l^{(p,p)}(k/T) \sum_{l=1}^{J} A_{jl} S_l^{(q,q)}(k/T) \\ &\quad + \left(\sum_{l=1}^{J} A_{jl} S_l^{(p,q)}(k/T)\right)^2 + \mathcal{O}(2^{2j}/T), \end{aligned}$$

where $A_{jl} =< \Psi_j, \Psi_l >= \sum_\tau \Psi_j(\tau)\Psi_l(\tau)$ for $j, l \in \mathbb{N}$ is the inner product matrix of discrete autocorrelation wavelets (see Nason et al. (2000) or Eckley and Nason (2005) for further details).

Proof: See Appendix 9.

As in the univariate setting of Nason et al. (2000), the above result establishes that the raw wavelet

periodogram matrix is both asymptotically biased and inconsistent. The bias has a particular form consisting of entries in the inner product matrix $\mathbf{A}$. In Cardinali and Nason (2010), the inner product matrix $\mathbf{A}$ is established to be invertible for all Daubechies' compactly supported wavelets. Consequently, the bias of the raw wavelet periodogram matrix estimator in Proposition 3.0.3 can be corrected. However, this would still be an inconsistent estimator. Thus, our proposal is to first apply a smoother on the raw wavelet periodogram matrix and then correct the bias. In particular, we use a rectangular kernel smoother with window of length $2M+1$ to produce the smoothed estimator,

$$\tilde{\mathbf{I}}_{j,k} = \frac{1}{2M+1} \sum_{m=-M}^{M} \mathbf{I}_{j,k+m}. \tag{10}$$

With such an estimator we establish the following result.

Proposition 3.0.4 *Assume that* $\sup_{z\in[0,1]} |\sum_{\tau} c(z,\tau)| \leq \infty$. *Then*

$$\begin{aligned} E\left[\tilde{I}_{j,k}^{(p,q)}\right] =& \sum_{l=1}^{J} A_{jl} S_l^{(p,q)}(k/T) + \mathcal{O}(MT^{-1}) + \mathcal{O}(T^{-1}) \\ \text{Var}\left\{\tilde{I}_{j,k}^{(p,q)}\right\} =& \mathcal{O}(2^{2j}/M) + \mathcal{O}(2^{2j}/T). \end{aligned}$$

Proof: See Appendix 10.

Remark. In the limit, as $T, M \to \infty$, $\text{Var}\left\{\tilde{I}_{j,k}^{(p,q)}\right\} \to 0$. Here, one observes the usual bias-variance trade-off: increasing M reduces the variance but also increases the bias. Moreover, with the additional condition that $M/T \to 0$, then $\left| E\left[\tilde{I}_{j,k}^{(p,q)}\right] - E\left[I_{j,k}^{(p,q)}\right] \right| \to 0$. Thus, one can correct the bias of the smoothed periodogram using the inverse of the inner product matrix $\mathbf{A^{-1}}$. The final smoothed bias-corrected estimator of the LWS matrix is then given by

$$\widehat{\mathbf{S}}_{j,k} = \sum_{l=1}^{J} A_{jl}^{-1} \tilde{\mathbf{I}}_{l,k}. \tag{11}$$

We will use the quantity $\widehat{\mathbf{S}}_{j,k}$ to estimate the wavelet coherence and partial coherence. Denote the (p,q)-th entry of $\widehat{\mathbf{S}}_{j,k}$ to be $\widehat{S}_{j,k}^{(p,q)}$ and let $\widehat{\mathbf{D}}_{j,k;T}$ be a diagonal matrix whose elements are $(\widehat{S}_{j,k}^{(p,p)})^{-(1/2)}$. Then, we define the estimator of the wavelet coherence matrix to be,

$$\widehat{\boldsymbol{\rho}}_{j,k} = \widehat{\mathbf{D}}_{j,k} \widehat{\mathbf{S}}_{j,k} \widehat{\mathbf{D}}_{j,k} \text{ for } j \in \{1,\ldots,J\}, k \in \{0,\ldots,T-1\}. \tag{12}$$

The (p,q)-th element of $\widehat{\boldsymbol{\rho}}_{j,k}$ is the estimated time-varying wavelet coherence between channels p and q at level j. Next, define $\widehat{\mathbf{G}}_{j,k} = (\widehat{\mathbf{S}}_{j,k})^{-1}$ and let $\widehat{\mathbf{H}}_{j,k}$ be a diagonal matrix whose

elements are $(\widehat{G}_{j,k}^{(p,p)})^{-(1/2)}$. Then, the estimator of the wavelet partial coherence matrix is defined to be,

$$\widehat{\mathbf{\Gamma}}_{j,k} = -\widehat{\mathbf{H}}_{j,k}\widehat{\mathbf{G}}_{j,k}\widehat{\mathbf{H}}_{j,k} \text{ for } j \in \{1,\ldots,J\}, k \in \{0,\ldots,T-1\}. \tag{13}$$

Thus, the (p,q)-th element of $\widehat{\mathbf{\Gamma}}_{j,k}$ is the estimated wavelet partial coherence between channels p and q. Note that the linear dependence of channels p and q on all the other channels are removed in the calculation of wavelet partial coherence. Finally we note that using Slutsky's theorem Slutsky (1925) it follows immediately that $\widehat{\boldsymbol{\rho}}_{j,k}$ and $\widehat{\mathbf{\Gamma}}_{j,k}$ are asymptotically unbiased and consistent estimators of the true wavelet coherence matrix and wavelet partial coherence matrix, respectively.

4 Applications of the Multivariate LSW model

To illustrate our proposed multivariate locally stationary wavelet process (MvLSW) we now consider two examples. Section 4.1 considers a simulated example whilst Section 4.2 presents an analysis of multivariate EEG data recorded during a visual-motor experiment.

4.1 Simulated Example

We simulate signals using a tri-variate model of the following form, $\mathbf{X}_t = \mathbf{A}_1\mathbf{X}_{t-1} + \mathbf{A}_2\mathbf{X}_{t-2} + \boldsymbol{\xi}_t$, where $\mathbf{A}_1 = 1.51\mathbf{I}_3$, $\mathbf{A}_2 = -0.83\mathbf{I}_3$ and $\boldsymbol{\xi}_t = [\xi_t^1\ \xi_t^2\ \xi_t^3]' \sim N(0, \mathbf{\Sigma}_t)$. Here $\mathbf{\Sigma}_t$ varies across time so that the cross-correlation structure changes from one time region to another. The channels of the series will therefore have a time-varying coherence structure which is known and constant over frequency. The structure is such that there is a peak in the spectral power at frequency $3\pi/16$ which corresponds to the mid point of wavelet level $j = 3$. We simulated 100 tri-variate signals from this model. Using the method proposed in Section 3 we estimate the coherence and partial coherence matrices for each simulated signal. In the

results reported the Haar wavelet was used in the analysis, although in other simulations we observed that the choice of wavelet made little practical difference for this example. For comparison we also calculate the coherence using both the SLEX method and the method of Ombao and Van Bellegem (OVB) in Ombao and Van Bellegem (2008). For direct comparisons, we have calculated these coherence values for the band of frequencies corresponding to wavelet level $j = 3$.

Figure 3 shows the results of the coherence estimation. In particular we note that of the three estimation methods, the proposed MvLSW coherence estimation scheme produces the most faithful overall estimate of the three. Most notably OVB fails to suitably capture the abrupt change in coherence which occurs within this simulated example. SLEX performs slightly better than OVB in terms of capturing the abrupt changes however it fails to consistently match the peaks and troughs of the coherence. The exception to this is the coherence between channels 1 and 2, where the spectral structure is constant. Here SLEX and OVB have both performed better then our MvLSW method. This is unsurprising given that for this pair the coherence is stationary. This is because OVB can adaptively choose the size of the window so that it matches any changes, if present, on the true spectral quantity. Similarly, the SLEX method chooses the best basis for representing signals and thus can adaptively select the stationary basis if the signal is indeed stationary. The results of partial coherence estimation using the proposed method are shown in Figure 4. We draw particular attention to how the wavelet partial coherence estimator is able to capture quite subtle time-localized changes in partial coherence. Comparison of this approach with SLEX and OVB equivalents for partial coherence is left as an avenue for future research, once such methods have been developed in the literature.

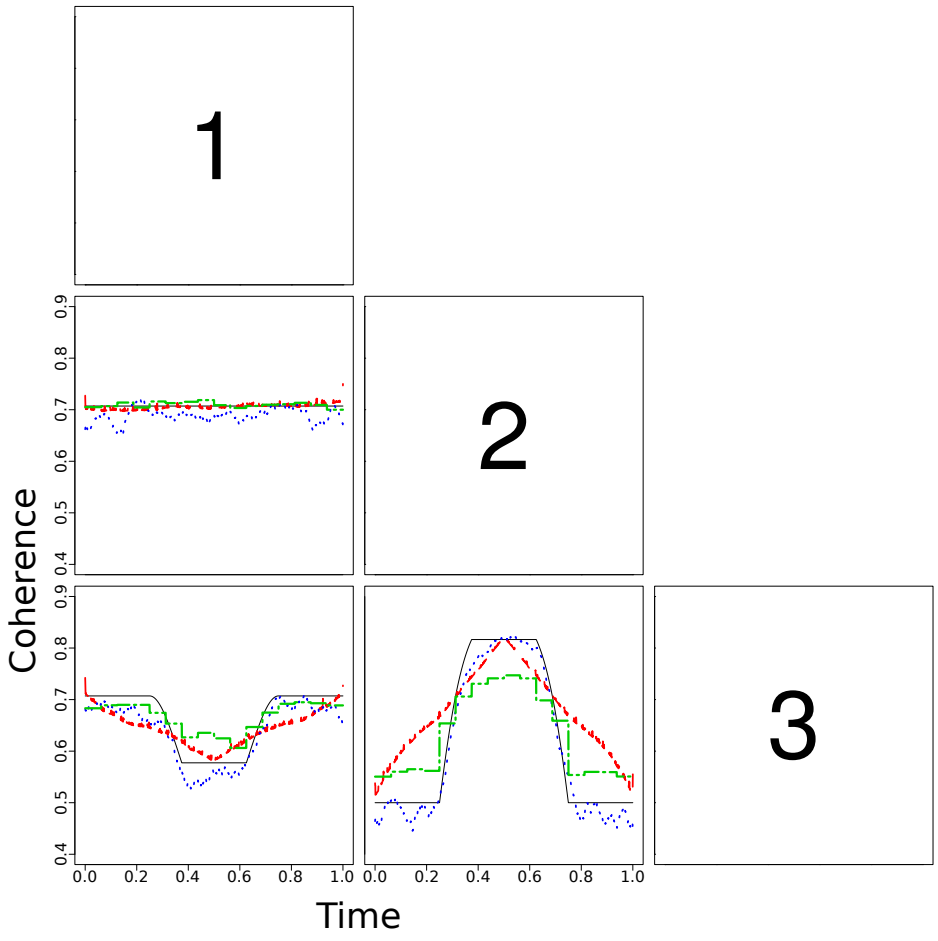


Figure 3: Coherence at level $j = 3$: truth (solid) and mean estimate of the coherence obtained from 100 simulations using MvLSW (dotted); SLEX (dotted and dashed) and OVB (dotted).

4.2 EEG Data

Our real data example is a multi-channel electroencephalogram (EEG) recorded from an experiment in which participants are instructed to move a hand held joystick to either the left or right. A 64-channel EEG was recorded at a sampling rate of 512 Hertz and then bandpass filtered at $(0.02, 100)$ Hertz. Each recording epoch was 1000 milliseconds; the instruction (left vs right) was given at time $t = 0$; and the subject responded with a wrist movement between 350 and 450 milliseconds. Here, we selected data for one participant

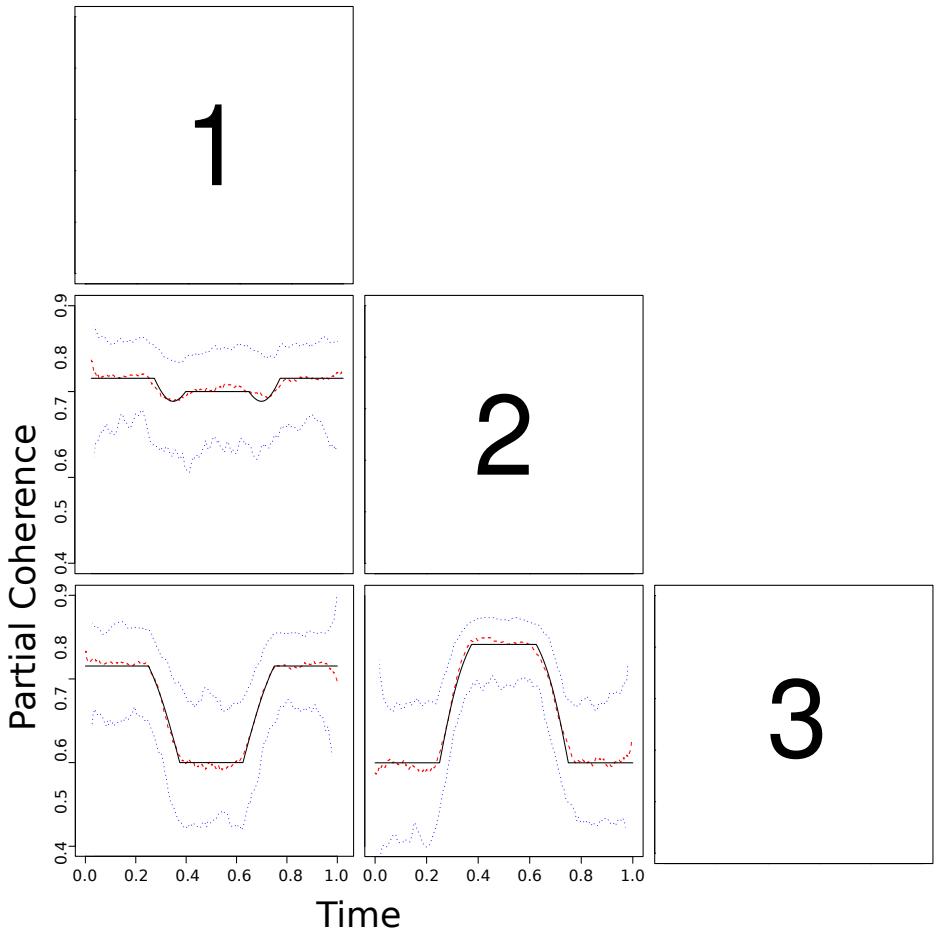


Figure 4: Partial coherence at level $j = 3$. Solid lines represent true values, dashed lines represent the mean of 100 simulations and the dotted lines denote approximate 95% pointwise confidence intervals.

and used 4 channels on the right hemisphere namely FC4 (right fronto-central), FC6 (also right parietal-fronto-central), P4 (right parietal), C4 (right central). This collection is a subset of the channels in Fiecas and Ombao (2011) believed to be engaged in visuo-motor tasks. The positions of these channels are shown in Figure 5. Here, we present an analysis of the wavelet spectral quantities computed for level $j = 2$ ($12.5 - 25$ Hertz), which is contained within the conventional beta band. To study the dynamics within each brain region, we estimated the time-varying and level dependent LWS by kernel smoothing the

wavelet auto- and cross-periodograms using a smoothing span that was objectively selected by generalized cross-validated gamma deviance criterion developed in Ombao et al. (2001). The Daubechies extremal-phase wavelet 10 vanishing moments was used as the analysing wavelet. We found that by using a smoother wavelet we were able to better capture the dynamics of the coherence and partial coherence of this recording.

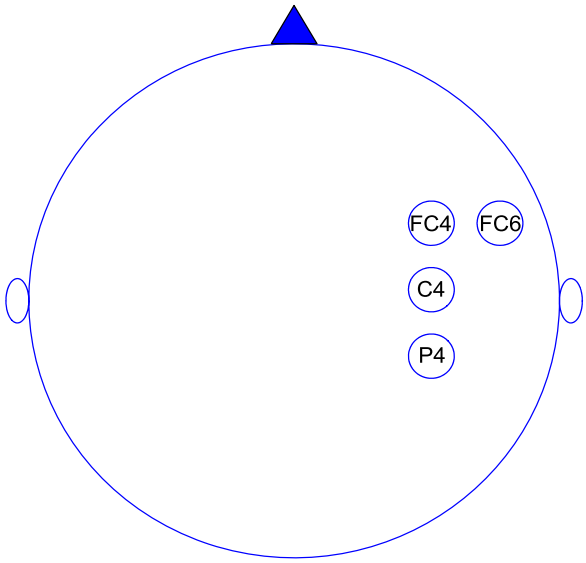
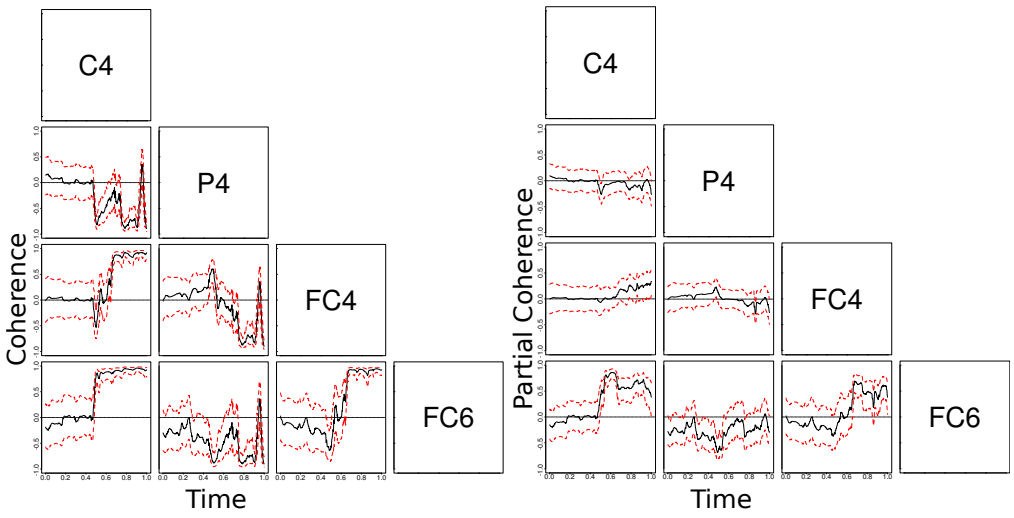


Figure 5: Placement of EEG channels included in analysis.

We investigated the dynamics of cross-dependence within the brain network by estimating the wavelet coherence and wavelet partial coherence. The point estimates of the wavelet coherence and partial coherence were computed using the quantities in the estimated LWS matrix. The approximate 95% pointwise confidence intervals for coherence and partial coherence were obtained by bootstrap resampling the stochastic component of the MvLSW model. Such an approach was used in Ombao et al. (2000) for inference on the evolutionary SLEX spectrum. Empirical distributions of the Fisher-z transformed wavelet coherence and partial coherence values were constructed based on B bootstrap replicates. Typically one might use $B = 1000$ such replicates. Following ideas from Fourier coherence, see for exam-

ple Ombao and Van Bellegem (2008), the wavelet coherence and partial coherence estimates were Fisher-z transformed in order to stabilize the variance of the estimator. The scale-shift specific variance of the empirical distribution of the Fisher-z transformed values were extracted and then utilized to compute the approximate 95% pointwise confidence intervals. For ease of interpretation these confidence intervals were then back-transformed to the scale $(-1, 1)$.



(a) Coherence (b) Partial Coherence

Figure 6: Coherence plot (left) and Partial Coherence plot (right) at level $j = 2$. Solid lines represent the estimated values and dashed the approximate 95% point-wise confidence intervals.

The plots displaying confidence bands on the wavelet coherence (see Figure 6(a)) suggest that, for the most part, brain activity captured by the P4 channel exhibited no linear dependence with brain activity at the central channels namely C4, FC6 and C4. In contrast, there appears to be a common temporal trend in coherence among the central channels. Early in the signals (immediately following visual instruction) there does not seem to be statistically significant connections. However, at about 400 milliseconds (approximately the

time the subject responds to the cue by moving), these central channels become strongly coherent with each other at the beta frequency band. It is interesting to see these brain dynamics during hand movement.

The natural follow-up question is whether or not the links between the central channels established by the coherence plots are *direct* or *indirect* (i.e., due to a connection0 with some common channel). We addressed this question by using the wavelet partial coherence within the framework of our proposed MvLSW model. In Figure 6(b), note that brain activity at FC4 was not directly linked to brain activity at the C4 channel but the link between FC4 and was FC6 was statistically significant beginning at around $t = 400$ milliseconds. Moreover, we observe that there was a statistically significant direct link between FC4 and FC6 – suggesting that the connection between FC4 and C4 observed in the coherence plot was not direct but was in fact related to their common link with the FC6 channel.

The results produced by the proposed MvLSW model are similar to the results from a Fourier-based approach in Fiecas et al. (2010). More importantly, we demonstrate that our proposed model and cross-dependence measure are able to identify an interesting result on the small network of central channels that suggest a direct link between activity at the FC6 channel and each of the FC4 and C4 channels during a visual-motor activity. This finding certainly requires further scientific experiments especially in how these direct connections might be crucial to preserving motor function as well as recovering lost motor function following a major traumatic brain injury. Of course, this analysis is done only on one subject and one will have to develop a more complex model that would take into account brain response variation across many subjects. Nevertheless, the analysis has demonstrated the potential utility and broad impact of the MvLSW model.

5 Concluding Remarks

In conclusion, we developed a rigorous, wavelet-based modeling framework which can capture the evolutionary scale-dependent cross-dependence between components of multivariate signals. An associated estimation theory was also established, demonstrating the uniqueness and asymptotic consistency of our spectral estimators. The particular construction which we proposed also permits the identification of time-scale localized coherence and partial coherence. The proposed wavelet partial coherence measure, in particular, can prove useful when considering the linear dependence between a pair of channels as it enables us to decouple the linear effects of other components of the multivariate signal.

References

Cardinali, A. and Nason, G. (2010). Costationarity of locally stationary time series. *Journal of Time Series Econometrics*, 2(2):1–35.

Cho, H. and Fryzlewicz, P. (2014). Multiple change-point detection for high-dimensional time series via sparsified binary segmentation. Technical report, The London School of Economics and Political Science.

Cohen, E. and Walden, A. (2010). A statistical study of temporally smoothed wavelet coherence. *IEEE Transactions on Signal Processing*, 58(6):2964–2973.

Cohen, E. and Walden, A. (2011). Wavelet coherence for certain nonstationary bivariate processes. *IEEE Transactions on Signal Processing*, 59(6):2522–2531.

Cohen, L. (1989). Time-frequency distributions - a review. *Proceedings of the IEEE*, 77:941–981.

Dahlhaus, R. (2000a). A likelihood approximation for locally stationary processes. *The Annals of Statistics*, 28(6):1762–1794.

Dahlhaus, R. (2000b). Graphical interaction models for multivariate time series1. *Metrika*, 51(2):157–172.

Dahlhaus, R. (2012). Locally stationary processes. *Handbook of Statistics*.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005.

Eckley, I. and Nason, G. (2005). Efficient computation of the discrete autocorrelation wavelet inner product matrix. *Statistics and Computing*, 15(2):83–92.

Fiecas, M. and Ombao, H. C. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics*, 5(2):1102–1125.

Fiecas, M., Ombao, H. C., and Linkletter, C. (2010). Functional connectivity: Shrinkage estimation and randomization test. *NeuroImage*, 49(4):3005–14.

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(1/2):134–139.

Kayhan, A., El-Jaroudi, A., and Chaparro, L. F. (1994). Evolutionary periodogram for nonstationary signals. *IEEE Transactions on Signal Processing*, 42(6):1527–1536.

Koopmans, L. H. (1964). On the multivariate analysis of weakly stationary stochastic processes. *The Annals of Mathematical Statistics*, 35:1765 – 1780.

Kumar, A. and Fuhrmann, D. (1992). A new transform for time-frequency analysis. *IEEE Transactions on Signal Processing*, 40(7):1697–1707.

Nason, G., von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society. Series B*, 62(2):271–292.

Ombao, H., von Sachs, R., and Guo, W. (2000). Estimation and inference for time-varying spectra of locally stationary SLEX processes In Memory of Jonathan A. Raz. *Proceedings of the 2nd International Symposium on Frontiers of Time Series Modeling.*

Ombao, H. C., Raz, J., Strawderman, R., and von Sachs, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*, 88(4):1186–1192.

Ombao, H. C. and Van Bellegem, S. (2008). Evolutionary coherence of nonstationary signals. *IEEE Transactions on Signal Processing*, 56(6):2259–2266.

Ombao, H. C., von Sachs, R., and Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, 100(470):519–531.

Priestley, M. (1988). *Non-linear and non-stationary time series analysis.* Academic Press.

Sanderson, J., Fryzlewicz, P., and Jones, M. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, 97(2):435–446.

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

Walden, A. and Cohen, E. (2012). Statistical Properties for Coherence Estimation From Evolutionary Spectra. *IEEE Transactions on Signal Processing*, 60(9):4586–4597.

6 Proof of Proposition 2.2.3

Suppose, by way of contradiction, that there exist two representations for the same process, $\mathbf{V}_j^{(1)}(u)$ and $\mathbf{V}_j^{(2)}(u)$. At each time point, u, there exists $\mathbf{S}_j^{(1)}(u)$ and $\mathbf{S}_j^{(2)}(u)$ such that,

$$\mathbf{c}(u,\tau) = \sum_{j=1}^{\infty} \mathbf{S}_j^{(1)}(u)\Psi_j(\tau) = \sum_{j=1}^{\infty} \mathbf{S}_j^{(2)}(u)\Psi_j(\tau). \tag{14}$$

Let $\mathbf{\Delta}_j(u)$ be a matrix representing the element-wise difference between the two representations, From equation (14) it is clear that,

$$\sum_{j=1}^{\infty} \mathbf{\Delta}_j(u)\Psi_j(\tau) = \mathbf{0}, \quad \forall u \in (0,1) \text{ and } \tau \in \mathbb{Z}. \tag{15}$$

To establish the uniqueness of the MvLWS representation we must show that (15) implies that, $\mathbf{\Delta}_j(u) = \mathbf{0}\ \forall j > 0,\ u \in (0,1)$. Using arguments similar to those set out by Nason et al. (2000) we use Parseval's relation and the definition of the inner product matrix to obtain, $A_{jl} = \sum_\tau \Psi_j(\tau)\Psi_l(\tau) = \frac{1}{2\pi}\int d\omega \hat{\Psi}_j(\omega)\hat{\Psi}_l(\omega)$, where $\hat{\Psi}_j(\omega) = \left|\hat{\psi}_j(\omega)\right|^2 = 2^j \left|m_1(2^{(j+1)}\omega)\right|^2 \prod_{l=0}^{(j-2)} \left|m_0(2^l\omega)\right|^2$, and $m_0(\omega) = 2^{-1/2}\sum_k h_k \exp(-i\omega k)$, with $\sum_k h_k^2 = 1$, $\frac{1}{\sqrt{2}}\sum_k h_k = 1$ and $|m_1(\omega)|^2 = 1 - |m_0(\omega)|^2$. From equation (15) we can say that for a general element:

$$\sum_l \sum_j \Delta_j^{(p,q)}(u)\Delta_l^{(p,q)}(u) \sum_\tau \Psi_j(\tau)\Psi_l(\tau) = 0$$

Hence it is easily shown that,

$$\int d\omega \left(\sum_j \Delta_j^{(p,q)}(u)\hat{\Psi}_j(\omega) \right)^2 = 0. \tag{16}$$

Since we have already made the assuption that, $\sum_j S_j^{(p,q)} < \infty\ \forall p,q$, we infer that $\sum_j \Delta_j^{(p,q)}(u)\hat{\Psi}_j(\omega)$ is continuous in $\omega \in [-\pi,\pi]$, because every $\hat{\Psi}_j(\omega)$ is and $\sum_j \left|\Delta_j^{(p,q)}(u)\right| < \infty$. Hence (16) implies that, $\sum_{j=1}^{\infty} \mathbf{\Delta}_j(u)\hat{\Psi}_j(\omega) = \mathbf{0}$. The remainder of the proof then follows similarly to Nason et al. (2000).

■

7 Proof of Proposition 2.2.4

Recall the definition of the wavelet representation of a multivariate series in equation (2).

$$\begin{aligned}
\operatorname{cov}\left(X_{uT}^{(p)}, X_{uT+\tau}^{(q)}\right) &= E\left[X_{uT}^{(p)} X_{uT+\tau}^{(q)}\right], \\
&= \mathrm{E}\left[\sum_{j=1}^{\infty}\sum_{k}\sum_{r=1}^{p} V_j^{(p,r)}(k/T)\psi_{j,k}(uT) z_{j,k}^{(r)}\right. \\
&\left.\times \sum_{j'=1}^{\infty}\sum_{k'}\sum_{r'=1}^{q} V_{j'}^{(q,r')}(k'/T)\psi_{j',k'}(uT+\tau) z_{j',k'}^{(r')}\right], \\
&= \sum_{j=1}^{\infty}\sum_{k}\sum_{r=1}^{\min p,q} V_j^{(p,r)}(k/T) V_j^{(q,r)}(k/T)\psi_{jk}(uT)\psi_{jk}(uT+\tau).
\end{aligned}$$

Recalling the definition of the LWS matrix we can say that,

$S_j^{(p,q)}(u) = \sum_{r=1}^{\min p,q} V_j^{(p,r)}(u) V_j^{(q,r)}(u)$. We also make the substitition $m = k - uT$ to obtain,

$$\operatorname{cov}\left(X_{uT}^{(p)}, X_{uT+\tau}^{(q)}\right) = \sum_{j}\sum_{m} S_j\left(\frac{uT+m}{T}\right)\psi_{jm}(0)\psi_{jm}(\tau).$$

Analogous to the approach considered by Nason et al. (2000) in the univariate setting, using the assumed Lipschitz continuous property of $V_j^{(p,q)}(z)$ and therefore $S_j^{(p,q)}(z)$we can consider the difference between this covariance and the function $c^{(p,q)}(u,\tau)$,

$$\begin{aligned}
\left|\operatorname{cov}\left(X_{uT}^{(p)}, X_{uT+\tau}^{(q)}\right) - c^{(p,q)}(u,\tau)\right| &= \left|\sum_{j}\sum_{m} S_j\left(\frac{uT+m}{T}\right)\psi_{jm}(0)\psi_{jm}(\tau) - c^{(p,q)}(u,\tau)\right| \\
&\leq T^{-1}\sum_{m}|m|\, L_j\,|\psi_{jm}(0)\psi_{jm}(\tau)| = \mathcal{O}(T^{-1}).
\end{aligned}$$

■

8 Proof of Proposition 2.2.5

To establish this result we firstly demonstrate that $\mathbf{S}_j^*(u)$ is positive definite. Since $\mathbf{S}_j(u)$ is positive definite, by Choleski, there exists a lower triangular matrix $\mathbf{V}_j(u)$ so that $\mathbf{S}_j(u) =$

$\mathbf{V}_j(u)\mathbf{V}_j'(u)$. Hence $\mathbf{S}_j^*(u) = \mathbf{M}\mathbf{V}_j(u)\mathbf{V}_j'(u)\mathbf{M}' = (\mathbf{M}\mathbf{V}_j(u))(\mathbf{M}\mathbf{V}_j(u))'$. Hence $\mathbf{S}_j^*(u)$ is positive definite. Second, since $\mathbf{S}_j^*(u)$ is positive definite, there exists a lower triangular matrix $\mathbf{V}_j^*(u)$ such that $\mathbf{S}_j^*(u) = \mathbf{V}_j^*(u)\mathbf{V}_j'^*(u)$. Thus $\mathbf{X}_t^*$ admits a MvLSW representation with transfer function $\mathbf{V}_j^*(u)$.

■

9 Proof of Proposition 3.0.3

Expectation: Recall that $d_{j,k}^{(p)} = \sum_t X_t^{(p)}\psi_{j,k}(t)$ and
$X_t^{(p)} = \sum_l \sum_m \sum_r V_l^{(p,r)}(m/T)\psi_{l,m}(t)z_{l,m}^{(r)}$. Hence

$$\begin{aligned} E\left[I_{j,k}^{(p,q)}\right] =& E\left[\left\{\sum_t X_t^{(p)}\psi_{j,k}(t)\right\}\left\{\sum_{t'} X_{t'}^{(q)}\psi_{j,k}(t')\right\}\right], \\ =& \sum_{l=1}^{J}\sum_m \sum_{r=1}^{\min\{p,q\}} V_l^{(p,r)}(m/T)\; V_l^{(q,r)}(m/T) \times \left\{\sum_t \psi_{l,m}(t)\psi_{j,k}(t)\right\}^2. \end{aligned} \tag{17}$$

Substituting $m = n + k$ into (17) we obtain,

$$E\left[I_{j,k}^{(p,q)}\right] = \sum_{l=1}^{J}\sum_n \left\{S_l^{(p,q)}\left(\frac{n+k}{T}\right)\right\}\left\{\sum_t \psi_{l,n+k-t}\psi_{j,k-t}\right\}^2.$$

Analogous to the univariate setting of Nason et al. (2000), since $S_j^{(p,q)}(z)$, is Lipschitz continuous with finite Lipschitz constant L_j, for some fixed n,
$\left|S_j^{(p,q)}((k+n)/T) - S_j^{(p,q)}(k/T)\right| \leq |n|\, L_j/T$, and therefore $S_j^{(p,q)}((n+k)/T) = S_j^{(p,q)}(k/T) + \mathcal{O}(T^{-1})$. Consequently

$$\begin{aligned} E\left[I_{j,k}^{(p,q)}\right] =& \sum_{l=1}^{J} S_l^{(p,q)}\left(\frac{k}{T}\right)\sum_t\sum_v \psi_{j,-t}\psi_{j,-v-t} \\ & \times \sum_n \psi_{l,n-t}\psi_{l,n-v-t} + \mathcal{O}(T^{-1}). \end{aligned} \tag{18}$$

Recalling the definition of the autocorrelation wavelets we find that,

$$E\left[I_{j,k}^{(p,q)}\right] = \sum_{l=1}^{J} S_l^{(p,q)}\left(\frac{k}{T}\right)\sum_v \Psi_l(v)\Psi_j(v) + \mathcal{O}(T^{-1}),$$
$$= \sum_{l=1}^{J} A_{jl} S_l^{(p,q)}\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).$$

■

Variance: To establish the variance of the raw periodogram, we begin by considering $E\left[(I_{j,k}^{(p,q)})^2\right] = E\left[\left(d_{j,k}^{(p)}\right)^2\left(d_{j,k}^{(p)}\right)^2\right]$.

$$E\left[(I_{j,k}^{(p,q)})^2\right] = \left(\sum_{l=1}^{J}\sum_m\sum_{r=1}^{p} V_l^{(p,r)}(m/T)\sum_t \psi_{l,m}(t)\psi_{j,k}(t)\right.$$
$$\left.\times \sum_{l'=1}^{J}\sum_{m'}\sum_{r'=1}^{q} V_{l'}^{(q,r')}(m'/T)\sum_{t'}\psi_{l',m'}(t')\psi_{j,k}(t')\right)^2$$
$$\times E\left[z_{l_1,m_1}^{(r_1)} z_{l_2,m_2}^{(r_2)} z_{l_3,m_3}^{(r_3)} z_{l_4,m_4}^{(r_4)}\right].$$

Using a result due to Isserlis (1918) the above expression can be re-written as the sum of three different elements $E\left[(I_{j,k}^{(p,q)})^2\right] = I_1 + I_2 + I_3$ where, for example,

$$I_1 = \prod_{i=1}^{4}\sum_{t_i,l_i,m_i,r_i} V_{l_i}^{(p_i,r_i)}(m_i/T)\psi_{l_i m_i}(t_i)\psi_{jk}(t_i) E\left[z_{l_1,m_1}^{(r_1)} z_{l_2,m_2}^{(r_2)}\right] E\left[z_{l_3,m_3}^{(r_3)} r_{l_4,m_4}^{(r_4)}\right].$$

Since $E\left[z_{l_1,m_1}^{(r_1)} z_{l_2,m_2}^{(r_2)}\right] = \delta_{l_1 l_2}\delta_{m_1 m_2}\delta_{r_1 r_2}$ this simplifies to:

$$I_1 = \sum_{l_1,m_1,r_1}\left(V_{l_1}^{(p,r_1)}(m_1/T)\right)^2 \times \sum_{t_1=0}^{T-1}\psi_{l_1,m_1}(t_1)\psi_{j,k}(t_1)$$
$$\times \sum_{t_2=0}^{T-1}\psi_{l_1,m_1}(t_2)\psi_{j,k}(t_2)\sum_{l_3,m_3,r_3}\left(V_{l_3}^{(q,r_3)}(m_3/T)\right)^2$$
$$\times \sum_{t_3=0}^{T-1}\psi_{l_3,m_3}(t_3)\psi_{j,k}(t_3)\sum_{t_4=0}^{T-1}\psi_{l_3,m_3}(t_4)\psi_{j,k}(t_4);$$
$$= E\left[I_{j,k}^{(p,p)}\right] E\left[I_{j,k}^{(q,q)}\right].$$

Similarly for I_2 we find that $I_2 = E\left[I_{j,k}^{(p,q)}\right]^2$ and $I_3 = E\left[I_{j,k}^{(p,q)}\right]^2$. Hence,

$$E\left[(I_{j,k}^{(p,q)})^2\right] = E\left[I_{j,k}^{(p,p)}\right] E\left[I_{j,k}^{(q,q)}\right] + 2E\left[I_{j,k}^{(p,q)}\right]^2,$$

and

$$\begin{aligned}\mathrm{Var}\left\{I_{j,k}^{(p,q)}\right\} &= \left(\sum_{l=1}^{J} A_{jl} S_l^{(p,p)}\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1})\right) \\ &\times \left(\sum_{l=1}^{J} A_{jl} S_l^{(q,q)}\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1})\right) \\ &+ \left(\sum_{l=1}^{J} A_{jl} S_l^{(p,q)}\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1})\right)^2.\end{aligned}$$

From Nason et al. (2000) it is known that $\sum_\tau |\Psi_j(\tau)| = \mathcal{O}(2^j)$, and hence $A_{jl} = \sum_\tau \Psi_j(\tau)\Psi_l(\tau) \leq (\sum_\tau |\Psi_j(\tau)|)^2 = \mathcal{O}(2^{2j})$. Hence it is easily verified that,

$$\begin{aligned}\mathrm{Var}\left\{I_{j,kT}^{(p,q)}\right\} =& \sum_{l=1}^{J} A_{jl} S_l^{(p,p)}\left(\frac{k}{T}\right) \sum_{l=1}^{J} A_{jl} S_l^{(q,q)}\left(\frac{k}{T}\right) \\ &+ \left(\sum_{l=1}^{J} A_{jl} S_l^{(p,q)}\left(\frac{k}{T}\right)\right)^2 + \mathcal{O}(2^{2j}/T).\end{aligned}$$

■

10 Proof of Proposition 3.0.4

Recall that the form of the smoothed periodogram is, $\tilde{\mathbf{I}}_{j,k} = (2M+1)^{-1} \sum_{m=-M}^{M} \mathbf{I}_{j,k+m}$.

Expectation:

$$E\left[\tilde{I}_{j,k}^{(p,q)}\right] = \frac{1}{2M+1} \sum_{m=-M}^{M} E\left[I_{j,k+m}^{(p,q)}\right].$$

Where $2M+1$ is the size of the smoothing window. Using the expected value of the periodogram previously calculated this becomes,

$$E\left[\tilde{I}_{j,k}^{(p,q)}\right] = \frac{1}{2M+1} \sum_{m=-M}^{M} \sum_{l=1}^{J} \left\{A_{jl} S_l^{(p,q)}\left(\frac{k+m}{T}\right) + \mathcal{O}(T^{-1})\right\}.$$

Due to the Lipschitz continuity assumed for the spectral components it follows that:

$$E\left[\tilde{I}_{j,k}^{(p,q)}\right] = \sum_{l=1}^{J} A_{jl} S_l^{(p,q)}\left(\frac{k}{T}\right) + \mathcal{O}(MT^{-1}).$$

As $T \to \infty$, $M \to \infty$ but $\frac{M}{T} \to 0$, the smoothed raw wavelet periodogram (auto and cross) is asymptotically biased in the usual way. As such it can be corrected by use of the inverse inner product matrix, A^{-1} to achieve an asymptotically unbiased estimate.

■

Variance: We begin by considering: $E\left[\left(\tilde{I}_{j,k}^{(p,q)}\right)^2\right]$.

$$E\left[\left(\tilde{I}_{j,k}^{(p,q)}\right)^2\right] = \frac{1}{(2M+1)^2} \sum_{m=-M}^{M} \sum_{m'=-M}^{M} E\left[I_{j,k+m}^{(p,q)} I_{j,k+m'}^{(p,q)}\right],$$

by substituting $\tau = m' - m$. Using arguments similar to those employed in the proof of the Expectation, it follows that:

$$\frac{1}{(2M+1)^2} \sum_{m=-M}^{M} \sum_{\tau=M-m}^{M+m} E\left[I_{j,k+m}^{(p,q)} I_{j,k+m+\tau}^{(p,q)}\right] = \frac{1}{(2M+1)^2} \sum_{m,\tau} E\left[d_{j,k+m}^{(p)} d_{j,k+m}^{(q)} d_{j,k+m+\tau}^{(p)} d_{j,k+m+\tau}^{(q)}\right],$$

Using Isserlis' Theorem Isserlis (1918), it can be shown that

$$\begin{aligned}
\mathrm{Var}\left\{\tilde{I}_{j,k}^{(p,q)}\right\} =& \frac{1}{(2M+1)^2} \left\{ \sum_{m,\tau} E\left[d_{j,k+m}^{(p)} d_{j,k+m+\tau}^{(p)}\right] E\left[d_{j,k+m}^{(q)} d_{j,k+m+\tau}^{(q)}\right] \right. \\
&\left. + \sum_{m,\tau} E\left[d_{j,k+m}^{(p)} d_{j,k+m+\tau}^{(q)}\right] E\left[d_{j,k+m}^{(q)} d_{j,k+m+\tau}^{(p)}\right] \right\}, \\
=& \frac{1}{(2M+1)^2} \sum_{m=-M}^{M} \left\{ \sum_{\tau} \sum_{l=1}^{J} S_l^{(p,p)}(k/T) A_{l,j}^{\tau} \right. \\
&\times \sum_{l'=1}^{J} S_{l'}^{(q,q)}(k/T) A_{l',j}^{\tau} + \sum_{\tau} \left(\sum_{l=1}^{J} S_l^{(p,q)}(k/T) A_{l,j}^{\tau} \right)^2 \\
&\left. + \sum_{\tau} (|m|+1)\mathcal{O}(T^{-1}) + \sum_{\tau} (|m|+1)^2 \mathcal{O}(T^{-2}) \right\}.
\end{aligned}$$

where $A_{l,j}^{\tau} = \sum_t \Psi_{l,j}(t)\Psi_{l,j}(t+\tau)$. Note that this is a form of inner product matrix but

with a given lag, τ. Examining the term,

$$\begin{aligned}
&\sum_{\tau}\sum_{l=1}^{J} S_l^{(p,p)}\left(k/T\right) A_{l,j}^{\tau} \sum_{l'=1}^{J} S_{l'}^{(q,q)}\left(k/T\right) A_{l',j}^{\tau} \\
&\qquad \leq \left(\sum_{\tau}\left|\sum_{l=1}^{J} S_l^{(p,p)}\left(k/T\right) A_{l,j}^{\tau}\right|\right)\left(\sum_{\tau}\left|\sum_{l'=1}^{J} S_{l'}^{(q,q)}\left(k/T\right) A_{l',j}^{\tau}\right|\right), \\
&\qquad = \left(\sum_{n}\left|c^{(p,p)}(k,n)\right|\sum_{\tau}\left|\Psi_{l,j}(n+\tau)\right|\right) \\
&\qquad \times \left(\sum_{n}\left|c^{(q,q)}(k,n)\right|\sum_{\tau}\left|\Psi_{l,j}(n+\tau)\right|\right) = \mathcal{O}(2^{2j}).
\end{aligned}$$

Similarly it can be shown that the second term is also equal to $\mathcal{O}(2^{2j})$ hence,

$$\mathrm{Var}\left\{\tilde{I}_{j,k}^{(p,q)}\right\} = \mathcal{O}(2^{2j}/M) + \mathcal{O}(2^{2j}/T). \tag{19}$$

Thus, the smoothed wavelet auto and cross periodogram is asympotically mean-squared consistent as $T \to \infty$, $M \to \infty$, $\frac{M}{T} \to 0$.

■