Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Adaptive Wavelet Domain Principal Component Analysis for Nonstationary Time Series

Marina I. Knight, Matthew A. Nunes & Jessica K. Hargreaves

To cite this article: Marina I. Knight, Matthew A. Nunes & Jessica K. Hargreaves (02 Jan 2024): Adaptive Wavelet Domain Principal Component Analysis for Nonstationary Time Series, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2023.2301069

To link to this article: https://doi.org/10.1080/10618600.2023.2301069

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2024, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/10618600.2023.2301069

Adaptive Wavelet Domain Principal Component Analysis for Nonstationary Time Series

Marina I. Knight[a], Matthew A. Nunes[b], and Jessica K. Hargreaves[a]

[a]Department of Mathematics, University of York, York, UK; [b]School of Mathematical Sciences, University of Bath, Bath, UK

ABSTRACT

High-dimensional multivariate nonstationary time series, that is, data whose second order properties vary over time, are common in many scientific and industrial applications. In this article we propose a novel wavelet domain dimension reduction technique for nonstationary time series. By constructing a time-scale adaptive principal component analysis of the data, our proposed method is able to capture the salient dynamic features of the multivariate time series. We also introduce a new time and scale dependent *cross-coherence* measure to quantify the extent of association between a multivariate nonstationary time series and its proposed wavelet domain principal component representation. Theoretical results establish that our associated estimation scheme enjoys good bias and consistency properties when determining wavelet domain principal components of input data. The proposed method is illustrated using extensive simulations and we demonstrate its applicability on a real-world dataset arising in a neuroscience study. Supplementary materials, with proofs of theoretical results, additional simulations and code, are available online.

ARTICLE HISTORY
Received March 2023
Accepted December 2023

KEYWORDS
Nonstationary time series; Principal component analysis (PCA); Time-dependent eigendecomposition; Wavelet spectrum

1. Introduction

Nonstationary time series, that is, series with time-varying second order structure, are observed in a number of fields, including environmental science (Das and Politis 2020; McGonigle, Killick, and Nunes 2022), finance (Fryzlewicz, Sapatinas, and Subba Rao 2006; Roueff and Von Sachs 2019) and the life sciences (Hargreaves et al. 2019; Wilson et al. 2021). Increasingly, it is of interest to mine information from series which are also high-dimensional in nature, representing objects with complex dependence or dynamics between their components.

In neurosciences, unobtrusive high quality data collection techniques have resulted in large high frequency, multichannel datasets from which important health insights could be gained. The work in this article is in particular motivated by the analysis of high-dimensional electroencephalogram (EEG) data, in which signals from multiple locations on the scalp measure brain activity and cognitive function (see Figure 1). Previous work has shown that such multichannel neurological datasets feature complex multicollinearities (Ombao and Ho 2006; Sanei and Chambers 2013; Park, Eckley, and Ombao 2014) which present analysts with the challenge of extracting features that capture the data dynamic dependencies. Formally, our EEG data setting demonstrably poses *additional challenges* through the presence of *nonstationary behavior* within channels as well as of *complex, time-evolving dependencies* across the multivariate time series.

Analysis of high-dimensional data featuring multicollinearity typically leads to statistical approaches involving dimension reduction via the well-known principal components analysis (PCA) technique (Jolliffe 2002) that seeks to construct uncorrelated linear combinations of the original variables that "best" explain the data variability.

For time series data, it is not optimal to consider traditional principal components, namely, decompositions of an instantaneous covariance matrix. Instead, it is more appropriate to use eigendecompositions of spectral density matrices, which account for the time-dependent structure present in the data (Brillinger 1975). The theoretical properties of spectral domain PCA are well studied and understood under the assumption of first- and second-order stationarity (Priestley 1983; Brockwell and Davis 1991; Shumway and Stoffer 2010), but even mild departures from this stationarity assumption can result in nonoptimal features and unsatisfactory dimension reduction, see for example, Lansangan and Barrios (2009). To handle nonstationarity, a time-dependent analogue of the spectral (frequency) domain PCA for stationary processes was developed in Ombao and Ho (2006) under the locally stationary Fourier processes framework. As a complementary modeling approach to Fourier-based analysis, wavelet techniques have enjoyed popularity in the statistical literature due to their ability to efficiently represent signals via time-scale decompositions, see for example, Nason (2008). However, despite the (wavelet) time–scale spectral signature being arguably more natural for nonstationary time series due to accurate time-localization of frequency (scale) components, to the best of our knowledge until the very recent work of Lim, Kwon, and Oh (2021) there was no available exploratory tool to perform wavelet spectral domain

CONTACT Matthew A. Nunes m.a.nunes@bath.ac.uk School of Mathematical Sciences, University of Bath, Bath, UK.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.

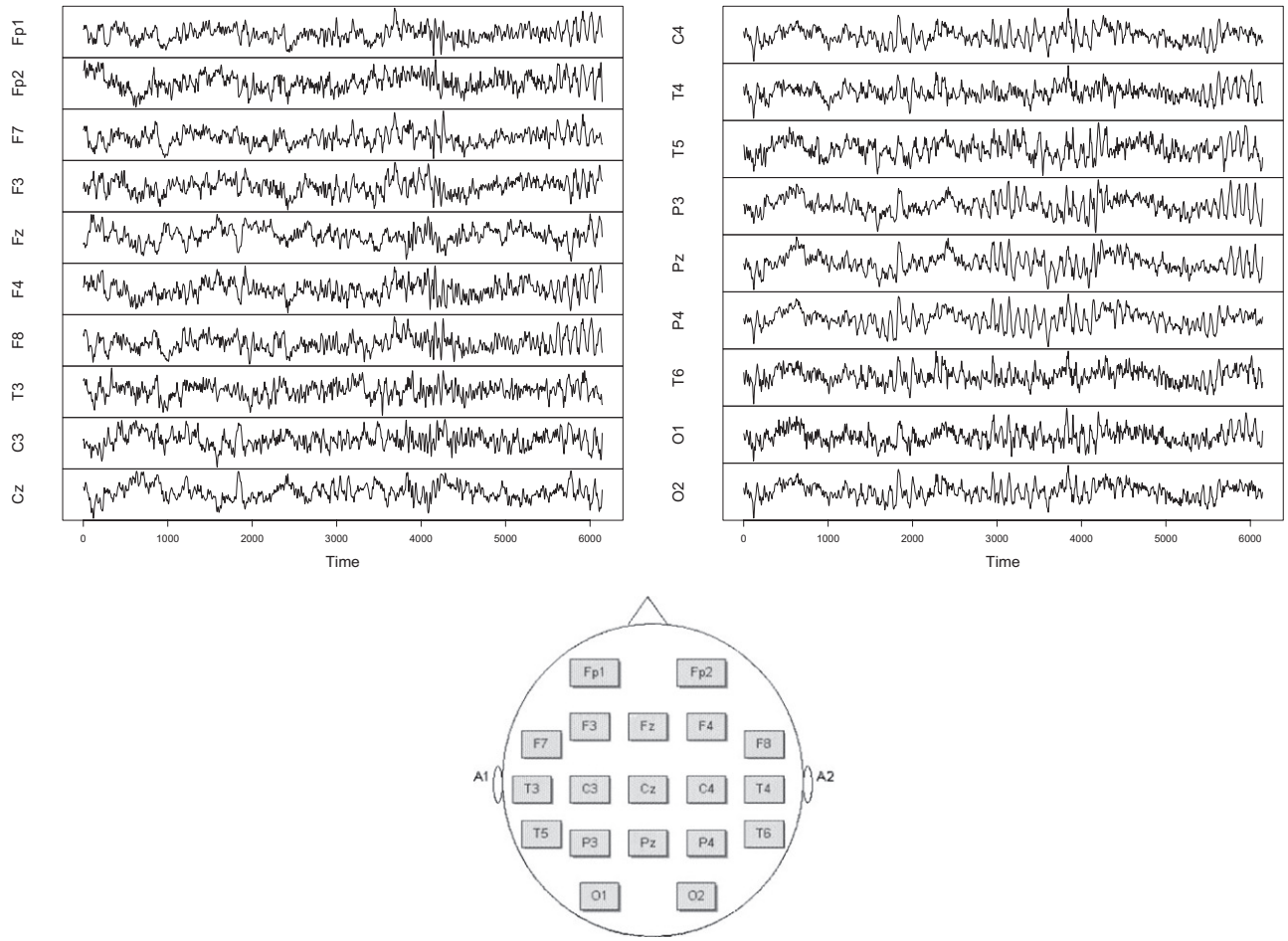

Figure 1. Multiple nonstationary time series, referred to as "channels," from our motivating EEG dataset. Vertical axis labels indicate brain sensor locations, depicted on the illustration in the bottom panel. The series exhibit visible nonstationarity (confirmed by testing with Nason 2013), with time-dependent within- and cross-channel frequency information.

dimension reduction for multivariate nonstationary time series data.

Motivated by a scarcity of appropriate tools in this context and by a limited current development of their associated theoretical and practical properties, we introduce a new wavelet domain principal component analysis tool. Our work adopts a formal, model-based approach for representing multivariate nonstationary series using the multivariate locally stationary wavelet process (MvLSW) framework of Park, Eckley, and Ombao (2014), which is capable of capturing a range of nonstationary dynamics. We achieve dimension reduction by means of an adaptive principal component analysis carried out on the process time–scale signature, as described by its wavelet decomposition under the MvLSW model. The primary aim of this analysis is to extract the salient spectral features that characterize the data, in a way that explicitly acknowledges the time-dependent (second order) statistical properties of the process, for example, within- and cross-channel dependence. Our proposed analysis departs from previous techniques that target decompositions of the associated Fourier (Ombao and Ho 2006) or wavelet spectrum (Lim, Kwon, and Oh 2021), and instead focuses on *maximizing the directions of process variation encompassed by its wavelet decomposition*. Notably, our proposed framework identifies *mathematically different and interpretable directions* of variation to those proposed by Lim, Kwon, and Oh (2021) and allows us to identify the specific scales and times at which original channels are responsible for most *(wavelet domain) power* and to characterize their effects.

Our proposed approach has the following advantages: (i) it constructs substantially better behaved estimators (smaller bias and faster decay rates) than its counterpart in Lim, Kwon, and Oh (2021) that directly targets the wavelet spectrum, (ii) its attractive practical properties are backed-up by theory, and (iii) it facilitates an intuitive interpretation of the proposed adaptive rotation. In this context, our work additionally (iv) introduces a new wavelet domain *cross-coherence measure* for pairs of rotated and original channels, thus, allowing practitioners to formally and meaningfully quantify the contributions between the original time series and its principal component representation.

This article is organized as follows. Section 2 introduces our adaptive wavelet domain dimension reduction technique, `mvLSW-PCA`, including algorithmic details (Sections 2.1–2.3) and the associated estimation theory for the wavelet domain eigenvalue and eigenvector quantities (Section 2.4). We demonstrate the efficacy of our estimation procedure by extensive simulations in Section 2.5. Subsequently, Section 3 introduces a new wavelet domain measure of association for our proposed principal component construction, together with a theoretically

justified estimation method. Section 4 illustrates our proposed `mvLSW-PCA` data analysis tool for the motivating neuroscience dataset introduced above in Section 1. Our concluding remarks are in Section 5.

2. Adaptive Wavelet Domain PCA for MvLSW Signals

In this section we introduce our new wavelet domain methodology to extract pertinent features and reduce the dimension of a multivariate nonstationary time series. In what follows, we consider a P-dimensional vector $\mathbf{X}_{t,T} = \left(X^{(1)}_{t,T}, X^{(2)}_{t,T}, \ldots, X^{(P)}_{t,T}\right)^T$, each element of which is an individual channel recorded at T times, $t = 0, \ldots, T-1$. The proofs of the results in this section can be found in Appendix A in the supplementary material.

2.1. Modeling Framework

Park, Eckley, and Ombao (2014) propose a multivariate locally stationary wavelet (MvLSW) model for the process $\{\mathbf{X}_{t,T}\}_{t=0}^{T-1}$, $T = 2^J \geq 1$ that allows for nonstationary behavior within- and across- channels. Locally stationary wavelet approaches have found a wealth of applications, with for example, Sanderson, Fryzlewicz, and Jones (2010), Park, Eckley, and Ombao (2014), and Embleton, Knight, and Ombao (2022), all showcasing the usefulness of wavelet-based methods for neurological data. The model construction, in the vein of Nason, von Sachs, and Kroisandt (2000), uses families of discrete nondecimated wavelets localized at scales j and times k, combined with amplitudes specified by scale-dependent functions of time, as follows

$$\mathbf{X}_{t,T} = \sum_{j=1}^{J} \sum_{k \in \mathbb{Z}} \mathbf{V}_j(k/T)\psi_{j,k}(t)\mathbf{z}_{j,k}. \tag{1}$$

In the above, $\{\psi_{j,k}(t) = \psi_{j,k-t}\}_{j,k}$ is a set of discrete nondecimated wavelets (see Nason, von Sachs, and Kroisandt (2000) for their description); $\mathbf{V}_j(k/T)$ is a $P \times P$-dimensional amplitude matrix (also known as transfer function), defined to have lower-triangular form and which in effect determines the inter-channel dependencies; the vectors $\{\mathbf{z}_{j,k}\}_{j,k}$ with $\mathbf{z}_{j,k} = \left(z^{(1)}_{j,k}, \ldots, z^{(P)}_{j,k}\right)^T$ are uncorrelated random vectors (across channels, scales and times), also known as process innovations, with mean vector $\mathbf{0}$ and variance-covariance matrix equal to the $P \times P$ identity matrix, for each scale $j = 1, 2, \ldots$ (finest scale, second finest and so on). Each element, say (m, q), of the transfer function matrix is assumed to behave as a Lipschitz continuous function, with Lipschitz constants $\{L_j^{(m,q)}\}_j$ satisfying $\sum_{j=1}^{\infty} 2^j L_j^{(m,q)} < \infty$ for any m, $q = 1, \ldots, P$ channels.

The *localized $(P \times P)$ wavelet spectral matrix* at scale $j = 1, \ldots, J$ and rescaled time $u = k/T \in (0, 1)$ describes the process power time–scale decomposition and is defined as

$$\mathbf{S}_j(u) = \mathbf{V}_j(u)\mathbf{V}_j^T(u), \tag{2}$$

where $\mathbf{V}_j^T(u)$ denotes the matrix transpose of $\mathbf{V}_j(u)$. The diagonal elements of the spectral matrix are the individual channel spectra, denoted in what follows by $\mathbf{S}_j^{(m,m)}(u)$, while the off-diagonal terms are the channel cross-spectra denoted by $\mathbf{S}_j^{(m,q)}(u)$, for any channels $m, q = 1, \ldots, P$. The multivariate spectrum provides a scale- and time-dependent decomposition for each individual channel variance, as well as for each channel cross-covariance (see Park, Eckley, and Ombao (2014) for details). In what follows, we will occasionally include a superscript, for example, $\boldsymbol{S}_j^{(X)}(u)$, to explicitly identify the corresponding multivariate process.

Key to our development, we shall focus on the transformed spectral quantity

$$\boldsymbol{\beta}_j^{(X)}(u) = \sum_{\ell=1}^{J} A_{\ell,j}\mathbf{S}_\ell^{(X)}(u), \tag{3}$$

where A is defined as a $J \times J$ symmetric and positive definite matrix with entries $A_{\ell,j} =< \Psi_\ell, \Psi_j >= \sum_n \Psi_\ell(n)\Psi_j(n)$, where $<\cdot,\cdot>$ denotes the ℓ^2-inner product, and $\Psi_j(n) = \sum_{k\in\mathbb{Z}} \psi_{j,k}(0)\psi_{j,k}(n)$ are the discrete autocorrelation wavelets (Nason, von Sachs, and Kroisandt 2000). As the matrix A is invertible (Nason, von Sachs, and Kroisandt 2000), we can equivalently retrieve the $\boldsymbol{S}$-spectrum from the $\boldsymbol{\beta}$-spectrum

$$\mathbf{S}_\ell^{(X)}(u) = \sum_{j=1}^{J} (A^{-1})_{\ell,j}\,\boldsymbol{\beta}_j^{(X)}(u). \tag{4}$$

A central quantity for wavelet-based spectral estimation, comprehensively treated in Section 2.4, is the raw wavelet periodogram matrix, denoted by $\mathbf{I}_{j,k;T}$ and defined as

$$\mathbf{I}_{j,k;T}^{(X)} = \mathbf{d}_{j,k;T}^{(X)} \left(\mathbf{d}_{j,k;T}^{(X)}\right)^T, \tag{5}$$

where $\mathbf{d}_{j,k;T}^{(X)} = \left(d^{(1)}_{j,k;T}, \ldots, d^{(P)}_{j,k;T}\right)^T$ is the wavelet coefficient vector obtained as $\mathbf{d}_{j,k;T}^{(X)} = \sum_{t=0}^{T-1} \mathbf{X}_{t,T}\psi_{j,k}(t)$. For ease of notation, where appropriate, in the remainder of this article we drop the explicit dependence on T and X for the wavelet coefficients and the periodogram. Note that for each channel pair (m, q), the raw cross-periodogram is simply given by $I_{j,k;T}^{(m,q)} = d_{j,k;T}^{(m)} d_{j,k;T}^{(q)}$.

We first establish the following properties of the $\boldsymbol{\beta}$-spectrum (with proofs in Appendix A.1–A.2) which will prove central to the mathematical justification of our construction.

Property 1. The $\boldsymbol{\beta}^{(X)}$-spectrum defined in (3) has the following properties.

Symmetry: $\boldsymbol{\beta}_j^{(X)}(u)$ is a symmetrical matrix for all scales j and rescaled times u.

Positive definiteness: $\boldsymbol{\beta}_j^{(X)}(u)$ is (at least) positive semidefinite for all scales j and rescaled times u.

Lipschitz continuity: Each entry (m, q) of the $\boldsymbol{\beta}_j^{(X)}(u)$ matrix is a Lipschitz continuous function of rescaled time (u) for all scales j and channels m and q, with the Lipschitz constants given by

$$\tilde{L}_j^{(m,q)} = \sum_{\ell=1}^{J} A_{\ell,j} L_\ell^{(m,q)}.$$

Property 1 ensures that our theoretical transformed spectral quantity admits an eigendecomposition with real-valued (positive) eigenvalues.

Property 2. The eigenvalues $\left(\lambda_j^{(i)}(u)\right)_{i=1}^{P}$ of the $\boldsymbol{\beta}^{(X)}$-spectrum are Lipschitz continuous functions of time (u) for any scale j.

The practical use of the $\boldsymbol{\beta}$-spectrum stems from the property of the raw wavelet periodogram matrix, $\boldsymbol{I}_{j,k}^{(X)}$, to be an asymptotically unbiased and consistent estimator of $\boldsymbol{\beta}_j^{(X)}(u)$ with $u = k/T$ (Park, Eckley, and Ombao 2014), rather than of the $\mathbf{S}_j^{(X)}(u)$ as for example, the Fourier case. The key to its successful use across a range of modeling and testing settings (Hargreaves et al. 2019; Embleton, Knight, and Ombao 2022) lies in its simpler and more accurate estimation, specifically the estimator convergence rate deteriorates from an order term T^{-1} associated to the $\boldsymbol{\beta}$-estimation to an order term $T^{\alpha-1}$ with $\alpha \in (0,1)$ for the wavelet spectrum $\mathbf{S}_j^{(X)}(u)$ (see Lemma A1 and Proposition 3 in Sanderson, Fryzlewicz, and Jones 2010). In the context of this work, an additional attractive practical relevance for using the $\boldsymbol{\beta}$-spectral structure and its estimator stems from realizing that

$$\text{var}(\mathbf{d}_{j,k;T}^{(X)}) = \mathbb{E}(\boldsymbol{I}_{j,k}^{(X)}) \approx \boldsymbol{\beta}_j^{(X)}(u), \tag{6}$$

where $\approx$ denotes equality up to $\mathcal{O}(T^{-1})$, and since $\mathbb{E}(\mathbf{d}_{j,k;T}^{(X)}) = 0$ as inherited from the zero-mean of $\{\boldsymbol{X}_{t,T}\}$. This expression motivates our proposal to identify the principal directions of process variation as encapsulated by the *variance of the wavelet coefficients*.

2.2. Proposed Locally Stationary Wavelet PCA

Motivated by our neurosciences data and the discussion of PCA for nonstationary time series data in Section 1, in what follows we treat the exploratory analysis of $\{\boldsymbol{X}_{t,T}\}_{t=0}^{T-1}$ and its dimension reduction entirely in the spectral domain. In line with previous work on spectral (frequency) domain PCA in the Fourier stationary and locally stationary contexts (Brillinger 1975; Ombao and Ho 2006), we aim to replace the original process by a new process scale- and time-dependent process whose spectral content is maximized at each scale and time pair and whose components are uncorrelated in the wavelet domain.

In the light of (6) and of Property 1, it is justified to propose to obtain the principal directions of process variation by performing an eigendecomposition of the $\boldsymbol{\beta}$-spectrum corresponding to the original process. This is equivalent to solving at each particular scale–(rescaled) time pair (j,u) the constrained quadratic optimization problem

$$\max_{\mathbf{c}} \left(\mathbf{c}^T \boldsymbol{\beta}_j^{(X)}(u)\mathbf{c}\right), \tag{7}$$

subject to the P-dimensional (eigen)vector $\mathbf{c}$ having ℓ_2-norm equal to 1. The well-known solution to problem (7) is given by the eigenvalue-eigenvector pairs of the spectral matrix $\boldsymbol{\beta}_j^{(X)}(u)$. We denote this set by $\left\{\left(\lambda_j^{(i)}(u), \mathbf{e}_j^{(i)}(u)\right)_{i=1}^{P}\right\}$, with the (positive) eigenvalues arranged in descending order from direction $i = 1$ to P for each (j,u) pair, and corresponding ith eigenvector $\mathbf{e}_j^{(i)}(u) = \left(e_{j,1}^{(i)}(u), \ldots, e_{j,P}^{(i)}(u)\right)^T$. Note the resulting eigenvectors are scale (j)- and (rescaled) time (u)- dependent quantities and connect the channel activation timing between the original and new signal. Their accurate estimation is thus paramount for performing meaningful inference on the basis of the spectral PCA.

By design, the original $\boldsymbol{\beta}_j^{(X)}(u)$ spectral structure is henceforth replaced by a much simplified channel dependence structure that leads to the identification of the strongest and most relevant modes of wavelet variation across channels, orthogonal to one another to ensure each captures nonredundant information

$$\boldsymbol{\beta}_j^{(PCA)}(u) = \text{diag}\left(\lambda_j^{(i)}(u)\right)_{i=1}^{P}, \tag{8}$$

where we denoted by "diag" a diagonal matrix, here of dimension $P \times P$, whose entries are the (j,u)-localized eigenvalues taken in descending order. Property 2 ensures that the eigenvalue structure mirrors the $\boldsymbol{\beta}$-spectrum properties assessed in Property 1.

We exploit the general connection (4) and propose the "corrected" eigenvalues

$$\boldsymbol{S}_j^{(PCA)}(u) = \text{diag}\left(\sum_{\ell=1}^{J} A_{j,\ell}^{-1}\lambda_\ell^{(i)}(u)\right)_{i=1}^{P}, \quad \forall\,(j,u). \tag{9}$$

Our proposal versus previous spectral PCA work. Crucially, our proposal departs from the previous Fourier and wavelet spectral domain eigendecompositions (Ombao and Ho 2006; Lim, Kwon, and Oh 2021) which target the original Fourier, respectively wavelet spectrum ($\boldsymbol{S}^{(X)}$), and we instead consider the related spectral quantity $\boldsymbol{\beta}^{(X)}$. In their work, Lim, Kwon, and Oh (2021) observe the superiority of their method when compared to Ombao and Ho (2006), particularly when identifying time-localized spectral activity, a feature that is also manifest when compared to our method. When mathematically comparing our proposal to that in the Lim, Kwon, and Oh (2021) construction, we note that while also yielding a diagonal spectral structure, by contrast theirs involves the eigenvalues of $\boldsymbol{S}_j^{(X)}(u)$ on the main diagonal, for each (j,u). Importantly, the resulting eigendecompositions of the two constructions are fundamentally different: since at scale j and rescaled time $u = k/T$ the quantity $\boldsymbol{\beta}_j^{(X)}(u)$ can be seen as the (asymptotic) variance of the wavelet coefficients, $\text{var}\left(\mathbf{d}_{j,k;T}^{(X)}\right)$, our problem can be reformulated as seeking scale- and time- dependent rotations of the process wavelet coefficients that maximize their variance; however, the Lim, Kwon, and Oh (2021) approach does not lend itself to be interpreted as a variance-based maximization problem. By centering our approach on the $\boldsymbol{\beta}$-spectrum, in effect we create linear combinations of the process wavelet coefficients that yield an intuitive interpretation of the proposed time- and scale-dependent rotation, nonexistent in the $\boldsymbol{S}$-spectrum approach of Lim, Kwon, and Oh (2021). We establish the connections between the two representations in Appendix A.3 for eigendecompositions, and in Appendix A.5 between the original and new PCA wavelet domains for our method. It is important to point out that mathematically the two decompositions on the true spectral quantities cannot be viewed as equivalent. Furthermore, we will see over the next sections that the inferior estimator convergence rates associated to the $\boldsymbol{S}$-spectrum as opposed to those of the $\boldsymbol{\beta}$-spectrum have serious consequent

effects for the corresponding eigenvector estimators, rendering the $\boldsymbol{S}$-based ones effectively non-interpretable.

2.3. Proposed `mvLSW-PCA` Data Analysis Tool

As all spectral quantities in the development above are unknown, we obtain the *sample* principal component series by replacing in the previous discussion the unknown spectral quantity $\boldsymbol{\beta}_j^{(X)}(\cdot)$ with the smoothed periodogram in (10) (Section 2.4). This is a well-behaved asymptotically unbiased and consistent estimator for the $\boldsymbol{\beta}$-spectrum, which we will denote by $\widehat{\boldsymbol{\beta}}_j^{(X)}(\cdot)$. Thus, the *estimators* of the time-varying eigenvalues and eigenvectors are simply the eigenvalues and eigenvectors of the time-varying spectral matrix estimator $\widehat{\boldsymbol{\beta}}_j^{(X)}(\cdot)$. The next section precisely defines these estimators and assesses their desirable properties.

While the dimension of the PCA domain quantities is still P, due to the uncorrelated nature of its channels, often only $Q < P$ channels are retained and used in practical tasks. We begin by assuming that Q is known, although in practice it will often have to be determined from the data.

Algorithm 1 summarizes the proposed wavelet domain PCA, whose output can be further used in order to investigate the identified important channels through for example, estimating their time-varying spectral activity and coherence. In what follows, we introduce and discuss each component of the proposed algorithm.

Algorithm 1 Proposed `mvLSW-PCA` algorithm for nonstationary time series.

Assume the observed data is a realization of a P-variate MvLSW process denoted as $\{\mathbf{X}_{t,T} = (X_{t,T}^{(1)}, \ldots, X_{t,T}^{(P)})\}, t = 0, \ldots, T-1$ where $T = 2^J$.

1. *Spectral estimation:* estimate the spectral content of the multivariate signal by using a model-based MvLSW smoothed (but not corrected) periodogram as defined in equation (10). Denote this estimate by $\{\widehat{\boldsymbol{\beta}}_j^{(X)}(u) = \tilde{\boldsymbol{I}}_{j,[uT]}\}_{j,k}$, for levels $j = 1, \ldots, J$ and locations $k = [uT] = 0, \ldots, T-1$.
2. *Eigendecomposition:* compute the eigenvalues and corresponding (unit norm) eigenvectors of $\widehat{\boldsymbol{\beta}}_j^{(X)}(u)$ for each level $j = 1, \ldots, J$ and location $k = [uT] = 0, \ldots, T-1$. Denote them in decreasing magnitude by $\{\left(\tilde{\lambda}_j^{(i)}(u), \tilde{\mathbf{e}}_j^{(i)}(u)\right)_{i=1}^{P}\}$ with $\tilde{\lambda}_j^{(1)}(u)$ having the largest magnitude. *The notation here is to be understood as discrete, that is* $\tilde{\lambda}_j^{(i)}(u) = \tilde{\lambda}_{j,k}^{(i)}$ *and similarly for the eigenvectors.*
3. *Dimension reduction:* if no a priori information on the number of components $Q \leq P$ is available, determine Q using empirical methods, for example using a screenplot of percentage spectral variation explained, $\tilde{\lambda}_j^{(i)}(u)/\mathrm{Tr}(\tilde{\boldsymbol{I}}_{j,[uT]})$.
4. *Time-dependent analysis:*
   - Examine the (possibly) reduced-dimension $(Q \times Q)$ and corrected eigenvalues $\widehat{\boldsymbol{S}}_\ell^{(PCA)}(u) = \mathrm{diag}\left(\sum_{j=1}^{J} A_{\ell,j}^{-1} \tilde{\lambda}_j^{(i)}(u)\right)_{i=1}^{Q}$, corresponding to the Q principal components retained, so that we may determine various quantities of interest, for example levels (say, $j^\star$) that account for most variation, time locations when certain features activate etc.
   - Examine the cross-coherence magnitude at the levels determined above, $\left(\widehat{\rho}_{j^\star}^{(i,p)}(u)\right)_{i=1}^{Q}$ as defined in (17) or (18), for inference on the activation timing (u) of the original channels (p) responsible for most spectral variation and interpretation of the principal components.

2.4. Estimation Theory for the Proposed Wavelet Domain PCA

Let us now address the theoretical properties for the proposed estimators of the wavelet $\boldsymbol{\beta}^{(X)}$-spectrum eigendecomposition outlined in Section 2.3. The theoretical results in this section are derived under the assumption of Gaussian innovations in (1) and the proofs can be found in Appendix A.4 of the supplementary material.

The raw wavelet periodogram matrix in (5) is an asymptotically biased and inconsistent estimator of the true spectrum, hence, Park, Eckley, and Ombao (2014) recommend first applying a smoother on the raw wavelet periodogram matrix and then correcting for its bias. In particular, in the spirit of Sanderson, Fryzlewicz, and Jones (2010), the authors recommend smoothing by means of a rectangular kernel smoother with window of length $(2M + 1)$ thus yielding a smoothed version of the raw periodogram

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

for a chosen M such that as $T \to \infty$, we have $M \to \infty$ and $M/T \to 0$.

Although not used in our proposal but useful for aiding the comparison with the Lim, Kwon, and Oh (2021) method, we recall that a consistent estimator for $\mathbf{S}_j^{(X)}(u)$ is obtained by bias-correcting the smoothed periodogram above, namely

$$\hat{\mathbf{S}}_j^{(X)}(u) = \sum_{\ell=1}^{J} A_{j,\ell}^{-1} \tilde{\mathbf{I}}_{\ell,[uT]}, \tag{11}$$

whose (m,q)th entry we denote $\hat{\mathbf{S}}_j^{(m,q)}(u)$ and where $A_{j,\ell}^{-1}$ is the (j,ℓ) entry of A^{-1}.

For our proposed algorithm in Section 2.3, for each scale-time pair (j,u), the unknown spectral quantity $\boldsymbol{\beta}_j^{(X)}(u)$ is replaced by the smoothed periodogram in (10); recall that to avoid notational clutter, we omit the superscript (X) for the periodogram and wavelet coefficients associated to the original process $\{\boldsymbol{X}_{t,T}\}$ when no confusion can arise.

Assuming that $\sup_{u\in(0,1)} \left|\sum_{\tau\in\mathbb{Z}} c^{(m,q)}(u,\tau)\right| < \infty$ for any channels (m,q), $\tilde{\boldsymbol{I}}_{j,[uT]}$ is an asymptotically unbiased and consistent estimator for $\boldsymbol{\beta}_j^{(X)}(u)$ (Park, Eckley, and Ombao 2014), namely, when retaining the highest order terms

$$\begin{aligned} \mathbb{E}(\tilde{\boldsymbol{I}}_{j,[uT]}) &= \boldsymbol{\beta}_j^{(X)}(u) + \mathcal{O}(MT^{-1}), \\ \mathrm{var}(\tilde{I}_{j,[uT]}^{(m,q)}) &= \mathcal{O}(2^{2j}M^{-1}), \quad \forall (j,u). \end{aligned} \tag{12}$$

Additionally, it is straightforward, but essential, to note that the smoothed periodogram above can be easily shown to be symmetric and positive definite; on a practical note, the use of (10) is crucial in order to remove noisy artifacts and ensure real-valued eigenvalues.

On estimating the eigendecomposition of $\boldsymbol{\beta}$ versus that of $\mathbf{S}$. Additional to the natural connection to classical PCA argued so far, there is an important mathematical argument in choosing to work with eigendecompositions of the estimator $\tilde{\boldsymbol{I}}_{j,[uT]}$ in (10) for $\boldsymbol{\beta}_j^{(X)}(u)$, rather than with eigendecompositions of its corrected counterpart $\hat{\mathbf{S}}_j^{(X)}(u)$ in (11) which estimates the original wavelet spectrum $\mathbf{S}_j^{(X)}(u)$. Closer inspection of the asymptotic behavior of these estimators reveals that while both are asymptotically unbiased and consistent (Park, Eckley, and Ombao 2014) due to the bias correction in (11), the convergence rates are much slower for $\hat{\mathbf{S}}_j^{(X)}(u)$, see for example the proof of Proposition 3, Sanderson, Fryzlewicz, and Jones (2010) and Theorem 2, Fryzlewicz and Nason (2006). In the context of locally stationary spectral estimation this is not problematic; however when the estimated spectral quantities are subsequently manipulated via an eigendecomposition, the rates of convergence of the corresponding estimated eigenvalues and eigenvectors to their true counterparts deteriorate further. Consequently, although not derived here, the asymptotic convergence rates of the eigendecompositions of $\hat{\mathbf{S}}_j^{(X)}(u)$ are markedly inferior to those of $\tilde{\boldsymbol{I}}_{j,[uT]}$. This point will be strikingly salient in practice, through both our simulations in Section 2.5.2 and real data analysis in Section 4 which show that the $\hat{\mathbf{S}}$-based eigenvectors proposed by Lim, Kwon, and Oh (2021) hardly exhibit any discernible structure. The failure to capture the crucial cross-process time-activation periods in turn renders these eigenvectors unusable for important tasks such as the meaningful interpretation of the principal components.

To aid the asymptotic derivations for our proposed estimated eigencomponents, we first show that the smoothed cross-periodograms are asymptotically uncorrelated, as follows.

Proposition 1. Under the assumption that $\sup_{u\in(0,1)} \left|\sum_{\tau\in\mathbb{Z}} c^{(m,q)}(u,\tau)\right| < \infty$ for any channels (m,q), we have for any channel pairs (m,q) and (m',q'),

$$\begin{aligned}&\operatorname{cov}\left(\tilde{I}_{j,[uT]}^{(m,q)}, \tilde{I}_{j,[uT]}^{(m',q')}\right) \\ &= (2M+1)^{-1}\left(\beta_j^{(m,m')}(u)\,\beta_j^{(q,q')}(u) + \beta_j^{(m,q')}(u)\,\beta_j^{(m',q)}(u)\right) \\ &\quad + \mathcal{O}(2^{2j}M^{-1}),\end{aligned} \tag{13}$$

where $T\to\infty$ and $M\to\infty$ such that $MT^{-1}\to 0$.

Under the notation $\left\{\left(\tilde{\lambda}_j^{(i)}(u), \tilde{\mathbf{e}}_j^{(i)}(u)\right)_{i=1}^{P}\right\}$ for the eigendecomposition of $\tilde{\boldsymbol{I}}_{j,[uT]}$, the following proposition shows that these are well-behaved estimators for the corresponding true quantities $\{\left(\lambda_j^{(i)}(u), \mathbf{e}_j^{(i)}(u)\right)_{i=1}^{P}\}$, the eigendecomposition of $\boldsymbol{\beta}_j(u)$.

Proposition 2. Under the assumptions of Proposition 1 and additionally assuming that (i) all moments of the process $\{X_{t,T}^{(p)}\}_t$ exist for any channel p and (ii) the $\boldsymbol{\beta}$-spectrum has bounded and uniformly continuous derivatives and unique eigenvalues at each scale-time (j,u), as $T\to\infty$, $M\to\infty$ such that $MT^{-1}\to 0$, the proposed eigenvalue and eigenvector estimators are consistent for the true ones, and

$$\begin{aligned}\mathbb{E}(\tilde{\lambda}_j^{(i)}(u)) &= \lambda_j^{(i)}(u) + \mathcal{O}(\sqrt{2^{2j}M^{-1}}),\\ \mathbb{E}(\tilde{\mathbf{e}}_j^{(i)}(u)) &= \mathbf{e}_j^{(i)}(u) + \boldsymbol{\mathcal{O}}(\sqrt{2^{2j}M^{-1}}),\\ \operatorname{var}(\tilde{\lambda}_j^{(i)}(u)) &= (M+1/2)^{-1}\left(\lambda_j^{(i)}(u)\right)^2 + \mathcal{O}(2^{2j}M^{-1}),\end{aligned}$$

$$\operatorname{cov}(\tilde{\lambda}_j^{(i)}(u), \tilde{\lambda}_j^{(i')}(u)) = \mathcal{O}(2^{2j}M^{-1}) \text{ when } i\neq i'$$

$$\begin{aligned}\operatorname{var}(\tilde{e}_{j,p}^{(i)}(u)) &= (2M+1)^{-1}\sum_{i'\neq i}\frac{\lambda_j^{(i')}(u)\,\lambda_j^{(i)}(u)}{\left(\lambda_j^{(i)}(u)-\lambda_j^{(i')}(u)\right)^2}\\ &\quad (e_{j,p}^{(i')}(u))^2 + \mathcal{O}(2^{2j}M^{-1}),\end{aligned}$$

$$\operatorname{cov}(\tilde{e}_{j,p}^{(i)}(u), \tilde{e}_{j,p'}^{(i'')}(u)) = \mathcal{O}(2^{2j}M^{-1}) \text{ when } i\neq i'',\ p\neq p'.$$

The eigenvalue results above coupled with the properties of the correction matrix A, ensure it is justified to examine the estimated corrected eigenvalues.

Proposition 3. Under the assumptions of Proposition 2, the (corrected) eigenvalue estimators, defined as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_j^{(PCA)}(u) &= \operatorname{diag}\left(\tilde{\lambda}_j^{(i)}(u)\right)_{i=1}^{P}, \quad \forall(j,u),\\ \widehat{\boldsymbol{S}}_\ell^{(PCA)}(u) &= \operatorname{diag}\left(\sum_{j=1}^{J} A_{\ell,j}^{-1}\,\tilde{\lambda}_j^{(i)}(u)\right)_{i=1}^{P}, \quad \forall(\ell,u),\end{aligned}$$

are consistent and asymptotically unbiased for the true eigenvalue quantities $\boldsymbol{\beta}_j^{(PCA)}(u)$ and $\boldsymbol{S}_\ell^{(PCA)}(u)$ of (3) and (4), respectively, at all scale-time pairs (j,u) and (ℓ,u), respectively, as $T\to\infty, M\to\infty$ such that $MT^{-1}\to 0$.

On the asymptotic normality of eigenvalues and eigenvectors. The entries of the smoothed (cross-) periodogram $\tilde{\boldsymbol{I}}_{j,[uT]}$ are typically modeled as asymptotically normal (Nason 2013; Hargreaves et al. 2019). Due to their asymptotically vanishing covariances (Proposition 1), it is reasonable to model them as asymptotically jointly normal and their asymptotic independence follows (see e.g., Brockwell and Davis 1991, p. 445). Hence, using similar arguments as for Fourier spectral decompositions (Brillinger 1969; Ombao and Ho 2006), the asymptotic normality of the eigenvalue and eigenvector estimators follow from perturbation theory (see (A.7) and (A.8) in Appendix A.4.2).

2.5. Empirical Estimation Performance of `mvLSW-PCA`

In this section we investigate our wavelet domain PCA approach on several simulated examples. For benchmarking our results,

we use the method of Lim, Kwon, and Oh (2021). In their paper, the authors demonstrate superior results to the Wigner-Ville transform approach proposed by Boashash, Azemi, and Khan (2015) as well as time-dependent Fourier domain PCA of Ombao and Ho (2006). Our own numerical experiments also reflect the inferiority of these techniques and thus we do not further explore these methods again here.

The processes in the simulation study were constructed to exhibit varying degrees of nonstationarity and coherence between series components in order to test our procedure on series arising in practical settings. In particular, we simulate realizations of length $T = 1024$ from the MvLSW model (1) with (i) diagonal spectral structure activated at various scales with time-constant behavior (Models 1 and 2); with varying intensities through time (Models 4, 5 and 6) and (ii) nonzero coherence within constant spectra through time (Model 3) and time-dependent cross-spectra (Model 7).

For completeness, we also replicate the simulation study performed in Lim, Kwon, and Oh (2021) who generate noncoherent series from the following models: (i) a bivariate model whose components are both concatenations of two Haar Moving Average (MA) processes, in the first component combining orders 1 and 3, and in the second one orders 1 and 5 (Model LKO1); (ii) a bivariate model whose components are a concatenation of four Haar MA processes of order 1, 2, 3, and 4 and a concatenation of two Haar moving average processes of order 2 and 3 (Model LKO2); (iii) a 20-dimensional process whose components are time-varying autoregressive series of order 2 (Model LKO3). For consistency with the other models, we simulate series of length $T = 1024$ for Models LKO1–LKO3. Full mathematical descriptions for all models we use can be found in Appendix B, along with visualizations of the spectral structure and corresponding process realizations for Models 4, 7 and LKO1.

For the results here, as well as the data analysis in Section 4, our nonstationary time series PCA methodology was implemented in the R statistical computing environment (R Core Team 2022), using modifications to the code from the `mvLSW` package (Taylor et al. (2017), see also Taylor, Park, and Eckley 2019). In what follows we denote our proposed method by `mvLSW-PCA`. We compare our procedure to the alternative estimation approach outlined in Lim, Kwon, and Oh (2021) (henceforth denoted `LKO`) using the authors' code available from *https://github.com/yaeji-lim/WDPCA*. For the choice of smoothing bandwidth in (10), for both methods we follow Taylor et al. (2017) and set $M = [\sqrt{T}]$. For each model, we evaluate the average performance of the two methods over $R = 100$ simulated realizations.

2.5.1. Estimation Assessment for Spectral Domain Eigenvalues

To assess how well our technique is able to estimate the true eigenvalues, we compute two indicative measures of fidelity of (corrected) eigenvalue estimation to the true structure, namely the time-scale-component averaged mean square error in estimation, and the time-scale-component bias incurred by the estimate, both averaged over the R simulated series. Expressions for these measures can be found in Appendix B, (B.1) and (B.2).

We compute these measures for all models except for Model LKO3, which does not have an inherent MvLSW spectral structure. The error results can be found in Table 1. It is striking that our estimation procedure on average achieves a 15% reduction in estimation error using the mean squared error metric (see (B.1)), while accompanied by a consistently lower uncertainty across all models. Similar observations can be made considering the average bias in the eigenvalue estimation (B.2). On careful examination, we observe that models which exhibit power across several scales are associated to most marked improvements in eigenvalue estimation, for example, 18%–19% for Models 2 and 3. This considerable reduction achievement is predominantly due to the much lower bias associated to our proposed approach, with bias improvements in favor of our method of 22% and 29% for Models 2 and 3, respectively. Further results appear in Appendix C.1 in supplementary materials accompanying this article.

We also repeated the eigenvalue estimation study for Models 1–7 with series lengths $T = 256$ and $T = 512$. Figure 2 and Appendix C.1 summarize the results and illustrate the consistent superiority of our proposed `mvLSW-PCA` method over the competitor, `LKO`.

When visually inspecting the average estimated eigenvalue structure, despite some leakage of information into neighboring scales, our procedure produces more reliable results when compared with the competitor `LKO` method, with this technique often drastically overestimating the values. An example of this behavior appears in Appendix C.2, which displays estimated eigenvalue structure (averaged over the $R = 100$ process realizations) for the `mvLSW-PCA` and `LKO` methods, respectively, for Models 4 and 7 (with reduced dimension $Q = 2$). Both estimates are illustrated against the eigenvalues corresponding to the true $\boldsymbol{S}$-spectrum, with ours rendering a more accurate depiction in the presence of inter-channel dependence, captured for example by the intensity agreement with the truth (see Appendix C.2 for Model 7). We attribute this consistently improved estimation of our procedure to the accurate estimation of the eigenvalues of $\boldsymbol{\beta}_j^{(X)}(u)$, prior to bias-correction via A^{-1} in Step 4 of our estimation procedure (Algorithm 1). An example

Table 1. Performance of the spectral domain eigenvalue estimation methods over $R = 100$ realizations of processes from Models 1–7 and Models LKO1 and LKO2.

Method	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model LKO1	Model LKO2
					MSE				
`mvLSW-PCA`	0.08(0.02)	5.04(1.48)	5.19(1.55)	0.14(0.03)	0.08(0.02)	4.39(1.36)	6.08(2.06)	0.11(0.06)	0.07(0.02)
`LKO`	0.10(0.02)	6.15(1.77)	6.42(1.85)	0.17(0.04)	0.10(0.02)	5.03(1.53)	6.86(2.37)	0.10(0.06)	0.07(0.02)
					Bias				
`mvLSW-PCA`	0.06(0.01)	0.46(0.12)	0.47(0.11)	0.08(0.02)	0.06(0.01)	0.42(0.11)	0.43(0.11)	0.07(0.03)	0.04(0.01)
`LKO`	0.07(0.01)	0.59(0.11)	0.66(0.12)	0.09(0.02)	0.07(0.01)	0.52(0.10)	0.57(0.12)	0.08(0.03)	0.07(0.02)

NOTE: Standard errors are shown in brackets.

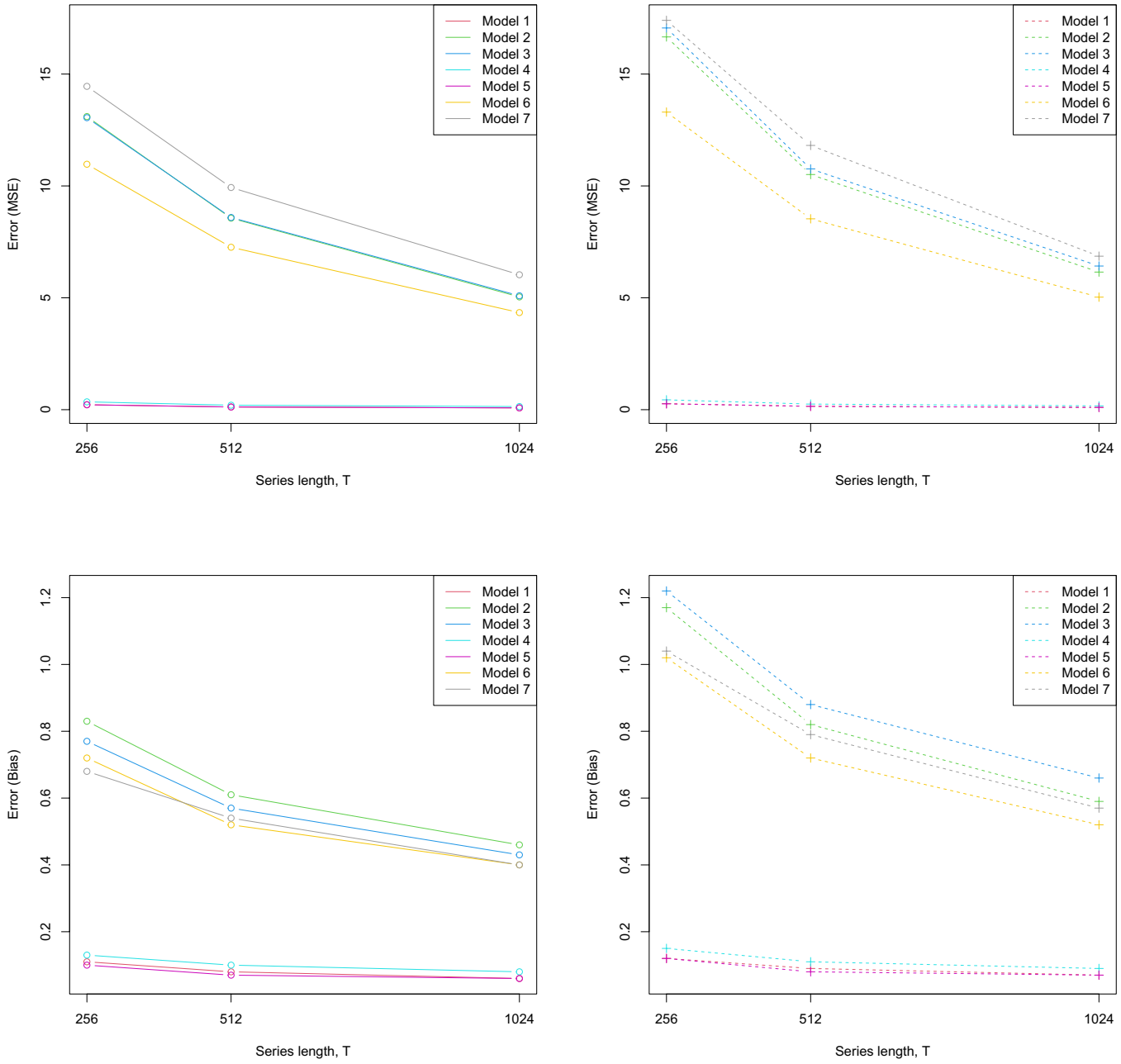

Figure 2. PC Spectra estimation performance with different series lengths, T.Top-left: MSE for `mvLSW-PCA`; top-right: MSE for `LKO`; bottom-left: bias for `mvLSW-PCA`; bottom-right: bias for `LKO`.

of the resulting (uncorrected) estimated eigenvalue structure for Model LKO1 with good agreement between the true and estimated $\boldsymbol{\beta}$-structure and intensity appears in Appendix C.2.

2.5.2. Estimation Assessment for Spectral Domain Eigenvectors

We next assess the quality of eigenvector structure estimation for both `mvLSW-PCA` and `LKO` methods. In particular, we compare the true PCA domain eigenvectors with the respective eigenvectors estimated by the two methods. This augments the investigation in Lim, Kwon, and Oh (2021) which only reports results regarding the estimated eigenvalue structures.

Contrasting `mvLSW-PCA` and `LKO` approaches, note that in general there is no direct connection between the eigenvectors of the spectral decomposition of $\boldsymbol{\beta}_j^{(X)}(u)$ and those of $\boldsymbol{S}_j^{(X)}(u)$ (see Appendix A.3). An exception occurs when the matrices $\boldsymbol{\beta}_j^{(X)}(u)$ and $\boldsymbol{S}_j^{(X)}(u)$ commute: they generate the same set of eigenvectors (Abadir and Magnus 2005, p. 180). For example, Models 1, 2, and 5 were designed to fulfill this condition, however this cannot be generally assumed.

Hence, in what follows we compare each method's performance with its respective true PCA domain eigenvector structure, that is, estimated and true eigenvectors of $\boldsymbol{\beta}_j^{(X)}(u)$ for our proposed `mvLSW-PCA` procedure (denoted $\mathbf{e}_j^{(i)}(u)$ in Section 2.2), and estimated and true eigenvectors of $\boldsymbol{S}_j^{(X)}(u)$ for the `LKO` method. As customary for analysis tasks involving PCA eigenvector structure, we also focus on their absolute value.

Table 2. Performance of the spectral domain eigenvector estimation methods over $R = 100$ realizations of processes from Models 1–7 and Models LKO1 and LKO2.

Method	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model LKO1	Model LKO2
					MSE				
mvLSW-PCA	0.10(0.02)	0.12(0.02)	0.10(0.02)	0.19(0.01)	0.08(0.01)	0.13(0.02)	0.04(0.00)	0.14(0.03)	0.28(0.05)
LKO	0.31(0.01)	0.32(0.01)	0.30(0.01)	0.31(0.01)	0.32(0.01)	0.31(0.01)	0.27(0.01)	0.34(0.02)	0.34(0.01)
					Bias				
mvLSW-PCA	0.12(0.01)	0.12(0.01)	0.09(0.01)	0.14(0.00)	0.10(0.00)	0.13(0.01)	−0.01(0.00)	0.08(0.01)	0.11(0.01)
LKO	0.15(0.00)	0.15(0.00)	0.13(0.00)	0.16(0.00)	0.14(0.00)	0.15(0.00)	0.13(0.00)	0.10(0.01)	0.12(0.01)

NOTE: Standard errors are shown in brackets.

The results in Table 2 highlight the dramatic MSE improvement attained by our method which reduces the estimation error 2- to 3-fold across all models. This is further emphasized by the results for the models designed to have the same eigenvectors regardless of whether we carry out a usual spectrum or a $\boldsymbol{\beta}$-spectrum decomposition. As the eigenvector estimates obtained via our method also incur a lower bias than those obtained by LKO method, we conclude that our mvLSW-PCA method yields eigenvector estimates with both lower bias and variance. These results come as no surprise, given the inferior asymptotic convergence rates of the estimated spectrum to the wavelet spectrum $\boldsymbol{S}$ adopted in the LKO method, compared to those of the smoothed periodogram to $\boldsymbol{\beta}$ for mvLSW-PCA. Additionally, the regularization often needed in practice for the spectral estimator of $\boldsymbol{S}$ (Taylor, Park, and Eckley 2019) also compounds the problems exhibited by the eigendecomposition of $\hat{\boldsymbol{S}}$ as proposed by Lim, Kwon, and Oh (2021).

Appendix C.3 displays the average absolute value of the estimated eigenvectors (over the $R = 100$ realizations) for Model 7 for our mvLSW-PCA method, together with the true eigenvector absolute value (of $\boldsymbol{\beta}_j^{(X)}(u)$), for principal components PC1 – PC2, respectively. Similarly, the average absolute values of the estimated eigenvectors (over the $M = 100$ realizations) for Model 7 for LKO method are also presented, together with the true eigenvector absolute value (of $\boldsymbol{S}_j^{(X)}(u)$), for principal components PC1 – PC2, respectively. While our estimation procedure is able to identify scale and/or temporal changes in the structure of the eigenvectors across all principal components, the LKO method struggles to find any key structure in the true eigenvectors of $\boldsymbol{S}_j^{(X)}(u)$. In particular, the information in the second and third principal component is often estimated poorly. These visual appraisals indeed reflect the numerical results reported in Table 2.

2.5.3. Estimation Assessment for the Percentage Variance Explained

We now explore whether the two competitor methods mvLSW-PCA and LKO select an appropriate number of principal components Q. We point out here that this quantity may exhibit time-dependency. For this reason, we choose to investigate the scale-time-component percentage variance explained as well as the behavior of Q when driven by a chosen percentage variance explained for Models 4 and 7. In this context, we have discovered that fueled by the poorer eigendecomposition estimation, LKO is inferior to mvLSW-PCA in adequately reflecting the time-varying and scale traits of the data (see Appendix C.4 for Models 4 and 7). Most importantly, we note that the LKO-based explained variance appears to be dominated by PC1 and to be less sensitive to the time-dependent aspect of the number of principal components required in order to explain a chosen amount of variation in the data.

3. Principal Component Cross-Coherence

A key component to a successful PCA is the interpretation of the emerging principal components. In this section we explore the wavelet domain connection between the (transformed) adaptively rotated and original processes, which will motivate the principled construction of a new measure of association. Recall that coherence is a commonly used tool that extracts the common scale/frequency content of two or more time series (Shumway and Stoffer 2010). Current locally stationary wavelet modeling literature has focused on the coherence between channels in a *single* MvLSW process (see Sanderson, Fryzlewicz, and Jones 2010; Park, Eckley, and Ombao 2014). However, our construction crucially differs from that work in quantifying dependence between different *processes* as opposed to different channels, hence, our use of the term *cross*-coherence. Our proposed cross-coherence measure essentially provides a normalized quantification of the contribution of the original process components to each wavelet domain principal component and will form an important component in the practical application of our proposed principal component methodology. Nevertheless, we note that this measure is not developed for use in a general cross-MvLSW process coherence.

3.1. Cross-Coherence as a Measure of Dependence

To formulate our new cross-coherence measure that will quantify the dependence between the rotated and original channels, we first consider the representation of an MvLSW process, following our adaptive principal component rotation. Mathematically, for each scale j and rescaled time $u = k/T$, our proposed mvLSW-PCA technique essentially constructs the process

$$\tilde{\mathbf{Y}}_{t,T}^{(j,k)} = \sum_{j'} \sum_{k'} \mathbf{V}_{j'}^{(\tilde{Y};j,k)}(k'/T)\psi_{j',k'}(t)\mathbf{z}_{j',k'}, \tag{14}$$

where the new process amplitudes at scale j' and time k' are defined as $\mathbf{V}_{j'}^{(\tilde{Y};j,k)}(k'/T) = \mathbf{E}_j^{(X)}(k/T)\mathbf{V}_{j'}^{(X)}(k'/T)$ and $\mathbf{E}_j^{(X)}(u)$ is the orthonormal $P \times P$ matrix of (row) eigenvectors corresponding to $\boldsymbol{\beta}_j^{(X)}(u)$. Note the amplitudes are the adaptively rotated transfer function of the original process, with the rotation specific to the scale j and time k at which the eigendecomposition was performed. Observe that the resulting process, whose individual channels are the principal components at a

fixed scale-time (j,k) (ordered by eigenvalue magnitude), can also be expressed as $\tilde{\mathbf{Y}}_{t,T}^{(j,k)} = \mathbf{E}_j^{(X)}(k/T)\mathbf{X}_{t,T}$. Details on the connection between the empirical wavelet coefficients of this process (the analogue of the vector in (5)) to those of $\mathbf{X}_{t,T}$ can be found in Appendix A.5.

Connection between the Principal Components and the Original Process

In order to derive the cross-coherence, which in this context is not only a scale (frequency) dependent quantity but is also time-evolving, we first construct the *cross-spectrum of the new, adaptively rotated process at each fixed scale j and time* $k = [uT]$ $\{\tilde{\mathbf{Y}}_{t,T}^{(j,k)}\}$ *and the original process* $\{\mathbf{X}_{t,T}\}$ as

$$\begin{aligned}\mathbf{S}_{j'}^{(\tilde{\mathbf{Y}}^{(j,k)},\mathbf{X})}(k'/T) &= \mathbf{V}_{j'}^{(\tilde{Y};j,k)}(k'/T)\left(\mathbf{V}_{j'}^{(X)}(k'/T)\right)^T,\\ &= \mathbf{E}_j^{(X)}(k/T)\mathbf{V}_{j'}^{(X)}(k'/T)\left(\mathbf{V}_{j'}^{(X)}(k'/T)\right)^T,\\ &= \mathbf{E}_j^{(X)}(k/T)\mathbf{S}_{j'}^{(X)}(k'/T),\ \forall(j',k'). \qquad (15)\end{aligned}$$

Definition 1. We define the cross-coherence between the ith principal component (obtained through an adaptive rotation at fixed scale j and time $k = [uT]$) and the original pth channel as the (i,p) entry of the matrix

$$\begin{aligned}\rho_j^{(\tilde{\mathbf{Y}}^{(j,k)},\mathbf{X})}(k/T) =& \operatorname{diag}\left(\left[\mathbf{S}_j^{(\tilde{Y}^{(j,k)})}(k/T)\right]^{-1/2}\right)\left(\mathbf{S}_j^{(\tilde{Y}^{(j,k)};X)}(k/T)\right)\\ &\operatorname{diag}\left(\left[\mathbf{S}_j^{(X)}(k/T)\right]^{-1/2}\right), \qquad (16)\end{aligned}$$

where "diag" extracts the diagonal entries of the spectral matrices corresponding to the new and original processes, see equations (A.9) (Appendix A.5 details its derivation) and (2), respectively, and the cross-spectrum is given by (15).

Note that the diagonal quantities are indeed positive since they are the spectra for the rotated and original processes. Additionally, the cross-coherence is only defined at those scales and locations (j,k) which correspond to nonzero spectral quantities.

As in previous developments for locally stationary processes, a coherence measure can be thought of as a scale- and time-dependent "correlation"-like quantity. It can be easily verified that it ranges from -1, indicating an absolute negative correlation, to 1 indicating an absolute positive correlation. In a multi-dimensional set of nonstationary time series observations, we can therefore reduce the information complexity by understanding which combinations of channels respond with highest spectral power across scale and time.

3.2. Estimating the Proposed Cross-Coherence Measure

We now provide theoretical and practical results underlying the estimation of the cross-coherence measure introduced above (see also Appendix A.6).

Proposition 4. Under the assumptions of Proposition 2 and for each fixed scale j and rescaled time u, the (i,p) entry of the cross-coherence estimator matrix

$$\begin{aligned}\widehat{\rho}_j^{(\tilde{\mathbf{Y}}^{(j,[uT])},\mathbf{X})}(u) = \operatorname{diag}&\left(\left[\tilde{\mathbf{E}}_j^{(X)}(u)\hat{\mathbf{S}}_j^{(X)}(u)\left(\tilde{\mathbf{E}}_j^{(X)}(u)\right)^T\right]^{-1/2}\right)\\ &\times\left(\tilde{\mathbf{E}}_j^{(X)}(u)\hat{\mathbf{S}}_j^{(X)}(u)\right)\\ &\times\operatorname{diag}\left(\left[\hat{\mathbf{S}}_j^{(X)}(u)\right]^{-1/2}\right) \qquad (17)\end{aligned}$$

is consistent for the true cross-coherence $\rho_j^{(\tilde{\mathbf{Y}}^{(j,[uT])},\mathbf{X})}(u)$ between the ith principal component adaptively rotated at (j,u) and the pth original channel.

Cross-Coherence Estimation Assessment via Simulation

To verify our coherence measure estimation procedure outlined above, we assessed the accuracy in estimating the true cross-coherence between the original and rotated processes for Models 3 and 7, which contain dependence between the original channels. Similar to Section 2.5.1, we compute the empirical average mean squared error and bias when estimating the absolute true cross-coherence across $R = 100$ simulated processes, where we average over the levels which contain coherence. In the case of Models 3 and 7 these are levels 2 and 4 (see Appendix B). While the cross-coherence is slightly overestimated, the estimates remain accurate despite the challenging complex time-varying dependence between the original and principal component processes. The average MSE and bias values were 0.14 (standard error: 0.01) and 0.09 (standard error: 0.009) for Model 3 and 0.06 (standard error: 0.007) and 0.03 (standard error: 0.009) for Model 7. An illustration of the estimated cross-coherence against the truth for Model 7 is shown in Appendix D for PC1–PC2. We also repeated the simulation experiment for series lengths $T = 256$ and $T = 512$; reassuringly, the MSE and bias decrease with increasing sample size.

Cross-Coherence Estimation for Experiments with Repeated Trials

In our motivating data example (revisited below in Section 4), participants were asked to repeat a cognitive task several times, resulting in multiple measurements of neurological activity for each participant. In such an experiment, we wish to pool information across the per-trial multivariate time series, and hence we propose to estimate the cross-coherence in this data setting via the trial-averaged spectral estimates

$$\begin{aligned}\overline{\widehat{\rho}}_j^{(\tilde{\mathbf{Y}}^{(j,[uT])},\mathbf{X})}(u) = \operatorname{diag}&\left(\left[\overline{\tilde{\mathbf{E}}}_j^{(X)}(u)\overline{\hat{\mathbf{S}}}_j^{(X)}(u)\left(\overline{\tilde{\mathbf{E}}}_j^{(X)}(u)\right)^T\right]^{-1/2}\right)\\ &\times\left(\overline{\tilde{\mathbf{E}}}_j^{(X)}(u)\overline{\hat{\mathbf{S}}}_j^{(X)}(u)\right)\\ &\times\operatorname{diag}\left(\left[\overline{\hat{\mathbf{S}}}_j^{(X)}(u)\right]^{-1/2}\right), \qquad (18)\end{aligned}$$

where $\bar{\cdot}$ refers to the averages over trials of the quantities in (17). This is a consistent estimator of the true participant cross-coherence and in practice we have found this estimator is better behaved than its trial-averaged coherence counterpart.

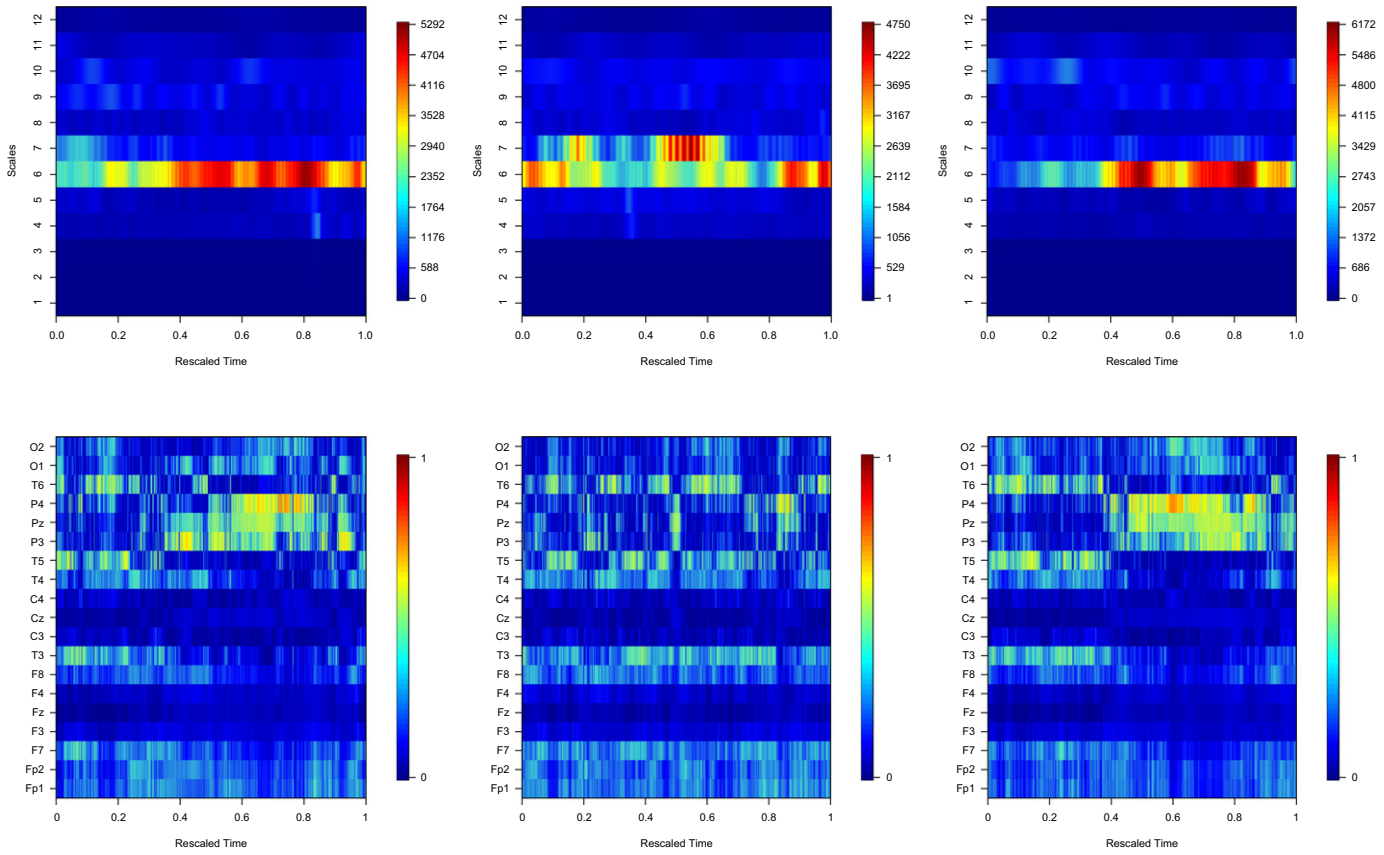

Figure 3. Participant 1 EEG activity patterns for level 6 (PC1): Estimated eigenvalues for left-tapping, resting state, and right-tapping conditions (top row); estimated principal component cross-coherence for left-tapping, resting state, and right-tapping conditions, displayed by channel (bottom row).

4. Case Study: EEG Data Analysis

We now apply the proposed wavelet domain dimension reduction methodology and implement the `mvLSW-PCA` tool (Sections 2.2 and 2.3) on the motivating neuroscience dataset from Section 1. The experiment consists of 120 EEG trials recorded at 1024Hz during which the participant was required to perform one single index finger tap following a cue that randomly instructed the use of left hand, right hand or to rest (40 trials per condition).

The data contains brain activity recorded across 19 channels, with placements illustrated in Figure 1, and features complex nonstationarities and channel-interdependence. Two additional channels recorded the precise duration of the right and left tap, respectively, via a microcontroller tapping device. The analysis of temporal dynamics in neurological data from such tasks is important for a number of reasons, for example to quantify rehabilitation in patients suffering from neuropathologies, or to contribute to the more fundamental question of determining which frequencies and areas of the brain dominate neuroactivity in different task-scenarios. It has been well-established that neuronal activity does not happen in specific brain regions in isolation, and that those regions can change over time. Such interhemispheric interaction has been observed in the context of neurological disorders, for example, seizures (Duckrow and Spencer 1992; Murase et al. 2004; Ombao and Ho 2006).

The data is publicly available from *https://researchdata.reading.ac.uk/117/* and is further described in Wairagkar (2017). After pre-processing, the resulting data are zero mean, in accordance to our MvLSW assumption. For each trial, the signal we analyze spans 4 sec (4096 observations) and is extracted such that the onset of the finger tap is at rescaled time 0.45 across all trials. For the resting state trials, the last 4 sec from each 10 sec trial were extracted. In what follows we analyze data from participant 1, a left-handed healthy individual. The (rescaled) tapping time is approximately 0.45–0.55 across trials. For each condition we follow Algorithm 1 and retain the first two principal components ($Q = 2$) which explain in excess of 82% of the process power. As several trials are available per condition, we use the cross-coherence estimator proposed in (18) in order to facilitate an understanding of the salient component features and time-dependent channel activation.

Resting State

Using the proposed principal component analysis for the resting data as baseline reference for the analyses using the right- and left-trials, otherwise hidden brain activation patterns emerge. The power across all principal components is primarily evident at scale 6 (encompassing alpha band, 8–12Hz, indicating alertness, and low beta band, 12–16Hz, connected to specific motor actions). Additional flares are present at coarser scale 7 (theta band, 4–8Hz characteristic of slower activity). Recalling the participant is left-handed, we note these are also present when left-tapping but not for right taps. The resting data is interestingly characterized by a lack of discernible time patterns, except for a peak at the movement onset and tapping time within theta band frequencies (Figure 3, top-middle).

Resting state coherence patterns. The first principal component (PC1) coherence analysis reveals that during resting time there are no temporal patterns of activation within alpha and low beta band frequencies. Left and right front, temporal and parietal channels exhibit some middling level of activation (absolute coherence values around 0.3–0.5) with no particular emergent features. The coherence patterns for the second principal component (PC2) in Appendix E show uniform primary activation of left-side channels Fp1, F7, and T3.

Left-Tapping

By contrast to the resting state, trials that involve tapping reveal emergent temporal patterns. Left-tapping PC1 is associated to periods of high activation at alpha and low beta band frequencies, with peaks observed just before the movement onset and during tapping (rescaled time 0.4–0.55, recall onset is at 0.45), as well as to short intense bursts in the post-tapping period (0.65–0.85). Notable power is also present in PC1 at theta band at the beginning of the trial sequence (up to rescaled time 0.3) in anticipation of the movement (see Figure 3, top-left). For PC2 the activity is peaking within rescaled time 0.65–0.75 at alpha and low beta band (Appendix E).

Left-tapping coherence patterns. The coherence patterns for PC1 reveal temporal activation windows within the alpha and low beta frequencies, with uniform activation across the theta band (Figure 3, bottom-left). Namely, for the alpha and low beta band we observe an initial activation in the left frontal F7 and neighboring left temporal lobes T3 and T5, followed by a shift to right temporal lobe activation in T6 (up to rescaled time 0.45). During movement onset and tapping time, the activity becomes localized in the left and central parietal lobes, P3 and Pz. The right parietal lobe P4 becomes activated in the immediate period following the tapping (until about 0.85 rescaled time). The prefrontal lobes (Fp1 and Fp2) exhibit weak coherence with left-tapping PC1 in the period preceding and during movement. Within theta band, no temporal patterns are present along the primary activation channels, right temporal (T6) and left temporal T3 and T5 (coherence values of roughly 0.6). Weaker coherence is also present within right temporal (T4) lobe and left and right frontal (F7, F8) channels, as well as in occipital lobes (O1 and O2). Within alpha and low beta band, uniformly distributed coherence appears between PC2 and the left frontal (F7) and temporal (T3) channels, with some activity present outside tapping time in the left pre-frontal (Fp1) lobe. During tapping time, there is an activation in left temporal T5 lobe coupled with a deactivation of Fp1.

Right-Tapping

Right-tapping PC1 has power primarily at alpha and low beta frequencies with activity peaks at movement onset and tapping time followed by the post-tapping period (0.65–0.85), and for PC2 the activity is peaking within rescaled time 0.65–0.75. While these traits are broadly similar to those found in the left-tapping trials, the intense activity periods are shorter and power is present for the right-hand tapping PC1 at theta band at the same peaks as for alpha and low beta frequencies (as opposed to the pre-movement time for left-tapping PC1) and at higher beta band waves around cue and tapping time for PC2.

Right-tapping coherence patterns. The coherence patterns for PC1 highlight activation around the onset and tapping time. Specifically, for alpha and low beta band an initial activation occurs in both right and left temporal lobes T5 and T6 with activation present in the left temporal lobe T3 until movement time. This is followed by a shift to right and central parietal lobe activation (P4 and Pz) up to rescaled time 0.85. Following tapping, the left parietal lobe P3 becomes activated until about rescaled time 0.85. Within theta waves, akin to the left-tapping, there are no patterns in the frontal (F7 and F8), temporal (T3, T4, T5, and T6) and occipital (O1 and O2) lobes that exhibit main activity. Within alpha and low beta frequencies, stronger coherence appears between PC2 and the left temporal (T3) and left frontal (F7) channels up to movement onset time, followed by uniform activity. The left pre-frontal (Fp1) cortex appears engaged throughout, while immediately following tapping there is an activation in the left temporal T5 lobe. Coherence at higher beta-band waves between PC2 and left pre-frontal (Fp1), frontal (F7) and temporal (T3) channels is sustained throughout (Appendix E).

Task-Oriented Comparisons

For this left-handed individual, the salient left-tapping PC1 is associated to longer periods of high activation at alpha and low beta bands when compared to right-tapping PC1. As opposed to the left-tapping PC1, the activity at alpha and low beta frequencies in the left and central parietal lobes is sustained almost entirely during the trial times that follow right-tapping; the temporal T5 lobe appears strongly associated with the movement period for PC2 and completely disengaged during the resting trials. Activity in PC1 at alpha and low beta band frequencies appears associated to the left frontal F7 channel for left-tapping and resting, while F7 is only associated with right-tapping at theta band (PC1) and high beta frequencies (PC2). Interestingly, the left pre-frontal (Fp1) lobe appears activated throughout the right tapping trials (PC2 at alpha and beta bands) and only outside the tapping period for the left-tapping trials (PC2 at alpha and low beta bands). The activity entailed by this experiment does not appear to engage the central lobes, regardless of the left/right nature of the tapping or resting. See a schematic of the salient features in Appendix E.

Comparisons with an analysis using the* `LKO` *method. To illustrate the difference between our proposed method and that of Lim, Kwon, and Oh (2021), we repeated the data analysis using the `LKO` technique. The resulting eigenvalues, together with the corresponding eigenvectors (in absolute value) are shown for PC1–PC2 in Appendix E. Note the very little difference in structure between PC1 and PC2, as well as the poor temporal localization of tapping activity during the experiment. The lack of discernible structure in the eigenvectors equate to weak activation-driven association of the PCs to the original process. In light of the results from Section 2.4, empirically verified in Section 2.5, we would argue that the real data analysis using our proposed `mvLSW-PCA` adaptive rotation procedure and

associated cross-coherence measure leads to a more informative analysis.

5. Conclusions

This article has proposed a new framework for wavelet domain principal component analysis for nonstationary time series, and also introduced a new cross-coherence measure between the process in the time domain and its principal component domain representation. We detailed a full estimation procedure for the wavelet domain PCA quantities, considering asymptotic properties. We demonstrated the high accuracy of our estimation approach through simulated examples, highlighting its improved performance over a state-of-the-art competitor method while being able to cope well with processes featuring a range of nonstationary dependence. We also illustrated the utility of our proposed methodology on a dataset from an experimental study on EEG dynamics. While the motivation for the work in this article arose in a neuroscience setting, we envisage that the tools developed here have the potential to provide valuable data insights in a wide range of applications which generate high-dimensional nonstationary time series that exhibit complex dynamics.

Supplementary Materials

The file `supp.zip` contains further results and code associated to this work, as described below (see also its readme file).

Proofs of theoretical results: This document (Appendix A) provides details of the results on our proposed methodology in Sections 2 and 3. (`proofs.pdf`)

Further simulation details: Here we provide model specifications for the simulations in Section 2.5 (Appendix B) as well as additional simulation results for different series lengths and supporting plots for the estimation performance in Sections 2 and 3 (Appendix C–D). In Appendix E we also include supporting figures for the EEG case study in Section 4. (`simstudy.pdf`)

code: *R* code implementing the proposed methods in this article. (`mvLSWPCA.R`)

Acknowledgments

The authors would like to gratefully thank the Editor, Associate Editor, and two anonymous reviewers for their constructive comments and suggestions that helped improve the quality of the manuscript.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

MIK and MAN gratefully acknowledge support from EPSRC grant EP/X002195/1.

ORCID

Marina I. Knight http://orcid.org/111.555
Matthew A. Nunes http://orcid.org/111.555
Jessica K. Hargreaves http://orcid.org/111.555

References

Abadir, K. M., and Magnus, J. R. (2005), *Matrix Algebra* (Vol. 1), Cambridge: Cambridge University Press. [8]

Boashash, B., Azemi, G., and Khan, N. A. (2015), "Principles of Time–Frequency Feature Extraction for Change Detection in Non-stationary Signals: Applications to Newborn EEG Abnormality Detection," *Pattern Recognition*, 48, 616–627. [7]

Brillinger, D. R. (1969), "The Canonical Analysis of Stationary Time Series," in *Multivariate Analysis* (Vol. II), ed. P. R. Krishnaiah, pp. 331–350, New York: Academic Press. [6]

——— (1975), *Time Series Data Analysis and Theory*, New York: Holt, Rinehart and Winston. [1,4]

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer. [1,6]

Das, S., and Politis, D. N. (2020), "Predictive Inference for Locally Stationary Time Series with An Application to Climate Data," *Journal of the American Statistical Association*, 116, 1–16. [1]

Duckrow, R. B., and Spencer, S. S. (1992), "Regional Coherence and the Transfer of Ictal Activity During Seizure Onset in the Medial Temporal Lobe," *Electroencephalography and Clinical Neurophysiology*, 82, 415–422. [11]

Embleton, J., Knight, M., and Ombao, H. (2022), "Multiscale Spectral Modelling for Nonstationary Time Series Within An Ordered Multiple-Trial Experiment," *Annals of Applied Statistics*, 16, 2774–2803. [3,4]

Fryzlewicz, P., and Nason, G. (2006), "Haar–Fisz Estimation of Evolutionary Wavelet Spectra," *Journal of the Royal Statistical Society*, Series B, 68, 611–634. [6]

Fryzlewicz, P., Sapatinas, T., and Subba Rao, S. (2006), "A Haar–Fisz Technique for Locally Stationary Volatility Estimation," *Biometrika*, 93, 687–704. [1]

Hargreaves, J., Knight, M., Pitchford, J., Oakenfull, R., Chawla, S., Munns, J., and Davis, S. (2019), "Wavelet Spectral Testing: Application to Nonstationary Circadian Plant Rhythms," *Annals of Applied Statistics*, 13, 1817–1846. [1,4,6]

Jolliffe, I. T. (2002), *Principal Component Analysis*, New York: Springer Verlag. [1]

Lansangan, J. R. G., and Barrios, E. B. (2009), "Principal Components Analysis of Nonstationary Time Series Data," *Statistics and Computing*, 19, 173–187. [1]

Lim, Y., Kwon, J., and Oh, H.-S. (2021), "Principal Component Analysis in the Wavelet Domain," *Pattern Recognition*, 119, 108096. [1,2,4,5,6,7,8,9,12]

McGonigle, E. T., Killick, R., and Nunes, M. A. (2022), "Trend Locally Stationary Wavelet Processes," *Journal of Time Series Analysis*, 43, 895–917. [1]

Murase, N., Duque, J., Mazzocchio, R., and Cohen, L. G. (2004), "Influence of Interhemispheric Interactions on Motor Function in Chronic Stroke," *Annals of Neurology*, 55, 400–409. [11]

Nason, G. P. (2008), *Wavelet Methods in Statistics with R*, New York: Springer. [1]

——— (2013), "A Test for Second-Order Stationarity and Approximate Confidence Intervals for Localized Autocovariances for Locally Stationary Time Series," *Journal of the Royal Statistical Society*, Series B, 75, 879–904. [2,6]

Nason, G. P., 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, 271–292. [3]

Ombao, H. C., and Ho, M. R. (2006), "Time-Dependent Frequency Domain Principal Components Analysis of Multichannel Non-stationary Signals," *Computational Statistics & Data Analysis*, 50, 2339–2360. [1,2,4,6,7,11]

Park, T., Eckley, I. A., and Ombao, H. C. (2014), "Estimating Time-Evolving Partial Coherence between Signals via Multivariate Locally Stationary Wavelet Processes," *IEEE Transactions on Signal Processing*, 62, 5240–5250. [1,2,3,4,5,6,9]

Priestley, M. B. (1983), *Spectral Analysis and Time Series*, London: Academic Press. [1]

R Core Team. (2022), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [7]

Roueff, F., and Von Sachs, R. (2019), "Time-Frequency Analysis of Locally Stationary Hawkes Processes," *Bernoulli*, 25, 1355–1385. [1]

Sanderson, J., Fryzlewicz, P., and Jones, M. W. (2010), "Estimating Linear Dependence between Nonstationary Time Series Using the Locally Stationary Wavelet Model," *Biometrika*, 97, 435–446. [3,4,5,6,9]

Sanei, S., and Chambers, J. A. (2013), *EEG Signal Processing*, Chichester: Wiley. [1]

Shumway, R. H., and Stoffer, D. S. (2010), *Time Series Analysis and Its Applications*, New York: Springer. [1,9]

Taylor, S., Park, T., and Eckley, I. A. (2019), "Multivariate Locally Stationary Wavelet Analysis with the mvLSW R package," *Journal of Statistical Software*, 90, 1–19. [7,9]

Taylor, S., Park, T., Eckley, I. A., and Killick, R. (2017), "mvLSW: Multivariate Locally Stationary Wavelet Process Estimation," R package version 1.2.1. *https://CRAN.R-project.org/package=mvLSW* [7]

Wairagkar, M. (2017), "EEG Data for Voluntary Finger Tapping Movement." University of Reading. *https://researchdata.reading.ac.uk/117/* [11]

Wilson, R., Eckley, I., Nunes, M., Park, T. (2021), "A Wavelet-based Approach for Imputation in Nonstationary Multivariate Time Series," *Statistics and Computing*, 31, Article 18. [1]