Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Spectral Dependence

Hernando Ombao[a,*], Marco Pinto[b]

[a] Statistics Program, King Abdullah University of Science and Technology (KAUST), Saudi Arabia
[b] Oslo Metropolitan University, Norway

ARTICLE INFO

Article history:
Received 30 March 2021
Revised 24 June 2022
Accepted 31 October 2022
Available online xxx

Keywords:
Causality
Cramér representation
Cross-coherence
Dual-frequency coherence
Fourier transform
Harmonizable processes
Multivariate time series
Spectral Analysis

ABSTRACT

A general framework for modeling dependence in multivariate time series is presented. Its fundamental approach relies on decomposing each signal inside a system into various frequency components and then studying the dependence properties through these oscillatory activities. The unifying theme across the paper is to explore the strength of dependence and possible lead-lag dynamics through filtering. The proposed framework is capable of representing both linear and non-linear dependencies that could occur instantaneously or after some delay (lagged dependence). Examples for studying dependence between oscillations are illustrated through multichannel electroencephalograms. These examples emphasized that some of the most prominent frequency domain measures such as coherence, partial coherence, and dual-frequency coherence can be derived as special cases under this general framework. Related approaches for modeling dependence through phase-amplitude coupling and causality of (one-sided) filtered signals are also introduced.

1. Introduction

One of the key goals in the analysis of multivariate time series data is to study the interactions between its different components. Multivariate time series data is being collected and analyzed in fields including environmental and climate science (Sethi and Mittal, 2020; Usman et al., 2020; Mahecha et al., 2020; Pham et al., 2020; Tatsis et al., 2020; Harvill et al., 2017), finance (Koutlis and Kugiumtzis, 2021; Hlávka et al., 2020; Reboredo and Ugolini, 2020; Qin et al., 2021; Bai et al., 2020; Ravagli and Boshnakov, 2020), computer science and engineering (Lin and Michailidis, 2018; Avendaño-Valencia and Chatzi, 2020; Blázquez-García et al., 2020; Gupta et al., 2020; Park and Jung, 2020; Huang et al., 2021), public health (Del Giorno et al., 2020; Aburto et al., 2020; Baum et al., 2020; Martín Cervantes et al., 2020; Lin et al., 2020), and neuroscience (Chen et al., 2020; Steinmann et al., 2020; Goyal and Garg, 2020; Soleimani et al., 2020; Manomaisaowapak et al., 2020; Suotsalo et al., 2020; Schoenberg, 2020; Manomaisaowapak and Songsiri, 2020; Pinto-Orellana and Hammer, 2020a; Nascimento et al., 2020; Ting et al., 2021b; Jiao et al., 2021; Maia et al., 2020; Sundararajan et al., 2020; Guerrero et al., 2021; Ombao, 2019). Considering the inherent complexity of those studied phenomena, one of the most common challenges and tasks is identifying and explaining the inter-relationship between the various components of the multivariate data. Thus, the purpose of this paper is to provide a summary of the various characterizations of dependence between the elements

* Corresponding author.
E-mail address: hernando.ombao@kaust.edu.sa (H. Ombao).

https://doi.org/10.1016/j.ecosta.2022.10.005

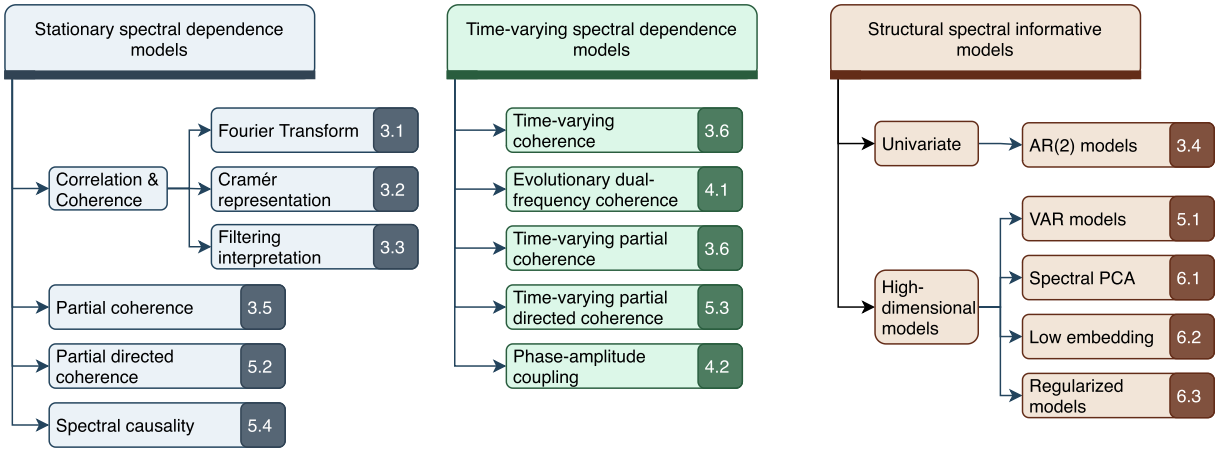

Fig. 1. Categories of dependence models that are covered in the current paper. The section in which the type of dependence measure is explained is indicated by a shaded box.

of a multivariate time series. The emphasis will be on the spectral measures of dependence which essentially examines the cross-relationships between the various oscillatory activities in these signals. Many of these measures are already being used to study dependence in multivariate time series data. One contribution of this paper is to demonstrate a unified framework under which these measures can be derived. The proposed framework here is to study the inter-relationships through the oscillatory activity derived from linear filtering. This proposed framework is in fact very intuitive and one of its potential advantage is that it can serve as a mechanism for studying more complex dependence structures in the multivariate time series data.

Define $\mathbf{X}(t) = [X_1(t), \ldots, X_P(t)]'$ to be a multivariate time series with P components. In environmental studies, components $X_1(t), \ldots, X_P(t)$ could represent recordings from various air pollution sensors at time t. Similarly, the same framework could describe wind velocity recordings at P different geographical locations. In a neuroscience experiment, a component X_p could be the measurement of brain electrical, or hemodynamic, activity from a specific sensor (electrode) which is placed either on the scalp or on the surface of the brain cortex. The key question that will be addressed through various statistical models and data analysis tools is to understand the a) nature of marginal dependence between a pair of channels $X_p(t)$ and $X_q(t)$ or b) between $X_p(t)$ and $X_q(t)$ conditional on the other components in the data.

This work is largely motivated from a neuroscience perspective. The brain circuitry can be conceived as the integration of a sensory, motor, and cognitive system that receives, processes, and reacts to external impulses or internal autoregulation activities (Swanson, 2003, pp. 80-91). Communication and feedback between those structures enable brain functions. For instance, memory is believed to rely on the hippocampus because it is at the center in the signal flow from and to the cortical areas, such as the orbitofrontal cortex, olfactory bulb, and superior temporal gyrus (Witter et al., 2017; McGaugh et al., 1990, p. 160-165). From a macro perspective, those memory flows also imply a high activity in the temporal brain region (McGaugh et al., 1990, p. 164). From an analytic perspective, these signal interactions can be studied as *undirected* dependencies (functional connectivity) $X_p(t) \Leftrightarrow X_q(t)$; or *directed*, or *causal*, networks (effective connectivity) $X_p(t) \Rightarrow X_q(t) \wedge X_q(t) \Rightarrow X_p(t)$. Additionally, current imaging techniques allow us to understand those interactions at different biological levels: (a.) through hemoglobin changes (energy consumption) using functional magnetic resonance imaging (fMRI) or functional near-infrared spectroscopy (fNIRS); (b.) through measurements of the electrical activity at the scalp, (electroencephalogram, EEG), at the cortical surface (electrocorticogram, ECoG), or in the extracellular environment (local field potentials, LFPs).

We first note that there are several models that aim to represent the different characterizations of dependence (Fig. 2) The most common measure of dependence is cross-covariance (or cross-correlation). In the simple case where $\mathbb{E}[X_p(t)] = 0$ and $\text{Var}[X_p(t)] = 1$ for all $p = 1, \ldots, P$ and all time t, then the cross-correlation (at lag 0) between X_p and X_q is

$$\rho_{pq}(0) = \mathbb{E}[X_p(t)X_q(t)] = \int u_p u_q f_{pq}(u_p, u_q)\, du_p du_q \tag{1}$$

where $f_{pq}(u_p, u_q)$ is the joint probability density function of $(X_p(t), X_q(t))$ and the integral is evaluated globally over the entire support of $(X_p(t), X_q(t))$. Cross-correlation provides a simple metric that measures the linear synchrony between a pair of components across the entire support of their joint distribution. When $|\rho_{pq}(0)|$ has a value close to 1, we conclude that there is a strong linear dependence between X_p and X_q. It is obvious that cross-correlation does not completely describe the nature of dependence between X_p and X_q. First, dependence goes beyond linear associations, and hence this paper addresses some types of non-linear dependence between components. Second, dependence may vary across the entire support. That is, the association between X_p and X_q at the "bulk" of the distribution may be different from its "tails" (e.g., extrema). Third, since time series data can be viewed as superpositions of sine and cosine waveforms with random ampli-

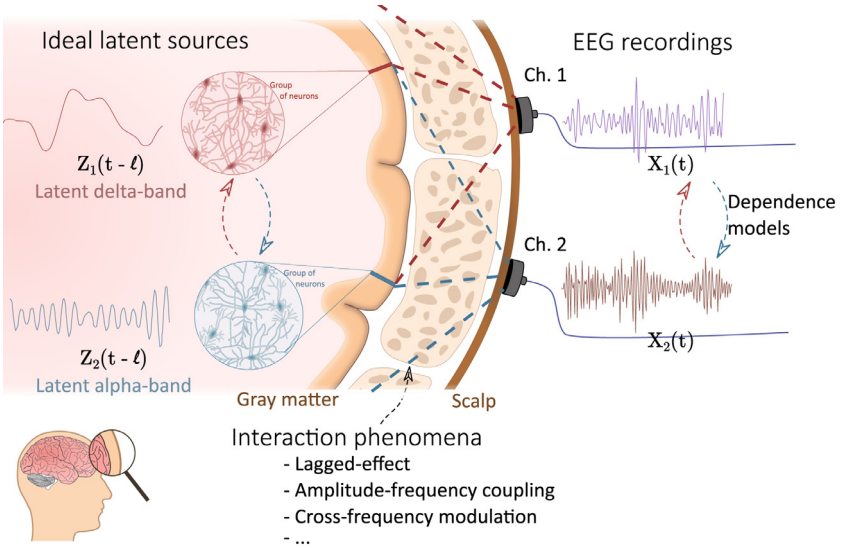

Fig. 2. Abstraction of dependence in brain signals. In this model, the hypothetical latent sources are assumed to generate neural oscillations (with their own frequency characteristics) that can interact among them before being recorded at the scalp. Different abstractions that model the sources' interactions and the mixing models are explained in this paper.

tudes, it is natural to seek to identify the specific oscillations that drive the linear relationship. For example, Guerrero et al. Guerrero et al. (2021) proposed a method for investigating the dependence observed by the impact of extreme behavior in specific frequency band on the brain network structure. In particular, some types of non-linear dependence in time domain can have spectral linear relationships. These types of models are the main focus of the models and methods covered in this paper.

To consider spectral measures of dependence, that is, dependence between oscillatory components, the starting point will be the Cramér representation of stationary time series. Under stationarity, both X_p and X_q can be decomposed into oscillations at various frequency bands. The key elements of the Cramér representation are the Fourier basis waveforms $\left\{\exp(i2\pi\omega t), \omega \in (-\frac{1}{2}, \frac{1}{2})\right\}$ and the associated random coefficients $\{dZ(\omega)\}$ which is an orthogonal increment random process (Priestley, 1981, pp. 244-246), (Li, 2014, pp. 20-21), (Thomson, 1982) that satisfies $\mathbb{E}[dZ(\omega)] = 0$ and $\mathbb{C}\text{ov}(dZ(\omega), dZ(\omega')) \neq 0$ for $\omega \neq \omega'$. The Cramér representation for $(X_p(t), X_q(t))$ is given by

$$\begin{pmatrix} X_p(t) \\ X_q(t) \end{pmatrix} = \begin{pmatrix} \int_{-0.5}^{0.5} \exp(i2\pi\omega t)\, dZ_p(\omega) \\ \int_{-0.5}^{0.5} \exp(i2\pi\omega t)\, dZ_q(\omega) \end{pmatrix}. \tag{2}$$

The Cramér representation above can be used to characterize the nature of synchrony between the ω-oscillation in X_p and the ω-oscillation in X_q which are, respectively, $\exp(i2\pi\omega t)dZ_p(\omega)$ and $\exp(i2\pi\omega t)dZ_q(\omega)$. This idea will be further developed in the next sections of this paper. Note here that the quasi-stationarity assumption is reasonable over brief segments of both resting-state and task conditions (Ombao et al., 2017, p. 188) or within reasonable short time intervals of EEGs (Sanei and Chambers, 2007, p. 20).

Most brain signals exhibit non-stationary behavior, which may be reflected changes in either (a.) the mean level, or (b.) the variance at some channels, or (c.) the cross-covariance structure or dependence between some pairs of channels. Note here that (a.) is a condition on the first moment while (b.) and (c.) are conditions on the second moment. Moreover, (a.) and (b.) are properties within a channel, while (c.) is a property that describes dependence between a pair of channels. There is no singular measure that can completely describe the *nature of dependence* between channels. The most common pair of measures consists of the cross-covariance and cross-correlation, whose equivalent measures in the frequency domain are the cross-spectrum and cross-coherence, respectively. This paper will focus on the frequency domain measures and thus dig deeper into being able to identify the oscillations that drive the dependence between a pair (or group) of channels. Under non-stationarity, the dependence structure can change over time. Our approach here is to slide a localized window across time and estimate the spectral properties within each window. This approach is proposed in (Priestley, 1965) and then reformulated in (Dahlhaus, 2012) to establish an asymptotic framework for demonstrating the theoretical properties of the estimators.

2. EEG spectral characteristics

To illustrate the various spectral dependence measures, we shall focus on the analysis of EEG signals. EEG is a noninvasive technique that collects electrical potential at the scalp gathered from synchronized responses of groups of neurons ("signal

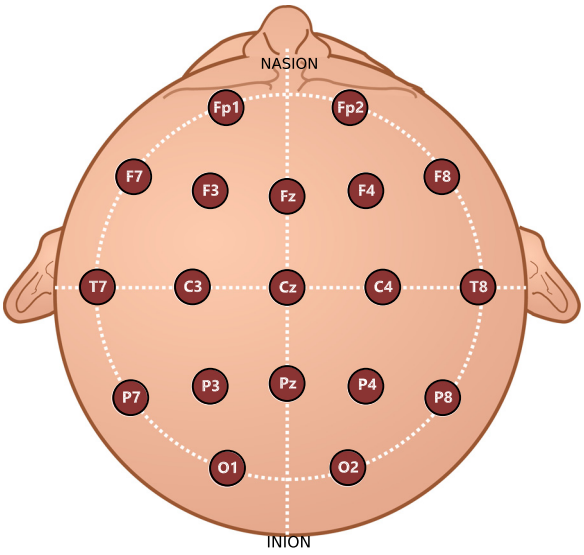

Fig. 3. Electrode positioning in the 10-20 standard layout.

sources") (Tatum, 2008, p. 4) (Ombao et al., 2017, p. 555) that are perpendicular to the scalp and dynamically organized in neighborhoods with scales of a few centimeters (Stern, 2013, p. 5). Naturally, EEG is affected by the volume conduction of the signals over tissue and skull (Nunez and Srinivasan, 2006, p. 144). Despite its limitations, the portability and inexpensiveness allow the integration of EEG in clinical settings and cognitive experiments that require naturalistic environments (in contrast, fMRI experiments require the participants to be in a supine position in a restricted space). In addition, the high temporal resolution of EEGs enables to capture the temporal dynamics of the neuronal activity. Therefore, through spatial, temporal, and morphological EEG patterns, some neurological and mental conditions can be diagnosed (Stern, 2013, p.3, 19-22). For instance, a specific metric obtained from EEG frequency properties measured on the channel Cz is considered an FDA-approved clinical method to assess attention deficit hyperactivity disorder (ADHD) (Snyder et al., 2015). Thus, EEGs have been extensively used in cognitive neuroscience, neurology, and psychiatry to study the neurophysiological basis of cognition and neuropsychiatric disorders: motor abilities (Wu et al., 2014a), anesthetic similarities with comma (Brown et al., 2010), encephalopathy (Jacob et al., 2019), schizophrenia (Steinmann et al., 2020), addictions (Newson and Thiagarajan, 2019), autism spectrum disorder (Newson and Thiagarajan, 2019; Ibrahim et al., 2018), depression (Mahato and Paul, 2019; Liao et al., 2017; Mantri et al., 2015), and ADHD (Snyder et al., 2015; Mohammadi et al., 2016). Here, we will illustrate some methods that characterize the dynamics of the inter-relationships between the activity measured at different channels.

In this paper, we use the EEG dataset related to a mental health disorder collected by Nasrabadi et al. (Moti Nasrabadi et al., 2020). This data consists of EEGs, sampled at 128Hz, from 60 children with attention deficit hyperactivity disorder (ADHD) and 60 children with no registered psychiatric disorder as a control group. These electrical recordings were collected from 19 channels evenly distributed on the head in the 10-20 standard layout (Fig. 3) (Oostenveld and Praamstra, 2001). Average recording from both ear lobes (A1 and A2) was used as electrode references. The experiment was intended to show potential differences in the brain response under visual attention task (Moti Nasrabadi et al., 2020; Allahverdy et al., 2011; Mohammadi et al., 2016). Therefore, the 120 participants were presented series of images and the task was to count the number of objects in each image. The number of objects in a set ranged from 5 to 16 - which is reasonable size for the children participants. Each collection of pictures was displayed without interruptions in order to prevent distraction from the subjects.

Prior to analysis, some standard pre-processing steps were conducted. These procedures are not unique to EEGs, or brain signals, as most data often need some cleaning before statistical modeling. Several attempts have been performed to standardize those pre-processing steps (Bigdely-Shamlo et al., 2015; Jas et al., 2017; Pedroni et al., 2019; Debnath et al., 2020; Desjardins et al., 2021). In general, pre-processing aims to increase the quality of the recorded signal by (Robbins et al., 2020): (a.) removing the effect of the electrical line (electrical interference at 50Hz or 60Hz due to the electrical source); (b.) removing artifacts due to eye movements, eye blinks or muscular movements; (c.) detecting, removing or repairing bad quality channels; (d.) filter non-relevant signal components; and (e.) re-referencing the signal to improve topographical localization. In this scenario, filtering is a crucial step that removes components of the signal that could not be related to cognitive or physiological processes, such as extremely high-frequency components. In this dataset, we applied a band-pass filter on the frequency interval (0.50-70.00) Hz and segmented the signal into the main "brain rhythms" (Ombao et al., 2017, p.610Bsemi Stern, 2013, p. 33-34,169-170,413-414Bsemi Nunez and Srinivasan, 2006, p. 12): frequency range into the delta band: (0.5, 4.0) Hertz, theta band: (4.0, 8.0) Hz, alpha band: (8.0, 12.0) Hertz, beta band: (12.0, 30.0) Hertz, and gamma band: (30.0, 50.0) Hertz. A non-causal second-order Butterworth filter was used to perform this frequency filtering: Fig. 4

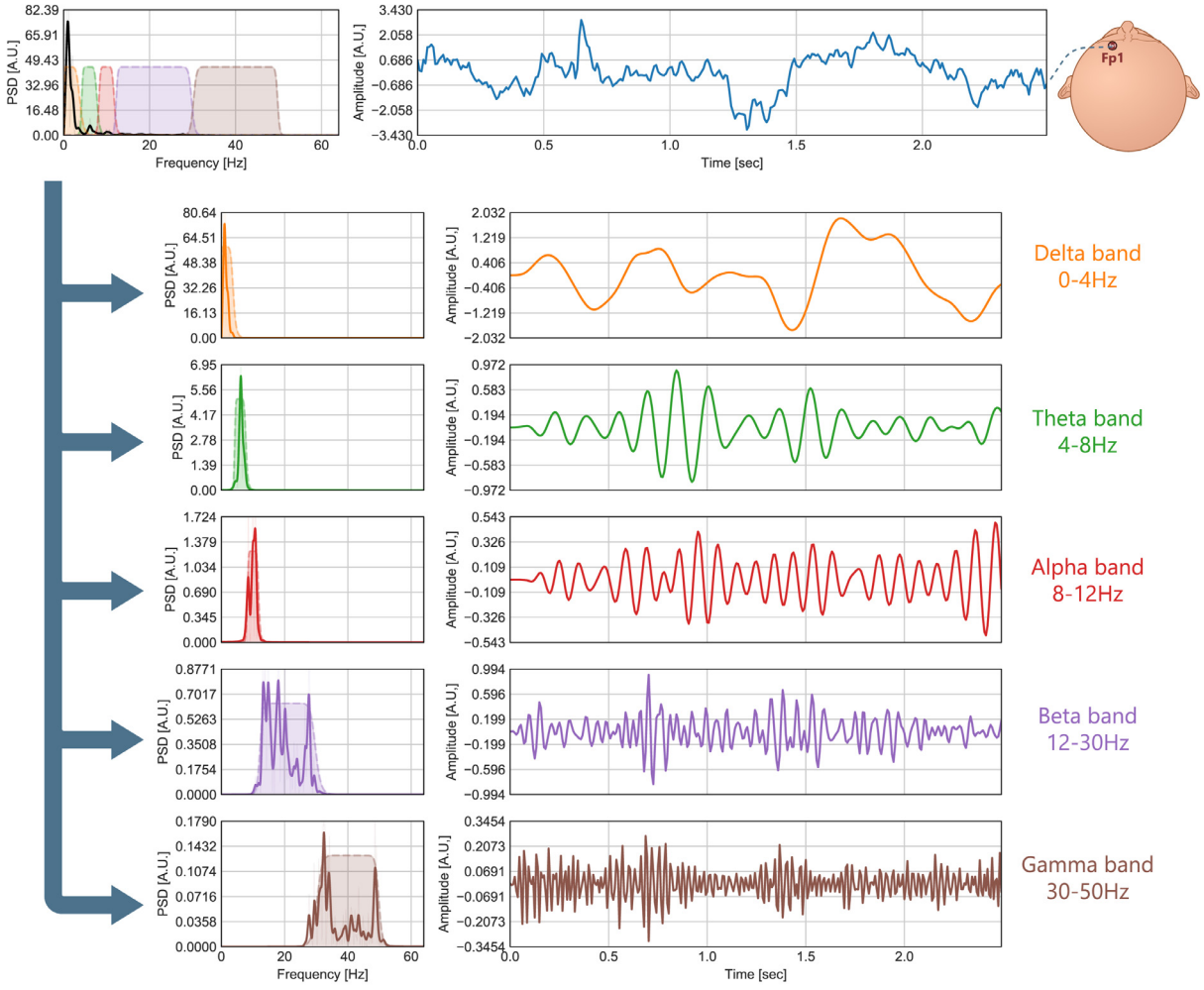

Fig. 4. Frequency decomposition of an EEG, collected at channel Fp1, into the five brain rhythms: delta, theta, alpha, beta and gamma.

shows an example of this decomposition where the shadow regions show the magnitude response of the applied filter. Finally, the EEGs are also often segmented into epochs of sixty seconds.

3. Coherence and Partial Coherence

In this section, a formal description of the two most common measures of spectral dependence, namely, coherence and partial coherence will be derived under the context of the Cramér representation of weakly stationary processes. One of the contributions of this paper is to provide an intuitive framework (i.e., via linear filtering) under which these common metrics as well as more complex dependence measures can be derived. Finally, these measures will be derived for the general case where the signals exhibit non-stationarity, e.g., under conditions where the dependence between signals may evolve over time.

3.1. Coherence and Correlation via the Fourier transform

Suppose that $\mathbf{X(t)}$ is a $P-$dimensional weakly stationary process with mean $\mathbb{E}\mathbf{X}(t) = \underline{0}$ for all time t and sequence of covariance matrices $\{\Sigma(h),\ h = 0, \pm1, \pm2, \ldots\}$ where $\Sigma(h)$ is defined by

$$\Sigma(h) = \mathbb{C}\text{ov}\left(\mathbf{X}(t+h), \mathbf{X}(t)\right) = \mathbb{E}\Big(\mathbf{X}(t+h)\mathbf{X}'(t)\Big).$$

Here,every element of $\Sigma(h)$ will be assumed to be absolute summability (over lag h): let $\sigma_{pq}(h)$ be the (p,q)-th entry of the matrix $\Sigma(h)$, then $\sum_{h=-\infty}^{\infty} |\sigma_{pq}(u)| < \infty$ for all (p,q). This condition ensures that auto-correlation and cross-correlation decay to 0 sufficiently fast.

There is a one-to-one relationship between the covariance matrix $\Sigma(h)$ and the spectral matrix via the Fourier transform. The spectral matrix, denoted $\mathbf{f}(\omega)$, is defined as

$$\mathbf{f}(\omega) = \sum_{h=-\infty}^{\infty} \Sigma(h) \exp(-i2\pi\, \omega\, h) \quad \text{where } \omega \in \Omega_0 = (-0.5, 0.5).$$

$\mathbf{f}(\omega)$ is a $P \times P$ Hermitian semi-positive definite matrix.

To analyze the properties of a single component of the multivariate time series, one should focus on the diagonal elements of the spectral matrix $\mathbf{f}(\omega)$. In fact, the auto-spectrum (univariate spectrum) of the p-th channel, denoted as $f_{pp}(\omega)$, is the (p, p)-th entry of $\mathbf{f}(\omega)$. In addition, the autocovariance sequence $\sigma_{pp}(h)$ can be derived from auto-spectrum $f_{pp}(\omega)$ via the inverse-Fourier transform

$$\sigma_{pp}(h) = \int_{\omega\in\Omega_0} f_{pp}(\omega)\, \exp\left(i2\pi\, \omega, h\right) d\omega.$$

One very important case is when $h = 0$, i.e., $\sigma_{pp}(0) = \mathrm{Var}\big(X_p(t)\big)$. This corollary provides the intuition that the auto-spectrum is a decomposition of the signal variance across all frequencies $\omega \in \Omega_0 = (-0.5, 0.5)$:

$$\sigma_{pp}(0) = \int_{\omega\in\Omega_0} f_{pp}(\omega)\, d\omega = \mathrm{Var}(X_p(t)). \tag{3}$$

We now introduce correlation as a dependence metric. Correlation between two components X_p and X_q at time lag h is defined by

$$r_{p,q}(h) = \mathbb{C}\mathrm{or}\left(X_p(t+h), X_q(t)\right) = \frac{\mathbb{C}\mathrm{ov}\left(X_p(t+h), X_q(t)\right)}{\sqrt{\mathrm{Var}X_p(t)\,\mathrm{Var}X_q(t)}} = \frac{\sigma_{pq}(h)}{\sqrt{\sigma_{pp}(0)\,\sigma_{qq}(0)}}.$$

It is clear that $r_{p,q}(h) \in [-1, 1]$ will reach its extreme value when one of the signals is perfectly linearly related to the other (one time series is a scaled and/or shifted version of the other). Thus, correlation $r_{p,q}(h)$ is known for being the simplest measure the quantifies the linear dependence, or synchrony, at a lag h between the pair of time series.

The spectral matrix $\mathbf{f}(\omega)$ provides more information about the interactions of their components. For instance, it allows us to identify the cross-spectrum (at frequency ω) between any pair of components, X_p and X_q, through its (p, q)-th entry. In a similar manner to the correlation in the time domain, we can define another measure that quantifies the similarity between the simultaneous spectral response of X_p and X_q. This metric is known as coherency which is formally defined as

$$\tau_{pq}(\omega) = \frac{f_{pq}(\omega)}{\sqrt{f_{pp}(\omega) f_{qq}(\omega)}}. \tag{4}$$

Coherency is complex-valued whose magnitude lies in [0,1]. However, it is more common to use cross-coherence which is the square magnitude of the coherency:

$$\rho_{pq}(\omega) = |\tau_{pq}(\omega)|^2 = \frac{|f_{pq}(\omega)|^2}{f_{pp}(\omega) f_{qq}(\omega)}. \tag{5}$$

3.2. Coherence and Correlation under the Cramér representation

We now develop an intuitive framework under which we can derive coherency and coherence. First, recall that the Cramér representation of a zero-mean $P-$dimensional weakly stationary process $\mathbf{X}(\mathbf{t})$ is given by

$$\mathbf{X}(t) = \int_{-0.5}^{0.5} \exp\left(i2\pi\, \omega\, t\right) d\mathbf{Z}(\omega)$$

where $d\mathbf{Z}(\omega) = [dZ_1(\omega), \ldots, dZ_P(\omega)]'$ is a vector of random coefficients associated with the Fourier waveform $\exp(i2\pi\omega t)$ for each of the P components. The theoretical aspects of this stochastic integration, we refer to (Brockwell and Davis, 1996, Section 11.2) Here, $\{d\mathbf{Z}(\omega)\}$ is a random process defined on $\omega \in (-0.5, 0.5)$ that satisfies $\mathbb{E}[d\mathbf{Z}(\omega)] = 0$ for all frequencies ω and

$$\mathbb{C}\mathrm{ov}\left(d\mathbf{Z}(\omega), d\mathbf{Z}(\omega')\right) = \begin{cases} 0, & \omega \neq \omega' \\ \mathbf{f}(\omega) d\omega, & \omega = \omega' \end{cases}$$

The estimation of the second moment of $dZ(\omega)$ from a finite sequence is discussed in detail in (Thomson, 1982). From the above formulation, we note that the correlation and modulus-squared correlation with the coefficients are, in fact, coherency and coherence as they were defined in Eqs. 4 and 5:

$$\mathbb{C}\mathrm{or}\left[d\mathbf{Z}_p(\omega), d\mathbf{Z}_q(\omega)\right] = \frac{f_{pq}(\omega)}{\sqrt{f_{pp}(\omega) f_{qq}(\omega}}$$

$$\|\mathbb{C}\text{or}\,[d\mathbf{Z}_p(\omega), d\mathbf{Z}_q(\omega)]\|^2 = \frac{|f_{pq}(\omega)|^2}{f_{pp}(\omega) f_{qq}(\omega)}.$$

Note that the Cramér representation of a time series (i.e., it is a linear mixture of infinitely many sinusoidal waveforms or Fourier waveforms with random amplitudes), provides an interesting perspective on the interpretation of the above-mentioned dependence metrics. Consider the components X_p and X_q, both of them contain Fourier oscillations in a continuum of frequencies, and let us now focus only on the specific frequency ω in the two components:

$$X_{p,\omega}(t) = \exp(i2\pi\omega t)\, dZ_p(\omega)$$
$$X_{q,\omega}(t) = \exp(i2\pi\omega t)\, dZ_q(\omega).$$

Since these random coefficients have zero mean, the variance at the ω-oscillatory activity of X_p is

$$\text{Var}[X_{p,\omega}(t)] = |\exp(i2\pi\,\omega\, t)|^2 \text{Var}[dZ_p(\omega)].$$

Since the random coefficients $\{dZ_p(\omega)\}$ are uncorrelated across $\omega \in (-0.5, 0.5)$, it follows that

$$\text{Var}[X_p(t)] = \int \text{Var}[X_{p,\omega}(t)] = \int \text{Var}[dZ_p(\omega)].$$

Based on this relationship, we can introduce an alternative interpretation of the spectral decomposition denoted along with Eq. 3: the total variance of a weakly stationary signal X_p at any time point can be viewed as an infinite sum of the variance of each of the random coefficients. Furthermore, we can think of the relationship between the variance of the random coefficients and the spectrum as follows

$$\text{Var}[dZ_p(\omega)] = f_{pp}(\omega)\, d\omega.$$

We now study the relationship between the signals X_p and X_q as a function of the oscillations. The covariance between the $\omega-$oscillation in X_p and the $\omega-$oscillation in X_q is

$$\mathbb{C}\text{ov}\,[X_{p,\omega}(t), X_{p,\omega}(s)] = \exp(i2\pi\omega(t-s))\, \mathbb{C}\text{ov}\,[dZ_p(\omega), dZ_q(\omega)].$$

Moreover, since $|\exp(i2\pi\omega t)| = 1$ for all t, the variances of these respective ω-oscillations are

$$\text{Var}[X_{p,\omega}(t)] = \text{Var}[dZ_p(\omega)] \quad \text{and} \quad \text{Var}[X_{q,\omega}(t)] = \text{Var}[dZ_q(\omega)].$$

In addition, the correlation between these ω-oscillations is

$$\mathbb{C}\text{or}\,[X_{p,\omega}(t), X_{p,\omega}(s)] = \exp(i2\pi\omega(t-s))\, \mathbb{C}\text{or}\,[dZ_p(\omega), dZ_q(\omega)]$$

and the respective modulus-squared correlation is given by

$$|\mathbb{C}\text{or}\,[X_{p,\omega}(t), X_{p,\omega}(s)]|^2 = |\mathbb{C}\text{or}\,[dZ_p(\omega), dZ_q(\omega)]|^2$$

which is identical to the definition of coherence given in Eq. 5.

In summary, it is possible to affirm that coherence between a pair of weakly stationary signals at frequency ω is the square-magnitude of the correlation between the ω-oscillations of these signals.

3.3. Coherence and filtering

In practical EEG analysis, coherence between a pair of channels is defined and estimated at some frequency bands - rather than in a singleton frequency. The standard frequency bands are delta $(0.5, 4.0)$ Hertz, theta $(4.0, 8.0)$ Hertz, alpha $(8.0, 12.0)$ Hertz, beta $(12.0, 30.0)$ Hertz and gamma $(30.0, 50.0)$ Hertz. This segmentation of the frequency axis has been widely accepted for many decades. However, there is a growing direction towards a more specific (narrower) frequency band analysis and a more data-adaptive approach to determining (a.) the number of frequency peaks, (b.) the location of these peaks, and (c.) the bandwidth associated with each one of them. This will be necessary for a finer differentiation between experimental conditions and patient diagnosis groups. The immediate task at hand is to point out the connection between linear filters and spectral and coherence estimation.

To estimate coherence, the first step is to apply a linear filter on time series components X_p and X_q so that the resulting filtered time series will have spectra whose power is concentrated around a pre-specified band Ω. In essence, a k-th order linear frequency-filter is comprised of a set of coefficients $\{a_0, \ldots, a_K\}$ and $\{c_1, \ldots, c_K\}$ under the constraint $\sum_k |a_k| < \infty$ and $\sum_k |c_k| < \infty$. The filtered signal $y^*(t)$ is obtained as a linear combination of the previous values of the unfiltered time series $y(t), \ldots, y(t-K)$ and its previous values $y^*(t-1), \ldots, y^*(t-K)$:

$$y^*(t) = \sum_{k=0}^{M} c_k y(t-k) - \sum_{k=1}^{M} a_k y^*(t-k) \tag{6}$$

Filters with $a_k = 0\,\forall k$ are known as finite impulse response (FIR) filters or infinite impulse response (IIR) filters when $\exists a_k \neq 0$ for some k. For an extensive analysis of linear time-invariant filtering, we refer to (Hayes, 1996). Even though for

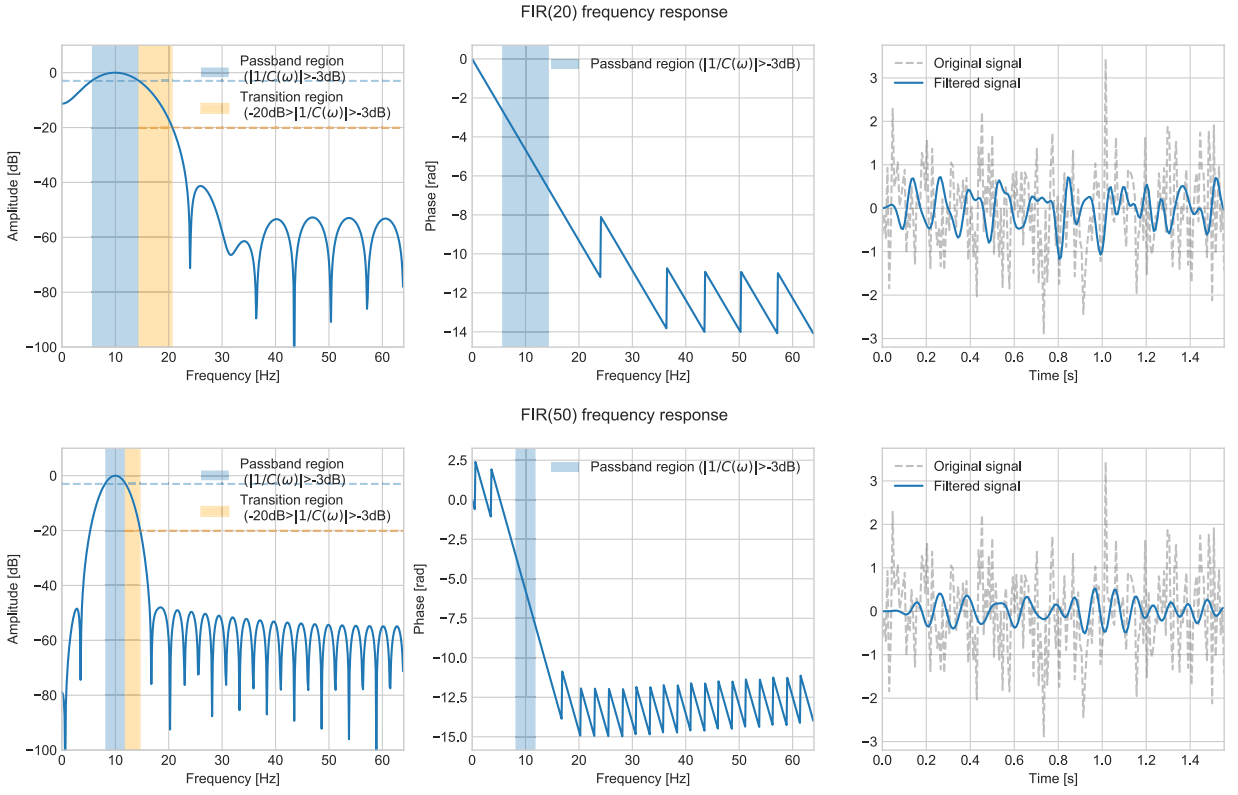

Fig. 5. Sample filter response in magnitude and phase for a 20th-order and a 50th-order FIR filter.

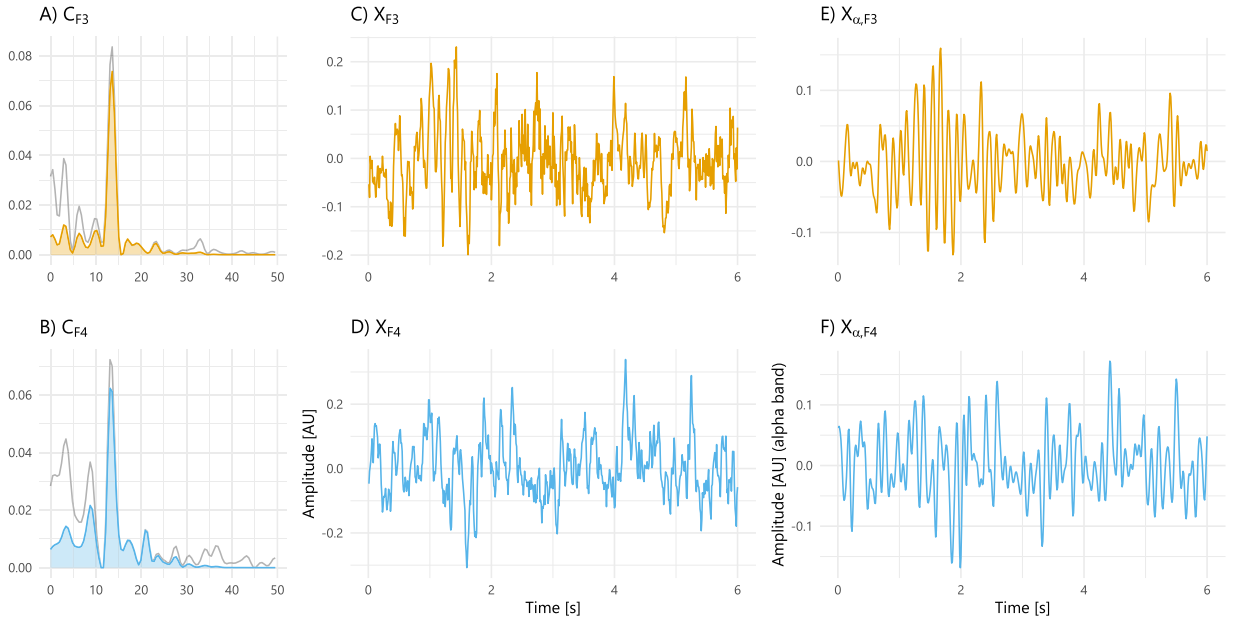

Fig. 6. Alpha-filtering of EEG signals in channels F3 and F4: A,B) Power spectrum density C_{F4} and C_{F3} after applying a band-pass filter. Note that the power is concentrated only in the interval 8-12Hz. The gray line in the background corresponds to the unfiltered spectra. C,D) Signals X_{F3} and X_{F4} without before filtering; and E.F) Alpha waves in both channels: $X_{F3,\alpha}$ and $X_{F4,\alpha}$.

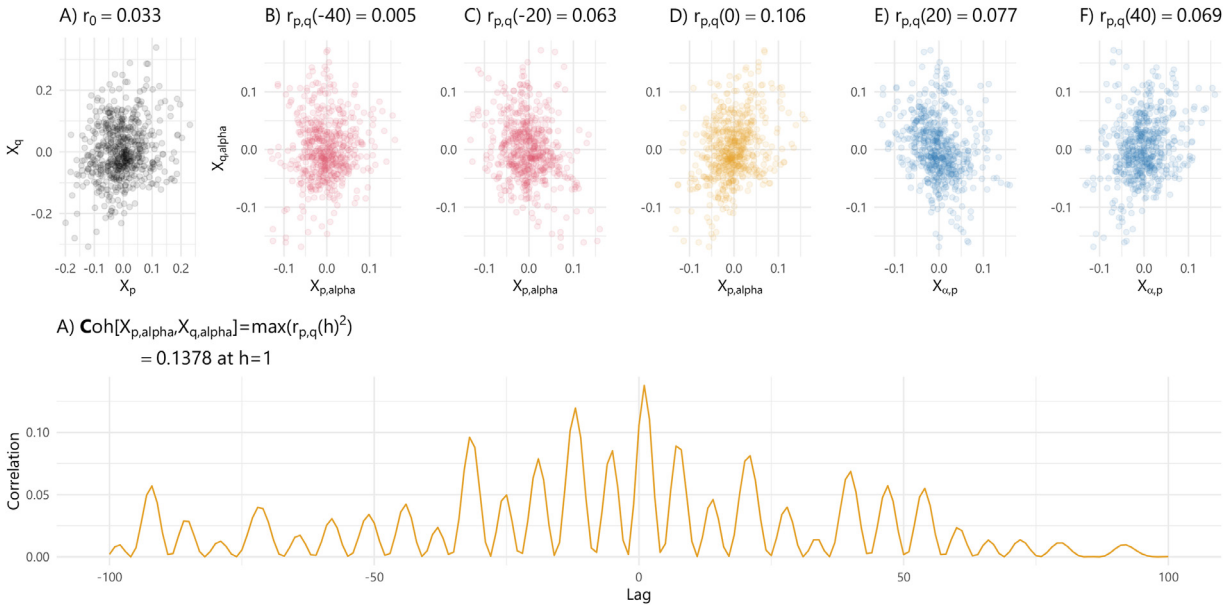

Fig. 7. Linear relationships between the channels F3 and F4: A) Scatterplot of the unfiltered signals, with a correlation of 0.033. B-F) Scatterplots of alpha-filtered channels at different lags: -40, -20, 0, 20, and 40. Note that lags are denoted from reference to F3 (X_p). Thus, negative lags imply that F4 is leading F3 during the correlation estimation. G) Cross-correlation as a function of the lags. The maximum cross-correlation (0.138) is located at lag 1.

the biomedical purposes discussed in this paper, this frequency bandwidth is the most relevant filtering feature. We also need inherent frequency-domain properties associated with the selected set of coefficients, such as the frequency transient response, passband ripples, and phase response (Fig. 5). Ripples in the bandpass frequency interval are small magnitude oscillations that could appear in the desired bandpass. Transient response is the type of attenuation that the filter can provide for "intermediate frequencies" between the bandstop region and the bandpass. Faster transient responses are typically linked with higher ripples, and therefore, higher distortion in the time domain. In addition, the phase response is the frequency-associated delay that the filter naturally induces in the output filtered signal. Linear responses are associated with lower distortion and are ensured for FIR filters. We refer to (Widmann et al., 2015) for a more detailed discussion of the strategies for designing filters in biomedical signals. In this paper, we will focus on the FIR family of filters.

Consider the filter so that $\{c_k, k = 0, \pm1, \pm2, \ldots\}$. A one-sided linear filter will be used to examine causality between the different oscillations, a further discussion about the reasons behind this condition is given in Section 5.4. The Fourier transform of the sequence of FIR filter coefficients $\{c_j\}$ is $C(\omega) = \sum_k c_k \exp(-i2\pi\omega k)$ The set of filter coefficients are selected so that $|C(\omega)|$ has a peak that is concentrated in the neighborhood of the frequency band Ω (band-pass region). In the frequency intervals outside of Ω (stop-band region), $|C(\omega)|$ is expected to has relatively small values. Thus, to extract the component of X_p and X_q that is associated with the Ω-oscillation, we apply the linear filter $\{c_j, j = 0, \pm1, \pm2, \ldots\}$ to obtain the convolution

$$X_{p,\Omega}(t) = \sum_{j=-\infty}^{\infty} c_j X_p(t-j) \quad \text{and} \quad X_{q,\Omega}(t) = \sum_{j=-\infty}^{\infty} c_j X_q(t-j).$$

The auto-spectra of each filtered series, $\{X_{p,\Omega}\}$ and $\{X_{q,\Omega}\}$ are

$$f_{pp,\Omega}(\omega) = |C(\omega)|^2 \, f_{pp}(\omega) \quad \text{and} \quad f_{qq,\Omega}(\omega) = |C(\omega)|^2 \, f_{qq}(\omega),$$

respectively, which has spectral power attenuated outside of the band Ω.

Example 1 (EEG example)**.** Recall the EEG-ADHD dataset described in Section 2. Let us focus in the signals recorded at channels F3 and F4, collected from the control subject S041. We denoted them by X_p and X_q (Fig. 7.C-D). These biomedical signals are sampled at 128 Hertz and with a 60-second epoch. In consequence, the total number of time points available for the analysis is $T = 128 \times 60 = 7680$. We extract the alpha-band component ($\Omega = 8 - 12$Hertz) using a 10-th order FIR band-pass filter defined by the coefficients:

$$\{c_i\} = \{-0.0272, -0.0468, -0.0423, 0.0771, 0.2677, \\ 0.3629, 0.2677, 0.0771, -0.0423, -0.0468, -0.0272\}$$

This 10-order filter was chosen as a compromise between filter delay and attenuation. The filtering results are depicted in Fig. 7.E-F with the spectrum after the filtering process in Fig. 7.A-B.

The next step is to study the dependence between the time series X_p and X_q through their filtered components. In practice, dependence can be examined at various frequency bands but in this example, we focus only on the alpha band for illustration purposes. As noted by Eq. 5, coherence is well defined for a single frequency ω. Nevertheless, for a frequency set Ω, we present two approaches to estimation:

$$\widehat{\rho}^{(A)}_{p,q}(\Omega) = \frac{1}{\|\Omega\|}\sum_{f\in\Omega}\frac{|P_{xy}(f)|^2}{|P_{xx}(f)||P_{yy}(f)|},$$

and

$$\widehat{\rho}^{(B)}_{p,q}(\Omega) = \frac{\sum_{f\in\Omega}|P_{xy}(f)|^2}{\sum_{f\in\Omega}|P_{xx}(f)|\sum_{f\in\Omega}|P_{yy}(f)|}.$$

This ambiguity was also raised and discussed by Kulaichev (2011). Thatcher et al., Thatcher et al. (1986, 2008); Thatcher (2011), also established the procedure of calculating coherence as "first computing the power spectra" over all frequencies, and later, "computing the normalized cross-spectra" as in $\rho^{(B)}_{p,q}(\Omega)$ in our notation. In this manuscript, we rely on this procedure. Note also that, from Ombao et al. (2006), coherence at the alpha-band is estimated to be

$$\widehat{\rho}_{p,q}(\alpha) \propto \max_{\ell}\left|\sum_{t} X_{p,\alpha}(t)\, X_{q,\alpha}(t-\ell)\right|^2 \tag{7}$$

where $X_{p,\alpha}(t)$ is the α-oscillatory activity in channel X_p.

Example 2 (Contemporaneous mixture)**.** We now illustrate that coherence gives information beyond cross-correlation (or the square cross-correlation). Consider the setting where there are two latent sources Z_1 and Z_2 where Z_1 is a high-frequency source and Z_2 is a low-frequency source. Define the observed time series X_1 and X_2 to be mixtures of these two sources, e.g.,

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} Z_1(t) \\ Z_2(t) \end{pmatrix} + \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{pmatrix}. \tag{8}$$

Suppose that $c_{21} = 0$ and the other entries of the mixing matrix are all non-zero. This implies that X_1 contains both low-frequency component Z_1 and high-frequency component Z_2. However, X_2 contains only the high-frequency component Z_2. Thus, it is the high-frequency component that drives the dependence between X_1 and X_2. The scatterplot of $X_2(t)$ vs $X_1(t)$ in Fig. 8.E shows that these two time series are correlated and the sample cross-correlation is computed to be 0.4776. However, correlation is limited in the information it can convey about the relationship between a pair of signals. For instance, it does not indicate what frequency band(s) drive that relationship.

To now investigate deeper the relationship between X_1 and X_2, we apply a low-pass filter (on band Ω_0) and a high-pass filter (on band Ω_1) and denote these filtered signals to be

$$X_{1,\Omega_0}(t),\ X_{1,\Omega_1}(t),\ X_{2,\Omega_0}(t),\ X_{2,\Omega_1}(t).$$

Under stationarity, the random coefficients in the Cramér representation are uncorrelated across frequencies (i.e., $\mathbb{C}\text{ov}[dZ_1(\omega), dZ_2(\omega')] = 0$ when $\omega \neq \omega'$). Thus,

$$\mathbb{C}\text{ov}[X_{1,\Omega_0}(s), X_{2,\Omega_1}(t)] = 0 \quad \text{and} \quad \mathbb{C}\text{ov}[X_{1,\Omega_1}(s), X_{2,\Omega_0}(t)] = 0.$$

For some non-stationary processes, there could be possible linear and non-linear dependence between different frequency bands. For the moment, we focus only on the correlation between the low-frequency components $X_{1,\Omega_0}(s)$ and $X_{2,\Omega_0}(t)$ and the high-frequency components $X_{1,\Omega_1}(s)$ and $X_{2,\Omega_1}(t)$. The lag-0 scatterplots are shown in 8.F-I. It is clear here that the linear relationship between the low-frequency components is weaker than the dependence on high-frequency components. This was to be expected from the data generating model in Eq. 2, which specifies that X_1 and X_2 both share the common high-frequency latent source. Assuming that the sample mean of these oscillations are all 0, then the coherence estimate at the low-frequency band is the squared cross-correlation between the X_{1,Ω_0} and X_{2,Ω_0}, i.e.,

$$\widehat{\rho}_{12}(\Omega_0) = \frac{\left\{\sum_{t=1}^{T}(X_{1,\Omega_0}(t)X_{2,\Omega_0}(t))\right\}^2}{\sum_{t=1}^{T}(X_{1,\Omega_0}(t))^2\sum_{t=1}^{T}(X_{2,\Omega_0}(t))^2}.$$

Note that due to the symmetric property of cross-correlation between X_{1,Ω_0} and X_{2,Ω_0}, coherence $\widehat{\rho}_{12}(\Omega_0)$ is also symmetric. The coherence estimate at the high-frequency band is computed similarly. The estimated values for coherence at the low and high-frequency bands are, respectively, 0.0608 and 0.8583.

Example 3 (Lagged mixture)**.** The data generating process in the previous example (Eq. 2) assumes an instantaneous mixture, i.e., the observed signals at a specific time t, $\{X_1(t), X_2(t)\}$ depend explicitly on the latent processes $Z_1(t)$ and $Z_2(t)\}$ also at the *same* time t. However, we can extend this model towards cases when there is some lag in the mixtures, e.g., the latent source has a delayed effect on some components of the observed signals. As Nunez et al. pointed out

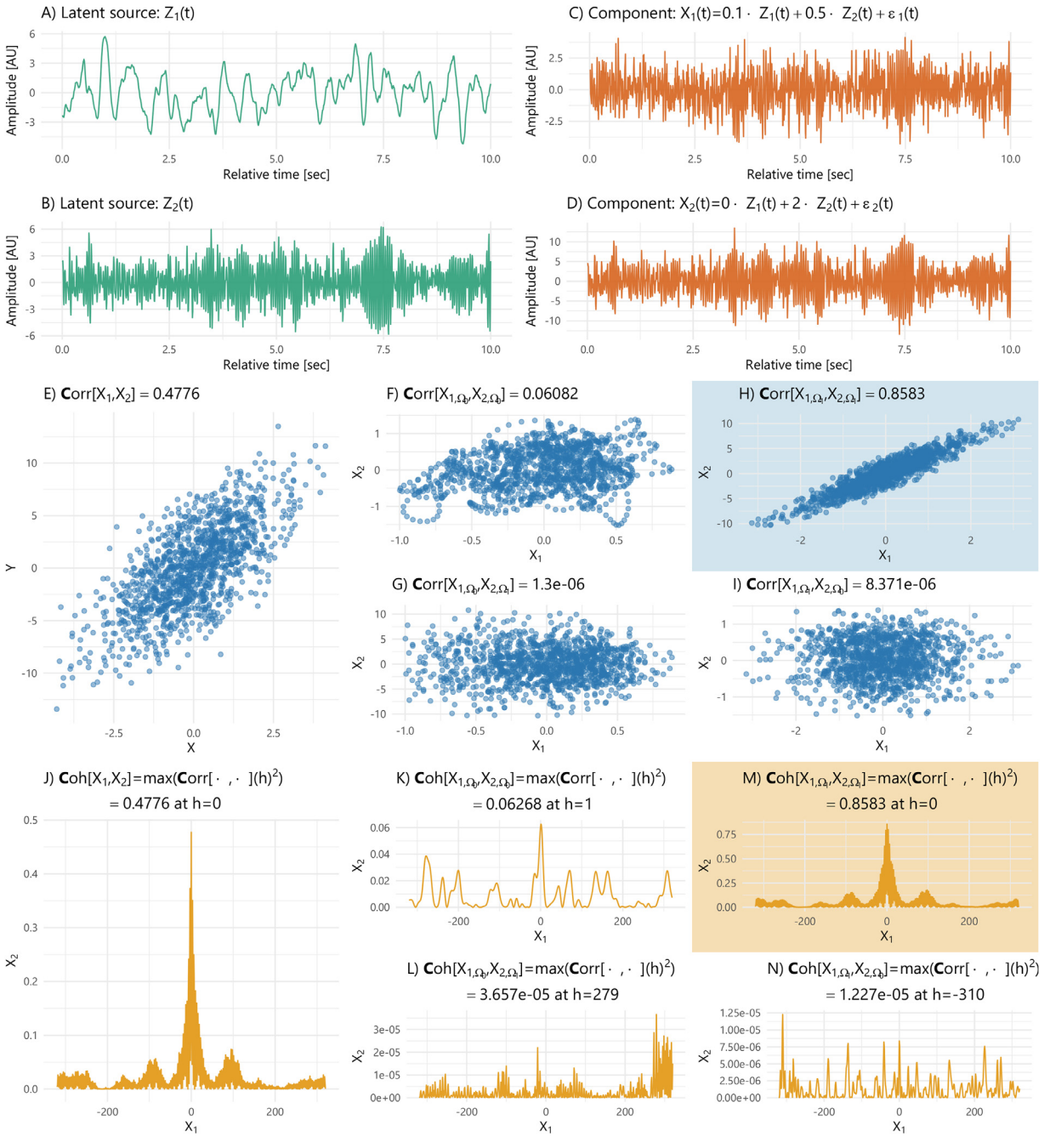

Fig. 8. Simulation from Example 2: system with an instantaneous mixture with latent sources $(Z_1(t), Z_3(t))$ and mixed components $(X_1(t), X_2(t))$. Note the differences in the correlation between the unfiltered, and the Ω_0- and Ω_1-filtered components. $\mathbb{C}\text{or}[X_1(t), X_2(t)] = 0.4776$ while $\mathbb{C}\text{or}[X_{\Omega 0}(t), X_{\Omega 0}(t)] = 0.0608$ and $\mathbb{C}\text{or}[X_{\Omega 1}(t), X_{\Omega 1}(t)] = 0.8583$.

(Nunez et al., 2015), axons can have propagation speeds in the range 600-900$\frac{cm}{s}$, and considering the average distance from the cortex to the scalp is 14.70 mm for middle-aged humans (Lu et al., 2019), we presume that delays of a few milliseconds (from any neural source to the scalp) can be feasible in the EEGs. Statistically, in order to handle these lagged mixtures, we introduce the backshift operator B^h (where $h = 0, 1, \ldots$) to be $B^h Z(t) = Z(t-h)$. Consider now the lagged mixture

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} c_{11}B^{h_{11}} & c_{12}B^{h_{12}} \\ c_{21}B^{h_{21}} & c_{22}B^{h_{22}} \end{pmatrix} \begin{pmatrix} Z_1(t) \\ Z_2(t) \end{pmatrix} + \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{pmatrix}. \tag{9}$$

Here, suppose that the mixture weight $c_{21} = 0$; and the lags for the latent sources are $h_{12} = 0$ and $h_{22} = 10$. Thus, the two observed time series are

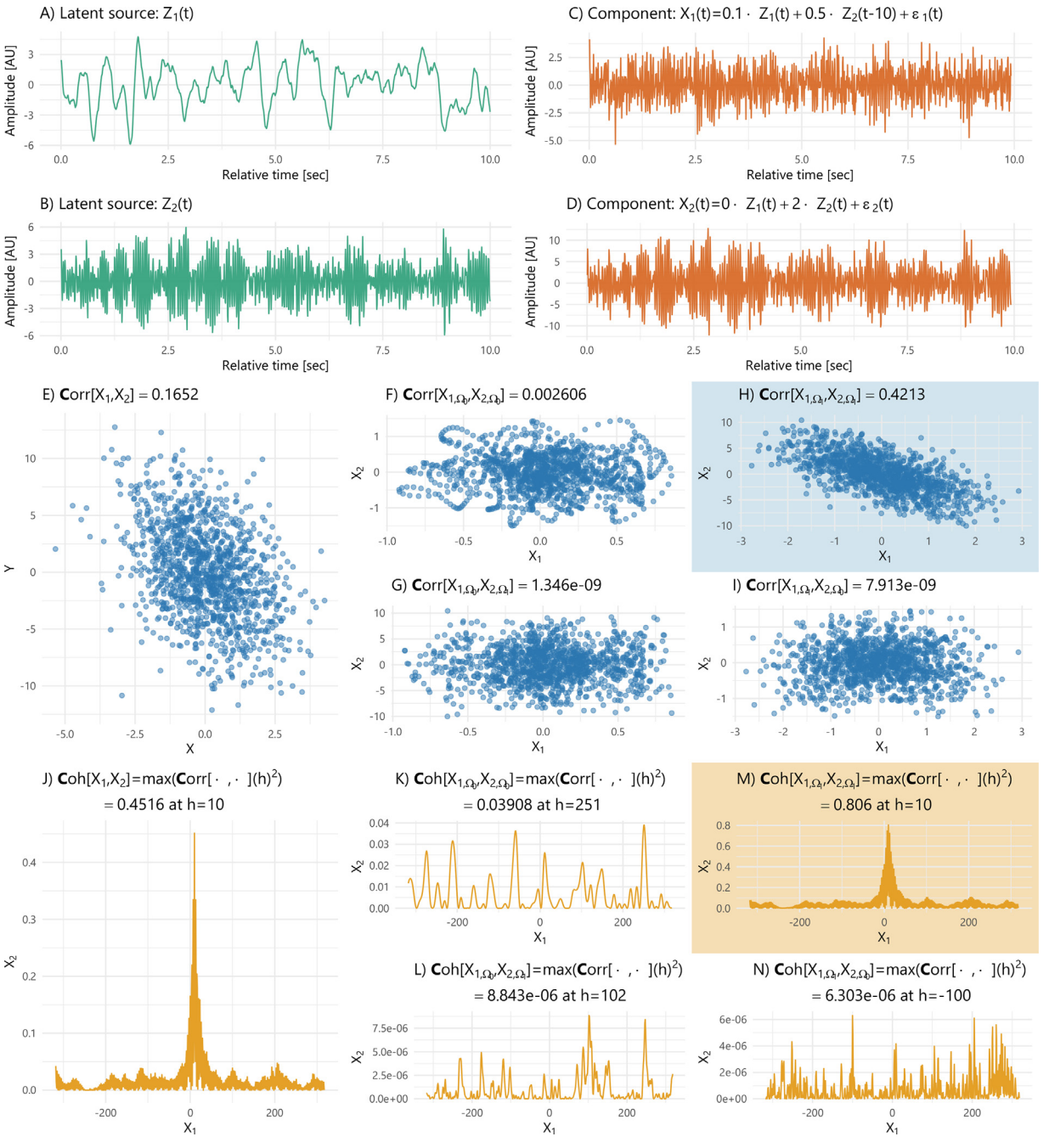

Fig. 9. Simulated example 3: system with latent sources $(Z_1(t), Z_3(t))$ and lagged mixed components $(X_1(t), X_2(t))$. Note the high contrast between correlation and coherence in low- and high-frequency components. $\mathbb{C}\mathrm{or}\,[X_{\Omega 0}(t), X_{\Omega 0}(t)] = 0.0026$ and $\mathbb{C}\mathrm{or}\,[X_{\Omega 1}(t), X_{\Omega 1}(t)] = 0.4213$ while $\mathbb{C}\mathrm{oh}\,\{X_{\Omega 0}(t), X_{\Omega 0}(t)\} = 0.039$ and $\mathbb{C}\mathrm{oh}\,\{X_{\Omega 1}(t), X_{\Omega 1}(t)\} = 0.8060$. Furthermore, the maximum correlation in the high-frequency components is located at lag 10.

$$X_1(t) = c_{11}Z_1(t) + c_{12}Z_2(t-10) + \epsilon_2(t)$$
$$X_2(t) = c_{22}Z_2(t) + \epsilon_1(t).$$

As in the previous example, the observed signals are driven by the high-frequency latent source Z_2, but the effect of Z_2 on X_2 is delayed by 10 time units. Consider now the scatterplots of the high-frequency filtered time series at (a.) lag 0: $X_{1,\Omega_1}(t)$ vs. $X_{2,\Omega_1}(t)$ and (b.) lag 10: $X_{1,\Omega_1}(t)$ and $X_{2,\Omega_1}(t-10)$. In Fig. 9, it is clear that the linear relationship at lag 10 appears stronger compared to the contemporaneous correlation. Denote the cross-correlation estimate at lag ℓ to be

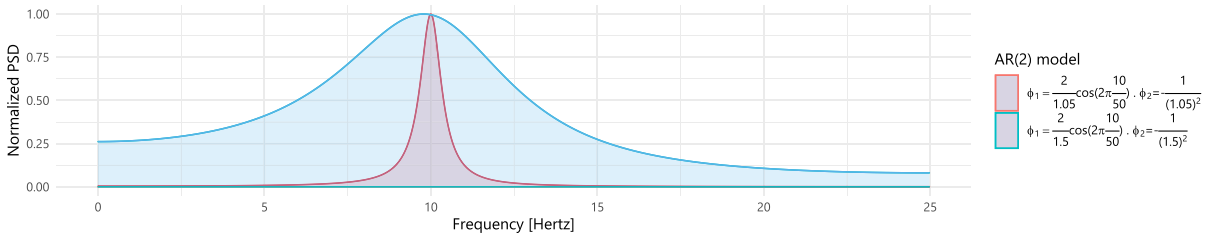

Fig. 10. Autoregressive process spectrum $AR(2)$ as a function of the magnitude of its roots. The closer the magnitude is to one, the narrower the spectrum.

$$\widehat{r}_{12}(\Omega_1, \ell) = \frac{\left\{\sum_{t=1}^{T}(X_{1,\Omega_1}(t)X_{2,\Omega_0}(t-\ell))\right\}^2}{\sum_{t=1}^{T}(X_{1,\Omega_1}(t))^2 \sum_{t=1}^{T}(X_{2,\Omega_1}(t))^2}.$$

Then the estimated coherence at the frequency band Ω_1 is

$$\widehat{\rho}_{12}(\Omega_0) = \max_{\ell=0,\pm1,\ldots} |\widehat{r}_{12}(\Omega_1, \ell)|^2.$$

3.4. AR(2) processes - discretized Cramér representation

One can approximate the Cramér representation of weakly stationary processes by a representation based on latent sources with identifiable spectra where each spectrum has its own unique peak frequency and bandwidth (or spread). Here, we will consider the class of weakly stationary second-order autoregressive, or simply AR(2), models to serve as "basis" latent sources. A process $Z(t)$ is AR(2) if it admits a representation

$$Z(t) - \phi_1 Z(t-1) - \phi_2 Z(t-2) = W(t)$$

where $\{W(t)\}$ is white noise with $\mathbb{E}[W(t)] = 0$ and $\text{Var}[W(t)] = \sigma_W^2$ and the AR(2) coefficients ϕ_1 and ϕ_2 must satisfy that the roots of the AR(2) polynomial function (denoted u_1, u_2)

$$\Phi(u) = 1 - \phi_1 u - \phi_2 u^2$$

must satisfy $|u_c| > 1$ for both $c = 1, 2$. In particular, we will consider the subclass of $AR(2)$ models whose roots are non-real complex-valued so that they are complex-conjugates of each other $u_2 = u_1^*$ and thus they can be reparametrized as

$$u_1 = M\exp(i2\pi\psi) \quad \text{and} \quad u_2 = M\exp(-i2\pi\psi)$$

where $M > 1$ and $\psi \in (-0.5, 0.5)$. Note that the AR(2) model can be parametrized by the coefficients (ϕ_1, ϕ_2) or by the roots (u_1, u_2) or by the magnitude and phase of the roots (M, ψ). In fact, the one-to-one relationship between the coefficients and the roots is given by

$$\phi_1 = \frac{2}{M}\cos(2\pi\psi) \quad \text{and} \quad \phi_2 = -\frac{1}{M^2} \tag{10}$$

One very important and interesting property of an AR(2) process with complex-valued roots is that its spectrum has a peak at ψ and the spread of this peak is governed by the magnitude M. When the root magnitude $M \longrightarrow 1+$ then the bandwidth of the peak around ψ becomes narrower. Conversely, when M becomes much larger than 1 then the bandwidth around ψ becomes wider (Fig. 10).

Example 4 (AR(2) with the peak at the alpha band)**.** We now describe how to specify an AR(2) process whose spectrum has a peak at 10 Hertz. Assume that the sampling rate is 100 Hertz with a consequent Nyquist frequency of 50 Hertz. The roots of the AR(2) processes are then complex-valued with phase $\psi = \frac{10}{50}$ and magnitude L. To model an AR(2) spectra with a narrowband response around 10 Hertz, we set the magnitude of the roots to be $L = 1.05$. Moreover, to model a broadband response, we set $L = 1.50$.

Therefore, the AR(2) coefficient parameters for the narrowband process are

$$\phi_1 = \frac{2}{1.05}\cos\left(2\pi\frac{10}{50}\right) \quad \text{and} \quad \phi_2 = -\frac{1}{(1.05)^2} \tag{11}$$

while, for the broadband signal, the AR(2) coefficients are given by

$$\phi_1 = \frac{2}{1.50}\cos\left(2\pi\frac{10}{50}\right) \quad \text{and} \quad \phi_2 = -\frac{1}{(1.50)^2}. \tag{12}$$

The corresponding auto-spectra of these two AR(2) processes are displayed in Fig. 10.

We now construct a representation for $\mathbf{X}(t)$ that is a linear mixture of uncorrelated AR(2) latent processes $Z_1, \ldots, Z_K$ whose spectra are identifiable with peaks within the bands $\Omega_1, \ldots, \Omega_K$. Define $g_k(\omega)$ to be the spectrum of the latent

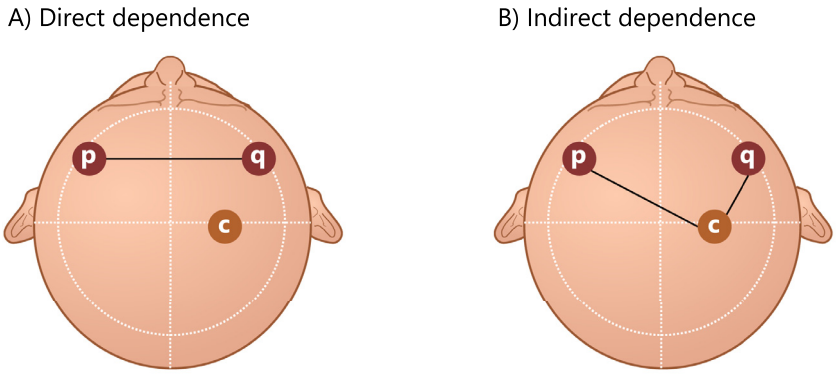

Fig. 11. Direct and indirect dependence between two channels p and q Indirect links could appear if both channels depend on a common channel c.

AR(2) process Z_k. These spectra will be standardized so that $\int_{-0.5}^{0.5} g_k(\omega)d\omega = 1$ for all $k = 1, \ldots, K$. Moreover, these spectra have peak locations that are at unique frequencies and that these are sufficiently separated. The choice of the number of components K may be guided by the standard in neuroscience where the $K = 5$ bands are modeling components in the delta, theta, alpha, beta, and gamma band (as they were defined in a previous section). Another approach is developed in Granados-Garcia et al. (2021), which data-adaptively selects K, the peak locations, and bandwidths. Therefore, the mixture of AR(2) latent sources is given by

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ \cdots \\ X_P(t) \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1K} \\ A_{21} & A_{22} & \ldots & A_{2K} \\ \cdots & & & \\ A_{P1} & A_{P2} & \ldots & A_{PK} \end{pmatrix} \begin{pmatrix} Z_1(t) \\ Z_2(t) \\ \cdots \\ Z_K(t) \end{pmatrix} \tag{13}$$

.

Consider components X_p and X_q. Suppose that, for a particular latent source Z_k, the coefficients $A_{pk} \neq 0$ and $A_{qk} \neq 0$. Then, these two signals share Z_k as the same common component, and that X_p and X_q are coherent at frequency band Ω_k.

This stochastic representation in terms of the AR(2) processes as building blocks is motivated by the result from Gao et al. (2020) for univariate processes where the essential idea is the following: define the true spectrum of a weakly stationary process X_p to be $f_{pp}(\omega)$ and set $f^A_{pp,K}(\omega)$ to be the mixture (or weighted average) of the spectrum of the K AR(2) latent processes that gives the minimum discrepancy, i.e.,

$$f^A_{pp,K} = \inf_S \int \left| f_{pp}(\omega) - S^A_{pp,K}(\omega) \right| d\omega \tag{14}$$

over all candidate spectra from a mixture of K AR(2) processes, S_K. This discrepancy decreases as K increases, and therefore, the model provides a better approximation.

3.5. Partial coherence

One of the key questions of interest is determining whether the dependence between two components X_p and X_q is pure or it is indirect through another component X_c (or a set of channels $c \neq p, q$). In Fig. 11, we show the distinction between the pure vs. indirect dependence between X_p and X_q. In Fig. 11.B, if we remove the link between X_c and X_p and the link between X_c and X_q, then there is no longer any dependence between X_p and X_q.

A standard approach in the time domain is to calculate the partial correlation between X_p and X_q ((Fried and Didelez, 2005; Fried et al., 2005)), which is essentially the cross-correlation between them after removing the linear effect of X_c on both X_p and X_q. This is also achieved by taking the inverse of the covariance matrix, denoted $\Sigma^{-1}(0)$ and then standardizing this matrix by a pre- and post-multiplication of a diagonal matrix whose elements are the inverse of the square root of the diagonal elements of $\Sigma^{-1}(0)$. The procedure for the frequency domain follows in a similar manner, as outlined in Fiecas et al. (2011): define the inverse of the spectral matrix to be $\mathbf{g}(\omega) = \mathbf{f}^{-1}(\omega)$ and let $\mathbf{h}(\omega)$ be a diagonal matrix whose elements are $\frac{1}{\sqrt{g_{rr}(\omega)}}$. Next, define the matrix

$$\Lambda(\omega) = -\mathbf{h}(\omega)\mathbf{g}(\omega)\mathbf{h}(\omega). \tag{15}$$

Then, partial coherence (PC) between components X_p and X_q at frequency ω is $|\Lambda_{pq}(\omega)|^2$. This particular characterization of PC requires inverting the spectral matrix. This can be time computationally demanding when the dimension P is large, and it also can be prone to numerical errors when the condition number of the spectral matrix $\mathbf{f}(\omega)$ is very high (i.e., the ratio of the largest to the smallest eigenvalue is large). This happens when there is a high degree of multicollinearity between the $\omega-$oscillations of the components of $\mathbf{X}$. To alleviate this problem, Fiecas et al. (2010) and Fiecas et al. (2011) developed a

class of spectral shrinkage procedures. The procedure consists in constructing a well-conditioned estimator for the spectral matrix and hence produces a numerically stable estimate of the inverse. The starting point is to construct a "smoothed" periodogram matrix which is a nonparametric estimator of $\mathbf{f}(\omega)$. From a stretch of time series $\mathbf{X}(1), \ldots, \mathbf{X}(T)\}$ (T even), compute the Fourier coefficients

$$\mathbf{d}(\omega_k) = \sum_{t=1}^{T} \mathbf{X}(t) \exp(-i2\pi\omega_k t)$$

at the fundamental frequencies $\omega_k = \frac{k}{T}$ where $k = -(\frac{T}{2} - 1), \ldots, \frac{T}{2})$. The $P \times P$ periodogram matrix is

$$\mathbf{I}(\omega_k) = \frac{1}{T}\mathbf{d}(\omega_k)\mathbf{d}^*(\omega_k)$$

where $\mathbf{d}^*$ is the complex-conjugate transpose of $\mathbf{d}$.

Though the periodogram matrix is asymptotically unbiased, it is not consistent. We can mitigate this issue by constructing a smoothed periodogram matrix estimator

$$\widetilde{\mathbf{f}}(\omega) = \sum_{\ell} Q_b(\omega_\ell - \omega)\mathbf{I}(\omega_\ell)$$

where the kernel weights $Q_b(u)$ are non-negative and sum to 1. The bandwidth b can be obtained using automatic bandwidth selection methods for periodogram smoothing, such as the least-squares in Lee (1997) or the gamma-deviance-GCV in Ombao et al. (2001b). The shrinkage targets are either the scaled identity matrix or a parametric matrix (that can be derived from a VAR model).

Suppose that a VAR(L) model is fit to the signal $\mathbf{X}(t), t = 1, \ldots, T\}\Phi(B)\mathbf{X}(t) = \mathbf{W}(t)$ where $\mathbf{W}(t)$ is a zero-mean $P-$variate white noise with $\mathbb{C}\text{ov}\,\mathbf{W}(t) = \Sigma_{\mathbf{W}}$ and $\Phi(B) = I - \Phi_1 B - \ldots - \Phi_L B^L$. Then the spectrum of this VAR(L) process is

$$\mathbf{h}(\omega) = \Phi(\exp(-i2\pi\omega))\ \Sigma_{\mathbf{W}}\ \Phi^*(\exp(-i2\pi\omega))$$

.

The parametric estimate of the spectral matrix, denoted $\widetilde{\mathbf{h}}(\omega)$, is obtained replacing Φ_ℓ, $\ell = 1, \ldots, L$, by the maximum likelihood or conditional maximum likelihood estimators. The shrinkage estimator for the spectrum takes the form

$$\widehat{\mathbf{f}}(\omega) = W_1(\omega)\widetilde{I}(\omega) + W_2(\omega))\widetilde{\mathbf{h}}(\omega)$$

where the weights $W_s(\omega)$ fall in (0,1) and $W_1(\omega) + W_2(\omega)$ for each ω. Moreover, the weight for the smoothed periodogram is proportional to the mean-squared error of the parametric estimator, i.e., $W_1(\omega) \propto \mathbb{E}|\widetilde{h}(\omega) - \mathbf{f}(\omega)|^2$.

This method can be interpreted as a spectrum estimator that automatically chooses the estimator with better "quality". Therefore, when the parametric estimator is "poor" (high MSE), the weights shift the estimator towards the nonparametric estimator. However, when the parametric model gives a good fit, the shrinkage shift to the parametric spectral estimate. Thus, the resulting spectral estimator has, in general, a good condition number and can be used further for the estimation of PC where the quality of the spectral information is critical.

An alternative view to the above approach in constructing partial coherence is through analyzing the oscillations. For simplicity, we consider the ω-oscillations for channels X_p, X_q and X_c, which we denote to be $X_{p,\omega}$, $X_{q,\omega}$ and $X_{c,\omega}$, respectively. At this stage, we shall consider only the contemporaneous (i.e., zero-phase or zero-lag) partial cross-correlation

$$\mathbb{C}\text{or}\{X_{p,\omega}(t), X_{q,\omega}(t) \mid X_{c,\omega}(t)\}.$$

To proceed, regress of $X_{p,\omega}(t)$ against $X_{c,\omega}(t)$ and extract the residuals, denoted $R_{p\dot{c},\omega}(t)$ and also $X_{q,\omega}(t)$ against $X_{c,\omega}(t)$, and extract the residuals, named as $R_{q\dot{c},\omega}(t)$.

Then, the zero-phase partial coherence between X_p and X_q at frequency ω is the quantity

$$\rho_{pq\dot{c}}(\omega) = \left|\mathbb{C}\text{or}\left[R_{p\dot{c},\omega}, R_{q\dot{c},\omega}\right]\right|^2.$$

Example 5 (Partial coherence on a gamma-interacting system)**.** Consider the setting where the independent latent processes are specific AR(2)'s that mimic the delta, alpha, and gamma activity which we denote by $Z_\delta(t)$, $Z_\alpha(t)$ and $Z_\gamma(t)$. Suppose that the observed time series are $X_1(t)$, $X_2(t)$ and $X_3(t)$, which are defined by the mixture

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} = \begin{pmatrix} a_{1\delta} & a_{1\alpha} & a_{1\gamma} \\ a_{2\delta} & a_{2\alpha} & a_{2\gamma} \\ a_{3\delta} & a_{3\alpha} & a_{3\gamma} \end{pmatrix} \begin{pmatrix} Z_\delta(t) \\ Z_\alpha(t) \\ Z_\gamma(t) \end{pmatrix}. \tag{16}$$

We now examine the dependence between X_1 and X_2 under the following setting. First, suppose that X_3 contains the gamma-oscillatory activity Z_γ purely, that is, $a_{3\delta} = a_{3\alpha} = 0$. Next, suppose that X_1 contains only Z_α and Z_γ, that is, $a_{1\delta} = 0$; and X_2 contains only Z_δ and Z_γ, that is, $a_{2\alpha} = 0$. A realization of such a system is shown in Fig. 12. Under this construction, X_1 and X_2 have zero coherence at the δ and α frequency band. However, X_1 and X_2 both contain the common gamma-oscillatory activity and hence have a non-zero coherence at the gamma-band. In our particular example, $\mathbb{C}\text{oh}\left\{X_{1,\gamma}, X_{2,\gamma}\right\} = 0.759$, $\mathbb{C}\text{oh}\left\{X_{1,\gamma}, X_{3,\gamma}\right\} = 0.770$, and $\mathbb{C}\text{oh}\left\{X_{2,\gamma}, X_{3,\gamma}\right\} = 0.893$.

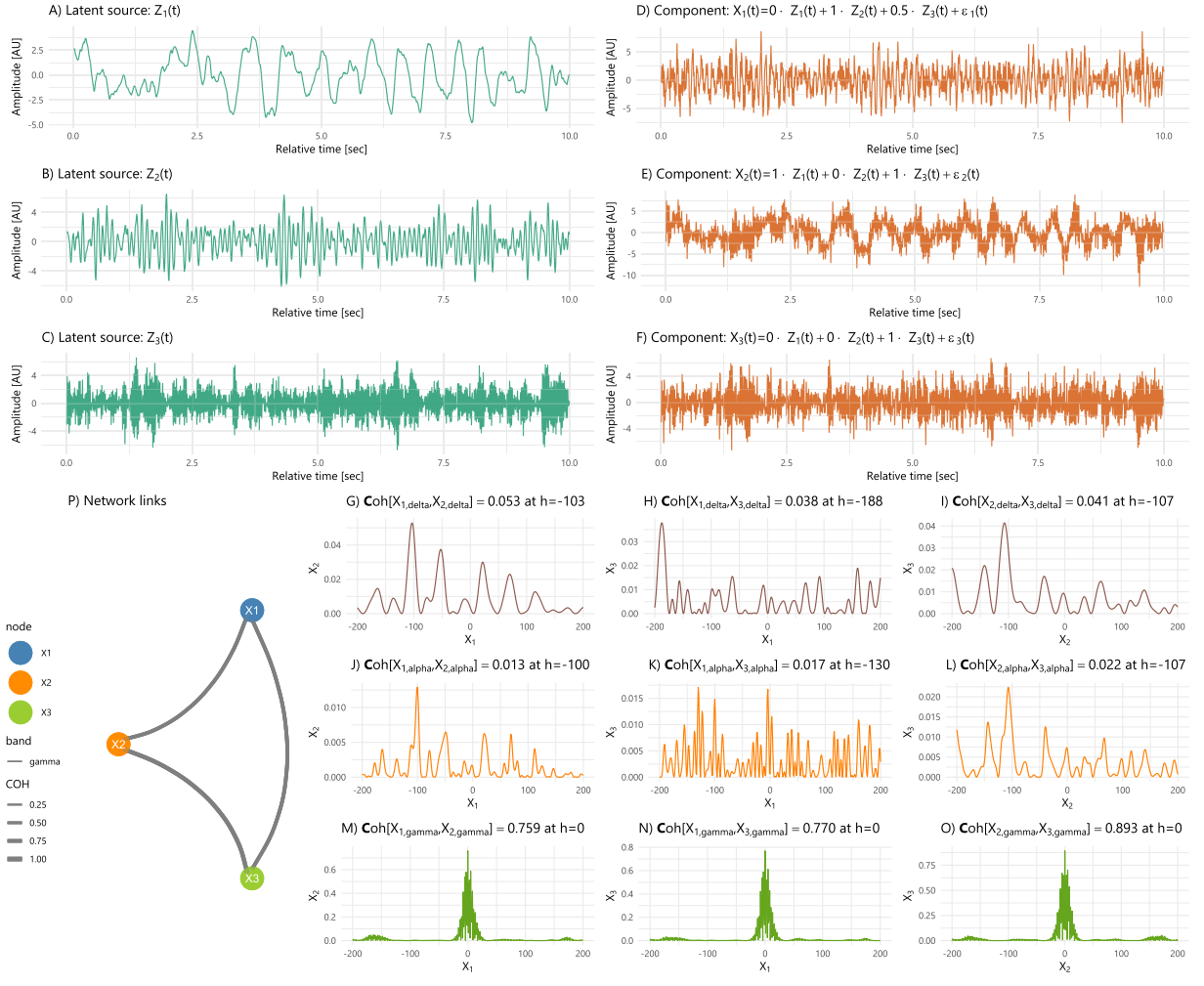

Fig. 12. Coherence network of a system with a dominating gamma component (Example 5). Three latent sources (delta-, alpha- and gamma-band waves) are linearly into combined three observed components X_1, X_2 and X_3. As expected, the network map (P.) denotes that all components are related within the gamma-band. The correlation across lags is also displayed to emphasize the coherence magnitudes.

However, if we compare coherence and partial coherence (Fig. 13), we can unveil some indirect dependence effects. Partial coherence between X_1 and X_2 (conditional on X_3) at the gamma-band is zero. This result indicates that if we remove X_3 (the gamma-band activity) from X_1 and X_2, then their "residuals" will no longer contain any common latent source.

Example 6 (Partial coherence)**.** As a continuation of Example 5, suppose that (a.) X_3 contains only gamma-oscillatory activity Z_γ, i.e., $a_{3\delta} = a_{3\alpha} = 0$; (b.) X_1 contains only Z_α and Z_γ: $a_{1\delta} = 0$; and (c.) X_2 contains Z_δ, Z_α and Z_γ. In such a dependence structure, we can observe the following frequency-dependent effects on X_1 and X_2:

(i.) At the delta-band, the coherence between X_1 and X_2 is zero; and hence the partial coherence (conditioned on X_3) is also zero;
(ii.) At the alpha-band, the coherence and partial coherence between X_1 and X_2 are both non-zero;
(iii.) At the gamma band, the coherence between X_1 and X_2 is non-zero but the partial coherence is zero.

These observations can be summarized in the coherence and partial coherence networks of Fig. 14.

3.6. Time-varying coherence and partial coherence

As noted, many brain signals exhibit non-stationarity. In some cases, the autospectra varies over time which indicates that the contributions of the various oscillations to the total variance change across the entire recording. In others, the autospectra might remain constant, but the strength and nature of the association between components can change. In fact, in Fiecas and Ombao (2016), the coherence between a pair of tetrodes from a local field potential implanted in a monkey

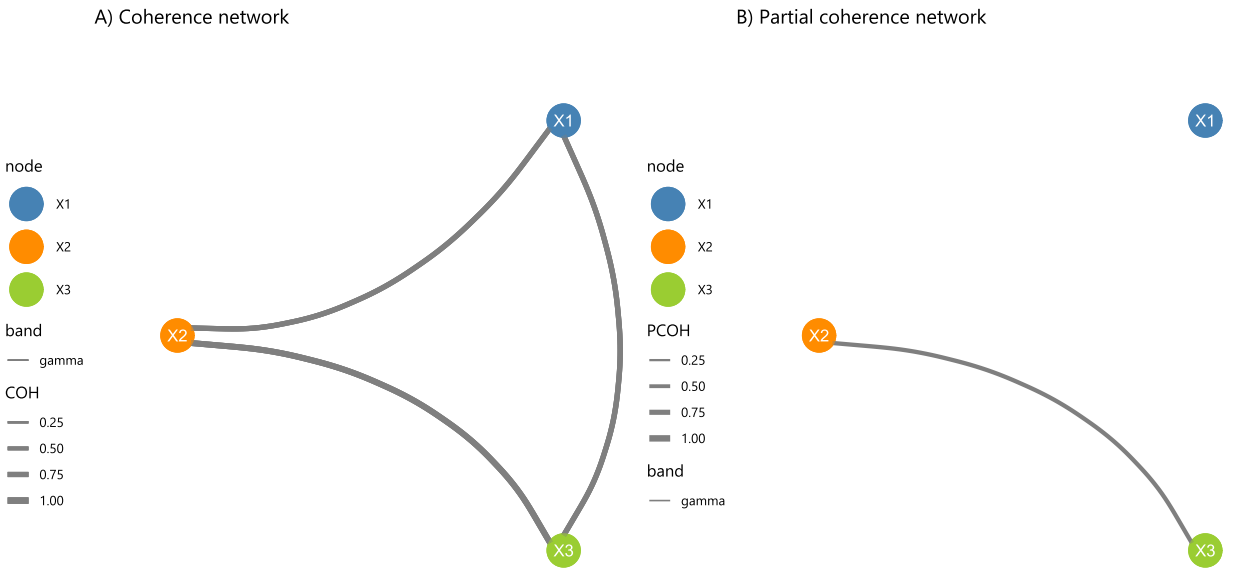

Fig. 13. Comparison between coherence and partial coherence in Example 5.

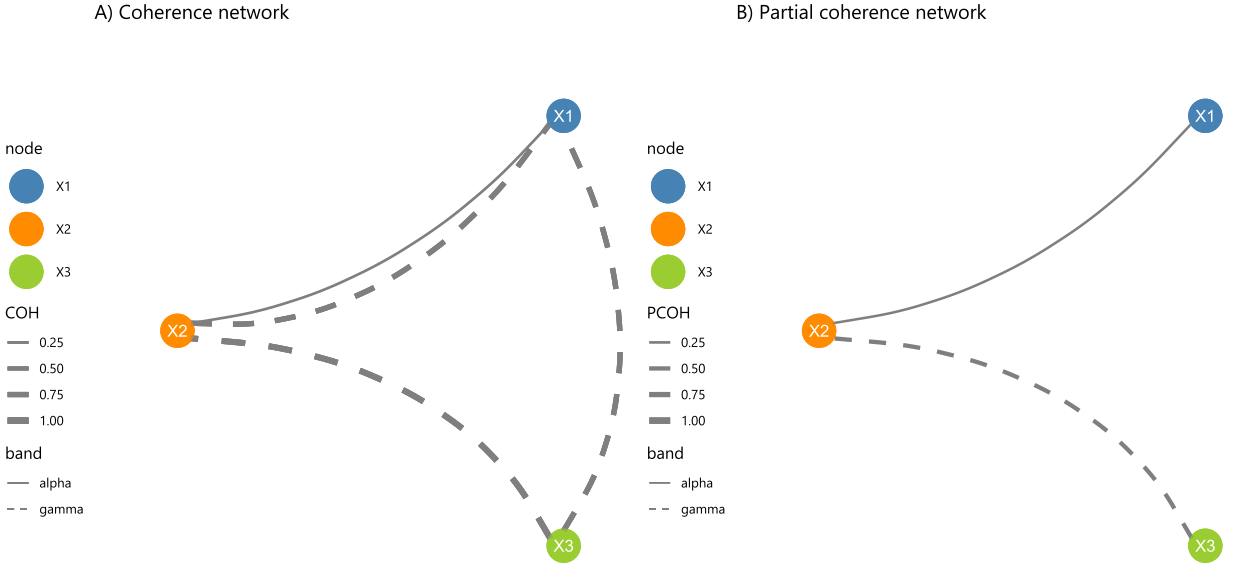

Fig. 14. Comparison between coherence and partial coherence in Example 6.

evolves both within a trial and even across trials in an experiment. Here, we will follow the models in Priestley (1965) and in Dahlhaus (2012) to define and estimate the time-varying coherence and partial coherence. The methods for non-stationary signals are discussed in Section 7.

4. Dual-frequency dependence

In this paper, we view the components of a multivariate time series $\mathbf{X}$ to consist of oscillations at various frequencies with random amplitudes and phases (complex-valued coefficients). Thus, we present a framework for modeling dependence between the components of $\mathbf{X}$ through thes e random oscillations. This is motivated by the Cramér representation $\mathbf{X}(t) = \int_{-0.5}^{0.5} \exp(i2\pi\omega t) d\mathbf{Z}(\omega)$ where, for weakly stationary processes, the random increments $d\mathbf{Z}(\omega)$ are uncorrelated across the frequency range in the unit interval $\omega \in (-0.5, 0.5)$. Hence, for $\omega \neq \omega'$, the $\omega-$oscillations in component X_p and the ω'-oscillations in X_q are, by default, uncorrelated (or independent if $d\mathbf{Z}(\omega)$ is Gaussian). However, there are situations, in the non-stationary setting, when the interesting dependence structure is between different frequency oscillations.

Loéve, in Loève (1955), introduced the class of harmonizable processes, which now allow for dependence between oscillatory activities at different oscillations. Here, we shall explore different characterizations of dependence between the random

oscillations $X_{p,\omega}(t) = \exp(i2\pi\omega t)dZ_p(\omega)$ and $X_{q,\omega'}(t) = \exp(i2\pi\omega' t)dZ_q(\omega')$. The first natural measure of dependence is the linear association between these two oscillations, which we call the dual-frequency coherence

$$\rho_{(p,\omega),(q,\omega')} = \mathbb{C}\text{or}\left[X_{p,\omega}(t), X_{q,\omega'}(t)\right].$$

This perspective opens up many possibilities of characterizations of dependence that are beyond that by classical coherence. We will now further generalize this notion of dual-frequency coherence to the situation where this measure could evolve over time. For example, during the course of one trial recording of an electroencephalogram, it is possible for coherence between the theta and gamma oscillations to be to be low upon the presentation of a stimulus but could gradually strengthen over the course of a trial. We will summarize the notion of time-dependent dual-frequency coherence in Section 4.1. In addition, note that coherence, partial coherence, and dual-frequency coherence all capture only the linear dependence between the various oscillations. Here, we will examine other non-linear measures of dependence, such as the phase-amplitude coupling. In Section 4.2, we will illustrate how the amplitude of a gamma-oscillatory activity in channel p might be change according to the phase of the alpha-activity in channel q. One such example could be the increase in the gamma-oscillation amplitude when the theta oscillation reaches its peak.

4.1. Evolutionary dual-frequency coherence

The modeling framework for developing the notion of a dual-frequency that evolves over time was first proposed in Cristina et al. (2018) via a frequency-discretized harmonizable process which we now describe. Here, we shall focus on the practical aspect of modeling and analyzing dual-frequency coherence in the observed brain signals. The technical details such as the asymptotic theory required for defining the population-specific quantity are developed in Cristina et al. (2018). Suppose that we observe the time series $\mathbf{X}$ and focus on a local window centered time point t with N observations. Define the *local* Fourier coefficient vector at frequency ω at time point t to be

$$\mathbf{d}(t,\omega) = \sum_{s=t-(\frac{N}{2}-1)}^{t+\frac{N}{2}} \mathbf{X}(s)\exp(-i2\pi\omega s) \tag{17}$$

and define the local dual-frequency periodogram matrix at frequencies ω_j and ω_k to be

$$\boldsymbol{I}(t;(\omega_j,\omega_k)) = (1/N)\boldsymbol{d}(t,\omega_j)\boldsymbol{d}(t,\omega_k). \tag{18}$$

When the data consists of several time-locked trials, one can compute trial-specific local dual-frequency periodogram matrices and then average them (across trials) in order to obtain some population-specific measure of the evolutionary dual-frequency spectral matrix. In the absence of replicated trials (i.e., the data is only from a single trial), then one can smooth the local dual-frequency periodogram matrices over time t within that single trial and also over frequencies a two-dimensional neighborhood around $(\omega_j,\omega)k)$. Denote the averaged dual-frequency periodogram matrix to be $\widehat{\mathbf{f}}(t,\omega_j,\omega_k)$. One measure of the strength of linear dependence between the ω_j-oscillations at component X_p and the ω_k-oscillations at component X_q is the time-localized dual-frequency coherence

$$\widehat{\rho}_{(p,\omega_j),(q,\omega_k)}(t) = \frac{|\widehat{f}_{(p,\omega_j),(q,\omega_k)}(t)|^2}{\widehat{f}_{(p,\omega_j),(p,\omega_j)}(t)\widehat{f}_{(q,\omega_k,(q,\omega_k)}(t)}. \tag{19}$$

Consistent with the approach adopted in the paper, linear dependence between the different oscillations will be examined also via linear filtering which we now describe. Consider two frequency bands Ω_1 and Ω_2 and the zero-mean filtered signals to be X_{1,Ω_1} and X_{2,Ω_2}. Under the proposed framework, the time-varying dual-frequency coherence estimate at a local time t by first computing the local cross-covariance and local variance estimate over the window $\{t-L,\ldots,t+L\}$ (for some $L>0$) as follows

$$\begin{aligned}
\widehat{f}_{(1,\Omega_1),(2,\Omega_2)}(t) &= \tfrac{1}{N}\sum_{s=t-(\frac{N}{2}-1)}^{t+\frac{N}{2}} X_{1,\omega_1}(s)X_{2,\omega_2}(s)\\
\widehat{f}_{(1,\Omega_1),(1,\Omega_1)}(t) &= \tfrac{1}{N}\sum_{s=t-(\frac{N}{2}-1)}^{t+\frac{N}{2}} X_{1,\omega_1}(s)X_{1,\omega_1}(s)\\
\widehat{f}_{(2,\Omega_2),(2,\Omega_2)}(t) &= \tfrac{1}{N}\sum_{s=t-(\frac{N}{2}-1)}^{t+\frac{N}{2}} X_{2,\omega_2}(s)X_{2,\omega_2}(s).
\end{aligned}$$

Consequently, the estimated local dual-frequency coherence at time t is

$$\widehat{\rho}_{(1,\Omega_1),(2,\Omega_2)}(t) = \frac{|\widehat{f}_{(1,\Omega_1),(2,\Omega_2)}(t)|^2}{\widehat{f}_{(1,\Omega_1),(1,\Omega_1)}(t)\widehat{f}_{(2,\Omega_2),(2,\Omega_2}(t)}. \tag{20}$$

4.2. Phase-Amplitude coupling

In the previous sections, we first examined coherence and partial coherence - both of which aim to measure the strength of linear dependence between a pair of channels or components X_p and X_q through the oscillatory activity at the *same* frequency. As a generalization, in Section 4, we examined dependence at different frequencies and how this dependence may change over time through the evolutionary dual-frequency coherence. However, all major limitation of these measures is

that they aim to capture only *linear* dependence. In the neuroscience literature, it is acknowledged that cross-frequency modulation could also appear as a different category of dependence in the brain networks. Similarly to coherence, this dependence type implies interactions within the frequency domain: spectral components in Ω_H are *modulated* by components resonating at frequencies in Ω_L. When the modulator Ω_H works at a high frequency, and the modulated signal Ω_L, this type of dependence is named phase-amplitude coupling (PAC). Baseline PAC networks in the human brain changes since birth (Mariscal et al., 2019), and specific delta-theta PAC patterns can be altered due to anesthetic effects (Molaee-Ardekani et al., 2007).

A common method to quantify PAC in time series was introduced in (Tort et al., 2010) through the modulation index (MI) for univariate signals. As a clarification, this metric is not the same as the homonym "modulation index" defined for amplitude-modulated systems (Oppenheim et al., 1997, p. 591). To estimate MI, let X_Ω be a filtered time series with its spectrum concentrated only in the interval Ω. Then, its analytic signal Y_Ω will be defined via the Hilbert transform:

$$Y_\Omega(t) = X_\Omega(t) + j\left(\frac{1}{\pi t} * X_\Omega(t)\right),$$

where the analytic $Y_\Omega(t)$ can be expressed using an exponential form,

$$Y_\Omega(t) = A_\Omega(t)\exp\left(j\phi_\Omega(t)\right)$$

such that the instantaneous phase and amplitude are identified through $\phi_\Omega(t)$ and $A_\Omega(t)$, respectively. Now obtain $Y_{\Omega_L}(t)$ and $Y_{\Omega_H}(t)$ for a target low-frequency range Ω_L and a high-frequency interval Ω_H (such as delta and gamma, respectively). Now, consider the joint signals $T = \left(A_{\Omega_L}, \phi_{\Omega_H}\right)$. The original algorithm, described by Tort et al., suggests creating a partition of the phase domain Q: $Q = \cup_{j=1}^{N} Q_j \quad Q_j = \left[2\pi\frac{j-1}{N} 2\pi\frac{j}{N}\right)$ and estimate the marginal mean value in each partition

$$P(j) = \mathbb{E}\left(T \mid \phi_{\Omega_H} = \phi\right) \quad \phi \in Q_j$$

Under the non-modulation hypothesis, $P(j)$ should resemble a uniform distribution $P_U(j) = 1\,\forall j$. MI is then introduced as a normalized measure of the divergence between of the difference between $P(j)$ and the uniform alternative $P_U(j)$: $\text{MI} = [\log(N)]^{-1} D_{KL}(P, U)$ where $D_{KL}(\cdot)$ is the Kullback-Leibler divergence:

$$D_{KL}(P, Q) = \sum_{k=1}^{N} P(k)\log\left(\frac{P(i)}{Q(i)}\right)$$

Note that in cases where no specific frequency band is known, our suggestion is to start by exploring PAC phenomena in the data using the "smoothed" signals (low-frequency) and their residuals (high-frequency components). We refer to (Fried et al., 2012; Schettlinger et al., 2009; Fried and George, 2011) for a comprehensive description of time-memory efficient estimation methods.

Example 7 (Phase-amplitude coupling)**.** Define the theta-band and gamma-band latent sources to be Z_θ and Z_γ, respectively. Assume that some modulation effects are observed as a result of non-linear mixtures of the latent sources:

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} A_{1,\delta}(t) & A_{1,\gamma}(t) \\ A_{2,\delta}(t) & A_{2,\gamma}(t) \end{pmatrix} \begin{pmatrix} Z_\theta(t) \\ Z_\gamma(t) \end{pmatrix} + \varepsilon(t)$$

where $\varepsilon(t) \sim \mathcal{N}(0, \Sigma_\varepsilon)$ with a covariance matrix $\Sigma_\varepsilon = 0.1\mathbf{I}$ such that $\mathbf{I}$ is an identity matrix. Next, we assume that two phase-amplitude coupling effects are observed with the following properties:

(i.) $X_1(t)$ denotes an amplitude-modulation effect where the amplitude of $Z_\theta(t)$ instantaneously leads to the amplitude's changes in the γ-oscillations. Therefore, the mixture functions are defined as $A_{1,\delta}(t) = Z_\gamma(t) + 1$ and $A_{1,\gamma}(t) = 2$.
(ii.) $X_2(t)$ shows a modulation effect where the $Z_\gamma(t)$ component has a low impact on the δ-oscillations: $A_{2,\delta}(t) = 4$ and $A_{2,\gamma}(t) = Z_\theta$.

A simulation of this process along with the modulation indexes for $Z_\delta(t)$, $Z_\gamma(t)$, $\varepsilon(t)$ and the observed components $X_1(t)$ and $X_2(t)$ are shown in Fig. 15. It is visually apparent that $P(j)$ is closer to $P_U(j)$ in scenarios without modulation effects (Fig. 15.A-F). In addition, the modulated process $X_2(t)$ implies an alteration on the extreme values of the process, and it could also be modeled using statistical models for extreme values (Guerrero et al., 2021).

5. Partial directed coherence and spectral causality

All previously mentioned spectral measures of dependence (namely, coherence, partial coherence, dual-frequency coherence, and evolutionary dual-frequency coherence) focus only on contemporaneous dependence and ignore the lead-lag relationships between oscillatory components. This notion of lead-lag is very important in neuroscience, particularly in identifying effective connectivity between brain regions or channels. In fact, many pioneering models for brain connectivity were applied to functional magnetic resonance imaging data and therefore took into account the spatial structure in the brain data (Bowman et al. (2008)). To address the computational issues for spatial covariance Castruccio et al. (2018) introduced a scalable multi-scale approach (local for voxels within a region of interest; global for regions of interest in the entire network).

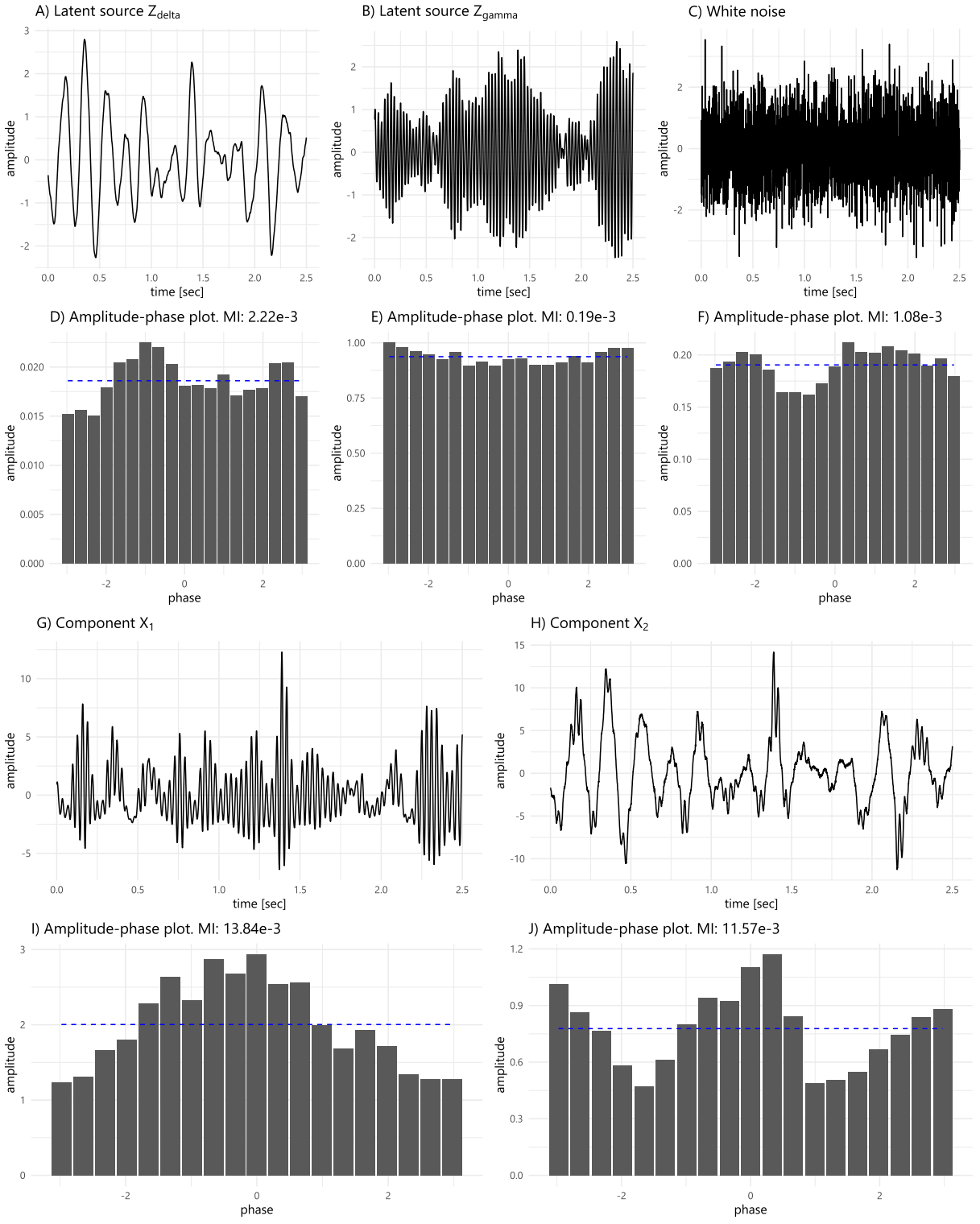

Fig. 15. Phase-amplitude coupling in Example 7: phase-amplitude graphs $P(j)$ and modulation indexes (MIs) for the latent sources $Z_\delta(t)$, $Z_\gamma(t)$ and the observations $X_1(t)$ and $X_2(t)$. Note that MI values in $X_1(t)$ and $X_2(t)$ are up to five times greater than MI magnitudes of the latent sources or the noise $\varepsilon(t)$.

In Kang et al. (2012), the temporal covariance structure is diagonalized (and hence sparsified) by applying a Fourier transform on the voxel-specific time series while taking into account variation across subjects through a mixed-effects modeling framework. In a recent work, Zhao and Luo (2019) developed a method for mediation analysis in fMRI data, and (Zhang et al., 2016) proposed a variational Bayes algorithm with reduced computational costs.

There have also been a number of statistical methods for modeling dynamic connectivity in fMRI: dynamic correlation in Cribben et al. (2012a); switching vector autoregressive (VAR) model in Balqis et al. (2017), regime-switching factor models in Ting et al. (2018c) and Shappell et al. (2019); Bayesian model of brain networks in (Zhang et al., 2016; 2015; Mumford and Ramsey, 2014; Kook et al., 2019; Rembach et al., 2015; Zhang et al., 2015); models for high dimensional networks in Ting et al. (2018b); detection of dynamic community structure in (Bassett et al., 2011), Ting et al. (2020) and Ting et al. (2021a). In Yu et al. (2016), a Bayesian VAR model was used to assess stroke-induced changes in the functional connectivity structure.

This section will cover approaches to identify lead-lag structures in brain signals using the vector autoregressive (VAR) models. Modeling these lead-lag dynamics is crucial to understanding the nature of brain systems, the impact of a shock (such as stroke) on the dynamic configuration of such systems, and the downstream effect of such disruption on cognition and behavior. However, the emphasis of this section will be developing a statistical framework for investigating frequency-band specific lead-lag dynamics. To further elaborate, note that while the classical VAR model characterizes the effect of the past observation of the signal X_q on a future observation on the signal X_p, our emphasis here will be on modeling and assessing the impact of the previous oscillatory activity X_{q,ω_1} on the future oscillations X_{p,ω_2}.

5.1. Vector autoregressive (VAR) models

A P-variate time series $\mathbf{X}$ is said to be a VAR model of order L (denoted VAR(L)) if it is weakly stationary and can be expressed as $\Phi(B)\mathbf{X}(t) = \mathbf{W}(t)$ where $\{\mathbf{W}(t)\}$ is white noise with $\mathbb{E}[\mathbf{W}(t)] = 0$ for all time t and $\mathbb{C}\text{ov}[\mathbf{W}(t)] = \Sigma_{\mathbf{W}}$ and $\Phi(B) = I - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_L B^L$ and Φ_ℓ, $\ell = 1, \ldots, L$ are the $P \times P$ VAR coefficient matrices. A comprehensive exposition of VAR models, including conditions for causality and methods for estimation of the coefficient matrices $\{\Phi_\ell\}$, are provided in (Lutkepohl, 2005). Note that from classical time series literature, the notion of "causality" is different from that of Granger causality. A time series $\mathbf{X}(t)$ is causal if it does not depend on future innovations, i.e., it depends only on the current or past white noise $\{\mathbf{W}(s),\ s = t, t-1, t-2, \ldots\}$.

Consider the two components X_p and X_q. From the point of view of forecasting, we say that "X_q Granger-causes X_p" (we write $X_q \longrightarrow X_p$) if the past values of X_q helps to improve prediction of the future values of X_p, in particular, if the squared error for the forecast of X_p that uses the past values of X_q is lower than that for the forecast that does not use the past values of X_q. Under the context of VAR models, note that

$$X_p(t) = \sum_{\ell=1}^{L} \Phi_{pq,\ell} X_q(t-\ell) + \sum_{r \neq q} \sum_{\ell=1}^{L} \Phi_{pr,\ell} X_r(t-\ell) \tag{21}$$

where $\Phi_{pq,\ell}$ is the (p,q) element of the matrix Φ_ℓ. For $X_p(t)$, it is the coefficient associated with the past value $X_q(t-\ell)$. Thus, $X_q \longrightarrow X_p$) if there exists some lag $\ell*$ where $\Phi_{pq,\ell*} \neq 0$. There is a large body of work on causality, starting with the seminal paper by Granger (1969). This was further studied in Geweke (1982), Geweke (1984), Hosoya (1991). Additional applications for subject- and group-level analysis was also analyzed in (Chiang et al., 2017). Recently, under non-stationarity, the nature of Granger-causality could evolve over time and this was investigated in Liu et al. (2021). Despite the fact that the concept of causality is derived from the spectral representation, there has not been a keen focus on the interpretations for causality based on the actual oscillatory activities. Hence, the goal in this section is to refocus the spotlight on this very important role of the oscillations in determining causality and, in general, directionality between a pair of signals.

5.2. Partial directed coherence

As noted, all previously discussed measures of coherence lack the important information on directionality. The concept of partial directed coherence (PDC), introduced in Baccalá and Sameshima (2001) and Baccalá et al. (2013), gives this additional information. Denote the VAR(L) inverse transfer function as $\Phi(\omega) = I - \sum_{\ell=1}^{P} \Phi_\ell \exp(-i2\pi\omega\ell)$ and denote $\Phi_{pq}(\omega)$ to be the (p,q) element of the inverse transfer function matrix $\Phi(\omega)$. Then the PDC, from component X_q to X_p, at frequency ω, is

$$\pi_{pq}(\omega) = \frac{|\Phi_{pq}(\omega)|^2}{\sum_{rq} |\Phi_{rq}(\omega)|^2}.$$

Note that $\pi_{pq}(\omega)$ lies in [0,1] and measures the amount of information flow, at frequency ω, from component X_q to X_p, relative to the total amount of information flow from X_q to all components. When $\pi_{pq}(\omega)$ is close to 1 then most of the ω-information flow from X_q goes directly to component X_p. Note that (normalized) total PDC on a close interval $\Omega = (\omega_0, \omega_1)$ can therefore be obtained by integrating the PDC on the regionΩ :

$$\Pi_{pq}(\Omega) = \frac{1}{\omega_1 - \omega_0} \int_{\omega_0}^{\omega_1} \pi_{pq}(\omega)\, d\omega$$

Table 1
Comparison of the connectivity metrics between channels and frequency in Example 8: coherence (COH), partial coherence (PCOH), and partial directed coherence (PDC). PDC was estimated through four different vector autoregressive models: VAR(2) and VAR(15) and regularized models: LASSLE(2) and LASSLE(15).

Band	Link	COH	PCOH	PDC: VAR(2)	PDC: VAR(15)	PDC: LASSLE(2)	PDC: LASSLE(15)
δ	$X_2 \Rightarrow X_1$	**0.9996**	**0.9749**	**0.5096**	**0.5334**	**0.5097**	**0.5097**
0-4	$X_1 \Rightarrow X_2$			0.0000	0.0070	0.0000	0.0000
Hertz	$X_3 \Rightarrow X_1$	**0.9959**	**0.2074**	0.0001	0.0438	0.0000	0.0000
	$X_1 \Rightarrow X_3$			**0.1754**	**0.6465**	**0.1878**	**0.4634**
	$X_4 \Rightarrow X_1$	0.0257	0.0347	0.0001	0.0440	0.0000	0.0000
	$X_1 \Rightarrow X_4$			0.0061	0.3040	0.0115	0.0031
	$X_3 \Rightarrow X_2$	**0.9961**	**0.3320**	**0.5774**	**0.5786**	**0.5774**	**0.5773**
	$X_2 \Rightarrow X_3$			0.2799	0.4561	0.0900	0.1495
	$X_4 \Rightarrow X_2$	0.0259	0.0336	**0.5773**	**0.5789**	**0.5773**	**0.5773**
	$X_2 \Rightarrow X_4$			0.0061	0.4376	0.0041	0.0096
	$X_4 \Rightarrow X_3$	0.0239	**0.2715**	**0.4152**	**0.4152**	0.0448	**0.3549**
	$X_3 \Rightarrow X_4$			0.0052	0.6050	0.0000	0.0000
	$X_2 \Rightarrow X_1$	0.9857	**0.9863**	**0.3854**	**0.2901**	**0.3854**	**0.3855**
	$X_1 \Rightarrow X_2$			0.0000	0.0547	0.0000	0.0000
γ	$X_3 \Rightarrow X_1$	0.0280	0.0404	0.0001	0.1165	0.0000	0.0000
30-50	$X_1 \Rightarrow X_3$			0.0009	**0.3611**	0.0009	0.0036
Hertz	$X_4 \Rightarrow X_1$	**0.9860**	**0.3938**	0.0000	0.1170	0.0000	0.0000
	$X_1 \Rightarrow X_4$			0.1233	**0.6449**	0.0599	0.0882
	$X_3 \Rightarrow X_2$	0.0232	0.0394	**0.5773**	**0.5800**	**0.5774**	**0.5773**
	$X_2 \Rightarrow X_3$			0.0012	0.4270	0.0006	0.0023
	$X_4 \Rightarrow X_2$	**0.9996**	**0.3967**	**0.5774**	**0.5808**	**0.5773**	**0.5773**
	$X_2 \Rightarrow X_4$			0.0212	**0.5365**	0.0216	0.0323
	$X_4 \Rightarrow X_3$	0.0236	0.0179	0.0017	**0.5550**	0.0003	0.0025
	$X_3 \Rightarrow X_4$			**0.2119**	**0.3476**	0.0000	0.0000

Under this framework, PDC is estimated by fitting a VAR model to the data $\mathbf{X}(t), t = 1, \ldots, T\}$ where the optimal order L can be selected using some objective criterion such as the Akaike information criterion (AIC) or the Bayesian-Schwartz information criterion (BIC).

Example 8 (Connectivity network comparison)**.** Consider a system with four channels that can be described by the following sparse VAR(2) model:

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \end{pmatrix} = \begin{pmatrix} \phi_{\beta,1} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \phi_{\delta,1} & 0 \\ 0 & 0 & 0 & \phi_{\gamma,1} \end{pmatrix} \begin{pmatrix} X_1(t-1) \\ X_2(t-1) \\ X_3(t-1) \\ X_4(t-1) \end{pmatrix} + \begin{pmatrix} \phi_{\beta,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \phi_{\delta,2} & 0 \\ 0 & 0 & 0 & \phi_{\gamma,2} \end{pmatrix} \begin{pmatrix} X_1(t-2) \\ X_2(t-2) \\ X_3(t-2) \\ X_4(t-2) \end{pmatrix} + \varepsilon(t)$$

where $(\phi_{\delta,1}, \phi_{\delta,2})$, $(\phi_{\gamma,1}, \phi_{\gamma,2})$ and $(\phi_{\beta,1}, \phi_{\beta,2})$ were calculated as described in Eq. 10 with $M = 1.049787$ and $\psi_\delta = \frac{2}{128}$, $\psi_\gamma = \frac{40}{128}$ and $\psi_\beta = \frac{20}{128}$, respectively. The noise $\varepsilon(t) \sim \mathcal{N}(0, \Sigma)$ where Σ is a diagonal matrix. Consequently, $X_3(t)$ and $X_4(t)$ are independent delta and gamma components.

In this multivariate system, direct and indirect lagged dependence links are denoted: X_3 and X_4 *directly* lead X_2: $X_4(t-2) \Longrightarrow X_2(t)$, $X_3(t-1) \Longrightarrow X_2(t)$, while both *indirectly* affect X_1 through X_2: $X_2(t-1) \Longrightarrow X_1(t)$. Previously, coherence (COH) and partial coherence (PCOH) were applied to identify *instantaneous* direct dependencies. Fig. 16 and Table 1 show the connectivity networks that can be estimated using PDC in addition to both dependence metrics along with their magnitudes.

It is important to note that VAR mismodeling can have a considerable affect on the dependence metrics (as can be observed in Fig. 16.F). However, these effects can be mitigated with regularization techniques that will be discussed in Section 6.3. For a comprehensive empirical analysis of the mismodeling phenomena in connectivity, we refer to (Stephanie, 2017).

5.3. Time-varying PDC

A natural question to ask, when one suspects non-stationarity in the signals, would be how to characterize and estimate PDC when it is changing over time. As noted, during the course of a trial, an experiment or even within an epoch, the brain functional network is dynamic (Ting et al. (2021a)). Thus, one may characterize a time-varying PDC through a time-varying VAR model

$$\Phi_t(B) = I - I - \Phi_{t,1}B - \Phi_{t,2}B^2 - \ldots - \Phi_{t,L}B^L$$

where $\Phi_{t,\ell}$ is the VAR coefficient matrix for lag ℓ at time t. The dimensionality of the parameters for at any time t for a time-varying VAR(L) model is P^2L. Thus, in order to have a sufficient number of observations at any time t, the time-varying coefficient matrices $\{\Phi_{t,\ell}, \ell = 1, \ldots, L\}$ are estimated by fitting a local conditional least squares estimate to a local

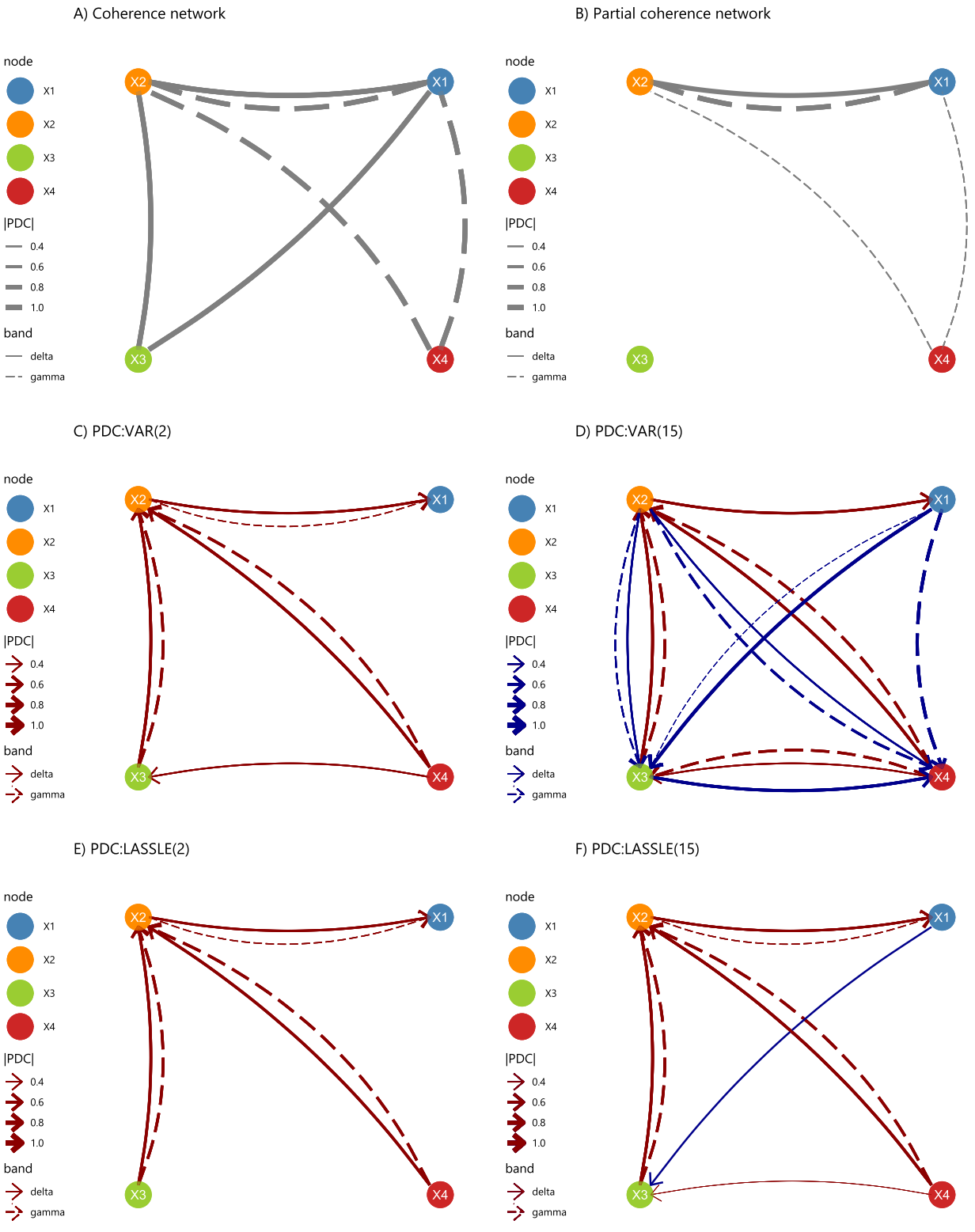

Fig. 16. Connectivity networks for the system in Example 8. Dependence metrics were obtained from six methods: coherence (COH), partial coherence (PCOH), partial directed coherence (PDC) estimated with an OLS-estimated VAR(2) and VAR(15) models, and PDC estimated with a regularized second- and fifteenth-order LASSLE models.

data at time $\{t - \left(\frac{N}{2} - 1\right), \ldots, \frac{N}{2}\}$. In addition to borrowing information from observations within a window (to increase the number of observations), it is also advisable to apply some regularization which will be discussed in Section 6.3. After obtaining estimates $\{\widehat{\Phi}_{t,\ell}\}$, we compute the estimate of the time-varying transfer function $\widehat{\Phi}(t,\omega)$, which then leads to a time-varying PDC estimate

$$\pi_{pq}(t,\omega) = \frac{|\Phi_{pq}(t,\omega)|^2}{\sum_{rq}|\Phi_{rq}(t,\omega)|^2}.$$

PDC has provided the important frequency-specific information of directionality from one component to another - which is already beyond what coherence, partial coherence, and dual-frequency coherence can offer. One can compare the PDC from X_q to X_p of the alpha-band vs. the gamma-band since this relative strength is a directed flow of communication that can vary across frequency oscillations.

5.4. Spectral causality

There are limitations with PDC as a measure of dependence. First, it does not indicate the phase or the physical time lag between oscillations. While it is useful to know the direction $X_q \Longrightarrow X_p$ at frequency ω, it would be crucial to identify the time lag $\tau_{pq,\omega}$ via some relationship such as

$$X_{p,\omega}(t) = A_{pq,\omega}X_{q,\omega}(t - \tau_{pq,\omega}) + \epsilon_{p,\omega}(t)$$

where the coefficient $A_{pq,\omega}$ and time lag $\tau_{pq,\omega}$ can vary depending on the channels and also on the frequency (or frequency bands). Second, the classical PDC captures directionality only for the *same* frequency (or frequency bands). It only models how the past of ω-oscillation in channel X_q could impact the future ω-oscillation in channel X_p. This is a limitation because, in practice, it would be more desirable to capture between-frequencies directionality (as in the dual-frequency setting). Under our proposed framework for modeling dependency via oscillations, we propose to study

$$X_{p,\omega_1}(t) = A_{pq,\omega_1}X_{q,\omega_1}(t - \tau_{pq,\omega_1}) + A_{pq,\omega_2}X_{q,\omega_2}(t - \tau_{pq,\omega_2}) + \epsilon_{p,\omega_1}(t).$$

The third limitation of the classical PDC is that it captures only the linear associations between the oscillations. We overcome the first and second limitations through the spectral vector autoregressive (Spectral-VAR) model. This current form of the model is linear. However, (Pinto et al., 2021b), introduce non-linear variants that are based on biophysical models.

Consider the situation where the goal is to investigate the potential causality from channel X_q to the gamma-oscillation of channel X_p. As used in previous examples, we will denote the delta, theta, alpha, beta, and gamma oscillations of X_q to be, respectively, $X_{q,\delta}, X_{q,\theta}, X_{q,\alpha}, X_{q,\beta}$ and $X_{q,\gamma}$ The oscillations for channel X_p are denoted in a similar manner. The key distinction here is that these oscillations are obtained from a **one-sided** filter $X_{q,\delta}(t) = \sum_{j=0}^{\infty} C_{\delta,j}X_q(t-j)$. Note that the one-sided feature is needed in order to properly capture these lead-lag relationships.

Example 9 (Two-sided filter lead-lag distortion)**.** Let $Z_\delta(t)$, $Z_\beta(t)$ and $Z_\gamma(t)$ be three latent oscillatory signals with main frequencies at 2, 15 and 30 Hertz, respectively. Now assume that two signals are observed:

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} Z_\delta(t) & & \\ Z_\delta(t-10) & +Z_\beta(t-10) & +Z_\gamma(t-10) \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix}$$

where $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are uncorrelated white noise with unit variance. Assume a one-sided and two-sided FIR(100) filter and denote their output filtered signals as $X_{k,\delta}^{(1-sided)}$ and $X_{k,\delta}^{(2-sided)}$ for a given channel X_k. Even though coherence between the delta-filtered $X_1(t)$ and $X_2(t)$ maintains a reasonable similar magnitude, observe in (Fig. 17(C-D) that the lead-lag relationship is not preserved.

As noted, under the proposed framework, one can study more complex dependence strucures. In particular, the spectral-VAR model of order L for predicting the gamma-oscillation of X_p is defined to be

$$\begin{aligned} X_{p,\gamma}(t) = &\sum_{\ell=1}^{L}\left\{A_{(p,\gamma),(p,\delta),\ell}X_{p,\delta}(t-\ell) + A_{(p,\gamma),(q,\delta)}X_{q,\delta}(t-\ell)\right\} + \ldots \\ &+ \sum_{\ell=1}^{L}\left\{A_{(p,\gamma),(p,\gamma),\ell}X_{p,\gamma}(t-\ell) + A_{(p,\gamma),(q,\gamma)}X_{q,\gamma}(t-\ell) + \epsilon_{p,\gamma}(t)\right\}. \end{aligned}$$

Under this set-up, we say that there is a Granger-causality relation from the past alpha-oscillatory activity in X_q to the future gamma-oscillation in X_p if there exists a time lag $\ell*$ such that $A_{(p,\gamma),(q,\alpha),\ell*} \neq 0$.

As a final remark, the usual VAR model does not address the need for frequency-specific lead-lag (or Granger-causality) relationships. Suppose that, from the usual VAR model, we conclude that $X_q \Longrightarrow X_p$. This information is helpful but not sufficiently comprehensive in the sense that it lacks the information on the specific frequency band (or bands) in X_q that Granger-cause the specific band(s) in the X_p as well as the precise channel-specific and frequency bands-specific time-lag between these oscillations.

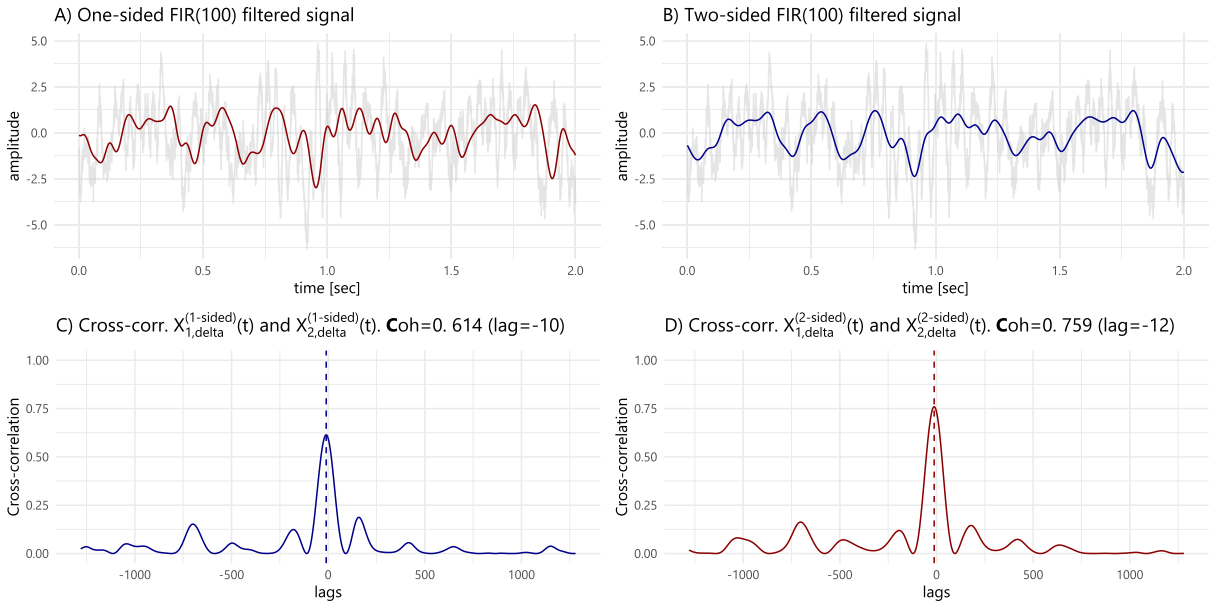

Fig. 17. Effects of an one-sided and two-sided filter in the lead-lag relationship (Example 9).

6. High-Dimensional Signals

Most brain signals are high-dimensional in space and time. For example, fMRI data are typically recorded over 10^5 brain voxels and across hundreds of time points sampled at a speed of one-complete image every 2 seconds (the sampling rate is 0.50 Hertz). Therefore, fMRI data are highly dense in space, though they have poor temporal resolution. In contrast, EEG data are more sparse in space as the number of channels can vary from about 20 to 256, depending on the recording system. However, the range of the number of temporal observations can be in the millions (with sampling frequency typically ranging from $100 - 1000$ Hertz). The goal now is to present two approaches to dealing with high-dimensional signals: (a.) by creating low-dimensional signal summaries such as principal component analysis; and (b.) by including a penalization in estimating parameters in statistical models.

6.1. Spectral principal components analysis

It is often the case that different components of a P-variate signal X exhibit some level of multi-collinearity, especially for signals recorded at adjacent locations. It is therefore natural to seek a low-dimensional summary that captures the main characteristics of these signals. One way is through the classical principal component analysis (PCA) – or linear auto-encoder/decoder in machine learning jargon – which is described as follows.

The auto-encoder algorithm described in (Rumelhart et al., 1988) is a general approach to learning compressed (low-dimensional) representations of the input data, which in this particular scenario are high-dimensional brain signals. The algorithm consists of two parts: (a.) encoder and (b.) decoder. The encoder function $F : \mathbf{X} \to \mathbf{Y}$ is a mapping from the original high-dimensional space $\mathbf{X}$ to a lower dimension space $\mathbf{Y}$. Since F effectively summarizes the information in $\mathbf{X}$ then it effectively reduces dimensionality and hence F is also called compression step. The decoder or reconstruction function, defined as $D : \mathbf{F} \to \mathbf{X}$ is a mapping from the encoded (low-dimensional) space to the original high-dimensional space. Consider a time series $\{\mathbf{X}(t), t = 1, \ldots, T\}$ generated by some process $p(\mathbf{X})$. The optimal mapping encoder F and decoder D jointly minimize the expected reconstruction error defined as

$$L(F, D) = \mathbb{E}_{p(\mathbf{X})} \|\mathbf{X} - D(F(\mathbf{X}))\|_F^2 \tag{22}$$

where $\|\cdot\|_F$ is the Frobenius norm, which is defined as $\|\|E\|\|_F = \sqrt{\text{Trace}(EE')}$.

Here, we will consider only the special cases where both the encoder and decoder are linear transformations of the original signal, which can be either instantaneous mixing or filtering. For these types of functions, the solution is closely related to principal component analysis (PCA) defined under the Frobenius norm based on the squared error of reconstruction.

Now consider the first family of encoder-decoders: instantaneous mixture processes of the observed signal $\{\mathbf{X}(t),\ t = 1, \ldots, T\}$. Under this model, the compressed signal is obtained as $\mathbf{Y}(t) = \mathbf{A}'\mathbf{X}(t)$. Denote the dimension of $\mathbf{Y}(t)$ to be Q; the dimension of $\mathbf{X}(t)$ to be P and $Q < P$. Similarly, the signal is reconstructed by the decoder $\widehat{\mathbf{X}}(t) = \mathbf{B}\mathbf{Y}(t)$. For identifiability purposes, $A'A = I_Q$, and $\mathbb{C}\text{ov}\,\mathbf{Y}(t)$ is diagonal so that the components of the compressed signal X are uncorrelated. The op-

timal low-dimensional representation maximizes the best reconstruction accuracy (or minimizes the squared error loss) via the following steps:

Step 1. Obtain the eigenvalues-eigenvectors of the covariance matrix of $\mathbf{X}$: $\Sigma^{\mathbf{X}}(0)$. Denote e-value and e-vector pair as $\{(\lambda_p, \mathbf{v}_p)\}_{p=1}^{P}$ where $\lambda_1 > \ldots, > \lambda_P$ and $\|\mathbf{v}_p\| = 1$ for all p. When $\Sigma^{\mathbf{X}}(0)$ is not known, we obtain an estimator from the observed signal $\{\mathbf{X}(t),\ t = 1, \ldots, T\}$.

Step 2. The solution for the optimal encoder-decoder is derived as

$$\widehat{A}' = \widehat{B} = [\mathbf{v}_1, \ldots, \mathbf{v}_Q] \quad \text{and} \quad \widehat{\mathbf{Y}}(t) = \widehat{A}\mathbf{X}(t).$$

Under the squared reconstruction error loss function, the solution is identical to applying PCA on the input signals using the covariance matrix at lag zero. Indeed, the solution accounts for most of the variation of the time series (or gives the minimum squared reconstruction error), among all instantaneous linear projections with the same dimension.

When the goal is to obtain summaries from time series data, it is important for the encoded (compressed) and decoded (expanded) components to capture the entire temporal dynamics (lead-lag structure) of the signal. The previous approach is a contemporaneous mixture and thus could miss important dynamics in the data. The second category of linear encoder-decoders relies on the idea of applying linear filters on $\{\mathbf{X}(t)\}$ instead of applying a merely instantaneous mixture. In contrast to the contemporaneous mixture, this encoder is more flexible and its lower-dimensional representation is written as

$$\mathbf{Y}(t) = \sum_{\ell=\infty}^{\infty} A'(\ell)\mathbf{X}(t-\ell) \tag{23}$$

where $A'(\ell) \in \mathbb{C}^{Q\times P}$ with $Q < P$. The components of the summarized signal, $Y_p(t)$ and $Y_q(t)$, have zero coherency. That is, the components $Y_p(t)$ and $Y_q(s)$ are uncorrelated at all time points t and s, and hence also for all lags. The reconstruction (decoder) function has the following form

$$\widehat{\mathbf{X}}(t) = \sum_{\ell=-\infty}^{\infty} B(\ell)\mathbf{Y}(t-\ell) \tag{24}$$

where $B(\ell) \in \mathbb{C}^{P\times Q}$ is the transformation coefficient matrix.

The optimal values of $A(\ell)$ and $B(\ell)$ are chosen to minimize the reconstruction error defined in Eq. 22. The solution is obtained via principal components analysis of the spectral matrix $\mathbf{f}(\omega)$ of the process $\mathbf{X}$ – rather than the lag-0 covariance matrix. Denote the eigenvalues of the spectral matrix at frequency ω to be $\{\lambda_1(\omega) > \lambda_2(\omega), \ldots, \lambda_P(\omega)\}$, and the corresponding eigenvectors to be $\{\mathbf{v}_1(\omega), \mathbf{v}_2(\omega), \ldots, \mathbf{v}_n(\omega)\}$. Then, the solution is

$$A(\ell) = \int_{-1/2}^{1/2} A(\omega)\exp(i2\pi\ell\omega)d\omega \tag{25}$$

$$B(\ell) = \int_{-1/2}^{1/2} B(\omega)\exp(i2\pi\ell\omega)d\omega \tag{26}$$

where $B(\omega) = A^*(\omega)$ and $A(\omega) = [\mathbf{v}_1(\omega), \ldots, \mathbf{v}_Q(\omega)]$.

This dimension reduction procedure was originally described in (Brillinger, 1964) and here we shall refer to this as the "Spectral-PCA" method (SPCA). This method was extended to various nonstationary settings, including Ombao et al. (2005b), where the stochastic representation of a multichannel signal was selected from a library of orthogonal localized Fourier waveforms (SLEX). In Ombao and Ringo Ho (2006), the time-varying spectral PCA was developed under the context of the Priestley-Dahlhaus model, which was further refined in Motta and Ombao (2012) for the experimental setting where there are replicated multivariate nonstationary signals. In practical data analysis, the interest is on the magnitude of the components of the eigenvector (or eigenvectors) with the largest eigenvalues because they represent the loading or weights given by the components of the observed signal. We motivate this in the example below.

Example 10 (Example on spectral PCA)**.** Suppose that we have latent sources that are oscillations at the delta, alpha, and gamma bands, denoted by Z_δ, Z_α and Z_γ, respectively. The observed signal $\mathbf{X}(t)$ is a mixture of the latent sources

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \\ X_4(t) \\ X_5(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} Z_\delta(t) \\ Z_\alpha(t) \\ Z_\gamma(t) \end{pmatrix} + \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \\ \epsilon_3(t) \\ \epsilon_4(t) \\ \epsilon_5(t) \end{pmatrix}. \tag{27}$$

The first two compressed components of the instantaneous mixture (PCA) and the spectral PCA (SPCA) are shown in Fig. 18. Note that all three first PCA principal components capture a mixture of the latent sources, and therefore, contain the delta, alpha, and gamma oscillations in different proportions. This oscillation mixture is made evident from its loadings (Fig. 19.A) where $X_1, \ldots, X_6$. Nevertheless, the first component emphasizes the delta and gamma sources, whereas the second component highlights the alpha and gamma bands.

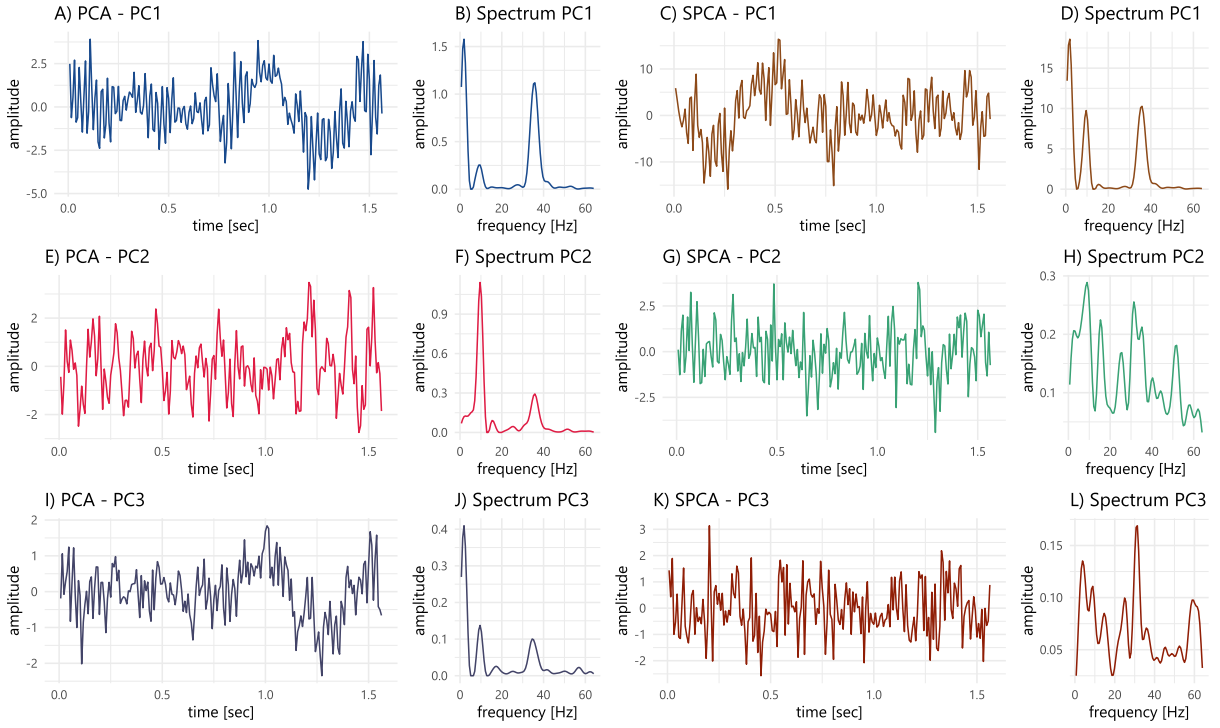

Fig. 18. Summary signals $\mathbf{Y}(t)$ using PCA and SPCA (Example 10). The components in the time and frequency domain are shown for both encoding algorithms.

On the other hand, the first SPCA component captures all the main spectral information of the latent oscillations (Fig. 18.D). Due to the properties of SPCA, mean loadings can be evaluated as a function of the frequency, and the contribution of the components can be evaluated for each frequency band (through the loadings in Fig. 19.B-D). For instance, consider the encoding processes on the alpha band: the main descriptive components are X_4 and X_5, but SPCA compensates the effects of the other signals in such a way that the sum of X_1 and X_2 (delta and gamma) will mitigate the effect of X_3 (also delta and gamma).

6.2. Brain connectivity analysis through low-dimensional embedding

As noted, neuronal populations behave in a coordinated manner both during resting-state and while executing tasks such as learning and memory retention. One of the major challenges to modeling connectivity in brain signals is the high dimensionality. In the case of fMRI data with $V = 10^5$ voxels, one would need to compute connectivity from in the order of 10^{10} pairs of voxels (or $V(V-1)/2$ pairs). To alleviate this problem, one approach in fMRI is to parcellate the entire brain volume into distinct regions of interest (ROIs) and hence connectivity is computed between ROIs rather than between voxels. This approach effectively reduces the dimensionality since connectivity is computed between broad regions rather than at the voxel level. This is also justified by the fact that neighboring voxels tend to behave similarly and thus it would be redundant to calculate connectivity between all pairs of voxels. Motivated by the ROI-based approach in fMRI, one procedure to study connectivity in EEG signals is to first create groups of channels using some anatomical information. Depending on the parcellation of the brain cortex, a cortical region could correspond to 15-25 EEG channels (Wu et al., 2014b). Within each group, we compute the signal summaries using spectral principal components analysis. In the second step, we model connectivity between groups of channels by computing dependence between the summaries. More precisely, suppose that the P-dimensional time series $\mathbf{X}$ is segmented into R groups denoted by $\mathbf{X}_r$, $r = 1, \ldots, R$. In each group $\mathbf{X}_r$, summaries are computed, which we denote by $\mathbf{Y}_r$. Thus, connectivity between groups r and r' will be derived from the summaries $\mathbf{Y}_r$ and $\mathbf{Y}_{r'}$

There are many possible methods for computing summaries. When biomedical signals can be modeled as functional data, principal components analysis (PCA) extensions have been formulated, as it is shown in (Kokoszka et al., 2019). Under sampled time series, a naïve solution is to compute the average across all channels within a group. In fact, connectivity analyses of functional magnetic resonance imaging (fMRI) are usually conducted by taking the time series averaged across voxels in pre-defined ROIs ((Fiecas et al., 2013; Gott et al., 2015)). However, simple averaging is problematic, especially when some of the signals are out of phase often observed in EEGs due to averaging can lead to signal cancellation. Sato et al. (Sato et al., 2010) already pointed the pitfalls and suggests a data-driven approach via conventional PCA, which essentially

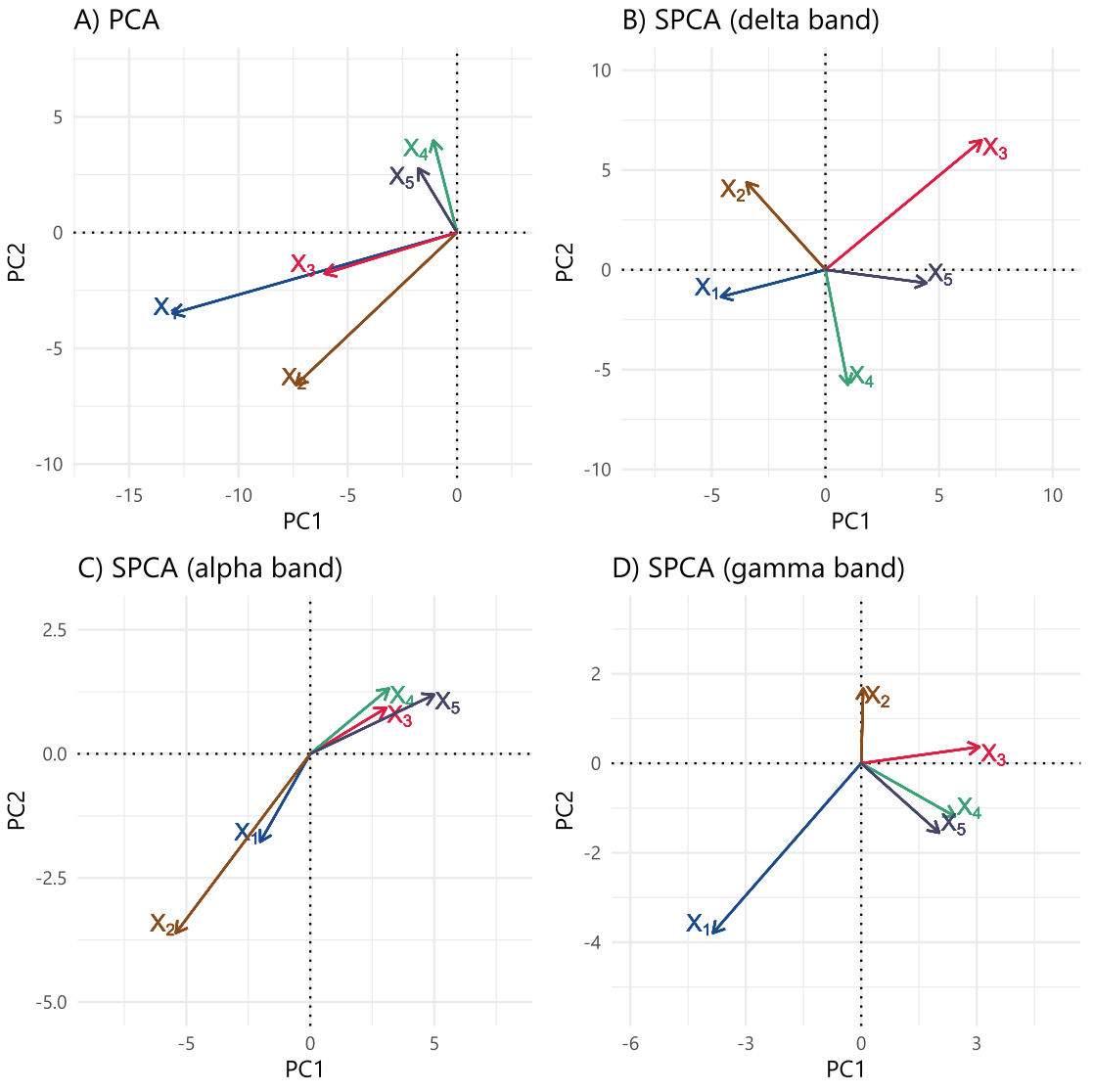

Fig. 19. Two-dimensional loadings of Example 10: PCA loadings ($\widehat{A}$) and mean SPCA loadings ($A(\omega)$) in the frequency bands: delta (0-4 Hertz), alpha (8-12 Hertz), and gamma (30-45 Hertz).

provides an instantaneous (or contemporaneous) mixing of time series. Other approaches for modeling brain connectivity from high-dimensional brain imaging data include Dynamic Connectivity Regression (DCR) (Cribben et al., 2012b), Dynamic Conditional Correlation (DCC) (Lindquist et al., 2014), group independent component analysis (ICA) (Calhoun et al., 2014; Calhoun and Adalı, 2012), and sparse vector autoregressive (VAR) modeling (Davis et al., 2016). Here, we propose to extract summaries $\mathbf{Y}_r$ from each group of channels via the spectral PCA method in Eqs. 23 and 24 above.

6.3. Regularized vector autoregressive models

Recall that connectivity measures such as coherence, partial coherence and partial directed coherence are based on the frequency domain but they can be motivated under the context of parametric models. In fact, PDC was developed within the framework of a vector autoregressive (VAR) model. Here, we discuss the challenges of fitting a VAR model when the number of channels P and the VAR order D are large. In this setting there will be P^2D number of VAR parameters that have to be estimated. The goal here is to introduce some regularization in the estimation procedure.

The classic method for estimating the VAR parameters is via the least squares estimator (or conditional likelihood for Gaussian signals), which is generally unbiased. However, the least-squares estimators (LSE) are problematic because of the high computational demand and that it does not possess specificity for coefficients whose true values are zero. In many ap-

plications, brain networks are high-dimensional structures that are assumed to be sparse but interconnected. To address the problem of high dimensional parameter space, a common estimation approach is by penalized regression (Tibshirani (1996), Fu (1998), Zhao et al. (2009), Hesterberg et al. (2008)), and a specific method is the LASSO (least absolute shrinkage and selection operator). Compared to the LSE approach, the LASSO has a lower computational cost (Mairal and Yu (2012)) and has higher specificity of zero-coefficients. However, the main limitation of the LASSO (and, in general, most regularization methods) is the bias of the non-zero coefficients' estimators. Thus, it could lead to misleading results when investigating the true strength of brain connectivity. By leveraging the strengths of each of the LSE and the LASSO, Hu et al. (Hu et al., 2019) proposes a (hybrid) two-step estimation procedure: the LASSLE method. This approach suggest a two-phase estimation process. In the first stage, the LASSO is applied to identify coefficients whose estimates are set to 0. In the second stage, the coefficients that survived the thresholding from the first stage are re-estimated via LSE. LASSLE was shown to have inherent low-bias for non-zero estimates, high specificity for zero-estimates, and significantly lower mean squared error (MSE) in the simulation study.

An example of the impact of using regularized models in connectivity estimation was visually shown in Fig. 16 E-F (from Example 8).

7. Modeling dependence in non-stationary brain signals

Contrary to common intuition, the brain is active and dynamic even during resting period (Menon and Krishnamurthy, 2019). Brain signals from various modalities, such as magnetoencephalograms, electroencephalograms, local field potentials, near-infrared optical signals, and functional magnetic resonance imaging, all show statistical properties that evolve during rest period and also in various task-related settings (memory, somatosensory, audio-visual). Such time-evolving characteristics are also observed across species: laboratory rats, macaque monkeys, and humans. These changes are seen in the variance, auto-correlation, cross-correlation, coherence, partial coherence, partial directed coherence, or graph-network properties (Chiang et al., 2016). It is worth noting that in some experimental settings, changes in the cross-channel dependence may be more pronounced than changes within a channel (e.g., auto-spectrum and variance). This phenomenon was observed in (Fiecas and Ombao, 2016), where changes in cross-coherence were observed and prominent than changes in the auto-spectrum in a macaque monkey local field potentials and correlated with the learning task.

In this section, an overview of the different approaches to analyzing non-stationary signals will be discussed. The first class of approaches gives stochastic representations using the Fourier waveforms or some multi-scale orthonormal basis such as wavelets, wavelet packets, and the smoothed localized complex exponentials (SLEX). In the second class of methods, the signals are segmented into quasi-stationary blocks, and the time-varying spectral properties such as the auto-spectra, coherence, and partial coherence are computed within each time block. This class of approaches produces a specific tiling of the time-frequency plane. The third class of approaches assumes that the dynamic brain activity fluctuates or switches between a finite number of "states". Each state is uniquely defined by a vector autoregressive model or a stochastic block model and thus gives a unique characterization of the brain functional network. This class of models depicts brain responses to a stimulus (or background activity during resting state) as switching between these states.

7.1. Stochastic representations

Priestley-Dahlhaus model. The major theme in this paper is the characterization of brain signals as mixtures of randomly oscillating waveforms. So far, the emphasis has been on stochastic representations in terms of the Fourier waveforms. As already noted, the Cramér representation of a P-dimensional stationary time series $\mathbf{X}_t$

$$\mathbf{X}(t) = \int_{-0.50}^{0.50} \mathbf{A}(\omega) \exp(i2\pi\omega t) d\mathbf{Z}(\omega)$$

where $\mathbf{A}(\omega)$ is a $P \times P$ the transfer function matrix and $\{d\mathbf{Z}(\omega)\}$ is a random increment process with $\mathbb{E}d\mathbf{Z}(\omega) = 0$ and $\mathbb{C}\text{ov}[d\mathbf{Z}(\omega), d\mathbf{Z}(\omega')] = \mathbf{I}\delta(\omega - \omega)$ where δ is the Dirac-delta function. To illustrate the role of the transfer function, define $A_{pq}(\omega)$ to be the (p, q) element of $\mathbf{A}(\omega)$; and $dZ_p(\omega)$ to be the p-th element of the random vector $d\mathbf{Z}(\omega)$. The spectral matrix is $\mathbf{f}(\omega) = \mathbf{A}(\omega)\mathbf{A} * (\omega)$. The univariate time series at channels p and q can be written as

$$X_p(t) = \sum_{r=1}^{P} \int_{-0.50}^{0.50} A_{pr}(\omega) \exp(i2\pi\omega t) dZ_r(\omega) \text{ and } X_q(t) = \sum_{r=1}^{P} \int_{-0.50}^{0.50} A_{qr}(\omega) \exp(i2\pi\omega t) dZ_r(\omega).$$

They are coherent at frequency ω if there exists some r where $A_{pr}(\omega) \neq 0$ and $A_{qr}(\omega) \neq 0$. In this case, $\mathbf{A}_p(\omega)\mathbf{A}_q^*(\omega) \neq 0$ (where $\mathbf{A}_r(\omega)$ is the r-th row of the transfer function matrix $\mathbf{A}(\omega)$. Under stationarity, spectral cross-dependence between signals (e.g., coherence) is constant over time. For brain signals, dependence between channels varies across time. A model for non-stationary time series is the time-dependent generalization of the Cramér representation which we refer in this paper as the Priestly-Dahlhaus model (Priestley (1965) and Dahlhaus (1997)). For a time series of length T, the Dahlhaus-Priestley model uses a time-dependent transfer function which gives a representation

$$\mathbf{X}(t) = \int_{-0.50}^{0.50} \mathbf{A}(t/T, \omega) \exp(i2\pi\omega t) d\mathbf{Z}(\omega).$$

where the transfer function $\mathbf{A}(\frac{t}{T},\omega)$ is defined on rescaled time $\frac{t}{T}\in(0,1)$ and frequency $\omega\in(-0.50,0.50)$. Under this model, the mixture changes over time because the random coefficient vector $\mathbf{A}(\frac{t}{T},\omega)\ d\mathbf{Z}(\omega)$ also changes with time. The time-varying spectrum is $\mathbf{f}(u,\omega)=\mathbf{A}(u,\omega)\mathbf{A}*(u,\omega)$ and, consequently, the time-varying coherence between channels p and q is

$$\rho_{pq}(u,\omega)=\frac{|f_{pq}(u,\omega)|^2}{f_{pp}(u,\omega)f_{qq}(u,\omega)}.$$

To estimate the time-varying spectrum at a particular rescaled time $u\in(0,1)$, local observations around this time point are used to form local periodograms, which are then smoothed over frequency. Alternatively, one can fit a localized semi-parametric estimator as shown in Fiecas et al. (2011). Note that a change-point detection extension was introduced in (Górecki et al., 2018).

Locally stationary wavelet process. An alternative to the Priestley-Dahlhaus model is the locally stationary wavelet process (LSW) proposed in Nason et al. (2000), which uses the wavelets as building blocks. Under the LSW model, a scale- and time-dependent wavelet spectrum is defined. The original LSW model is univariate and has been extended to the multivariate setting in Park et al. (2014) as follows. A time series $\mathbf{X}(t), t=1,2,\ldots,T$ is a multivariate LSW process if it has the representation

$$\mathbf{X}(t)=\sum_j\sum_k\mathbf{V}_j(\frac{k}{T})\psi_{j,t-k}\mathbf{Z}_{jk}$$

where $\{\psi_{j,t-k}\}$ is a set of discrete non-decimated wavelets; $\mathbf{V}_j(\frac{k}{T})$ is the transfer function matrix which is lower-triangular; and $\mathbf{Z}_{jk}$ are uncorrelated zero mean with covariance matrix equals to the identity matrix. Note that $\mathbf{Z}_{jk}$ in the multivariate LSW is the analog of $d\mathbf{Z}(\omega)$ in the Fourier-based stochastic representation. The classification procedure for the LSW model was developed in Fryzlewicz and Ombao (2009) for univariate time series and in Park et al. (2018) for multivariate time series. Given training data (signals with known group membership), these methods extract the wavelet scale-shift features or projections that separate the different classes of signals. These features are then used to classify a test signal. These wavelet-based classification methods are demonstrated to be asymptotically consistent, i.e., the probability of misclassification decreases to zero as the length of the test signal increases. **The SLEX model.** There are other time-localized bases that are also well-suited for representing non-stationary time series. In particular, the SLEX (smooth localized complex exponentials) waveforms are ideal for a comprehensive analysis that is a time-dependent generalization of Fourier-based methods. The SLEX waveforms are time-localized versions of the Fourier waveforms and they are constructed by applying a projection operator (Ombao et al. (2005a)). The starting point is to build the SLEX library which is a collection of many bases. These bases consist of functions defined on a dyadic support on rescaled time. In Fig. 20, see the plot of a specific SLEX waveform with support on the second quarter of the rescaled time and with approximately 2 complete cycles within that block. In this example, a tree is grown to level $J=2$ (i.e., the finest blocks have support with width $\frac{1}{2^2}$). There are 5 bases in this library and each basis corresponds to a specific dyadic segmentation of [0,1]. One particular basis is represented by the magenta-colored blocks which are denoted by $B(2,0), B(2,1), B(1,1)$. This basis also corresponds to the specific segmentation $[0,\frac{1}{4})\cup[\frac{1}{4},\frac{1}{2})\cup[\frac{1}{2},1]$. Define $\mathcal{B}$ to be a set of indices (j,b) that make up one particular basis. In this particular example, $\mathcal{B}=\{(2,0),(2,1),(1,1)\}$. Denote an SLEX waveform with support on block $B(j,b)$ oscillating at frequency ω is denoted as $\Psi_{j,b,\omega}(t)$. Then the SLEX model corresponding to this particular basis $\mathcal{B}$ is

$$\mathbf{X}(t)=\sum_{(j,b)\in\mathcal{B}}\int_{-0.50}^{0.50}\Theta_{j,b}(\omega)\psi_{j,b,\omega}(t)d\mathbf{Z}_{j,b}(\omega) \tag{28}$$

where $\Theta_{j,b}(\omega)$ is the SLEX transfer function defined on time block $B(j,b)$ and $\{d\mathbf{Z}_{j,b}(\omega)\}$ is an increment random process that is orthonormal across time blocks (j,b) and frequencies ω. For a given SLEX basis, the time-dependent SLEX spectral matrix is

$$\mathbf{f}_s(t/T,\omega)=\Theta_{j,b}(\omega)\Theta^*_{j,b}(\omega)$$

for t in the time block $B(j,b)$. Thus, the SLEX auto-spectra and the SLEX-coherence are defined in a similar way as the classical Fourier approach.

The main advantage of the SLEX methods for analyzing non-stationary signals is the flexibility offered by the library of bases – which is a collection of many possible bases. Depending on the particular problem of interest, a "best" basis can be selected from this collection of bases using the Coifman and Wickerhauser best basis algorithm. Each basis in the SLEX library gives rise to a unique segmentation and hence a unique SLEX model (or SLEX representation of the signal). To select the best basis for signal representation, Ombao et al. (2001a), developed penalized Kullback-Leibler criterion. This criterion jointly minimizes: (i.) the error of discrepancy between the empirical time-varying SLEX spectrum and the candidate true SLEX spectrum, and (ii.) the complexity in the SLEX model measured by the number of blocks for each candidate basis. For the problem of modeling high dimensional time series, (Ombao et al., 2005a) developed a procedure for model selection and dimension reduction through the SLEX principal components analysis. In some applications, the goal is to discriminate between classes of signals and to classify the test signal. Under the SLEX framework, there is a rich set of time-frequency features derived from the many potential bases. The SLEX method for discrimination selects the basis that gives the maximal

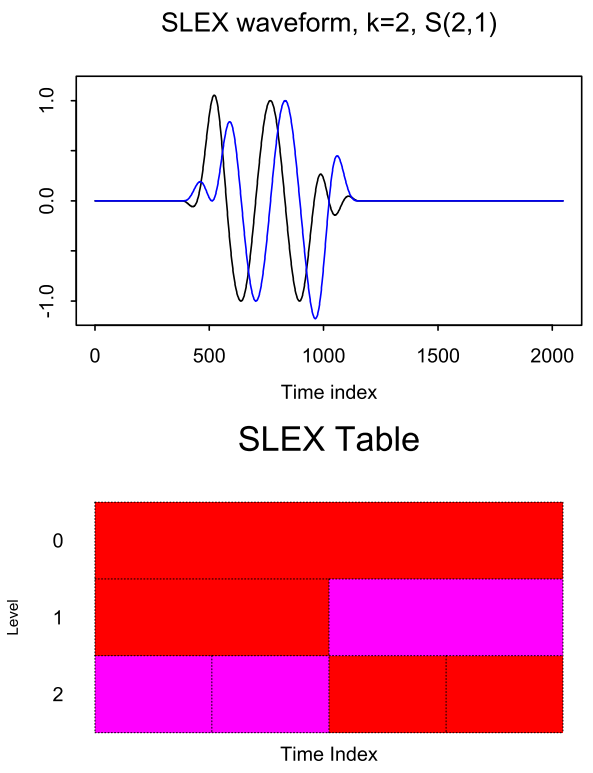

Fig. 20. Top: SLEX waveform (real and imaginary parts) with local support at dyadic block $B(2,1)$. Bottom: SLEX library with level $J = 2$. The SLEX library consists of a finite number of bases and each basis gives a dyadic segmentation of the time axis. The set of magenta-colored blocks represent one particular basis with time blocks $B(2,0)B(2,1), B(1,1)$ which corresponds to the segmentation on rescaled time: $\left[0, \frac{1}{4}\right) \cup bigl[\frac{1}{4}, \frac{1}{2}\right) \cup \left[\frac{1}{2}, 1\right]$.

discrepancy between classes of signals (e.g., signals from healthy controls vs. disease groups). A classification procedure for univariate signals was proposed in (Huang et al., 2004) and for multivariate signals in Böhm et al. (2010). The SLEX method for classification is also demonstrated to be consistent, i.e., the probability of misclassification converges to zero as the length of the test signal increases.

7.2. Methods for change-points analysis

One broad class of methods for analyzing non-stationary signals identifies the change-points and consequently segments the signals into quasi-stationary time blocks. There are many change-points methods and the most relevant to brain signal analysis are discussed here.

One line of segmentation methods is based on dyadic segmentation. In Adak (1998), a procedure for dyadic segmentation of a non-stationary univariate signal is proposed. A signal is split dyadically up to some specified level J so that at each level j there are exactly 2^j. Each block at level j (mother block) is split into two blocks (children blocks). Starting from the deepest level of the tree, the spectra at the adjoining children blocks are estimated and then compared. When they are similar (based on some discrepancy metric), then these children blocks are combined to form one mother block. Otherwise, they are kept as distinct blocks. A large discrepancy between the spectral estimates suggest that there is a change-point. The Adak method also imposes a penalty for complexity in order to prevent a stationary block from unnecessarily splitting into two. The SLEX methods in (Ombao et al., 2005a) and Ombao et al. (2001a) can also be viewed as dyadic segmentation methods even though the segmentation is merely a by-product of the best model (or best SLEX basis) selected from the penalized Kullback-Leibler criterion. Another method based on a similar dyadic division is presented in (Pinto-Orellana and Hammer, 2020b) and (Pinto et al., 2021a): the dyadic aggregated autoregressive model (DASAR). In this representation, where each dyadic block at level j is modeled using an aggregated autoregressive model (with degrees of freedom in the number of maximum components and the maximum order of each autoregressive component). Rules for its tree expansion are similar to SLEX. By construction, DASAR blocks represents one resonating component through two parameters: central frequency and bandwidth (exponential of the negative root modulus M in Eq. 10). This compact representation shows notable relevance for classification purposes (Pinto-Orellana and Hammer, 2020b) and frequency tracking (Pinto et al., 2021a).

For completeness, we also enumerate the other interesting change-point methods that have been either applied to brain signals or are potentially useful in analyzing brain signals: binary segmentation for transformed autoregressive conditionally heteroscedastic models in Fryzlewicz and Subba-Rao (2014); group lasso in Chan et al. (2014); score-type test statistics for

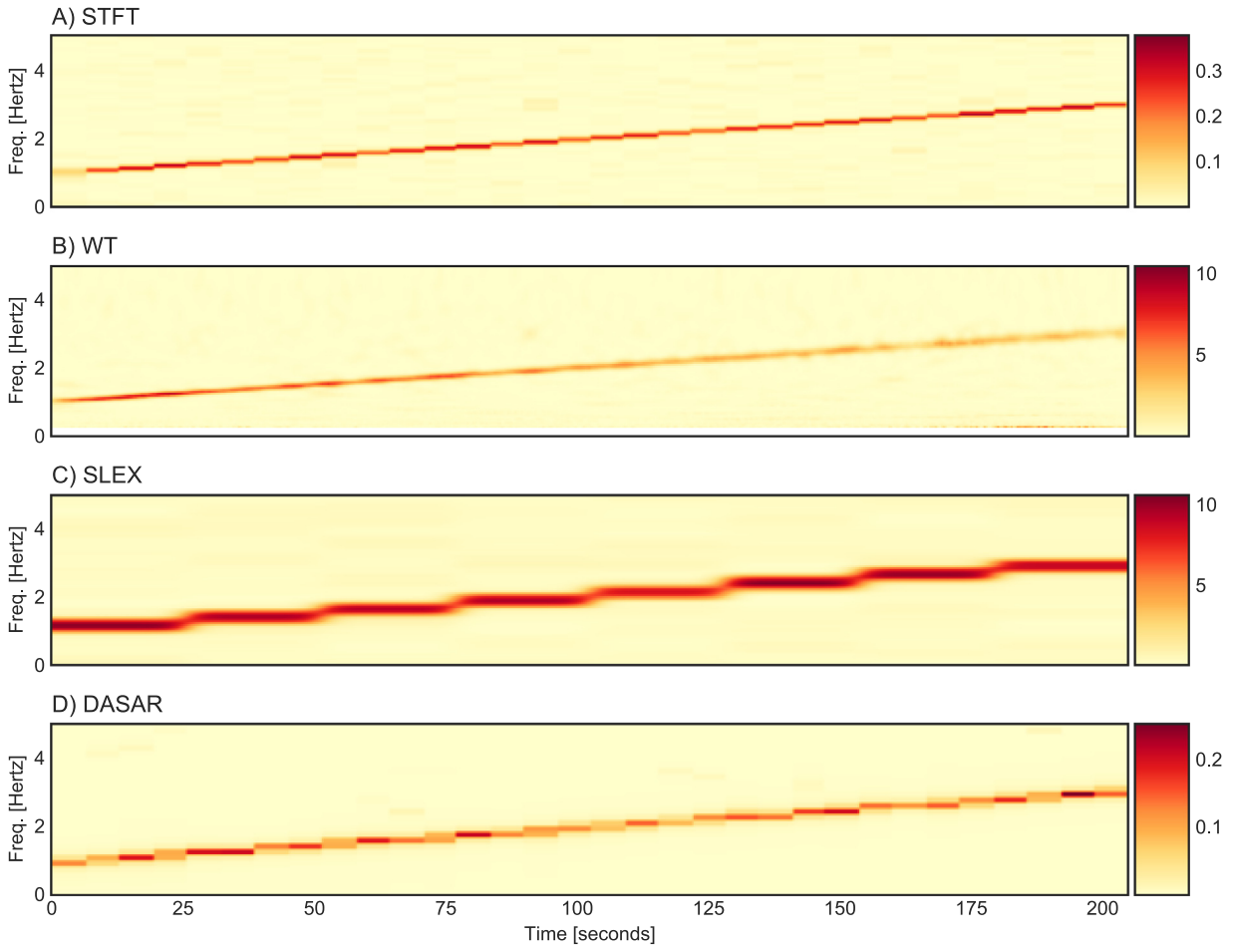

Fig. 21. Spectral estimation of a linearly frequency evolving signal using four representation methods: short-time Fourier transform, wavelet transform, SLEX and DASAR.

VAR-based models in Kirch et al. (2015); test of non-stationarity based on the discrete Fourier transform in Subba-Rao and Jentsch (2015); and the frequency-specific change-point detection (FreSpeD) Schröder and Ombao (2019).

Example 11 (Non-stationary signal with linearly frequency evolution)**.** Consider a non-stationary time series $X(t)$ that has a continuously-evolving frequency, for example, a chirp signal, with a frequency starting at ω_0 and increasing with increment $\Delta\omega$ Hertz each second

$$C(t) = 2\sin\big(2\pi\left(\omega_0 + \Delta\omega t\right)t\big) + \varepsilon(t)$$

where $\varepsilon(t) \sim \mathcal{N}(0, 1)$.

These types of signals have been observed in seismic events (Flandrin, 2018, p.10-11), and also during epileptic episodes as Schiff et al. observed (Schiff et al., 2000). Fig. 21 shows the spectral estimations obtained through short-time Fourier transform, wavelet transform, SLEX and DASAR. Note that the four of them were capable of detecting the time-evolving resonator.

Example 12 (Time-frequency EEG representation)**.** Recall the ADHD-EEG dataset described in Section 2. In Fig. 22, we show the time-varying spectral response of the signals at left and right frontal channels (F3 and F4) in the frequency interval 10–30 Hertz (beta band). Note that the power at the 10-Hertz and 20-Hertz components are changing over time: 20-Hertz oscillations appear stronger at 8-12 seconds and 16-21 seconds with higher magnitudes in the left hemisphere of the brain.

7.3. Switching processes and community detection

Another class of models characterizes changes in brain connectivity networks via a Markov-switching between distinct brain states. Many approaches that use the sliding window or fit some time-varying coefficient models are usually incapable

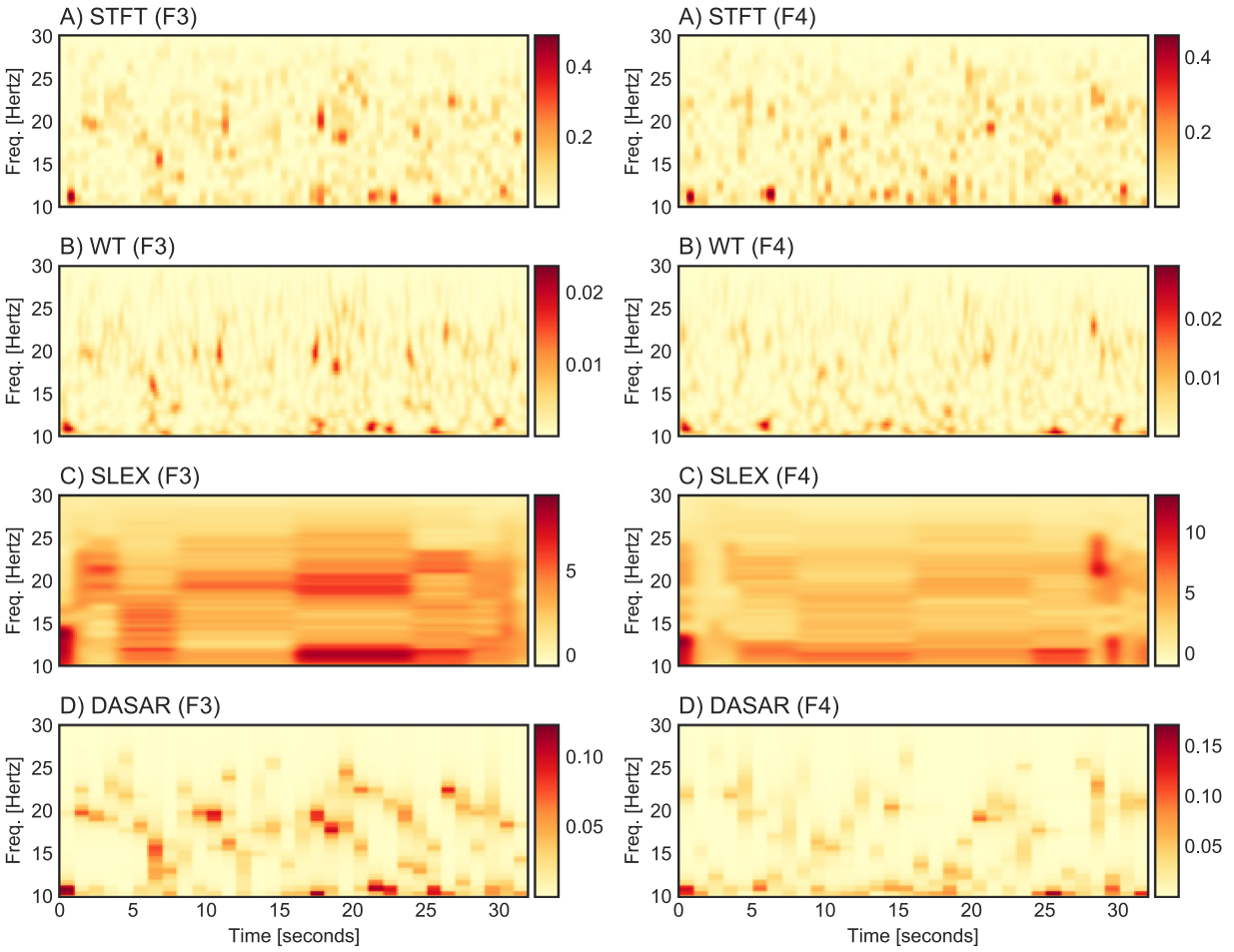

Fig. 22. Time-frequence representation of 32-second EEG recordings from channels on the frontal left (F3) and right (F4) hemisphere using short-time Fourier transform, wavelet transform, SLEX and DASAR.

of capturing abrupt changes. To deal wit high dimensionality of brain signals, a Markov-switching dynamic factor model was proposed in (Ting et al., 2018a). This model captures dynamic connectivity states multivariate brain signals driven by lower-dimensional latent factors. A regime-switching vector autoregressive (SVAR) factor process was used to measure the dynamic directed connectivity. There is also an emerging line of models for brain connectivity that are based on stochastic block models. Indeed there is accumulating evidence that suggests that functional connectivity patterns are dynamically evolving over multiple time scales during rest and while performing a cognitive task. However, functional connectivity tends to be temporally clustered into a finite number of connectivity states. These are distinct connectivity patterns that recur over the course of the experiment. Recent work in (Ting et al., 2021a) characterizes dynamic functional connectivity - specifically the state-driven changes in community organization of brain networks. A key contribution is that the approach takes into account variation across individuals. Many key innovations are being developed in this line of modeling, and they are anticipated to set the trend given the ability of these methods to handle high dimensionality datasets while providing results that are easily interpretable.

7.4. Time-varying dependence between filtered signals

Previously described methods for analyzing frequency-specific dependence between channels $X_p(t)$ and $X_q(t)$ can be extended to the non-stationary setting where this dependence can evolve over time. Recall that $X_{p,\omega}(t)$ and $X_{q,\omega}(t)$ are the ω-oscillations derived from the two channels. Thus, the time-varying correlation between these oscillations is defined to be

$$\rho_{pq,\omega}(t^*) = \mathbb{C}\text{or}\left(X_{p,\omega}(t^*), X_{q,\omega}(t^*)\right).$$

The natural approach for estimating $\rho_{pq,\omega}(t^*)$ is to form a sliding window and then compute the local correlation within each window. This approach is motivated by a time-dependent generalization of the mixture model in Eqs. 2, 3 and 3.4,

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ \dots \\ X_P(t) \end{pmatrix} = \begin{pmatrix} A_{11}(t) & A_{12}(t) & \dots & A_{1K}(t) \\ A_{21}(t) & A_{22}(t) & \dots & A_{2K}(t) \\ \dots & & & \\ A_{P1}(t) & A_{P2}(t) & \dots & A_{PK}(t) \end{pmatrix} \begin{pmatrix} Z_1(t) \\ Z_2(t) \\ \dots \\ Z_K(t) \end{pmatrix}. \tag{29}$$

The above model specifies the mixture matrix to have components $A_{pq}(t)$ that change with time. The time-varying mixture model in Eq. 7.4 can be further generalized by allowing for temporal delay $A_{pq}(t)B^{d_{pq}}$, which will be captured by the backshift operator $B^{d_{pq}}$.

8. R package ECOSTASpecDepRef and data reproducibility

An R package (ECOSTASpecDepRef) was developed to provide a reference implementation for the algorithms discussed in the current paper, as well as the procedures for replicating the results shown in Examples 1–10. The source code is available at the repository https://github.com/biostatistics-kaust/ecosta-spectral-dependence.

9. Conclusion

This paper presents a general framework to modeling dependence in multivariate signals through the oscillatory activities extracted from each channel. Under this framework, classical notions of dependence can be derived: coherence, partial coherence, and dual-frequency coherence. This framework opens up the possibility of exploring more complex measures of dependence. The unifying theme is to explore the strength of dependence and possible lead-lag dynamics through filtering. The proposed framework is sufficiently comprehensive given its capability to model and study both linear and non-linear dependence in both contemporaneous and lagged configurations.

There are numerous open problems in analyzing time series from designed experiments, with the spectral dependence approach, that were not addressed in this paper due to space constraints. Among them, we mention the following possible experimental scenarios: (i.) There could be well-defined labeled groups, with their own dependence characteristics, according to disease (controls vs. ADHD), stimulus types (reading vs. math), cognitive load (easy vs. difficult). (ii.) There could be several participants within each group, and consequently, the models should account for the cross-subject variation of the brain's functional response within groups. (iii.) There could be repeated expositions (or trials) to the same stimulus, and the model should be able to extract the common information across responses while simultaneously modeling trial-wise variations of brain response. (iv.) There could be different temporal scales in the experiment, and the model should take into account transient responses (i.e., within a trial or epoch), long-term responses (across the trials in an experiment) while considering the developmental changes in the human brain on longitudinal scales.

Statistical models should also allow for assessing differences in the fixed effects (e.g., control vs. disease) while also accounting for between-subject variation. A candidate approach can be performed via mixed-effects models, such as the framework proposed in Gorrostieta et al. (2012) and (Chiang et al., 2017), or fully Bayesian approaches, as the methods proposed in Yu et al. (2016) and (Warnick et al., 2018). In addition to the need for formal statistical models, there is also a need to develop new approaches for studying possible non-linear dependence through high-dimensional models. These methods should leverage advances in dimensionality reduction, statistical/machine learning, and optimization in order to examine these complex dependence structures from high-dimensional models. There are also open problems on inference. Some of the inference on these dependence measures may be explicitly addressed under the context of mixed effects models or Bayesian hierarchical models. Some of these might be more suited under resampling-based procedures which will be developed in our future work.

References

Aburto, J.M., Villavicencio, F., Basellini, U., Kjærgaard, S., Vaupel, J.W., 2020. Dynamics of life expectancy and life span equality. Proceedings of the National Academy of Sciences 117 (10), 5250–5259. doi:10.1073/pnas.1915884117.

Adak, S., 1998. Time-dependent spectral analysis of nonstationary time series. Journal of the American Statistical Association 93 (444), 1488–1501.

Allahverdy, A., Nasrabadi, A.M., Mohammadi, M.R., 2011. Detecting ADHD children using symbolic dynamic of nonlinear features of EEG. In: 2011 19th Iranian Conference on Electrical Engineering, p. 1-1.

Avendaño-Valencia, L.D., Chatzi, E.N., 2020. Multivariate GP-VAR models for robust structural identification under operational variability. Probabilistic Engineering Mechanics 60, 103035. doi:10.1016/j.probengmech.2020.103035.

Baccalá, L.A., de Brito, C.S.N., Takahashi, D.Y., Sameshima, K., 2013. Unified asymptotic theory for all partial directed coherence forms. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371 (1997), 20120158. doi:10.1098/rsta.2012.0158.

Baccalá, L.A., Sameshima, K., 2001. Partial directed coherence: a new concept in neural structure determination. Biological cybernetics 84 (6), 463–474.

Bai, P., Safikhani, A., Michailidis, G., 2020. Multiple Change Points Detection in Low Rank and Sparse High Dimensional Vector Autoregressive Models. IEEE Transactions on Signal Processing 68, 3074–3089. doi:10.1109/TSP.2020.2993145.

Balqis, S., Ting, C.-M., Ombao, H., Sh-Salleh, H., 2017. A unified estimation framework for state- related changes in effective brain connectivity. IEEE Transactions on Biomedical Engineering 64, 844–858.

Bassett, D., Wymbs, N., Porter, M., Mucha, P., Carlson, J., Grafton, S., 2011. Improved state change estimation in dynamic functional connectivity using hidden semi-markov models. Proc. Nat. Acad. Sci. USA 108 (18), 7641–7646.

Baum, A., Aguilar-Gomez, S., Lightwood, J., Bruzelius, E., Glantz, S.A., Basu, S., 2020. Estimating the long-run relationship between state cigarette taxes and county life expectancy. Tobacco Control 29 (1), 81–88. doi:10.1136/tobaccocontrol-2018-054686.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., Robbins, K.A., 2015. The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. Frontiers in Neuroinformatics 9. doi:10.3389/fninf.2015.00016.
Blázquez-García, A., Conde, A., Mori, U., Lozano, J.A., 2020. A review on outlier/anomaly detection in time series data. arXiv:2002.04236 [cs, stat].
Böhm, H., Ombao, H., von Sachs, R., Sanes, J., 2010. Classification of multivariate non-stationary signals: The SLEX-shrinkage approach. Journal of Statistical Planning and Inference 140 (12), 3754–3763. doi:10.1016/j.jspi.2010.04.040.
Bowman, F., Caffo, B., Bassett, D., Kilts, C., 2008. A bayesian hierarchical framework for spatial modeling of fmri data. NeuroImage 39, 146–156.
Brillinger, D., 1964. A frequency approach to the techniques of principle components, factor analysis and canonical variates in the case of stationary series. In: Royal Statistical Society Conference, Cardi, Wales.
Brockwell, P.J., Davis, R.A., 1996. Time Series: Theory and Methods, 2nd ed Springer, New York.
Brown, E.N., Lydic, R., Schiff, N.D., 2010. General Anesthesia, Sleep, and Coma. New England Journal of Medicine 363 (27), 2638–2650. doi:10.1056/NEJMra0808281.
Calhoun, V.D., Adalı, T., 2012. Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery. IEEE reviews in biomedical engineering 5, 60–73. doi:10.1109/RBME.2012.2211076.
Calhoun, V.D., Miller, R., Pearlson, G., Adalı, T., 2014. The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. Neuron 84 (2), 262–274. doi:10.1016/j.neuron.2014.10.015.
Castruccio, S., Ombao, H., Genton, M., 2018. A scalable multiresolution spatiotemporal model for brain activation and connectivity in fmri data. Biometrics 74, 823–833.
Chan, N., Yau, C., Zhang, R., 2014. Group lasso for structural break time series. Journal of the American Statistical Association 109 (506), 590–599.
Chen, M., Chow, S.-M., Hammal, Z., Messinger, D.S., Cohn, J.F., 2020. A Person- and Time-Varying Vector Autoregressive Model to Capture Interactive Infant-Mother Head Movement Dynamics. Multivariate Behavioral Research 1–29. doi:10.1080/00273171.2020.1762065.
Chiang, S., Cassese, A., Guindani, M., Vannucci, M., Yeh, H.J., Haneef, Z., Stern, J.M., 2016. Time-dependence of graph theory metrics in functional connectivity analysis. NeuroImage 125, 601–615. doi:10.1016/j.neuroimage.2015.10.070.
Chiang, S., Guindani, M., Yeh, H.J., Haneef, Z., Stern, J.M., Vannucci, M., 2017. Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data: Bayesian Multi-Modal VAR Model. Human Brain Mapping 38 (3), 1311–1332. doi:10.1002/hbm.23456.
Cribben, I., Haraldsdottir, R., Atlas, L.Y., Wager, T.D., Lindquist, M.A., 2012. Dynamic connectivity regression: Determining state-related changes in brain connectivity. NeuroImage 61, 907–920.
Cribben, I., Haraldsdottir, R., Atlas, L.Y., Wager, T.D., Lindquist, M.A., 2012. Dynamic connectivity regression: Determining state-related changes in brain connectivity. NeuroImage 61 (4), 907–920. doi:10.1016/j.neuroimage.2012.03.070.
Cristina, G., Ombao, H., von Sachs, R., 2018. Time-dependent dual frequency coherence in multivariate non-stationary time series. Journal of Time Series Analysis 40, 3–22.
Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. The annals of Statistics 25 (1), 1–37.
Dahlhaus, R., 2012. Locally stationary processes. Handbook of Statistics, Time Series Analysis: Methods and Applications 351–408.
Davis, R.A., Zang, P., Zheng, T., 2016. Sparse Vector Autoregressive Modeling. Journal of Computational and Graphical Statistics 25 (4), 1077–1096. doi:10.1080/10618600.2015.1092978.
Debnath, R., Buzzell, G.A., Morales, S., Bowers, M.E., Leach, S.C., Fox, N.A., 2020. The Maryland analysis of developmental EEG (MADE) pipeline. Psychophysiology 57 (6). doi:10.1111/psyp.13580.
Del Giorno, R., Ottini, A., Greco, A., Stefanelli, K., Kola, F., Clivio, L., Ceschi, A., Gabutti, L., 2020. Peer-pressure and overuse: The effect of a multimodal approach on variation in benzodiazepine prescriptions in a network of public hospitals. International Journal of Clinical Practice 74 (3). doi:10.1111/ijcp.13448.
Desjardins, J.A., van Noordt, S., Huberty, S., Segalowitz, S.J., Elsabbagh, M., 2021. EEG Integrated Platform Lossless (EEG-IP-L) pre-processing pipeline for objective signal quality assessment incorporating data annotation and blind source separation. Journal of Neuroscience Methods 347, 108961. doi:10.1016/j.jneumeth.2020.108961.
Fiecas, M., Ombao, H., 2016. Modeling the evolution of dynamic brain processes during an associative learning experiment. Journal of the American Statistical Association 1440–1453.
Fiecas, M., Ombao, H., Linkletter, C., Thompson, W., Sanes, J., 2010. Functional connectivity: Shrinkage estimation and randomization test. Neuroimage 49 (4), 3005–3014.
Fiecas, M., Ombao, H., van Lunen, D., Baumgartner, R., Coimbra, A., Feng, D., 2013. Quantifying temporal correlations: A test–retest evaluation of functional connectivity in resting-state fMRI. NeuroImage 65, 231–241. doi:10.1016/j.neuroimage.2012.09.052.
Fiecas, M., Ombao, H., et al., 2011. The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. The Annals of Applied Statistics 5 (2A), 1102–1125.
Flandrin, P., 2018. Explorations in Time-Frequency Analysis.
Fried, R., Didelez, V., 2005. Latent variable analysis and partial correlation graphs for multivariate time series. Statistics & Probability Letters 73 (3), 287–296. doi:10.1016/j.spl.2005.04.002.
Fried, R., Didelez, V., Lanius, V., 2005. Partial Correlation Graphs and Dynamic Latent Variables for Physiological Time Series. In: Baier, D., Wernecke, K.-D. (Eds.), Innovations in Classification, Data Science, and Information Systems. Springer-Verlag, Berlin/Heidelberg, pp. 259–266. doi:10.1007/3-540-26981-9_30.
Fried, R., George, A.C., 2011. Median Filters and Extensions. In: Lovric, M. (Ed.), International Encyclopedia of Statistical Science. Springer Berlin Heidelberg, Berlin, Heidelberg, p. 806-806. doi:10.1007/978-3-642-04898-2_361.
Fried, R., Raabe, N., Thieler, A., 2012. On the robust analysis of periodic nonstationary time series. In: COMPSTAT: Proceedings in Computational Statistics, Physica, pp. 245–257.
Fryzlewicz, P., Ombao, H., 2009. Consistent classification of non-stationary signals using stochastic wavelet representations. J Amer Stat Assoc 104, 299–312.
Fryzlewicz, P., Subba-Rao, S., 2014. Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. Journal of the Royal Statistical Society, Series B 903–924.
Fu, W.J., 1998. Penalized regressions: the bridge versus the lasso. Journal of computational and graphical statistics 7 (3), 397–416.
Gao, X., Shahbaba, B., Fortin, N., Ombao, H., 2020. Evolutionary state-space models with applications to time-frequency analysis of local field potentials. Statistica Sinica 30, 1561–1582.
Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association 77, 304–313.
Geweke, J., 1984. The decomposition and measurement of the interdependency between second-order stationary processes. Journal of the American Statistical Association 779, 907–915.
Górecki, T., Horváth, L., Kokoszka, P., 2018. Change point detection in heteroscedastic time series. Econometrics and Statistics 7, 63–88. doi:10.1016/j.ecosta.2017.07.005.
Gorrostieta, C., Ombao, H., Bedard, P., Sanes, J.N., 2012. Investigating stimulus-induced changes in connectivity using mixed effects vector autoregressive models. NeuroImage 59, 3347–3355.
Gott, A.N., Eckley, I.A., Aston, J.A.D., 2015. Estimating the population local wavelet spectrum with application to non-stationary functional magnetic resonance imaging time series. Statistics in Medicine 34 (29), 3901–3915. doi:10.1002/sim.6592.
Goyal, A., Garg, R., 2020. Effective EEG Connectivity by Sparse Vector Autoregressive Model. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. ACM, Hyderabad India, pp. 37–45. doi:10.1145/3371158.3371163.
Granados-Garcia, G., Fiecas, M., Shahbaba, B., Fortin, N., Ombao, H., 2021. Modeling brain waves as a mixture of latent processes http://arxiv.org/abs/2102.11971.

Granger, C., 1969. Investigating causal relations by econometric models and crossspectral methods. Econometrica 37, 424–438.
Guerrero, M.B., Huser, R., Ombao, H., 2021. Conex-Connect: Learning Patterns in Extremal Brain Connectivity From Multi-Channel EEG Data. arXiv:2101.09352 [q-bio, stat].
Gupta, A., Gupta, H.P., Biswas, B., Dutta, T., 2020. A Fault-Tolerant Early Classification Approach for Human Activities using Multivariate Time Series. IEEE Transactions on Mobile Computing 1-1. doi:10.1109/TMC.2020.2973616.
Harvill, J.L., Kohli, P., Ravishanker, N., 2017. Clustering Nonlinear, Nonstationary Time Series Using BSLEX. Methodology and Computing in Applied Probability 19 (3), 935–955. doi:10.1007/s11009-016-9528-1.
Hayes, M.H., 1996. Statistical Digital Signal Processing and Modeling. John Wiley and Sons, New York.
Hesterberg, T., Choi, N.H., Meier, L., Fraley, C., et al., 2008. Least angle and l1 penalized regression: A review. Statistics Surveys 2, 61–93.
Hlávka, Z., Hušková, M., Meintanis, S.G., 2020. Change-point methods for multivariate time-series: Paired vectorial observations. Statistical Papers 61 (4), 1351–1383. doi:10.1007/s00362-020-01175-3.
Hosoya, Y., 1991. The decomposition and measurement of the interdependency between second-order stationary processes. Probability Theory and Related Fields 88, 429–444.
Hu, L., Fortin, N., Ombao, H., 2019. Vector autoregressive models for multivariate brain signals. Statistics in the Biosciences 11, 91–126. doi:10.1007/s12561-017-9210-3.
Huang, A., Qiu, L., Li, Z., 2021. Applying deep learning method in TVP-VAR model under systematic financial risk monitoring and early warning. Journal of Computational and Applied Mathematics 382, 113065. doi:10.1016/j.cam.2020.113065.
Huang, H.-Y., Ombao, H., Stoffer, D.S., 2004. Discrimination and Classification of Nonstationary Time Series Using the SLEX Model. Journal of the American Statistical Association 99 (467), 763–774. doi:10.1198/016214504000001105.
Ibrahim, S., Djemal, R., Alsuwailem, A., 2018. Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis. Biocybernetics and Biomedical Engineering 38 (1), 16–26. doi:10.1016/j.bbe.2017.08.006.
Jacob, J.E., Nair, G.K., Cherian, A., Iype, T., 2019. Application of fractal dimension for EEG based diagnosis of encephalopathy. Analog Integrated Circuits and Signal Processing 100 (2), 429–436. doi:10.1007/s10470-019-01388-z.
Jas, M., Engemann, D.A., Bekhti, Y., Raimondo, F., Gramfort, A., 2017. Autoreject: Automated artifact rejection for MEG and EEG data. NeuroImage 159, 417–429. doi:10.1016/j.neuroimage.2017.06.030.
Jiao, S., Shen, T., Yu, Z., Ombao, H., 2021. Change-point detection using spectral PCA for multivariate time series. arXiv:2101.04334 [stat].
Kang, H., Ombao, H., Linkletter, C., Long, N., Badre, D., 2012. Spatio-spectral mixed-effects model for functional magnetic resonance imaging data. Journal of the American Statistical Association 107 (498), 568–577. doi:10.1080/01621459.2012.664503.
Kirch, C., Muhsal, B., Ombao, H., 2015. Detection of changes in multivariate time series with application to eeg data. Journal of the American Statistical Association 110. 2297–1216
Kokoszka, P., Stoev, S., Xiong, Q., 2019. Principal components analysis of regularly varying functions. Bernoulli 25 (4B). doi:10.3150/19-BEJ1113.
Kook, J.H., Guindani, M., Zhang, L., Vannucci, M., 2019. NPBayes-fMRI: Non-parametric Bayesian General Linear Models for Single- and Multi-Subject fMRI Data. Statistics in Biosciences 11 (1), 3–21. doi:10.1007/s12561-017-9205-0.
Koutlis, C., Kugiumtzis, D., 2021. The Effect of a Hidden Source on the Estimation of Connectivity Networks from Multivariate Time Series. Entropy 23 (2), 208. doi:10.3390/e23020208.
Kulaichev, A.P., 2011. The Informativeness of Coherence Analysis in EEG Studies. Neuroscience and Behavioral Physiology 41 (3), 321–328. doi:10.1007/s11055-011-9419-4.
Lee, T., 1997. A simple span selector for periodogram smoothing. Biometrika 84, 965–969.
Li, T.-H., 2014. Time Series with Mixed Spectra. CRC Press/Chapman & Hall, Boca Raton, Fla..
Liao, S.-C., Wu, C.-T., Huang, H.-C., Cheng, W.-T., Liu, Y.-H., 2017. Major Depression Detection from EEG Signals Using Kernel Eigen-Filter-Bank Common Spatial Patterns. Sensors 17 (6), 1385. doi:10.3390/s17061385.
Lin, J., Michailidis, G., 2018. Regularized Estimation of High-dimensional Factor-Augmented Vector Autoregressive (FAVAR) Models. SSRN Electronic Journal doi:10.2139/ssrn.3615069.
Lin, Y.-L., Chen, W.-Y., Shieh, S.-H., 2020. Age Structural Transitions and Copayment Policy Effectiveness: Evidence from Taiwan's National Health Insurance System. International Journal of Environmental Research and Public Health 17 (12), 4183. doi:10.3390/ijerph17124183.
Lindquist, M.A., Xu, Y., Nebel, M.B., Caffo, B.S., 2014. Evaluating dynamic bivariate correlations in resting-state fMRI: A comparison study and a new approach. NeuroImage 101, 531–546. doi:10.1016/j.neuroimage.2014.06.052.
Liu, Y., Taniguchi, M., Ombao, H., 2021. Statistical inference for local granger causality. arXiv: https://arxiv.org/abs/2103.00209 Submitted.
Loève, M., 1955. Probability Theory. The University series in higher Mathematics.
Lu, H., Lam, L.C.W., Ning, Y., 2019. Scalp-to-cortex distance of left primary motor cortex and its computational head model: Implications for personalized neuromodulation. CNS Neuroscience & Therapeutics 25 (11), 1270–1276. doi:10.1111/cns.13204.
Lutkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. New York Springer, Berlin.
Mahato, S., Paul, S., 2019. Electroencephalogram (EEG) Signal Analysis for Diagnosis of Major Depressive Disorder (MDD): A Review. In: Nath, V., Mandal, J.K. (Eds.), Nanoelectronics, Circuits and Communication Systems, Vol. 511. Springer Singapore, Singapore, pp. 323–335. doi:10.1007/978-981-13-0776-8_30.
Mahecha, M.D., Gans, F., Brandt, G., Christiansen, R., Cornell, S.E., Fomferra, N., Kraemer, G., Peters, J., Bodesheim, P., Camps-Valls, G., Donges, J.F., Dorigo, W., Estupinan-Suarez, L.M., Gutierrez-Velez, V.H., Gutwin, M., Jung, M., Londoño, M.C., Miralles, D.G., Papastefanou, P., Reichstein, M., 2020. Earth system data cubes unravel global multivariate dynamics. Earth System Dynamics 11 (1), 201–234. doi:10.5194/esd-11-201-2020.
Maia, G.O., Barreto-Souza, W., Bastos, F.S., Ombao, H., 2020. Semiparametric time series models driven by latent factor. arXiv:2004.11470 [stat].
Mairal, J., Yu, B., 2012. Complexity analysis of the lasso regularization path. arXiv preprint arXiv:1205.0079.
Manomaisaowapak, P., Nartkulpat, A., Songsiri, J., 2020. Granger Causality Inference in EEG Source Connectivity Analysis: A State-Space Approach. Preprint. Neuroscience doi:10.1101/2020.10.07.329276.
Manomaisaowapak, P., Songsiri, J., 2020. Learning A Common Granger Causality Network Using A Non-Convex Regularization. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Barcelona, Spain, pp. 1160–1164. doi:10.1109/ICASSP40776.2020.9054430.
Mantri, S., Patil, D., Agrawal, P., Wadhai, V., 2015. Non invasive EEG signal processing framework for real time depression analysis. In: 2015 SAI Intelligent Systems Conference (IntelliSys). IEEE, London, United Kingdom, pp. 518–521. doi:10.1109/IntelliSys.2015.7361188.
Mariscal, M.G., Levin, A.R., Gabard-Durnam, L.J., Tager-Flusberg, H., Nelson, C.A., 2019. Developmental Changes in EEG Phase Amplitude Coupling and Phase Preference over the First Three Years After Birth. Preprint. Neuroscience doi:10.1101/818583.
Martín Cervantes, P.A., Rueda López, N., Cruz Rambaud, S., 2020. Life Expectancy at Birth: A Causal Analysis of the Health Sector in Spain. In: Flaut, D., Hošková-Mayerová, Š., Ispas, C., Maturo, F., Flaut, C. (Eds.), Decision Making in Social Sciences: Between Traditions and Innovations, Vol. 247. Springer International Publishing, Cham, pp. 373–410. doi:10.1007/978-3-030-30659-5_21.
, 1990. In: McGaugh, J.L., Weinberger, N.M., Lynch, G., University of California, Irvine (Eds.), Brain Organization and Memory: Cells, Systems, and Circuits. Oxford University Press, New York.
Menon, S.S., Krishnamurthy, K., 2019. A Comparison of Static and Dynamic Functional Connectivities for Identifying Subjects and Biological Sex Using Intrinsic Individual Brain Connectivity. Scientific Reports 9 (1), 5729. doi:10.1038/s41598-019-42090-4.
Mohammadi, M.R., Khaleghi, A., Nasrabadi, A.M., Rafieivand, S., Begol, M., Zarafshan, H., 2016. EEG classification of ADHD and normal children using non-linear features and neural network. Biomedical Engineering Letters 6 (2), 66–73. doi:10.1007/s13534-016-0218-2.
Molaee-Ardekani, B., Senhadji, L., Shamsollahi, M.-B., Wodey, E., Vosoughi-Vahdat, B., 2007. Delta waves differently modulate high frequency components of EEG oscillations in various unconsciousness levels. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE doi:10.1109/iembs.2007.4352534.

Moti Nasrabadi, A., Allahverdy, A., Samavati, M., Mohammadi, M. R., 2020. EEG data for ADHD / Control children. https://ieee-dataport.org/open-access/eeg-data-adhd-control-children. doi:10.21227/RZFH-ZN36.
Motta, G., Ombao, H., 2012. Evolutionary Factor Analysis of Replicated Time Series. Biometrics 68 (3), 825–836. doi:10.1111/j.1541-0420.2012.01744.x.
Mumford, J., Ramsey, J., 2014. Bayesian networks for fmri: a primer. NeuroImage 86 (19), 572–583.
Nascimento, D.C., Pinto-Orellana, M.A., Leite, J.P., Edwards, D.J., Louzada, F., Santos, T.E.G., 2020. BrainWave Nets: Are Sparse Dynamic Models Susceptible to Brain Manipulation Experimentation? Frontiers in Systems Neuroscience 14, 527757. doi:10.3389/fnsys.2020.527757.
Nason, G.P., von Sachs, R., Kroisandt, G., 2000. Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 62 (2), 271–292.
Newson, J.J., Thiagarajan, T.C., 2019. EEG Frequency Bands in Psychiatric Disorders: A Review of Resting State Studies. Frontiers in Human Neuroscience 12, 521. doi:10.3389/fnhum.2018.00521.
Nunez, P.L., Srinivasan, R., 2006. Electric Fields of the Brain: The Neurophysics of EEG, 2nd ed Oxford University Press, Oxford ; New York.
Nunez, P.L., Srinivasan, R., Fields, R.D., 2015. EEG functional connectivity, axon delays and white matter disease. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology 126 (1), 110–120. doi:10.1016/j.clinph.2014.04.003.
Ombao, H., 2019. Spectral Approach to Modeling Dependence in Multivariate Time Series. Journal of Physics: Conference Series 1417, 012007. doi:10.1088/1742-6596/1417/1/012007.
Ombao, H., Lindquist, M., Thompson, W., Aston, J., 2017. Handbook of Neuroimaging Data Analysis.
Ombao, H., Ringo Ho, M.-h., 2006. Time-dependent frequency domain principal components analysis of multichannel non-stationary signals. Computational Statistics & Data Analysis 50 (9), 2339–2360. doi:10.1016/j.csda.2004.12.011.
Ombao, H., von Sachs, R., Guo, W., 2005. SLEX Analysis of Multivariate Non-Stationary Time Series. J Amer Stat Assoc 100 (470), 519–531.
Ombao, H., Van Bellegem, S., et al., 2006. Coherence analysis of nonstationary time series: a linear filtering point of view. IEEE Trans on Signal Proc 56 (6), 2259–2266.
Ombao, H., von Sachs, R., Guo, W., 2005. SLEX Analysis of Multivariate Nonstationary Time Series. Journal of the American Statistical Association 100 (470), 519–531. doi:10.1198/016214504000001448.
Ombao, H.C., Raz, J.A., von Sachs, R., Malow, B.A., 2001. Automatic statistical analysis of bivariate nonstationary time series. Journal of the American Statistical Association 96 (454), 543–560.
Ombao, H.C., Raz, J.A., Strawderman, R., von Sachs, R., 2001. A simple generalised crossvalidation method of span selection for periodogram smoothing. Biometrika 88, 1186–1192.
Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clinical Neurophysiology 112 (4), 713–719. doi:10.1016/S1388-2457(00)00527-7.
Oppenheim, A.V., Willsky, A.S., Nawab, S.H., 1997. Signals & Systems, 2nd ed Prentice Hall, Upper Saddle River, N.J.
Park, H., Jung, J.-Y., 2020. SAX-ARM: Deviant event pattern discovery from multivariate time series using symbolic aggregate approximation and association rule mining. Expert Systems with Applications 141, 112950. doi:10.1016/j.eswa.2019.112950.
Park, T., Eckley, I., Ombao, H., 2014. Estimating the time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. IEEE Transactions on Signal Processing 62, 5240–5250.
Park, T., Eckley, I., Ombao, H., 2018. Dynamic classification using multivariate locally stationary wavelets. Signal Processing 152, 118–129.
Pedroni, A., Bahreini, A., Langer, N., 2019. Automagic: Standardized preprocessing of big EEG data. NeuroImage 200, 460–473. doi:10.1016/j.neuroimage.2019.06.046.
Pham, N.M., Huynh, T.L.D., Nasir, M.A., 2020. Environmental consequences of population, affluence and technological progress for European countries: A Malthusian view. Journal of Environmental Management 260, 110143. doi:10.1016/j.jenvman.2020.110143.
Pinto, M., Mirtaheri, P., Hammer, H., 2021a. Dyadic aggregated autoregressive (DASAR) model for time-frequency representation of biomedical signals.
Pinto, M., Mirtaheri, P., Hammer, H., 2021b. SCAU: Modeling spectral causality for multivariate time series with applications to electroencephalograms.
Pinto-Orellana, M.A., Hammer, H.L., 2020. Analysis of Optical Brain Signals Using Connectivity Graph Networks. Machine Learning and Knowledge Extraction 12279, 485–497. doi:10.1007/978-3-030-57321-8_27.
Pinto-Orellana, M.A., Hammer, H.L., 2020. Dyadic Aggregated Autoregressive Model (DASAR) for Automatic Modulation Classification. IEEE Access 8, 156096–156103. doi:10.1109/ACCESS.2020.3019243.
Priestley, M., 1965. Evolutionary spectra and non-stationary processes. Journal of the Royal Statistical Society Series B. (Methodological) 27 (2), 204–237.
Priestley, M.B., 1981. Spectral Analysis and Time Series. Academic Press, London ; New York.
Qin, Y., Chen, J., Dong, X., 2021. Oil prices, policy uncertainty and travel and leisure stocks in China. Energy Economics 96, 105112. doi:10.1016/j.eneco.2021.105112.
Ravagli, D., Boshnakov, G.N., 2020. Portfolio optimization with mixture vector autoregressive models. arXiv:2005.13396 [stat].
Reboredo, J.C., Ugolini, A., 2020. Price spillovers between rare earth stocks and financial markets. Resources Policy 66, 101647. doi:10.1016/j.resourpol.2020.101647.
Rembach, A., Stingo, F.C., Peterson, C., Vannucci, M., Do, K.-A., Wilson, W.J., Macaulay, S.L., Ryan, T.M., Martins, R.N., Ames, D., Masters, C.L., Doecke, J.D., 2015. Bayesian Graphical Network Analyses Reveal Complex Biological Interactions Specific to Alzheimer's Disease. Journal of Alzheimer's Disease 44 (3), 917–925. doi:10.3233/JAD-141497.
Robbins, K.A., Touryan, J., Mullen, T., Kothe, C., Bigdely-Shamlo, N., 2020. How Sensitive Are EEG Results to Preprocessing Methods: A Benchmarking Study. IEEE Transactions on Neural Systems and Rehabilitation Engineering 28 (5), 1081–1090. doi:10.1109/TNSRE.2020.2980223.
Rumelhart, D., Hinton, G., Williams, R., 1988. Learning Internal Representations by Error Propagation. In: Readings in Cognitive Science. Elsevier, pp. 399–421. doi:10.1016/B978-1-4832-1446-7.50035-2.
Sanei, S., Chambers, J., 2007. EEG Signal Processing. John Wiley & Sons, Chichester, England ; Hoboken, NJ.
Sato, J.R., Fujita, A., Cardoso, E.F., Thomaz, C.E., Brammer, M.J., Amaro, E., 2010. Analyzing the connectivity between regions of interest: An approach based on cluster Granger causality for fMRI data analysis. NeuroImage 52 (4), 1444–1455. doi:10.1016/j.neuroimage.2010.05.022.
Schettlinger, K., Fried, R., Gather, U., 2009. Real-time signal processing by adaptive repeated median filters. International Journal of Adaptive Control and Signal Processing n/a–n/a. doi:10.1002/acs.1105.
Schiff, S.J., Colella, D., Jacyna, G.M., Hughes, E., Creekmore, J.W., Marshall, A., Bozek-Kuzmicki, M., Benke, G., Gaillard, W.D., Conry, J., Weinstein, S.R., 2000. Brain chirps: Spectrographic signatures of epileptic seizures. Clinical Neurophysiology 111 (6), 953–958. doi:10.1016/S1388-2457(00)00259-5.
Schoenberg, P.L.A., 2020. Linear and Nonlinear EEG-Based Functional Networks in Anxiety Disorders. In: Kim, Y.-K. (Ed.), Anxiety Disorders, Vol. 1191. Springer Singapore, Singapore, pp. 35–59. doi:10.1007/978-981-32-9705-0_3.
Schröder, A.L., Ombao, H., 2019. FreSpeD: Frequency-Specific Change-Point Detection in Epileptic Seizure Multi-Channel EEG Data. Journal of the American Statistical Association 114 (525), 115–128. doi:10.1080/01621459.2018.1476238.
Sethi, J.K., Mittal, M., 2020. Analysis of Air Quality using Univariate and Multivariate Time Series Models. In: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, Noida, India, pp. 823–827. doi:10.1109/Confluence47617.2020.9058303.
Shappell, H., Caffo, B., Pekar, J., Lindquist, M., 2019. Improved state change estimation in dynamic functional connectivity using hidden semi-markov models. NeuroImage 191 (19), 243–357.
Snyder, S.M., Rugino, T.A., Hornig, M., Stein, M.A., 2015. Integration of an EEG biomarker with a clinician's ADHD evaluation. Brain and Behavior 5 (4). doi:10.1002/brb3.330.
Soleimani, B., Das, P., Kulasingham, J., Simon, J.Z., Babadi, B., 2020. Granger Causal Inference from Indirect Low-Dimensional Measurements with Application to MEG Functional Connectivity Analysis. In: 2020 54th Annual Conference on Information Sciences and Systems (CISS). IEEE, Princeton, NJ, USA, pp. 1–5. doi:10.1109/CISS48834.2020.1570617418.

Steinmann, S., Nolte, G., Mulert, C., 2020. EEG Connectivity Pattern: A Window into the Schizophrenia Mind? In: Kubicki, M., Shenton, M.E. (Eds.) Neuroimaging in Schizophrenia. Springer International Publishing, Cham, pp. 227–240. doi:10.1007/978-3-030-35206-6_13.
Stephanie, L., 2017. Regularized Structural Equation Modeling for Individual-Level Directed Functional Connectivity. The University of North Carolina at Chapel Hill University Libraries.
Stern, J., 2013. Atlas of EEG Patterns. Wolters Kluwer/Lippincott Williams & Wilkins Health, Philadelphia.
Subba-Rao, S., Jentsch, C., 2015. A test for second order stationarity of a multivariate time series. Journal of Econometrics 185, 124–161.
Sundararajan, R.R., Frostig, R., Ombao, H., 2020. Modeling Spectral Properties in Stationary Processes of Varying Dimensions with Applications to Brain Local Field Potential Signals. Entropy 22 (12), 1375. doi:10.3390/e22121375.
Suotsalo, K., Xu, Y., Corander, J., Pensar, J., 2020. High-dimensional structure learning of sparse vector autoregressive models using fractional marginal pseudo-likelihood. arXiv:2011.01484 [stat].
Swanson, L.W., 2003. Brain Architecture: Understanding the Basic Plan. Oxford University Press, Oxford ; New York.
Tatsis, K., Dertimanis, V., Ou, Y., Chatzi, E., 2020. GP-ARX-Based Structural Damage Detection and Localization under Varying Environmental Conditions. Journal of Sensor and Actuator Networks 9 (3), 41. doi:10.3390/jsan9030041.
Tatum, W., 2008. Handbook of EEG Interpretation. Demos Medical Pub, New York.
Thatcher, R., Krause, P., Hrybyk, M., 1986. Cortico-cortical associations and EEG coherence: A two-compartmental model. Electroencephalography and Clinical Neurophysiology 64 (2), 123–143. doi:10.1016/0013-4694(86)90107-0.
Thatcher, R.W., 2011. Neuropsychiatry and quantitative EEG in the 21st Century. Neuropsychiatry 1 (5), 495–514. doi:10.2217/npy.11.45.
Thatcher, R.W., North, D.M., Biver, C.J., 2008. Development of cortical connections as measured by EEG coherence and phase delays. Human Brain Mapping 29 (12), 1400–1415. doi:10.1002/hbm.20474.
Thomson, D., 1982. Spectrum estimation and harmonic analysis. Proceedings of the IEEE 70 (9), 1055–1096. doi:10.1109/PROC.1982.12433.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 267–288.
Ting, C., Ombao, H., Samdin, S., Salleh, S., 2018. Estimating dynamic connectivity states in fmri using regime-switching factor models. IEEE Transactions on Medical Imaging 37, 1011–1023.
Ting, C.-M., Ombao, H., Salleh, S.H., Abd Latif, A., 2018. Multi-scale factor analysis of high- dimensional functional connectivity in brain networks. IEEE Transactions on Network Science and Engineering 7, 449–465.
Ting, C.-M., Ombao, H., Salleh, S.-H., Latif, A.Z.A., 2020. Multi-Scale Factor Analysis of High-Dimensional Functional Connectivity in Brain Networks. IEEE Transactions on Network Science and Engineering 7 (1), 449–465. doi:10.1109/TNSE.2018.2869862.
Ting, C.-M., Ombao, H., Salleh-Sh, H., 2018. Estimating dynamic connectivity states in fmri using regime-switching factor models. IEEE Transactions in Medical Imaging 37, 1011–1023.
Ting, C.-M., Samdin, B., Tang, M., Ombao, H., 2021. Detecting dynamic community structure in functional brain networks across individuals: A multilayer approach. IEEE Transactions in Medical Imaging 7, 468–480.
Ting, C.-M., Samdin, S.B., Tang, M., Ombao, H., 2021. Detecting Dynamic Community Structure in Functional Brain Networks Across Individuals: A Multilayer Approach. IEEE Transactions on Medical Imaging 40 (2), 468–480. doi:10.1109/TMI.2020.3030047.
Tort, A.B.L., Komorowski, R., Eichenbaum, H., Kopell, N., 2010. Measuring Phase-Amplitude Coupling Between Neuronal Oscillations of Different Frequencies. Journal of Neurophysiology 104 (2), 1195–1210. doi:10.1152/jn.00106.2010.
Usman, O., Alola, A.A., Sarkodie, S.A., 2020. Assessment of the role of renewable energy consumption and trade policy on environmental degradation using innovation accounting: Evidence from the US. Renewable Energy 150, 266–277. doi:10.1016/j.renene.2019.12.151.
Warnick, R., Guindani, M., Erhardt, E., Allen, E., Calhoun, V., Vannucci, M., 2018. A bayesian approach for estimating dynamic functional network connectivity in fmri data. Journal of the American Statistical Association 113 (521), 134–151.
Widmann, A., Schröger, E., Maess, B., 2015. Digital filter design for electrophysiological data – a practical approach. Journal of Neuroscience Methods 250, 34–46. doi:10.1016/j.jneumeth.2014.08.002.
Witter, M.P., Doan, T.P., Jacobsen, B., Nilssen, E.S., Ohara, S., 2017. Architecture of the Entorhinal Cortex A Review of Entorhinal Anatomy in Rodents with Some Comparative Notes. Frontiers in Systems Neuroscience 11, 46. doi:10.3389/fnsys.2017.00046.
Wu, J., Srinivasan, R., Kaur, A., Cramer, S.C., 2014. Resting-state cortical connectivity predicts motor skill acquisition. NeuroImage 91, 84–90. doi:10.1016/j.neuroimage.2014.01.026.
Wu, J., Srinivasan, R., Kaur, A., Cramer, S.C., 2014. Resting-state cortical connectivity predicts motor skill acquisition. NeuroImage 91, 84–90.
Yu, Z., Prado, R., Burke, E., Cramer, S., Ombao, H., 2016. A hierarchical bayesian model for studying the impact of stroke on brain motor function. Journal of the American Statistical Association 111, 549–562.
Zhang, L., Guindani, M., Vannucci, M., 2015. Bayesian models for functional magnetic resonance imaging data analysis. Wiley Interdisciplinary Reviews: Computational Statistics 7 (1), 21–41. doi:10.1002/wics.1339.
Zhang, L., Guindani, M., Versace, F., Engelmann, J.M., Vannucci, M., 2016. A spatiotemporal nonparametric Bayesian model of multi-subject fMRI data. The Annals of Applied Statistics 10 (2). doi:10.1214/16-AOAS926.
Zhao, P., Rocha, G., Yu, B., 2009. The composite absolute penalties family for grouped and hierarchical variable selection. The Annals of Statistics 3468–3497.
Zhao, Y., Luo, X., 2019. Granger mediation analysis of functional magnetic resonance imaging time series. Biometrics 75, 788–798.