RESEARCH · **Open Access**

Non-identical smoothing operators for estimating time-frequency interdependence in electrophysiological recordings

Saeid Mehrkanoon[1,2*], Michael Breakspear[1,2,3], Andreas Daffertshofer[4] and Tjeerd W Boonstra[1,2,4]

Abstract

Synchronization of neural activity from distant parts of the brain is crucial for the coordination of cognitive activities. Because neural synchronization varies both in time and frequency, time–frequency (T-F) coherence is commonly employed to assess interdependences in electrophysiological recordings. T-F coherence entails smoothing the cross and power spectra to ensure statistical consistency of the estimate, which reduces its T-F resolution. This trade-off has been described in detail when the cross and power spectra are smoothed using identical smoothing operators, which may yield spurious coherent frequencies. In this article, we examine the use of non-identical smoothing operators for the estimation of T-F interdependence, i.e., phase synchronization is characterized by phase locking between signals captured by the cross spectrum and we may hence improve the trade-off by selectively smoothing the auto spectra. We first show that the frequency marginal density of the present estimate is bound within [0,1] when using non-identical smoothing operators. An analytic calculation of the bias and variance of present estimators is performed and compared with the bias and variance of standard T-F coherence using Monte Carlo simulations. We then test the use of non-identical smoothing operators on simulated data, whose T-F properties are known through construction. Finally, we analyze empirical data from eyes-closed surface electroencephalography recorded in human subjects to investigate alpha-band synchronization. These analyses show that selectively smoothing the auto spectra reduces the bias of the estimator and may improve the detection of T-F interdependence in electrophysiological data at high temporal resolution.

Keywords: Coherence, Time–frequency analysis, Electroencephalography, Functional connectivity, Neural synchronization

1 Introduction

Coherence analysis has widely been used for investigating functional connectivity between electrophysiological signals to study the role of neural synchronization in perception, action, and cognition [1-3]. Neural synchronization has been reported in different frequency bands and to vary over time and is hence commonly assessed using time–frequency (T-F) coherence. T-F coherence has been used in many fields of science, including neuroscience [4-10]. Short-time Fourier transform (STFT) and wavelet transformation are typically used to estimate the T-F distribution of the signals of interest, which can be interpreted as the spectral density estimation. Other approaches include the quadratic T-F distribution from Cohen's class [11], such as the Winger-Ville [11,12] and the Choi-Williams distribution [13].

Smoothing the cross and auto spectra is required for the estimation of T-F coherence [7,14]. Given an arbitrary T-F decomposition method, smoothing the cross and auto spectra can be performed via one of the following approaches; (a) Smoothing of periodigrams via ensemble averaging, which is based on the Welch's overlapped segment averaging (WOSA) method [15]. This approach yields magnitude-squared coherence (MSC), see also [16]. (b) Smoothing in one or both of the time and frequency domains. In wavelet coherence, an explicit scale-dependent 2D smoothing operator has previously been

*Correspondence: s.mehrkanoon@unsw.edu.au
[1] School of Psychiatry, The University of New South Wales, Sydney, NSW, Australia
[2] Black Dog Institute, Hospital Rd, Randwick, NSW 2031, Australia
Full list of author information is available at the end of the article

used, e.g., a Gaussian envelope of the Morlet wavelet function as the time-smoothing operator and a box-car filter with a specific width as the scale (or frequency) smoothing operator [7,17,18]. (c) Smoothing of both cross and auto spectra via averaging a set of spectra estimated by multiple orthogonal taper functions [10,19]. Multiwavelet coherence was proposed based on the concept of Thomson's multitaper technique [20] and involves a set of orthogonal wavelet taper functions, the so-called Morse wavelet. Smoothing in the T-F plane is implicit when the cross and auto spectra are averaged across the set of taper functions [20-22]. It has been shown that multiwavelet coherence has a lower variance compared to explicit 2D smoothing operators [21].

All these approaches estimate coherence by using identical smoothing operators for the cross and auto spectra. The use of identical smoothing operators indicates that T-F coherence satisfies the Cauchy-Schwarz inequality and is hence bounded within [0,1] (see also [20,21]). Smoothing entails an inherent trade-off between statistical consistency and T-F resolution, i.e., wider smoothing windows result in improved accuracy and precision at the cost of reduced resolution in T-F plane [23]. When using identical smoothing operators the estimator fails when smoothing goes to 1, as the numerator and denominator become identical. Because phase synchronization is characterized by phase locking between signals, captured by the cross spectrum, we may improve the temporal resolution by selectively smoothing the auto spectra. When using non-identical smoothing operators, the bias of the estimator may not go to 1 when the cross spectrum is not smoothed, resulting in an estimator with improved temporal resolution. In particular, we consider the case in which only the auto spectra are smoothed. The resulting estimate is not bound within [0,1], but it can be shown that its frequency marginal density is equivalent to standard coherence as a function of frequency.

We first test the effect of non-identical smoothing operators in simulated data and quantify its performance based on the sensitivity, specificity, and z-score. We then apply this method to empirical data of surface electroencephalography (EEG) and assess cortico-cortical connectivity between occipital channels estimated using identical and non-identical smoothing operators. The marginal-covariance densities are used to compare the observed correlations in the time and frequency domains. These analyses show that the use of non-identical smoothing operators may help to detect weak correlations in electrophysiological data and open up new possibilities for quantifying T-F interdependences.

2 Methods
We first provide a short overview of T-F decomposition methods before introducing a generalized T-F coherency

function. We then derive two specific estimators based on non-identical smoothing operators. Two simulated datasets are generated for testing the estimators. The first example involves signals whose correlations vary in time and in the second example correlations vary in time and frequency. The empirical dataset consists of EEG data from two occipital channels in three subjects to investigate synchronization in the alpha frequency band (8–12 Hz), which is commonly observed over occipital cortex.

2.1 Spectral decomposition
Both the STFT [24] and the wavelet transform [5,25-27] have extensively been used for spectral decomposition. There is a trade-off between spectral and temporal resolution depending on window length: short windows give poor spectral but good temporal resolution, and vice versa for long windows. Temporal and spectral resolutions remain constant over frequencies for STFT, as a fixed window length is used. In contrast, the wavelet transform has varying T-F resolution, because signal are decomposed by rescaling and shifting a mother wavelet function. Irrespective of the differences between spectral decomposition methods, it has been shown that the Fourier, Hilbert, and wavelet-based techniques are in fact mathematically equivalent when using the most frequently employed class of wavelets [28]. Recently, statistical properties including the bias and variance of the phase estimators such as Fourier, Hilbert, and wavelet-based techniques have been discussed in [29].

Using WOSA, the signal is first partitioned into L equivalent segments. Each segment is then weighted with a suitable window $w[m]$ (positive, square integrable, integrates to one), e.g., a unit power Hamming window. The Fourier transform of each weighted segment is computed and the estimation of power spectral density (PSD) is obtained by averaging over L overlapping segments [24], estimating the cross spectral density (CSD), PSD, and the discrete Fourier transform of signal $x_l[n]$ as

$$\hat{p}_{xy}[k] = \frac{1}{L} \sum_{l=1}^{L} X_l[k]\, Y_l^*[k],$$

$$\hat{p}_{xx}[k] = \frac{1}{L} \sum_{l=1}^{L} \left| X_l[k] \right|^2 = \frac{1}{L}\hat{p}_x[k], \qquad (1)$$

$$X_l[k] = \frac{1}{M} \sum_{m=1}^{M} x_l[m]\, w[m]\, e^{-j\frac{2\pi}{M}km},$$

where k denotes the discrete frequency, asterisk $(^*)$ the complex conjugate, $x_l[n] = x[l(M-q)+n]$ and $y_l[n] = y[l(M-q)+n]$ where $l = 1, 2, \ldots, L$ and $n = 1, 2, \ldots, N$, $j = \sqrt{-1}$. L and N are, respectively, the total number of overlapping windows and data samples in $x[n]$, M and q denote the width of a weighting window function $w[m]$

and the overlapping samples, respectively. Alternatively, the wavelet transform and wavelet-PSD of discrete signal $x[n]$ at scale or frequency a, $(0 < a < \infty)$ and translation or time b, $(b \in \mathbb{R})$ can be expressed as

$$W(a,b;x,\psi) = \sqrt{\frac{\Delta t}{|a|}} \sum_{n=1}^{N} x[n]\, \psi^*\left(\frac{(n-b)\Delta t}{a}\right), \quad (2)$$

$$\hat{p}_{xx_W}(a,b) = |W(a,b;x,\psi)|^2,$$

where Δt denotes the sampling interval, and $\psi(.)$ a permissible wavelet function such as a complex Morlet wavelet function [20,21].

2.2 T-F coherence

Coherence quantifies linear correlations between two stochastic processes, or two observations $x[n]$ and $y[n]$, from a linear time-invariant system as a function of frequency, or time and frequency. Here, we define a generalized coherency function as

$$\hat{\Gamma}_{xy}[l,k] = \frac{S_1\left\{\hat{p}_{xy}[l,k]\right\}}{\sqrt{S_2\left\{\hat{p}_{xx}[l,k]\right\} S_2\left\{\hat{p}_{yy}[l,k]\right\}}}$$

$$l = 1,2,\ldots,L, \quad (3)$$

where $\hat{p}_{xy}[l,k]$ denotes the CSD estimate between $x_l[n]$ and $y_l[n]$, $\hat{p}_{xx}[l,k]$ the PSD estimate, and $S_1\{.\}$ and $S_2\{.\}$ the smoothing operators for the numerator (CSD) and denominator (PSDs), respectively. If we assume identical smoothing operators (i.e., $S_1\{.\} \equiv S_2\{.\}$), we obtain standard T-F coherency, which is normative (i.e., $0 \leq |\hat{\Gamma}_{xy}[l,k]| \leq 1$), see also [7,15,17,30].

The smoothing operators $S_1\{.\}$ and $S_2\{.\}$ can be 1D function of time [16], or an explicit 2D smoothing operator in both time and frequency [7,17,22], or an average over a set of orthogonal-based spectral estimates such as multitaper approaches [10,19,22]. Using a time-domain smoothing operator, $S_{\text{time}}(\bullet|k)$, we obtain standard MSC as

$$\hat{\gamma}_{xy}^2[k] = \frac{\left|\hat{p}_{xy}[k]\right|^2}{\hat{p}_{xx}[k]\,\hat{p}_{yy}[k]}, \quad (4)$$

and T-F coherence as

$$\hat{\Gamma}_{xy}^2[l,k] = \frac{\left|\hat{p}_{xy}[l,k] \circledast v[m]\right|^2}{\left\{\left|\hat{p}_{xx}[l,k]\right|^2 \circledast v[m]\right\}\left\{\left|\hat{p}_{yy}[l,k]\right|^2 \circledast v[m]\right\}} k$$

$$= 1,2,\ldots,K, \quad (5)$$

where the numerator and denominator of Equation (4) are estimated as in Equation (1). The smoothing window $v[m]$, a suitable smoothing window (positive, square integrable, integrates to one)—e.g., a unit power Hamming window function of length M_1—is used as time-domain smoothing operator $S_{\text{time}}(\bullet|k)$, where $\circledast$ denotes the convolution operator. Note that smoothing process is

implemented by convolving the T-F CSD and PSDs with smoothing window $v[m]$ at any given frequency k. In the present context and for comparison purposes, we use a weighting window in the STFT, i.e., normalized Hamming window of 0.5 s in duration, and smoothing window, $v[m]$, of 0.75 s duration. In this study, a normalized Hamming window was also used for smoothing the cross and auto spectra defined in Equations (5) and (11), and refer to T-F interdependence using identical smoothing operators as Method 1.

Smoothing in both time and frequency has been proposed to improve the consistency of the T-F coherence [22], which can be expressed as $S\left\{p_{xy}[l,k]\right\} = S_{\text{freq}}(S_{\text{time}}(p_{xy}|k)|l)$. In the wavelet coherence approach, the Gaussian envelope of the Morlet wavelet with unit power is used as the time-domain smoothing operator, and a frequency-domain filter, a box-car function of the width δ, as the frequency-domain smoothing operator,

$$S\left\{p_{xy}[l,k]\right\} = S_{\text{freq}}\left(S_{\text{time}}(p_{xy}|k)\Big|l\right)$$

$$= S_{\text{freq}}\left(p_{xy}[l,k]\, c_1 e^{-l^2/(2\sigma_l k)^2}\Big|l\right), \quad (6)$$

where $S_{\text{freq}}(p_{xy}|l) = p_{xy}[l,k]\, c_2 \prod(\sigma_k \delta a)$. Here $\prod(.)$ denotes the box-car function, σ_l and σ_k the time and frequency localization parameters, respectively. The coefficients c_1 and c_2 normalize the filters to unit energy. Subsequently, identical smoothing operators are applied to the temporal PSDs $p_{xx}[l,k]$ and $p_{yy}[l,k]$ [22]. Alternatively, multiple orthogonal tapers are used to estimate the spectral densities [10,19] and smoothing is performed implicitly by averaging over these tapers. The use of a set of orthogonal wavelet functions allows a low variance estimation of the spectral density [21,22,31], and T-F coherence can hence be estimated as

$$\hat{\Lambda}_{xy}^2[l,k] = \frac{\left|\hat{Q}_{xy}[l,k]\right|^2}{\hat{Q}_{xx}[l,k]\,\hat{Q}_{yy}[l,k]}, \quad (7)$$

$$\hat{Q}_{xy}[l,k] = \frac{1}{P}\sum_{p=1}^{P}\hat{p}_{xy}[l,k;p], \quad (8)$$

where $\hat{p}_{xy}[l,k;p]$ is the pth spectral density estimate obtained from pth orthogonal taper or wavelet function.

2.3 Non-identical smoothing operators

We now relax the constraint of identical smoothing operators and investigate Equation (3) when $S_1\{.\} \neq S_2\{.\}$. Note that by relaxing this constraint the estimator does not satisfy the Cauchy-Schwarz inequality and hence is not bound within [0,1]. We will first consider the case in which we only smooth the PSDs, i.e., $S_{\text{time}}(\bullet|k)$, is applied to the denominator of Equation (3) only. In particular, we rescale

the temporal complex-valued CSD between signals $x_l[n]$ and $y_l[n]$ by the ensemble averaged PSDs as

$$
\hat{\theta}_{xy}[l,k] = \frac{\hat{p}_{xy}[l,k]}{\sqrt{\hat{p}_{xx}[k]\,\hat{p}_{yy}[k]}}
$$
$$
= L\frac{\hat{p}_{xy}[l,k]}{\sqrt{\hat{p}_x[k]\,\hat{p}_y[k]}}, \quad l = 1,2,\dots,L, \tag{9}
$$

where $\hat{p}_{xx}[k] = \frac{1}{L}\hat{p}_x[k]$, which is given in Equation (1). Although this estimator (Method 2) is not bound within [0,1], we can easily show that the frequency-marginal density of $|\hat{\theta}_{xy}[l,k]|^2$ approaches the MSC given in Equation (4) and hence is bound within [0,1] as shown by

$$
\hat{\gamma}^2_{xy}[k] = \left| \underset{l}{\mathbb{E}}\{\hat{\theta}_{xy}[l,k]\} \right|^2,
$$
$$
= \left| \underset{l}{\mathbb{E}}\{\hat{\theta}_R[l,k]\} + j\underset{l}{\mathbb{E}}\{\hat{\theta}_I[l,k]\} \right|^2,
$$
$$
= \left| \frac{1/L\sum_{l=1}^L \hat{\theta}_R[l,k]}{\sqrt{\hat{p}_{xx}[k]\,\hat{p}_{yy}[k]}} + j\frac{1/L\sum_{l=1}^L \hat{\theta}_I[l,k]}{\sqrt{\hat{p}_{xx}[k]\,\hat{p}_{yy}[k]}} \right|^2,
$$
$$
= \left| \Re\{\hat{\gamma}_{xy}[k]\} + j\Im\{\hat{\gamma}_{xy}[k]\} \right|^2, \tag{10}
$$

where $\mathbb{E}$ is the mathematical expectation, $\hat{\theta}_R[l,k]$ and $\hat{\theta}_I[l,k]$ denote the real and imaginary parts of $\hat{\theta}_{xy}[l,k]$. That is, Equation (9) represents the rescaled T-F interdependence estimate in which the MSC can be obtained by temporally smoothing $\hat{\theta}_{xy}[l,k]$.

Since the numerator in Equation (9) is not smoothed the variance of this estimate will be higher than the variance of Equation (5), but its temporal resolution is also higher than estimates obtained from Equation (5). To facilitate comparison between both approaches, we hence smooth $\hat{\theta}_{xy}[l,k]$ by substituting the numerator of Equation (5) in the numerator of Equation (9), which yields

$$
\hat{\Xi}_{xy}[l,k] = \frac{\hat{p}_{xy}[l,k]\circledast\upsilon[m]}{\sqrt{\hat{p}_{xx}[k]\,\hat{p}_{yy}[k]}} = L\frac{\hat{p}_{xy}[l,k]\circledast\upsilon[m]}{\sqrt{\hat{p}_x[k]\,\hat{p}_y[k]}}. \tag{11}
$$

That is, the smoothed CSD is rescaled by the inner product of the ensemble-averaged PSDs (Method 3). (The MATLAB-function of Methods 1,2, and 3 can be found in the following link: *http://www.sng.org.au/Downloads*)

2.3.1 The bias and variance of T-F interdependence estimators, $\hat{\theta}_{xy}[l,k]$ and $\hat{\Xi}_{xy}[l,k]$

It is well known that the standard T-F coherence function, (see Equation (5)), between two random Gaussian processes is distributed as a chi-squared random variable with ν degrees of freedom, shown as χ^2_ν, [7,17]. Statistical properties of MSC, (see Equation (4)), was shown to have a specific probability density function (PDF) and cumulative distribution function [16,32]. Recently, Cohen

and Walden [18,31] have shown that temporally smoothed wavelet coherence, (see Equation (7)), of two stationary processes with Normal distribution and zero mean is distributed as the Goodman's distribution with an appropriate degrees of freedom. Therefore, the bias and variance of the standard T-F coherence and MSC have explicitly been stated and addressed in the literature. In this section, the aim is to analytically calculate the bias and variance of T-F interdependence estimators, $|\hat{\theta}_{xy}[l,k]|$, see Equation (9), and $|\hat{\Xi}_{xy}[l,k]|$, see Equation (11). We define the bias and variance of the estimators $|\hat{\theta}_{xy}[l,k]|$ and $|\hat{\Xi}_{xy}[l,k]|$ as follows (derivations of analytic calculation of the bias and variance are given in Appendix)

$$
\mathbb{E}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\} \approx \frac{\mathbb{E}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]} - \frac{\mathrm{cov}(\mathbf{A},\mathbf{B})}{\mathbb{E}[\mathbf{B}]^2} + \frac{\mathbb{E}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]^3}\mathrm{Var}[\mathbf{B}], \tag{12a}
$$
$$
\mathrm{Var}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\} \approx \frac{\mathrm{Var}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]^2} - \frac{2\mathbb{E}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]^3}\mathrm{cov}(\mathbf{A},\mathbf{B}) + \frac{\mathbb{E}[\mathbf{A}]^2}{\mathbb{E}[\mathbf{B}]^4}\mathrm{Var}[\mathbf{B}], \tag{12b}
$$

where $\mathbf{A} \triangleq |\hat{p}_{xy_l}[k]|$ and $\mathbf{B} \triangleq \sqrt{\hat{p}_{xx}[k]\,\hat{p}_{yy}[k]} = \sqrt{L^2_{xy}[k]+Q^2_{xy}[k]} = \hat{p}_{xy}[k]$, where $L_{xy}[k]$ and $Q_{xy}[k]$ denote the real and imaginary parts of the ensemble cross spectra, respectively. Using Equation (12a) approximates the bias of estimators $|\hat{\theta}_{xy}[l,k]|$ and $|\hat{\Xi}_{xy}[l,k]|$ as $\mathbb{E}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\} \approx \frac{2.1}{\sqrt{\pi D_{w_d}}}$ and $\mathbb{E}\left\{\frac{\bar{\mathbf{A}}}{\mathbf{B}}\right\} \approx \frac{2.1}{M_1\sqrt{\pi D_{w_d}}}$, where $\bar{\mathbf{A}}$ and M_1 denote the cross spectra and the width of a smoothing window, $\upsilon[m]$, respectively (see Appendix). $D_{w_d} = R_w[0] + \frac{8}{\pi^2}R_w[1] + (1-\frac{8}{\pi^2})R_w[3]$ is the scaled correlation function of the Fourier transform of a weighting-window function, such that $R_w[d] = \sum_{m=1}^{M-d} W[m]\,W[m+d]$, where $W[k] = \mathcal{F}\{w(m)\}$ denotes the Fourier transform of a weighting window function $w(m)$ of length M [16]. In this article, $D_{w_d} \approx 1.8$ (see Appendix). Accordingly, using Equation (12b) approximates the variance of estimators $|\hat{\theta}_{xy}[l,k]|$ and $|\hat{\Xi}_{xy}[l,k]|$ as $\mathrm{Var}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\} \approx \frac{0.54}{D_{w_d}}$ and $\mathrm{Var}\left\{\frac{\bar{\mathbf{A}}}{\mathbf{B}}\right\} \approx \frac{0.54}{2M_1 D_{w_d}}$, respectively (see Appendix).

In order to verify the calculation of the bias and variance, Monte Carlo simulations have been performed for 1,000 realizations. Resulting bias and variance of the T-F coherence estimators, $|\hat{\Gamma}_{xy}[l,k]|$ and $|\hat{\Xi}_{xy}[l,k]|$, are given in Figure 1. It is shown that the bias of estimator $|\hat{\Xi}_{xy}[l,k]|$ (Method 3) (≈ 0.8) is lower than that of estimator $|\hat{\Gamma}_{xy}[l,k]|$ (Method 1) where a smoothing-window is not used. Naturally, the variance of estimator $|\hat{\Gamma}_{xy}[l,k]|$ (Method 1) goes to zero, because it attributes the value of 1 to all T-F coherence estimates when no smoothing is applied. On the other hand, the variance of estimator $|\hat{\Xi}_{xy}[l,k]|$ (Method 3) is not zero (≈ 0.4) when no smoothing is applied to the cross spectrum. This indicates that method 3 can differentiate between coherent
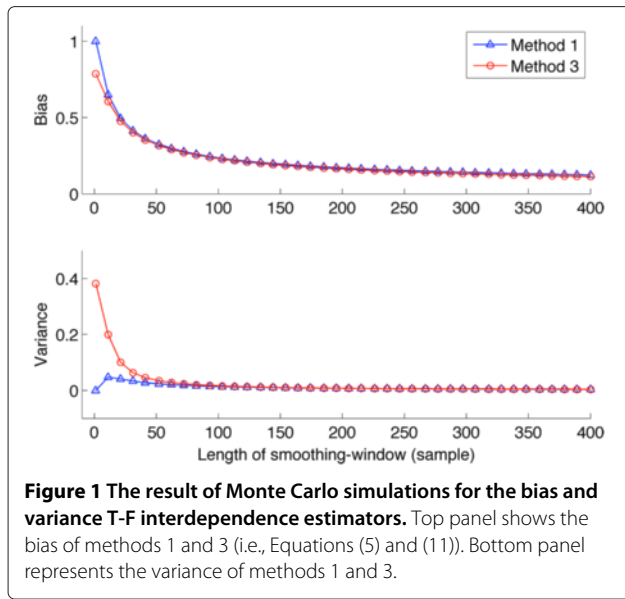
Figure 1 The result of Monte Carlo simulations for the bias and variance T-F interdependence estimators. Top panel shows the bias of methods 1 and 3 (i.e., Equations (5) and (11)). Bottom panel represents the variance of methods 1 and 3.

and non-coherent T-F points at maximal temporal resolution, as smoothing is not required. The bias and variance obtained from an analytic calculation (i.e., Equations (28a), (28b)) closely agree with simulation results shown in Figure 1, verifying the analyses performed to approximate the bias and variance of the present T-F interdependence estimators.

2.4 Simulated data

To compare three methods and test the use of non-identical smoothing operators, we derived from two simulated datasets with different properties:

(1) Signal pairs with time-varying coherence, i.e., Gaussian noise modulated at 0.6 and 1.2 Hz.
(2) Signal pairs with frequency-varying coherence, i.e., a jump in synchronous frequencies from 10 to 20 Hz.

After comparing the methods in these basic examples, we then explore their ability to characterize dynamic correlations in human EEG data, where the ground truth is not known in advance.

2.4.1 Simulated dataset 1

We first consider an example proposed by Lachaux et al. [8] to evaluate T-F coherence. Two independent discrete noise sequences $\{\alpha_{1_n}\}_{n=1}^N$ and $\{\alpha_{2_n}\}_{n=1}^N$ with mean 0 and variance 1 were drawn from normal distribution and modulated by a sinusoid signal $\beta[n]$, where N is the total number of samples. This modulation produces two discrete signals $S_1[n] = \alpha_1[n] + \beta[n]\alpha_2[n]$, and $S_2[n] = \alpha_2[n] + \beta[n]\alpha_1[n]$, which are synchronized and desynchronized over time. We defined a pure sinusoid signal $\beta[n]$ as $\beta[n] = a(1 + \sin[2\pi fn])$. Following Lachaux

et al. [8], we choose $f = 0.6$ Hz. In addition, to assess the temporal resolution of the different methods, we also consider $f = 0.9$ Hz and $f = 1.2$ Hz, i.e., in this example coherence is modulated at a higher frequency. The amplitude a determines the boundaries of the coherence, see Equation (13). Since $\alpha_1[n]$ and $\alpha_2[n]$ are Gaussian noise with variance $\sigma^2 = 1$, the power spectra of $\alpha_1[n]$ and $\alpha_2[n]$ are both equal to $\sigma^2 = 1$ for all frequencies, and their cross spectra are equal to 0. The coherence between signals $S_1[n]$ and $S_2[n]$ can analytically be defined as a function of $\beta[n]$ as follows:

$$\Upsilon[a, n] = \frac{4a^2 F^2[n]}{(1 + a^2 F^2[n])^2}, \tag{13}$$

where $F[n] = \frac{1}{a}\beta[n]$, and $a \in [0.01, 0.5]$. By substituting $a_{max} = 0.5$ and $a_{min} = 0.01$ into Equation (13), one respectively obtain two bounds for time-varying coherence, one is $0 \leq \Upsilon[0.5, n] \leq 1$ and the other is $0 \leq \Upsilon[0.01, n] \leq 1.6 \times 10^{-3}$.

The methods were assessed for different levels of coherence by considering $a \in [0.01, 0.5]$. To test for statistical significance of the estimated T-F interdependence, sets of surrogate data of linear independent signals were created. For each amplitude a, 100 surrogates were generated by randomizing the corresponding Fourier phases [33], i.e., by destroying the temporal structure in the data while leaving amplitudes intact. T-F interdependence was computed between the surrogate sets and 95% confidence intervals were determined for each method independently. The analytic coherence is continuously distributed and we assess the performance of the different methods by regressing estimated T-F interdependences on the analytic coherence.

2.4.2 Simulated dataset 2

In the second example, coherence varies both in time and frequency and we assess the methods by varying the signal-to-noise ratio (SNR). We simulated two independent discrete signals $S_3[n] = \left[x_1[n] + \eta_1[n], x_2[n] + \eta_2[n] \right]$ and $S_4[n] = \left[x_1[n] + \eta_3[n], x_2[n] + \eta_4[n] \right]$ of 20-s duration. Note that n denotes the discrete time index. $\eta[n]$ is Gaussian noise: $\eta[n] \sim \mathcal{N}(\mu_\eta = 0, \sigma_\eta^2 = 1)$. Two sinusoid signals $x_1[n] = A\sin[2\pi 10n]$ and $x_2[n] = A\sin[2\pi 20n]$ were contaminated by four independent Gaussian noise sequences $\eta_i[n], i = 1, 2, 3, 4$ as stated above. To produce signals with different amplitudes, we defined a range of amplitudes A by using a wide range of SNR,

$$A = A_{noise}\left(10^{\frac{SNR}{20}}\right). \tag{14}$$

The three methods were compared for different levels of SNR $\in [-30, 10]$ dB, where dB is a logarithmic unit that

indicates the ratio of a physical quantity (noise strength, signal strength, etc). Here, we set $A_{noise} = 0.2$. Performance of the methods was assessed by quantifying the sensitivity, specificity, and z-score. Sensitivity and specificity are statistical measures of the performance of a binary classification test defined as

$$\begin{aligned} Sensitivity &= \frac{TP}{TP + FN}\%, \\ Specificity &= \frac{TN}{TN + FP}\%, \end{aligned} \quad (15)$$

where TP, TN, FP, and FN denote number of true positives, true negatives, false positives, and false negatives, respectively. By construction, signals $S_3[n]$ and $S_4[n]$ are either fully correlated or uncorrelated for each point in the T-F plane, i.e., they are binary. To binarize the estimated T-F interdependence we again determined the 95% confidence interval by phase-randomizing the data (see Section 2.4.1), assigning all T-F points exceeding the threshold to 1 and the rest to 0. In addition, the z-scores were determined at each specific SNR as

$$Z = \frac{\bar{X} - \mu}{\sigma_x}, \quad (16)$$

where $\bar{X}$ denotes mean of the estimate at T-F points where true coherence is present and μ and σ_x the mean and standard deviation of the estimate at T-F points were true coherence is absent (null distribution). Note that the true coherence is represented in Figure 2.

2.4.3 Experimental data (human EEG)
We then sought an exploratory assessment of the different estimators of T-F interdependence in characterizing synchronization in an exemplar physiological dataset, namely human EEG. We studied the correlations between two scalp channels over occipital cortex, i.e., O_1 and O_2. EEG was acquired from three healthy human subjects while they were sitting still with their eyes closed for 10 min. EEG data were sampled at $Fs = 500$ Hz and digitized by an analog-to-digital convertor with a resolution of

16 bits/sample. The protocol was approved by the Human Research Ethics Committee of The University of New South Wales and all subjects gave their voluntary and informed consent. EEG data recorded over occipital areas were chosen because it is known to contain complex patterns of intermittent synchronization within the alpha frequency range (8–13 Hz), see [34].

Because the true coherence between empirically recorded signals is not known, we assess the methods based on the marginal densities of the covariance matrix. We first determine the time-domain and frequency-domain zero-lag covariance matrices. The time-domain covariance matrix of $|\hat{\Xi}_{xy}[l,k]|$ is estimated as $\hat{R}_{xy}^{(L \times L)} = \mathbb{E}\{|\hat{\Xi}_{xy}[l,k]| |\hat{\Xi}_{xy}^{H}[l,k]|\}$ where H denotes Hermitain transposition (i.e., conjugate transposition). Analogously, the frequency-domain covariance matrix is estimated as $\hat{R}_{xy}^{(K \times K)} = \mathbb{E}\{|\hat{\Xi}_{xy}^{H}[l,k]| |\hat{\Xi}_{xy}[l,k]|\}$. Therefore, the time-domain zero-lag covariance matrix of $|\hat{\Xi}_{xy}[l,k]|$ can be estimated as

$$\begin{aligned} \hat{R}_{xy}^{(L \times L)} = \frac{1}{K} \sum_{k=1}^{K} & \left(|\hat{\Xi}_{xy}[l_1,k]| - \mu_{l_1|k} \right) \\ & \times \left(|\hat{\Xi}_{xy}^{H}[l_2,k]| - \mu_{l_2|k} \right), l_1 \neq l_2 = 1,2,\ldots,L, \end{aligned} \quad (17)$$

where $\mu_{l_1|k}$ denotes the mean value of $|\hat{\Xi}_{xy}[l_1,k]|$ at time-window l_1 and given frequencies $k = 1,2,\ldots,K$. Similarly, the frequency-domain zero-lag covariance matrix $|\hat{R}_{xy}^{(K \times K)}|$ is expressed as

$$\begin{aligned} \hat{R}_{xy}^{(K \times K)} = \frac{1}{L} \sum_{l=1}^{L} & \left(|\hat{\Xi}_{xy}^{H}[l,k_1]| - \mu_{k_1|l} \right) \\ & \times \left(|\hat{\Xi}_{xy}[l,k_2]| - \mu_{k_2|l} \right), k_1 \neq k_2 = 1,2,\ldots,K. \end{aligned} \quad (18)$$

The diagonal of matrices $\hat{R}_{xy}^{(L \times L)}$ and $\hat{R}_{xy}^{(K \times K)}$ reflect the variance and reveal the dominant temporal interaction
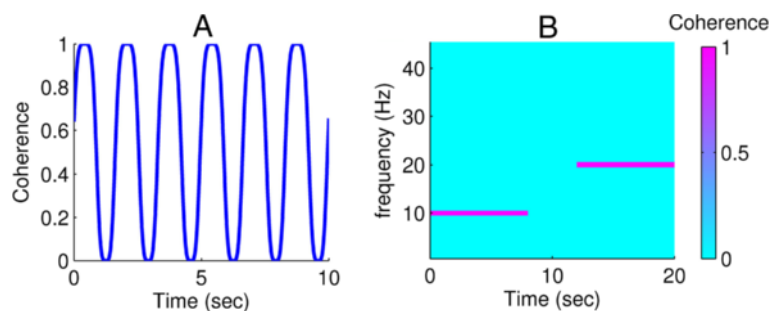

Figure 2 The analytic coherence for two simulated data sets. Panel **A** represents the analytic time-varying coherence in Example 1 obtained from Equation (13) at amplitude $a = 0.5$, and panel **B** shows the desired T-F coherence for simulated data of Example 2.

and spectral correlation between signals $x[n]$ and $y[n]$, which is used to compare the different methods.

3 Results

3.1 Simulations

Two basic datasets were generated to compare three methods on signals with known properties. Figure 2 shows the analytic (or the true-valued) time-varying coherence obtained from Equation (13) at amplitude $a = 0.5$ for dataset 1, and the desired T-F coherence for the frequency-varying coherence described in dataset 2. T-F interdependence was assessed with different estimators as defined in Equations (5), (9), and (11), which use either identical and non-identical smoothing operators. In

the first simulated dataset, the time course of all three methods follows the analytic time-varying coherence, although some differences can be observed (Figure 3). Using identical smoothing operators, estimated correlation levels are generally quite high (Figure 3A). Indeed, if we determine the number of statistically significant T-F points using phase-randomized surrogate data, we see that the null hypothesis is rejected in at least $> 50\%$ of the simulations even for T-F points where true coherence is absent (Figure 3B). These are false positive observations as the apparent statistical significance is high but the true rate is zero. Because analytic coherence is identical across frequencies, we investigate time-marginal density of all three methods, which again confirms the raised lower
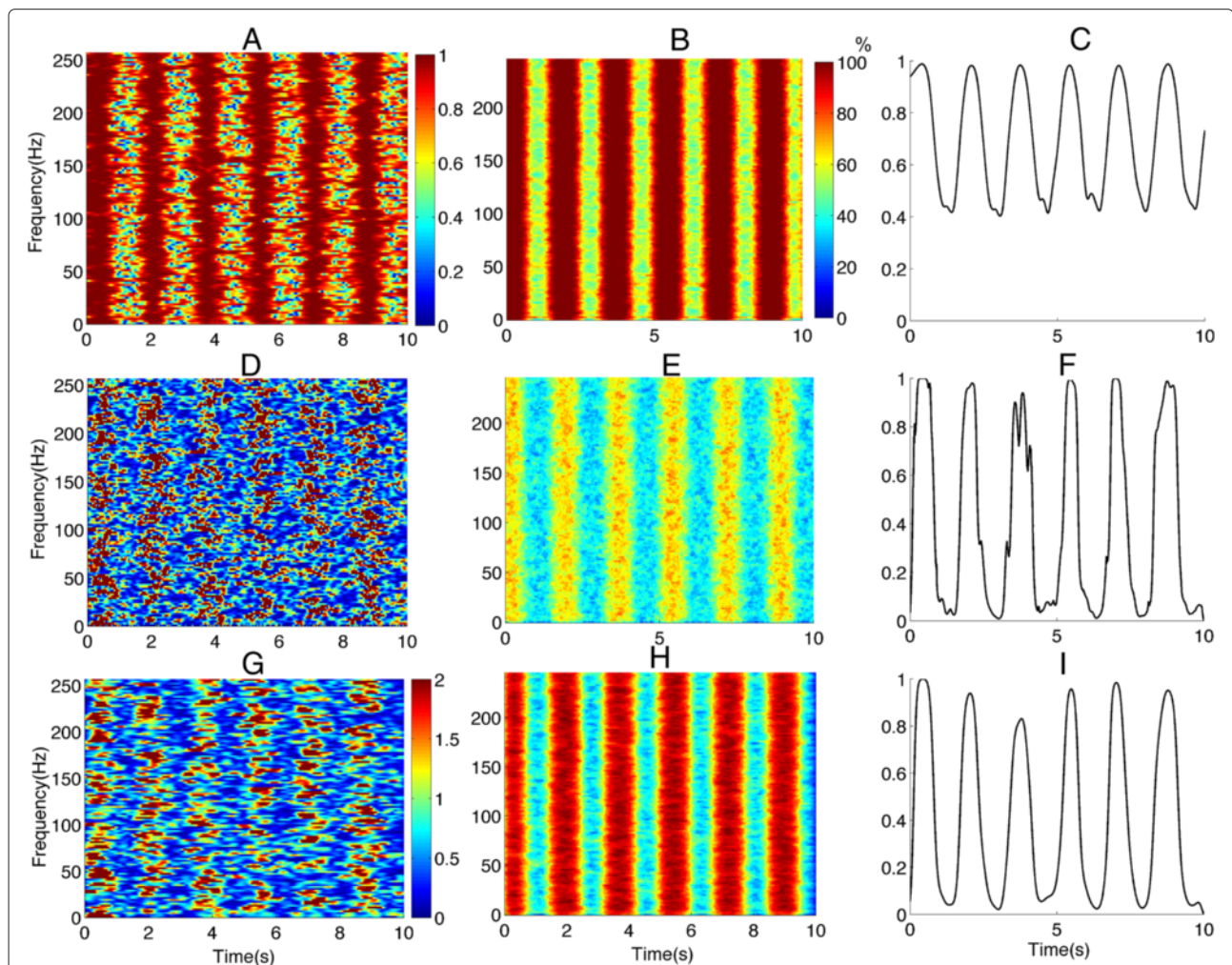

Figure 3 The resulting T-F coherence estimates using Equations (5), (9), and (11) for the first simulated dataset. All results were estimated using a 0.5-s Hamming weighting window and a 0.75-s Hamming smoothing window. The first simulated dataset ($a = 0.5$) compares three different estimators of T-F interdependence: method 1 using identical smoothing operators (top row) and methods 2 and 3 using non-identical smoothing operators (bottom two rows). Panels **A**, **D**, and **G** show the magnitude of the T-F estimate obtained from single simulation; panels **B**, **E**, and **H** show the percentage of simulations in which the null hypothesis was rejected (estimated across 100 simulations), and panels **C**, **F**, **I** the time-marginal density reflecting the time-course of the coherence estimate. In panels **F** and **I**, T-F coherence interdependence is projected within [0,1] by using $\tanh^{-1}(.)$ function for comparison purposes.

bound of the coherency estimate when using identical smoothing operators (Figure 3C).

As expected, not smoothing the CSDs, as described in Equation (5), strongly increases the variance of the estimate (Figure 3D). However, the estimator still reflects the analytic coherence as shown by the statistical analysis (Figure 3E) and its time-course (Figure 3F). The variance of the estimate is reduced after smoothing the CSDs, as described in Equation (11) (Figure 3G). Although the variance of the estimate is still higher compared to the first method based on identical smoothing operators, the range of the estimate has increased. The percentage of simulations in which the null hypothesis is rejected varies between 30 and 100% (Figure 3H) and the time course reveals a lower value around 0.15 (Figure 3I).

We then assessed the performance of the three estimators by linear regression of the estimated interdependence (Figure 4) against the analytic coherence (Figure 2). The top row of Figure 4 shows an example of the regression lines for the three methods. Using identical smoothing operators, estimated coherency is bounded in [0,1], the estimate is largely restricted to values > 0.5 as reflected by the large offset (0.55) and small slope (0.37) of the regression line (Figure 4A). The estimator predicts analytic coherence well, as shown by the correlations coefficient (r) of 0.67. Using non-identical smoothing operators, estimated coherence is no longer bounded in [0,1] and reveals increased variance (Figure 4B,C). However, the regression line is closer to that of an ideal estimator ($X = Y$), as revealed by a lower offset (0.41) and a larger slope (0.69)

(Figure 4C). Due to larger variation in the estimate, the correlation coefficient for non-identical smoothing operators is lower ($r = 0.47$) than for identical smoothing operators. This is true for the whole range of amplitudes ($a \in [0.01, 0.5]$) when coherence is modulated at 0.6 Hz (Figure 4D). Methods 1 and 3 use a temporal smoothing window with a length of 0.75 s. If coherence is modulated at 1.2 Hz (or 0.833 s period length), the changes in coherence within each cycle get smoothed out as the length of the smoothing window is close to the period length of the modulation. Method 2 does not involve smoothing of the cross spectrum and hence is still able to capture the cyclic changes in coherence. When coherence is modulated at lower rates (e.g., 0.6 Hz) methods 1 and 3 perform better as the smoothing window is shorter than the period length of the modulation of coherence and smoothing improves the SNR. This result confirms the superior temporal resolution of method 2 in this setting. As Figure 4D–F shows, there is a decreasing trend in the r values of methods 1 and 3 due to the smoothing window. This indicates that method 1 reduces the variance of estimates when using a smoothing operator, but this comes at the expense of reducing the temporal resolution. This is revealed by a decreasing value of r. As shown in Figure 4D–F, method 2 does not deviate far from $r \sim 0.2$ while the modulation frequency increases.

In the second dataset, coherence is varied in time and frequency reflecting a jump in time from 10 to 20 Hz (Figure 2B). Again, all three methods are able to detect this coherence pattern (Figure 5, top row), i.e., showing correlations at frequencies $f_1 = 10$ Hz (from 0 to
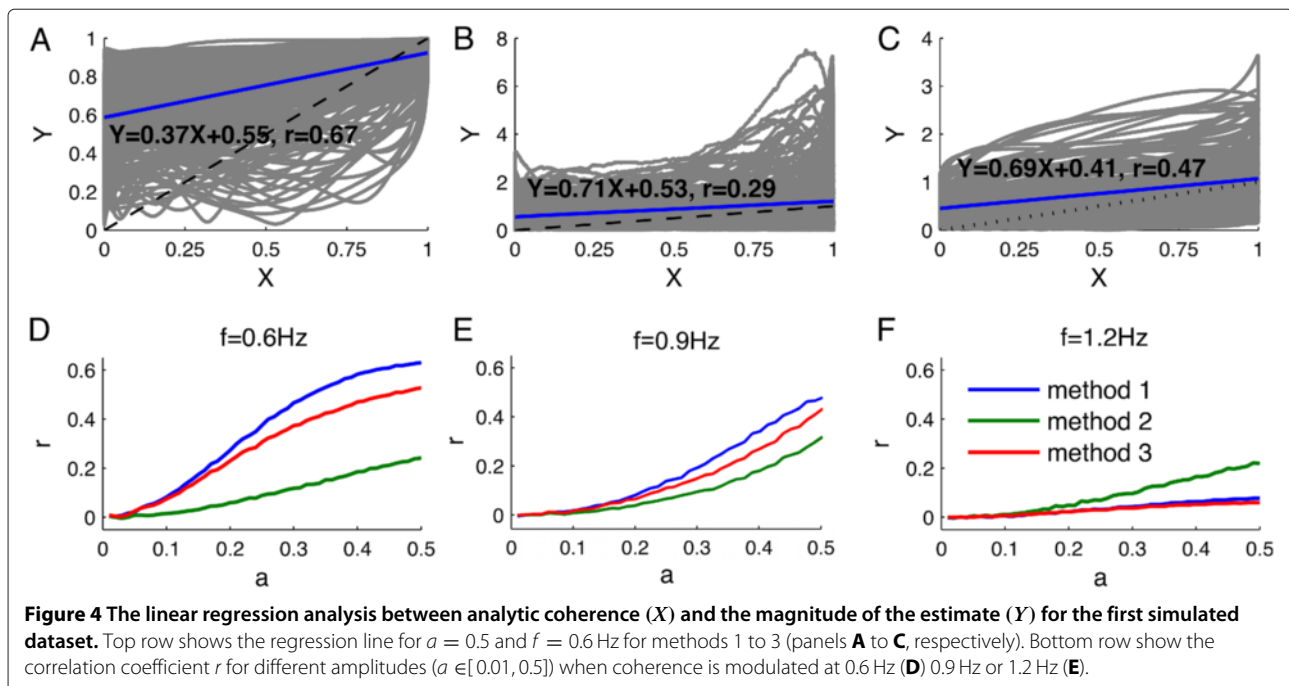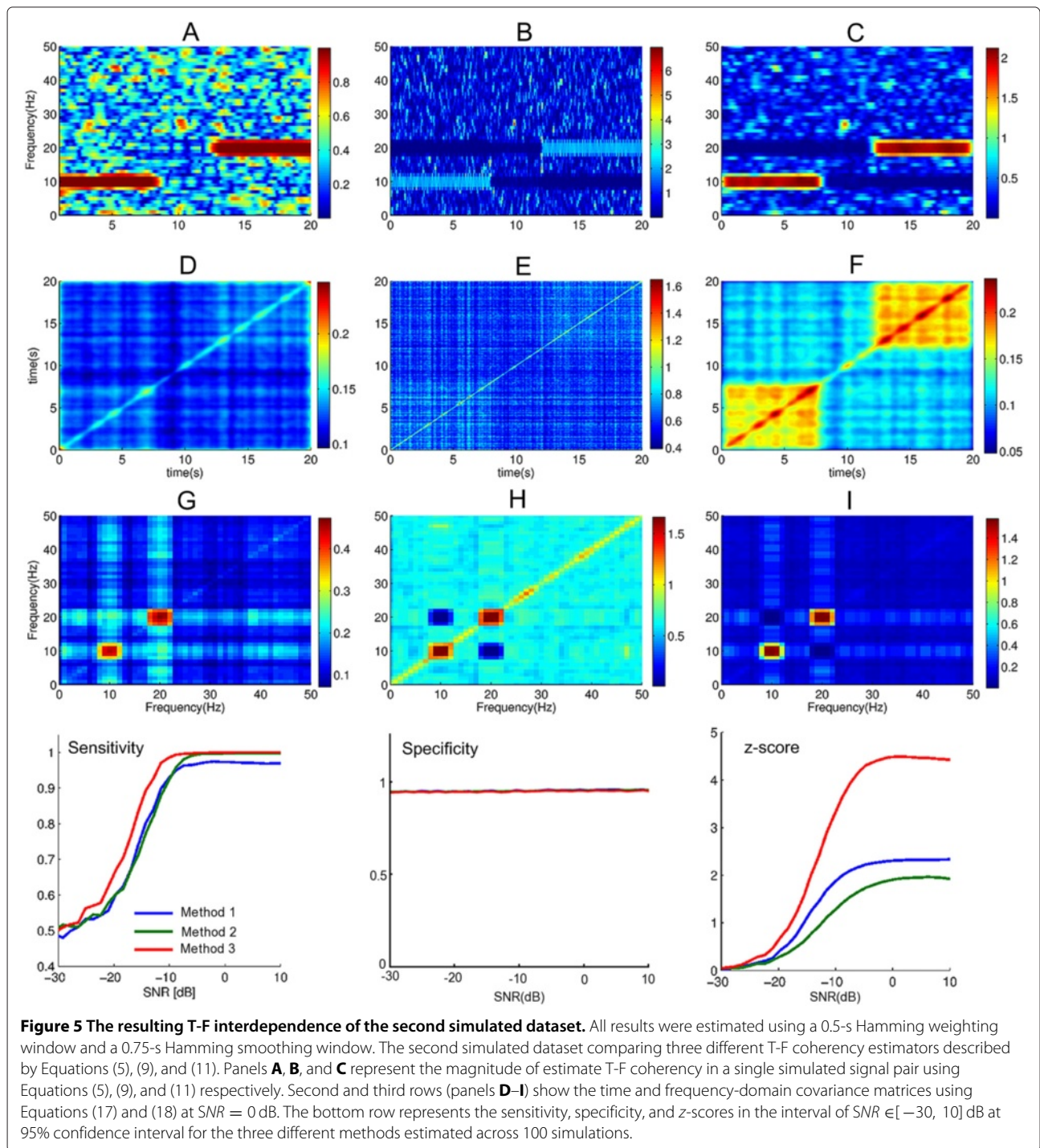

Figure 4 The linear regression analysis between analytic coherence (X) and the magnitude of the estimate (Y) for the first simulated dataset. Top row shows the regression line for $a = 0.5$ and $f = 0.6$ Hz for methods 1 to 3 (panels **A** to **C**, respectively). Bottom row show the correlation coefficient r for different amplitudes ($a \in [0.01, 0.5]$) when coherence is modulated at 0.6 Hz (**D**) 0.9 Hz or 1.2 Hz (**E**).

Figure 5 The resulting T-F interdependence of the second simulated dataset. All results were estimated using a 0.5-s Hamming weighting window and a 0.75-s Hamming smoothing window. The second simulated dataset comparing three different T-F coherency estimators described by Equations (5), (9), and (11). Panels **A**, **B**, and **C** represent the magnitude of estimate T-F coherency in a single simulated signal pair using Equations (5), (9), and (11) respectively. Second and third rows (panels **D**–**I**) show the time and frequency-domain covariance matrices using Equations (17) and (18) at $SNR = 0$ dB. The bottom row represents the sensitivity, specificity, and z-scores in the interval of $SNR \in [-30, 10]$ dB at 95% confidence interval for the three different methods estimated across 100 simulations.

8 s) and $f_2 = 20$ Hz (from 12 to 20 s). Similar to the results from dataset 1, method 1 (based on identical smoothing operators) shows the highest level of background coherence (Figure 5A). In contrast, coherence levels for method 2 are relatively reduced for both the synchronous T-F points as the non-coherent background (Figure 5B). The smoothed version of method 2 shows

higher levels at the coherent T-F points combined with low background levels (Figure 5C). The time and frequency marginal densities of method 3 also correctly reflect the coherent time points (Figure 5F) and frequencies (Figure 5I), whereas method 1 shows more spurious off-diagonal correlations (Figure 5G). In this example, the input variable in binary, i.e., coherence is either 0 or 1, and

we use signal detection theory to quantify performance. Estimated T-F interdependence was binarized by defining a 95% confidence interval using a resampling technique and all T-F points were set at 0 or 1 depending on this threshold. We then quantified the sensitivity and specificity of the three methods. The specificity was constant at 0.95 because we used a 95% confidence interval to binarize the data. The sensitivity increased for all three methods with increasing SNRs showing a jump from 0.5 to approximately 1 around SNR of -10 dB. Method 3 had slightly higher sensitivity at these intermediate SNR levels, although the difference was marginal. The z-score, reflecting the ratio between estimated interdependence at correlated and uncorrelated T-F points (Equation (16)), showed a similar increase at higher SNR levels. Method 3 has the highest z-score across all levels of SNR reflecting the increased contrast between correlated and uncorrelated T-F points.

These two basic simulations show that using non-identical smoothing operators and selectively smoothing the numerator and denominator of the T-F coherence function increases the range of the estimate and allows detecting a broader range of correlations. The increased range helps to decrease the number of false positives and reduces spurious correlation patterns. However, the increased range comes at the cost of increased variance of the estimate.

3.2 EEG data
We finally apply three methods to empirical data to characterize dynamic patterns of interdependences between left and right of human occipital cortexes, as reflected in surface EEG recordings at channels O_1 and O_2. Data segments for three subjects are shown in the top row of Figure 6 revealing intermittent episodes of alpha activity. The T-F interdependences estimated with method 1

show high levels of correlations uniformly across time and frequency (Figure 7, row A). In contrast, methods 2 and 3 reveal a more heterogeneous T-F estimate reflecting a larger range of estimated values (Figure 7, rows B and C, respectively). Because the underlying properties of the signals are unknown, we use the frequency-marginal densities of the estimates (see Equation (18)) to obtain the correlated frequencies. Method 1 revealed uniformly correlated frequencies, i.e., interdependences are not restricted to any particular frequency and numerous off-diagonal correlations are present. In contrast, the frequency-domain covariance matrix of methods 2 and 3 shows specific correlated frequencies between the two EEG signals, O_1 and O_2 isolated to the alpha frequency range (8–12 Hz) in all three subjects (Figure 7 at rows E and F). Although some off-diagonal correlations are present, they are strongly reduced compared to method 1.

4 Discussion
In this article, we assess the use of non-identical smoothing operators for estimating T-F interdependence. Traditionally, the cross and auto spectra are smoothed using identical smoothing operators to assure that the estimate is bound within [0,1]. By relaxing this constraint, we examine the possibility of selectively smoothing the auto spectra to improve the trade-off between T-F resolution and statistical consistency. Analytic and simulation analysis of the bias and variance showed that the use of non-identical smoothing operators reduces the bias of the estimate when smoothing vanishes, which allows to design an estimator with maximal temporal resolution. However, the reduced bias comes at the cost of increased variance of the estimate and the estimate is no longer bound within [0,1]. Two stimulated datasets showed that the reduced bias results in an increased range and reduced
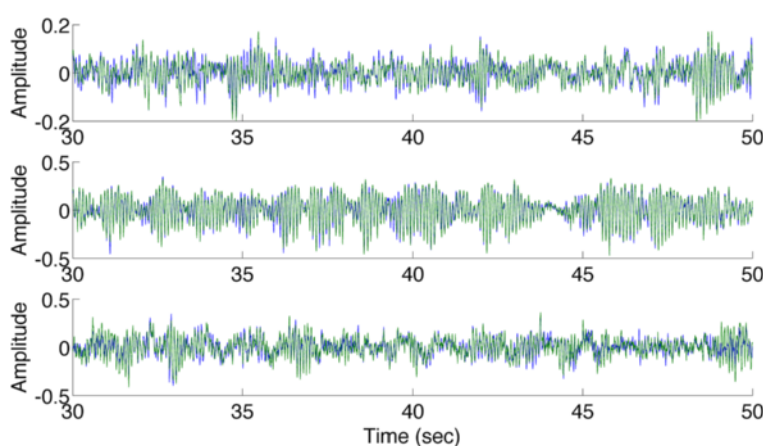

Figure 6 The occipital EEG signals O_1 and O_2 recorded from three human subjects showing intermitted episodes of alpha activity.
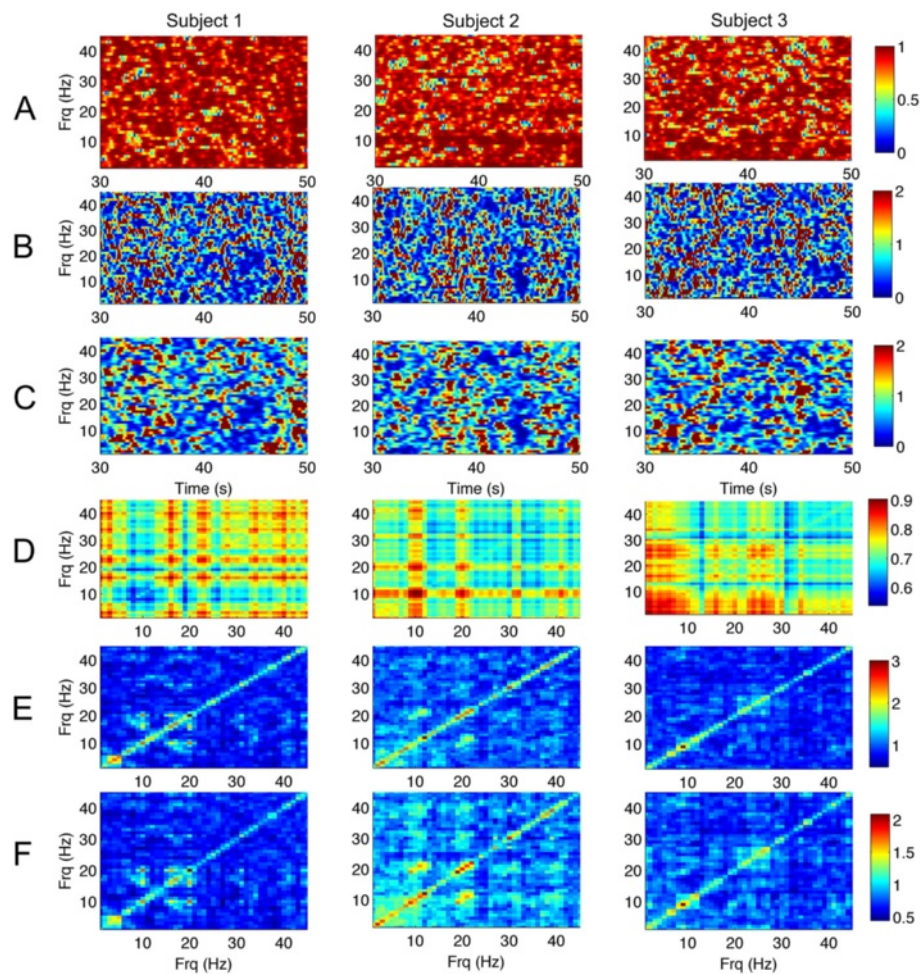
Figure 7 T-F interdependences between occipital EEG signals O_1 and O_2 illustrated here for three human subjects. All results were estimated using a 0.5-s Hamming weighting window and a 0.75-s Hamming smoothing window. Each column represents the results for each subject. Upper panels—taken place in rows **A**, **B**, and **C**—represent the magnitude T-F interdependence estimates using Equations (3), (9), and (11), respectively, for three subjects. Lower panels—taken place in rows **D**, **E**, and **F**—represent the frequency-domain zero-lag covariance matrix (using Equation (18)) of the magnitude T-F interdependence estimates shown in rows of **A**, **B**, **C**, respectively.

false positive of the estimate, but lower correlation with the true distribution due to increased variance. When applying these methods to two occipital EEG channels, the reduced range of the estimate obtained with identical smoothing operators resulted in spurious correlation patterns. In contrast, the estimator based on non-identical smoothing operators revealed a larger range and selectively detected correlations in the alpha frequency band. Together these results show that non-identical smoothing operators may be useful when high temporal resolution or low false positive rates are required.

All three estimators are based on time-domain smoothing operators. In the first method, the cross and auto spectra are smoothed with a short temporal window. In the second method, the temporal CSD is normalized by the ensemble averaged PSDs. Finally, in the third method,

the smoothed CSD is normalized by the ensemble averaged PSDs. Two datasets were constructed to compare the three estimators. In the first dataset coherence changed continuously over time. Using non-identical smoothing operators (method 3) resulted in an increased range and reduced bias compared to identical smoothing operators (method 1), as reflected by a reduction in false positives (Figure 3H) and a regression line with a smaller off-set and increased slope (Figure 4C). However, the regression coefficient, r, was lower for method 3 than for method 1 due to increased variance of the estimate. Using identical smoothing operators (method 1), the combination of higher bias and lower temporal resolution may lead to spurious correlations (false positives), as high correlations spread in the T-F plane. In particular, we found that wider smoothing window yields spurious coherent

frequencies where a narrower weighting-window is used (see Figures 7 and 8). In this framework, the role of non-identical smoothing operator is to avoid production of spurious coherent frequencies and improve the temporal resolution of the estimate.

In the second dataset, coherence changes discretely reflecting a jump from one frequency to another over time. Method 3 showed a slightly higher sensitivity and increased z-score reflecting the increase in contrast between correlated T-F points and background levels (Figure 5). When applying these estimators to resting-state EEG to detect T-F interdependences between two occipital channels, using non-identical smoothing operators suggested a broader range of estimated correlations across frequencies and an apparent reduction of uniform correlations across all frequencies (Figure 7, row D), isolating the alpha rhythm as the dominant synchronous oscillation in these scalp EEG data. The increased range and reduced bias of method 3 may hence be useful to detect T-F interdependences in electrophysiological data, such as EEG, in which correlations are sparse. Under these circumstances it may be more important to reduce the number of false positives by reducing background coherences levels.

The two estimators using non-identical smoothing operators were derived by normalizing the temporal CSD by the ensemble-averaged PSDs, instead normalizing each T-F area in the CSD by the PSDs in the same area. This results in an increase range of the estimate when the PSDs are not constant over time. That is, because the CSDs are normalized by the average spectral power, estimated T-F interdependences will fluctuate with temporal fluctuations in power. Although this may not be desirable for all types of signals, several studies have shown that modulations in synchronization and spectral power often coincide in neural data [35,36], which may improve the detection of T-F interdependence using non-identical smoothing operators. Another consequence of normalizing the CSD by the ensemble-averaged PSD is that only the frequency marginal density of the (squared) estimate is bound within [0,1] but the time-resolved estimate itself not. Although the scaling of the time-resolved measure is essentially identical to coherence as shown by its frequency marginal density, it can be projected on the interval of [0,1] using Fisher transform to facilitate interpretation (see Figure 3F, I and [37]). For the sake of simplicity we used a normalized Hamming window as a temporal smoothing window to compare the use of non-identical smoothing operators. There is however no limitations for using more sophisticated smoothing functions, such as 2D smoothing operators [18] or a multitaper technique [21].

5 Conclusion

By relaxing the constraint of using identical smoothing operators for the cross and auto spectra, we open up
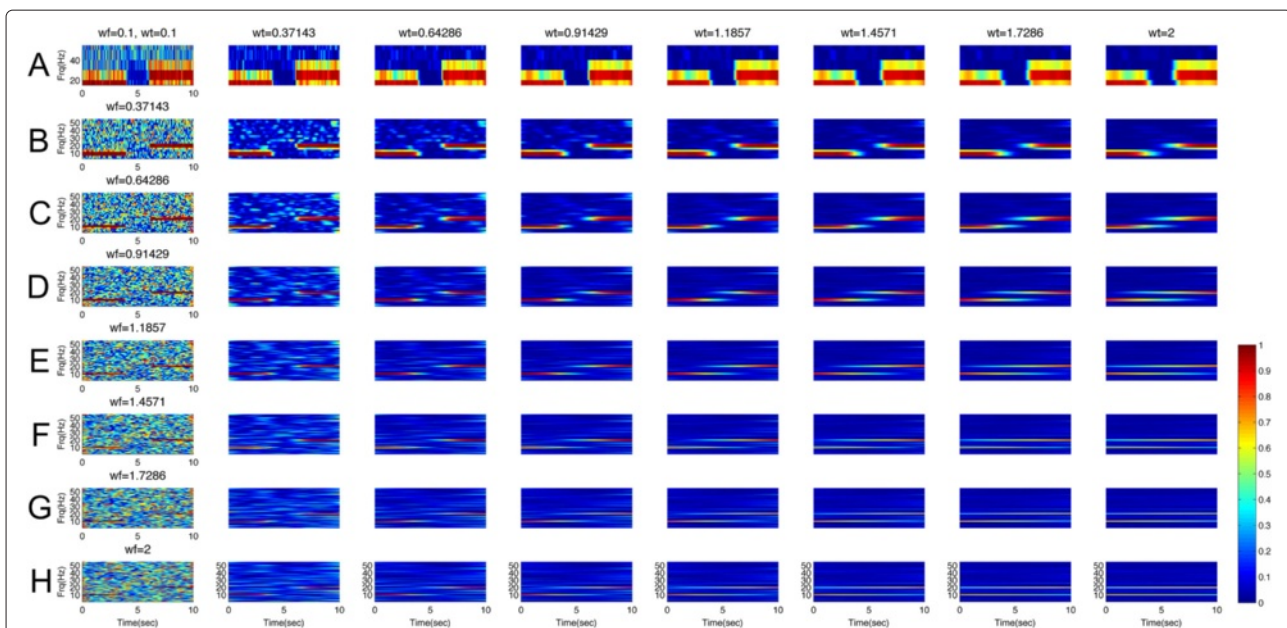

Figure 8 The effect of the width of the smoothing window, *wt*, on the T-F interdependence derived from Equation (5) as a function of the length of the weighting-window, *wf*. (**A**) T-F interdependence estimates using Hamming-weighting window of the width $wf = 0.1$ s and varying-width smoothing window, $wt = 0.1, 0.37143, \ldots, 2$ s, rows (**B–G**) illustrate the T-F interdependence estimates using different lengths of the weighting and smoothing windows, respectively.

broader possibilities for the estimation of interdependence in the T-F space. We specifically considered an example of non-identical smoothing operator by normalizing the temporal CSD by the ensemble-averaged PSDs. This method revealed a reduced bias and increased range at the cost of increased variance of the estimate and may be useful for detecting T-F interdependences in resting-state EEG where correlations between signals are sparse and changes over time.

Appendix

In this section, we provide the analytic calculation of the bias and variance of the present T-F interdependence estimators. Let us consider two discrete, independent real-valued random processes, drawn from a Gaussian distribution with mean of zero and variance of σ_x^2. We denote these two random processes $\{x_n\} \sim \mathcal{N}(0, \sigma_x^2)$ and $\{y_n\} \sim \mathcal{N}(0, \sigma_y^2)$. We now derive the bias and variance of the present T-F interdependence estimators of these two processes.

Bias and variance of the T-F interdependence estimator $|\hat{\theta}_{xy}[l,k]|$

We define the bias of $|\hat{\theta}_{xy}[l,k]|$ as

$$\text{Bias} \triangleq \mathbb{E}\left\{|\hat{\theta}_{xy}[l,k]|\right\} - |\theta_{xy}[l,k]| = \mathbb{E}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\}, \quad (19a)$$

$$\mathbf{A} \triangleq |\hat{p}_{xy_l}[k]| = \sqrt{L_{xy_l}^2[k] + Q_{xy_l}^2[k]} = |X_l[k]\, Y_l^*[k]|, \quad (19b)$$

$$\mathbf{B} \triangleq \sqrt{\hat{p}_{xx}[k]\,\hat{p}_{yy}[k]} = \sqrt{L_{xy}^2[k] + Q_{xy}^2[k]} = \sqrt{\hat{p}_{xy}[k]}, \quad (19c)$$

where $\mathbb{E}$ denotes the mathematical expectation, $\theta_{xy}[l,k]$ the true T-F interdependence between two independent random variables x and y, i.e., $\theta_{xy}[l,k] = 0$, $\mathbf{A}$ and $\mathbf{B}$ the two random variables denoting magnitude of cross spectra at lth segment and magnitude of the product of power spectra, respectively, and $L_{xy}[k]$ and $Q_{xy}[k]$ the real and imaginary parts of the ensemble cross spectra, respectively. Therefore, the key point is to approximate the term $\mathbb{E}\{\frac{\mathbf{A}}{\mathbf{B}}\}$. Analogously, we define the variance of $|\hat{\theta}_{xy}[l,k]|$ as

$$\begin{aligned} \text{V} = \text{Var}\left\{|\hat{\theta}_{xy}[l,k]|\right\} &= \mathbb{E}[\,|\hat{\theta}_{xy}[l,k]|^2\,] \\ &- \mathbb{E}[\,|\hat{\theta}_{xy}[l,k]|\,]^2 = \text{Var}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\}. \end{aligned} \quad (20)$$

Since $|\hat{\theta}_{xy}[l,k]|$ is a function of two random variables, and hence we may write $|\hat{\theta}_{xy}[l,k]|$ as a form of $|\hat{\theta}_{xy}[l,k]| = f(\mathbf{A}, \mathbf{B})$. We can generalize the expectation and variance of a function of two or more random variables using Taylor series expansion. In statistical signal processing, the expansion is evaluated at the expected value and variance of a random variable, X. For example, for an univariate

random variable, evaluation of the expansion of $f(X)$ at $\mathbb{E}[X]$ is calculated as

$$\begin{aligned} f(X) = {}& f(\mathbb{E}[X]) + f'(\mathbb{E}[X])(X - \mathbb{E}[X]) \\ & + \frac{f''(\mathbb{E}[X])(X - \mathbb{E}[X])^2}{2!} \\ & + \frac{f'''(\mathbb{E}[X])(X - \mathbb{E}[X])^3}{3!} + \text{H.O.T.}, \end{aligned} \quad (21)$$

where H.O.T. stands for the higher-order terms, and $f'(.)$ and $f''(.)$ denote the first and second derivatives of a function of random variable X, respectively. It is easy to show the expected value and variance of $f(X)$ by using the Taylor expansion

$$\begin{aligned} \mathbb{E}\{f(X)\} &= \mathbb{E}[f(\mu_X + (X - \mu_X))], \\ &\approx \mathbb{E}\Big[f(\mu_X) + f'(\mu_X)(X - \mu_X) \\ &\qquad + \frac{1}{2}f''(\mu_X)(X - \mu_X)^2 \Big], \\ &\approx f(\mu_X) + \frac{f''(\mu_X)}{2}\sigma_X^2, \end{aligned} \quad (22)$$

noting that $\mathbb{E}[X - \mu_X] = 0$, where μ_X and σ_X^2 are the mean and variance of X. The variance of $f(X)$ is derived from variance definition and the Taylor expansion:

$$\begin{aligned} \text{Var}\{f(X)\} &= \mathbb{E}[f^2(X)] - \mathbb{E}^2[f(X)] \\ &\approx \text{Var}(X)f'(\mathbb{E}[X])^2 \approx \sigma_X^2 f'(\mathbb{E}[X])^2. \end{aligned} \quad (23)$$

It is now possible to generalize this concept to function of more than one random variable using multivariate Taylor expansion, which is referred to as delta method [38-40]. Using the delta method, the variance of a function of two random variables X_1 and X_2 are defined as

$$\text{Var}[f(X_1, X2)] \approx \left(\frac{\partial f}{\partial X_1}\right)\left(\frac{\partial f}{\partial X_2}\right)\text{Cov}(X_1, X_2). \quad (24)$$

Applying multivariate Taylor expansions of Equation (22), and (24) to a ratio of two random variables $\mathbf{A}$, see Equation (19b), and $\mathbf{B}$, see Equation (19c), we define the bias and variance of $\hat{\theta}_{xy}[l,k]$ as [39-41]

$$\mathbb{E}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\} \approx \frac{\mathbb{E}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]} - \frac{\text{cov}(\mathbf{A}, \mathbf{B})}{\mathbb{E}[\mathbf{B}]^2} + \frac{\mathbb{E}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]^3}\text{Var}[\mathbf{B}], \quad (25a)$$

$$\text{Var}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\} \approx \frac{\text{Var}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]^2} - \frac{2\mathbb{E}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]^3}\text{cov}(\mathbf{A}, \mathbf{B}) + \frac{\mathbb{E}[\mathbf{A}]^2}{\mathbb{E}[\mathbf{B}]^4}\text{Var}[\mathbf{B}]. \quad (25b)$$

The terms shown in Equations (25a) and (25b) give an approximation of the bias and variance of the T-F interdependence estimator $|\hat{\theta}_{xy}[l,k]|$. It was shown that the standardized $\mathbf{A}$–i.e., $\sqrt{\dfrac{L_{xy_l}^2[k]}{\sigma_{1_l}^2} + \dfrac{Q_{xy_l}^2[k]}{\sigma_{2_l}^2}}$, where $\sigma_{1_l}^2 = \text{Var}[L_{xy_l}[k]] = \sigma_{2_l}^2 = \text{Var}[Q_{xy_l}[k]] = \frac{1}{2}\sigma_x^4\sigma_y^4$ (see [42])—has a Rayleigh-distribution with the PDF $f_Z(z) =$

$ze^{-z^2/2}$, $z \geq 0$ whose mean and variance are $\sqrt{\pi/2}$ and $(4 - \pi)/2$, respectively [43]. So, the mean and variance of **A** are calculated as

$$\mathbb{E}[\mathbf{A}] = \sqrt{\frac{\pi}{2}} \sigma_{1_l} = \sqrt{\frac{\pi}{2}} \left(\frac{1}{\sqrt{2}} \sigma_x^2 \sigma_y^2 \right) = \frac{\sqrt{\pi}}{2} \sigma_x^2 \sigma_y^2, \tag{26a}$$

$$\mathrm{Var}[\mathbf{A}] = \frac{4 - \pi}{2} \left(\frac{1}{2} \sigma_x^4 \sigma_y^4 \right) = 0.21 \sigma_x^4 \sigma_y^4. \tag{26b}$$

Analogously, since random variable **B** is the square-root of the product of the ensemble power spectra $\hat{p}_{xx}[k]$ and $\hat{p}_{yy}[k]$ derived from the WOSA approach, it can be shown that the standardized **B**—i.e., $\sqrt{\frac{L_{xy}^2[k]}{\sigma_1^2} + \frac{Q_{xy}^2[k]}{\sigma_2^2}}$, where $\sigma_1^2 = \mathrm{Var}[L_{xy}[k]] = \sigma_2^2 = \mathrm{Var}[Q_{xy}[k]] = \frac{1}{2}\sqrt{D_{w_d}} \sigma_x^4 \sigma_y^4$—has a Rayleigh-distribution with PDF $f_Z(z) = ze^{-z^2/2}$, $z \geq 0$ whose mean and variance are $\sqrt{\pi/2}$ and $(4-\pi)/2$, respectively. Therefore, the mean and variance of **B** are given by

$$\mathbb{E}[\mathbf{B}] = \sqrt{\frac{\pi}{2}} \sigma_1 = \sqrt{\frac{\pi D_{w_d}}{2}} \left(\frac{1}{\sqrt{2}} \sigma_x^2 \sigma_y^2 \right) \tag{27a}$$

$$= \frac{\sqrt{\pi D_{w_d}}}{2} \sigma_x^2 \sigma_y^2, \tag{27b}$$

$$\mathrm{Var}[\mathbf{B}] = \frac{4 - \pi}{2} (\frac{1}{2} D_{w_d} \sigma_x^4 \sigma_y^4) = 0.21 D_{w_d} \sigma_x^4 \sigma_y^4, \tag{27c}$$

where $D_{w_d} = R_w[0] + \frac{8}{\pi^2} R_w[1] + (1 - \frac{8}{\pi^2}) R_w[3]$ is the scaled correlation function of the Fourier transform of a weighting window function, such that $R_w[d] = \sum_{m=1}^{M-d} W[m] W[m+d]$, where $W[k] = \mathcal{F}\{w(m)\}$ denotes the Fourier transform of a weighting-window function $w(m)$ of length M [16]. We approximated the value of D_{w_d} when $M = 100$ as $D_{w_d} \approx 1.8$ by using a simulation. Note that $\mathrm{cov}(\mathbf{A}, \mathbf{B}) = 0$ because Rayleigh random variables **A** and **B** are uncorrelated since processes $\{x_n\}$ and $\{y_n\}$ are drawn from Gaussian distribution [42]. Therefore, Equations (25a) and (25b) can be rewritten as

$$\mathbb{E}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\} \approx \frac{\mathbb{E}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]} + \frac{\mathbb{E}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]^3} \mathrm{Var}[\mathbf{B}] \approx \frac{2.1}{\sqrt{\pi D_{w_d}}} \approx 0.88, \tag{28a}$$

$$\mathrm{Var}\left\{\frac{\mathbf{A}}{\mathbf{B}}\right\} \approx \frac{\mathrm{Var}[\mathbf{A}]}{\mathbb{E}[\mathbf{B}]^2} + \frac{\mathbb{E}[\mathbf{A}]^2}{\mathbb{E}[\mathbf{B}]^4} \mathrm{Var}[\mathbf{B}] \approx \frac{0.54}{D_{w_d}} \approx 0.3. \tag{28b}$$

Bias and variance of T-F interdependence estimator $|\hat{\Xi}_{xy}[l, k]|$

T-F interdependence, $|\hat{\Xi}_{xy}[l, k]|$, has a smoothed cross spectra (see Equation (11)), whereas $|\hat{\theta}_{xy}[l, k]|$ does not. The mean and variance of denominators of Equations (9) and (11) are however identical. Torrence and Compo [7] and Torrence and Webster [17] showed smoothing in both time and frequency domains, whereas we only use a time domain smoothing operator, which is implemented by a
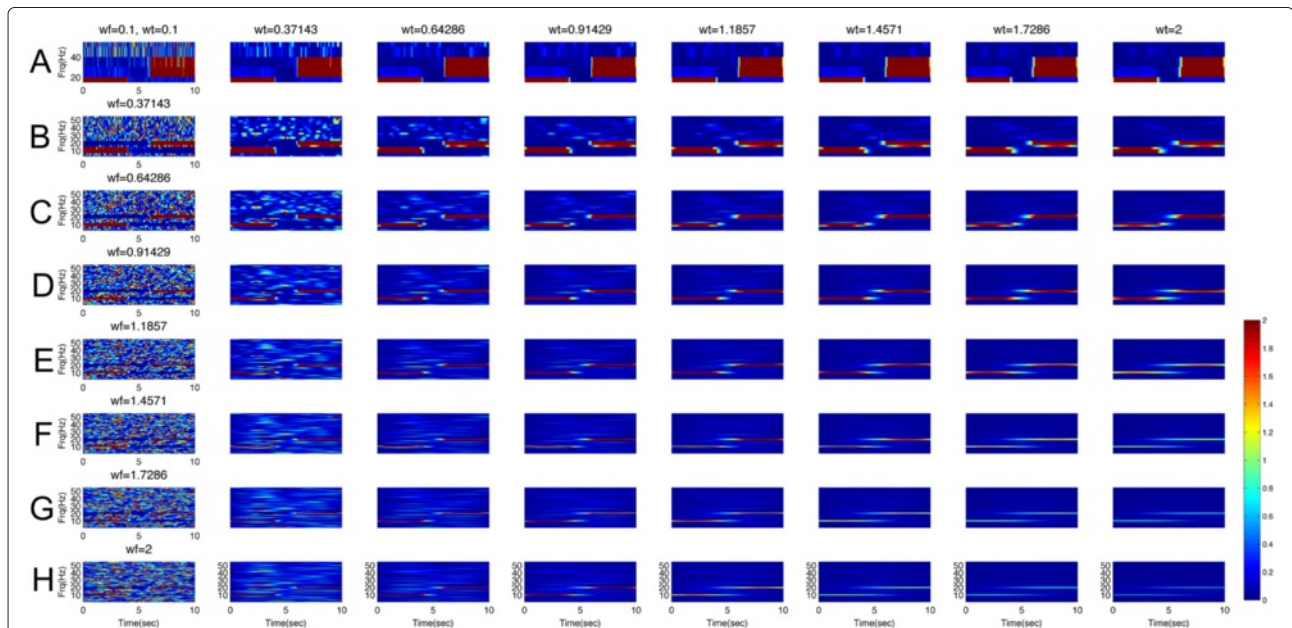

Figure 9 The effect of the width of the smoothing window, *wt*, on the T-F interdependence derived from Equation (11) as a function of the length of the weighting-window, *wf*. (A) T-F interdependence estimates using Hamming-weighting window of the width $wf = 0.1$ s and varying-width smoothing window, $wt = 0.1, 0.37143, \ldots, 2$ s, rows (**B–G**) illustrate the T-F interdependence estimates using different lengths of the weighting and smoothing windows, respectively.

running average (or convolution). It was originally shown that the variance of the smoothed cross spectra can be reduced by the factor of $[\approx \frac{1}{2M_1}]$, where M_1 is the length of smoothing window in seconds (or samples) [42]. We show the effect of smoothed cross spectra in the bias and variance of T-F interdependence estimator $|\hat{\Xi}_{xy}[l,k]|$. A normalized Hamming window was used as the smoothing window in this study. Therefore, the bias and variance of the T-F interdependence estimator $|\hat{\Xi}_{xy}[l,k]|$ can be approximated by using Equations (25a) and (25b)

$$\mathbb{E}\left\{\frac{\bar{A}}{\mathbf{B}}\right\} \approx \frac{2.1}{M_1\sqrt{\pi D_{w_d}}}, \tag{29a}$$

$$\mathrm{Var}\left\{\frac{\bar{A}}{\mathbf{B}}\right\} \approx \frac{0.54}{2M_1 D_{w_d}}. \tag{29b}$$

Noting that $\mathbb{E}[\bar{A}] = \frac{1}{M_1}\mathbb{E}[\mathbf{A}]$, where $\bar{A}$ denotes the smoothed cross spectra using a moving average (or convolution, see the numerator of Equation (11)). Equation (29b) indicates that the bias and variance are reduced by a factor of $\frac{1}{M_1}$ and $\frac{1}{2M_1}$ compared to Equations (28a) and (28b), respectively. The main disadvantage of large M_1 is the reduction of the time resolution by changing the distribution of time points of the T-F plane at a given frequency k. That is, larger M_1 induces uniform distribution to the temporal evolution of T-F points at any given frequency k (Figure 8 bottom rows of the right columns).

Empirically, the length of the smoothing window M_1 should be close to but longer than weighting window used in the Fourier transform. Figure 8 shows the results using identical smoothing windows (method 1), while Figure 9 shows the results for non-identical smoothing windows (method 3). The columns represent changes in length of the weighting window in the T-F decomposition, whereas the rows represent changes in the length of the smoothing window. It can be seen that applying a wider smoothing window to the T-F plane reduces the variance of the estimated at the cost of the reduction of the temporal resolution.

Competing interests

The authors declare that they have no competing interests.

References

1. A Schnitzler, J Gross, Normal and pathological oscillatory communication in the brain. Nat. Rev. Neurosci. **6**(4), 285–296 (2005)
2. F Varela, JP Lachaux, E Rodriguez, J Martinerie, The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. **2**(4), 229–239 (2001)
3. G Buzsaki, A Draguhn, Neuronal oscillations in cortical networks. Science **304**(5679), 1926–1929 (2004)
4. A Bruckstein, D Donoho, M Elad, From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. **51**, 34–81 (2009)
5. A Grinsted, J Moore, S Jevrejeva, Application of the cross wavelet transform and wavelet coherence to geophysical times series. Nonlinear Process. Geophys. **11**(5–6), 561–566 (2004)
6. K Spencer, P Nestor, M Niznikiewicz, D Salisbury, M Shenton, R McCarley, Abnormal neural synchrony in schizophrenia. J. Neurosci. **23**(19), 7407–7411 (2003)
7. C Torrence, G Compo, A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. **79**, 61–78 (1998)
8. JP Lachaux, A Lutz, D Rudrauf, D Cosmelli, M Le Van Quyen, J Martinerie, F Varela, Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. Neurophysiologie Clinique. **32**(3), 157–174 (2002)
9. M Plett, Transient detection with cross wavelet transforms and wavelet coherence. IEEE Trans. Signal Process. **55**(5I), 1605–1611 (2007)
10. Y Xu, S Haykin, R Racine, Multiple window time-frequency distribution and coherence of EEG using Slepian sequences and Hermite functions. IEEE Trans. Biomed. Eng. **46**(7), 861–66 (1999)
11. L Cohen, Time-frequency distributions—a review. Proc. IEEE. **77**(7), 941–981 (1989)
12. A Tzallas, M Tsipouras, D Fotiadis, Epileptic seizure detection in EEGs using time-frequency analysis. IEEE Trans. Inf. Technol. Biomed. **13**(5), 703–710 (2009)
13. HI Choi, WJ Williams, Improved time-frequency representation of multicomponent signals using exponential kernels. IEEE Trans. Acoust. Speech Signal Process. **37**(6), 862–871 (1989)
14. P Liu, Wavelet spectrum analysis and ocean wind waves. Wavelets Geophys, 151–166 (1994)
15. G Carter, Coherence and time delay estimation. Proc. IEEE. **75**(2), 236–255 (1987)
16. R Bortel, P Sovka, Approximation of statistical distribution of magnitude squared coherence estimated with segment overlapping. Signal Process. **87**(5), 1100–1117 (2007)
17. C Torrence, P Webster, Interdecadal changes in the ENSO-monsoon system. J. Climate. **12**(8 PART 2), 2679–2690 (1999)
18. E Cohen, A Walden, A statistical study of temporally smoothed wavelet coherence. IEEE Trans. Signal Process. **58**(6), 2964–2973 (2010)
19. DJ Thomson, Spectrum estimation and harmonic analysis. Proc. IEEE. **70**(9), 1055–1096 (1982)
20. I Daubechies, Time-frequency localization operators: a geometric phase space approach. IEEE Trans. Inf. Theory. **34**(4), 605–612 (1988)
21. S Olhede, A Walden, Generalized Morse wavelets. IEEE Trans. Signal Process. **50**(11), 2661–2670 (2002)
22. JS Brittain, D Halliday, B Conway, J Nielsen, Single-trial multiwavelet coherence in application to neurophysiological time series. IEEE Trans. Biomed. Eng. **54**(5), 854–862 (2007)
23. DR Brillinger, *Time Series Data Analysis and Theory Holt*. (Rinehart&amp Winston, New York, 1975)
24. PD Welch, The use of fast-Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. **AU-15**, 70–73 (1967)
25. X Jiang, S Mahadevan, Wavelet spectrum analysis approach to model validation of dynamic systems. Mech. Syst. Signal Process. **25**(2), 575–590 (2011)
26. M Breakspear, CJ Stam, Dynamics of a neural system with a multiscale architecture. Philos. Trans. R. Soc. B. **36**, 1051–1074 (2005)
27. TW Boonstra, A Daffertshofer, M Breakspear, PJ Beek, Multivariate time-frequency analysis of electromagnetic brain activity during bimanual motor learning. NeuroImage. **36**(2), 370–377 (2007)
28. A Bruns, Hilbert Fourier-, and wavelet-based signal analysis: are they really different approaches? J. Neurosci. Methods. **137**, 321–332 (2004)
29. KQ Lepage, MA Kramer, UT Eden, Some sampling properties of common phase estimators. Neural Comput. **25**, 1–21 (2013)
30. S Aydin, Comparison of power spectrum predictors in computing coherence functions for intracortical EEG signals. Ann. Biomed. Eng. **37**, 192–200 (2009)

31. E Cohen, A Walden, A statistical analysis of Morse wavelet coherence. IEEE Trans. Signal Process. **58**(3 PART 1), 980–989 (2010)
32. G Carter, CH Knapp, AH Nuttall, Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing. IEEE Trans. Audio Electroacoust. **AU-21**(4), 337–344 (1973)
33. H Kants, T Schreiber, vol. 1. (Cambridge University Press, Cambridge, MA, 2003), pp. 109–112
34. F Freyer, K Aquino, PA Robinson, P Ritter, M Breakspear, Bistability and non-Gaussian fluctuations in spontaneous cortical activity. J. Neurosci. **29**(26), 8512–8524 (2009)
35. A Daffertshofer, BCM van Wijk, On the influence of amplitude on the connectivity between phases. Front. Neuroinf. **5**(6), 1–12 (2011)
36. BCM van Wijk, PJ Beek, A Daffertshofer, Neural synchrony within the motor system: what have we learned so far? Front. Human Neurosci. **6**, 252–263 (2012)
37. AM Amjad, DM Halliday, JR Rosenberg, BA Conway, An extended difference of coherence test for comparing and combining several independent coherence estimates: theory and application to the study of motor units and physiological tremor. J. Neurosci. Methods. **73**, 69–79 (1997)
38. ES Lee, *Analyzing Complex Survey Data No. 071 in 07*, 2nd edn. (SAGE Publications, London, 2006)
39. L Kish, *Survey Sampling*. (Wiley, New York, 1965)
40. G Kalton, *Introduction to Survey Sampling: Quantitative Applications in the Social Sciences, No. 035 in 07*. (SAGE Publications, London, 1983)
41. AC Davison, *Statistical Models*. (Cambridge University Press, Cambridge, Ma, 2008)
42. G Jenkins, D Watts, *Spectral Analysis and Its Applications*. (Holden Day, San Francisco, CA, 1968)
43. AI Günter, in *Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis, IEEE ISPA*. Bias and variance of averaged and smoothed periodogram-based log-amplitude spectra (Pula-Croatia), (University Computing Center Zagreb-Croatia, 2001), pp. 452–457