Canonical information flow decomposition among neural structure subsets

Daniel Y. Takahashi[1], Luiz A. Baccalá[2]* and Koichi Sameshima[3]

[1] Psychology Department, Neuroscience Institute, Princeton University, Princeton, NJ, USA
[2] Telecommunications and Control Department, Escola Politécnica, University of São Paulo, São Paulo, Brazil
[3] Department of Radiology and Oncology, Faculdade de Medicina, University of São Paulo, São Paulo, Brazil

Partial directed coherence (PDC) and directed coherence (DC) which describe complementary aspects of the directed information flow between pairs of univariate components that belong to a vector of simultaneously observed time series have recently been generalized as bPDC/bDC, respectively, to portray the relationship between subsets of component vectors (Takahashi, 2009; Faes and Nollo, 2013). This generalization is specially important for neuroscience applications as one often wishes to address the link between the set of time series from an observed ROI (region of interest) with respect to series from some other physiologically relevant ROI. bPDC/bDC are limited, however, in that several time series within a given subset may be irrelevant or may even interact opposingly with respect to one another leading to interpretation difficulties. To address this, we propose an alternative measure, termed cPDC/cDC, employing canonical decomposition to reveal the main frequency domain modes of interaction between the vector subsets. We also show bPDC/bDC and cPDC/cDC are related and possess mutual information rate interpretations. Numerical examples and a real data set illustrate the concepts. The present contribution provides what is seemingly the first canonical decomposition of information flow in the frequency domain.

Keywords: directed connectivity measures, canonical decomposition, frequency domain, information flow, generalized coherence

1. INTRODUCTION

Human behavior is primarily thought as a property that emerges from the interaction of several brain areas, body parts, and the environment. Understanding how these elements dynamically interact is one of major themes of systems neuroscience. Several multivariate time series methods—old and new—have been introduced to describe the interdependence between brain areas using signal modalities like EEG, BOLD signals, MEG and LFP—and are collectively called connectivity measures. *Partial directed coherence* (PDC) (Baccalá and Sameshima, 2001) and *directed coherence/directed transfer function* (DC/DTF) (Kamiński and Blinowska, 1991) are two examples of such connectivity measures. Both describe complementary aspects (see Baccalá and Sameshima, 2014 for an in depth discussion) of how information flows between pairs of univariate time series components that belong to a multivariate vector of simultaneously observed time series (Takahashi et al., 2010). Recently, PDC and DC have been generalized (as bPDC/bDC, respectively) to describe how subsets (blocks) of components within a time series vector interrelate (Takahashi, 2009; Faes and Nollo, 2013). This is specially important for neuroscience applications as one often wants to investigate the interaction between sets of time series that are circumscribed to an observed region of interest (ROI) with respect to another physiologically relevant ROI (Nedungadi et al., 2011). The potential relevance of this type of question alone justifies looking for their deeper meaning in terms of information theoretical quantities.

Despite their practical importance, bPDC/bDC suffer from the limitation that several time series within a given subset may be irrelevant or interact in opposition to one another thereby posing interpretation difficulties. Also, in several situations, a researcher may be interested in just the few "best" descriptions of interaction between two sets of time series but not in the total amount of information flowing between them. For a more concrete example, assume that two brain areas interact and that bPDC is large. In this situation, it does not straightforwardly follow that all brain region components are interacting in the same way, or even whether some such components may be ignored. One way to address this limitation is to decompose bPDC/bDC into different components weighed according to relevance.

The aim of this article is twofold: (a) to provide a proper information theoretic interpretation for bPDC/bDC and (b) to introduce a canonical decomposition of information flows, henceforth termed, respectively, canonical PDC/DC (cPDC/cDC). These new decompositions allow us to closely mimic classical canonical correlation analysis so that different dynamically relevant interaction modes between brain areas can be exposed. Due to PDC interpretability in terms of Granger causality (Baccalá and Sameshima, 2014), a consequence of the present formulation is that cPDC represents a long sought frequency domain counterpart to time domain canonical decompositions of Granger causality (Sato et al., 2010; Ashrafulla et al., 2013).

The article is organized as follows. We first introduce the background and notation necessary for the rest of the article

(section 2). In the results section (section 3), we first show that both bPDC and bDC between two subsets of processes are block coherences between suitably defined underlying processes. Then, we demonstrate that such coherences are nothing but monotonic transformations of the mutual information rate between the respective processes (Gelfand and Yaglom, 1959; Takahashi et al., 2010; Nedungadi et al., 2011) leading immediately to their interpretability as mutual information rates. Next, we introduce cPDC and cDC and prove that they are the non-zero eigenvalues of the matrices whose determinants underlie the respective bPDC and bDC definitions (section 4). Using simulated examples and publicly available data we illustrate the usefulness of cPDC/cDC (section 5) followed by a brief discussion (section 6). Proof details are left to the Appendix.

2. BACKGROUND

Let $X_1, \ldots, X_K$ be K distinct multivariate time series vectors with dimension $M_1, \ldots, M_K$. Using T to indicate matrix transposition, let $X(t) = \left[X_1(t)^T, \ldots, X_K(t)^T\right]^T$ for each time $t \in \mathbb{Z}$ be a second order stationary time series with spectral density matrix $S(\omega)$ at each frequency $\omega \in [-\pi, \pi)$. To justify our formal computation, we assume that $S(\omega)$ is uniformly bounded from below and above and invertible at all frequencies (Hannan, 1970). This is called the *boundedness* condition which guarantees that the following autoregressive (AR) representation of X holds in the mean square sense

$$X(t) = \sum_{l=1}^{+\infty} A(l)X(t-l) + \epsilon(t), \quad (1)$$

where $\epsilon(t) = \left[\epsilon_1(t)^T \ldots \epsilon_K(t)^T\right]^T$ stands for a zero mean innovation process, i.e., $\mathbb{E}\left[\epsilon(t)\epsilon(t)^T\right] = \Sigma$ and $\mathbb{E}\left[\epsilon(t)\epsilon(l)^T\right] = 0$ for $l \neq t$. For $l \geq 1$, $A(l)$ are $(M_1 + \ldots + M_K)^2$-dimensional matrices. Let $A_{pq}(l)$ for $p, q \in \{1, \ldots, K\}$ and $l \geq 1$ be $M_p \times M_q$-dimensional matrices so that $A(l)$ has the following structure

$$A(l) = \begin{bmatrix} A_{11}(l) & \ldots & A_{1M}(l) \\ \vdots & \ddots & \vdots \\ A_{M1}(l) & \ldots & A_{MM}(l) \end{bmatrix}$$

We define $\bar{A}(\omega) = I - \sum_{l \geq 1} A(l)e^{-\sqrt{-1}\omega l}$.

Under the boundedness condition, the following moving average (MA) mean square sense representation for the process X also holds

$$X(t) = \sum_{l=0}^{+\infty} H(l)\epsilon(t-l), \quad (2)$$

where $H(l)$ for $l \geq 0$ are $(M_1 + \ldots + M_K)^2$-dimensional matrices. Let $\bar{H}(\omega) = \sum_{l \geq 0} H(l)e^{-\sqrt{-1}\omega l}$. We have that $\bar{A}^*(\omega) = \bar{H}^{-1}(\omega)$ for all $\omega \in [-\pi, \pi)$. The superscript $*$ indicates the matrix complex conjugate.

Let $P(\omega) = S^{-1}(\omega)$. bPDC from the multivariate process X_j to the process X_i at frequency ω, denoted $\pi_{ij}^{(b)}(\omega)$, is defined

(Takahashi, 2009; Faes and Nollo, 2013) by

$$\pi_{ij}^{(b)}(\omega) = 1 - \det\left(P_{jj}(\omega) - \bar{A}_{ij}^*(\omega)\Sigma_{ii}^{-1}\bar{A}_{ij}(\omega)\right) \det\left(P_{jj}(\omega)\right)^{-1}, \quad (3)$$

where det indicates the determinant and the subscript indices relate to the natural block structure associated with the matrices.

Let $\Theta = \Sigma^{-1}$. bDC from the multivariate process X_j to the process X_i at frequency ω, denoted $\gamma_{ij}^{(b)}(\omega)$, is defined (Takahashi, 2009; Faes and Nollo, 2013) by

$$\gamma_{ij}^{(b)}(\omega) = 1 - \det\left(S_{ii}(\omega) - \bar{H}_{ij}(\omega)\Theta_{jj}^{-1}\bar{H}_{ij}^*(\omega)\right) \det\left(S_{ii}(\omega)\right)^{-1}. \quad (4)$$

Note that the present bDC definition differs slightly from the one in Faes and Nollo (2013). We removed the unnecessary condition of strict causality, i.e., diagonality of Σ, simply by substituting Σ_{jj}^{-1} by Θ_{jj}^{-1} in their definition of bDC as it is more suited for formulating information theoretic results as shown ahead.

Consider a second-order stationary multivariate process $W(t) = \left[Y(t)^T Z(t)^T\right]^T$. The block coherence between Y and Z at frequency ω is defined as (Nedungadi et al., 2011)

$$C_{YZ}^{(b)}(\omega) = 1 - \det\left(S_{WW}(\omega)\right) \det\left(S_{YY}(\omega)\right)^{-1} \det\left(S_{ZZ}(\omega)\right)^{-1}. \quad (5)$$

Observe that we used the process name in the subscript of the power spectrum S to indicate the corresponding spectral density matrices. In the rest of the article, we will use interchangeably the process name or the corresponding indices in the subscript whenever there is no ambiguity.

Another important definition is that of mutual information rate (MIR) between two multivariate strictly stationary processes Y and Z is

$$\text{MIR}_{YZ} = \lim_{t \to +\infty} \frac{1}{t} \mathbb{E}\left[\log \frac{d\mathbb{P}(Y(1), \ldots, Y(t), Z(1), \ldots, Z(t))}{d\mathbb{P}(Y(1), \ldots, Y(t))d\mathbb{P}(Z(1), \ldots, Z(t))}\right]. \quad (6)$$

The classical relationship between block coherence (Equation 5) and mutual information rate (Equation 6) follows from

Theorem. *(Gelfand and Yaglom, 1959; Pinsker, 1964) If Y and Z are jointly stationary Gaussian processes satisfying the boundedness condition, we have that the MIR between Y and Z is given by*

$$MIR_{YZ} = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log\left(1 - C_{YZ}^{(b)}(\omega)\right) d\omega. \quad (7)$$

Now, following Takahashi et al. (2010), we define, for $i \in \{1, \ldots, K\}$, the partialized process η_i by

$$\eta_i(t) = X_i(t) - \mathbb{E}\left[X_i(t)|\left\{X_j(l), j \neq i, l \in \mathbb{Z}\right\}\right], \quad (8)$$

where $\mathbb{E}[\odot|\odot]$ henceforth denotes the best linear conditional predictor. Likewise the partialized innovation process ζ_i for $i \in \{1, \ldots, K\}$ is

$$\zeta_i(t) = \epsilon_i(t) - \mathbb{E}\left[\epsilon_i(t)|\left\{\epsilon_j(t), j \neq i\right\}\right]. \quad (9)$$

Observe that both partialized process and partialized innovation process were defined in Takahashi et al. (2010) but for the special case of scalar η_i and ζ_i.

3. RELATION BETWEEN bPDC/bDC AND MUTUAL INFORMATION RATE

Our first result establishes the relationship between bPDC and block coherence and is analogous to Theorem 1 in Takahashi et al. (2010).

Theorem 1. *Let X satisfy the boundedness condition. For all $i, j \in \{1, \ldots, K\}$ and all frequencies $\omega \in [-\pi, \pi)$ we have that*

$$\pi_{ij}^{(b)}(\omega) = C_{\epsilon_i \eta_j}^{(b)}(\omega). \tag{10}$$

A straightforward corollary is

Corollary 1. *Let X be a stationary Gaussian process and satisfy the boundedness condition. For all $i, j \in \{1, \ldots, K\}$ we have that*

$$MIR_{\epsilon_i \eta_j} = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log\left(1 - \pi_{ij}^{(b)}(\omega)\right) d\omega. \tag{11}$$

Similar results also hold for bDC.

Theorem 2. *Let X satisfy the boundedness condition. For all $i, j \in \{1, \ldots, K\}$ and all frequencies $\omega \in [-\pi, \pi)$ we have that*

$$\gamma_{ij}^{(b)}(\omega) = C_{X_i \zeta_j}^{(b)}(\omega). \tag{12}$$

and

Corollary 2. *Let X be a stationary Gaussian process and satisfy the boundedness condition. For all $i, j \in \{1, \ldots, K\}$, we have that*

$$MIR_{X_i \zeta_j} = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log\left(1 - \gamma_{ij}^{(b)}(\omega)\right) d\omega. \tag{13}$$

4. CANONICAL PDC AND DC

Canonical correlation is a classical method developed initially by Hotelling (1936) to address the relationship between random vectors. Brillinger (1981) generalized the method for time series and gave an excellent account of the relationship between canonical correlation analysis and different ideas in multivariate statistics. Our formulation of canonical coherence is equivalent to the definition introduced by Brillinger (1981).

Let Y and Z be respectively M_Y- and M_Z-dimensional jointly second order stationary processes. To better understand the relationship between Y and Z, we can ask the following question: Which components of Y and Z are most representative of the interaction between the processes? One way to formalize this is to consider filtering matrices $B_Y(l)$ ($1 \times M_Y$) and $B_Z(l)$ ($1 \times M_Z$), for all $l \in \mathbb{Z}$ and define the scalar processes b_Y and b_Z by

$$b_Y(t) = \sum_{l \in \mathbb{Z}} B_Y(l) Y(t - l) \tag{14}$$

and

$$b_Z(t) = \sum_{l \in \mathbb{Z}} B_Z(l) Z(t - l), \tag{15}$$

so that $C_{b_Y b_Z}(\omega)$ is maximized for all $\omega \in [-\pi, \pi)$. If furthermore Y and Z are jointly stationary Gaussian processes, then this is equivalent to maximizing $MIR_{b_Y b_Z}$.

Following the above idea, we define the first canonical coherence between Y and Z at frequency ω by

$$C_{YZ}^{(c_1)}(\omega) = \sup_{B_Y, B_Z} C_{b_Y b_Z}(\omega). \tag{16}$$

Assume that the supremum (Equation 16) is achieved for $\bar{b}_Y$ and $\bar{b}_Z$, which we call *first canonical time series*. Consider the residual processes $Y^1(t) = Y(t) - \mathbb{E}\left[Y(t) | \{\bar{b}_Y(l), l \in \mathbb{Z}\}\right]$ and $Z^1(t) = Z(t) - \mathbb{E}\left[Z(t) | \{\bar{b}_Z(l), l \in \mathbb{Z}\}\right]$. Observe that Y^1 and Z^1 are uncorrelated to the processes $\bar{b}_Y$ and $\bar{b}_Z$, respectively. The second canonical coherence $C_{YZ}^{(c_2)}(\omega)$ is defined recursively on the residues by $C_{YZ}^{(c_2)}(\omega) = C_{Y^1 Z^1}^{(c_1)}(\omega)$.

Analogously, for $2 \leq m \leq \min\{M_Y, M_Z\}$, considering the residual processes

$$Y^m(t) = Y^{m-1}(t) - \mathbb{E}\left[Y^{m-1}(t) | \{\bar{b}_{Y^k}(l), l \in \mathbb{Z}, \right.$$
$$\left. k \in \{1, \ldots, m-1\}\}\right]$$

and

$$Z^m(t) = Z^{m-1}(t) - \mathbb{E}\left[Z^{m-1}(t) | \{\bar{b}_{Z^k}(l), l \in \mathbb{Z}, \right.$$
$$\left. k \in \{1, \ldots, m-1\}\}\right],$$

one may define the m-th canonical coherence as

$$C_{YZ}^{(c_m)}(\omega) = C_{Y^{m-1} Z^{m-1}}^{(c_1)}(\omega). \tag{17}$$

In this way, it is possible to construct a hierarchy of coherences where each element captures the dependence structure that is not explained by the other elements.

Finally, we introduce cPDC and cDC. For $m \leq \min\{M_i, M_j\}$, the *m-th canonical* PDC from j to i at frequency ω denoted $\pi_{ij}^{(c_m)}(\omega)$ is defined by

$$\pi_{ij}^{(c_m)}(\omega) = C_{\epsilon_i \eta_j}^{(c_m)}(\omega). \tag{18}$$

Similarly, the *m-th canonical* DC from j to i at frequency ω denoted $\gamma_{ij}^{(c_m)}(\omega)$ is defined by

$$\gamma_{ij}^{(c_m)}(\omega) = C_{X_i \zeta_j}^{(c_m)}(\omega). \tag{19}$$

At first sight, it is unclear whether the canonical PDC and DC exist at all or even if they are uniquely defined. More importantly, nor is it obvious that it is possible to compute them. Despite these initial uncertainties, we show next that canonical coherences are consistently defined as the non-null eigenvalues of some specific matrices.

Let $\lambda^m(Q)$ denote its m-th eigenvalue from matrix Q ordered from its largest to its smallest value. The following theorem furnishes a practical way to calculate cPDC and cDC.

Theorem 3. *Under the boundedness condition for X the following identities hold:*

$$\pi_{ij}^{(c_m)}(\omega) = \lambda^m \left(\bar{A}_{ij}^*(\omega) \Sigma_{ii}^{-1} \bar{A}_{ij}(\omega) P_{jj}^{-1}(\omega) \right) \quad (20)$$

and

$$\gamma_{ij}^{(c_m)}(\omega) = \lambda^m \left(S_{ii}^{-1}(\omega) \bar{H}_{ij}(\omega) \Theta_{jj}^{-1} \bar{H}_{ij}^*(\omega) \right). \quad (21)$$

Furthermore it is possible to relate bPDC/bDC and cPDC/cDC via

Theorem 4. *Under the same conditions of Theorem 3 the following identities hold:*

$$\pi_{ij}^{(b)}(\omega) = 1 - \prod_{m=1}^{\min\{M_i, M_j\}} \left(1 - \pi_{ij}^{(c_m)}(\omega) \right) \quad (22)$$

and

$$\gamma_{ij}^{(b)}(\omega) = 1 - \prod_{m=1}^{\min\{M_i, M_j\}} \left(1 - \gamma_{ij}^{(c_m)}(\omega) \right). \quad (23)$$

A simple consequence of Equations (22), (23) is that for stationary Gaussian processes satisfying the boundedness condition, we now have a decomposition of the mutual information rates

$$MIR_{\epsilon_i \eta_j} = \sum_{m=1}^{\min\{M_i, M_j\}} -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log \left(1 - \pi_{ij}^{(c_m)}(\omega) \right) d\omega \quad (24)$$

and

$$MIR_{X_i \zeta_j} = \sum_{m=1}^{\min\{M_i, M_j\}} -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log \left(1 - \gamma_{ij}^{(c_m)}(\omega) \right) d\omega. \quad (25)$$

Note how the quantities being summed in Equations (24), (25) are formally themselves contributions to the mutual information written in terms of their canonical coherence contributions.

5. ILLUSTRATIONS

5.1. SIMULATED MODELS

Example 1. To provide insight into cPDC behavior, we begin with a very simple example that can be fully and explicitly solved.

Let a vector of observed time series $[Y_1, Y_2, Y_3, Y_4]$ be a real valued autoregressive process of order $p = 1$ and $\Sigma = I$. The autoregressive coefficients of the model are described by

$$A(1) = \begin{pmatrix} .5 & f & 0 & 0 \\ e & .5 & 0 & 0 \\ a & b & .5 & h \\ c & d & g & .5 \end{pmatrix}, \quad (26)$$

as in **Figure 1**.
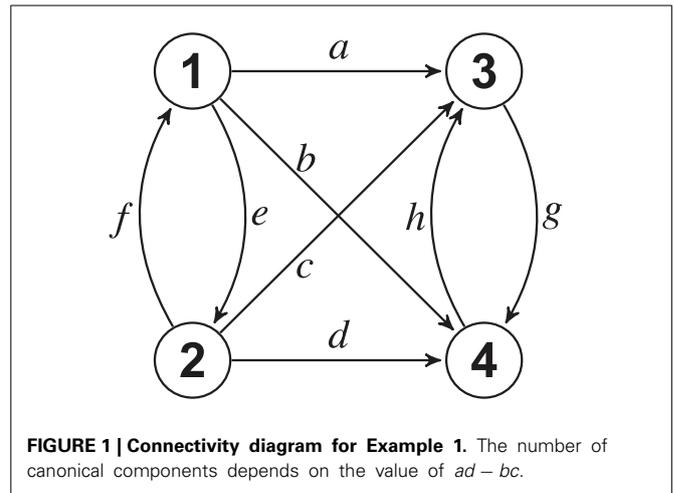
FIGURE 1 | Connectivity diagram for Example 1. The number of canonical components depends on the value of $ad - bc$.

By adopting time series blocks as $X_1 = [Y_1 \ Y_2]$ and $X_2 = [Y_3 \ Y_4]$, when $e = f = g = h = 0$, direct computation shows that the canonical PDC from block X_2 to X_1 is zero, i.e., $\pi_{12}^{(c_1)}(\omega) = \pi_{12}^{(c_2)}(\omega) = 0$ for all ω (reflecting the nullity of the 2×2 $A(l)$ right side upper block), whereas the coupling in the opposite direction contributes two distinct components:

$$\pi_{21}^{(c_1)}(\omega) = \frac{a^2 + b^2 + c^2 + d^2 + \sqrt{(a^2 + b^2 + c^2 + d^2)^2 - 4(ad - bc)^2}}{2.5 - 2\cos(\omega)} \quad (27)$$
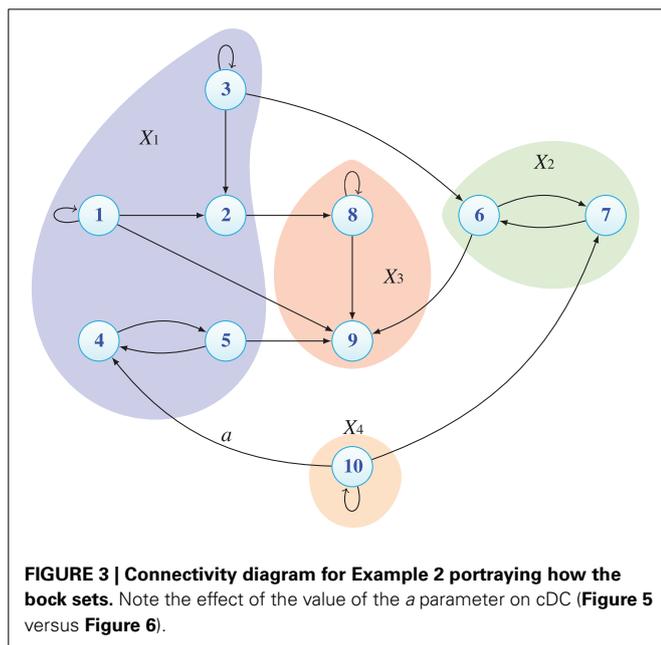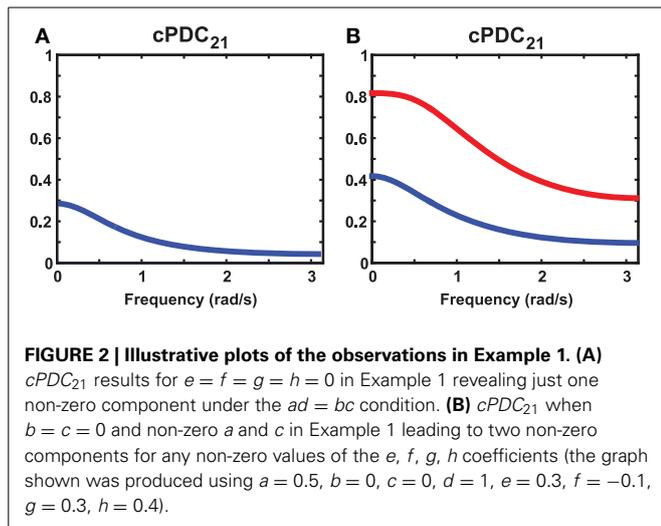
and

$$\pi_{21}^{(c_2)}(\omega) = \frac{a^2 + b^2 + c^2 + d^2 - \sqrt{(a^2 + b^2 + c^2 + d^2)^2 - 4(ad - bc)^2}}{2.5 - 2\cos(\omega)}. \quad (28)$$

For $ad = bc$—i.e., if the lower left 2×2 block determinant of $A(l)$ is zero as well, the total number of non-zero cPDC components reduces to just 1.

Even if e, f, g, h are non-zero, i.e., regardless of intrablock dynamics, $a = b = 0$ suffices to produce the single non-zero $\pi_{21}^{(c_1)}(\omega)$ component (shown in **Figure 2A**) since block X_1 interacts with block X_2 exclusively through Y_4, i.e., $\pi_{21}^{(c_2)}(\omega) \equiv 0$. In this case, since only Y_4 is directly impacted by the interaction, only one combined source of variance exists even though two links exist between the blocks. Likewise if $b = d = 0$, even though two links leave X_1, there is only one dynamical component that counts.

This contrasts with the situation when $b = c = 0$ where two non-zero $\pi_{21}^{(c_2)}(\omega)$ coexist (**Figure 2B**) regardless of the values of e, f, g, h which, nonetheless, contribute to the relative size of the components.

Example 2. In the next example, a 10-variate time series $(Y_1, \ldots, Y_{10})$ follows the connectivity diagram represented in **Figure 3**. The multivariate time series is divided into four blocks $(X_1, X_2, X_3, \text{and } X_4)$, where X_4 only sends information and X_3, which is an integrative block, only receives information. Block X_1 has two functionally distinct internal parts, and only one is reached by outside influence. The scenario is fairly complicated and we next illustrate cPDC/cDC usefulness for understanding the underlying dynamic interaction between blocks.

FIGURE 2 | Illustrative plots of the observations in Example 1. (A) $cPDC_{21}$ results for $e = f = g = h = 0$ in Example 1 revealing just one non-zero component under the $ad = bc$ condition. **(B)** $cPDC_{21}$ when $b = c = 0$ and non-zero a and c in Example 1 leading to two non-zero components for any non-zero values of the e, f, g, h coefficients (the graph shown was produced using $a = 0.5$, $b = 0$, $c = 0$, $d = 1$, $e = 0.3$, $f = -0.1$, $g = 0.3$, $h = 0.4$).

FIGURE 3 | Connectivity diagram for Example 2 portraying how the bock sets. Note the effect of the value of the a parameter on cDC (**Figure 5** versus **Figure 6**).

To help interpret the results, we begin by describing the non-zero model coefficients and their dynamical effects. Observe that the model subscript indices in this example indicate the corresponding scalar process and not the block number.

1. Block $X_1 = [Y_1 \ Y_2 \ Y_3 \ Y_4 \ Y_5]$

$$A_{1,1}(1) = 1.98 \ cos(\pi/50), A_{1,1}(2) = -(.99)^2,$$

(low frequency oscillator in Y_1) (29)

$$A_{2,3}(1) = 1,$$ (30)

$$A_{3,3}(1) = 1.98cos(\pi/2), A_{3,3}(2) = -(.99)^2,$$

(oscillator at midband ($\pi/2$) in Y_3) (31)

$$A_{5,4}(1) = .99, A_{4,5}(1) = -.99,$$

(oscillator at midband in $[Y_4 \ Y_5]$) (32)

$$A_{8,2}(1) = 1, A_{8,2}(3) = 1, \quad \text{(midband notch)}$$ (33)

$$A_{6,3}(1) = 1, A_{6,3}(3) = 1, \quad \text{(midband notch)}$$ (34)

$$A_{9,1}(1) = 1, A_{9,5}(1) = 1.$$ (35)

2. Block $X_2 = [Y_6 \ Y_7]$

$$A_{6,7}(1) = .99, A_{7,6}(1) = -.99,$$

(oscillator identical to the $[Y_4 \ Y_5]$) (36)

$$A_{9,6}(1) = 1.$$ (37)

3. Block $X_3 = [Y_8 \ Y_9]$

$$A_{8,8}(1) = -1, \ A_{9,8}(1) = .5.$$ (38)

4. Block $X_4 = [Y_{10}]$

$$A_{10,10}(1) = 1.98 \ cos(2\pi/3), \ A_{10,10}(2) = -(.99)^2,$$

(high frequency oscillator in Y_{10}) (39)

$$A_{4,10}(1) = a,$$ (40)

$$A_{7,10}(1) = 1.$$ (41)

The resulting cPDC components can be appreciated in **Figure 4** for $|a| = 1$. Among their interesting features is the existence of the notch filtered link from X_1 to X_2 and to X_3 at midband. The effects of the low frequency dynamics due to Y_1 and the midband resonance due to $[Y_4$ and $Y_5]$ manifests itself as the strongest component from X_1 to X_3. Likewise the single link effect from X_2 to X_3 is readily apparent as the higher frequency resonances from X_4 toward both X_1 and X_2. Both X_3 components are identically equal to 1 since nothing leaves the block.

The corresponding cDCs are portrayed in **Figure 5** for $a = -1$ with no signal reachability from X_4 to X_3. This contrasts markedly with **Figure 6** for $a = 1$ where X_4's indirect effects on X_3 are not balanced out.

The effects of the notch connections are readily apparent in both cases. For example, the power associated with the notch frequencies are the local components to X_2 and X_3 and cannot be attributed to outside influence. For block X_1 only one of the five components is different from 1 reflecting the contribution coming from X_4.

5.2. EMPIRICAL DATA

This example is based on EEG data borrowed from Sameshima et al. (2014) (Ex. 7.7), which describes a left mesial temporal ictal episode monitored using an extended 10–20 system. The midline electrodes were excluded and left (L) and right (R) side electrodes were grouped as to whether they were frontal (F), central (C), parietal (P), temporal (T) or occipital (O) leading to the canonical PDCs portrayed in **Figure 7** where the most important connecting blocks share a dominant low pass frequency canonical component of fairly identical shape pointing to the existence
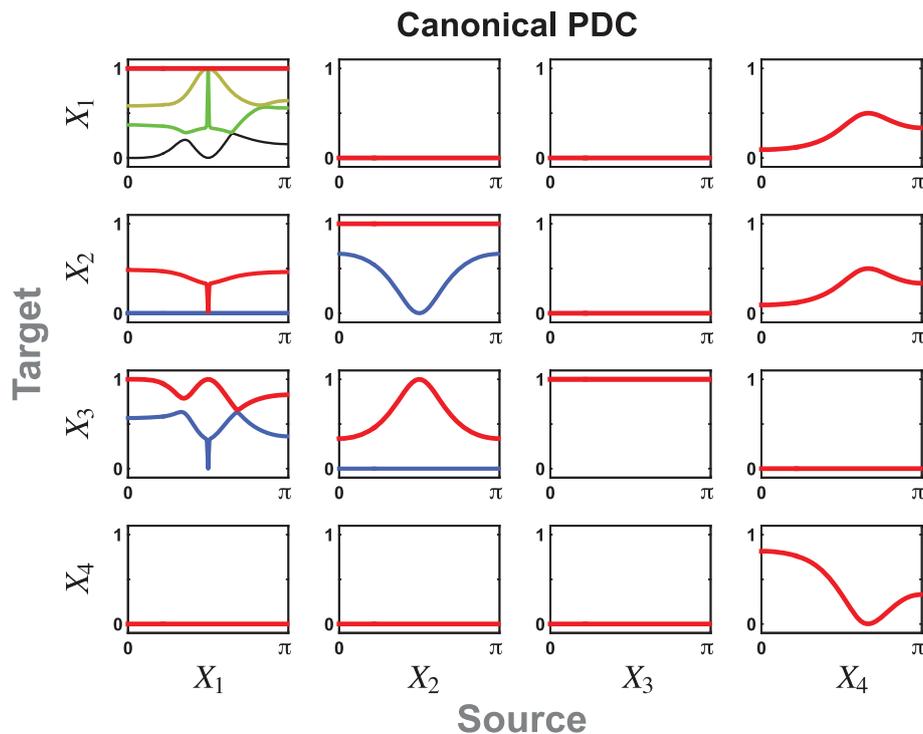
FIGURE 4 | The cPDC for Example 2 reflects the existence of the notch connecting filters from X_1 to X_2 and to X_3. The intrinsic dynamics of the oscillators from a subregion of X_1 into X_3 is apparent in the resonances of the largest cPDC component. The resonance within block X_2 manifest itself in the single non-zero component into X_3 while the effect of X_4 reaches symmetrically into X_1 and X_2 via its single dynamic component. In this and following two figures, each subfigure may contain up to five cPDC/cDC components, given by $min(M_i, M_j)$ as in Equations (22)/(23), represented in red, blue, yellow, green, or black lines in decreasing order of magnitude.

of a shared dominant connectivity dynamics behind the observation, see **Figure 8A**. Their connectivity is further summarized in **Figure 8B**.

6. DISCUSSION

We showed that bPDC/bDC introduced in Takahashi (2009) and Faes and Nollo (2013) are block coherences between properly chosen vector time series. When the time series are Gaussian, this implies that bPDC/bDC represent mutual information rates between well defined underlying vector time series. This fully generalizes the results presented in Takahashi et al. (2010). To enhance the understanding of the possibly complex interaction between multiple time series and overcome some bPDC/bDC limitations, we showed that the latter can be decomposed in canonical terms that we call cPDC/cDC. These decompositions represent the various different modes of interaction whereby sets of time series interact. We introduced an explicit way to compute these new quantities and proved some of their properties. The usefulness of cPDC/cDC was illustrated by three examples.

6.1. bPDC AND bDC AS BLOCK COHERENCES

Takahashi et al. (2010) showed that PDC from the j-th scalar time series to the i-th scalar time series is the coherence between the i-th innovation process and the j-th partialized process with a similar result for DC. It is natural to ask whether an analogous

result holds for bPDC and bDC. We showed that this is indeed the case where bPDC/bDC represent block coherences relating subsets of adequately defined innovations/partialization processes (Takahashi, 2009; Nedungadi et al., 2011). At first sight these identities may seem surprising as both bPDC and bDC are fully multivariate and directional measures of dependence, whereas block coherences are at once block-pairwise and symmetric measures of dependence. Yet careful reading of Theorems 1 and 2 highlights that bPDC/bDC from j to i and bPDC/bDC from i to j are, in general, block coherences between distinct pairs of vector processes which explains their asymmetric nature and lends them their directed connectivity character. Also, we note that for both bPDC and bDC, the coherences involve innovation process subsets which explains their fully multivariate characteristic as measures. Another interesting observation is that since the innovation processes are uncorrelated to the past of the partialized processes by construction, in the case of bPDC only innovations in the past of the partialized process contribute to the coherence which explains why bPDC is a directed measure of dependence. An analogous observation holds for bDC. In the Gaussian case, the bPDC/bDC representation as a block coherence allows relating them to the mutual information rate between suitably chosen time series. Formally this justifies the idea that these quantities are *de facto* measures of information flow. For an interesting comparison between bPDC/bDC and Geweke's measure of linear feedback see Faes
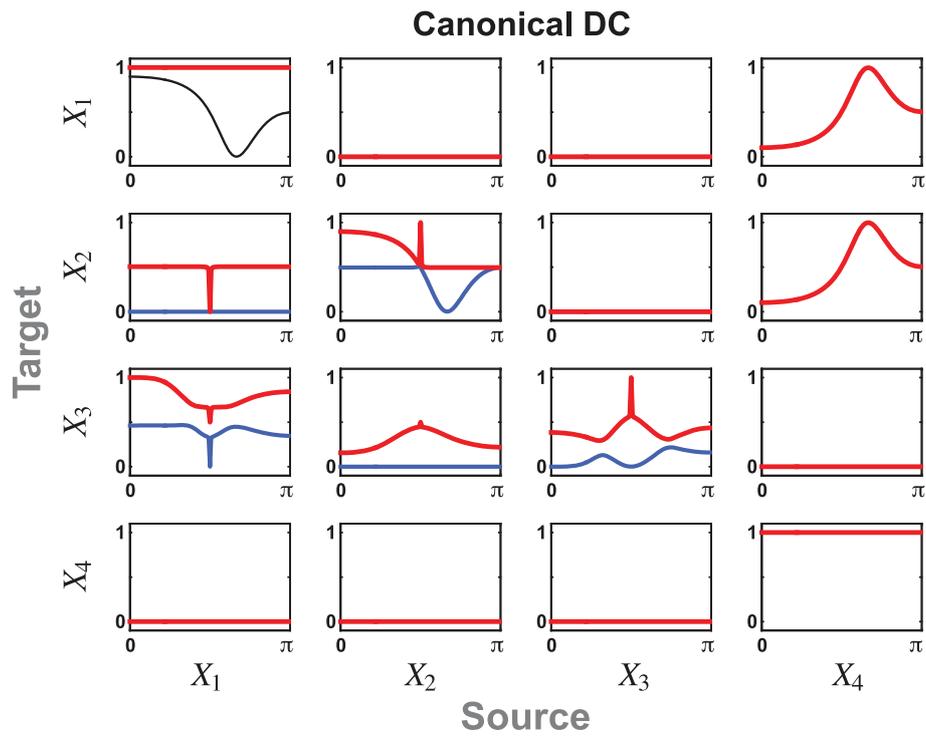
FIGURE 5 | cDC for Example 2 for $a = -1$ leading to a cancelation of the effect of X_4 on X_3 as the signal travels indirectly through two exactly identical structures but with opposite phases before reaching X_3. The notch filtering action is also apparent from the cDCs from X_1 to X_2 and X_3.
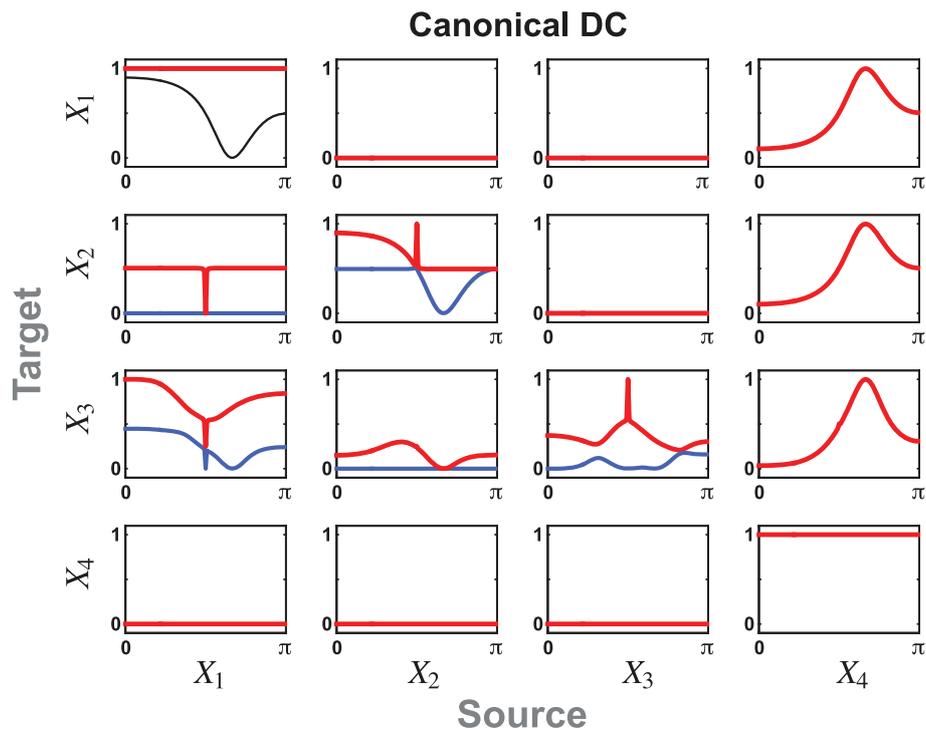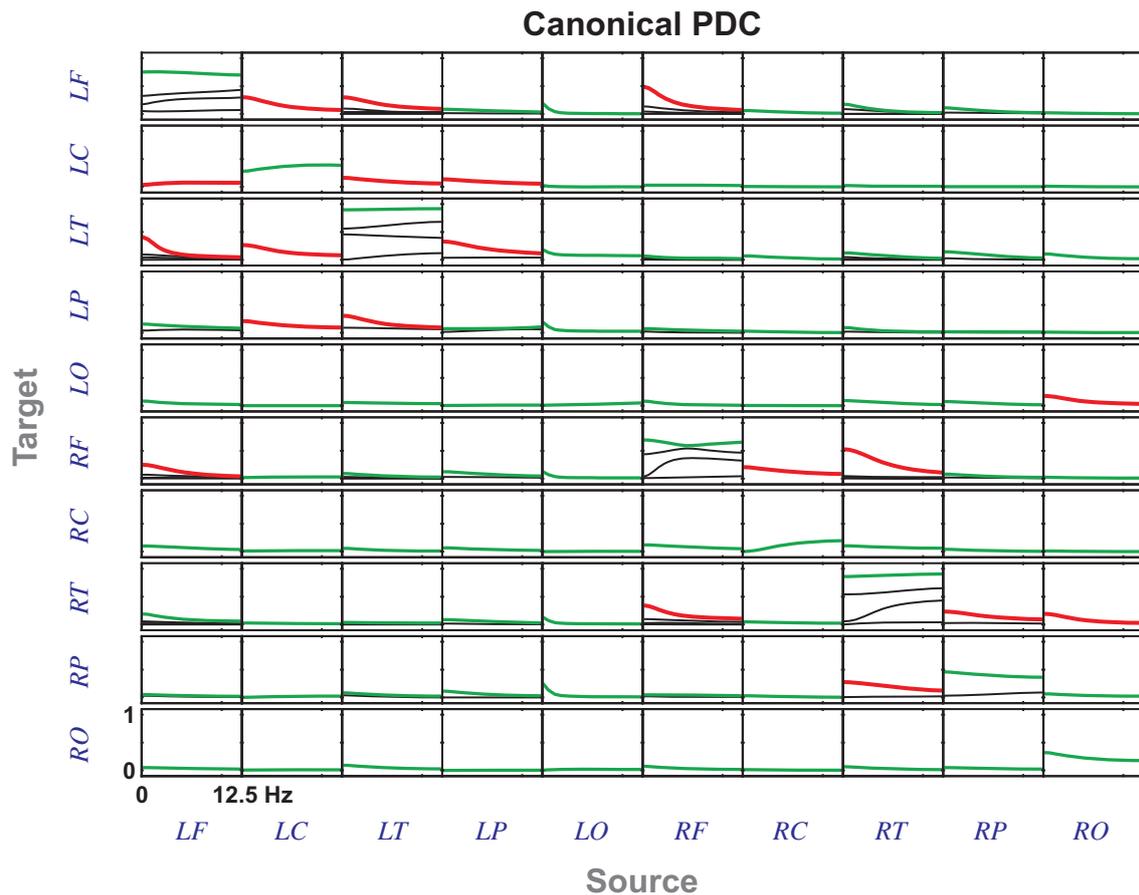

FIGURE 6 | cDC for Example 2 with $a = 1$ which differs from Figure 5 in the effect from X_4 to X_3 which no longer cancels out.

FIGURE 7 | cPDC from the Empirical Data example (section 5.2) from the left mesial ictal episode where the largest components are represented either in red or green. cPDC values in red were arbitrarily considered significant and were pictorially summarized in **Figure 8B**.

and Nollo (2013). As a small note for the reader, we observe that our definition of bPDC/bDC is slightly more general than the one proposed by Faes and Nollo (2013) because the covariance matrix of the innovations does not need to be diagonal as they assumed.
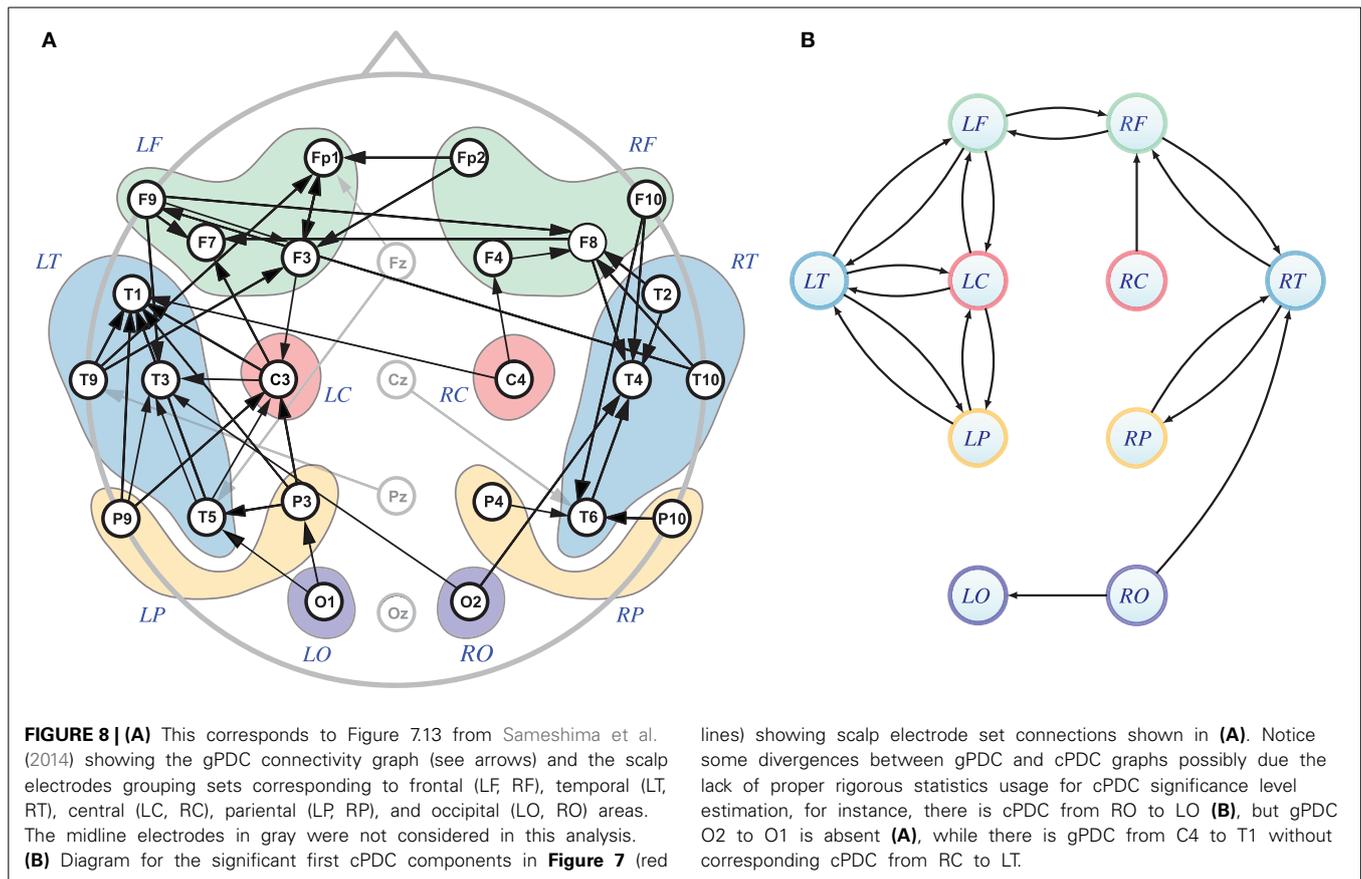
6.2. CANONICAL DECOMPOSITION OF DIRECTIONAL MEASURES

Given a pair of random vectors, it is natural to ask how to measure/represent dependence between them. In statistics, there are two main methods, both inspired by the basic Pearson correlation, to address this. The first one generalizes Pearson correlation directly using the determinants of the covariance matrix between and within each set of random variables. For time series, the equivalent measure in the frequency domain is the block coherence and the directed versions are bPDC and bDC. A second generalization rests on the idea of canonical correlation introduced by Hotelling (1936). There are several generalizations of canonical correlation for time series taylored specifically to infer Granger causality in the time domain (Sato et al., 2010; Wu et al., 2011), but, to the best of our knowledge, cPDC and cDC are the first proposals of canonical measures of directed dependence in the frequency domain.

One advantage of cPDC/cDC over bPDC/bDC is that canonical decomposition allows inferring the various different existing modes of interaction between sets of time series in close analogy to what is done for classical canonical correlation and principal component analyses. One should expect this to be useful when several signals are redundant, generated by similar mechanisms, or when there are several time series that do not significantly contribute to the interaction between sets of time series, e.g., when there are many brain areas that are not interacting with each other during some specific behavior. Besides, as we show in Theorem 4, we can recover the bPDC/bDC from the cPDC/cDC.

6.3. INTERPRETING cPDC/cDC

The main practical interest of cPDC/cDC is to allow the simplification of connectivity interpretations whilst giving new insights into the dynamical interaction between neural structures. We illustrated the achievable simplification using an EEG data set from an epileptic patient. We also showed how cPDC is related to the number of "modes" of interaction between sets of time series through the simple numerical Example 1 and via the slightly more complex Example 2. We expect that cPDC/cDC together

FIGURE 8 | (A) This corresponds to Figure 7.13 from Sameshima et al. (2014) showing the gPDC connectivity graph (see arrows) and the scalp electrodes grouping sets corresponding to frontal (LF, RF), temporal (LT, RT), central (LC, RC), pariental (LP, RP), and occipital (LO, RO) areas. The midline electrodes in gray were not considered in this analysis. **(B)** Diagram for the significant first cPDC components in **Figure 7** (red lines) showing scalp electrode set connections shown in **(A)**. Notice some divergences between gPDC and cPDC graphs possibly due the lack of proper rigorous statistics usage for cPDC significance level estimation, for instance, there is cPDC from RO to LO **(B)**, but gPDC O2 to O1 is absent **(A)**, while there is gPDC from C4 to T1 without corresponding cPDC from RC to LT.

with bPDC/bDC become useful tools for handling high dimensional data sets that are increasingly being recorded by several researchers.

We propose that a reasonable way to understand the usefulness of cPDC/cDC is to make an analogy with classical principal component and canonical correlation analyses. Therefore, similar heuristics could be applied in practical situations, for example, to decide the number of different components to include in the interpretation. The canonical time series $\bar{b}_Y$ and $\bar{b}_Z$ from section 4 (see also Brillinger, 1981) are analogous to the canonical variables from the classical canonical correlation analysis and can play a similar role for result interpretation.

Finally we remark that the computational procedures used for the present paper will be made available the PDC homepage at http://www.lcs.poli.usp.br/~baccala/pdc/canon together with the data used in section 5.2.

REFERENCES

Ashrafulla, S., Haldar, J. P., Joshi, A. A., and Leahy, R. M. (2013). Canonical Granger causality between regions of interest. *Neuroimage* 83, 189–199. doi: 10.1016/j.neuroimage.2013.06.056

Baccalá, L. A., and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* 84, 463–474. doi: 10.1007/PL00007990

Baccalá, L. A., and Sameshima, K. (2014). "Multivariate time series brain connectivity: a sum up," in *Methods in Brain Connectivity Inference Through Multivariate Time Series Analysis*, eds K. Sameshima and L. A. Baccalá (Boca Raton: CRC Press), 245–251. doi: 10.1201/b16550-18

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory. Classics in applied mathematics*. Vol. 36. San Francisco, CA: SIAM, Society for Industrial and Applied Mathematics; Holden-Day.

Faes, L., and Nollo, G. (2013). Measuring frequency domain Granger causality for multiple blocks of interacting time series. *Biol. Cybern.* 107, 217–232. doi: 10.1007/s00422-013-0547-5

Gelfand, I. M., and Yaglom, A. M. (1959). Calculation of amount of information about a random function contained in another such function. *Am. Math. Soc. Transl. Ser.* 2, 3–52.

Hannan, E. J. (1970). *Multiple Time Series (Wiley Series in Probability and Mathematical Statistics)*. New York, NY: Wiley. doi: 10.1002/9780470316429

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* 28, 321–377. doi: 10.2307/2333955

Kamiński, M., and Blinowska, K. J. (1991). A new method of the description of the information flow in brain structures. *Biol. Cybern.* 65, 203–210. doi: 10.1007/BF00198091

Lütkepohl, H. (1996). *Handbook of Matrices*. Chichester: John Wiley.

Nedungadi, A. G., Ding, M., and Rangarajan, G. (2011). Block coherence: a method for measuring the interdependence between two blocks of neurobiological time series. *Biol. Cybern.* 104, 197–207. doi: 10.1007/s00422-011-0429-7

Pinsker, M. S. (1964). *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden-Day.

Sameshima, K., Takahashi, D. Y., and Baccalá, L. A. (2014). "Asymptotic PDC properties," in *Methods in Brain Connectivity Inference through Multivariate Time Series Analysis*, eds K. Sameshima and L. A. Baccalá (Boca Raton: CRC Press), 113–131. doi: 10.1201/b16550-9

Sato, J. R., Fujita, A., Cardoso, E. F., Thomaz, C. E., Brammer, M. J., and Amaro, E. Jr. (2010). Analyzing the connectivity between regions of interest: an approach based on cluster Granger causality for fMRI data analysis. *Neuroimage* 52, 1444–1455. doi: 10.1016/j.neuroimage.2010.05.022

Takahashi, D. Y. (2009). *Medidas de Fluxo de Informação com Aplicação em Neurociência*. Ph.D. thesis, University of São Paulo. Available online at: http://www.teses.usp.br/teses/disponiveis/95/95131/tde-07062011-115256/en.php

Takahashi, D. Y., Baccalá, L. A., and Sameshima, K. (2010). Information theoretic interpretation of frequency domain connectivity measures. *Biol. Cybern.* 103, 463–469. doi: 10.1007/s00422-010-0410-x

Wu, G. R., Chen, F., Kang, D., Zhang, X., Marinazzo, D., and Chen, H. (2011). Multiscale causal connectivity analysis by canonical correlation: theory and application to epileptic brain. *IEEE Trans. Biomed. Eng.* 58, 3088–3096. doi: 10.1109/TBME.2011.2162669

A. APPENDIX

A.1 PROOF OF THEOREMS 1 AND 2 AND COROLLARIES 1 AND 2

The proofs in this section follow the pattern of those in Takahashi et al. (2010). The chief difference lies in the care needed regarding the order of the products between the defining matrices. Here we exhibit the main proof ingredients for reader convenience, with further details available in Takahashi (2009) and Takahashi et al. (2010).

Proof of Theorem 1 and Corollary 1. Let $W = [Y^T\ Z^T]^T$ be a second order stationary process satisfying the boundedness condition, using the following well known identity for determinants (Lütkepohl, 1996)

$$\det\left(S_W(\omega)\right) = \det\left(S_{ZZ}(\omega)\right. \\ \left. - S_{ZY}(\omega)S_{YY}^{-1}(\omega)S_{YZ}(\omega)\right)\det\left(S_{YY}(\omega)\right) \quad (A1)$$

leads to

$$C_{YZ}^{(b)}(\omega) = 1 - \det\left(S_{ZZ}(\omega)\right. \\ \left. - S_{ZY}(\omega)S_{YY}^{-1}(\omega)S_{YZ}(\omega)\right)\det\left(S_{ZZ}^{-1}(\omega)\right) \quad (A2)$$

$$= 1 - \det\left(S_{YY}(\omega)\right. \\ \left. - S_{YZ}(\omega)S_{ZZ}^{-1}(\omega)S_{ZY}(\omega)\right)\det\left(S_{YY}^{-1}(\omega)\right), \quad (A3)$$

under Equation (5).

Rewrite bPDC as

$$\pi_{ij}^{(b)}(\omega) = 1 - \det\left(P_{jj}^{-1}(\omega)\right. \\ \left. - P_{jj}^{-1}(\omega)\bar{A}_{ij}^*(\omega)\Sigma_{ii}^{-1}\bar{A}_{ij}(\omega)P_{jj}^{-1}(\omega)\right)\det\left(P_{jj}(\omega)\right), (A4)$$

so that using the following identities proved in Takahashi et al. (2010)

$$P_{jj}(\omega) = S_{\eta_j\eta_j}^{-1}(\omega), \quad (A5)$$

$$S_{\epsilon_i\eta_j}(\omega) = \bar{A}_{ij}(\omega)S_{\eta_j\eta_j}(\omega), \quad (A6)$$

and for all $\omega \in [-\pi, \pi)$

$$S_{\epsilon_i\epsilon_i}(\omega) = \Sigma_{ii}, \quad (A7)$$

back substituted into Equation (A4) leads to

$$\pi_{ij}^{(b)}(\omega) = 1 - \det\left(S_{\eta_j\eta_j}(\omega)\right. \\ \left. - S_{\eta_j\epsilon_i}(\omega)\Sigma_{ii}^{-1}S_{\epsilon_i\eta_j}(\omega)\right)\det\left(S_{\eta_j\eta_j}^{-1}(\omega)\right), \quad (A8)$$

so that using Equation (A2) shows that the right-hand side of Equation (A8) actually is $C_{\epsilon_i\eta_j}^{(b)}(\omega)$ as we set out to prove. Corollary 1 is immediate from Theorem 1 and Equation (7). □

Proof of Theorem 2 and Corollary 2. Theorem 2 is obtained by rewriting $C_{X_i\zeta_j}^{(b)}(\omega)$ using Equation (A3) noting that

$$S_{X_i\zeta_j}(\omega) = \bar{H}_{ij}(\omega)S_{\zeta_j\zeta_j}(\omega) \quad (A9)$$

and for all $\omega \in [-\pi, \pi)$

$$S_{\zeta_j\zeta_j}(\omega) = \Theta_{jj}^{-1}. \quad (A10)$$

Corollary 2 follows from Theorem 2 and Equation (7). □

A.2 PROOF OF THEOREMS 3 AND 4

Brillinger (1981, chapter 10) introduced the idea of canonical coherence for time series. We restate his result under our notation as the following theorem.

Theorem 5 (Brillinger, Theorem 10.3.2). *Let X and Y be m_1 and m_2-dimensional time-series jointly satisfying the boundedness condition. For $m \leq \min\{m_1, m_2\}$, the following identity holds:*

$$C_{XY}^{(c_m)}(\omega) = \lambda^m(S_{YY}^{-1}(\omega)S_{YX}(\omega)S_{XX}^{-1}(\omega)S_{XY}(\omega)) \quad (A11)$$

$$= \lambda^m(S_{XX}^{-1}(\omega)S_{XY}(\omega)S_{YY}^{-1}(\omega)S_{YX}(\omega)). \quad (A12)$$

Proof of Theorem 3. From Equations (18), (A11), we have

$$C_{\epsilon_i\eta_j}^{(c_m)}(\omega) = \lambda^m(S_{\eta_j\eta_j}^{-1}(\omega)S_{\eta_j\epsilon_i}(\omega)S_{\epsilon_i\epsilon_i}^{-1}(\omega)S_{\epsilon_i\eta_j}(\omega)). \quad (A13)$$

Now, from Equations (A5), (A6), and (A7) it follows that

$$S_{\eta_j\eta_j}^{-1}(\omega)S_{\eta_j\epsilon_i}(\omega)S_{\epsilon_i\epsilon_i}^{-1}(\omega)S_{\epsilon_i\eta_j}(\omega) = \bar{A}_{ij}^*(\omega)\Sigma_{ii}^{-1}\bar{A}_{ij}(\omega)P_{jj}^{-1}(\omega), \quad (A14)$$

which proves Equation (20).

To prove Equation (21), we use Equations (19), (A12) to obtain

$$C_{X_i\zeta_j}^{(c_m)}(\omega) = \lambda^m(S_{X_iX_i}^{-1}(\omega)S_{X_i\zeta_j}(\omega)S_{\zeta_j\zeta_j}^{-1}(\omega)S_{\zeta_jX_i}(\omega)). \quad (A15)$$

Finally, from Equations (A9), (A10), we have

$$S_{X_iX_i}^{-1}(\omega)S_{X_i\zeta_j}(\omega)S_{\zeta_j\zeta_j}^{-1}(\omega)S_{\zeta_jX_i}(\omega) = S_{ii}^{-1}(\omega)\bar{H}_{ij}(\omega)\Theta_{jj}^{-1}\bar{H}_{ij}^*(\omega), \quad (A16)$$

which concludes the proof. □

Proof of Theorem 4. Rewrite bPDC as

$$1 - \pi_{ij}^{(b)}(\omega) = \det\left(I - \bar{A}_{ij}^*(\omega)\Sigma_{ii}^{-1}\bar{A}_{ij}(\omega)P_{jj}^{-1}(\omega)\right). \quad (A17)$$

Now, Equation (22) is a straightforward consequence of the relationship between eigenvalues and the determinant of a matrix. A similar argument proves Equation (23). □