# COHERENCE-BASED TIME SERIES CLUSTERING FOR STATISTICAL INFERENCE AND VISUALIZATION OF BRAIN CONNECTIVITY

BY CAROLINA EUÁN, YING SUN AND HERNANDO OMBAO

*King Abdullah University of Science and Technology (KAUST)*

We develop the hierarchical cluster coherence (HCC) method for brain signals, a procedure for characterizing connectivity in a network by clustering nodes or groups of channels that display a high level of coordination as measured by "cluster-coherence." While the most common approach to measure dependence between clusters is through pairs of single time series, our method proposes cluster coherence which measures dependence between pairs of whole clusters rather than between single elements. Thus it takes into account both the dependence between clusters and within channels in a cluster. The identified clusters contain time series that exhibit high cross-dependence in the spectral domain. Simulation studies demonstrate that the proposed HCC method is competitive with the other feature-based clustering methods. To study clustering in a network of multichannel electroencephalograms (EEG) during an epileptic seizure, we applied the HCC method and identified connectivity on alpha $(8, 12)$ Hertz and beta $(16, 30)$ Hertz bands at different phases of the recording: before an epileptic seizure, during the early and middle phases of the seizure episode. To increase the potential impact of HCC in neuroscience, we also developed the HCC-Vis, an R-Shiny app (RStudio), which can be downloaded from https://carolinaeuan.shinyapps.io/hcc-vis/.

**1. Introduction.** Brain signals, for example, electroencephalograms (EEGs), are often viewed as mixtures of oscillations with varying amplitudes across locations on the scalp. Thus, it is natural to study these signals using the frequency domain approach of time series. There exists a direct relationship between energy distribution among different frequency bands and brain activity. The frequency bands of interest in neuroscience are in the range of $(0, 50)$ Hertz (Hz), where the frequency bands are named as follows: delta $(0, 4)$ Hz, theta $(4, 8)$ Hz, alpha $(8, 12)$ Hz, beta $(12, 30)$ Hz and gamma $(30, 50)$ Hz (Buzsáki and Draguhn (2004)). Energy in the lower frequency bands, delta and theta, are typically related to resting state of the brain, while activities in the higher frequency bands, alpha and beta, are more likely linked to cognitive activity. A first approach to study brain connectivity (i.e., the dependence between brain signals in a network) is via

---

covariance, correlation or precision matrices, which is commonly used on functional magnetic resonance imaging (fMRI) data. Among dependency measures in the frequency domain, coherence and partial coherence are the most commonly used in the study of EEG signals (Bowyer (2016), Cribben and Yu (2017), Decker, Fillmore and Coherence (2017), Fiecas et al. (2010), Ombao and Van Bellegem (2008)). Coherence is the analog of cross-correlation between two time series in the frequency domain. Since there is a one to one correspondence between the covariance matrix and the spectral matrix, the dependence between time series can be characterized either through the temporal or frequency domain.

EEGs are multivariate times series recordings from many electrodes placed on the scalp that reflect electrical activity of neuronal populations on the scalp. To quantify connectivity between brain regions, we need to measure dependency or coherence between sets of time series. One approach to explore the connectivity between signals is to set a threshold for the observed coherence values; then, two signals are connected if the observed coherence exceeds the threshold. This procedure is simple and computationally fast. However, this procedure could produce groups having time series or signals with low levels of dependence. As an alternative, we propose to study brain connectivity in a network through clustering. Compared to the first approach, clustering methods will produce groups of time series that show a high dependence between channels within a cluster and low dependence between different clusters. With this approach, clusters of time series with high dependence in the frequency domain correspond to connected brain regions. Moreover, a clustering method based on the spectral features of the EEG data will produce groups of brain signals that are easily interpreted. In this paper, we propose a new time series clustering method based on coherence to describe brain connectivity.

The primary contributions of this paper are the following: (1) We propose a divergence measure that is based on coherence between groups rather than pairwise between channels in each group. Thus, the measure automatically takes into account the dependence between channels within each cluster. (2) The HCC algorithm identifies highly connected brain regions through the clusters produced. (3) We propose new visualization tools for the clustering produced by the HCC method. These are implemented in the HCC-Vis, an interactive Shiny App, which is useful for the neuroscience research.

The use of spectral features of time series for discrimination and classification has attracted the attention of many researchers. A review of different methods for clustering time series can be found in Caiado, Maharaj and D'Urso (2016). Among the existing clustering methods for time series, Caiado, Crato and Peña (2006) were the first to propose the use of periodograms of time series to define similarity in hierarchical clustering. A more recent study proposed the hierarchical spectral clustering (HSM) method (Euán, Ombao and Ortega (2018)), which also considers the spectral density for the classification of time series. Unlike classic hierarchical methods, the HSM method does not need a linkage function between clusters,

since when merging clusters all the time series contained in the cluster are used
to update the spectral density estimation, and the dissimilarity between clusters is
computed by using the updated spectral densities. The general goal is to identify
clusters of time series with similar spectral densities or similar oscillatory patterns
(spectrally synchronized). Another classification method is based on the discrete
wavelet transform (DWT) of the time series (see Maharaj, D'Urso and Galaged-
era (2010), D'Urso and Maharaj (2012)), for a more robust clustering result un-
der nonstationary variances. The DWT dyadically segments the frequency axis
through the repeated application of filtering. This method also produces clusters
of time series with similar spectra. Model-based clustering (D'Urso, De Giovanni
and Massari (2016)) and fuzzy clustering methods for time series on frequency
domain (Maharaj and D'Urso (2011)) are also present in the literature. However,
the similarity in spectra or waveforms does not necessarily imply a dependence
between the time series; it is possible to have two time series generated from the
same model with the same spectral density but independent.

Our goal is to develop a method that identifies clusters of channels or time series
such that the within-cluster dependence (as measured by coherence) is high and
the dependence between clusters is low. Therefore, coherence plays a central role
in this paper. Maharaj and D'Urso (2010) proposed coherence-based clustering
procedures for classifying time series, hierarchical and nonhierarchical methods;
among the hierarchical clustering, they proposed to measure the similarity between
clusters with average coherence or minimum coherence. They showed that coher-
ence could identify linearly related groupings of time series, even in cases with
nonstationary variances. However, their measures based on average and minimum
coherence are pairwise measures between elements of different clusters and do not
take into account the dependence among all channels within a cluster. To address
the within-cluster dependence, we need to generalize coherence to measure the
dependence between two clusters of time series.

In the literature, we can find two proposed extensions of coherence: block
coherence (Nedungadi, Ding and Rangarajan (2011)) and canonical coherence
(Takahashi, Baccalá and Sameshima (2014)). Block coherence at frequency
$\omega$ between two sets of multivariate time series, $\mathbf{X}$ and $\mathbf{Y}$, is defined as $1 -
\frac{\det(S_{\mathbf{X},\mathbf{Y}}(\omega))}{\det(S_{\mathbf{X}}(\omega))\det(S_{\mathbf{Y}}(\omega))}$, where $S(\omega)$ denotes the spectral matrix. However, block co-
herence could sometimes mischaracterize the nature of dependence between a pair
of clusters because it would give a high value even when the dependence is driven
only by a small number of channels (see Illustration 1 in Section 3). To overcome
this limitation, canonical coherence was proposed. Canonical coherence projects
a block of time series into canonical variables, then computes the block coherence
between the canonical variables from different blocks. This procedure improves
the block coherence, but the identified clusters may lack interpretability. In this
paper, we develop a new notion of coherence, cluster coherence, that measures the
dependence between groups of time series. Since cluster coherence is defined by

the pairwise coherence matrix, the observed dissimilarity values can be interpreted regarding the original signals. Then, we apply the cluster coherence in a hierarchical algorithm. The HCC method can be applied to any set stationary or locally stationary times series in the time domain to identify clusters whose members show high dependency.

The rest of this paper is organized as follows. Section 2 introduces the frequency analysis of multivariate time series. In Section 3, we first present the cluster coherence measure with comparisons to average coherence and introduce the HCC clustering algorithm, and then we propose a bootstrap procedure to test the significance of the HCC clustering results. In Section 4, we conduct simulation studies to compare the performance of the HCC method with other time series clustering methods and to evaluate the power of the proposed bootstrap test. Finally, in Section 5, we study an EEG recorded during an epileptic seizure to identify brain regions connected in the alpha and beta bands.

**2. Spectral analysis for multivariate time series.** Let $\mathbf{X}(t) = (X_1(t), \ldots, X_N(t))^T$ be a $N$-variate stationary time series with mean $\mathbf{0}$ and covariance matrix function $\Gamma(\cdot)$ with absolutely summable elements denoted by $\gamma_{j,k}(\cdot)$. Then, for $j, k = 1, \ldots, N$, the function $f_{j,k}(\omega) = \sum_{-\infty}^{\infty} \gamma_{j,k}(h) e^{-i2\pi\omega h}$, $-\frac{1}{2} \leq \omega \leq \frac{1}{2}$, is called the cross spectrum or cross spectral density of $X_j(t)$ and $X_k(t)$ for $j \neq k$ and the auto-spectrum for $j = k$ (see Brockwell and Davis (2006), Shumway and Stoffer (2011)). The matrix

$$\mathbf{S}(\omega) = \begin{pmatrix} f_{1,1}(\omega) & f_{1,2}(\omega) & \cdots & f_{1,N}(\omega) \\ f_{2,1}(\omega) & f_{2,2}(\omega) & \cdots & f_{2,N}(\omega) \\ \vdots & \ddots & \ddots & \vdots \\ f_{N,1}(\omega) & f_{N,2}(\omega) & \cdots & f_{N,N}(\omega) \end{pmatrix}$$

is called the spectral density matrix of $\mathbf{X}(t)$ at frequency $\omega$. Statistical inference in frequency analysis of multivariate time series is based on the periodogram which we now define. Let $\{\mathbf{X}(1), \ldots, \mathbf{X}(T)\}$ be an observed time series; then the periodogram matrix at the frequencies $w_j = j/T$ is defined to be $\mathbf{I}(\omega_j) = \frac{1}{T}(\sum_{t=1}^{T} \mathbf{X}(t) e^{-it2\pi\omega_j})(\sum_{t=1}^{T} \mathbf{X}(t) e^{-it2\pi\omega_j})^*$, where $*$ denotes complex conjugate transpose. Although the periodogram is not a consistent estimator of the spectral density, consistent estimators can be constructed by smoothing the periodogram (see Brockwell and Davis (2006), Chapter 11, Shumway and Stoffer (2011), Chapter 4). Notice that contrary to the univariate case, the cross spectrum is not constrained to be positive. It is in fact complex-valued. However, the spectral matrix needs to be a positive definite spectral density matrix. To preserve positive-definiteness of the spectral matrix estimator, the classic procedure is to use the same degree of smoothness in all the spectral estimates. If it is not reasonable to assume the same degree of smoothness for all spectral functions, there exist nonparametric methods based on the Cholesky to estimate the spectral matrix with different degree of smoothness (see Pawitan (1996), Rosen and Stoffer (2007)).

Let $\{dW_1(\omega), \ldots, dW_N(\omega)\}$ be the zero mean orthogonal increment processes in the univariate spectral representation, that is, $X_j(t) = \int e^{it2\pi\omega} \, dW_j(\omega)$ for $j = 1, \ldots, N$. The cross spectrum and auto-spectrum satisfy that $f_{j,k}(\omega) \, d\omega = \mathbb{E}[dW_j(\omega) \, \overline{dW_k(\omega)}] = \text{cov}(dW_j(\omega), dW_k(\omega))$, for $j, k = 1, \ldots, N$. If $\omega_1 \neq \omega_2$, then $\mathbb{E}[dW_j(\omega_1) \, \overline{dW_k(\omega_2)}] = 0$. Thus, the coherence at frequency $\omega$ is defined as

$$(2.1) \qquad \kappa_{j,k}(\omega) = \frac{|f_{j,k}(\omega)|^2}{f_{j,j}(\omega) f_{k,k}(\omega)}, \qquad -\frac{1}{2} \leq \omega \leq \frac{1}{2},$$

which measures the correlation between $dW_j(\omega)$ and $dW_k(\omega)$. In this paper, we call that two time series are dependent if they are correlated in the frequency domain, that is, $\kappa_{j,k}(\omega) > 0$ at a single frequency or in a frequency band. The coherence matrix of $\mathbf{X}(t)$ at frequency $\omega$, $\mathbf{C}(\omega)$, is equal to one in the diagonal and $\kappa_{j,k}(\omega)$ in the off-diagonal elements.

Under stationarity, coherence could vary across frequencies but remains constant across time. The concept of evolutionary coherence has been studied by Ombao and Van Bellegem (2008). An alternative to coherence for studying functional connectivity is partial coherence, which measures the direct linear association between any pair components of a multivariate time series after removing the linear effects of the other components (Fiecas and Ombao (2011), Fiecas et al. (2010)). In this paper, we will focus on coherence and how to extend this concept to measure dependency between clusters of EEG signals (or multivariate time series).

**3. Hierarchical cluster coherence method.** In this paper, we will refer to the electrode located at the scalp as a channel and in the case of multiple electrodes as a multichannel.

We propose the hierarchical cluster coherence (HCC) method to describe brain connectivity. The HCC method uses a hierarchical algorithm to identify clusters containing highly dependent time series in the frequency domain, based on our coherence-based measure of similarity. First, we define the cluster coherence between two groups of time series and illustrate the advantages over other measures of dependency between clusters. Then, we present the HCC clustering algorithm.

3.1. *Cluster coherence.* Consider the $N$-variate stationary time series $\mathbf{X}(t)$ as two different blocks of time series, that is, $\mathbf{X}(t) = [\mathbf{X}_1(t) \ \mathbf{X}_2(t)]$, where $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ are stationary multivariate time series of dimension $n_1$ and $n_2 = N - n_1$, respectively. We can think of $\mathbf{X}_1(t)$ as a collection of time series in one cluster and $\mathbf{X}_2(t)$ as a second collection of time series in a different cluster. Consider the coherence matrix of $\mathbf{X}(t)$, as a block matrix for each $\omega$, that is,

$$(3.1) \qquad \mathbf{C}(\omega) = \begin{pmatrix} \mathbf{C}_{1,1}(\omega) & \mathbf{C}_{1,2}(\omega) \\ \mathbf{C}_{2,1}(\omega) & \mathbf{C}_{2,2}(\omega) \end{pmatrix},$$

where $\mathbf{C}_{j,j}(\omega)$ is the coherence matrix of $\mathbf{X}_j(t)$, $j = 1, 2$, and $\mathbf{C}_{j,k}(\omega)$ is the coherence between elements of $\mathbf{X}_j(t)$ and elements of $\mathbf{X}_k(t)$, $j \neq k$. Let $\{\lambda_1^1(\omega), \ldots, \lambda_{n_1}^1(\omega)\}$ and $\{\lambda_1^2(\omega), \ldots, \lambda_{n_2}^2(\omega)\}$ be the normalized eigenvalues of the within cluster coherence matrices $\mathbf{C}_{1,1}(\omega)$ and $\mathbf{C}_{2,2}(\omega)$, and let $\{\lambda_1(\omega), \ldots, \lambda_N(\omega)\}$ be the normalized eigenvalues of $\mathbf{C}(\omega)$. We define the cluster coherence between $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ as

$$(3.2) \qquad \text{CCo}(\omega) = \left( \sum_{j=1}^{N} |\lambda_{[j]}(\omega) - \lambda_{[j]}^*(\omega)|^p \right)^{(1/p)}$$

where $\lambda_{[j]}(\omega)$ is the $j$th largest eigenvalue of $\{\lambda_1(\omega), \ldots, \lambda_N(\omega)\}$ and $\lambda_{[j]}^*(\omega)$ is the $j$th largest eigenvalue of $\{\lambda_1^1(\omega), \ldots, \lambda_{n_1}^1(\omega), \lambda_1^2(\omega), \ldots, \lambda_{n_2}^2(\omega)\}$. The eigenvalues are normalized by using the corresponding $L^p$ norm. We define cluster coherence in terms of the $L^p$ norm, but we consider only the cases where $p = 1$ or 2.

PROPERTIES OF CCo($\omega$).

P1. CCo($\omega$) *is bounded.* The cluster coherence between $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ is bounded between 0 and 1, that is, $0 \leq \text{CCo}(\omega) \leq 1$. This can be deduced from the definition of cluster coherence between $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$, where the norm of all of the eigenvalues is equal to 1 by construction.

P2. *If* $\mathbf{X}_1(t)$ *and* $\mathbf{X}_2(t)$ *are uncorrelated, that is, the coherence value between any two time series among clusters is zero, then* CCo($\omega$) $= 0$ *for all* $\omega$. If $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ are independent at frequency $\omega$, then $\mathbf{C}_{1,2}(\omega) = \mathbf{C}_{2,1}(\omega) = \mathbf{0}$. As a result, the set of eigenvalues of the whole coherence matrix $\mathbf{C}(\omega)$ is equal to merging the eigenvalues of $\mathbf{C}_{1,1}(\omega)$ and $\mathbf{C}_{2,2}(\omega)$, $\lambda_{[j]}(\omega) = \lambda_{[j]}^*(\omega)$ for $j = 1, \ldots, N$. Then, cluster coherence is equal to zero.

P3. *If* $\mathbf{X}_1(t)$ *and* $\mathbf{X}_2(t)$ *are perfectly correlated, then* CCo($\omega$) $= 1$ *for all* $\omega$. When $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ are perfectly correlated, $\mathbf{C}_{j,k}(\omega) = \mathbf{1}_{n_j \times n_k}$ (all-ones matrix of dimension $n_j \times n_k$) for all $\omega$ and $j, k = 1, 2$. Then, $\lambda_{[1]}(\omega) = 1$ and $\lambda_{[j]}(\omega) = 0$ for $j = 2, \ldots, N$ and $\lambda_{[1]}^*(\omega) = \lambda_{[2]}^*(\omega) = 1/2$ and $\lambda_{[j]}^*(\omega) = 0$ for $j = 3, \ldots, N$.

P4. *If* $n_1 = n_2 = 1$ *and* $p = 1$, *then* CCo($\omega$) $= \kappa_{1,2}(\omega)$ (equation (2.1)). This is because $n_1 = n_2 = 1$ and $p = 1$, hence $\{\lambda_{[1]}(\omega), \lambda_{[2]}(\omega)\} = \{\frac{1+\kappa_{1,2}(\omega)}{2}, \frac{1-\kappa_{1,2}(\omega)}{2}\}$ and $\{\lambda_{[1]}^*(\omega), \lambda_{[2]}^*(\omega)\} = \{\frac{1}{2}, \frac{1}{2}\}$. Finally, by computing (3.2), CCo($\omega$) $= \kappa_{1,2}(\omega)$.

REMARK 1. Cluster coherence compares the eigenvalues of the coherence matrix between clusters with the uncorrelated case, that is, when $\mathbf{C}_{1,2}(\omega) = \mathbf{C}_{2,1}(\omega) = \mathbf{0}$. If the between-cluster coherence is close to 1, then the clusters are highly dependent in the frequency domain.

REMARK 2. When $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ are uncorrelated, the set of eigenvalues of the whole coherence matrix $\mathbf{C}(\omega)$ is the combined set of eigenvalues of $\mathbf{C}_{1,1}(\omega)$

and $\mathbf{C}_{2,2}(\omega)$ for all $\omega$. This fact produces a value of cluster coherence equal to zero.

REMARK 3.   In contrast, if $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ are perfectly correlated (P3), we can characterize them linearly via a latent signal $W(t)$ that is shared by components of $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ in the frequency domain. This will produce that the largest eigenvalue of $\mathbf{C}(\omega)$ is close to 1 and the rest of the eigenvalues are close to zero. Consequently, cluster coherence will be close to 1.

REMARK 4.   Cluster coherence is based on a distance between matrices, which is defined as a function of eigenvalues. In general, two matrices, $\mathbf{A}$ and $\mathbf{B}$, can have the same set of eigenvalues even though their entries are very different (e.g., when $\mathbf{B}$ is a rotation of $\mathbf{A}$). However, this is not an issue with the problem that is addressed in this paper. We define CCo between a chosen pair of matrices as follows, the coherence matrix $\mathbf{C}(\omega)$ and the uncorrelated-clusters coherence matrix $\tilde{\mathbf{C}}(\omega)$ which is equal to $\mathbf{C}_{1,1}(\omega)$ and $\mathbf{C}_{2,2}(\omega)$ in the block diagonal, and zero in the off block diagonal. Then, one other possible matrix that could have equal eigenvalues with $\tilde{\mathbf{C}}(\omega)$ is the one when an orthogonal transformation of the time series within each cluster is applied. Then, CCo$(\omega)$ does not depend on the time series order within each cluster.

3.2. *Illustrative examples*: *Comparison of the* CCo *with average coherence and block coherence.*   In hierarchical clustering, the most commonly used measure for cluster similarity is the single linkage that corresponds to average coherence (AC). AC between $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ is defined as mean$(\mathbf{C}_{1,2}(\omega))$ in (3.1), where mean$(\mathbf{A})$ is the average over all elements of the matrix $\mathbf{A}$. Therefore, AC considers only the pairwise coherence in $\mathbf{C}_{1,2}(\omega)$, neglecting within coherences $\mathbf{C}_{1,1}(\omega)$ and $\mathbf{C}_{2,2}(\omega)$. An alternative to AC is Block Coherence (BC). BC between $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ is defined as $1 - \det(\mathbf{C}(\omega))/\det(\mathbf{C}_{1,1}(\omega))\det(\mathbf{C}_{2,2}(\omega))$ in (3.1). However, BC tends to overestimate the dependency between clusters (see Illustration 1). In contrast, CCo measures dependency using the complete information of the coherence matrix represented by eigenvalues. CCo has shown (under different simulated scenarios) a better balance between the within-cluster dependency and the between-cluster dependency, enhancing the clustering performance.

Our first illustrative example is based on a mixture of autoregressive processes to capture the oscillations in brain signals (Gao et al. (2016)). The goal is to show that when the within-cluster dependency is high, then average coherence and block coherence overestimate the dependency among clusters.

ILLUSTRATION 1.   Let $Z_1(t)$, $Z_2(t)$ and $Z_3(t)$ be independent AR(2) latent processes with the unimodal spectral density peaks at 3, 5 and 9 Hz, respectively.
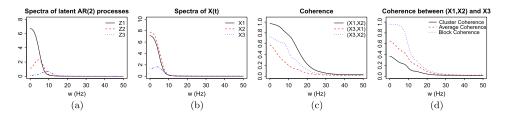
FIG. 1.    *Estimated spectral densities of*: (a) *latent* AR(2) *processes and* (b) *observed signals. Estimated coherence between signals*: (c) *pairwise coherence and* (d) *coherence measures between clusters* $\{X_1(t), X_2(t)\}$ *and* $\{X_3(t)\}$.

Let $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t))^T$ be a three-variate time series generated by a mixture of these latent processes,

$$X_1(t) = Z_1(t) + 0.2Z_2(t) + \varepsilon_1(t),$$
$$X_2(t) = Z_1(t) + 0.6Z_2(t) + \varepsilon_2(t),$$
$$X_3(t) = 0.3Z_1(t) + 0.7Z_2(t) + 0.3Z_3(t) + \varepsilon_3(t),$$

where $\varepsilon_i(t)$ are white noise sequences. We simulate 1000 replicates with 1000 time points and sampling frequency 100 Hz, and compute the cluster coherence as $\sum_{j=1}^{N} |\lambda_{[j]}(\omega) - \lambda_{[j]}^*(\omega)|$ ($p = 1$). Figure 1(a) shows the estimated spectra of the latent variables; these latent variables represent activity on the delta, theta and alpha bands. Figure 1(b) shows the estimated spectral densities for univariate time series $X_j(t)$. $X_1(t)$ and $X_2(t)$ have more similar spectral densities, which is a consequence of being more highly correlated in the low frequencies than $X_3(t)$. Figure 1(c) shows the mean curve over the 1000 replicates of the estimated coherence on frequency bands delta, theta and alpha. We observe that $X_1(t)$ and $X_2(t)$ are highly correlated signals in the low frequency band. However, $X_3(t)$ is highly correlated with $X_2(t)$ but less correlated with $X_1(t)$. In this sense, it is reasonable to have two clusters, $\{X_1(t), X_2(t)\}$ and $\{X_3(t)\}$. Figure 1(d) shows the mean curve over the 1000 replicates of the estimated functions of cluster coherence, average coherence and block coherence, respectively. Consider 0.5 as a threshold to decide whether or not two clusters should be merged. When using the average coherence or block coherence, it is very likely that $\{X_1(t), X_2(t)\}$ and $\{X_3(t)\}$ will be merged as one cluster, since these values are larger than 0.5 on delta and theta bands, resulting in a bigger cluster than expected. However, when using cluster coherence, $\{X_1(t), X_2(t)\}$ and $\{X_3(t)\}$ may not be merged, since cluster coherence is lower than 0.4.

There exist different factors that influence the similarity measures between clusters. It is desirable to select a measure that will be less influenced by these factors. The next illustrative example explores the effect of noise variability, cluster size and low dependency between clusters on the CCo and AC.

ILLUSTRATION 2. Let $\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_{n_1}(t), X_{n_1+1}(t), \ldots,$ $X_{n_1+n_2}(t))^T$ be a multivariate time series, where each $X_j(t)$ is an AR(2) process with a unimodal spectral density peak at 10 Hz and correlated innovations, that is, $\mathbf{X}(t)$ follows a VAR(2) process with correlated innovations. In matrix notation,

$$\mathbf{X}(t) = A_1 \mathbf{X}(t-1) + A_2 \mathbf{X}(t-2) + \boldsymbol{\varepsilon}(t),$$

where $A_m$ is a diagonal matrix with $\phi_m$ (the autorregresive coefficientes) on the diagonal, $m = 1, 2$, and $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \ldots, \varepsilon_{n_1+n_2}(t))$ is white noise with covariance matrix

$$(3.3) \qquad \Sigma_\varepsilon = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \ldots & \rho_1 & \delta & \delta & \ldots & \delta \\ \rho_1 & 1 & \ldots & \rho_1 & \delta & \delta & \ldots & \delta \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \rho_1 & \rho_1 & \ldots & 1 & \delta & \delta & \ldots & \delta \\ \delta & \delta & \ldots & \delta & 1 & \rho_2 & \ldots & \rho_2 \\ \delta & \delta & \ldots & \delta & \rho_2 & 1 & \ldots & \rho_2 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \delta & \delta & \ldots & \delta & \rho_2 & \rho_2 & \ldots & 1 \end{pmatrix}.$$

Then, the multivariate time series $\mathbf{X}(t)$ has two independent clusters $C_1 = \{X_1(t), \ldots, X_{n_1}(t)\}$ and $C_2 = \{X_{n_1+1}(t), \ldots, X_{n_1+n_2}(t)\}$ when $\delta = 0$ and low dependent clusters when $0 < \delta \ll 1$. We simulate 1000 replicates of the multivariate time series with 1000 time points at sampling frequency 100 Hz. The correlation coefficients are fixed as $\rho_1 = 0.5$ and $\rho_2 = 0.7$, then any value of $|\delta| > \sqrt{0.35}$ guarantees the positive definitness of $\Sigma_\varepsilon$. We select the parameter values with different purposes as follows: (1) Noise variability, $\sigma^2 = 1, 2, 4, 8$; (2) Cluster size, $n_1 = 10, 20, 50, 100$ and $n_2 = 12, 24, 60, 120$; (3) Zero/Low dependency between clusters, $\delta = 0, 0.1, 0.2, 0.3$.

We assume that $C_1$ and $C_2$ are known, and we compute the CCo and AC between the two clusters. Figure 2 shows the boxplots of the CCo and AC under the different sets of values for the parameters. Since the within correlation of the clusters depends on the correlation between the innovations, the coherence shows low variability across frequencies. Then, the boxplots are computed with all values across frequencies and replicates. Within each plot, the noise variability, $\sigma^2$, changes from 1 to 8, and the CCo values are on the left side while the AC values are on the right side. Between plots, we vary per column the zero/low dependency between clusters, $\delta$, and per row the cluster size, $n_1$ and $n_2$.

When there is no dependence between $C_1$ and $C_2$ (i.e., $\delta = 0$) and for a fixed value of $\sigma^2$, increasing the clusters size does not influence on the distribution of the CCo and AC values. However, a small change in the dispersion of the CCo and AC values is observed when $\sigma^2$ increases. If the clusters are weakly correlated, we expect that the similarity measures increase, especially when $\delta \geq 0.2$. The CCo
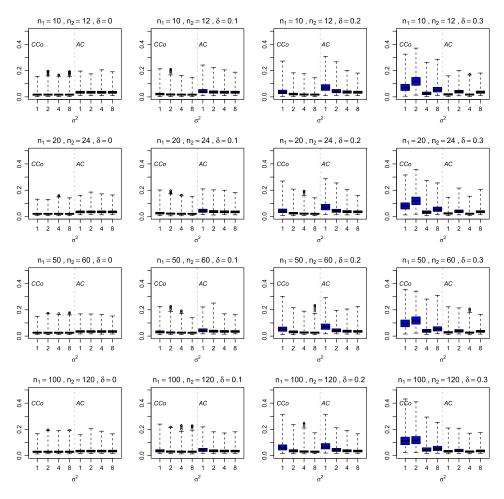
FIG. 2. *Boxplots of the Cluster Coherence* (CCo) *and Average Coherence* (AC) *with different parameters value in Illustration* 2 *for all ω. By column, we vary the correlation δ, and by row, we vary the cluster size, $n_1$, and $n_2$.*

values are higher than the AC when $\delta \geq 0.2$ and $\sigma^2 = 1, 2, 4$, this suggests a higher power of CCo to identify dependent clusters. In most of the cases, the cluster size has a positive impact on the CCo values, and the noise variability affects the performance when it is large. This experiment was repeated to consider the case of unbalanced clusters. Overall, CCo measures dependency among clusters better than AC with moderate unbalanced cluster size ($n_1 = 50$ and $n_2 = 12$).

3.3. *Clustering algorithm.* The HCC method uses cluster coherence as a measure of similarity between clusters in each iteration of the clustering algorithm.

HIERARCHICAL CLUSTER COHERENCE (HCC) ALGORITHM. Let $\mathbf{X}(t) = (X_1(t)X_2(t)\cdots X_N(t))^T$, $t = 1, \ldots, T$ be a set of time series and let $\Omega_{12} = (\omega_1, \omega_2)$ be the frequency band of interest. The procedure starts with clusters $\{C_1, C_2, \ldots, C_k\}$ where the initial value of $k = N$ so that each cluster contains one signal.

**Step 1.** Estimate the coherence matrix $\mathbf{C}(\omega) = \{\kappa_{l,m}(\omega)\}_{l,m=1,\ldots,k}$ for each frequency $\omega \in \Omega_{12}$.

**Step 2.** Compute the initial dissimilarity between clusters at band $\Omega_{12}$ as

$$d_{l,m} = d(C_l, C_m) = 1 - \frac{1}{\omega_2 - \omega_1} \int_{\omega_1}^{\omega_2} \kappa_{l,m}(\omega) \, d\omega.$$

**Step 3.** Find the two clusters with the lowest dissimilarity and merge these clusters. Assume WLOG that $C_{k-1}$ and $C_k$ are the two most similar clusters, then the new clusters will be $\{C_1, C_2, \ldots, C_{k-1} \cup C_k\}$.

**Step 4.** Reduce the number of clusters by one, that is, $k = k - 1$.

**Step 5.** Compute the new dissimilarity matrix as $d_{l,m} = d(C_l, C_m) = 1 - \frac{1}{\omega_2 - \omega_1} \times \int_{\omega_1}^{\omega_2} \mathrm{CCo}(\omega) \, d\omega$.

**Step 6.** Repeat Steps 3–5 until $k = 1$.

Note that in Step 5 the new dissimilarity between the updated cluster setting will correspond to coherence when the clusters have only one member each or cluster coherence otherwise.

In real applications, we need to choose the number of clusters. We propose to use the scree plot of the minimum dissimilarity value to decide the number of clusters. Specifically, we choose the smallest number of clusters $k$ such that the dissimilarity value will not decrease significantly from $k$ to $k + 1$. The method of scree plot is similar to the gap statistics (Tibshirani, Walther and Hastie (2001)), and was also used by Euán, Ombao and Ortega (2018) to choose the number of clusters for the HSM clustering method.

3.4. *Clustering uncertainty.* When applying the HCC method to real data sets, we shall take into account the two sources of uncertainty: (1) The estimation of the cluster coherence values and (2) The selection of the number of clusters. To quantify the clustering uncertainty, Kimes et al. (2017) proposed a Monte Carlo based approach for hierarchical clustering of multivariate data. With a similar idea, we propose a bootstrap procedure for the HCC method.

*Estimation of the spectra.* The HCC method requires estimating the spectral matrix. A class of estimators for the spectral matrix is the smoothed periodogram (Brockwell and Davis (2006)), $\hat{\mathbf{S}}(\omega_j) = \sum_{|k| \le m_T} W_T(k) \mathbf{I}(\omega_{j+k})$, where $W_n$ is a kernel function, $m_T$ is the smoothing window size and $T$ denotes the time series length. To ensure the mean square consistency of $\hat{\mathbf{S}}(\omega_j)$, the kernel function

must satisfy the condition that $\sum_{|k| \le m_T} W_T^2(k) \to 0$ when $T \to \infty$. Moreover, to preserve positive-definiteness, it is sufficient for $m_T$ to be the same for all entries of the spectral estimates. Under regularity conditions, the distribution of $\hat{\mathbf{S}}(\omega_j)$ can be approximated by a Normal distribution (see Brockwell and Davis (2006), Chapter 11) or a Wishart distribution (see Brillinger (1975), Chapter 7). The asymptotic covariance of the multivariate distribution depends on $W_n$, $m_T$ and the true spectral matrix $\mathbf{S}(\omega)$. In our implementation, we choose the Fejér kernel (see Brockwell and Davis (2006), Chapter 2). In practice, we need to choose the size of the smoothing window, $m_T$. We apply the generalized cross-validation (GCV) criteria proposed by Ombao et al. (2001) to select the size of the smoothing window. The selection of the spectral density estimator can be modified considering that the quality of the estimator could influence the accuracy of the clustering results. An alternative nonparametric estimator of the spectral density is the multitaper spectral estimator (Walden (2000)) which also has important optimality properties. A review of multitaper spectral analysis can be found in Babadi and Brown (2014).

Then, the auto-spectral estimates are the elements of the diagonal of $\hat{S}(\omega_j)$ and the cross-spectral estimates are the elements of the off diagonals of $\hat{S}(\omega_j)$. Let us denote by $\hat{f}_{k,l}(\omega_j)$, $k, l = 1, \dots, N$, the auto-spectral and cross-spectral estimates, then the estimator for pairwise coherence at frequency $\omega_j$ is

$$\hat{\kappa}_{k,l}(\omega_j) = \frac{|\hat{f}_{k,l}(\omega_j)|^2}{\hat{f}_{k,k}(\omega_j)\hat{f}_{l,l}(\omega_j)}.$$

The estimator of cluster coherence, $\hat{C\text{Co}}$, is computed based on the eigenvalues of the estimated coherence matrix $\hat{\mathbf{C}}(\omega)$. To make inference on the observed CCo values, we propose a bootstrap procedure as follows.

*Bootstrapping of* CCo. For each omega, let $S(\omega)$ be the estimated spectral matrix for two clusters, $C_1$ and $C_2$,

$$\hat{\mathbf{S}}(\omega) = \begin{pmatrix} \hat{\mathbf{S}}_{1,1}(\omega) & \hat{\mathbf{S}}_{1,2}(\omega) \\ \hat{\mathbf{S}}_{2,1}(\omega) & \hat{\mathbf{S}}_{2,2}(\omega) \end{pmatrix}.$$

Then, if $C_1$ and $C_2$ are uncorrelated at frequency $\omega$, $\mathbf{S}_{1,2}(\omega) = (\mathbf{S}_{2,1}(\omega))^t = \mathbf{0}$. To resample values of $\hat{C\text{Co}}(\omega)$ between clusters $C_1$ and $C_2$, we assume independence between them at frequency $\omega$. We first simulate a realization of the spectral matrix $\mathbf{S}_j^B(\omega)$ from either the Normal or Wishart distribution with

$$\mathbf{S}_{H_0}(\omega) = \begin{pmatrix} \hat{\mathbf{S}}_{1,1}(\omega) & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{S}}_{2,2}(\omega) \end{pmatrix},$$

by assuming independency at frequency omega, and we compute the value of the cluster coherence, $\text{CCo}_j^B$, for the simulated matrix. We repeat $M$ times to obtain a

bootstrap sample $\{CCo_1^B, \ldots, CCo_M^B\}$, and approximate the distribution of cluster coherence. Then, a bootstrap test can be performed to test if two clusters are independent at frequency $\omega$. Since the HCC method uses a hierarchical algorithm to identify potential clusters, the clusters are ordered and nested. Suppose the number of clusters chosen by the scree plot is $k$. Denote the set of clusters to be $\{C_1, \ldots, C_{k-1}, C_k\}$, where $C_{k-1}$ and $C_k$ are assumed to be the most dependent clusters. If there is significant evidence against independence among $C_{k-1}$ and $C_k$, we merge clusters $C_{k-1}$ and $C_k$. Then, the total number of clusters becomes $k-1$ instead of $k$, and the set of the clusters becomes $\{C_1, \ldots, C_{k-2}, C_{k-1} \cup C_k\}$. In summary, to quantify the significance of the identified clusters with the HCC method, we take the advantage of the hierarchal structure of the clusters and develop a bootstrap test for independence between the last merged two clusters.

**4. Simulation study.** In this section we consider different experiments to test the performance of the HCC method. Experiment 1 evaluates the power of the bootstrap test proposed in Section 3.4. Experiment 2 compares the HCC method with other nonparametric clustering methods. Finally, Experiment 3 explores the performance of the HCC method to cluster multiple correlated time series when the clusters are not entirely independent and under the presence of possible contaminated data.

To improve the comparison of the clustering results, we develop a visualization tool. This tool contains two plots: a scree plot and a cluster merging plot. In the scree plot, we consider the dissimilarity value obtained when merging two clusters. For a fixed number of clusters $k$, we plot this dissimilarity value; this graph is located in the displayed upper plot. The clustering merging plot shows the dynamic of the clustering method; on the $x$-axis, we shall have the number of clusters and on the $y$-axis colors denote time series belonging to the same cluster. From $k$ to $k-1$ clusters, one color should disappear, which means that the time series in that cluster were merged with another cluster.

4.1. *Power of the bootstrap test.* Our first experiment is similar to Illustration 2 in Section 3.2.

EXPERIMENT 1. Let $\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_{n_1}(t), X_{n_1+1}(t), \ldots, X_{n_1+n_2}(t))^T$ be a multivariate time series that follows the VAR(2) model with correlated innovations as in Illustration 2. $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \ldots, \varepsilon_{n_1+n_2}(t))$ is white noise with covariance matrix as in (3.3). We fix $\sigma^2 = 1$, $n_1 = 20$ and $n_2 = 24$, and we vary $\delta$ from 0 to 0.7. When $\delta = 0$, there are two independent clusters $\mathcal{C}_1 = \{X_1(t), X_2(t), \ldots, X_{20}(t)\}$ and $\mathcal{C}_2 = \{X_{21}(t), X_{22}(t), \ldots, X_{44}(t)\}$, where the within cluster dependency depends on $\rho_1$ and $\rho_2$, respectively. When $\delta$ increases, the clusters $\mathcal{C}_1$ and $\mathcal{C}_1$ become more dependent. We consider two cases: (1) $(\rho_1, \rho_2) = (0.7, 0.5)$ and (2) $(\rho_1, \rho_2) = (0.9, 0.7)$.
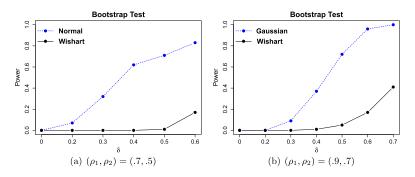
FIG. 3. *Power curve of the bootstrap test. The blue dashed curve corresponds the case when using the Normal approximation and the black curve is for the case of the Wishart approximation of the estimated spectral matrix.*

We simulate 100 replicates of the process $\mathbf{X}(t)$. For each replicate, we apply the HCC method and test the independence between the two estimated clusters $C_1$ and $C_2$. We set a significance level $\alpha = 0.05$ and compute the proportion of rejection of the hypothesis of independence between the two clusters. Figure 3 shows the results for both cases and under different values of $\delta$. In both cases, when $\delta$ increases, the proportion of rejection increases. If $\delta < .3$, there is not enough evidence to reject the independence hypothesis, suggesting two clusters rather than only one cluster. In this experiment, the test based on the Normal approximation shows better power than the test based on the Wishart approximation.

4.2. *Comparison study.* We compare the HCC method with other time series clustering methods. The goal is to compare the right classification of the HCC method with other clustering methods assuming the same number of clusters. We consider only nonparametric and model free clustering methods. We compare the HCC method with three hierarchical clustering methods with the single linkage function (Caiado, Maharaj and D'Urso (2016)): (1) *HA*—dissimilarity between time series computed as $d(X, Y) = \sum_h (\gamma_X(h) - \gamma_Y(h))^2$, where $\gamma(h)$ denotes the autocorrelation function, (2) *HS*—dissimilarity between time series computed as $d(X, Y) = \sum_\omega (f_X(\omega) - f_Y(\omega))^2$, where $f(\omega)$ denotes the spectral density function, and (3) *HC*—dissimilarity measure computed as $d(X, Y) = 1 - \kappa_{X,Y}(\omega)$, where $\kappa(\omega)$ is the pairwise coherence. We consider also three fuzzy clustering methods (D'Urso and Maharaj (2009), Maharaj and D'Urso (2011), Lafuente-Rego, D'Urso and Vilar (2018)): (1) *FA*—fuzzy $k$-means between autocorrelation functions, (2) *FS*—fuzzy $k$-means between the spectral densities, and (3) *PAMC*—partitioning clustering with distance based on the coherence function. Finally, we include the *HSM* clustering method that uses a hierarchical merger clustering method and computes similarity based on spectral densities. Note that HC uses average coherence, as introduced in Section 3.2, to measure the similarity between clusters.
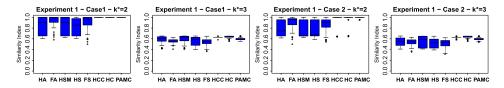
FIG. 4. *Comparative study under Experiment 1 setting. Each boxplot corresponds to the similarity values for each method with* 100 *replicates.*

We assume that the true clustering setting is known. We evaluate the performance of each clustering method by computing the following similarity index. Let $C = \{C_1, \ldots, C_k\}$ and $G = \{G_1, \ldots, G_{k^*}\}$ be the set of the $k$ true clusters and a $k^*$-cluster solution, respectively. Then, $\text{Sim}(G, C) = \frac{1}{k} \sum_{l=1}^{k} \max_{1 \leq m \leq k^*} \text{Sim}(G_m, C_l)$, where $\text{Sim}(G_m, C_l) = \frac{2|G_m \cup C_l|}{|G_m| + |C_l|}$. $\text{Sim}(G, C) = 1$ when the clusters setting $G$ is equal to the true cluster setting $C$. We also consider the Rand Index (Rand (1971)) to compare the methods and the results were very similar.

We consider Experiment 1 with $n_1 = 10$ and $n_2 = 12$ and two possible cases for $\delta$, Case 1: $\delta = 0$ and Case 2: $\delta = 0.4$. For all clustering methods, we consider cluster settings with two or three clusters. We compare the identified clusters with $C_1 = \{X_1(t), X_2(t), \ldots, X_{10}(t)\}$ and $C_2 = \{X_{11}(t), X_{12}(t), \ldots, X_{22}(t)\}$. Case 1 represents two well separated (independent) clusters while Case 2 considers two clusters with low dependency between them. Figure 4 shows the boxplot of the similarity values computed with 100 replicates. When the number of clusters agrees with the truth, that is, $k^* = k = 2$, the HCC method as well as the HA and PAMC methods show the best right classification index. When the number of clusters is overestimated, that is, $k^* = 3 > k$, the HCC method shows the highest similarity to the true clusters. When the clusters are not completely independent, as expected in practice, the HCC method can recover the right cluster structure.

EXPERIMENT 2. Let $Z_1(t)$, $Z_2(t)$ and $Z_3(t)$ be three independent AR(2) processes. Then, $\mathbf{X}(t)$ is a 9-multivariate time series generated by $\mathbf{X}(t) = \mathbf{A}\mathbf{Z}(t) + \boldsymbol{\varepsilon}(t)$, where $\mathbf{Z}(t) = (Z_1(t), Z_2(t), Z_3(t))^T$,

$$\mathbf{A}^T = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix},$$

and $\boldsymbol{\varepsilon}(t)$ is a 9-dimensional white noise with covariance matrix $\sigma^2 \mathbf{I}$. We consider two cases: Case 1: $Z_1(t)$, $Z_2(t)$ and $Z_3(t)$ have the same unimodal spectral density concentrated at 2 Hz and Case 2: $Z_1(t)$, $Z_2(t)$ have unimodal spectral density concentrated at 2 Hz and $Z_3(t)$ has unimodal spectral density at 4 Hz. Under this setting, $C_1 = \{X_1(t), X_2(t), X_3(t)\}$, $C_2 = \{X_4(t), X_5(t), X_6(t)\}$ and $C_3 = \{X_7(t), X_8(t), X_9(t)\}$ are the three independent clusters. We simulate time series of length $T = 1000$ with a sampling frequency of 100 Hz and $\sigma^2 = 50$, and
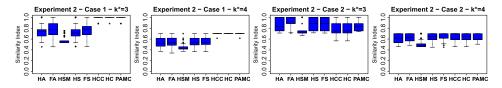
FIG. 5. *Comparative study under Experiment* 2 *setting. Each boxplot corresponds to the similarity values for each method with* 100 *replicates.*

apply the different 8 clustering methods with $k^* = 3, 4$ clusters. We consider 100 replicates and compute the similarity index for each replicate. Figure 5 shows the boxplot of the similarity values for each method. In each plot, the first two boxplots (HA and FA) correspond to clustering method based on the ACF, and the next three boxplots (HSM, HS and FS) show the clustering methods based on the spectra. These five clustering methods do not consider measures of cross dependence between clusters. In Case 1, the ACF and spectral densities are the same for all time series. Therefore, the similarity of the identified clusters with $\{C_1, C_2, C_3\}$ is low. HCC, HC and PAMC methods show a good classification performance even when the number of clusters might be overestimated ($k^* = 4$). In Case 2, elements in $C_3$ have different features compared to the other two clusters. This helps to improve the classification performance of the HA, FA, HSM, HS and FS methods.

In summary, when the clusters are completely independent as in Case 1 of Experiment 1 and Experiment 2, and the true number of clusters is known, the HCC method is as good as other clustering methods based on coherence. However, when the clusters are not completely independent, or the number of clusters is overestimated, the HCC method identifies clusters more accurately than the other methods. We explore the performance of the HCC method if we have a bigger number of time series. We present the simulation results in the Supplementary Material (Euán, Sun and Ombao (2019a)).

4.3. *Clustering of spatial correlated time series.* The next experiment involves spatially correlated signals to mimic real brain processes. We locate the time series on a 19-EEG array to visualize the clustering results. The HCC method does not use the locations of the EEG channels but computes the cluster coherence to identify the potential clusters.

EXPERIMENT 3. Let $X_s(t)$ be the signal at channel $s$. Figure 6(a) shows the spatial location of the channels in a 2D projection. We assume that $\{X_s(t)\}$ is generated by a mixture of independent time series $Z_{s_i}(t)$, $i = 1, \ldots, 4$, located at $C3$, $C4$, $Pz$ and $Fz$ channels, respectively. To simulate $X_s(t)$: First, we simulate four independent AR(2). $Z_{s_1}(t)$ and $Z_{s_2}(t)$ are located at $C3$ and $C4$ channels and have a spectral density with power concentrated at 9 Hz. $Z_{s_3}(t)$ and $Z_{s_4}(t)$ are located at
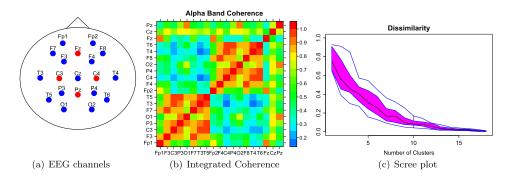
FIG. 6.    (a) *Location of the EEG channels on the scalp* (2*D projection*). *Location of latent signals are highlighted in red.* (b) *Integrated coherence on the alpha band* (*average over the* 100 *replicates*). (c) *Functional boxplot of the scree plot obtained from* 100 *replicates.*

*Pz* and *Fz* channels and have a spectral density with power concentrated at 10 Hz. Then, we compute $X_s(t) = \sum_{i=1}^{4} a(s, s_i) Z_{s_i}(t)$, with $a(s, s_i) = \exp(-\frac{\|s - s_i\|}{\kappa})$, where $\kappa = 1/3$. We consider $T = 1000$ time points and $M = 100$ replicates of this experiment.

We applied the HCC clustering method and the HC (hierarchical clustering with single linkage) to the simulated EEG signals for the alpha band, 8–12 Hz. Figure 6(b) shows the integrated coherence on the alpha band, averaged over the 100 replicates. There are two clusters of highly correlated signals and almost independent among clusters located on different sides of the brain, the left central with the left temporal channels, and the right central with the right temporal channels. We expect that channels located near the same source will be clustered together. We use the scree plot criteria to decide the number of clusters. There is one curve of the minimum dissimilarity value per each replicate. To visualize variability across replicates, Figure 6(c) presents the functional boxplot of the scree plot. This plot suggests the presence of five or six clusters in most of the replicates.

To compare the clustering results, we fix the number of clusters to six and plot the affinity matrices. The affinity matrix $A$ is defined as follows, let $A_{ij}^m = 1$ if channels $i$ and $j$ are in the same cluster at replicate $m$, then $A_{ij} = \sum_{m=1}^{M} A_{ij}^m / M$. Figure 7(a) and (b) show the affinity matrix over 100 replicates for each method. In general, the clustering results are very consistent. The HC method tends to have two large clusters and isolate the channels $O1$, $O2$ and $Cz$. In contrast, the HCC method tends to have six clusters of similar size.

The clustering results between replicates differ little; we consider replicate 85 as a representative example. This replicate corresponds to the median curve in the functional boxplot (see Figure 6(c)). In this case, the HC method assigns $Cz$ to one single cluster. In contrast, the HCC method assigns $\{Cz, Pz, O1, O2\}$ together in one cluster, which seems more reasonable. Another difference is the assignment of $\{F4, F8\}$. The HC method assigns these channels to the same cluster as
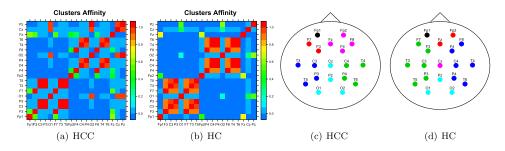
(a) HCC    (b) HC    (c) HCC    (d) HC

FIG. 7. (a) *and* (b) *Affinity matrices over* 100 *replicates with six clusters.* (c) *and* (d) *Cluster locations on the scalp by choosing six clusters of replicate* 85. *Different clusters are represented by colors.*

$\{C4, T4, P4, T6\}$, while the HCC assigns them to the same cluster as $\{Fz, Fp2\}$. These two different clustering results are an example of the advantage of the HCC method, since $\{F4, F8\}$ are strongly correlated with $\{C4, T4\}$ but weakly correlated with $\{P4, T6\}$. Therefore, the within-clusters dependency obtained by using the HCC method are higher than those from the HC method.

4.4. *Effect of the physiological artifacts in coherence-based clustering.* Finally, we explore the possible effect of physiological artifacts in the HCC and HC clustering methods. The presence of physiological artifacts on EEG data is very common. In fact, raw EEG data usually is very noisy, and some preprocessing of the EEG data is needed. Physiological artifacts on EEG data can be produced by eye movement, muscular movement, technical problems, etc. (Viqueira, García Zapirain and Mendez Zorrilla (2013)).

We consider an illustration of this situation with simulation setting similar to Experiment 3. We keep $Z_{s_1}(t)$, $Z_{s_2}(t)$ and $Z_{s_3}(t)$ at the corresponding channels $C3$, $C4$ and $Pz$. We add $Z_{s_4}(t)$, $Z_{s_5}(t)$ located at $FP1$ and $FP2$ channels with spectral density with power concentrated at 5 Hz. The eye blinking artifact shape was generated as the difference between two gamma functions. We simulate one EEG signal with 19 channels as in Experiment 3 and then we contaminate chan-
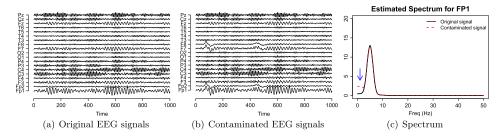


(a) Original EEG signals    (b) Contaminated EEG signals    (c) Spectrum

FIG. 8. *Contamination of EEG signals by the eye blinking artifact.* (a) *and* (b) *EEG signals.* (c) *Estimated spectra of FP1 channel.*

(a) HCC - Original EEG signals

(b) HCC - Contaminated EEG signals

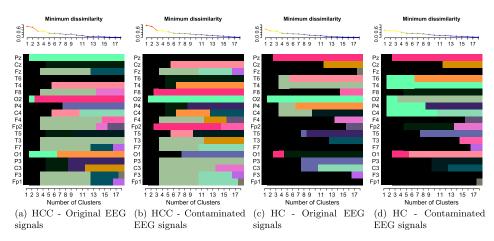(c) HC - Original EEG signals

(d) HC - Contaminated EEG signals

FIG. 9. *Visualization tool of clustering results on theta band with original EEG signals and contaminated signals.*

nels $FP1$, $FP2$, $F7$ and $F8$. Figures 8(a) and (b) show the EEG simulated signals without and with the contamination, respectively. The eye blinking artifact affects lower frequencies due to the high amplitude in less than a second, see Figure 8(c).

We applied the HCC and HC clustering methods in the alpha and theta bands. We observed that there is no significant effect of the eye blinking artifact in the clustering results on the alpha frequency band. However, the theta band clustering is different. Figure 9 shows the clustering results with the original (simulated) signals and the contaminated with eye blinking artifact on the theta band. If we consider five clusters, for both methods, the eye blinking artifact will separate $FP2$ and $F8$ from the rest of the frontal channels. It seems that the presence of the eye blinking artifact could produce its cluster in low frequencies in this case. A more exhaustive simulation study needs to be performed to have a more general conclusion. But, we need to know that artifacts could affect the obtained clusters when using coherence based clustering methods.

**5. Data analysis using the HCC-Vis toolbox.** We developed the HCC method to cluster signals from many channels in a brain network. Then, the identified clusters represent connected brain regions. We studied an EEG recorded from a patient of Dr. Malow (neurologist formerly at the University of Michigan) during an epileptic seizure. The main interest is to identified connectivity on high frequency bands. Therefore, we present the results obtained by the HCC method on the alpha and beta bands. To visualize the clustering results we developed the HCC-Vis, a Shiny app (RStudio) https://carolinaeuan.shinyapps.io/hcc-vis/.

5.1. *Analysis of epileptic EEG seizure.* Dataset corresponds to EEG data recorded from a female patient during spontaneous epileptic seizure. The recording lasted for 500 seconds and was digitized at 100 Hz. The data array has 21
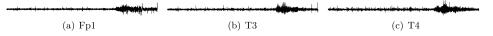
FIG. 10. *EEG traces from channels Fp1, located in the frontal region, and T3 and T4 located in the temporal region (left and right).*

channels with 19 bipolar scalp electrodes placed according to the 10–20 system and two sphenoidal electrodes placed intracranially at the base of the temporal lobe. Figure 10 shows four of the 21 EEG signals. The seizure onset was recorded around 340 seconds. The seizure begins in the left temporal region of the brain, where channel $T3$ was located. Therefore, our analyses used $T3$ as the main focal point (main node). We address the following questions: (1) Which channels are connected to $T3$ during different seizure phases, such as before and after the seizure? (2) Does the connectivity structure between channels differ between the alpha activity and the beta activity? (3) Does seizure modify brain connectivity?

To answer these questions, we divide the EEG recording into disjointed 10-second segments, and applied the HCC method to each segment. Additionally, we show the results from the HC method to compare with the HCC the results, and from the HSM method in the Supplementary Material (Euán, Sun and Ombao (2019b)) to complement the interpretation of the HCC results. Here, we show the results for time segment 5, before seizure (40 to 50 seconds); time segment 35, early seizure (340 to 350 seconds); and time segment 38, middle seizure (370 to 380 seconds).

Figure 11(a) shows the initial screen of the HCC-Vis. We select the three clustering methods to be displayed and time 50 seconds (before seizure) that correspond
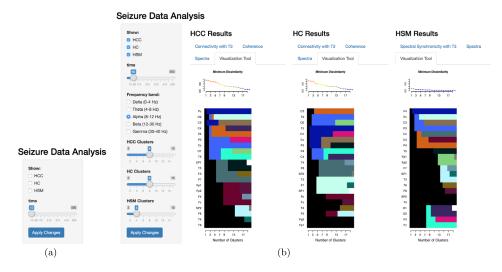


FIG. 11. *HCC-Vis. (a) Initial screen. (b) Visualization Tools for clustering on alpha band.*

TABLE 1
*HCC Clusters for alpha and beta bands*

|       | Before seizure | Early seizure | Middle seizure |
|-------|----------------|---------------|----------------|
| Alpha band | {O2}, {F7, F3, T3, SP1}, {P4, C4}, {T6, T4, F8, SP2, Fz, F4, Fp2, Fp1}, {Pz, C3}, {Cz, P3, T5, O1} | {C3}, {P4}, {Pz}, {T6, T4, F8, SP2}, {C4, F4}, {T5, T3, F7, SP1, Fz,F3,Fp2, Fp1}, {Cz, P3, O2, O1} | {C4}, {P4}, {SP1}, {P3, C3}, {T6, SP2, O2, O1, Pz, F3, Fz, F4, T3, F7, Cz, T5, Fp2, Fp1, T4, F8} |
| Beta band | {P4}, {F8, Fp2, T4, SP2}, {Cz, P3}, {Pz, Fz}, {F7, Fp1, T3, SP1, T6, O2, T5, O1}, {C3, F3, C4, F4} | {C3}, {O2, P4, T5, O1}, {F8, Fp2, T4, SP2}, {Cz, P3}, {Pz, F3}, {Fz, F4} {F7, Fp1, T3, SP1}, {T6, C4} | {C4, C3}, {F3, Fp1, F7}, {T3, SP1}, {F4, Fp2, F8} {T6, O2, Pz, Fz, Cz, T5, O1, P3, P4}, {T4, SP2} |

to before seizure scenario. Then, we select the alpha band to execute the clustering methods. Figure 11(b) shows the visualization tools to decide the number of clusters and to observe the clustering dynamics between different clustering methods. We repeat this for early seizure (340 to 350 seconds) and middle seizure (370 to 380 seconds) on alpha and beta bands. In general, the clustering results from different methods, scenarios and frequency bands differ in some cases but are similar in others. The clustering dynamics between the HCC method and the HC method are different for all of the channels. HSM method identifies fewer clusters which means that many channels have similar spectral densities, that is, they are spectrally synchronized. Our goal is to identify connectivity based on coherence so we will focus on the HCC clustering results.

The scree plot for each time segment suggests five to seven possible clusters. We apply the bootstrap test to select the most significant clustering setting in this range. Take the clustering results before the seizure on the alpha band as an example. First, we apply the bootstrap test for $H_0$: 7 clusters vs. $H_A$: 6 clusters. In this case, the $p$-value was close to zero which means that we have strong evidence to merge two out of the seven clusters. Then, we proceed to test $H_0$: 6 clusters vs. $H_A$: 5 clusters. The observed $p$-value was almost equal to 1 and we do not have enough evidence to merge two out of the six clusters. Thus, we accept the six-cluster setting. Using the same testing procedure, we find strong evidence to select seven clusters on the alpha band in the early seizure, and five clusters for the middle seizure. In the beta band, our tests suggest six, eight and six clusters for the before, early and middle seizure scenarios, respectively. Table 1 lists the corresponding clusters. In general, there is no easy way to label or color the changing clusters clearly, because two clusters from different time segments do not necessarily contain identical members even though the total number of the clusters is the same. Therefore, to illustrate the time evolving clustering, we have chosen one cluster that contains T3 to avoid the labeling issue. Now, we show the clustering

(a) Before Seizure - Alpha band
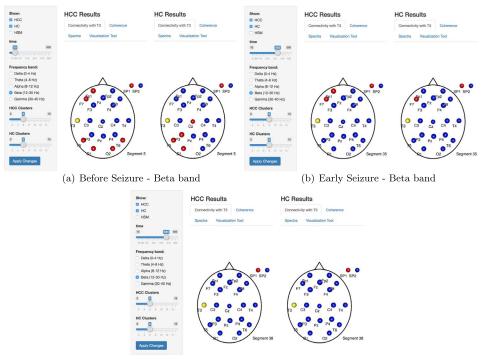
(b) Early Seizure - Alpha band

(c) Middle Seizure - Alpha band

FIG. 12. *Connectivity with $T3$ based on coherence on the alpha band. In each case, the HCC results are on the left side and the HC results are on the right side. This analysis was produced using the HCC-Vis (*[https://carolinaeuan.shinyapps.io/hcc-vis/](https://carolinaeuan.shinyapps.io/hcc-vis/)*).*

results from $T3$ in the following sense. The yellow dot in each figure represents $T3$. If a channel belongs to the same cluster with $T3$, it is represented in red; otherwise, it is in blue.

Figures 12 and 13 show clustering based on coherence of $T3$ and other channels, on alpha and beta bands. Clustering results from the HCC method on the alpha band are different depending on the scenario. Connectivity before seizure was between $T3$ and $F7$, $F3$ and $SP1$, which is reasonable since they are located closely on the scalp. After the seizure started, more channels have a higher tendency to belong to the same cluster at $T3$. As a result, the left frontal and temporal regions become a single highly correlated cluster. In the middle seizure, the connectivity of $T3$ expands to more channels located in the frontal region and temporal region. If we consider HC, the results in the first scenario are similar to HCC results.

The clustering results for the beta band are more similar between the two methods, HCC and HC. The results also suggest that there is a strong dependence between the channel $T3$ and $SP1$ on the beta band. $F7$ also shows connectivity at this amplitude. However, this connectivity vanishes after the seizure. The differences between the clustering results on different frequency bands could be related to the

(a) Before Seizure - Beta band

(b) Early Seizure - Beta band



(c) Middle Seizure - Beta band

FIG. 13. *Connectivity with T3 based on coherence in the beta band. In each case, the HCC results are on the left side and the HC results are on the right side. These results were produced using the proposed HCC-Vis.*

spectral profile in each case, clustering results obtained by using the HSM method (see Euán, Sun and Ombao (2019b)) confirm that many of the channels show activity in same frequencies. In particular, we identify that $T3$ is not the only channel with energy in high frequency bands, alpha and beta. The HSM method and additional plots to help the interpretation are available on the HCC-Vis (see Euán, Sun and Ombao (2019b)). To conclude, HCC clusters change between frequency bands and evolve under different scenarios. These suggest that the brain network changes between frequency bands and during different stages of the seizure episode. Also, we identify many channels that showed activity in higher frequency bands as $T3$, but they are not necessarily correlated with $T3$.

**6. Discussion and conclusions.** The main contribution of this paper is the hierarchical coherence clustering (HCC) method which uses the notion of a cluster-based coherence rather than coherence between a pair of univariate time series. The simulation studies showed the advantages of the HCC method compared to commonly used clustering methods. In cases where the clusters are independent,

the HCC method performs as well as the other clustering methods. In the presence of noisy signals, for example, in EEG data, the HCC method performs better. We studied the properties of cluster coherence when the spectral matrix and the number of clusters are known. We also proposed a bootstrap test for detecting significantly independent clusters and investigated the power of the test based on simulations. Under independence or moderate dependence level, the simulation studies demonstrated that the proposed method has a good power to detect the significant clusters.

We applied the HCC method to EEG data recorded during an epileptic seizure on the alpha and beta bands. The clustering results show that many channels were active during the seizure but not all of them were correlated. Also, by considering the changes in the connectivity before and after clustering, we found that the seizure possibly affected the connectivity on the alpha band.

Cluster coherence measures the similarity between clusters by taking into account both the within-cluster structure and between clusters via the eigenvalues. One limitation of this method is the computational time required to compute the eigenvalues when the size of the clusters increases. A truncated version of the cluster coherence could be used to speed up these computations; in this case, a fast algorithm could be used to compute only the $p$ first eigenvalues.

## SUPPLEMENTARY MATERIAL

**Supplement A: Scalability of the HCC Method** (DOI: 10.1214/18-AOAS1225SUPPA; .pdf). We provide additional supporting plots that show the scalability of the HCC method to a larger number of time series and compare to HAC method. The results reflects that average coherence could overestimate the within cluster correlation. In contrast, cluster coherence measures more reasonably the within-cluster dependency.

**Supplement B: HCC-Vis toolbox** (DOI: 10.1214/18-AOAS1225SUPPB; .pdf). We developed the HCC-Vis to visualize the clustering results. This supplementary material shows some additional plots to enhance the potential of the visualization tools.

## REFERENCES

BABADI, B. and BROWN, E. N. (2014). A review of multitaper spectral analysis. *IEEE Trans. Biomed. Eng.* **61** 1555–1564.

BOWYER, S. M. (2016). Coherence a measure of the brain networks: Past and present. *Neuropsychiatric Electrophysiology* **2** 1.

BRILLINGER, D. R. (1975). *Time Series*: *Data Analysis and Theory*. *International Series in Decision Processes*. Holt, Rinehart and Winston, New York. MR0443257

BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series*: *Theory and Methods*. *Springer Series in Statistics*. Springer, New York. Reprint of the second (1991) edition. MR2839251

BUZSÁKI, G. and DRAGUHN, A. (2004). Neuronal oscillations in cortical networks. *Science* **304** 1926–1929.

CAIADO, J., CRATO, N. and PEÑA, D. (2006). A periodogram-based metric for time series classification. *Comput. Statist. Data Anal.* **50** 2668–2684. MR2227325

CAIADO, J., MAHARAJ, E. A. and D'URSO, P. (2016). Time-series clustering. In *Handbook of Cluster Analysis. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 241–263. CRC Press, Boca Raton, FL. MR3644715

CRIBBEN, I. and YU, Y. (2017). Estimating whole-brain dynamics by using spectral clustering. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 607–627. MR3632344

D'URSO, P., DE GIOVANNI, L. and MASSARI, R. (2016). GARCH-based robust clustering of time series. *Fuzzy Sets and Systems* **305** 1–28. MR3557847

D'URSO, P. and MAHARAJ, E. A. (2009). Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems* **160** 3565–3589. MR2563306

D'URSO, P. and MAHARAJ, E. A. (2012). Wavelets-based clustering of multivariate time series. *Fuzzy Sets and Systems* **193** 33–61. MR2880740

DECKER, S., FILLMORE, P. T. and COHERENCE, A. R. (2017). The measurement and application of brain connectivity. *NeuroRegulation* **4** 3–13.

EUÁN, C., OMBAO, H. and ORTEGA, J. (2018). The hierarchical spectral merger algorithm: A new time series clustering procedure. *J. Classification* **35** 71–99. MR3790113

EUÁN, C., SUN, Y. and OMBAO, H. (2019a). Supplement to "Coherence-based time series clustering for statistical inference and visualization of brain connectivity." DOI:10.1214/18-AOAS1225SUPPA.

EUÁN, C., SUN, Y. and OMBAO, H. (2019b). Supplement to "Coherence-based time series clustering for statistical inference and visualization of brain connectivity." DOI:10.1214/18-AOAS1225SUPPB.

FIECAS, M. and OMBAO, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *Ann. Appl. Stat.* **5** 1102–1125. MR2840188

FIECAS, M., OMBAO, H., LINKLETTER, C., THOMPSON, W. and SANES, J. (2010). Functional connectivity: Shrinkage estimation and randomization test. *NeuroImage* **49** 3901–3915.

GAO, X., SHAHBABA, B., FORTIN, N. and OMBAO, H. (2016). Evolutionary state-space model and its application to time-frequency analysis of local field potentials. Available at arXiv:1610.07271.

KIMES, P. K., LIU, Y., HAYES, D. N. and MARRON, J. S. (2017). Statistical significance for hierarchical clustering. *Biometrics* **73** 811–821. MR3713115

LAFUENTE-REGO, B., D'URSO, P. and VILAR, J. A. (2018). Robust fuzzy clustering based on quantile autocovariances. *Statist. Papers* 1–56.

MAHARAJ, E. A. and D'URSO, P. (2010). A coherence-based approach for the pattern recognition of time series. *Phys. A* **389** 3516–3537. MR2659314

MAHARAJ, E. A. and D'URSO, P. (2011). Fuzzy clustering of time series in the frequency domain. *Inform. Sci.* **181** 1187–1211.

MAHARAJ, E. A., D'URSO, P. and GALAGEDERA, D. U. A. (2010). Wavelet-based fuzzy clustering of time series. *J. Classification* **27** 231–275. MR2726320

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. Cybernet.* **104** 197–207. MR2795598

OMBAO, H. and VAN BELLEGEM, S. (2008). Evolutionary coherence of nonstationary signals. *IEEE Trans. Signal Process.* **56** 2259–2266. MR2516630

OMBAO, H. C., RAZ, J. A., STRAWDERMAN, R. L. and VON SACHS, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika* **88** 1186–1192. MR1872229

PAWITAN, Y. (1996). Automatic estimation of the cross-spectrum of a bivariate time series. *Biometrika* **83** 419–432. MR1439793

RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* **66** 846–850.

ROSEN, O. and STOFFER, D. S. (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika* **94** 335–345. MR2331489

SHUMWAY, R. H. and STOFFER, D. S. (2011). *Time Series Analysis and Its Applications. With R Examples*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2721825

TAKAHASHI, D. Y., BACCALÁ, L. A. and SAMESHIMA, K. (2014). Canonical information flow decomposition among neural structure subsets. *Frontiers in Neuroinformatics* **8** 49.

TIBSHIRANI, R., WALTHER, G. and HASTIE, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. MR1841503

VIQUEIRA, M., GARCÍA ZAPIRAIN, B. and MENDEZ ZORRILLA, A. (2013). Ocular movement and cardiac rhythm control using eeg techniques. In *Medical Imaging in Clinical Practice* (O. F. Erondu, ed.), Chapter 6. InTech, Rijeka.

WALDEN, A. T. (2000). A unified view of multitaper multivariate spectral estimation. *Biometrika* **87** 767–788. MR1813974

CEMSE DIVISION
KING ABDULLAH UNIVERSITY OF SCIENCE
  AND TECHNOLOGY (KAUST)
THUWAL 23955
SAUDI ARABIA
E-MAIL: carolina.euancampos@kaust.edu.sa