Supplement A: Scalability of the HCC Method

Carolina Euan, Ying Sun and Hernando Ombao

September 11, 2018

Supplement of *Coherence-based Time Series Clustering for Brain Connectivity Visualization*

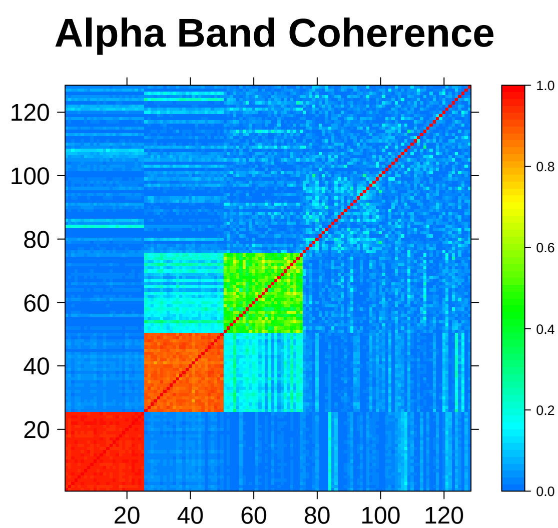
Since we are motivated by the study of EEG signals, it is important to explore the scalability of our method. In other words, which is the performance of the HCC method compared to HAC if we have a larger number of time series or EEG channels. We performed a similar experiment to **Experiment 2**, but with larger number of time series. We consider the following example with 128 time series which is a usual size of a multichannel EEG.

Experiment. Let $Z_j(t)$ be an AR(2) latent process for $j = 1, \ldots, 5$, with the unimodal spectral density concentrated around $2, 6, 10, 15$ and 40 Hz, respectively. Each $Z_j(t)$ represents a latent process in the different frequency bands (delta, theta, alpha, beta, and gamma). $\mathbf{X}(t)$ is a 128-multivariate time series generated by

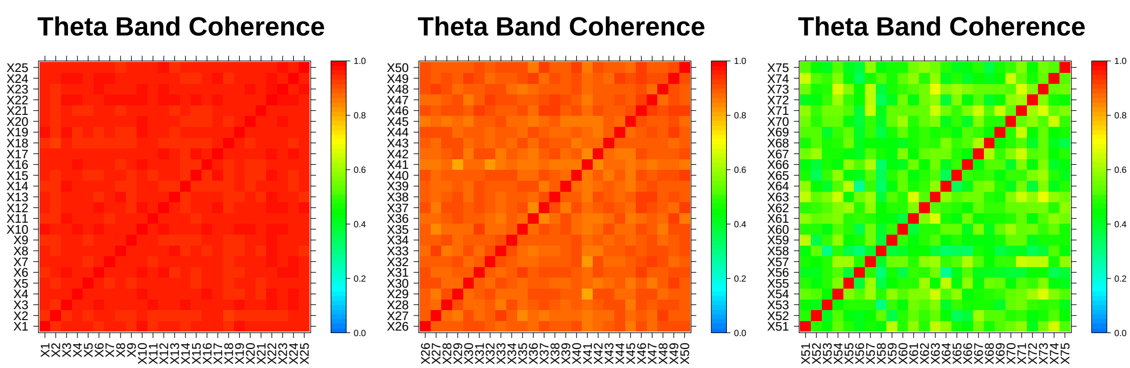
$$\mathbf{X}(t) = \mathbf{A}\mathbf{Z}(t) + \boldsymbol{\varepsilon}(t),$$

where $\mathbf{Z}(t) = (Z_1(t), \ldots, Z_5(t))^T$ and $\boldsymbol{\varepsilon}(t)$ is white noise. The rows of the coefficient matrix $\mathbf{A}$ are built as follows: $A_{l.} = c(1, .2, 0, 0, 0)$ for $l = 1, \ldots, 25$, $A_{l.} = c(0, 1, 0, 0, 0)$ for $l = 26, \ldots, 50$, $A_{l.} = c(0, .2, 1, 0, 0)$ for $l = 51, \ldots, 75$, $A_{l.} = c(0, 0, 0, 1, 0)$ for $l = 76, \ldots, 100$ and $A_{l.} = c(0, 0, 0, 0, 1)$ for $l = 101, \ldots, 128$.

We simulate time series of length $T = 1000$ with a sampling frequency of 100 Hz. Figure 1 shows the estimated integrated coherence on beta band. On beta band, there are two highly correlated clusters $C_1 = \{X_1(t), \ldots, X_{25}(t)\}$ and $C_2 = \{X_{26}(t), \ldots, X_{50}(t)\}$. There is a third cluster $C_3 = \{X_{51}(t), \ldots, X_{75}(t)\}$ which has lower within correlation. Figure 2 shows the clustering result for experiment 3. Overall, the two methods show similar results. These three main clusters were recovered by the HCC and HAC methods. One difference is that if we underestimate the number of clusters, the HAC method will merge (when $k = 37$) C_3 and C_2 before the HCC method (when $k = 23$). This reflects that average coherence could overestimate the within cluster correlation. In contrast, cluster coherence measures more reasonable the within-cluster. Here, we show one simulation of this experiments but the clustering results were very consistent through different replicates, similar to Experiment 2.

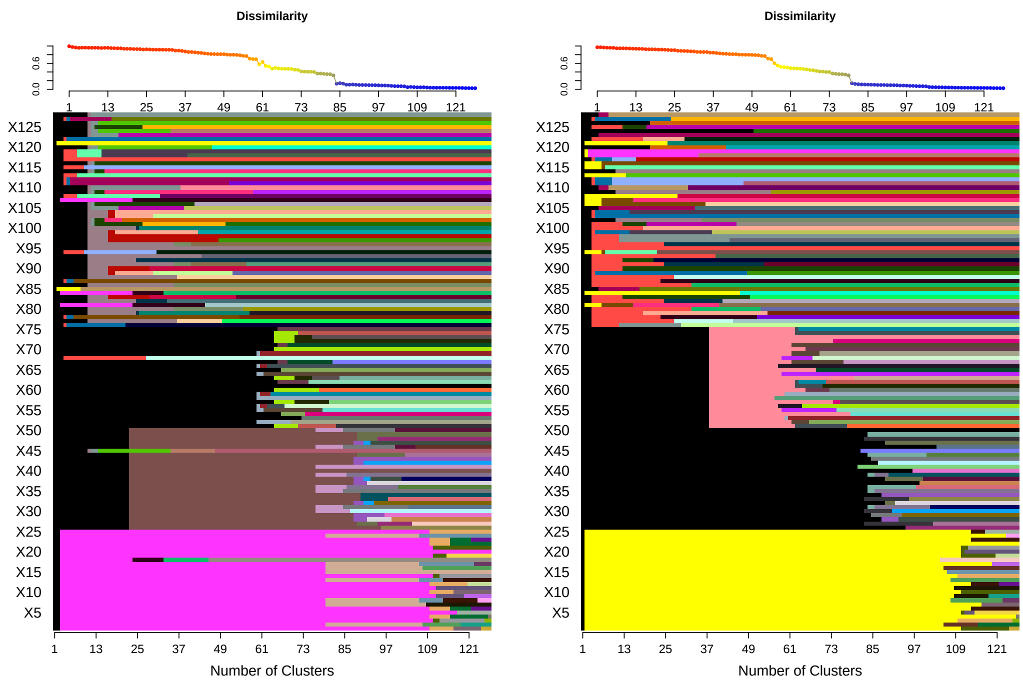


(a) Coherence matrix



(b) Coherence submatrix

Figure 1: Estimated coherence of one simulation of Experiment 3.



(a) HCC (b) HAC

Figure 2: Clustering results for Experiment 3. Each figure represents the clustering results with different methods; the upper plot shows the minimum dissimilarity value and the lower plot represents which signals belong to same cluster using same colors.