Supplementary Results for the Manuscript: Inferring Brain Signals Synchronicity from a Sample of EEG Readings

Qian Li[1*]
Damla Şentürk[1,2],
Catherine A. Sugar[1,2,3],
Shafali Jeste[3],
Charlotte DiStefano[3],
Joel Frohlich[3],
and
Donatello Telesca[1,*].
[1] Department of Biostatistics, University of California, Los Angeles
[2] Department of Statistics, University of California, Los Angeles
[3] Department of Psychiatry and Biobehavioral Sciences,
University of California, Los Angeles
[*] *email: dtelesca@ucla.edu*

*This work was supported by the grant R01 GM111378-01A1 (DS, DT, CS) from the National Institute of General Medical Sciences.

1 Introduction

This document accompanies our main article and provides a detailed discussion of technical points. In particular, Section **??** extends on some technical details of dimensional reduction, estimation and inference procedures: the spectral density estimator; step-by-step Gibbs sampler used for posterior inference and information criteria that were considered for model selection under the context of MIC. Section **??** establishes the ground for our simulation studies, how the EEG signals were engineered, based on which we further investigate the cases of non-stationarity, varying sample sizes and Signal-to-Noise Ratios (SNRs). We conclude with an extended report of our case study, covering EEG preprocessing and quality examination steps, post analysis of MIC clusters on spectral domain and an assessment of MCMC convergence, in Section **??**. The implementation of MIC is fulfilled in R and Rcpp, which is publicly available on Github (`https://github.com/Qian-Li/MIC2`)

2 Estimation and Model Selection Procedures

2.1 Spectral Density Estimation

Let $\{Z_\tau,\ \tau = 0, \pm 1, \pm 2, \ldots\}$ be a zero mean, weakly stationary time series, with autocovariance $\mathcal{C}_Z(h) = E(Z_\tau, Z_{\tau+h})$, $(h = 0, \pm 1, \pm 2, \ldots)$. We define the spectral density function $\phi_Z(\omega)$ of Z_τ as in [**?**], so that:

$$\phi_Z(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \mathcal{C}_Z(\tau) \exp(-i\tau\omega), \ \ \omega \in [0, \pi].$$

Let $\widehat{\mathcal{C}}_Z(\tau) = 1/(N-h)\sum_{h+1}^{N} x(t) \cdot x(t-h)$ be the sample autocovariance function. Note that $\mathcal{C}_Z(\cdot)$ is symmetric around 0, therefore it suffices to define the one-sided estimator for $h \in \{0, 1, 2, \ldots\}$. By introducing a smoothing kernel $w(u)$, and a truncation parameter $a > 0$, we obtain a consistent estimator of the spectral density function at Fourier frequency ξ as follows:

$$\hat{\phi}(\xi) = \sum_{|h| \le a} w(h/a)\widehat{\mathcal{C}_Z}(h) \exp(-i2\pi\xi h), \ \ \xi \in [0, 1/2]. \tag{1}$$

And here we use the Parzen window with bandwidth a for smoothing,

$$w(u) = \begin{cases} 1 - 6|u|^2 + 6|u|^3, & \text{if } |u| < \frac{1}{2} \\ 2(1-|u|)^3, & \text{if } \frac{1}{2} \le |u| \le 1 \\ 0, & \text{otherwise.} \end{cases} .$$

The estimator in (**??**) is a natural non-parametric estimator of the spectral density, which has been well recognized in both statistics and engineering literatures ([**?**], [**?**]). It smoothes the periodogram by means of both the Parzen kernel and a pre-determined bandwidth a. **?** suggests data-driven bandwidth selection based on generalized cross-validations, however, we proceed with the maximal bandwidth $a = N - 1$ when estimating spectral densities for each segmented EEG signals. The considerations come in two-fold: bias weigh more than

inconsistency when estimating spectral densities on short time series; variance reduction is feasible afterwards by considering the Welch's method ([**?**]).

To be precise under the setting of EEG studies, we consider an ensemble of time-series segments denoted with $Y_{ij}(\boldsymbol{s}_i)$. Therefore for subject i ($i = 1, \ldots, n$), at electrode j ($j = 1, \ldots, p$), we obtain EEG signals of every 1024ms on a common set of segments $\boldsymbol{s}_i = \{s \colon 1, \ldots, q_i\}$ which also denotes time under the context of ongoing experiments. Pipelined data quality control, including artifacts removal and rejections, proceeds linear de-trending and tapering of 10% at both ends on segment level. Thereafter, we obtain spectral density estimates as in (**??**), on subject specific set of segments $\boldsymbol{s}_i$, and denote with $\hat{\phi}_{ij}(\omega, s)$, $s \in \boldsymbol{s}_i$. Finally, as we operate over epochs composed of γ segments, the adjacent epochs overlay a percentage of δ with each other. For instance, the i-th epoch ($t = i$) averages segments within the range of $[(i-1)(1-\delta)\gamma + 1, (i-1)(1-\delta)\gamma + \gamma]$ until its upper bound run out of $\boldsymbol{s}_i$, as is depicted in Fig **??**. More formally,

$$\hat{\phi}_{ij}(\omega, t) = \frac{1}{\gamma} \sum_{s=(t-1)(1-\delta)\gamma+1}^{(t-1)(1-\delta)\gamma+\gamma} \hat{\phi}_{ij}(\omega, s) \cdot \mathbb{1}(s^{(u)} \in \boldsymbol{s}_i),$$

where $s^{(u)}$ indicates the upper bound of the smoothing window, i.e. $s^{(u)} = (i-1)(1-\delta)\gamma + \gamma$. The proposed epoch spectral estimator serves a moving average of the segment level periodograms, a special case of the Welch's reduced variance estimator.

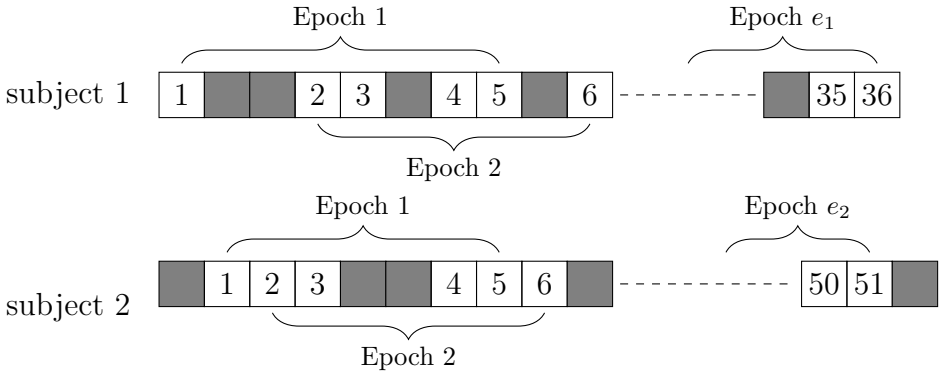

Figure 1: **Epoch smoothing on spectral estimates**: shaded blocks indicate rejected segments by quality control, which cannot be located on the original experimental time scale; segments (as numbered blocks) are averaged within a window of $\gamma = 5$ that overlaps $\delta = 4/5$,

2.2 Gibbs Sampler

The proposed MIC model is implemented under a fully Bayesian framework, with all the model parameters and latent variables sampled from a Gibbs sampler. By default, the algorithm is initialized with a K-groups partition using the `kmeans` of RcppArmadillo, then it proceeds hierarchically upward and samples sequentially from either full conditional or marginal posterior distributions. At the m-th iteration, the sampling scheme is detailed as below:

- $\boldsymbol{\Theta}_i(t)|\mathbb{X}_i(t), \mathbb{L}_i(t) \sim f(\boldsymbol{\theta}_{ik}(t)|\mathbb{X}_i(t), \mathbb{L}_i(t) = k)$ for $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T_i\}$.

  Specifically, the posterior distribution for $\boldsymbol{\theta}_{ik}^{(m)}(t)$ is conjugate:

$$\boldsymbol{\theta}_{ik}^{(m)}(t) \sim \text{NIW}(\boldsymbol{\mu}_{ik}^{(m)}(t), \boldsymbol{\lambda}_{ik}^{(m)}(t), \sigma_{ik}^{2\ (m)}(t), \nu_{ik}^{(m)}(t)|\mathbb{X}_i(t))$$

  And these quantities are first updated based on the current cluster assignment, $\mathbb{L}_i^{(m-1)}(t)$.

- $\mathbb{L}_i(t)|\mathbb{X}_i(t), \boldsymbol{\theta}_i(t), \boldsymbol{\beta}_i, \alpha_i, \Pi \sim \prod_j \Pr(k|X_{ij}(t), \boldsymbol{\theta}_{ij}(t), \boldsymbol{\beta}_i, \alpha_i, \Pi)$ for $j = 1, \ldots, p$ and $k = 1, \ldots, K$ where marginal distribution of $\mathbb{L}_i(t)$ is derived by integrating out latent labels of $\mathbb{C}_i$ and $\mathbb{S}$,

$$\Pr(k|X_{ij}(t), \boldsymbol{\theta}_{ij}(t), \beta_{ij}, \alpha_i, \Pi) \propto \pi(k, \Pi^{(m-1)}, \beta_{ij}^{(m-1)}, \alpha_i^{(m-1)}) f(X_{ij}(t)|\boldsymbol{\theta}_{ik}^{(m)}(t)),$$
$$\pi(k, \Pi^{(m-1)}, \boldsymbol{\beta}_i^{(m-1)}, \alpha_i^{(m-1)}) = \Pr(L_{ij}(t) = k|\Pi^{(m-1)}, \beta_{ij}^{(m-1)}, \alpha_i^{(m-1)})$$

  And $L_{ij}^{(m)}(t)$ will be sampled as $k \in \{1, \ldots, K\}$ with the probability proportional to the above expression.

- $\mathbb{C}_i|\mathbb{L}_i(t), \alpha_i, \boldsymbol{\beta}_i, \Pi \sim \prod_j \Pr(k|L_{ij}(t), \ldots, L_{ij}(T_i), \Pi, \alpha_i, \boldsymbol{\beta}_i)$, where marginal distribution of $\mathbb{C}_i$ is derived by integrating out latent labels $\mathbb{S}$,

$$\Pr(k|L_{ij}(1), \ldots, L_{ij}(T_i), \Pi, \alpha_i, \boldsymbol{\beta}_i) \propto \pi(k, \Pi^{(m-1)}, \alpha_i) \left(\prod_{t=1}^{T_i} \nu_e(L_{ij}(t)^{(m)}, k, \beta_i^{(m-1)}(t)) \right),$$
$$\pi(k, \Pi^{(m-1)}, \alpha_i) = \Pr(C_{ij} = k|\Pi^{(m-1)}, \alpha_i^{(m-1)})$$

- $\beta_i(t)|\mathbb{C}_i, \mathbb{L}_i(t) \sim \text{TBeta}(c_i + \tau_i, d_i + pT_i - \tau_i(t), 1/K)$ where $\tau_i(t)$ is the number of samples $(L_{ij}(t) = C_{ij})$ for $t \in \{1, \ldots, T_i\}$ and $j \in \{1, \ldots, p\}$. Also $\mathbb{C}_i = \mathbb{C}_i^{(m)}$ and $\mathbb{L}_i(t) = \mathbb{L}_i^{(m)}(t)$.

- $S_j|\Pi, C_{1j}, \ldots, C_{nj}, \alpha_i \sim \Pr(k|C_{1j}, \ldots, C_{nj}, \Pi, \alpha_i)$, where

$$\Pr(k|C_{1j}^{(m)}, \ldots, C_{nj}^{(m)}, \Pi^{(m-1)}, \alpha_i^{(m-1)}) \propto \pi_k^{(m-1)} \prod_{i=1}^{n} \nu_s(k, C_{ij}^{(m)}, \alpha_i^{(m-1)}).$$

- $\alpha_i|\mathbb{C}_i, \mathbb{S} \sim \text{TBeta}(a_i + \psi_i, b_i + p - \psi_i, 1/K)$ where ψ_i is the number of electrodes $(C_{ij}^{(m)} = S_j^{(m)})$ for $j \in \{1, \ldots, p\}$.

- $\Pi|\mathbb{S} \sim \text{Dirichlet}(\boldsymbol{\eta} + \boldsymbol{\rho})$ where ρ_k is the number of samples clustered k in $\mathbb{S}^{(m)}$ and $\eta_k = 3$ is chosen as a priori.

Markov chain Monte Carlo (MCMC) proceeds by hierarchically iterating through the posterior distributions, discarding 1/5 of the posterior samples for burn-in and keeping the rest for the posterior calculations. Parameter estimates and credible intervals are extracted and summarized, after which the information criteria are calculated to select the optimal K across models.

2.3 Model Selection and Information Criteria

In this section, we expatiate on the appropriate information criteria serving the purpose of model selection. We also report the details to the calculation of a rescaled coherence, $\bar{\alpha}^*$. To certain degree, this self-contained metric reflects the individual and group level fitting and stays robust to the input dimensionality, therefore it could be used for model assessment from a complementary perspective and across dimensions. The `dksearch` algorithm of `MIC2` package is steered by the BIC at a fixed dimension d, but reports all the information criteria and the adjusted coherence as output along searching trajectory.

- The BIC (**?**) is defined as

$$\mathrm{BIC}(K) = 2\log f(y|\widehat{\tau}, K) - d\log(n),$$

  where d is the number of free parameters.

  In our model:

$$\mathrm{BIC}(K) = \sum_i \sum_t^{T_i} 2\log f(\mathbb{X}_i(t)|(\widehat{\boldsymbol{\theta}}_i(t), \widehat{\mathbb{L}}_i(t)), K) - d_f \log(P),$$

  where d_f is the number of free parameters in the model,

$$d_f = 2K \cdot d \cdot N_{epoch} + N_{epoch} + n + K - 1,$$

  and d is the dimensionality of the eigen-Laplacian data.

- The DIC (**?**) uses *effective* model parameters p_d instead, which has a general form of,

$$\mathrm{DIC}(K) = -2\log p(y|\widehat{\tau}, K) + 2p_d,$$

$$\text{where } p_d = \mathbb{E}_{\tau|y}(\log p(y|\widehat{\tau})) - \log p(y|\widehat{\tau}).$$

  [**?**] suggested a form of DIC (as DIC_4 in their paper), which treats the labels $\mathbb{L}$ as missing data thus taking expectation with respect to them afterwards. The formal definition of DIC in our model is,

$$\begin{aligned}\mathrm{DIC}(K) &= \sum_i \sum_t^{T_i} \{\mathbb{E}_{\mathbb{L}_i(t)l|\mathbb{X}_i(t)}[\mathrm{DIC}(\mathbb{X}_i(t), \mathbb{L}_i(t), K)]\} \\ &= \sum_i \sum_t^{T_i} \{-4\,\mathbb{E}_{\boldsymbol{\theta}, \mathbb{L}_i(t)|\mathbb{X}_i(t)}[\log f(\mathbb{X}_i(t), \mathbb{L}_i(t)|\boldsymbol{\theta}_i(t))] \\ &\quad + 2\,\mathbb{E}_{\mathbb{L}_i(t)|\mathbb{X}_i(t)}[\log f(\mathbb{X}_i(t), \mathbf{L}|\widehat{\boldsymbol{\theta}}_i(t))]\}\end{aligned}$$

- The adjusted coherence (**?**) was suggested for choosing K,

$$\alpha_i^* = \frac{K \cdot \alpha_i - 1}{K - 1}; \quad \bar{\alpha}^* = \frac{K \cdot \bar{\alpha} - 1}{K - 1},$$

Because coherence parameters are restricted as $\alpha_i \in (1/K, 1)$ in our formulation, we normalize K-dependent measures so that α_i^* reflects adherence to group mean on a common range of $(0, 1)$. In our model, α_i quantify the level of agreement between the individual and group consensus labels, which complements the BIC and DIC that are defined on the epoch level. As pointed out by [**?**], this measure tends to favors small K for more coherent partitions between individuals and their consensus, therefore we only suggest its usage for informal assessment across dimensions but recommend either BIC or DIC as a formal device.

3 Extended Simulation Study

3.1 Simulating Band-Oscillating EEG Signals

EEG oscillations are often characterized by means of frequency bands, namely: delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz) and gamma (30-50 Hz). Following [**?**], we use a linear mixture of second order auto-regressive (AR(2)) processes to simulate signals that would feature certain desired properties in their power spectral densities. To be specific, an AR(2) process Z_t is defined as,

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \epsilon_t$$

where ϵ_t is a Gaussian white noise process. The oscillation properties of the procedure can be reflected by its characteristic polynomial, $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$. The root to this equation, denoted as $z_0^{(1)}$ and $z_0^{(2)}$, admit a polar representation that directly relates to the features in the frequency domain,

$$|z_0^{(1)}| = |z_0^{(2)}| = M, \qquad arg(z_0^{(\cdot)}) = \frac{2\pi\eta}{F_s},$$

where F_s is the sampling frequency, M is the amplitude or magnitude of the root ($M > 1$) and η is the frequency index. The spectrum of the AR(2) process with polynomial roots above will have peak frequency at η, and it gets broader as $M \to \infty$ and narrower as $M \to 1^+$. That is to say, with a predetermined set of spectral characteristics (F_s, η, M), one can simulate a finate sample of the corresponding AR(2) process parameterized by coefficients (ϕ_1, ϕ_2).

In our experiment, $M = 1.03$ and $F_s = 200$Hz are fixed and 5 peak frequencies are considered for linear mixture. Let Z_t^m be the m-th AR(2) process ,$m = 1, \ldots, 5$, that has single peak frequency at $\eta = 2, 6, 10, 21$ and 40 Hz respectively. We linearly mix up the simulated processes by a set of pre-determined weights $\mathbf{e}_{L(j)}$, that is specific to electrode j, therefore electrodes sharing the same weights vector have similar spectral profiles and should be grouped together. To exemplify,

$$X_{jt} = \mathbf{e}_{L(j)}^T \cdot \mathbf{Z}_{jt} + \epsilon_{jt}, \qquad \mathbf{Z}_{jt} = (Z_{jt}^1, Z_{jt}^2, Z_{jt}^3, Z_{jt}^4, Z_{jt}^5)^T,$$

where ϵ_{jt} is a white noise process independent from the true signal $\mathbf{e}_{L(j)}^T \cdot \mathbf{Z}_{jt}$. We simulated K groups of signal patterns, labeled as $L(j) \in \{1, \ldots, K\}$, by taking a linear combination of

five AR(2) processes of weights $\mathbf{e}_{L(j)}$. To be more specific, 4 groups of synchronized patterns ($K = 4$) are generated by considering weights,

$$e_1 = (1, 2, 0, 0, 0)^T,\ e_2 = (0, 1, 2, 0, 0)^T,\ e_3 = (0, 0, 1, 1, 0)^T,\ e_4 = (0, 0, 0, 1, 1)^T$$

3.2 Simulating Non-stationary EEGs

To assess the robustness of proposed procedure to the violations of stationarity, we construct a piecewise stationary process that randomly alternates between a *main* and an *off* states. Meanwhile, the labels stays the same before and after the transitions, such that linear weights bijectively match to its alternative within the same cluster. To be specific, the two sets of AR(2) mixing weights, $\mathbf{e}_{L(j)}$ and $\mathbf{e}'_{L(j)}$ characterizing the *main* and *off* states, are chosen to be,

$$\begin{aligned}
e_1 = (1, 2, 0, 0, 0)^T &\longleftrightarrow e'_1 = (1, 0, 0, 2, 0)^T \\
e_2 = (0, 1, 2, 0, 0)^T &\longleftrightarrow e'_2 = (0, 0, 2, 1, 0)^T \\
e_3 = (0, 0, 1, 1, 0)^T &\longleftrightarrow e'_3 = (0, 0, 1, 2, 0)^T \\
e_4 = (0, 0, 0, 1, 1)^T &\longleftrightarrow e'_4 = (0, 0, 0, 1, 2)^T
\end{aligned}$$

3.3 Extended Results

We evaluate the proposed procedure against multiple epoch window sizes of $\gamma = \{4, 6, 8, 10\}$ with an fixed amount of overlap of 50%, on a sample of 10 subjects each with 50 segments of simulated signals. Results reported here are based on a Monte Carlo study of 100 repeated datasets, from the perspective of α estimates and clustering accuracy relative to the known truth. Fig **??** demonstrates the results of both the cases: α estimates are stable and accurate over varying smoothing settings (Fig **??**(a)), clustering accuracy is reliable on the individual level regardless of its coherence to the group and the group labels are recovered accurately even when the sample coherence is low (Fig **??**(b)).

Furthermore, we extend the simulation study to asses the effects of varying sample sizes and signal-to-noise ratios (SNRs). To be exact, SNR of the simulated time series $X_j(t) = \mathbf{e}^T_{L(j)} \cdot \mathbf{Z}_j(t) + \epsilon_j(t)$, is defined,

$$\text{SNR} = \frac{\text{Var}(\mathbf{e}^T_{L(j)} \cdot \mathbf{Z}_j(\cdot))}{\text{Var}(\epsilon_j(\cdot))},$$

and we investigate the clustering accuracy when sample size is set to $n \in \{10, 20, 40\}$ with a SNR$\in \{10, 5, 1\}$. For each scenario, 100 Monte Carlo datasets are simulated with an random α uniformly chose between 0.25 and 0.9, however, the time series are set to be stationary at the *main* state. The signal length retains at 50 segments, smoothing fixed at a window size of 5 and overlapping of 2, along with the other technical parameters unchanged regarding the simulation. As shown in Table **??**, individual labels are perfectly estimated, group label accuracy is robust to low SNRs and is significantly improved as sample size increases.

The setup is repeated on simulated time series that are non-stationary, which randomly alternates between the *main* and *off* state. Clustering accuracy is reported in Table **??**,

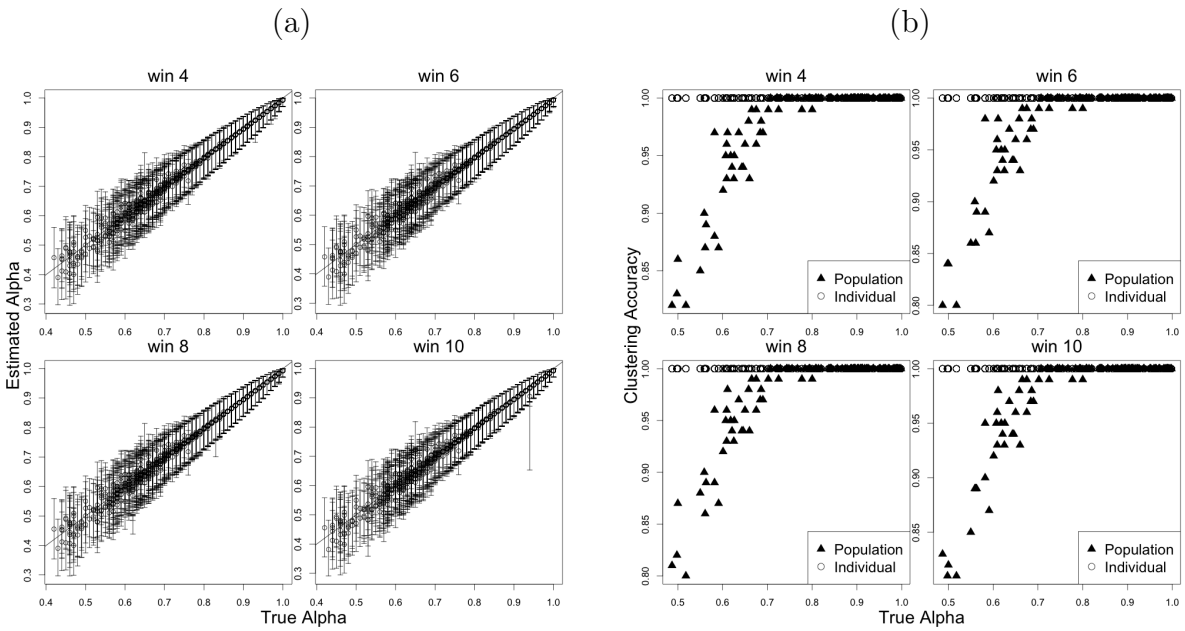

Figure 2: Extended simulation results: (a) $\widehat{\alpha}$ and its 90% credible interval from MCMC samples, against the true α's. (b) Clustering accuracy at both individual and group level, as a function of the true α's.

	$N = 10$		$N = 20$		$N = 40$	
SNR	grp	ind	grp	ind	grp	ind
10	0.8121	1.000	0.8684	1.000	0.9031	1.000
5	0.8073	1.000	0.8669	1.000	0.8975	1.000
1	0.8099	1.000	0.8691	1.000	0.9000	1.000

Table 1: Clustering accuracy relative to sample sizes and SNRs

where both individual and group accuracies approximates the stationary case consolidating the robustness of the proposed procedure to piecewise stationarity.

4 Case Study Extended Results

4.1 EEG measurements

The sample in our case study includes 9 participants (29-60 months of age) recognized as Typical Developing (TD), and 10 participants (27-99 months of age) diagnosed with Autism Spectrum Disorder (ASD). During the experiment, EEG was recorded at 250/500Hz using 129 channel geodesic nets with Ag/AgCl electrodes, while participants watched videos of soap bubbles and other non-social images on a computer monitor for 2 to 6 minutes.

The EEG recordings were bandpass filtered at 1-50 Hz using a finite impulse response (FIR) filter with EEGLAB toolbox (**?**). Recordings were then segmented into 1024 ms segments for preprocessing. Noisy or loose channels were spherically interpolated using

	$N = 10$		$N = 20$		$N = 40$	
SNR	grp	ind	grp	ind	grp	ind
10	0.8079	1.000	0.8687	1.000	0.8971	1.000
5	0.8122	1.000	0.8667	1.000	0.8985	1.000
1	0.8131	1.000	0.8660	1.000	0.9020	1.000

Table 2: Clustering accuracy relative to sample sizes and SNRs (non-stationary)

EEGLAB, and EEG recording segments with > 11 interpolated channels were rejected. 4 eye channels were physically removed from the net before the recording session even began, therefore excluded from clustering analysis along with another reference channel that was placed on the top of the scalp. All remaining segments were manually inspected for non-stereotyped artifacts, e.g., electromyogram (EMG), and rejected based on qualitative inspection; then a combined principal component analysis (PCA) and independent component analysis (ICA) approach was used to eliminate stereotyped artifacts, e.g., ocular artifacts. All EEG recordings were re-referenced to the average prior to power calculations.

EEG was recorded for 165.1($\pm$74.3) seconds among ASD and 139.9($\pm$30.1) among TD. After the aforementioned quality control, 102.7($\pm$19.8, ASD) and 95.9($\pm$32.8, ASD) segments entered the MIC analysis.

4.2 Model selection

The searching algorithm proposed in our main article is empowered to guide the choice of dimensionality d and the number of clusters k. Multiple smoothing settings (γ, δ) are evaluated, as shown in Table **??**, and overall $k = 5$ is favored within both TD and ASD cohort. It is also worth noticing that mean coherence estimated from TD is always higher than ASD, suggesting less heterogeneity in their estimated cluster results. As the window size γ increases, 6 clusters are favored instead of 5, but the adjusted coherence metric indicates worse fitting on the group level, therefore we focused on $\gamma = 6$. Similarly more clusters are suggested for smaller overlaps, as have been observed in the simulation study, which could indicates the non-stationary nature of the observed EEG readings. And eventually, we decide to proceed with $\gamma = 6, \delta = 4/6$ by weighing in considerations for non-stationarity and overfitting on the epoch level.

4.3 MCMC Mixing

We consider MCMC mixing by evaluating coherence parameter (α) and cluster estimates convergence at a fixed configuration of smoothing ($\gamma = 6$ and $\delta = 4$) and model setup ($d = 5$ and $K = 5$). MCMC draws 30k and 60k samples respectively, and a comparison is presented in Table **??** in terms of $\widehat{\alpha}$ and clustering results. To compare K-clusters partition, we used the `adjustedRandIndex` function of the `mclust` package in R (see [**?**], [**?**] and [**?**]).

The 90% credible intervals for α_i's are relatively close within the range of ± 0.01, and the cluster labels perfectly agree at group level and highly consistent for coherent individuals (subject 1, 2, 7, and 9). Therefore, for the purpose of facilitating group level inference,

smoothing		TD			ASD		
window (γ)	overlap ($\gamma\delta$)	d	k	mean($\widehat{\alpha}$)	d	k	mean($\widehat{\alpha}$)
4	2	5	5	0.783	5	5	0.751
6	3	5	5	0.747	5	5	0.723
8	4	6	5	0.696	5	5	0.701
10	5	6	6	0.691	6	5	0.661
12	6	6	6	0.669	6	5	0.659
6	5	5	5	0.786	5	5	0.754
6	4	5	5	0.763	5	5	0.722
6	3	5	5	0.747	5	5	0.723
6	2	6	6	0.725	5	5	0.696
6	1	6	5	0.717	5	5	0.695
6	0	6	6	0.698	5	5	0.679

Table 3: (d, k) search results on TD and ASD with varying smoothing settings

	α C.I. 20K	α C.I. 40K	adjRandIndex
population	·	·	1.0000
subject 1	(0.7249, 0.8643)	(0.7277, 0.8695)	0.9784
subject 2	(0.7053, 0.8480)	(0.6984, 0.8432)	0.9666
subject 3	(0.6579, 0.8188)	(0.6641, 0.8297)	0.9264
subject 4	(0.7158, 0.8549)	(0.7188, 0.8574)	0.8940
subject 5	(0.6003, 0.7669)	(0.6097, 0.7739)	0.7839
subject 6	(0.6745, 0.8295)	(0.6787, 0.8329)	0.7082
subject 7	(0.7118, 0.8549)	(0.7099, 0.8527)	0.9645
subject 8	(0.6817, 0.8337)	(0.6732, 0.8269)	0.7751
subject 9	(0.7328, 0.8656)	(0.7408, 0.8707)	0.9782

Table 4: MCMC mixing at 30k and 60k draws

the high consistency of the proposed method highlights its advantage of being robust to loosely coherent individuals, whose hard clustering label estimates are sensitive to the missing segments and experimental artifacts.

4.4 MIC separated Spectral Densities

To better understand how band powers are driving the MIC clusters, we relate the MIC labeled electrodes to their power spectral densities obtained on the epoch level. Fig **??** presents the epoch level spectral densities and the corresponding MIC determined cluster labels from two individuals belonging to the ASD group. First of all, every cluster manifests its spectral characteristics for both individuals, for example, frontal region (red) features a bimodal PSD in the theta and delta bands, and separable from each other visually. More importantly, the spectral characteristics differ dramatically across individual, even when the electrodes are clustered together. As an example, the mid-frontal region (red) is identified for both cases, even though it features a bimodal PSD in theta and delta in (a) but theta and alpha in

(b). It also highlights an advantage of the proposed MIC framework, that the clustering results depends on the proximity measure while allowing for information integration across individuals.

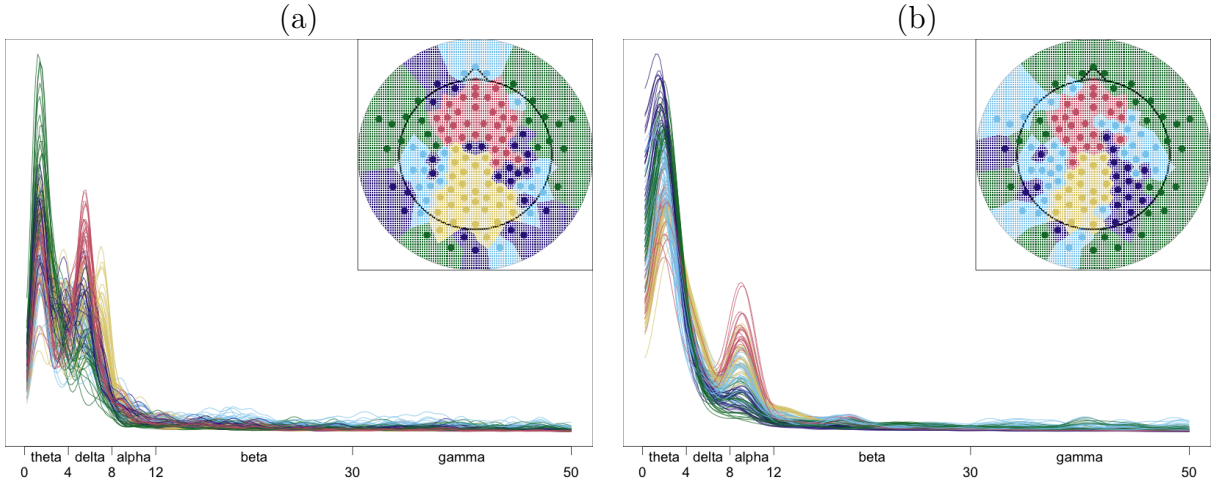

Figure 3: MIC labeled electrodes and underlying PSDs for two subjects from ASD group: (a) one epoch from subject 1. (b) one epoch from subject 2.

4.5 Cluster assessment using Random Forest

MIC provides estimates of cluster labels at the group, subject, and epoch level. Therefore, a formal attempt at covariate adjustment may require significant methodological extensions to account for the multilevel nature of clustering patterns. Here we conduct an informal investigation on how specific spectral features, and individual characteristics may explain clustering. Specifically, we consider a post-hock analysis of cluster label estimates at the subject level. Taking cluster labels as a categorical outcome, we explain cluster label variability in a classification setting. Each individual contributes 124 electrodes labels. Cluster label predictors include, cohort indicators (TD/ASD), age information, and band power estimates. In order, to obtain band power estimates at the subject level, we consider the most coherent (highest $\widehat{\beta}$) epoch from each individual, and summarize its spectral estimates by 5 frequency bands.

We use the `randomForest` function from the R package `randomForest` to predict MIC estimated cluster labels, using power bands, group and age. Classification errors are 0.2806, 0.2783, 0.2049, 0.2603 and 0.3450 for each cluster respectively. Table **??** shows the cluster specific predictor importance, and highest contribution comes from age against the lowest from group. As for band powers, theta band turns out the most important whereas beta band the least, but in general band powers are not predicting cluster labels well, or equivalently, the MIC estimated clusters are not driven by any band patterns.

cluster	1	2	3	4	5	MeanAccuDecre	MeanGiniDecre
theta	0.1651	0.1527	0.1497	0.1434	0.1328	0.1487	310.41
delta	0.1367	0.1383	0.1242	0.1444	0.1397	0.1359	323.87
alpha	0.1251	0.1223	0.0964	0.1567	0.1188	0.1222	303.63
beta	0.1026	0.1021	0.0576	0.1267	0.1141	0.0981	277.19
gamma	0.1118	0.1317	0.0859	0.1392	0.1247	0.1172	302.06
group	0.0440	0.0485	0.0176	0.0486	0.0424	0.0392	36.55
age	0.1523	0.1775	0.0853	0.1951	0.1827	0.1550	306.15

Table 5: Predictor importance by using a RandomForest to predict MIC labels

References

[] H. Akaike. Power spectrum estimation through autoregressive model fitting. *Annals of the institute of Statistical Mathematics*, 21(1):407–419, 1969.

[] D. R. Brillinger. *Time series: data analysis and theory*. Holden-Day Inc. Oakland, California, 1981.

[] G. Celeux, F. Forbes, C. P. Robert, D. M. Titterington, et al. Deviance information criteria for missing data models. *Bayesian analysis*, 1(4):651–673, 2006.

[] A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[] C. Euan, H. Ombao, and J. Ortega. Spectral synchronicity in brain signals. *arXiv preprint arXiv:1507.05018*, 2015.

[] C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631, 2002.

[] C. Fraley, A. Raftery, and L. Scrucca. Normal mixture modeling for model-based clustering, classification, and density estimation. *Department of Statistics, University of Washington*, 23:2012, 2012.

[] L. Hubert and P. Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

[] E. F. Lock and D. B. Dunson. Bayesian consensus clustering. *Bioinformatics*, 29(20): 2610–2616, 2013.

[] H. C. Ombao, J. A. Raz, R. L. Strawderman, and R. Von Sachs. A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*, 88 (4):1186–1192, 2001.

[] G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2): 461–464, 1978.

[] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

[] P. Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.