

# Adaptive Bayesian Spectral Analysis of High-dimensional Nonstationary Time Series

Zeda Li, Ori Rosen, Fabio Ferrarelli and Robert T. Krafty \*

December 19, 2020

---

\*Zeda Li is Assistant Professor, Paul H. Chook Department of Information System and Statistics, Baruch College, The City University of New York (zeda.li@baruch.cuny.edu), Ori Rosen is Professor, Department of Mathematical Sciences, University of Texas at El Paso (orosen@utep.edu), Fabio Ferrarelli is Assistant Professor, Department of Psychiatry, University of Pittsburgh (ferrarellif@upmc.edu), and Robert T. Krafty is Professor and Chair, Department of Biostatistics and Bioinformatics, Emory University (rkrafty@emory.edu). This work was funded in part by PSC-CUNY Research Award 63069-0051, a Eugene M. Lang Junior Faculty Research Fellowship, and National Institutes of Health grants R01GM113243, R01MH113827, and R21MH119543.

# Abstract

This article introduces a nonparametric approach to spectral analysis of a high-dimensional multivariate nonstationary time series. The procedure is based on a novel frequency-domain factor model that provides a flexible yet parsimonious representation of spectral matrices from a large number of simultaneously observed time series. Real and imaginary parts of the factor loading matrices are modeled independently using a prior that is formulated from the tensor product of penalized splines and multiplicative gamma process shrinkage priors, allowing for infinitely many factors with loadings increasingly shrunk towards zero as the column index increases. Formulated in a fully Bayesian framework, the time series is adaptively partitioned into approximately stationary segments, where both the number and locations of partition points are assumed unknown. Stochastic approximation Monte Carlo (SAMC) techniques are used to accommodate the unknown number of segments, and a conditional Whittle likelihood-based Gibbs sampler is developed for efficient sampling within segments. By averaging over the distribution of partitions, the proposed method can approximate both abrupt and slowly varying changes in spectral matrices. Performance of the proposed model is evaluated by extensive simulations and demonstrated through the analysis of high-density electroencephalography. Supplementary materials for this article are available online.

KEY WORDS: Factor Analysis; High-dimensional Time Series; Locally Stationary Process; Multiplicative Gamma Process; Penalized Splines; Spectral Analysis; Stochastic Approximation Monte Carlo.

## 1 Introduction

Technological advances have facilitated an explosion in the number of studies that simultaneously record a large number of processes over time. In many applications, important scientific information is contained in the variability attributable to oscillations at different frequencies, which can be quantified through the power spectrum. Simultaneous analyses of such data, which take into account all cross-time series dependencies as well as within-

series frequency patterns, provide a comprehensive understanding of the underlying process. However, such analyses are challenging when the data are high-dimensional and nonstationary, since unwieldy sizes of spectral matrices evolving over time present methodological and computational obstacles.

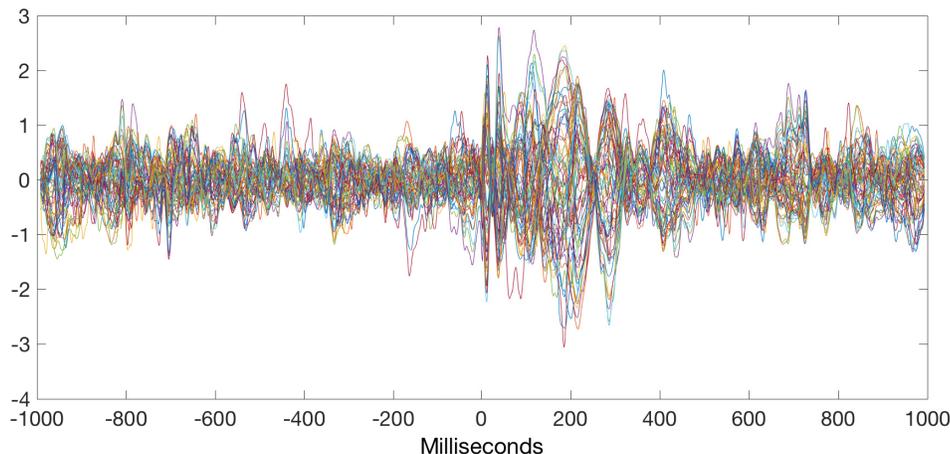


Figure 1: 64-channel TMS-evoked hdEEG from a first-episode psychosis patient.

Our motivating application comes from the analysis of 64-channel high-density electroencephalography (hdEEG) with transcranial magnetic stimulation (TMS) in a patient who was hospitalized during a first-break psychotic episode (Figure 1). TMS is a noninvasive procedure that uses magnetic fields to excite brain cells. EEG is used to measure electrophysiological activity simultaneously across multiple regions, or channels, of the brain. Traditional EEG measures activity at 2-16 different channels. More recently, hdEEG, which can measure activity from up to 256 channels, has been utilized to obtain a more comprehensive view of brain activity. The power spectrum of hdEEG during TMS and its evolution provide dynamic, high-resolution neurobiological information with regards to neurological mechanisms, disruptions of which have been observed in patients with a clinical diagnosis of schizophrenia (Kaskie and Ferrarelli, 2018). The goal of our analysis is to investigate hdEEG during TMS in our patient who is experiencing a first-break psychosis, a potential precursor to a future diagnosis of schizophrenia, which could serve as a potential pre-clinical biomarker (Ferrarelli et al., 2018).

Historically, a rolling-window procedure (Priestley, 1981) is used for the spectral analysis

of nonstationary time series whose second-order structure evolves over time. This procedure partitions the time series into prespecified overlapping time blocks and then estimates the power spectrum at each of these blocks by smoothing the local periodogram, which is a noisy estimate of the local power spectrum, across frequencies using tools such as local averaging (Shumway and Stoffer, 2011). More recently, a variety of approaches to the spectral analysis of multivariate nonstationary time series have been proposed. These methods can be roughly grouped into three categories: methods that assume, in a manner similar to the rolling window estimator, power spectra evolve smoothly over time (Dahlhaus, 2000; Guo and Dai, 2006; Sanderson et al., 2010; Park et al., 2014), methods that are based on data-driven piecewise stationary approximations (Ombao et al., 2005; Davis et al., 2006), and adaptive Bayesian methods that can approximate both abrupt and slowly varying dynamics (Zhang, 2016; Li and Krafty, 2019). However, these methods focus on the analysis of low-dimensional collections of time series, becoming theoretically unjustified or computationally infeasible when a large, or even moderate, number of simultaneous series are observed. The primary contribution of this article is introducing an adaptive method for the spectral analysis of high-dimensional time series that can capture both abrupt and slowly varying dynamics efficiently while nonparametrically modeling spectral matrices.

Various approaches have been developed for estimating the spectral matrix of a high-dimensional stationary time series. These include shrinkage estimators (Böhm and von Sachs, 2009; Fiecas et al., 2010; Fiecas and von Sachs, 2014; Schneider-Luftman and Walden, 2016) and thresholding estimators (Sun et al., 2018; Fiecas et al., 2019). Unfortunately, these methods are not readily extendable to the spectral analysis of nonstationary time series. A common approach to analyzing high-dimensional time series is to use factor models, as they not only induce a parsimonious and interpretable structure among the time series, but can also overcome the curse of dimensionality when estimating high-dimensional covariance or spectral matrices, see Ensor (2013) for a review. Although factor models have been used extensively for time-domain time series analysis, mostly in the econometrics literature, existing factor models for frequency-domain analysis of time series are relatively few. These methods are primarily framed in the stationary setting (Stoffer, 1999; Macaro and Prado, 2014), or

are based on time-domain parametric formulations (West et al., 1999; Prado et al., 2001). In addition to these time series factor models, factor models for the analysis of functional time series can also be adapted for the analysis of multivariate power spectra (Kowal et al., 2017). However, the direct application of such functional data-based procedures for spectral analysis can only estimate predefined functions, such as univariate log-spectra and squared coherences, but not the entire spectral matrix.

In this article, we present an adaptive nonparametric method for the spectral analysis of high-dimensional nonstationary time series. Our approach is based on a novel frequency-domain locally stationary factor model and scalable Markov chain Monte Carlo (MCMC) techniques. The factor model is general in the sense that it allows for both real- and complex-valued spectral matrices, which means that individual time series can fluctuate simultaneously or propagate in a lagged fashion. Real and imaginary parts of the local factor loading matrices are modeled independently using a novel prior formulated through the tensor product of penalized spline priors, which induce smooth structure across frequency, and multiplicative gamma process shrinkage priors (Bhattacharya and Dunson, 2011), which allow infinitely many factors with the loadings increasingly shrunk towards zero as the column index increases and mitigate the need for estimating the number of factors. The approach adaptively divides a high-dimensional time series into an unknown random number of approximately stationary segments of variable lengths through stochastic approximation Monte Carlo (SAMC). By approximating the likelihood function via products of local Whittle likelihoods, a scalable and efficient Gibbs sampling algorithm is developed. As in other adaptive Bayesian methods (Rosen et al., 2012; Zhang, 2016; Li and Krafty, 2019), by averaging estimates across the distribution of partitions, the method not only produces estimates that can capture abrupt changes, but also effectively approximates slowly varying processes.

The article is organized as follows. Section 2 presents the proposed frequency-domain factor model for high-dimensional nonstationary time series. Section 3 describes the prior distributions for the model parameters and an outline of the sampling scheme. Section 4 presents the results of extensive simulation studies. The proposed method is applied to real

data in Section 5. Section 6 concludes with some discussion and future work. A detailed description of the sampling scheme is given in the supplementary material.

## 2 The Model

### 2.1 The Locally Stationary Factor Model

This article considers  $P$ -dimensional vector-valued time series  $\mathbf{X}_t = (X_{1t}, \dots, X_{Pt})'$  that can be represented through a novel locally stationary factor model. The model possesses a Cramér representation with time-varying factor loadings, which decomposes  $\mathbf{X}_t$  into information accounted for by a set of  $Q$  common processes, or factors, plus an idiosyncratic component. Formally, we model an  $\mathbb{R}^P$ -valued time series of length  $T$ ,  $\{\mathbf{X}_t : t = 1, \dots, T\}$ , as

$$\mathbf{X}_t = \int_0^1 \Lambda(t/T, \omega) \exp(2\pi i \omega t) d\mathbf{Z}(\omega) + \boldsymbol{\epsilon}_t, \quad (1)$$

where  $\mathbf{Z}(\omega)$  is a  $Q$ -dimensional mean-zero orthogonal incremental process with independent and Hermitian latent factors, such that  $E\{d\mathbf{Z}(\omega)d\mathbf{Z}^*(\zeta)\}$  is the identity matrix if  $\omega = \zeta$  and zero otherwise,  $\boldsymbol{\epsilon}_t$  is a  $P$ -dimensional independent white noise, and  $\Lambda(u, \omega)$  is a  $P \times Q$  time-varying loading matrix that is a function of scaled time  $u \in [0, 1]$  and frequency  $\omega \in \mathbb{R}$ . The complex-valued loading matrix  $\Lambda(u, \omega)$  is periodic and Hermitian as a function of frequency such that  $\Lambda(u, \omega) = \Lambda(u, \omega + 1)$  and  $\Lambda(u, \omega) = \overline{\Lambda(u, -\omega)}$ , where  $\bar{\Lambda}$  is the conjugate of  $\Lambda$ . The time-varying power spectrum defined through the locally stationary factor Cramér representation in (1) is given by

$$f(u, \omega) = \Lambda(u, \omega)\Lambda(u, \omega)^* + \Sigma_\epsilon, \quad u \in [0, 1], \omega \in \mathbb{R},$$

where  $\Lambda^*$  is the conjugate transpose of  $\Lambda$ , and  $\Sigma_\epsilon$  is the  $P \times P$  diagonal covariance matrix of  $\boldsymbol{\epsilon}_t$ . Hence, the time-varying spectrum  $f(u, \omega)$  is a complex-valued positive definite  $P \times P$  Hermitian matrix. We assume that given  $u$ , each component of  $f(u, \cdot)$  possesses a square-integrable first derivative as a function of frequency; given  $\omega$ , each component of  $f(\cdot, \omega)$  is continuous as a function of scaled time at all but a possibly finite number of points.

Several aspects of the proposed locally stationary factor model should be noted. First, a time series defined by the locally stationary factor model is locally stationary in the sense of Li and Krafty (2019), which differs slightly from the definitions of local stationarity used by Dahlhaus (2000) and Guo and Dai (2006). The model of Dahlhaus (2000) assumes a series of transfer functions/loadings indexed by the length of the time series that converges to a fixed function as the length increases. This structure is adopted primarily to allow for the fitting of popular parametric models such as time-varying moving average models. Since we are considering nonparametric estimation, in a manner similar to that in Guo and Dai (2006) and Li and Krafty (2019), we directly use the limiting function. Further, similar to the method of Li and Krafty (2019), the proposed model allows for a finite number of discontinuities of the spectrum as a function of time. This is more flexible than the models of Dahlhaus (2000) and Guo and Dai (2006), which require continuity as a function of time. Second, through the introduction of the tensor product penalized spline and multiplicative gamma process shrinkage prior (see details in Section 3.1), the proposed model allows for an infinite number of factors with sufficiently decaying loadings. This enables any locally stationary process in the sense of Guo and Dai (2006) or Li and Krafty (2019) to be represented as a locally stationary factor model. The use of this prior also mitigates the sensitivity to the number of factors. Nevertheless, the finite factor representation is used to overcome the curse of dimensionality in estimating the power spectrum, where the number of factors  $Q$  is typically smaller than  $P$ , inducing a reduced-rank characterization of the spectral matrix. Lastly, a common concern about any factor analysis is that the loadings are not identifiable without appropriate constraints, such as lower triangular structure for the loading matrix (Geweke and Zhou, 1996; Carvalho et al., 2008). It should be emphasized that our goal is not to identify or interpret the factors themselves, but to estimate time-varying power spectral matrices. As noted by Bhattacharya and Dunson (2011) in the context of covariance matrices, the shrinkage provided by the multiplicative gamma process prior enables valid estimation, prediction, and inference on power spectrum despite the lack of identifiability of the loadings.

## 2.2 Piecewise Stationary Approximation

A locally stationary time series can be accurately approximated as a piecewise stationary process (Adak, 1998; Guo and Dai, 2006), and our procedure involves a piecewise stationary approximation to the locally stationary factor model. For a time series of length  $T$   $\{\mathbf{X}_t : t = 1, \dots, T\}$ , a collection of partition points of the time series into  $L$  segments is denoted by  $\boldsymbol{\xi} = (\xi_0, \dots, \xi_L)'$  where  $\xi_0 = 0$  and  $\xi_L = T$  such that  $\mathbf{X}_t$  is approximately stationary within the segments  $\{t : \xi_{\ell-1} < t \leq \xi_\ell\}$  for  $\ell = 1, \dots, L$ . Conditional on  $L$  and  $\boldsymbol{\xi}$ ,

$$\mathbf{X}_t \approx \sum_{\ell=1}^L \int_0^1 \Lambda_\ell(\omega) \exp(2\pi i \omega t) d\mathbf{Z}(\omega) + \boldsymbol{\epsilon}_t,$$

where for  $\ell = 1, \dots, L$ ,  $\Lambda_\ell(\omega) = \Lambda(u_\ell, \omega) \delta_\ell(t)$ ,  $\delta_\ell(t)$  is an indicator function such that  $\delta_\ell(t) = 1$  if  $t \in (\xi_{\ell-1}, \xi_\ell]$  and zero otherwise, and  $u_\ell = (\xi_\ell + \xi_{\ell-1})/2$  is the midpoint of the  $\ell$ th segment. Within the  $\ell$ th segment, the time series is approximately second-order stationary with local power spectrum  $f(u_\ell, \omega) = \Lambda_\ell(\omega) \Lambda_\ell(\omega)^* + \Sigma_\epsilon$ , where  $\Sigma_\epsilon$  is a diagonal covariance matrix of  $\boldsymbol{\epsilon}_t$ . It should be noted that the number of segments  $L$  and partition  $\boldsymbol{\xi}$  are random variables whose prior distributions are given in Section 3.2. Estimates and inferences will be obtained by averaging over the posterior distribution of the partitions and the number of segments.

Conditional on approximately stationary segments, define the local discrete Fourier transform (DFT) at frequency  $\omega_{k\ell}$  within segment  $\ell$  as

$$\mathbf{Y}_{k\ell} = T_\ell^{-1/2} \sum_{t=\xi_{\ell-1}+1}^{\xi_\ell} \mathbf{X}_t \exp(-2\pi i \omega_{k\ell} t), \quad k = 1, \dots, K_\ell, \ell = 1, \dots, L,$$

where  $T_\ell$  is the number of time points in the  $\ell$ th segment,  $\omega_{k\ell} = k/T_\ell$ ,  $k = 1, \dots, K_\ell$  are the Fourier frequencies, and  $K_\ell = \lfloor (T_\ell - 1)/2 \rfloor$ , which is the greatest integer that is less than or equal to  $(T_\ell - 1)/2$ . Under some regularity conditions (Brillinger, 2002, Theorem 4.4.1), the  $\mathbf{Y}_{k\ell}$  are approximately independent multivariate complex Gaussian random vectors with mean  $\mathbf{0}$  and covariance  $f(u, \omega_{k\ell})$ , denoted by  $\mathbf{Y}_{k\ell} \stackrel{\text{app}}{\sim} CN[\mathbf{0}, f(u, \omega_{k\ell})]$ . The sampling scheme follows from the distribution of the local DFTs conditional on the latent factors. In particular, we let  $\mathbf{D}_{k\ell} = \int_0^{\omega_{k\ell}} \mathbf{Z}(\omega) d\omega$ ,  $\mathbf{E}_{k\ell} = T_\ell^{-1/2} \sum_{t=\xi_{\ell-1}+1}^{\xi_\ell} \boldsymbol{\epsilon}_t \exp(-2\pi i \omega_{k\ell} t)$ , and  $\Lambda_{k\ell} = \Lambda(u_\ell, \omega_{k\ell})$ .

It then follows from Brillinger (2002) that

$$\mathbf{Y}_{k\ell} \approx \Lambda_{k\ell} \mathbf{D}_{k\ell} + \mathbf{E}_{k\ell},$$

where  $\mathbf{D}_{k\ell}$  is approximately independently distributed  $CN(\mathbf{0}, I_Q)$ ,  $I_Q$  is a  $Q \times Q$  identity matrix, and  $\mathbf{E}_{k\ell} \sim CN(\mathbf{0}, \Sigma_\epsilon)$ . This leads to the conditional Whittle likelihood

$$\mathcal{L}(\mathbf{Y} \mid \mathbf{\Lambda}, \mathbf{D}, \Sigma_\epsilon, L) \approx \prod_{\ell=1}^L \prod_{k=1}^{K_\ell} \prod_{p=1}^P \left\{ \sigma_{\epsilon,p}^{-1} \exp \left( -\sigma_{\epsilon,p}^{-2} \left| Y_{k\ell p} - \Lambda_{k\ell}^{(p)} \mathbf{D}_{k\ell} \right|^2 \right) \right\},$$

where  $\mathbf{Y}$  represents all local discrete Fourier transforms,  $\mathbf{\Lambda}$  and  $\mathbf{D}$  represent the collections of loadings and factors at all segments and Fourier frequencies,  $\sigma_{\epsilon,p}^2$  is the  $p$ th diagonal element of  $\Sigma_\epsilon$ ,  $Y_{k\ell p}$  is the  $p$ th element of  $\mathbf{Y}_{k\ell}$  and  $\Lambda_{k\ell}^{(p)}$  is the  $p$ th row of  $\Lambda_{k\ell}$ .

### 3 Adaptive Bayesian Spectral Analysis

In this section, we introduce an adaptive Bayesian approach that extends the approaches of Rosen et al. (2012), Zhang (2016), and Li and Krafty (2019) to spectral analysis of high-dimensional time series. Under this approach, a time series is adaptively partitioned into a random number of approximately stationary segments, local spectra are estimated within each segment, and time-varying power spectrum is obtained by averaging local estimates over the distribution of partitions. First, in Section 3.1, we present the Bayesian estimation procedure for a high-dimensional *stationary* time series. Then, in Section 3.2, we introduce our proposed SAMC-based adaptive Bayesian sampling scheme for temporal partitioning of a *nonstationary* time series.

#### 3.1 Spectral Estimation for Stationary Time Series

To aid the presentation, in this section, we focus on estimating the spectral matrix,  $f(\omega)$ , of a stationary time series. Under the factor formulation,  $f(\omega)$  is expressed as a sum of a nonnegative definite matrix  $\Lambda(\omega)\Lambda(\omega)^*$  induced by the loadings, and a positive-definite diagonal matrix  $\Sigma_\epsilon$ .

Given the factor model, the conditional Whittle likelihood for a stationary multivariate time series is given by

$$\mathcal{L}(\mathbf{Y} \mid \Lambda, \mathbf{D}, \Sigma_\epsilon) \approx \prod_{k=1}^K \prod_{p=1}^P \left\{ \sigma_{\epsilon,p}^{-1} \exp \left( -\sigma_{\epsilon,p}^{-2} \left| Y_{kp} - \Lambda_k^{(p)} \mathbf{D}_k \right|^2 \right) \right\},$$

where  $\mathbf{Y}$  contains the discrete Fourier transforms of the time series,  $\Lambda$  and  $\mathbf{D}$  denote the collections of loadings and factors, respectively,  $Y_{kp}$  is the  $p$ th element of  $\mathbf{Y}_k$  and  $\Lambda_k^{(p)}$  is the  $p$ th row of  $\Lambda_k$ . To complete our model, we place prior distributions on  $\Lambda$  and  $\sigma_{\epsilon,p}$ .

Prior distributions on the error variances are placed by assuming that  $\sigma_{\epsilon,p}$ , for  $p = 1, \dots, P$ , are independent Half- $t(\nu, G_\epsilon)$  with density  $p(\sigma_{\epsilon,p}) \propto [1 + (\sigma_{\epsilon,p}/G_\epsilon)^2/\nu]^{-(\nu+1)/2}$ ,  $\sigma_{\epsilon,p} > 0$ , where the hyperparameters  $\nu$  and  $G_\epsilon$  are known constants (Gelman, 2006). In practice, this distribution can be represented by a scale mixture of inverse gamma distributions (Wand et al., 2012):  $(\sigma_{\epsilon,p}^2 | g_{\epsilon,p}) \sim \text{IG}(\nu/2, \nu/g_{\epsilon,p})$ ,  $g_{\epsilon,p} \sim \text{IG}(1/2, 1/G_{\epsilon,p}^2)$ , where  $\text{IG}(a, b)$  denotes an inverse Gamma distribution with density  $p(x) \propto x^{-(a+1)} \exp(-b/x)$ ,  $x > 0$ .

A novel prior is placed on the  $P \times Q$  complex-valued loading matrix  $\Lambda(\omega)$ , such that the real and imaginary parts of its entries are modeled independently through tensor products of Bayesian penalized splines (Krafty et al., 2017; Li and Krafty, 2019) and multiplicative gamma process shrinkage priors (Bhattacharya and Dunson, 2011). In particular, we use the first  $S$  Demmler-Reinsch basis functions

$$\Re\{\Lambda_{pq}(\omega)\} = \alpha_{pq0} + \sum_{s=1}^{S-1} \alpha_{pqs} \sqrt{2} \cos(2\pi s\omega), \quad (2)$$

$$\Im\{\Lambda_{pq}(\omega)\} = \sum_{s=1}^S \beta_{pqs} \sqrt{2} \sin(2\pi s\omega). \quad (3)$$

In this formulation, the real and imaginary parts of the loading matrix are modeled by periodic even and odd linear splines, respectively, which has been found to improve performance compared to a model that only accounts for periodic patterns but does not restrict the functions to be odd or even (Krafty and Collinge, 2013). A Bayesian penalized spline can be formulated by placing independent  $N[0, (2\pi s)^{-1} \tau^2]$  priors on the coefficients conditional on a smoothing parameter  $\tau^2$ . This prior induces smoothness as a function of frequency by regularizing integrated squared first derivatives. However, we also desire the loadings to

decay for large  $q$ , so that the majority of information is captured by the first several factors. To achieve this, we introduce a prior that is a tensor product of this penalized spline prior with a gamma process shrinkage prior. Formally, we define the prior distributions as:

1.  $\alpha_{pq0} \sim N\left(0, \psi_{q,(re)}^{-1}\right)$ ,  $\alpha_{pqs} \sim N\left(0, (2\pi s)^{-1} \tau_{pq,(re)}^2 \psi_{q,(re)}^{-1}\right)$  for  $s = 1, \dots, S - 1$ .
2.  $\beta_{pqs} \sim N\left(0, (2\pi s)^{-1} \tau_{pq,(im)}^2 \psi_{q,(im)}^{-1}\right)$  for  $s = 1, \dots, S$ .
3.  $\tau_{pq,(re)}, \tau_{pq,(im)} \sim \text{Half-}t(\nu, G_\tau)$  are the square roots of the smoothing parameters of the real and imaginary parts, respectively, which control the roughness as a function of frequency.
4.  $\psi_{q,(\cdot)}$ ,  $q = 1, \dots, Q$ , and  $(\cdot)=(re)$  or  $(im)$ , are shrinkage parameters that control the decay of the columns of the loading matrix, where  $\psi_{q,(\cdot)} = \prod_{h=1}^q \phi_{h,(\cdot)}$ , with independent priors  $\phi_{1,(\cdot)} \sim \text{Ga}(a_1, 1)$ ,  $\phi_{q,(\cdot)} \sim \text{Ga}(a_2, 1)$  for  $q \geq 2$ , where  $a_1$  and  $a_2$  are fixed constants.

This formulation has three favorable properties. First, the choice of  $S$  provides a compromise between loss of accuracy relative to the full-rank Bayesian smoothing spline ( $S = K$ ) and computational feasibility. We select  $S = 10$  in our simulation studies and data analysis, which accounts for at least 99.975% of the total variance of the full smoothing spline when  $T \leq 10^4$  (Krafty et al., 2017), for considerable computational savings without sacrificing model fit. Second, the shrinkage parameters  $\psi_{q,(\cdot)}$  are stochastically increasing when  $a_2 > 1$ , which places more shrinkage toward zero as the column index increases. This formulation allows an infinite number of factors with sufficiently decaying loadings, which not only provides a good approximation to the spectral representation of multivariate stationary time series (Brillinger, 2002), but also reduces the sensitivity to the number of factors. Third, this formulation allows for sampling from the posterior distribution through an efficient Gibbs sampler which scales well to high-dimensional time series. The sampling scheme can also be performed in parallel to further increase computational speed. Details of the sampling scheme are provided in the supplementary material.

## 3.2 Adaptive Spectral Estimation for Nonstationary Time Series

As described in Section 2, for the spectral analysis of a nonstationary time series, we adaptively estimate the unknown number and locations of the partition points. Conditional on the partitions, local estimation is performed as in Section 3.1.

### 3.2.1 Priors

We first specify prior distributions for the number and locations of the partitions. Following Rosen et al. (2012), a discrete uniform prior is placed on the number of partitions, such that  $L \sim \mathcal{U}(1, L_{\max})$ , where  $L_{\max}$  is a fixed large integer that represents the maximum possible number of possible segments. In general,  $L_{\max}$  is set to be large enough to capture all possible approximately stationary segments, but if after running the procedure, we find that the conditional probability of  $L_{\max}$  is not approximately zero, we increase  $L_{\max}$ . To ensure that the large-sample local Whittle likelihood approximation holds, we choose a minimum number of time points per segment, denoted by  $t_{\min}$ . In our experience,  $t_{\min} = 200$  is sufficient for our simulation studies and real data application. It should be noted that the prespecified  $t_{\min}$  is larger than the one used in Rosen et al. (2012); Li and Krafty (2019), since we are working with high-dimensional time series which may require a larger sample size for the Whittle approximation. Given the number of segments  $L$ , equal weights on all possible locations of a partition point, conditional on the previous partition points, are placed. Specifically, the prior for the partition  $\boldsymbol{\xi}$  is

$$\Pr(\boldsymbol{\xi} \mid L) = \prod_{\ell=1}^{L-1} \Pr(\xi_{\ell} \mid \xi_{\ell-1}, \ell),$$

where  $\Pr(\xi_{\ell} \mid \xi_{\ell-1}, \ell) = 1/a_{\ell}$ , and  $a_{\ell} = T - \xi_{\ell-1} - (L - \ell + 1)t_{\min} + 1$  is the number of possible locations for the  $\ell$ th partition point.

### 3.2.2 Sampling Scheme

We develop a stochastic approximation Monte Carlo (Liang et al., 2007, SAMC) sampling scheme for automatic and adaptive estimation of the time-varying spectrum. Although reversible jump Markov chain Monte Carlo (RJMCMC) has been shown to be effective for adaptive spectral estimation of low-dimensional multivariate nonstationary time series (Li and Krafty, 2019), the self-adjusting property of SAMC is more suitable for high-dimensional nonstationary time series. We briefly describe the sampling scheme in this section and provide technical details in the supplementary material.

To allow a more compact presentation, we define  $\Xi$  as the collection of all parameters, including coefficients of basis functions, smoothing parameters, and shrinkage parameters. The SAMC sampling scheme at iteration  $j$  simulates parameters from the following distribution

$$p_{\vartheta_{j,\ell}}(L, \Xi) \propto \sum_{\ell=1}^{L_{max}} \frac{p(L, \Xi)}{e^{\vartheta_{j,\ell}}} \delta(L = \ell),$$

where  $\vartheta_{j,\ell} \in \Theta$ , for all  $j$ , and  $\Theta$  is a compact set. We consider moving the current value of the chain  $(L^j, \Xi_{L^j}^j)$  to a proposed value  $(L^{j+1}, \Xi_{L^{j+1}}^{j+1})$ . Each iteration consists of two types of moves, within-model and between-model moves, which we outline below.

1. Between-model moves: propose  $L^{j+1}$  by letting  $L^{j+1} = L^j + 1$  or  $L^j - 1$ .

- If  $L^{j+1} = L^j + 1$ , a birth move is proposed. A new partition point is drawn by first randomly selecting a current segment to split, and then randomly proposing a new partition point within that segment. Conditional on  $L^{j+1}$  and the new partition, new smoothing and shrinkage parameters are proposed. Then, factors and coefficients of the real and imaginary parts of the basis functions are updated by using (S1), (S2), and (S3) in the supplementary materials, respectively.
- Or, if  $L^{j+1} = L^j - 1$ , a death move is proposed. A partition point is randomly selected to be deleted and a new segment is formed by combining the two adjacent segments separated by this partition point. Then, the current smoothing and shrinkage parameters of these two segments are used to form new smoothing

and shrinkage parameters in the newly combined segment. Finally, factors and coefficients of the real and imaginary parts of the basis functions are updated by (S1), (S2), and (S3) in the supplementary materials, respectively.

- Let  $q(L^{j+1}, \Xi_{L^{j+1}}^{j+1} | L^j, \Xi_{L^j}^j)$  be the proposal density, the proposed move is then accepted with probability

$$A = \min \left\{ 1, \frac{e^{\vartheta_{j,L^j}}}{e^{\vartheta_{j+1,L^{j+1}}}} \frac{\pi(L^{j+1}, \Xi_{L^{j+1}}^{j+1} | \mathbf{x}) \times q(L^j, \Xi_{L^j}^j | L^{j+1}, \Xi_{L^{j+1}}^{j+1})}{\pi(L^j, \Xi_{L^j}^j | \mathbf{x}) \times q(L^{j+1}, \Xi_{L^{j+1}}^{j+1} | L^j, \Xi_{L^j}^j)} \right\},$$

where  $\mathbf{x}$  is the time series,  $\pi(\cdot)$  is the posterior distribution, and  $e^{\vartheta_{j,L^j}}$  and  $e^{\vartheta_{j+1,L^{j+1}}}$  are self-adjusting factors.

- Set  $\vartheta^* = \vartheta_j + \gamma_{j+1}(\mathbf{e}_{j+1} - p_0)$ , where  $p_0 = 1/L_{\max}$  is a predefined probability,  $\{\gamma_j\}$  is a positive sequence such that  $\gamma_j = j_0 / \max(j_0, j)$ ,  $j_0$  is a predefined constant,  $\mathbf{e}_{j+1} = (e_{j+1,1}, \dots, e_{j+1,L_{\max}})'$  and  $e_{j+1,\ell} = 1$  if  $L^{j+1} = \ell$ , and zero otherwise. If  $\vartheta^* \in \Theta$ , then we set  $\vartheta_{j+1} = \vartheta^*$ ; otherwise, set  $\vartheta = \vartheta^* + \mathbf{c}^*$ , where  $\mathbf{c}^*$  is chosen such that  $\vartheta^* + \mathbf{c}^* \in \Theta$ .

2. Within-model moves: involve no change in the number of segments, i.e.  $L^{j+1} = L^j$ . A partition point  $\xi_\ell$  is randomly selected to be relocated. Then, the factors and basis function coefficients are drawn. These two moves are jointly accepted or rejected in a Metropolis-Hasting step. Lastly, smoothing parameters, shrinkage parameters, and error variances are updated via Gibbs sampling steps.

### 3.2.3 Choosing the Number of Factors $Q$

The infinite frequency-domain factor model with the tensor product penalized spline and multiplicative gamma process shrinkage prior mitigates the sensitivity to the number of factors by ensuring that the effective number of factors is small. From a practical standpoint, we wish to choose a relatively small number of factors  $Q$  to save computational time without sacrificing the estimation accuracy. A possible approach to choosing the number of factors is to use a model selection criterion, such as WAIC (Watanabe, 2010). It is plausible to select  $Q$  through WAIC for spectral analysis of a stationary time series. However, the applicability

of WAIC to a nonstationary time series is doubtful since the values of locally Fourier-transformed data depend on both the number and locations of the partitions which are updated every MCMC iteration. It is also possible to treat  $Q$  as an unknown parameter, putting a prior on  $Q$  and proceeding with a reversible jump Markov chain Monte Carlo or SAMC. However, this strategy substantially increases computational complexity and may lead to slow chain mixing.

We follow the adaptive method of Bhattacharya and Dunson (2011) to choose the effective number of factors  $Q$ . At the  $j$ th iteration, suppose the current number of factors is  $Q^{(j)}$  and let  $\mathbf{e}_\ell$  be a  $Q$ -variate vector with the  $h$ th element equal to 1 if the  $h$ th column of  $|\Lambda_{k\ell}|$  has all its elements in a pre-specified small neighbourhood of zero, and 0 otherwise, for  $h = 1, \dots, Q$ , where  $|\cdot|$  denotes the complex modulus. Further, define  $\mathbf{e} = \mathbf{e}_1 \circ \dots \circ \mathbf{e}_L$ , where  $\circ$  represents the Schur or elementwise product of vectors and  $m^{(j)}$  as the number of 1's in  $\mathbf{e}$ , or the number of redundant columns. Then, if  $m^{(j)} \neq 0$ , we update the effective number of factors for the  $(j + 1)$ th iteration as  $Q^{(j+1)} = Q^{(j)} - m^{(j)}$  with probability  $p(j) = 1/\exp(a_0 + a_1j)$  and discard the redundant columns in  $\Lambda_{k\ell}$ ; if  $m^{(j)} = 0$ , we update  $Q^{(j+1)} = Q^{(j)} + 1$  with the same updating probability  $p(j)$ . When adding a factor, we sample all parameters from the prior distribution to fill in additional columns, and otherwise, retain the parameters corresponding to the nonredundant columns. The parameters  $a_0 = 1$  and  $a_1 = 0.001$  are chosen so that adaptation occurs around every 10 iterations at the beginning of the chain but decreases in frequency exponentially fast.

## 4 Simulation Studies

In this section, we evaluate the proposed method through simulated datasets. In Section 4.1, we investigate performance in estimating time-varying spectra of a piecewise stationary time series with one partition as well as of a slowly varying process. In Section 4.2, we investigate performance in accurately identifying partition points, both for a stationary time series, as well as for a piecewise stationary time series with multiple partition points.

## 4.1 Time-Varying Power Spectrum Estimation

Two processes are considered: piecewise stationary and slowly varying. For each process, we fix the number of observations at  $T = 2048$ , and consider two different dimensions:  $P = 24$  and  $P = 48$ . For each simulated time series, the time-varying power spectrum is estimated by two methods. First, the proposed method with a total of 10000 iterations, where the first 2000 iterations are treated as burn-in, and with various numbers of factors,  $Q = 8, 10, 16$ . We also applied the adaptive approach described in Section 3.2.3 to select  $Q$ . Second, the rolling-window method (Shumway and Stoffer, 2011) that partitions the time series into  $B$  overlapping time blocks and estimates the local power spectrum by smoothed periodogram matrices with bandwidth chosen via generalized cross-validation (GCV) (Ombao et al., 2001). In our simulations, we consider three temporal block sizes,  $B = 64, 128, 256$ .

We investigate the performance of an estimator of a spectral matrix  $f(u, \omega)$  through the mean integrated squared error (MISE), which can be obtained by averaging squared errors across equally spaced time points and frequency grid as follows

$$\text{MISE} = [T(K + 1)]^{-1} \sum_{t=1}^T \sum_{k=0}^K \left\| \widehat{f}(t/T, \omega_k) - f(t/T, \omega_k) \right\|_F^2,$$

where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix, and  $\widehat{f}(t/T, \omega_k)$  is the estimated posterior mean of  $f(t/T, \omega_k)$ . Moreover, we also consider the following quantities,

$$\begin{aligned} \text{MISE}_d &= [T(K + 1)]^{-1} \sum_{t=1}^T \sum_{k=0}^K \left\| \text{diag}[\widehat{f}(t/T, \omega_k)] - \text{diag}[f(t/T, \omega_k)] \right\|^2, \\ \text{MISE}_o &= [T(K + 1)]^{-1} \sum_{t=1}^T \sum_{k=0}^K \left\| \widehat{f}(t/T, \omega_k) - f(t/T, \omega_k) \right\|_{F^*}^2, \end{aligned}$$

where  $\|\cdot\|_{F^*}$  denotes the Frobenius norm, discarding the diagonal entries such that  $\|A\|_{F^*} = \sqrt{\sum_{i=1}^P \sum_{j \neq i} A_{ij}^2}$ . In other words,  $\text{MISE}_d$  and  $\text{MISE}_o$  are the mean squared errors of the diagonal and off-diagonal elements, respectively.

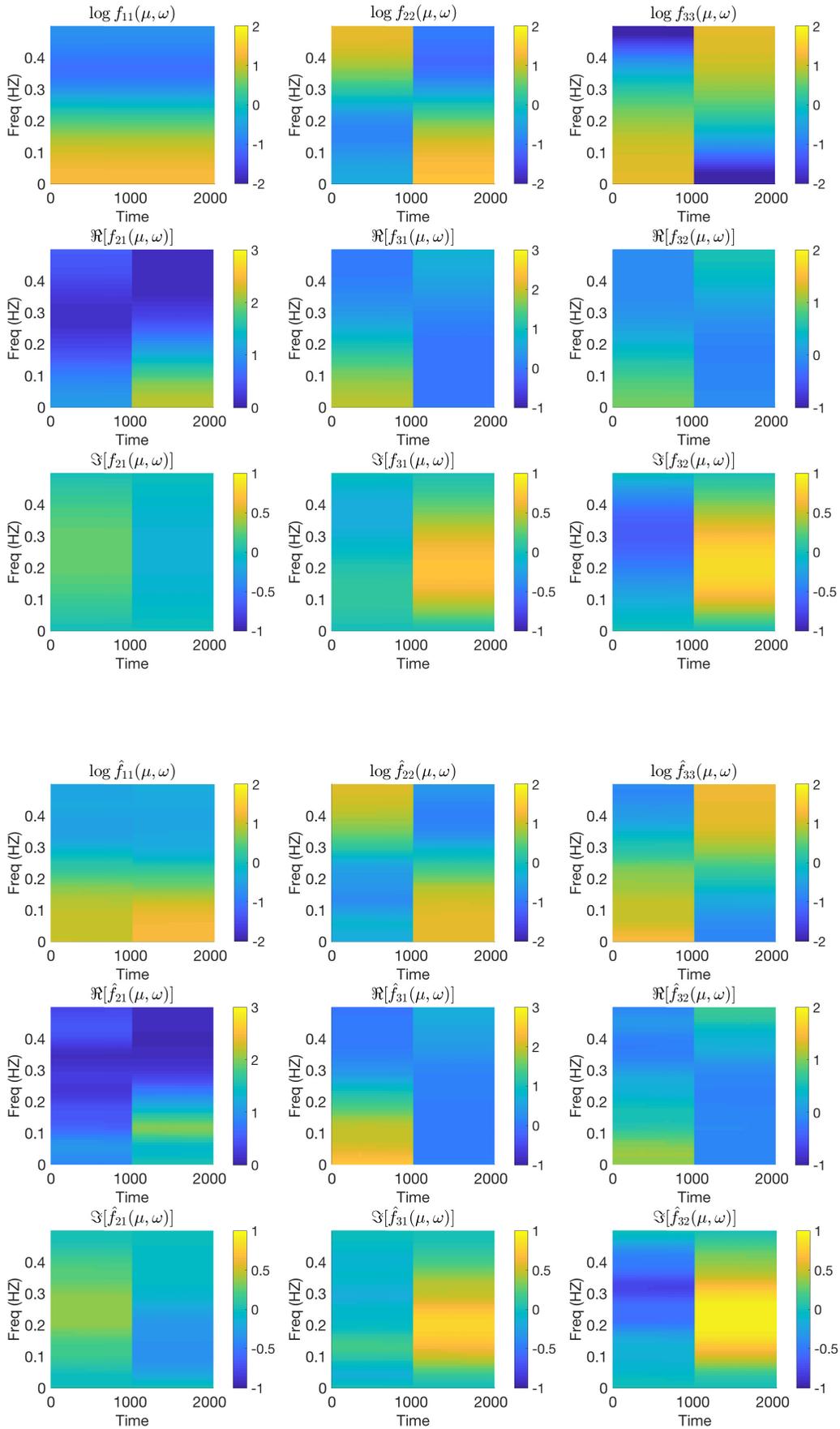


Figure 2: Rows 1-3: the true time-varying log power spectra:  $\log f_{11}(u, \omega)$ ,  $\log f_{22}(u, \omega)$ , and  $\log f_{33}(u, \omega)$ ; and the real and imaginary parts of the cross-spectra:  $f_{21}(u, \omega)$ ,  $f_{31}(u, \omega)$ , and  $f_{32}(u, \omega)$  of process (4). Rows 4-6: the corresponding estimates based on the proposed method.

### 4.1.1 Piecewise Stationary Process

We simulate 100 independent  $P$ -dimensional piecewise stationary time series of length  $T = 2048$  from

$$X_t = \begin{cases} \boldsymbol{\epsilon}_t + \Phi_{11}\boldsymbol{\epsilon}_{t-1} + \Phi_{12}\boldsymbol{\epsilon}_{t-2} & \text{if } 1 \leq t \leq 1024 \\ \boldsymbol{\epsilon}_t + \Phi_{21}\boldsymbol{\epsilon}_{t-1} + \Phi_{22}\boldsymbol{\epsilon}_{t-2} & \text{if } 1025 \leq t \leq 2048. \end{cases} \quad (4)$$

Each of the  $P \times P$  coefficient matrices are block diagonal with  $P/3$  blocks such that  $\Phi_{11} = I_{P/3} \otimes \Phi_{11}^0$ ,  $\Phi_{21} = I_{P/3} \otimes \Phi_{21}^0$ ,  $\Phi_{12} = I_{P/3} \otimes \Phi_{12}^0$ , and  $\Phi_{22} = I_{P/3} \otimes \Phi_{22}^0$ , where  $I_{P/3}$  is a  $P/3 \times P/3$  identity matrix,  $\otimes$  denotes the Kronecker product, and

$$\Phi_{11}^0 = \begin{pmatrix} 0.6 & 0 & 0 \\ 0.2 & -0.6 & 0 \\ 0.1 & 0.2 & 0.6 \end{pmatrix}, \Phi_{21}^0 = \begin{pmatrix} 0.6 & 0 & 0 \\ 0.2 & 0.6 & 0 \\ -0.1 & -0.2 & -0.6 \end{pmatrix}, \Phi_{12}^0 = \Phi_{22}^0 = \begin{pmatrix} 0.3 & 0 & 0 \\ 0 & -0.3 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The white noise  $\boldsymbol{\epsilon}_t \stackrel{i.i.d.}{\sim} N_p(\mathbf{0}, \Omega)$  with  $\Omega = I_{P/3} \otimes \Omega^0$ , where  $\Omega^0$  has 1's on the diagonal and 0.5 off the diagonal. The true  $P \times P$  time-varying power spectrum is  $f(u, \omega) = \Phi(u, \omega)\Omega\Phi(u, \omega)^*$  where  $\Phi(u, \omega) = I + \Phi_{11} \exp(-2\pi i\omega) + \Phi_{12} \exp(-4\pi i\omega)$  for  $u \in [0, 1/2]$ ,  $\Phi(u, \omega) = I - \Phi_{21} \exp(-2\pi i\omega) + \Phi_{22} \exp(-4\pi i\omega)$  for  $u \in (1/2, 1]$  (Priestley, 1981, Chapter 9.4). Some of the true time-varying log power spectra and cross-spectra, and their estimates under the proposed procedure are displayed in Figures 2, which shows that the true power spectrum changes abruptly at  $t = 1024$  and the proposed approach can accurately capture these dynamics.

### 4.1.2 Slowly Varying Process

We consider a slowly varying  $P$ -dimensional vector autoregressive process VAR(1) of length  $T = 2048$  and simulate 100 time series independently from

$$\mathbf{X}_t = \Theta_t \mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t, \quad (5)$$

where  $\boldsymbol{\epsilon}_t \stackrel{i.i.d.}{\sim} N_p(\mathbf{0}, \Omega)$ ,  $\Omega$  is as in Section 4.1.1 and  $\Theta_t = I_{P/2} \otimes \Theta_t^0$ , with

$$\Theta_t^0 = \begin{pmatrix} \theta_1^0(t) & 0.1 \\ 0 & \theta_2^0(t) \end{pmatrix},$$

and  $\theta_1^0(t) = -0.5 + t/T$ ,  $\theta_2^0(t) = 0.7 - 1.4t/T$  for  $t = 1, \dots, T$ . The true  $P \times P$  time-varying power spectrum is  $f(t/T, \omega) = \Theta^{-1}(t/T, \omega)\Omega\Theta^{-1}(t/T, \omega)^*$  where  $\Theta(t/T, \omega) = I +$

$\Theta_t \exp(-2\pi i\omega)$  (Priestley, 1981, Chapter 9.4). Some of the true time-varying spectra and cross-spectra, and their estimates under the proposed procedure are displayed in Figure 3, which shows that the true power spectrum changes smoothly over time and the proposed approach can successfully capture the slowly varying dynamics.

### 4.1.3 Results

The means and standard deviations of the three MISE criteria are presented in Table 1. For both processes, the proposed method outperforms the rolling-window estimator, as the former has smaller MISE,  $\text{MISE}_d$ , and  $\text{MISE}_o$  across both  $P = 24$  and  $P = 48$ . This is particularly true when estimating the off-diagonal elements as  $\text{MISE}_o$  values corresponding to the rolling-window method are 4-16 times larger than those corresponding to the proposed factor model. Obtaining accurate estimates of the off-diagonal elements of a spectral matrix is crucial since they can be used to derive many useful measures of between-series relationships, such as coherence and partial coherence, which provide valuable information for understanding many neurophysiological time series observed simultaneously. Moreover, it should be noted that the proposed method has similar MISE values across the selected number of factors  $Q$ . This can be attributed to using the multiplicative gamma process shrinkage priors, which reduce the sensitivity to the number of factors. We applied the approach outlined in Section 3.2.3 to select  $Q$  adaptively. For Process (4), the average of the estimated numbers of factors are 11.56 and 12.10 corresponding to  $P = 24$  and 48, respectively; for Process (5), the average of the estimated numbers of factors are 13.40 and 14.23 corresponding to  $P = 24$  and 48, respectively. These results are consistent with the findings in Table 1, indicating that the additional gains in estimation accuracy by using  $Q = 16$  is subtle.

Table 1 also reports the mean and standard deviation of run time per iteration of the proposed method for each setting using Matlab 2020a and Windows 10 on a desktop computer with a 3.6 GHz Intel Core i7 processor and 8 GB RAM. Although the run time increases linearly with the number of factors  $Q$  and the dimensionality  $P$ , the proposed method is

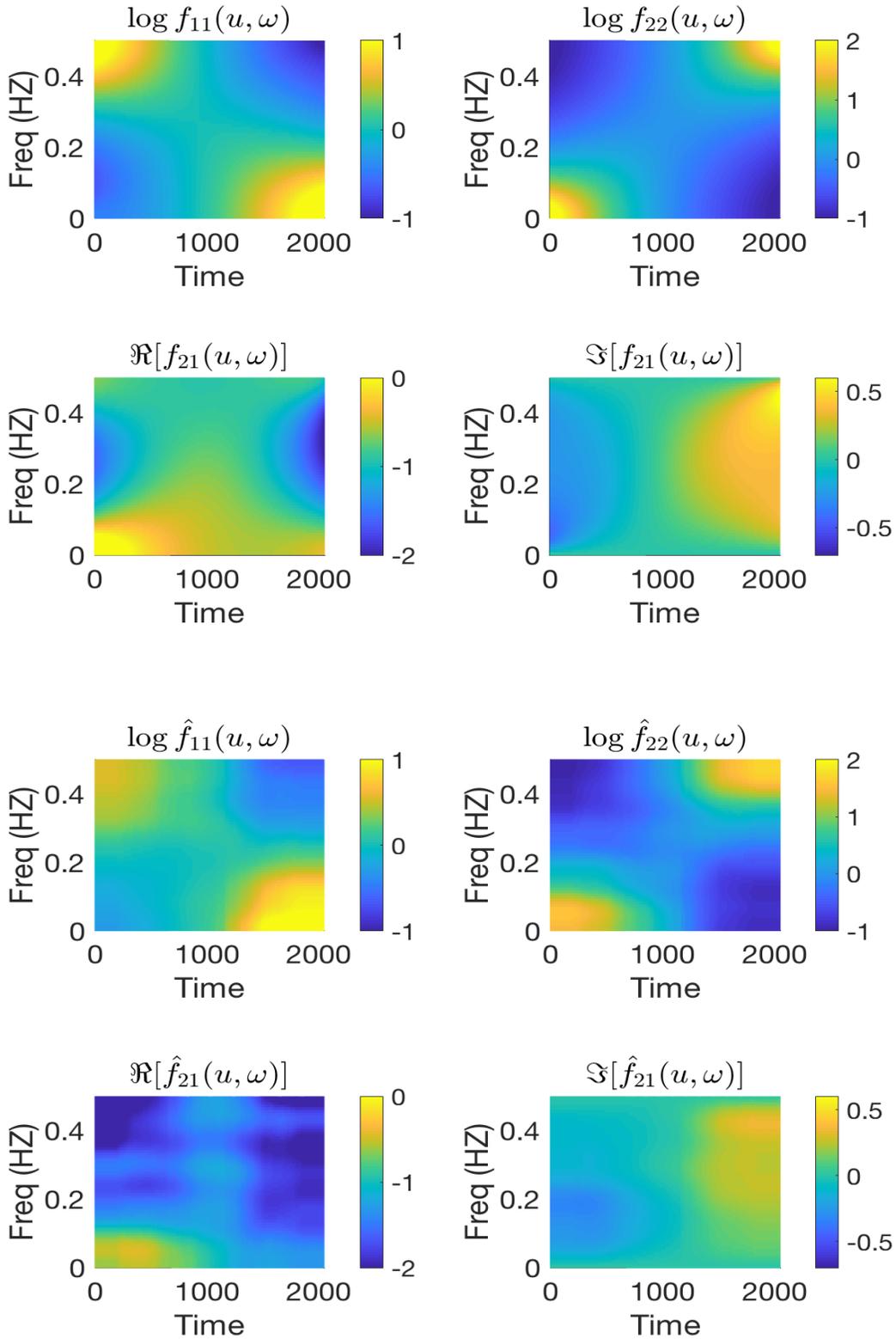


Figure 3: Rows 1-2: the true time-varying log power spectra:  $\log f_{11}(u, \omega)$  and  $\log f_{22}(u, \omega)$ ; and the real and imaginary parts of the cross-spectra  $f_{21}(u, \omega)$  of process (5). Rows 3-4: the corresponding estimates based on the proposed method.

significantly more efficient compared to other Bayesian methods (see Table 3 of Li and Krafty (2019) for a comparison), which are impractical in our simulation settings.

Table 1: Simulation results for the piecewise stationary and slowly varying processes. Based on 100 repetitions, means (standard deviation) of MISE, MISE<sub>d</sub>, MISE<sub>o</sub> obtained through the rolling-window and the proposed factor model, and run time in seconds of the proposed factor model.

Settings	$P$		Rolling-window			Factor Model			
			$B = 64$	$B = 128$	$B = 256$	$Q = 8$	$Q = 10$	$Q = 16$	
Process (4): Piecewise	24	MISE	23.75 (1.01)	14.73 (0.80)	9.29 (0.52)	4.24 (0.31)	3.33 (0.11)	3.27 (0.09)	
		MISE <sub>d</sub>	2.54 (0.17)	1.95 (0.14)	1.91 (0.09)	1.08 (0.06)	0.31 (0.04)	0.24 (0.06)	
		MISE <sub>o</sub>	21.21 (0.90)	12.78 (0.73)	7.38 (0.47)	3.16 (0.26)	3.02 (0.11)	3.02 (0.06)	
		Time				0.41 (0.03)	0.58 (0.06)	0.95 (0.08)	
	48	MISE	92.42 (2.65)	56.19 (2.26)	33.33 (1.58)	10.00 (0.36)	9.31 (0.12)	9.29 (0.14)	
		MISE <sub>d</sub>	5.07 (0.23)	3.89 (0.17)	3.81 (0.10)	3.61 (0.07)	2.79 (0.06)	2.81 (0.05)	
		MISE <sub>o</sub>	87.35 (2.53)	52.30 (2.17)	29.52 (1.53)	6.39 (0.34)	6.35 (0.22)	6.32 (0.21)	
		Time				0.76 (0.05)	0.91 (0.06)	1.62 (0.14)	
	Process (5): Slowly-varying	24	MISE	23.17 (2.58)	17.23 (1.74)	14.91 (1.17)	4.04 (0.06)	3.98 (0.06)	3.97 (0.06)
			MISE <sub>d</sub>	6.13 (0.37)	6.02 (0.30)	6.23 (0.31)	1.72 (0.02)	1.69 (0.02)	1.70 (0.03)
			MISE <sub>o</sub>	17.04 (2.26)	11.21 (1.50)	7.68 (0.99)	2.32 (0.05)	2.29 (0.04)	2.27 (0.02)
			Time				0.38 (0.07)	0.47 (0.15)	0.94 (0.20)
48		MISE	76.92 (7.00)	52.48 (4.99)	37.79 (3.27)	8.46 (0.12)	8.15 (0.08)	8.15 (0.10)	
		MISE <sub>d</sub>	12.30 (0.64)	12.03 (0.47)	13.04 (0.38)	4.24 (0.05)	4.12 (0.03)	4.13 (0.04)	
		MISE <sub>o</sub>	64.63 (6.45)	40.56 (4.61)	24.75 (3.00)	4.22 (0.09)	4.03 (0.06)	4.02 (0.07)	
		Time				0.57 (0.10)	0.76 (0.13)	1.33 (0.26)	

## 4.2 Estimation of Partitions

In this section, we focus on evaluating the frequentist properties of the proposed method in estimating the number and location of partition points. First, to demonstrate that the proposed method can correctly identify a stationary process, we use the first piece of process (4), that is, we simulate one hundred 48-dimensional time series of length  $T = 1024$  independently from

$$\mathbf{X}_t = \boldsymbol{\epsilon}_t + \Phi_{11}\boldsymbol{\epsilon}_{t-1} + \Phi_{12}\boldsymbol{\epsilon}_{t-2}, \quad (6)$$

where  $\Phi_{11}$ ,  $\Phi_{12}$ , and  $\boldsymbol{\epsilon}_t$  are as in Section 4.1.1. Second, we simulate one hundred 48-dimensional time series independently from

$$\mathbf{X}_t = \begin{cases} \boldsymbol{\epsilon}_t^{(1)} + \Phi_{11}\boldsymbol{\epsilon}_{t-1}^{(1)} + \Phi_{12}\boldsymbol{\epsilon}_{t-2}^{(1)} & \text{if } 1 \leq t \leq 500, \\ \boldsymbol{\epsilon}_t^{(1)} + \Phi_{21}\boldsymbol{\epsilon}_{t-1}^{(1)} + \Phi_{22}\boldsymbol{\epsilon}_{t-2}^{(1)} & \text{if } 501 \leq t \leq 1000, \\ \boldsymbol{\epsilon}_t^{(2)} + \Phi_{31}\boldsymbol{\epsilon}_{t-1}^{(2)} + \Phi_{32}\boldsymbol{\epsilon}_{t-2}^{(2)} & \text{if } 1001 \leq t \leq 2000, \\ \boldsymbol{\epsilon}_t^{(2)} + \Phi_{41}\boldsymbol{\epsilon}_{t-1}^{(2)} + \Phi_{42}\boldsymbol{\epsilon}_{t-2}^{(2)} & \text{if } 2001 \leq t \leq 4000, \end{cases} \quad (7)$$

where  $\Phi_{11}$ ,  $\Phi_{12}$ ,  $\Phi_{21}$ , and  $\Phi_{22}$  are as in Section 4.1.1, and  $\Phi_{31} = \Phi_{21}$ ,  $\Phi_{32} = \Phi_{22}$ . We define  $\Phi_{41}$  and  $\Phi_{42}$  as block diagonal matrices such that  $\Phi_{41} = I_{P/3} \otimes \Phi_{41}^0$  and  $\Phi_{42} = I_{P/3} \otimes \Phi_{42}^0$ , where

$$\Phi_{11}^0 = \begin{pmatrix} 1.32 & 0 & 0 \\ 0.2 & -0.6 & 0 \\ 0.1 & 0.2 & 0.6 \end{pmatrix}, \Phi_{42}^0 = \begin{pmatrix} -0.81 & 0 & 0 \\ 0 & -0.3 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The white noise terms  $\epsilon_t^{(1)}$ ,  $\epsilon_t^{(2)}$  are independent zero-mean  $P$ -dimensional Gaussian random variables whose covariance matrices have unit variances and pairwise correlations of 0.5 and 0.9, respectively. This process has both short segments (with 500 observations) and long ones (with 2000 observations), and the change, at  $t = 1000$ , in the off-diagonal elements of the covariance of the error term is subtle compared to the other changes.

Table 2: Simulation results for processes (6) and (7) based on 100 repetitions: means (standard deviations) of the posterior probabilities  $\Pr(L|\mathbf{X})$ , means of MISE obtained through the rolling-window estimator (RW), the local average estimator (LA), and the proposed factor model.

Settings	Number of segments $L$							MISE		
	$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$	Factor Model	RW	LA
Process (6)	0.940 (0.079)	0.057 (0.073)	0.030 (0.007)	0 (0)	0 (0)	0 (0)	0 (0)	4.29 (0.07)		11.42 (1.32)
Process (7)	0 (0)	0 (0)	0.051 (0.182)	0.889 (0.194)	0.060 (0.103)	0 (0)	0 (0)	14.42 (2.42)	63.22 (3.43)	

The proposed method was run using  $Q = 12$  factors for 10,000 iterations with a burn-in of 2,000. Table 2 reports the mean and standard deviation of the estimated posterior probabilities of the numbers of segments,  $\Pr(L|\mathbf{X})$ . For process (6), the proposed method correctly identifies a stationary process with a high probability:  $\widehat{\Pr}(L = 1|\mathbf{X}) = 0.940$ . For process (7), the proposed method also correctly assigns the highest posterior probability to  $L = 4$ , i.e.,  $\widehat{\Pr}(L = 4|X) = 0.889$ . In addition, we compare our method with the other two methods in terms of MISE. To ensure fair comparisons, the local averaging estimator that treats the time series as a stationary process in advance and then smooths the periodogram across frequencies via local averaging was used for Process (6); the rolling window estimator discussed in Section 1 with  $B = 256$  was used for Process (7). Consistent with the results in Section 4.1, the proposed method outperforms the rolling-window estimator for Process (7). More importantly, the proposed method has smaller MISE compared to the local averaging estimator for Process (6) even though it can be considered an oracle estimator by treating

the time series as stationary.

## 5 Analysis of TMS-Evoked High-Density EEG

Abnormal neurophysiological activity is consistently observed in patients with schizophrenia and is associated with poor cognitive ability. To examine biological correlates, identify potential biomarkers, and to guide individualized treatment of mental illness, physicians often use hdEEG to measure electrophysiological activity simultaneously across multiple regions, or channels, of the brain. Transcranial magnetic stimulation (TMS) is a noninvasive method of brain stimulation that uses an insulated coil placed over a chosen location of the scalp to induce a magnetic field that excites cells in the cortex of the brain. Concurrent TMS and hdEEG are used to record TMS-evoked local and general brain activity with millisecond precision. Time-varying spectral analysis of hdEEG during TMS provides important neurobiological information that can be used to assess and guide treatments of psychiatric disorders (Kaskie and Ferrarelli, 2018).

We demonstrate the proposed method through the analysis of hdEEG during TMS from a patient during hospitalization for their first psychotic episode. The epoch considered was taken from 1000 milliseconds before and 1000 milliseconds after a TMS delivery to the primary motor cortical area of the patient’s brain with a sampling rate of 5000 samples per second (5000 Hz). Data were preprocessed using the **TMSEEG** Matlab GUI (Atluri et al., 2016) where signals were downsampled to 1000 Hz. Independent component analysis (ICA) was used to remove pulse and ripple artifacts, IIR bandpass (1-80 Hz) and notch filters (60 Hz) were applied, and signals were referenced and standardized to unit variance. The resulting data, which are displayed in Figure 1, are a 64-dimensional time series, i.e.,  $P = 64$ , of length  $T = 2000$ .

The proposed method was run for 10,000 iterations with burn-in of 5,000 where the number of factors was selected adaptively. The goal of our study is to analyze the entire time-varying spectral matrix,  $f(u, \omega)$ , as well as certain functions of the power spectrum.

Specifically, three types of quantities are of interest. First, the time-varying power spectra  $f_{jj}(u, \omega)$ ,  $j = 1, \dots, P$  and the pairwise squared coherences

$$\rho_{jk}^2(u, \omega) = |f_{jk}(u, \omega)|^2 / [f_{jj}(u, \omega)f_{kk}(u, \omega)], \quad j, k = 1, \dots, P,$$

can be easily estimated. Second, the proposed adaptive Bayesian method allows one to conduct inference on frequency-collapsed functionals. In particular, power within several frequency bands, including the beta band (16-31 HZ), contains important neurophysiological information (Ferrarelli et al., 2018). Frequency-band collapsed measures can be computed as integrals of the power spectra. For example, the time-varying beta-band collapsed spectral matrices are given by

$$f^\beta(u) = \int_{16}^{31} f(u, \omega) d\omega. \quad (8)$$

Third, the time-varying frequency-collapsed squared coherences can be derived to investigate connectivity between brain regions across time at frequency bands of interest. A commonly used measure of local connectivity is the beta-band coherences. The beta-band squared coherence between channels  $j$  and  $k$  is defined as

$$\rho_{jk}^{2,\beta}(u) = \left| \int_{16}^{31} f_{jk}(u, \omega) d\omega \right|^2 / \left\{ f_{jj}^\beta(u) f_{kk}^\beta(u) \right\},$$

where  $f_{jj}^\beta(u) = \int_{16}^{31} f_{jj}(u, \omega) d\omega$  is the time-varying beta-band power of channel  $j$ . The following sections provide the results of our analyses of the TMS-evoked hdEEG.

## 5.1 Time-Varying Power Spectra and Squared Coherences

Figure 4 presents the estimated time-varying log-power spectra and the corresponding pairwise squared coherences from three channels: C1, F1, and P1, which are located in the primary motor cortex, the frontal cortex, and the parietal cortex, respectively. It is clear that the log-power spectra and squared coherences changed abruptly, resulting in three approximately stationary segments, which can be identified as three periods: before TMS (the baseline EEG), immediately after TMS, and recovering from TMS. In general, our results indicate that the estimated power spectra increased at all frequencies after TMS delivery. It

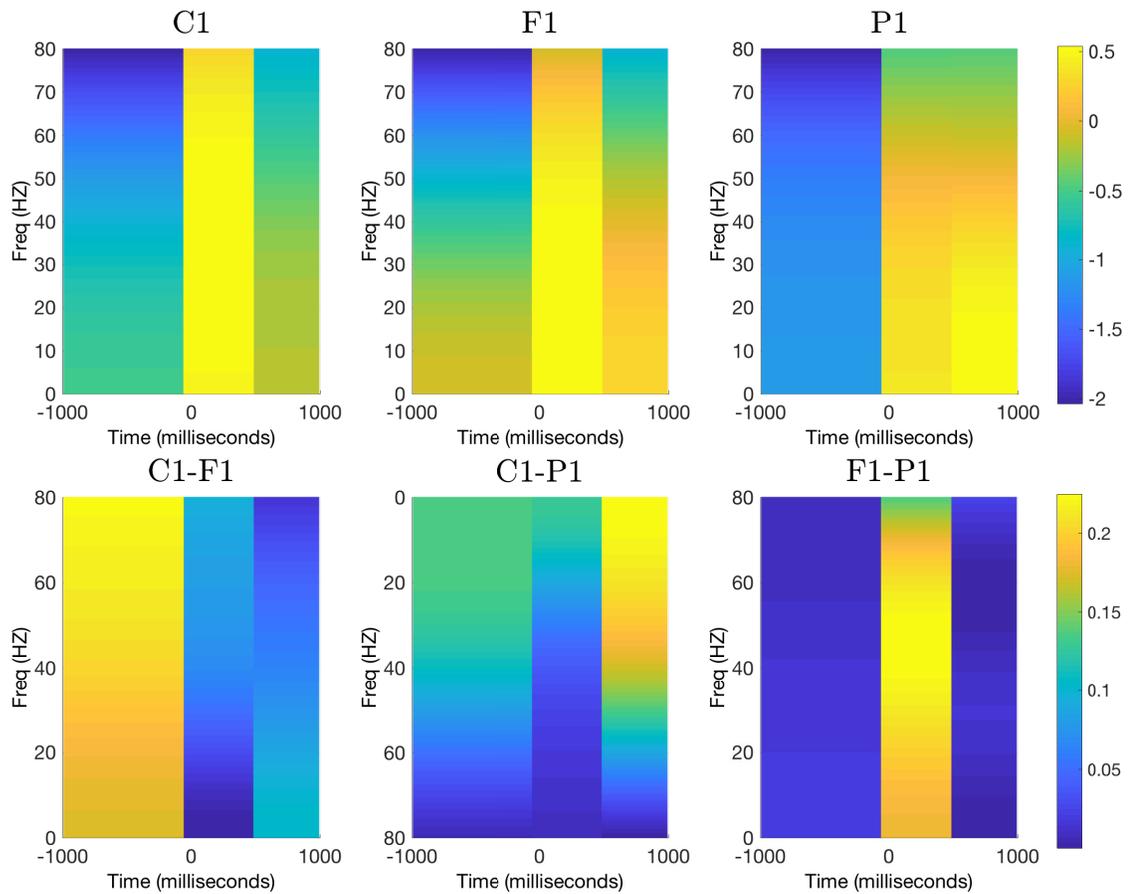


Figure 4: Estimates of the time-varying log spectra (first row) and pairwise squared coherences (second row) of the TMS evoked-EEG in channels C1, F1, and P1.

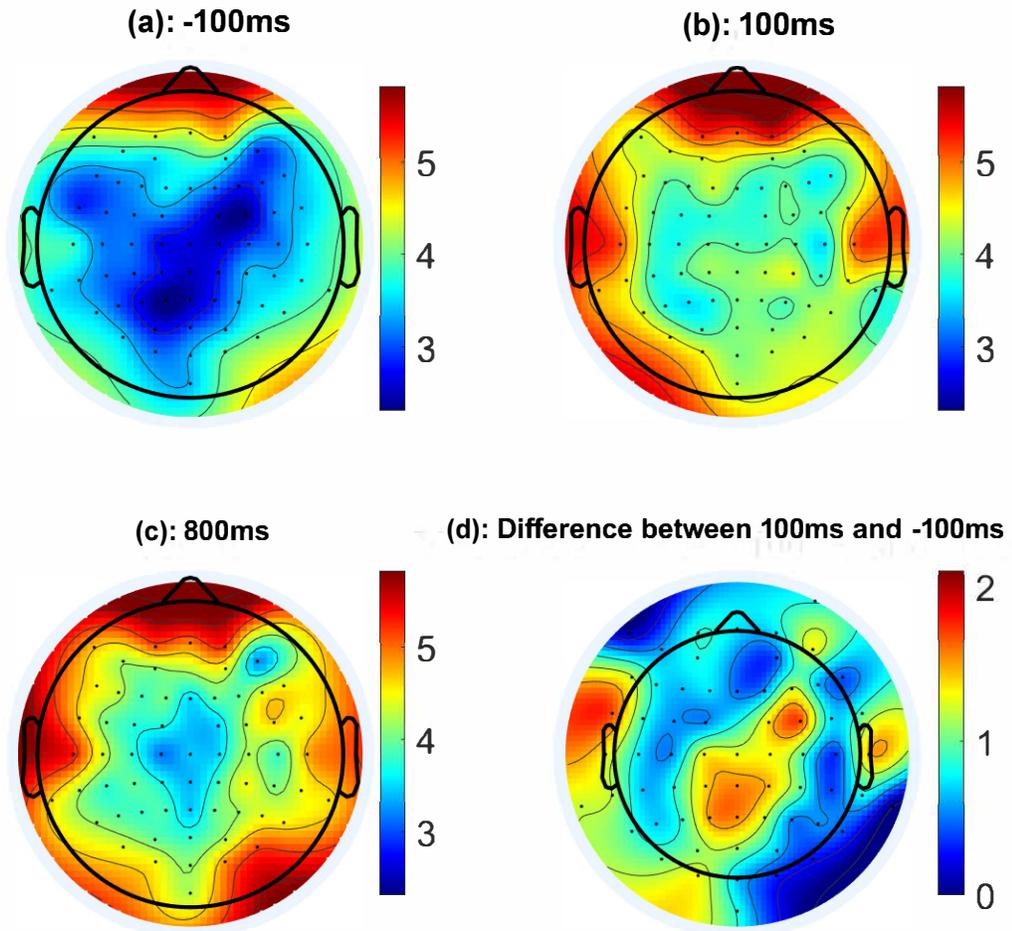


Figure 5: The estimated beta-band log power of 64-channel hdEEG at time points: (a) -100 ms, (b) 100 ms, and (c) 800 ms; and (d) the difference of beta-band log power between time points 100 ms and -100 ms.

should be noted that channel C1, which is located in the primary motor cortex (the region where TMS was induced), had a greater power increase compared to channels F1 and P1. The squared coherence between C1 and F1 and between C1 and P1 decreased dramatically at all frequencies after TMS delivery, indicating that the stimulation restricts the correlation between the primary motor cortical area and other regions of the brain. Interestingly, channels F1 and P1, which are located far apart, experienced an increase in the squared coherence after the TMS delivery. This calls for further research on how channels that are not in actively stimulated regions behave after TMS, which has not been explored in the biomedical literature.

## 5.2 Time-Varying Frequency-Collapsed Power Spectra

Figure 5 (a)-(c) presents topoplots of the estimated beta-band collapsed spectral matrices defined in (8) at different time points:  $t = -100, 100,$  and  $800$  milliseconds (ms). Comparing Figure 5 (a) and (b), we see that the log beta-band power increased drastically after the TMS delivery ( $t = 100$  ms) for all channels. However, as indicated by Figure 5 (d), which displays the differences of log beta-band power between time points 100 ms and -100 ms, the amount of increases of the beta-band power in the primary motor cortex is larger than that in the other areas. Moreover, some prefrontal and parietal cortices had the smallest amount of increase in beta-band power. It should be noted that similar characteristics in gamma-band spectral matrices have been observed in patients with schizophrenia (Huber et al., 2008). Our findings suggest that our patient, who has experienced a first psychotic episode but does not yet meet the clinical requirement for a diagnosis of schizophrenia, exhibits some neurophysiological characteristics that are similar to those in patients with schizophrenia compared to healthy controls, which could potentially serve as a subclinical biomarker.

## 5.3 Time-varying Frequency-Collapsed Coherences

Figure 6 displays the estimated beta-band squared coherence of C1 with respect to all other channels at time points:  $t = -100, 100,$  and  $800$  ms. Broadly, during the baseline period

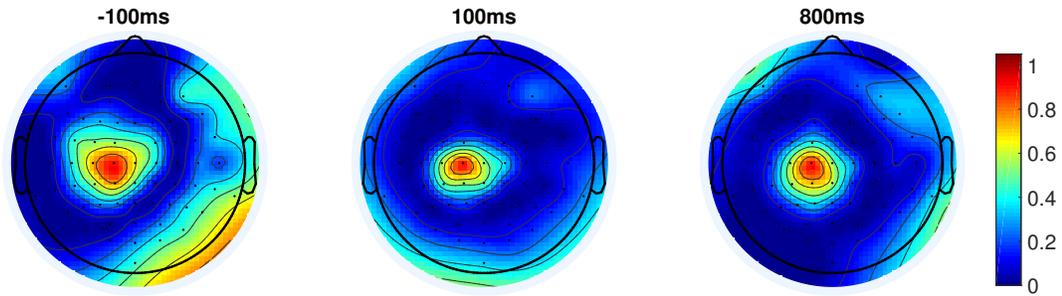


Figure 6: Estimates of beta-band squared coherences of C1 with respect to all other channels at different time points

( $t = -100$  ms), C1 is connected to the channels that are also within the primary motor cortex, as well as to channels within the right hemisphere of the parietal and frontal cortices. However, after the TMS ( $t = 100$  ms) delivery, C1 only retains high beta-band coherences within the primary motor cortex. Beta-band connectivity between C1 and other brain regions decreases or disappears. When  $t = 800$  ms, the connectivity between C1 and the right hemisphere of the parietal and frontal cortices begins to be restored. This phenomenon can be further illustrated in Figure 7, which shows the estimated time-varying beta squared coherences between C1 and Cz, and C1 and TP10 along with 95% credible intervals (as the 2.5th and 97.5th empirical percentiles of the MCMC iterates after the burn-in period). It should be noted that Cz is located close to C1, while TP10 is located at the right hemisphere of the parietal cortical area.

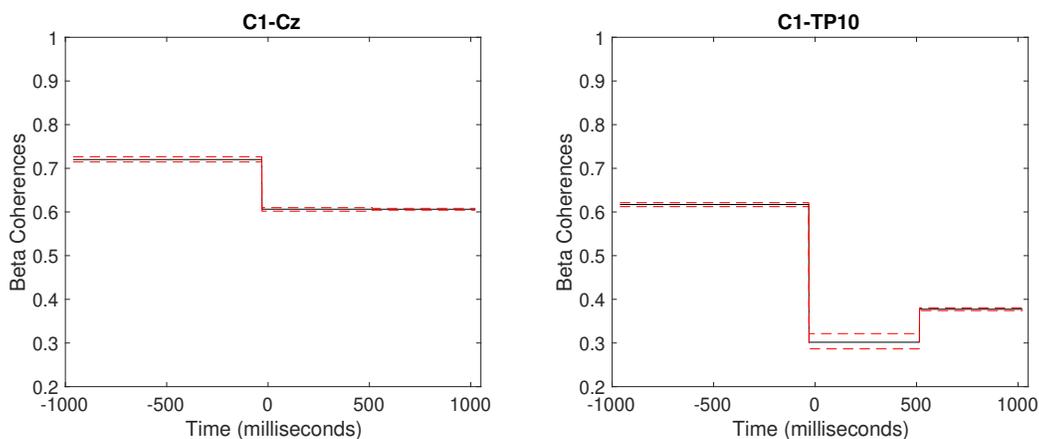


Figure 7: Estimated of beta-band squared coherences of C1 with respect to Cz (left) and TP10 (right) over time with 95% pointwise credible intervals.

## 6 Final Remarks

This article introduces the first adaptive approach to analyzing the time-varying power spectrum of a high-dimensional nonstationary time series. A novel frequency-domain factor model is developed using a prior that is the tensor product of penalized splines and multiplicative gamma process priors. A scalable MCMC algorithm is developed for model fitting, allowing for inference on both abrupt and slowly varying processes. We conclude this article by discussing some limitations and related future extensions. First, the proposed procedure is designed for the analysis of a single high-dimensional time series. However, in many applications, interest lies in the analysis of replicated high-dimensional time series in order to understand how time-varying spectra are associated with other covariates, such as clinical symptoms. A possible extension of the proposed method for the analysis of replicated time series could involve adaptively dividing the grid of time and covariate values into approximately stationary blocks in a manner similar to that proposed in the univariate setting by Bruce et al. (2018). Second, our analysis of high-dimensional time series assumes that all time series have the same sampling rate. However, time series can be sampled at different rates. For example, one could be interested in the joint spectral analysis of other physiological signals, such as heart rate variability, which is typically sampled around 1 Hz, with hdEEG. Future research will focus on developing adaptive spectral analysis of high-dimensional time series with different sampling rates.

## 7 Supplementary Materials

Supplementary materials are available online, including a pdf file that includes details of the sampling scheme, a convergence assessment tool for SAMC and its application to the TMS-evoked EEG data analysis, and additional simulation results. Matlab code for implementing the proposed method are provided.

## References

- Adak, S. (1998), “Time-dependent spectral analysis of non-stationary time series,” *Journal of the American Statistical Association*, 93, 1488–1501.
- Atluri, S., Frehlich, M., Mei, Y., Garcia Dominguez, L., Rogasch, N., Wong, W., Daskalakis, Z., and Farzan, F. (2016), “TMSEEG: A MATLAB-based graphical user interface for processing electrophysiological signals during transcranial magnetic stimulation,” *Front. Neural Circuits*, 10, 78–88.
- Bhattacharya, A. and Dunson, D. B. (2011), “Sparse Bayesian infinite factor models,” *Biometrika*, 98, 291–306.
- Böhm, H. and von Sachs, R. (2009), “Shrinkage estimation in the frequency domain of multivariate time series,” *Journal of Multivariate Analysis*, 100, 913–935.
- Brillinger, D. R. (2002), *Time Series: Data Analysis and Theory*, Philadelphia: SIAM.
- Bruce, S. A., Hall, M. H., Buysse, D. J., and Krafty, R. T. (2018), “Conditional adaptive Bayesian spectral analysis of nonstationary biomedical time series,” *Biometrics*, 74, 260–269.
- Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008), “High-dimensional sparse factor modeling: applications in gene expression genomics,” *Journal of the American Statistical Association*, 103, 1438–1456.
- Dahlhaus, R. (2000), “A likelihood approximation for locally stationary process,” *Annals of Statistics*, 28, 1762–1794.
- Davis, R., Lee, T., and Rodriguez-Yam, G. (2006), “Structural break estimation for nonstationary time series models,” *Journal of American Statistical Association*, 101, 223–239.
- Ensor, K. (2013), “Time series factor models,” *WIREs Computational Statistics*, 5, 97–104.

- Ferrarelli, F., Kaskie, R., Graziano, B., Reis, C., and Casali, A. (2018), “Abnormalities in the evoked frontal oscillatory activity of first-episode psychosis: A TMS/EEG study,” *Schizophrenia Research*, 206, 436–439.
- Fiecas, M., Leng, C., Liu, W., and Yu, Y. (2019), “Spectral analysis of high-dimensional time series,” *Electronic Journal of Statistics*, 13, 4079–4101.
- Fiecas, M., Ombao, H., Linkletter, C., Thompson, W., and Sanes, J. (2010), “Functional connectivity: shrinkage estimation and randomization test,” *NeuroImage*, 5, 1102–1125.
- Fiecas, M. and von Sachs, R. (2014), “Data-driven shrinkage of the spectral density matrix of a high-dimensional time series,” *Electronic Journal of Statistics*, 8, 2975–3003.
- Gelman, A. (2006), “Prior distributions for variance parameters in hierarchical models,” *Bayesian Analysis*, 1, 1–19.
- Geweke, J. and Zhou, G. (1996), “Measuring the price of the arbitrage pricing theory,” *The Review of Financial Studies*, 9, 557–587.
- Guo, W. and Dai, M. (2006), “Multivariate time-dependent spectral analysis using Cholesky decomposition,” *Statistica Sinica*, 16, 825–845.
- Huber, R., Esser, S. K., Ferrarelli, F., Massimini, M., Peterson, M. J., and Tononi, G. (2008), “Reduced evoked gamma oscillations in the frontal cortex in schizophrenia patients: a TMS/EEG study,” *American Journal of Psychiatry*, 165, 996–1005.
- Kaskie, R. E. and Ferrarelli, F. (2018), “Investigating the neurobiology of schizophrenia and other major psychiatric disorders with Transcranial Magnetic Stimulation,” *Schizophrenia Research*, 192, 30–38.
- Kowal, D., Matteson, D., and Ruppert, D. (2017), “A Bayesian multivariate functional dynamic linear model,” *Journal of the American Statistical Association*, 112, 733–744.
- Krafty, R. T. and Collinge, W. O. (2013), “Penalized multivariate Whittle likelihood for power spectrum estimation,” *Biometrika*, 100, 447–458.

- Krafty, R. T., Rosen, O., Stoffer, D. S., Buysse, D. J., and Hall, M. H. (2017), “Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology,” *Journal of the American Statistical Association*, 112, 1405–1416.
- Li, Z. and Krafty, R. T. (2019), “Adaptive Bayesian time-frequency analysis of multivariate time series,” *Journal of The American Statistical Association*, 114, 453–465.
- Liang, F., Liu, C., and Varroll, R. (2007), “Stochastic approximation in Monte Carlo computation,” *Journal of American Statistical Association*, 102, 305–320.
- Macaro, C. and Prado, R. (2014), “Spectral decompositions of multiple time series: A Bayesian nonparametric approach,” *Psychometrika*, 79, 105–129.
- Ombao, H., von Sachs, R., and Guo, W. (2005), “SLEX analysis of multivariate nonstationary time series,” *Journal of the American Statistical Association*, 100, 519–531.
- 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.
- Park, T., Eckley, I. A., and Ombao, H. C. (2014), “Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes,” *IEEE Transactions on Signal Processing*, 62, 5240–5240.
- Prado, R., West, M., and Krystal, A. (2001), “Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag-lead structure,” *Journal of the Royal Statistical Society, Series C*, 50, 95–109.
- Priestley, M. B. (1981), *Spectral Analysis and Time Series, Vol. 1 and 2*, London: Academic Press.
- Rosen, O., Wood, S., and Stoffer, D. (2012), “AdaptSPEC: Adaptive spectral estimation for nonstationary time series,” *Journal of the American Statistical Association*, 107, 1575–1589.

- Sanderson, J., Fryzlewicz, P., and Jones, M. W. (2010), “Estimating linear dependence between nonstationary time series using the locally stationary wavelet model,” *Biometrika*, 97, 435–446.
- Schneider-Luftman, D. and Walden, A. T. (2016), “Partial coherence estimation via spectral matrix shrinkage under quadratic loss,” *IEEE Transactions on Signal Processing*, 64, 5767–5777.
- Shumway, R. and Stoffer, D. (2011), *Time Series Analysis and Its Applications*, Springer: New York.
- Stoffer, D. (1999), “Detecting common signals in multiple time series using the spectral envelope,” *Journal of the American Statistical Association*, 94, 1341–1356.
- Sun, Y., Li, Y., Kuceyeski, A., and Basu, S. (2018), “Large spectral density matrix estimation by thresholding,” *arXiv: 0902.0885 [stat.ME]*.
- Wand, M., Ormerod, J., Padoan, S., and Frühworth, R. (2012), “Mean field variational Bayes for elaborate distribution,” *Bayesian Analysis*, 7, 847–900.
- Watanabe, S. (2010), “Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory,” *Journal of Machine Learning Research*, 11, 3571–3594.
- West, M., Prado, R., and Krystal, A. (1999), “Evaluation and comparison of EEG traces: latent structure in nonstationary time series,” *Journal of the American Statistical Association*, 94, 375–387.
- Zhang, S. (2016), “Adaptive spectral estimation for nonstationary multivariate time series,” *Computational Statistics & Data Analysis*, 103, 330–349.