UNIVERSITY OF CALIFORNIA
SANTA CRUZ

EFFICIENT BAYESIAN INFERENCE ON CLASSICAL AND QUANTILE SPECTRAL ANALYSIS OF MULTIVARIATE TIME SERIES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICAL SCIENCE

by

Zhixiong Hu

September 2022

The Dissertation of Zhixiong Hu
is approved:

Professor Raquel Prado, Chair

Professor Athanasios Kottas

Professor Robert Lund

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Efficient Bayesian Inference on Classical and Quantile Spectral Analysis of Multivariate Time Series

by

Zhixiong Hu

Spectral analysis has been widely used to characterize the properties of one or more time series in the frequency domain. Accurate inference of spectral density matrices is critical for understanding the structure underlying the components of a given multivariate temporal process, and for revealing potential relationships across its components. However, inference of spectral density matrices suffers from the curse of dimensionality. This dissertation first develops methods to estimate the spectral density matrix and functions of this matrix for high-dimensional stationary time series under a Bayesian framework. We consider a Whittle likelihood-based spectral modeling approach and impose a discounted regularized horseshoe prior on the coefficients that define a spline representation of the Cholesky factorization components of the inverse spectral density matrix. Next, we extend the model to estimate the time-varying power spectrum and its functions for high-dimensional nonstationary time series. Under a locally stationary basis representation, two types of priors on basis coefficients are developed: a slow-varying double gamma shrinkage prior is used to induce power spectral estimates to evolve smoothly over time, while a piecewise linear function with a global-local shrinkage prior is proposed for cases in which the power spectral estimates are expected to display abrupt changes. Finally, we further develop methods to conduct quantile spectral analysis for multivariate stationary time series by modeling the matrix of quantile cross-spectral density kernels via its low-rank factorization. Several customized

stochastic gradient variational Bayes (SGVB) approaches, supported by parallel computation and GPU accelerations, are developed to obtain fast approximate posterior inference in all the spectral modeling frameworks mentioned above. Extensive simulation studies and data analyses show that our models and methods for posterior inference are accurate and time efficient. Furthermore, our methods are superior compared to competing methods for standard and quantile spectral analysis of multivariate time series.

To my family, adviser and friends, I couldn't have done this without you. Thank you all for being with me along the way.

Acknowledgments

The completion of this work could not have been possible without the expertise of Professor Raquel Prado, my beloved thesis adviser. Her patient guidance and advise carried me though all the stages of writing my thesis. I would also like to thank my committee members, Professor Athanasios Kottas and Professor Robert Lund, for letting my research life be enjoyable, and for your brilliant comments and suggestions.

I am grateful for my lovely UCSC Statistics Department and the Baskin School of Engineer for providing all the fundamental learning resources. I would like to thank the whole department faculty for their meticulous care and excellent teaching. I would like to thank my graduate advisor, Theo-Alyce Gordon, for her kind assistant in my graduation process. I would also like to show my gratitude to UCSC International Student & Scholar Services team for guiding me through all the important steps to maintain my valid immigration status as an international student.

I would like to thank my family and friends for supporting me throughout my life. Thank my parents, Mr. Hu and Mrs. Cheng, without you I wouldn't even have a chance to study in the US or enter the graduate school. Finally, I would like to thank my girlfriend, Miss Beltran, for sharing with me all the emotional moments in our life and helping me maintain a healthy work-life balance. Thank you all for making me better and better.

Chapter 1

Introduction

Spectral analysis discovers trends, periodic and other characteristics of a time series by representing these features in the frequency domain. For a multivariate temporal process, accurate inference of its spectral density matrix is critical for understanding the structure underlying its components, and for revealing potential relationships across them. A multivariate stationary temporal process has its spectral density matrix remaining unchanged over time, while a nonstationary one has a time-varying power spectrum representing the evolution of the dynamic latent structure underlying the individual components, as well as the potential relationships across different components.

Bayesian spectral analysis has been successfully used to analyze temporal data in several applied settings. For instance, Rosen et al. (2012) analyzed the monthly El Niño–Southern Oscillation (ENSO) data from 1950 to 2010 by modeling the individual time-varying spectra of three ENSO indicators (SOI, Nino3.4 index, and DSLPA) using a Bayesian smoothing spline model for univariate nonstationary time series and found that the spectra of the three indicators were very similar and that they did not vary much over time, concluding that it is very unlikely that there has been a change in the frequency or intensity of ENSO over time.

Li and Krafty (2019) extended the univariate method of Rosen et al. (2012) to analyze multivariate nonstationary time series and modeled the spectral matrix of the three ENSO indicators jointly. The spectral density and coherence estimates provided by Li and Krafty (2019) suggested that a mild change might exist in ENSO during the 1980s. Using the same model, Li and Krafty (2019) also analyzed the neurophysiological activity during sleep based on EEG sleep data and provided evidence from the obtained time-varying power spectra estimates that the transition between sleep stages is sudden. In another neuroscience study, by means of a Bayesian approach that models the factor representation of the time-varying spectral matrix for multivariate nonstationary time series, Li et al. (2021) conducted an analysis of transcranial magnetic stimulation (TMS)-evoked high-density EEG from a patient with schizophrenia during hospitalization to examine biological correlates among brain regions, and concluded from their power spectra estimates that the entire process could be approximately broken into three periods: before TMS, immediately after TMS, and recovering from TMS. The Bayesian approaches for spectral analysis mentioned above are methodologically sound and practically relevant. However, they have limitations when a relative large number of time series components are considered. In the multivariate setting, as the number of components increases, the size of the spectral density matrix grows quadratically, making estimation and inference rather challenging. The majority of the previous methods focus on low dimensional modeling when the time series dimension $P < 10$, which are quite time-consuming for a larger P. However, efficient high dimensional spectral analysis is important for many real-world applications. For instance, in EEG data, the number of brain channels can be over 60, having a $P > 60$. This work is motivated by the need to develop scalable Bayesian methodologies for accurate and time-efficient spectral analysis

of high-dimensional time series.

The first contribution of this work is to develop a novel Bayesian model to obtain scalable and accurate inference on the spectral density matrix, and functions of this matrix, for high-dimensional stationary time series. So far, several frequency domain methods have been developed for Bayesian spectral analysis of stationary multivariate time series. Rosen and Stoffer (2007) proposed a Bayesian approach that uses Markov chain Monte Carlo (MCMC) techniques to obtain posterior inference in a model that considers smoothing splines to represent the real and imaginary components of the modified complex Cholesky decomposition of the inverse spectral density matrix. In earlier related work, Dai and Guo (2004) used a Cholesky factorization of the spectral density matrix. More recently, Krafty and Collinge (2013) derived an approach that considers a penalized Whittle log-likelihood to incorporate regularization on the multivariate power spectra, allowing for varying levels of smoothness among power spectral components. Meier et al. (2020) modeled the spectral density matrix with matrix-valued mixture weights induced by a Hermitian positive definite Gamma process, effectively extending the univariate Bernstein-Dirichlet process prior approach of Choudhuri et al. (2004) to the multivariate case. These approaches provide key modeling and inference tools in the spectral domain, however they also have very high computational costs and hence, lack of scalability in high-dimensional and even in relatively low or moderate-dimensional settings that involve joint analysis of a collection of time series components.

Following an approach similar to that of Rosen and Stoffer (2007), we model each component in the Cholesky decomposition of the inverse spectral density matrix via smoothing splines. We use a rich set of spline basis to represent the inverse spectral density matrix, preserving model flexibility. In addition, in order

to allow for model flexibility while also avoiding overfitting, we consider a modified version of the regularized horseshoe priors of Piironen and Vehtari (2017), referred to as discount regularized horseshoe priors, on the parameters that define the spline representation of the inverse of the spectral density matrix. Our discounted regularized horseshoe (DRH) prior includes a pre-specified discount factor on the hyper-parameter that controls the local shrinkage parameters on spline coefficients. The proposed prior varies the degree of regularization according to the smoothness of the basis functions. In addition to providing various levels of smoothness for the spectral components, the global and local shrinkage parameters in the proposed prior add significant model flexibility. A detailed discussion of the proposed method and its comparison to the alternatives can be found in Chapter 2.

The second contribution of this work is our novel Bayesian approach to obtain estimates on the time-varying power spectra for multivariate nonstationary time series. Several methods have been developed for the spectral analysis of multivariate nonstationary time series. The rolling-window procedure (Priestley, 1981) partitions the time series into fixed time blocks and uses local averaging (Shumway and Stoffer, 2011) to obtain estimates of local power spectra by smoothing the local periodogram. Similar to the rolling-window approach that assumes second-order frequency domain structures evolve smoothly over time, parametric (Dahlhaus, 2000; Guo and Dai, 2006) and nonparametric (Sanderson et al., 2010; Park et al., 2014) methods have been developed to capture changes of the power spectra over time. Some approaches divide a time series into approximately stationary segments and obtain estimates of local spectra within segments (Ombao et al., 2005; Davis et al., 2006), while other approaches, such as the adaptive Bayesian approaches of Zhang (2016); Li and Krafty (2019) and Li et al. (2021)

can approximate both abrupt and slowly varying changes also by partitioning the time series but treating the number of components in the partition, as well as the location of such components, as unknown parameters and obtain posterior estimates using Markov chain Monte Carlo (MCMC) techniques.

In practice, a drawback for the adaptive Bayesian approaches of Zhang (2016); Li and Krafty (2019) is that they are computationally expansive for high-dimensional analysis due to the large computational cost of estimating the unknown partition points. As an attempt to improve model scalability, Li et al. (2021) proposed a frequency domain locally stationary factor model which represents the spectral matrices in terms of low-rank parsimonious factor loading matrices whose real and imaginary parts are modeled independently with using the tensor product of penalized smoothing spline priors. Nevertheless, this approach still requires a large amount of run time to infer the unknown number of partition components and their corresponding locations. Additionally, the method may not be accurate enough due to the fact that estimating the number of partitions brings large model uncertainty during posterior inference, and it uses a single model to handle two types of intrinsically different nonstationary processes (slow varying processes and abrupt changing processes).

We use a Cholesky factorization of the inverse of the local spectral matrices and model the Cholesky components via smoothing splines (Zhang, 2016; Li and Krafty, 2019) to guarantee positive-definiteness and flexible smoothing of the power spectra. As mentioned earlier, previous Bayesian approaches (Rosen et al., 2012; Zhang, 2016; Li and Krafty, 2019; Li et al., 2021) consider time partitions as unknown parameters and estimate them jointly with the remaining model parameters, which leads to a highly expensive inference process from the computational viewpoint. Instead, to improve computation efficiency, we follow

Davis et al. (2006) and Guo and Dai (2006) and use pre-specified equally-spaced partitions of the time series process to obtain a locally stationary Whittle likelihood approximation. Within each partition component we assume a Cholesky decomposition of the local spectral density similar to that developed in Chapter 2 for the stationary case, and use a spline representation of the elements of this decomposition. Under this modeling framework, we consider two types of priors on the spline coefficients in order to handle both slow varying and abrupt changing time series. Slow-varying double gamma shrinkage priors (SV-Prior) inspired by Bitto and Frühwirth-Schnatter (2019) are used to allow the power spectra to evolve smoothly over time (e.g., for cases in which the power spectra changes slowly over time, or when it remains constant over time). Alternatively, piecewise linear functions inspired by Hosseini et al. (2021) are assumed on the time-varying spline coefficients with shrinkage priors (AC-Prior) to tackle abrupt changing patterns. An efficient model selection procedure based on the deviance information criterion (DIC; Spiegelhalter et al., 2014) is used to choose between the aforementioned models when analyzing a particular dataset. We note that the likelihood function used for posterior inference and model selection is approximated based on a given partition via products of local Whittle likelihood approximations. Details of the model specification and evaluation are included in Chapter 3.

The third contribution of this work is the development of a novel methodology for Bayesian quantile spectral analysis for multivariate stationary time series. Classical spectral analysis approaches as those mentioned above are limited to modeling 1st- and 2nd-order dynamics. Instead, quantile-based spectral methods are more robust to outliers and capture additional information about temporal dependence, such as conditional skewness, kurtosis or tail behavior (e.g., see discussion and examples in Birr et al., 2017 and Zhang, 2019). Classical spectral

inference is based on using the periodogram as input data. Quantile spectral inference relies on using non-standard versions of periodograms as input data, with some approaches based on the so called quantile periodograms, or integrated versions of these (Hagemann, 2013; Hong, 2000), while other approaches are based on the so called rank-based copula Laplace periodograms (Dette et al., 2015; Kley et al., 2016). Quantile periodograms were developed in Li (2012) based on the concept of Laplace periodograms (Li, 2008), which were used to handle heavy-tailed noise and nonlinear distortions in data acquisition and transmission systems. Dette et al. (2015) established the matrix of quantile cross-spectral density kernels (CSDKs) and introduced a class of estimators of CSDKs, the so-called smoothed rank-based Laplace periodograms (RLPs). Kley et al. (2016) provided a detailed asymptotic analysis of the smoothed RLPs for univariate time series, which was later extended to construct quantile spectral estimates of multivariate time series by Baruník and Kley (2019).

Previous research on Bayesian quantile spectral analysis is fairly limited. Zhang (2019) proposed a Bayesian copula spectral analysis for univariate stationary time series, where the Bayesian spectral approach of Rosen and Stoffer (2007) is used to model the CSDK matrix and the RLPs based on the discrete Fourier transform (DFT) of the clipped time series (Hong, 2000; Kley et al., 2016). However, the approach of Zhang (2019) only focuses on univariate time series and potential multivariate extensions lack feasibility for two main reasons reasons. First, it models the modified complex Cholesky decompostion of the inverse CSDK matrix, whose size grows quadratically with respect to the number of the chosen quantiles and the dimension of time series. Second, the Markov chain Monte Carlo (MCMC) posterior inference scheme of Zhang (2019) incurs in large computational costs even for univariate quantile spectral analysis.

In Chapter 4, we develop a novel scalable Bayesian approach for quantile spectral analysis of multivariate stationary time series that assumes a low-rank factorization of the target CSDK matrix and shrinkage priors on the parameters that define this factorization to avoid overfitting More specifically, our approach utilizes the low-rank factor model proposed in Li et al. (2021) to provide a parsimonious representation of the target complex-valued matrix of quantile cross-spectral density kernels defined by a complex-valued loading matrix, which is then modeled via smoothing splines. Furthermore, similar to our approach in Chapter 2, in order to allow for model flexibility while also avoiding overfitting, we consider discount regularized horseshoe (DRH) priors on the parameters that define the spline representation of the complex-valued loading matrix of the factor model. The DRH prior, imposed on the spline coefficients of the basis representation of the loading factor matrix, includes a pre-specified discount factor on the hyper-parameter that controls the local shrinkage parameters on spline coefficients, allowing us to vary the degree of regularization according to the smoothness of the basis functions. The proposed model is referred to as FM-DRH.

Finally, the last contribution of this work is our customized variational Bayes algorithms for fast, flexible posterior inference of all aforementioned models we have established (see details in Sections 2.2, 3.2, and 4.2). More specifically, in order to obtain model scalability while preserving the accuracy, we develop a set of variational inference (VI) (Blei et al., 2017) approaches, which are facilitated and scaled up by stochastic optimization (Robbins and Monro, 1951). We adapt the stochastic gradient variational Bayes (SGVB) approach of Kingma and Welling (2013) and Ong et al. (2018) that utilizes the gradient of the unnormalized log-posterior distribution through optimization. Assuming a multivariate Gaussian variational distribution, SGVB relies on the reparameterization trick to yield a

simple differentiable unbiased and low-variance gradient estimator (Xu et al., 2019; Domke, 2019). In comparison to MCMC, SGVB tends to be much faster and scalable for large-scale inference problems. The proposed framework is highly parallelizable and compatible with GPU acceleration, providing a computationally efficient framework for spectral modeling and inference of high-dimensional time series. We present extensive simulation studies in Sections 2.3, 3.3, and 4.3 that show that our inferences are accurate and time efficient, and that the proposed methods are superior to competing approaches for the multivariate stationary case, the multivariate non-stationary case and the multivariate quantile case.

The details of these topics are presented in the following chapters, organized as follows. Chapter 2 develops the novel model with customized SGVB algorithm for Bayesian inference for spectral analysis of multivariate stationary time series. In Chapter 3, we introduce the proposed modeling procedure for spectral analysis of multivariate nonstationary time series. In Chapter 4, we develop the FM-DRH model as well as the corresponding SGVB algorithm for quantile spectral analysis. Finally, in the last chapter, we conclude with a brief review of the methods covered and provide future directions.

Chapter 2

Spectral Analysis of Multivariate Stationary Time Series

In this chapter, we introduce a novel efficient Bayesian framework for spectral analysis of multivariate stationary time series. The remainder of the chapter is organized as follows. Section 2.1 specifies the model and priors used for spectral analysis of stationary multivariate time series. Section 2.2 describes the proposed stochastic gradient variational Bayes scheme for fast and scalable posterior inference. Section 2.3 reports results of extensive simulation studies that illustrate the accuracy and scalability of the proposed approach. Section 2.4 applies the proposed framework to the analysis of multi-location wind speed time series data from the Iowa Environmental Mesonet (IEM) database (Todey et al., 2002). This section also presents the results of the analysis of a multi-channel electroencephalogram dataset from the UCI Machine Learning Repository (Dua and Graff, 2017).

2.1 Model Specification

2.1.1 The Modified Complex Cholesky Decomposition

Let $\{\mathbf{x}_t\}$ denote a zero-mean P-dimensional time series process. Consider a realization $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ of such process, or $\mathbf{x}_{1:n}$. These data can be represented in the frequency domain via the discrete Fourier transform (DFT),

$$\boldsymbol{y}(\nu_k) = n^{-\frac{1}{2}} \sum_{t=1}^{n} \mathbf{x}_t \, \mathrm{e}^{-2\pi \, \mathrm{i} \, \nu_k t},$$

where $\nu_k = k/n$ denote Fourier frequencies, $k = 0, 1, ..., n-1$. Because the discrete Fourier transform is an even function of ν, there are only $[n/2]$ distinct $\boldsymbol{y}(\nu_k)$. If we assume $\{\mathbf{x}_t\}$ to be stationary with $P \times P$ autocovariance matrix, $\boldsymbol{\Gamma}(h) = \{\gamma_{il}(h)\}$, satisfying $\sum_{h=-\infty}^{\infty} |\gamma_{il}(h)| < \infty$ for all $i, l = 1, ..., P$ where $\gamma_{il}(h) = \mathrm{E}[\mathbf{x}_{i,t} \, \mathbf{x}_{l,t+h}]$, there exists a $P \times P$ spectral density matrix

$$\boldsymbol{f}(\nu) = \sum_{h=-\infty}^{\infty} \boldsymbol{\Gamma}(h) \, \mathrm{e}^{-2\pi \, \mathrm{i} \, \nu h},$$

where $\nu \in [-0.5, 0.5]$ is a frequency measured in cycles per unit time. Note that $\boldsymbol{f}(\nu) = \boldsymbol{f}^*(-\nu)$ where $*$ denotes the conjugate transpose. Additionally, $\boldsymbol{f}(\nu)$ is also positive definite. The diagonal entries of $\boldsymbol{f}(\nu)$, $\boldsymbol{f}_{jj}(\nu)$ for $j = 1, ..., P$, correspond to the spectral densities of each of the P components of the time series $\mathbf{x}_t$. The off-diagonal elements of the spectral matrix can be combined to define the squared coherence

$$\rho_{il}^2(\nu) = \frac{|\boldsymbol{f}_{il}(\nu)|^2}{\boldsymbol{f}_{ii}(\nu)\boldsymbol{f}_{ll}(\nu)},$$

with $\rho_{il}^2(\nu) \in [0,1]$, where i, $l = 1$, ..., P with $i \neq l$. The coherence is a frequency-domain analog of a the correlation between two components. It measures the strength of the association of the different time series components for a given frequency ν. Squared coherence values closer to one indicate strong association, while values close to zero show little or no association between components i and l.

Following an approach similar to that in Rosen and Stoffer (2007), we use a Cholesky factorization of the inverse of the spectral density matrix and then represent the components of this factorization via smoothing splines, as follows. Let $\mathbf{y}_k \equiv \boldsymbol{y}(\nu_k)$ and $\mathbf{f}_k \equiv \boldsymbol{f}(\nu_k)$ with $\boldsymbol{f}$ the spectral density matrix. Then, the multivariate extension of the Whittle likelihood approximation (Whittle, 1957) is given by

$$L(\mathbf{y}_{1:N}; \mathbf{f}_{1:N}) \approx \prod_{k=1}^{N} \det(\mathbf{f}_k)^{-1} \exp\left(-\mathbf{y}_k^* \mathbf{f}_k^{-1} \mathbf{y}_k\right),$$

with $N = [n/2]$. To ensure that $\boldsymbol{f}$ is positive definite, $\mathbf{f}_k^{-1}$ can be modeled through the modified complex Cholesky factorization:

$$\mathbf{f}_k^{-1} = \mathbf{T}_k^* \mathbf{D}_k^{-1} \mathbf{T}_k, \tag{2.1}$$

where $\mathbf{D}_k = \text{diag}\left(\delta_{1k}^2, \ldots, \delta_{pk}^2\right)$, and $\mathbf{T}_k$ is a complex unit lower triangular matrix such that

$$\mathbf{T}_k = \begin{pmatrix} 1 & & & & \\ -\theta_{21}^{(k)} & 1 & & & \\ -\theta_{31}^{(k)} & -\theta_{32}^{(k)} & 1 & & \\ \vdots & \vdots & & \ddots & \\ -\theta_{p1}^{(k)} & -\theta_{p2}^{(k)} & \ldots & -\theta_{p,p-1}^{(k)} & 1 \end{pmatrix}. \tag{2.2}$$

Consequently, the likelihood can be rewritten as:

$$L\left(\mathbf{y}_{1:N};\mathbf{f}_{1:N}\right) \propto \prod_{k=1}^{N}\prod_{j=1}^{P}\left[\delta_{jk}^{-2}\exp\left(-\frac{|y_{jk}-\sum_{l=1}^{j-1}\theta_{jl}^{(k)}y_{lk}|^2}{\delta_{jk}^2}\right)\right], \tag{2.3}$$

with y_{jk} denoting the j-th dimension of $\mathbf{y}_k$, and $|\cdot|$ denoting the absolute value. Note that (2.3) decomposes the likelihood into P components. For $j = 1, ..., P$, each component is written as

$$L_j\left(\mathbf{y}_{1:N};\mathbf{f}_{1:N}\right) = \prod_{k=1}^{N}\left[\delta_{jk}^{-2}\exp\left(-\frac{|y_{jk}-\sum_{l=1}^{j-1}\theta_{jl}^{(k)}y_{lk}|^2}{\delta_{jk}^2}\right)\right], \tag{2.4}$$

such that $L\left(\mathbf{y}_{1:N};\mathbf{f}_{1:N}\right) \propto \prod_{j=1}^{P} L_j\left(\mathbf{Y};\mathbf{f}_{1:N}\right)$ and so, posterior inference can be obtained in parallel over j.

2.1.2 Prior Specification

Rosen and Stoffer (2007) used a basis representation to model the entries of the $\mathbf{D}_k$'s and $\mathbf{T}_k$'s. Our proposed approach considers the same representation, which is briefly described below, but uses a different prior structure on the parameters that define this representation. More specifically, the Demmler-Reinsch basis functions (Eubank, 1999) are used to represent each $\log\delta_{jk}^2$ and the real and imaginary parts of each $\theta_{jl}^{(k)}$ in terms of M-truncated smoothing splines:

$$\begin{aligned}
\log\delta_{j,k}^2 &= \gamma_{j,0} + \gamma_{j,1}\nu_k + \sum_{s=1}^{M-1}\psi_s\left(\nu_k\right)\gamma_{j,s+1},\\
\Re(\theta_{jl}^{(k)}) &= \alpha_{jl,0} + \alpha_{jl,1}\nu_k + \sum_{s=1}^{M-1}\psi_s\left(\nu_k\right)\alpha_{jl,s+1},\\
\Im(\theta_{jl}^{(k)}) &= \beta_{jl,0} + \beta_{jl,1}\nu_k + \sum_{s=1}^{M-1}\psi_s\left(\nu_k\right)\beta_{jl,s+1}\,,
\end{aligned} \tag{2.5}$$

where $\psi_s(\nu_k) = \sqrt{2}\cos\left(s\pi\,\nu_k\right)$. M is a pre-specified constant determining the

number of basis functions to be included and therefore the model flexibility. Different options of spline basis (e.g., B-spline, P-spline) satisfying user-specific needs can be easily incorporated with our framework (for instance, a spectra lying within certain frequency bands can be handled by B-spline basis with selective knots). Let

$$\boldsymbol{\gamma}_j = (\gamma_{j,0}, ..., \gamma_{j,M})', \quad \boldsymbol{\alpha}_{jl} = (\alpha_{jl,0}, ..., \alpha_{jl,M})', \quad \boldsymbol{\beta}_{jl} = (\beta_{jl,0}, ..., \beta_{jl,M})',$$
$$\mathbf{X}_k = (1, \nu_k, \psi_1(\nu_k), ..., \psi_{M-1}(\nu_k)).$$

Then, the aforementioned smoothing splines in (2.5) can be rewritten as:

$$\log \delta_{j,k}^2 = \mathbf{X}_k \boldsymbol{\gamma}_j, \quad \Re(\theta_{jl}^{(k)}) = \mathbf{X}_k \boldsymbol{\alpha}_{jl}, \quad \Im(\theta_{jl}^{(k)}) = \mathbf{X}_k \boldsymbol{\beta}_{jl}. \tag{2.6}$$

For notation brevity, let $\boldsymbol{\alpha}_{j\cdot}$ denote $\{\boldsymbol{\alpha}_{j1}, ..., \boldsymbol{\alpha}_{jj-1}\}$ and $\boldsymbol{\beta}_{j\cdot}$ denote $\{\boldsymbol{\beta}_{j1}, ..., \boldsymbol{\beta}_{jj-1}\}$. By plugging (2.6) into (2.4), the above $L_j\left(\mathbf{y}_{1:N}; \mathbf{f}_{1:N}\right)$ can be reparameterized as $L_j\left(\mathbf{y}_{1:N}; \boldsymbol{\gamma}_j, \boldsymbol{\alpha}_{j\cdot}, \boldsymbol{\beta}_{j\cdot}\right)$, where for every $j = 1, ..., P$:

$$\log L_j\left(\mathbf{y}_{1:N}; \boldsymbol{\gamma}_j, \boldsymbol{\alpha}_{j\cdot}, \boldsymbol{\beta}_{j\cdot}\right) = -\sum_{k=1}^{N} \left[\mathbf{X}_k \boldsymbol{\gamma}_j + \frac{\left| y_{jk} - \sum_{l=1}^{j-1} \left(\mathbf{X}_k \boldsymbol{\alpha}_{jl} + \mathrm{i}\, \mathbf{X}_k \boldsymbol{\beta}_{jl}\right) y_{lk} \right|^2}{\mathrm{e}^{\mathbf{X}_k \boldsymbol{\gamma}_j}} \right]. \tag{2.7}$$

Note that in the case of $k = 1$, L_1 depends solely on $\boldsymbol{\gamma}_1$.

Our goal is to obtain posterior estimates of the $\boldsymbol{\gamma}_j$'s, $\boldsymbol{\alpha}_{jl}$'s, and $\boldsymbol{\beta}_{jl}$'s. Hence, priors on spline coefficients are needed. One option is to use normal priors as in Rosen and Stoffer (2007); Zhang (2019). We show later via simulation study that such priors can be inadequate to handle cases in which some of the time series components require a large number of bases in their representation, while others require a small number, thus failing to provide a flexible modeling framework that can adequately capture this situation. An ideal prior setting should allow the selection of enough basis terms to achieve a good fit, while also avoiding overfitting

for each component. To achieve this, our prior setting is based upon the so-called regularized horseshoe prior (Piironen and Vehtari, 2017). This is a global-local shrinkage prior with a global parameter that provides shrinkage towards zero for all the components sharing this parameter, and local, or component-specific parameters that allow some of individual components to escape from the shrinkage. Xie (2018) first explored the use of this prior in spectral analysis of univariate stationary time series. Here we extend this approach to develop a novel model and related inference procedure for accurate and computationally efficient Bayesian multivariate spectral analysis.

The proposed prior distribution structure is as follows. First, we impose no shrinkage on $\gamma_{j,0}$ and $\gamma_{j,1}$ and assume that $\gamma_{j,s} \sim N(0, 10)$ for all j and $s = 0, 1$. This is based on the idea that the individual spectral densities will have at the very least a baseline basis representation that is non-zero on the intercept and slope terms, and possibly a more sophisticated structure that can be captured by the regularized horseshoe prior on the remaining parameters as explained below. We have found that a prior variance of 10 (or any value within $[10, 10^5]$ that leads to a non-informative prior) for these terms typically works well in practice. Then, for $s = 2, \ldots, M$ we assume

$$\gamma_{j,s}|\lambda_{js}, \tau_j \sim \mathrm{N}\left(0, \frac{c^2\tau_j^2\lambda_{js}^2}{c^2 + \tau_j^2\lambda_{js}^2}\right),$$

where c is a fixed constant. We discuss the selection of c below. Similarly, for $s = 0, \ldots, M$,

$$\alpha_{jl,s}|\lambda_{jls,(re)}, \tau_{jl} \sim \mathrm{N}\left(0, \frac{c^2\tau_{jl}^2\lambda_{jls,(re)}^2}{c^2 + \tau_{jl}^2\lambda_{jls,(re)}^2}\right),$$
$$\beta_{jl,s}|\lambda_{jls,(im)}, \tau_{jl} \sim \mathrm{N}\left(0, \frac{c^2\tau_{jl}^2\lambda_{jls,(im)}^2}{c^2 + \tau_{jl}^2\lambda_{jls,(im)}^2}\right).$$

The parameters τ_j and τ_{jl} above control the the overall roughness of the spec-

tral density matrix for each component j as a function of the frequency. We assume $\tau_j, \tau_{jl} \sim \mathrm{C}^+(0, c_\tau)$, where C^+ denotes Half-Cauchy distribution and c_τ is a fixed constant. We also have local parameters λ_{js}, $\lambda_{jls,(re)}$ and $\lambda_{jls,(im)}$ which adjust the roughness as a function of the frequency according to individual Demmler-Reinsch basis with $\lambda_{js}, \lambda_{jls,(re)}, \lambda_{jls,(im)} \sim \mathrm{C}^+(0, c_s)$. Here, c_s is a pre-specified "discount effect" function varying with s. The choices of c_τ and c_s are described below.

The prior structure above has the following features. First, as it was previously mentioned, it assumes that the diagonal entries of a spectral density matrix are all non-zero, and so it does not impose a shrinkage prior on $\gamma_{j,0}$ and $\gamma_{j,1}$. On the other hand, a global-local shrinkage prior is proposed on the rest of $\gamma_{j,s}$'s.

Second, the sparsity of $\mathbf{f}_k^{-1}$'s is directly controlled by θ_{jl}, whose real and imaginary parts are modeled by splines with coefficients $\boldsymbol{\alpha}_{jl}$ and $\boldsymbol{\beta}_{jl}$ accordingly. To detect sparsity patterns on the off-diagonal entries in the spectral matrix, group shrinkage effects are proposed on each pair of $\{\boldsymbol{\alpha}_{jl}, \boldsymbol{\beta}_{jl}\}$ by letting every coefficient pair, $\alpha_{jl,s}$ and $\beta_{jl,s}$, share the same global parameter, τ_{jl}. In practice, the lack of association between the components of the multivariate time series will be inferred in terms of the shrinkage of the corresponding estimates of the $\{\boldsymbol{\alpha}_{jl}, \boldsymbol{\beta}_{jl}\}$ pairs towards zero. We have found results to be robust to the choice of hyperparameters, with $c \in (1, 10^5)$ and $c_\tau \in (10^{-5}, 10^{-1})$ providing indistinguishable estimates in all empirical examples considered. Thus, we simply set $c = 2$ and $c_\tau = 0.01$ by default.

Third, the discount effect function c_s is used to heuristically gain model flexibility. Previous studies do not provide a general criterion on how to optimally choose the number of basis, M. Rosen et al. (2009, 2012); Zhang (2016); Krafty et al. (2017) and Zhang (2019) suggest using $M = 10$. However, such choice is based on their specific prior structure and simulation studies, which lacks gener-

alization and may not be accurate in some settings. We want our approach to incorporate more basis functions when needed in order to preserve model flexibility in general situations. The cosine bases, $\psi_s(\nu_k)$, allow for modeling more volatile behavior with larger s, which is generally less likely to appear in practical settings. Therefore, in addition to the regularization obtained from the regularized horseshoe prior above, we further penalize the appearance of bases according to the value of s. Coefficients of basis terms with larger s will be more likely to be shrunk towards zero, unless the need to include them in the model is strongly supported by the data. In this way, one can simply choose a relatively large M and the proposed prior will then automatically preserve the informative bases and will shrink towards zero those that are not needed. Inspired by the shape of the sigmoid function $\text{Sig}(x) = [1 + \exp(-x)]^{-1}$, we choose the form of c_s as $c_s = \text{Sig}(-as + b)$ such that c_s is a monotone decreasing function of s and is bounded between 0 and 1. Here a and b are fixed constants defining the shape of the function. a determines how fast the shrinkage level grows. If a is small, c_s decreases slowly, meaning the shrinkage level grows slowly. On the other hand, a larger a indicates a faster growth of the shrinkage level. Meanwhile, b controls when the growth starts. A larger b causes later decrease of c_s, or equivalently, later growth of the shrinkage level. We proposed a default choice of $a = 1$ and $b = M/2$. This choice is justified through experimentation via simulation studies and real data analyses.

We refer to the prior structure above as the discounted regularized horseshoe (DRH), and the DRH model to the Bayesian smoothing splines model for multivariate spectral analysis that assumes the DRH prior. We show later via simulation studies that DRH can easily handle more than $M = 30$ bases, which results in much richer representations than those considered in previous studies.

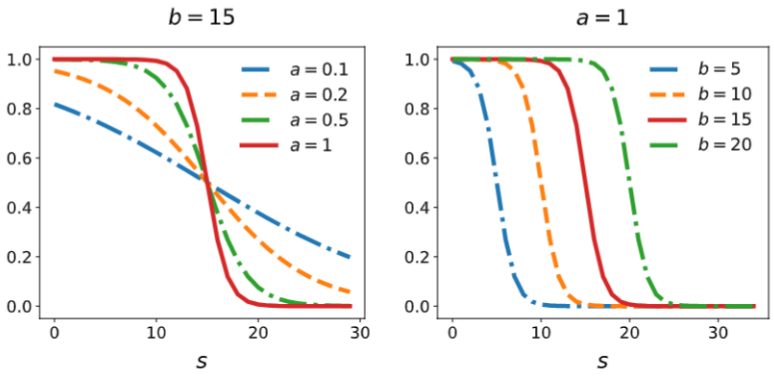

Figure 2.1: $\text{Sig}(-as + ab)$ as a function of s. Left: varying a values while fixing $b = 25$. Right: varying b values while fixing $a = 1$.

It should be noted that this notion of increasing the penalization as M increases was also used by Li and Krafty (2019); Li et al. (2021). The key difference is that our approach is based on a regularized horseshoe prior structure while that of Li and Krafty (2019) uses normal priors without a global-local shrinkage structure, and the prior in Li et al. (2021) is designed for a stationary factor model representation of the spectral density matrix that is totally different from the modified complex Cholesky representation used here. Section 2.3 includes a comparison between our approach and other currently available approaches in extensive simulation studies. In summary, the DRH joint prior for the j-th likelihood component can be written as:

$$\begin{aligned}\pi^{(j)}_{\text{DRH}}(\cdot) = &\prod_{s=0}^{1} p(\gamma_{j,s}) \prod_{s=2}^{M} p(\gamma_{j,s} \mid \lambda_{js}, \tau_j)\, p(\lambda_{js})\, p(\tau_j) \\ &\prod_{l=1}^{j-1} \prod_{s=0}^{M} p(\alpha_{jl,s} \mid \lambda_{jls,(re)}, \tau_{jl})\, p(\lambda_{jls,(re)}) \\ &\prod_{l=1}^{j-1} \prod_{s=0}^{M} p(\beta_{jl,s} \mid \lambda_{jls,(im)}, \tau_{jl})\, p(\lambda_{jls,(im)})\, p(\tau_{jl}).\end{aligned} \tag{2.8}$$

We further take the log transformation on all λ and τ to satisfy the positive constraint of these parameters.

2.2 Posterior Computation

In this section, we describe our approach to adapting variational Bayes techniques to obtain posterior inference of the proposed DRH spectral model. We take advantage of the power of modern computational resources, such as graphic processing units (GPUs), in order to provide a flexible, scalable and computationally efficient inference scheme for high-dimensional time series. Our posterior inference scheme is based on the stochastic gradient variational Bayes (SGVB) approach (Kingma and Welling, 2013). As explained above, the likelihood and priors can be decomposed into P components so we consider inference on each component in parallel. Let $\mathbf{v}_j$ denote all the model parameters for the jth component. The joint log density for the jth component can be written as:

$$\log p(\mathbf{v}_j,\, \mathbf{y}_{1:N}) = \log L_j\left(\mathbf{y}_{1:N}\,;\, \mathbf{v}_j\right) + \log \pi(\mathbf{v}_j), \tag{2.9}$$

where $\log L_j(\cdot)$ is defined in (2.7). $\pi(\cdot)$ is the proposed prior in (2.8). The true posterior is given by $p(\mathbf{v}_j \mid \mathbf{y}_{1:N}) = p(\mathbf{v}_j,\, \mathbf{y}_{1:N}) \,/\, p(\mathbf{y}_{1:N})$, with $p(\mathbf{y}_{1:N})$ an intractable normalizing constant, thus $p(\mathbf{v}_j \mid \mathbf{y}_{1:N})$ also becomes intractable. The goal of variational inference is to find a surrogate distribution that is most similar to the true posterior. Let $q_{\boldsymbol{\phi}_j}(\mathbf{v}_j)$ denote the surrogate distribution of $p(\mathbf{v}_j \mid \mathbf{y}_{1:N})$, where $\boldsymbol{\phi}_j$ are unknown learnable parameters. The so-called evidence lower bound (ELBO) between $q_{\boldsymbol{\phi}_j}(\mathbf{v}_j)$ and $p(\mathbf{v}_j \mid \mathbf{y}_{1:N})$ is defined as $\mathcal{L}\left(p_j,\, q_{\boldsymbol{\phi}_j}\right) = \mathrm{E}_{\mathbf{v}_j \sim q_{\boldsymbol{\phi}_j}(\mathbf{v}_j)}\left[\log p(\mathbf{v}_j,\, \mathbf{y}_{1:N}) - \log q_{\boldsymbol{\phi}_j}(\mathbf{v}_j)\right]$. The goal of variational inference is to find the values of $\boldsymbol{\phi}_j$ that maximize ELBO.

We consider $q_{\boldsymbol{\phi}_j}(\mathbf{v}_j) = \mathrm{N}(\mathbf{v}_j \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ as the default surrogate distribution for $p(\mathbf{v}_j \mid \mathbf{y}_{1:N})$. $\boldsymbol{\Sigma}_j$ is a diagonal matrix with $\mathrm{diag}(\boldsymbol{\Sigma}_j) = \boldsymbol{\sigma}_j^2$. This is one of the most popular choices of a surrogate distribution for SGVB, since the diagonal

structure of the variance-covariance matrix significantly simplifies computations. We show in experiments that the SGVB approximate inference works well in terms of covering the underlying true spectra. Having said this, the proposed framework is friendly to user-specified surrogate distributions, meaning that we can consider more sophisticated choices if needed in practice. For instance, the Gaussian variational approximation with a factor covariance structure (Ong et al., 2018) is an alternative that we have also implemented and that leads to improved uncertainty quantification in low-dimensional and moderate-dimensional practical settings as illustrated in Section 4.1.

Next, to improve run-time convergence, we utilize a reparameterization trick (Kingma and Welling, 2013) on $\mathbf{v}_j$ to obtain more accurate estimates of the gradient of the variational objective (Xu et al., 2019; Domke, 2019). More specifically, we set $\mathbf{v}_j = \boldsymbol{\mu}_j + \boldsymbol{\sigma}_j \odot \boldsymbol{\epsilon}_j$, such that $\boldsymbol{\epsilon}_j \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$ where $\mathbf{I}$ denotes the identity matrix, and $\odot$ denotes element-wise product. To guarantee that the components $\boldsymbol{\sigma}_j^2$ are positive, we take the log transformation, i.e., $\boldsymbol{\zeta}_j = \log \boldsymbol{\sigma}_j^2$. Accordingly, $\boldsymbol{\phi}_j = \{\boldsymbol{\mu}_j, \boldsymbol{\zeta}_j\}$, and the ELBO between $p(\mathbf{v}_j \mid \mathbf{y}_{1:N})$ and $q_{\boldsymbol{\phi}_j}(\mathbf{v}_j)$ can be written as:

$$\begin{aligned}\mathcal{L}\left(p_j, q_{\boldsymbol{\phi}_j}\right) = &\mathrm{E}_{\boldsymbol{\epsilon}_j \sim \mathbf{N}(\mathbf{0},\mathbf{I})}\left[\log p(\boldsymbol{\mu}_j + \exp\frac{\boldsymbol{\zeta}_j}{2} \odot \boldsymbol{\epsilon}_j\,,\, \mathbf{y}_{1:N})\right] \\ &- \mathrm{E}_{\boldsymbol{\epsilon}_j \sim \mathbf{N}(\mathbf{0},\mathbf{I})}\left[\log q_{\boldsymbol{\phi}_j}(\boldsymbol{\mu}_j + \exp\frac{\boldsymbol{\zeta}_j}{2} \odot \boldsymbol{\epsilon}_j)\right].\end{aligned} \tag{2.10}$$

With one sample $\boldsymbol{\epsilon}_j^{(1)} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$, the SGVB gradient estimator of (2.10) is given by:

$$\begin{aligned}\nabla_{\boldsymbol{\phi}_j}\mathcal{L}\left(p_j, q_{\boldsymbol{\phi}_j}\right) \simeq &\nabla_{\boldsymbol{\phi}_j} \log p(\boldsymbol{\mu}_j + \exp\frac{\boldsymbol{\zeta}_j}{2} \odot \boldsymbol{\epsilon}_j^{(1)}\,,\, \mathbf{y}_{1:N}) \\ &- \nabla_{\boldsymbol{\phi}_j} \log q_{\boldsymbol{\phi}_j}(\boldsymbol{\mu}_j + \exp\frac{\boldsymbol{\zeta}_j}{2} \odot \boldsymbol{\epsilon}_j^{(1)}).\end{aligned} \tag{2.11}$$

We found that (2.11) led to a fast convergence in our experiments. More discussions on the convergence of SGVB in our studies are included in supplementary

materials Section S.1. Alternatively, in the Gaussian variational approximation with a factor covariance structure of Ong et al. (2018), the surrogate posterior for a m-dimensional parameter vector $\mathbf{v}$ is defined as $\mathbf{v} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + \mathbf{I}\boldsymbol{\sigma}^2$. $\mathbf{B}$ is a $m \times q$ full rank matrix with $q << m$. The reparameterization trick for the factor covariance structure can be conducted by calculating $\mathbf{v} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$ with $\boldsymbol{\xi} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}_q)$ and $\boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}_m)$, where $\mathbf{I}_q, \mathbf{I}_m$ denote the q, m-dimensional identity matrices. The factor covariance structure allows the variational Gaussian approximation to also capture dependency among the parameters, however, estimating $\mathbf{B}$ can be computational intense with longer run time and more parameter storage when m and q are large. When $q = 0$, the factor covariance structure reduces to the diagonal covariance structure. A comparison between the aforementioned two surrogate choices is presented in Section 2.3.1.

In practice, we propose the so-called three-phase Variational Bayes (TPVB) for posterior inference, as illustrated in Algorithm 1, to stabilize the inference process. Note that Algorithm 1 presents the version with the default Gaussian surrogate distribution with a diagonal covariance matrix, but this algorithm can be easily modified to incorporate the surrogate distribution with a factor covariance structure. In particular, Phase 1 uses (2.9) as the objective function to obtain a point approximation, $\tilde{\mathbf{v}}_j = \arg\max_{\mathbf{v}_j} \log p(\mathbf{v}_j, \mathbf{y}_{1:N})$ by gradient ascent. Phase 2 fixes $\boldsymbol{\mu}_j = \tilde{\mathbf{v}}_j$ and only updates $\boldsymbol{\zeta}_j$ by SGVB to maximize (2.10). Finally, both previously updated $\boldsymbol{\mu}_j$ and $\boldsymbol{\zeta}_j$ are fine-tuned at Phase 3 by SGVB to maximize (2.10). Results from our extensive simulation studies below show that our proposed TPVB achieves high numerical stability. We built the model and implemented the experiments in `Python 3.7` with `Tensorflow-Probability` and `Tensorflow` (Abadi et al., 2015) packages. All gradients are computed via auto-differentiation modules in `Tensorflow`. At each iteration, we use `Adam` (Kingma

Algorithm 1: Three-Phase Variational Bayes (TPVB) Inference

Input : DFT transformed observations of P-dimensional time series, $\mathbf{y}_{1:N}$.
Output: variational posteriors $q_{\boldsymbol{\phi}_j}(\mathbf{v}_j)$, $j = 1, ..., P$.
Params: (parameters) $\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_p$, with $\boldsymbol{\phi}_j = \{\boldsymbol{\mu}_j, \boldsymbol{\zeta}_j\}$, $j = 1, ..., P$.

for $j \in \{1, ..., p\}$ **do**
  Phase 1–Point Approximation: maximizes (2.9) w.r.t $\mathbf{v}_j$, **while** *in iteration* **do**
    Compute the gradients of $\mathbf{v}_j$;
    Update $\mathbf{v}_j$ via gradient ascent;
  end
  Fix $\boldsymbol{\mu}_j$ equal to the last updated $\mathbf{v}_j$ in Phase 1.
  Phase 2–Uncertainty Quant: maximize (2.10) w.r.t $\boldsymbol{\zeta}_j$, **while** *in iteration* **do**
    Draw $\boldsymbol{\epsilon}_j \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$ and compute the SGVB gradients of $\boldsymbol{\zeta}_j$ using (2.11);
    Update $\boldsymbol{\zeta}_j$ via gradient ascent;
  end
  Phase 3–Fine-tuning (Optional): **while** *in iteration* **do**
    Draw $\boldsymbol{\epsilon}_j \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$ and compute the SGVB gradients of both $\boldsymbol{\mu}_j$ and $\boldsymbol{\zeta}_j$ using (2.11);
    Update $\boldsymbol{\mu}_j$ and $\boldsymbol{\zeta}_j$ via gradient ascent;
  end
end

and Ba, 2017) to update model parameters via gradient ascent.

2.3 Simulation Studies

In this section, we show the accuracy and efficiency of our methods through simulations. In all cases below, for all smoothing splines, we set the number of basis to $M = 30$. This provides much richer representations in comparison to those used in previous studies where $M = 10$ regardless of the time series length or their underlying spectral complexity. We show that the DRH spectral model can effectively handle a large M and lead to accurate spectral inference for various

spectral density shapes. During the TPVB inference process, the learning rate is used as a hyper-parameter to control the rate at which the algorithm updates the parameter estimates. The learning rates in `Adam` optimizers are set to be 0.0005, 0.05 and 0.005, respectively for each phase, to obtain gradient descent updates for trainable parameters. We decide the number of iterations by monitoring the values of the objective function, and stop the training when a typical convergent pattern is reached. That leads to approximately 5000, 500, and 500 iterations required for each of the 3 phases. The convergence of the TPVB algorithm is monitored via plots of the log-posterior and the evidence lower bound, as shown in for in Section S.1 of the Supplementary Material. The simulation experiments are executed on a x64-based PC with a 2.60-GHz Intel® Core™ i7-9750H CPU and a Nvidia GeForce GTX 1660 Ti GPU card. We illustrate the accurate performance of the proposed framework for both low and large dimensional spectral analysis, provide comparisons to other available approaches, and discuss model scalability.

2.3.1 Simulation Study 1

In this study, we show that the inference obtained by the proposed TPVB is an excellent approximation to the full posterior inference obtained from Markov Chain Monte Carlo (MCMC) methods. We simulate 100 datasets containing $\mathbf{x}_t = (x_{1,t}, x_{2,t})'$, for $t = 1, \cdots, 1024$, from the following identifiable and stationary bivariate vector autoregressive–moving-average VARMA(2,2) process:

$$\begin{aligned} \mathbf{x}_t = &\begin{bmatrix} 0.2 & 0.5 \\ 0 & -0.2 \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.5 & -0.2 \end{bmatrix} \mathbf{x}_{t-2} + \mathbf{w}_t \\ &+ \begin{bmatrix} 0.6 & 0 \\ 0.2 & -0.5 \end{bmatrix} \mathbf{w}_{t-1} + \begin{bmatrix} 0.3 & 0 \\ 0 & 0.3 \end{bmatrix} \mathbf{w}_{t-2}, \end{aligned} \tag{2.12}$$

where $\mathbf{w}_t$ are independent zero-mean bivariate Gaussian random variables whose components have unit variance and pairwise correlation 0.8. The average squared error (ASE) was used to numerically summarize the performance of the estimators of the spectral components. The ASE of a given posterior spectral estimator is obtained by averaging squared errors across an equally spaced grid of 500 frequencies. For example, the ASE of $\hat{f}_{11}(\nu)$, a posterior mean estimator of $f_{11}(\nu)$, is computed as: $\text{ASE} = \frac{1}{500}\sum_{k=1}^{500}\left[\hat{f}_{11}(\nu_k) - f_{11}(\nu_k)\right]^2$, with $\nu_k = k/1000$.

For each of the 100 datasets, we fit the same DRH model, with the same prior structure, and compare the posterior estimates given by Hamiltonian Monte Carlo (HMC; Neal, 2011), TPVB with the default Gaussian surrogate having a diagonal covariance structure ($q = 0$) defined in Section 2.2, and TPVB with Gaussian surrogate having a factor covariance structure (Ong et al., 2018) defined in Section 2.2 with $q = 5$ and $q = 10$ factors.

We averaged the ASEs for each inference scheme over the 100 datasets to obtain means and standard deviations of ASEs for the two spectral densities, $\boldsymbol{f}_{11}$ and $\boldsymbol{f}_{22}$, and for the squared coherence, ρ^2_{12}. Table 2.1 shows the ASE and run time results obtained by all the methods. In general, we note that all three TPVB approximations slightly underestimate the posterior variability, as expected, however, the results are pretty similar to those obtained under the HMC scheme. The incorporating a factor covariance structure in the variational surrogate distribution mitigates the uncertainty underestimation issue in variational inference, with posterior variability closer to HMC as q increases. We also compared the GPU running time for all schemes. TPVB ($q = 0$) takes on average 42 seconds to run per dataset. By utilizing the factor covariance structure in the variational surrogate distribution, the average run time for TPVB increases to 62 seconds when $q = 5$, and 72 seconds when $q = 10$, indicating that even though such factor

covariance structure provides more accurate estimates on posterior variability, it also requires higher computation costs than the default choice. Note that, the run time of TPVB does not significantly change by the posterior sample size. The run time for HMC, however, grows with respect to the required posterior sample size. On average, HMC takes roughly 380 seconds per dataset to collect ten thousand posterior samples. Given that we do not know in advance how many posterior samples are required, and usually more than ten thousand samples are required for HMC convergence, TPVB ends up being much faster in practice.

Table 2.1: Means (standard deviations) of the run time and ASE values of posterior estimates for spectral components obtained from the DRH model using TPVB and HMC inference schemes with 10000 posterior samples based on 100 simulated datasets of the VARMA process (2.12). In TPVB, the Gaussian variational surrogate distributions with diagonal covariance ($q = 0$) and factor covariance structures under $q = 5, 10$ are compared.

	$\boldsymbol{f}_{11}$	$\boldsymbol{f}_{22}$	ρ^2_{12} $(\times 10^3)$	Runtime [sec]
HMC	0.44 (0.37)	0.28 (0.23)	0.26 (0.18)	382 (65)
TPVB (q=10)	0.44 (0.36)	0.29 (0.21)	0.27 (0.16)	72 (16)
TPVB (q=5)	0.46 (0.32)	0.28 (0.18)	0.26 (0.12)	62 (15)
TPVB (q=0)	0.45 (0.31)	0.27 (0.16)	0.26 (0.10)	46 (12)

Figure 2.2 displays the posterior estimates for the individual spectral densities and the coherence given by TPVB ($q = 0$). It can be seen that the estimates approximate the true curves smoothly. In comparison to the factor covariance structure, TPVB ($q = 0$) is more computational efficient, and preserves similar accuracy regarding to power spectrum estimates (similar mean ASEs). Considering that the main focus of this paper is the high-dimensional model scalability, we focus on the TPVB ($q = 0$) in the following sections of this chapter, noting that the variational Bayeses approximation can be improved by using the factor covariance structure in the surrogate distribution.

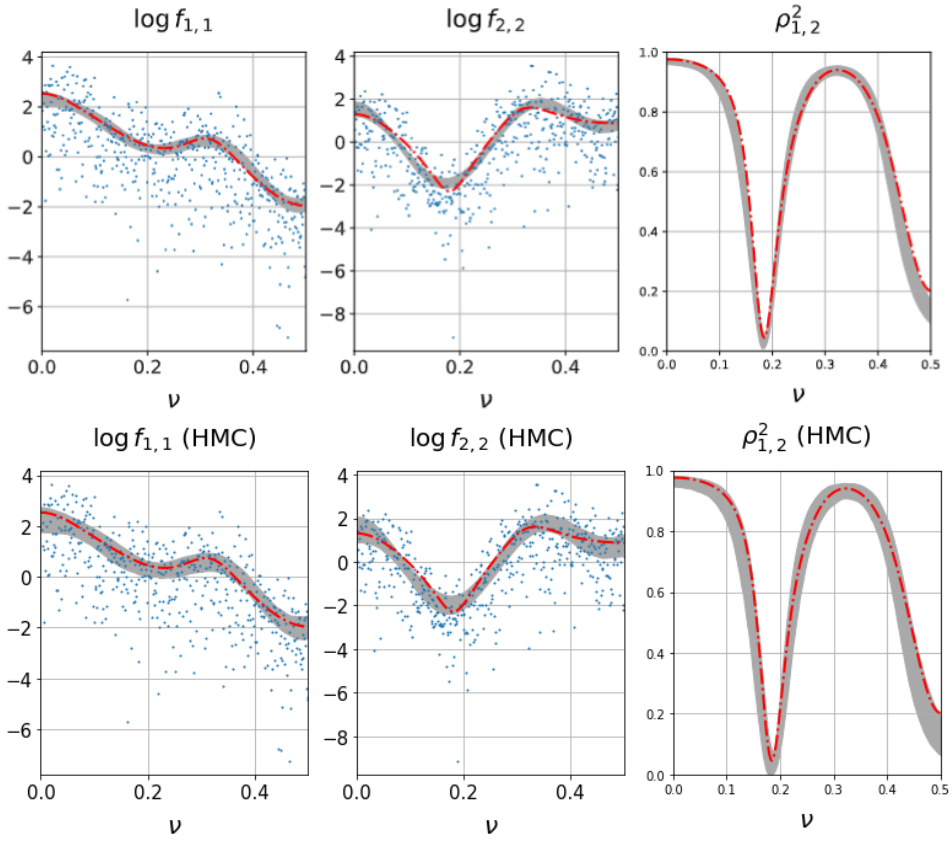

Figure 2.2: Posterior inference of log spectral densities and squared coherence for the VARMA process (2.12) given by the proposed DRH TPVB framework (top frames) and HMC algorithm (bottom frames). Grey regions are 95% posterior intervals. Lines are true values. Dots represent log periodogram.

Similar results in terms of the accuracy of the variational approximations were obtained in another simulation study that considers a VAR(2) bivariate structure (see supplementary materials S.3 for details).

2.3.2 Simulation Study 2

The computation efficiency gained by TPVB allows us to analyze higher dimensional time series in the spectral domain. We simulated 20 datasets containing 21-dimensional time series where each multivariate time series was constructed by combining 7 groups of 3-dimensional time series. We generated each group of time

series from one of 7 3-dimensional VAR processes of order 3. The explicit data generation procedure is detailed in Section S.2 of the Supplementary Material.

Each of the 20 simulated datasets consists of 21 time series components, $\mathbf{x}_t = (x_{1,t}, x_{2,t}, ..., x_{21,t})'$ for $t = 1 : n$, where components within each of the 7 groups allow non-zero coherence, but components that are not in the same group have zero coherence. We considered different scenarios with n set to 128, 256, 384, 512, and 640. For each n, we calculate the ASEs for our DRH model, the SPEC model of Rosen and Stoffer (2007), and the factor model (FM) for stationary time series of Li et al. (2021), all fitted by TPVB on every replication. We set $M = 30$ for all models. Note that, the SPEC model does not consider a global-local regularization in the prior, and instead assigns normal priors to spline coefficients. Unlike DRH and SPEC which model the modified complex Cholesky factorized components of the inverse spectral matrix, the FM tries to learn a low-rank decomposition of the spectral matrix.

Considering that there are 21 spectral densities and 210 squared coherences in total, it is impractical to list all the results individually. Instead, we visualize the ASEs for all the spectral densities and squared coherences in the two boxplots shown in Figure 2.3. This figure presents the boxplots of ASEs varying by the length n of the time series. It can be seen that in comparison to SPEC and FM, DRH provides much better estimates of pairwise squared coherences (see right plot in Figure 2.3), and still offers competitive results in terms of the inference for spectral densities for all n (left plot in the same figure). To further illustrate the accuracy of the proposed approach, we also provide inference results for some spectral densities and some coherences. Figure 2.4 shows DRH inference for the first 8 spectral densities in one of the 21-dimensional datasets of length $n = 640$, while Figure 2.5 displays DRH inference for 8 randomly chosen squared coherence

terms. In general, we see that the proposed model adequately captures the behaviour of the spectral densities and coherences (both zero and non-zero terms). We note that obtaining results using the TPVB approximation for the proposed DRH for a 21-dimensional dataset with a length of 640 observations took less than 9 minutes in the same hardware platform described above.

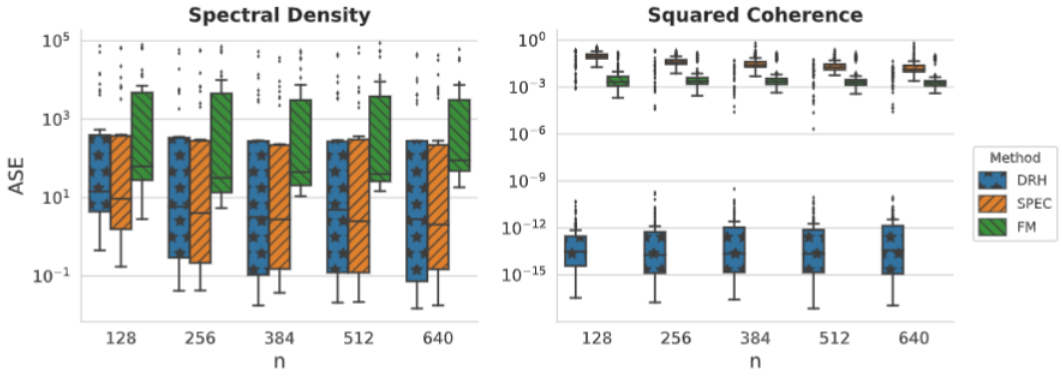

Figure 2.3: Log-scaled boxplots of averaged squared errors (ASEs) of spectral densities (left frame) and squared coherence estimates (right frame) given by our DRH model, the SPEC model, and factor model (FM), all fitted by TPVB for simulated 21-dimensional time series with length n.

2.3.3 Simulation Study 3

This section discusses model scalability as time series dimension considerably increases ($P > 100$). None of the alternative approaches reviewed in the Introduction can be feasibly scaled to consider spectral analysis of multivariate time series of this dimension. For this study, we independently generated $[P/3]$ times from the 3-dimensional VAR(3) process used in Section 2.3.2. The resulting $[P/3]$ sets of 3-dimensional times series were combined into a P-dimensional time series, and were modeled jointly. Specific details of the data generation procedure are included in supplementary materials Section S.2. We consider values of $P = 30, 45, 60, 75, 90$ and 105. Figure 2.6 shows the boxplots of ASE values for the spectral densities

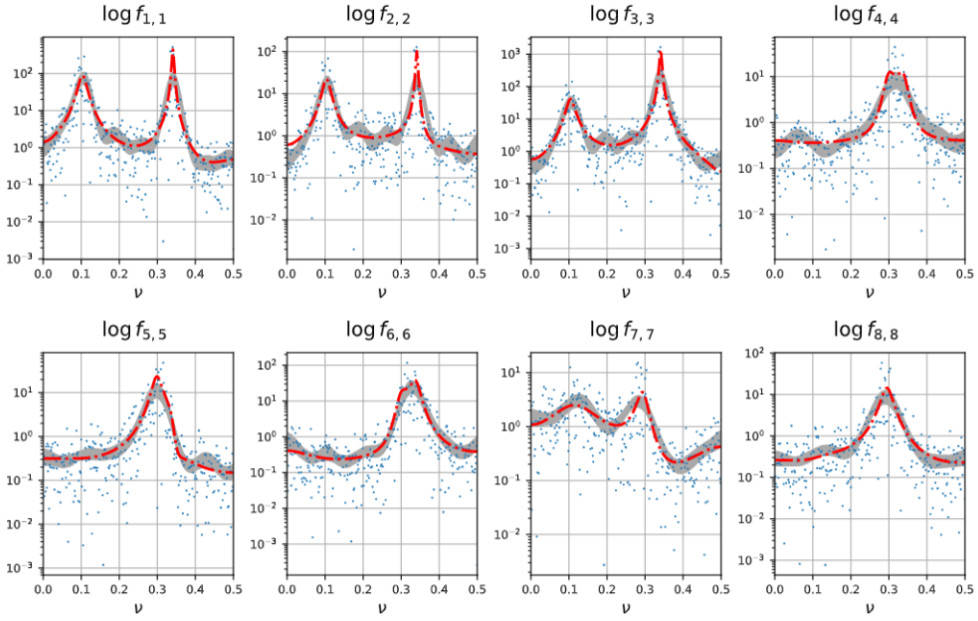

Figure 2.4: Posterior inference of the first 8 log spectral densities given by the proposed DRH TPVB framework for a 21-dimensional dataset of length $n = 640$. Grey regions correspond to 95% posterior intervals. Lines are true values. Dots represent log periodograms.

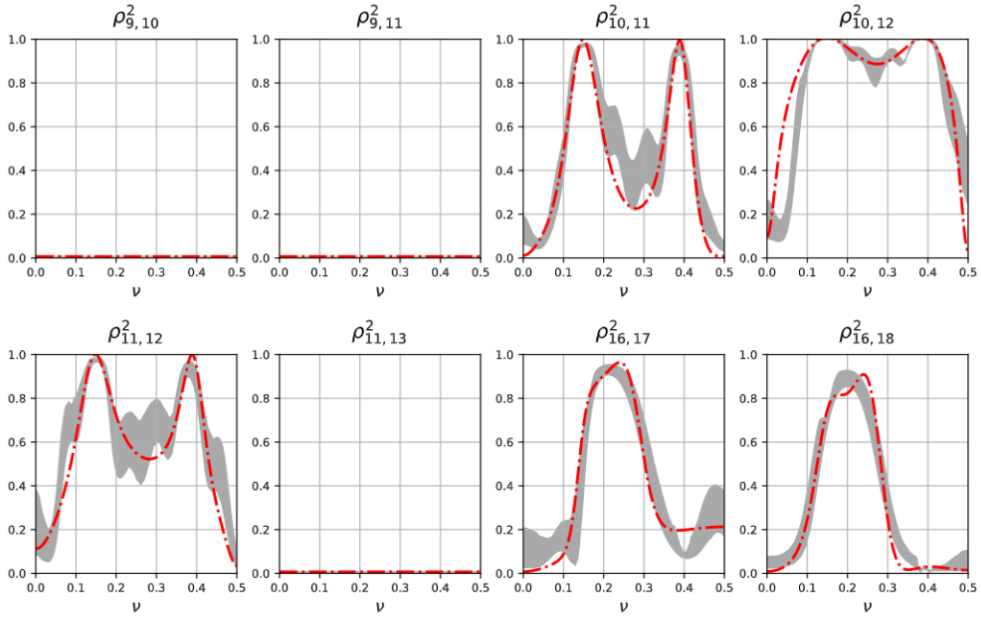

Figure 2.5: Posterior inference of 8 randomly chosen squared coherences obtained from the proposed DRH TPVB framework for a 21-dimensional dataset of length $n = 640$. Grey regions are 95% posterior intervals. Lines are true values.

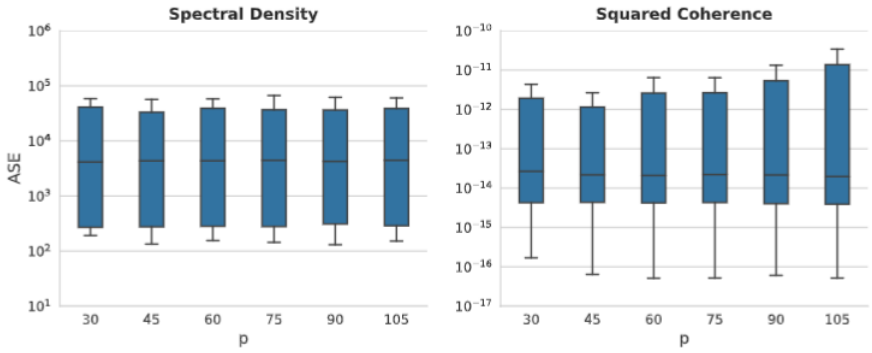

Figure 2.6: Log-scaled boxplots of averaged squared errors (ASEs) of spectral densities (left frame) and squared coherences (right frame) given by the DRH TPVB framework fitted on multivariate time series of dimension P, with fixed length $n = 1024$.

and coherencies for different values of P for datasets of length $n = 1024$. It can be seen that the ASEs for the spectral densities remain in the same range of values as the dimension of the time series increases. For the pairwise squared coherences, the median ASE, indicated by the dark horizontal line in each boxplot, stays mostly constant, but there is slightly more variation as the number of time series increases, which is expected as the number of pairwise comparisons increases from 435 for $P = 30$ to 5,460 for $P = 105$.

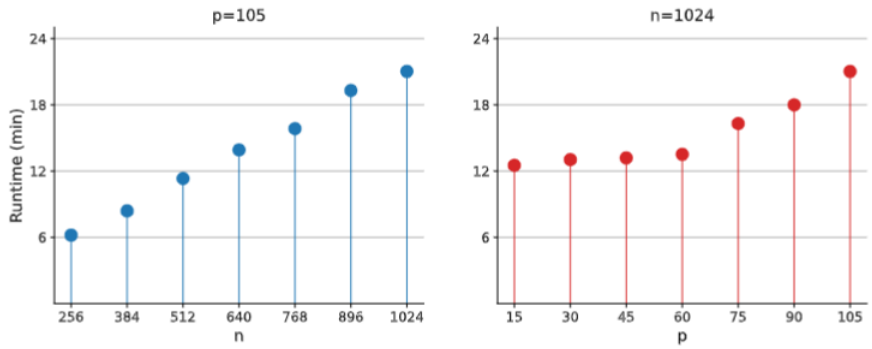

Figure 2.7: DRH TPVB framework runtime (in minutes) per dataset, where n denotes length of observed time series, and P denotes the multivariate time series dimension. Left: runtime change by n with $P = 105$. Right: runtime change by P with $n = 1024$.

To illustrate the time efficiency of the DHR TPVB, we also evaluate how the model runtime is affected by the length, n, and the dimension, P of the multivariate time series. Figure 2.7 shows that the runtime increases as n and P increase, but is still quite affordable: our framework can analyze a 105-dimensional time series with 1024 observations in less than 24 mins. Considering that the inference process can use parallel computing and we have so far only used one GPU for computation, the increase of run time for even larger dimensional analyses could be made up by using multiple GPUs, or Tensor Processing Units (TPUs). One of our future goals is developing reliable code for effective multi-GPU utilization.

2.3.4 Simulation Study 4

The DRH prior defined in Section 2.1.2 is based upon the regularized horseshoe prior, which is one of the global local shrinkage priors. As an alternative, the normal gamma prior also belongs to the global local shrinkage prior family (Griffin and Brown, 2010; Bitto and Frühwirth-Schnatter, 2019). In this study, we compare the DRH with the normal gamma shrinkage prior. In our context, we impose the normal gamma (NG) prior on basis coefficients such that

$$\begin{aligned}
\gamma_{j,s} \mid \lambda_{js}^2 &\sim \mathrm{N}\left(0, \lambda_{js}^2\right), \quad \lambda_{js}^2 \sim \mathrm{Ga}(1, \tau_j^2/2), \\
\alpha_{jl,s} \mid \lambda_{jls,(re)} &\sim \mathrm{N}\left(0, \lambda_{jls,(re)}^2\right), \quad \lambda_{jls,(re)}^2 \sim \mathrm{Ga}(1, \tau_{jl}^2/2), \\
\beta_{jl,s} \mid \lambda_{jls,(im)} &\sim \mathrm{N}\left(0, \lambda_{jls,(im)}^2\right), \quad \lambda_{jls,(im)}^2 \sim \mathrm{Ga}(1, \tau_{jl}^2/2),
\end{aligned}$$

where λ_{js}^2, $\lambda_{jls,(re)}^2$, $\lambda_{jls(im)}^2$ are local parameters and τ_j^2, τ_{jl}^2 are global parameters that τ_j^2, $\tau_{jl}^2 \sim$ Ga(0.001, 0.001) (an informative prior). To compare the performance of the proposed DRH and NG. We conduct a simulation study to simulate

50 P-dimensional time series of length 1024 independently from

$$\boldsymbol{X}_t = \boldsymbol{w}_t + \Phi_1 \boldsymbol{w}_{t-1} + \Phi_2 \boldsymbol{w}_{t-2}, \tag{2.13}$$

where $\boldsymbol{w}_t \overset{iid}{\sim} \mathrm{N}(\boldsymbol{0}, \Omega)$ with $\Omega = I_{P/3} \otimes \Omega^0$, and Ω^0 is a 3×3 matrix that has 1's on the diagonal and 0.5 off the diagonal. $\otimes$ denotes the Kronecker product. $I_{P/3}$ is a $P/3 \times P/3$ identity matrix. $\Phi_1 = I_{P/3} \otimes \Phi_1^0$, and $\Phi_2 = I_{P/3} \otimes \Phi_2^0$ with

$$\Phi_1^0 = \begin{pmatrix} 0.6 & 0 & 0 \\ 0.2 & -0.5 & 0 \\ 0.1 & 0.3 & 0.4 \end{pmatrix}, \quad \Phi_2^0 = \begin{pmatrix} 0.3 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

To allow for a more concise summarization of the accuracy given by different models, instead of calculating ASEs for every component of the power spectrum individually, we calculate the mean integrated squared error (MISE) (Li et al., 2021) for stationary process over the whole spectral matrix $\boldsymbol{f}(\omega)$ by

$$\mathrm{MISE} = \frac{1}{500} \sum_{k=1}^{500} \left\| \hat{\boldsymbol{f}}\left(\nu_k\right) - \boldsymbol{f}\left(\nu_k\right) \right\|_F^2,$$

where $\|\cdot\|_F$ is the matrix Frobenius norm. $\nu_k = k/1000$ for $k = 1, ..., 500$ are the prespecified equally spaced frequency grids. $\hat{\boldsymbol{f}}$ is the obtained posterior estimate of $\boldsymbol{f}$. Furthermore, the performance of diagonal and off-diagonal estimators are investigated separately by

$$\mathrm{MISE}_d = \frac{1}{500} \sum_{k=1}^{500} \left\| \mathrm{diag}[\hat{\boldsymbol{f}}\left(\nu_k\right)] - \mathrm{diag}[\boldsymbol{f}\left(\nu_k\right)] \right\|^2,$$
$$\mathrm{MISE}_o = \frac{1}{500} \sum_{k=1}^{500} \left\| \hat{\boldsymbol{f}}\left(\nu_k\right) - \boldsymbol{f}\left(\nu_k\right) \right\|_{F*}^2,$$

Table 2.2: Simulation results for the process (2.13) based on 50 repetitions: means (standard deviations) of the MISE, MISE_o, and MISE_d given by the proposed DRH, NG, and the baseline SPEC by Rosen and Stoffer (2007).

P	Model	MISE	MISE_d	MISE_o
3	DRH	0.38 (0.04)	0.22 (0.03)	0.16 (0.02)
	NG	0.42 (0.03)	0.24 (0.03)	0.18 (0.02)
	SPEC	0.47 (0.06)	0.27 (0.03)	0.20 (0.03)
12	DRH	1.66 (0.09)	0.82 (0.04)	0.84 (0.06)
	NG	1.82 (0.09)	0.89 (0.04)	0.93 (0.06)
	SPEC	2.24 (0.14)	1.02 (0.08)	1.24 (0.11)
24	DRH	3.27 (0.13)	1.59 (0.08)	1.68 (0.09)
	NG	3.66 (0.14)	1.76 (0.08)	1.90 (0.09)
	SPEC	5.15 (0.26)	2.21 (0.13)	2.94 (0.18)
48	DRH	7.27 (0.30)	3.04 (0.16)	4.23 (0.21)
	NG	7.90 (0.27)	3.39 (0.16)	4.51 (0.19)
	SPEC	10.57 (0.50)	4.03 (0.22)	6.54 (0.35)
96	DRH	16.39 (0.32)	6.22 (0.22)	10.17 (0.28)
	NG	18.57 (0.34)	6.85 (0.21)	11.72 (0.27)
	SPEC	23.10 (0.64)	7.38 (0.39)	15.62 (0.47)

where $\|A\|_{F*} = \sqrt{\sum_{i=1}^{P} \sum_{j \neq i} A_{ij}^2}$. We use MISE, MISE_d and MISE_o to jointly evaluate the performance of candidate models.

We compare the proposed DRH model, NG model and the baseline SPEC model (Rosen and Stoffer, 2007) in Table 2.2. It can be seen that both DRH and NG outperform the baseline, but DRH performs better than NG. Two comments should be noted with this study. First, both DRH and NG outperforming the baseline indicates that, the global-local shrinkage prior is indeed useful to automatically shrink non-informative basis towards zeros such that we can set a large basis number (30 basis used in this simulation) in advance to preserve model flexibility at the beginning and worry less about overfitting the data. Second, the better outcomes given by DRH in comparison to NG imply that, even though both

priors belong to the global-local shrinkage family that includes both individual smoothing and global shrinkage components altogether, the DRH provides better estimates due to the benefits from (1) the discounted effect that varies the degree of regularization according to the smoothness of the spectral components; (2) the regularized horseshoe structure where parameters with significantly large values could escape from the shrinkage.

2.4 Data Analyses

2.4.1 Analysis of California Wind Profile

We test our method on data from the Iowa State University Environmental Mesonet (IEM) Automated Surface System (ASOS) database (Todey et al., 2002; Mannarano, 1998). This is a publicly available repository of automated airport weather observations and general basic weather reports from the National Weather Service (NWS), the Federal Aviation Administration (FAA) and the Department of Defense (DOD). We consider wind speed data taken from 6 airports in California: EDU (Davis), SAC and SMF (Sacramento), MRY (Monterey), SNS (Salinas), and WVI (Watsonville). Note that EDU, SAC and SMF are located in Sacramento area, while MRY, SNS and WVI are located near the Monterey Bay in the Monterey and Santa Cruz counties. Our goal is to infer 6 spectral densities and 15 pairs of squared coherences in order to provide insight into the temporal relationships across different locations over a particular period of time. We consider the median wind speed every 2 hours, starting from 06/01/2019 12:00 am to 08/31/2019 11:59 pm, for each location. Given that our method is based on the assumption of stationarity, we consider a period of time within the summer months to avoid extreme values and non-stationarities related to rainfall and

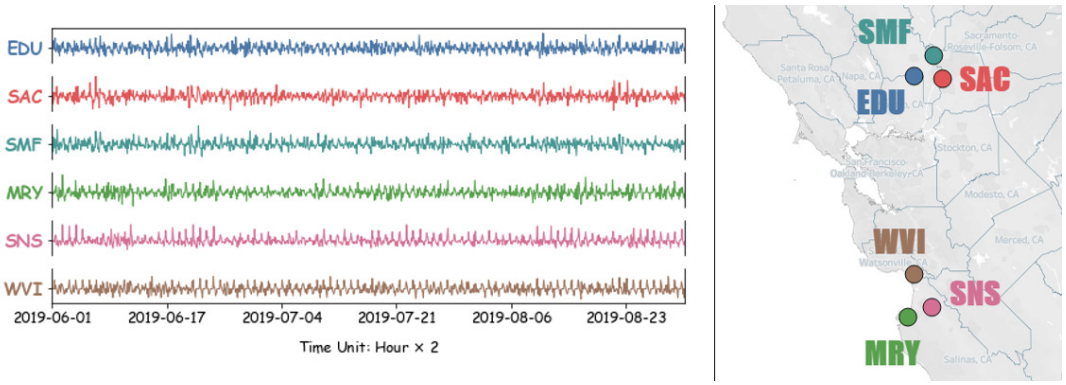

Figure 2.8: Standardized first differences of median wind speed measurements every two hours between 06/01/2019 12:00 am and 08/31/2019 11:59 pm from selected California airports.

storms that occur in other months. We note that there is essentially no rainfall during this period within these locations. Prior analyses of wind profiles in some of these locations have shown quasi-periodic patterns every 24-hours (Garcia et al., 2020). Here we focus on spectral inference on frequency values that are away from zero, therefore, in order to remove any local trends we jointly analyze 6-dimensional data corresponding to the first order differences for each time series. We further standardized each detrended time series by subtracting its mean and dividing by its standard deviation to enhance computation stability. The resulting 6-dimensional time series data with 1,104 observations along with their locations are shown in Figure 2.8. Additional stationarity diagnostics of the differenced time series are included in Section S.4 of the Supplementary Materials. For this analysis we set $M = 30$. The prior hyperparameters and algorithmic settings for the DHR TPVB inference are the same as those discussed in Section 2.3.

Figures 2.9 and 2.10 show the posterior estimates and 95% posterior intervals given by TPVB under the proposed DHR model. It can be seen that our estimates smoothly follow the trajectories of the observed log periodograms. Note that all the estimated spectral densities have a peak around frequency 0.08 (ap-

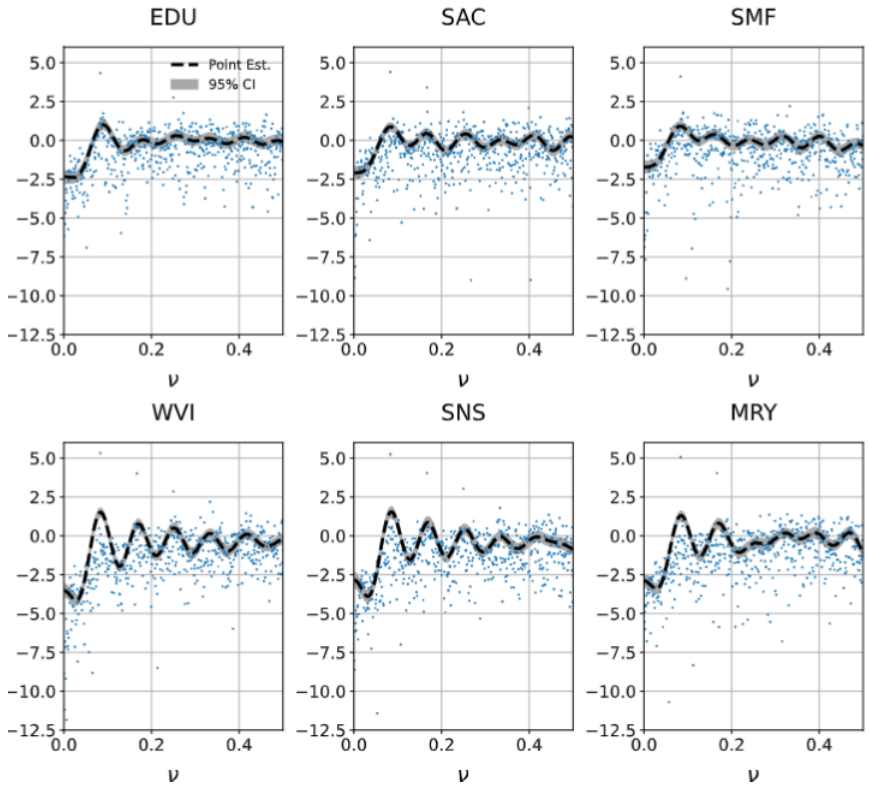

Figure 2.9: Results of the wind profile analysis: Posterior inference of log spectral densities given by the proposed DRH TPVB framework. Lines are posterior mean estimates. Grey regions are 95% posterior intervals. Dots represent observed log periodograms.

proximately 1/12), indicating a strong daily periodic behavior at each location, with higher power around 0.08 for those locations closer to coastal areas (MRY, SNS, WVI). This is consistent with the quasi-periodic nature of the daily wind profile during summer months in such locations (Garcia et al., 2020). Figure 2.10 shows that squared coherences are inferred be non-zero only for pairs of sites that are geographically close to each other, i.e., non-zero squared coherences between pairs in the group (EDU, SAC, SMF), and between pairs in the group (MRY, SNS, WVI), but zero coherence between pairs in which the two series are in different groups. This shows that locations that are next to each other share strong coherences in the 1-st order differences of their wind profile patterns at certain

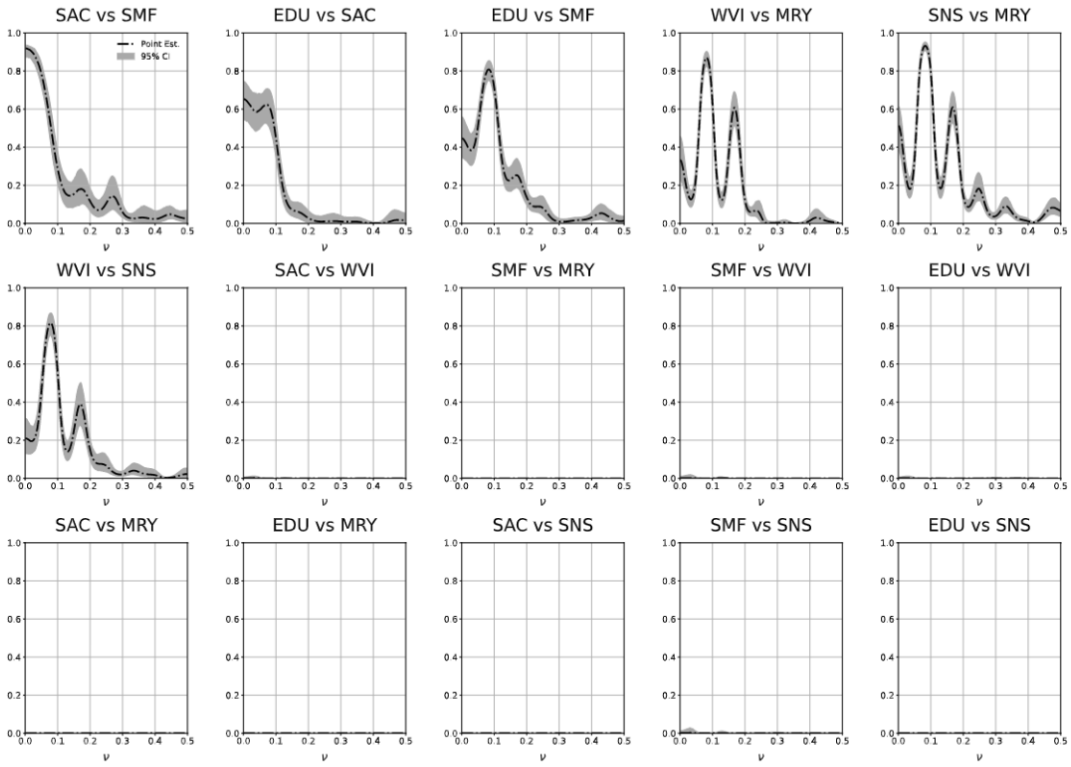

Figure 2.10: Result of the wind profile analysis: posterior inference of squared coherences given by the proposed DRH TPVB framework. Lines are posterior mean eatimates. Grey regions are 95% posterior intervals.

frequencies, while this is not the case for locations that are far apart. We also see that the coherence is higher for lower frequencies and, for most pairs, the highest coherence occurs near the 0.08 frequency (daily quasi-periodicity).

2.4.2 Analysis of Multi-Channel EEG Data

We demonstrate the proposed methodology through the analysis of a multi-channel EEG data set available at the UCI Machine Learning Repository (Dua and Graff, 2017). These data arise from a large study aiming to examine EEG correlates of genetic predisposition to alcoholism. The dataset includes measurements from 61 electrodes placed on the scalp of each participant (Zhang et al., 1995) sampled at 256 Hz for 1 second. The full dataset contains a total of 122

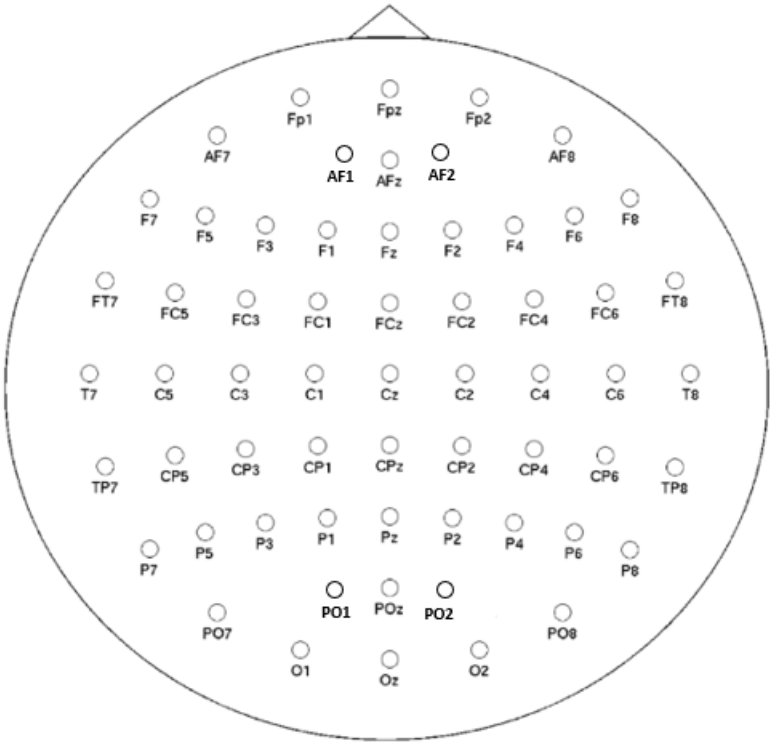

Figure 2.11: Schematic representation of the 61-channel EEG locations.

subjects with two groups of subjects, alcoholic subjects and control subjects. Each subject completed 120 trials under different stimuli. During each trial, each subject was exposed to either a single stimulus (S1) or to two stimuli (S1 and S2) which were pictures of objects chosen from the 1980 Snodgrass and Vanderwart picture set (Snodgrass and Vanderwart, 1980).

We analyze the small dataset version that includes EEG recordings for one alcoholic subject and one control subject. In particular, we consider data from the first trial for each of the subjects under a single stimulus, consisting of two 256-length, 61-dimensional time series, one for an alcoholic subject and another one for a control subject. Figure 2.11 presents the schematic representation of the channel locations. Table 2.3 lists the analyzed 61 EEG channels included in the dataset and their corresponding indices used in this paper. Figure 2.12

displays the detrended time series by taking the 1-st order difference from raw data for both alcoholic and control subjects, where we removed local trends by taking the 1-st order difference. The detrended time series were treated as being stationary, which is not an unreasonable assumption considering their short length ($n = 256$). We fit our proposed DRH multivariate spectral model with $M = 10$ and the same hyperparameters and TPVB inference setting discussed in Section 2.3 to the two 61-dimensional time series separately. We obtain fast results leading to approximate posterior inference of the spectral density matrix for the two 61-dimensional time series.

Table 2.3: 61 selective EEG channels included in EEG Database Data Set, UCI Machine Learning Repository (Dua and Graff, 2017) that are used in our study . Row and column indices indicate the corresponding channel ID (row index + column index) used in our model. For instance, AF1, F1, and P1 are accordingly the 1-st, 20-th, and 44-th dimension of the model inputs.

	0	1	2	3	4	5	6	7	8	9
1	AF1	AF2	AF7	AF8	AFZ	C1	C2	C3	C4	C5
11	C6	CP1	CP2	CP3	CP4	CP5	CP6	CPZ	CZ	F1
21	F2	F3	F4	F5	F6	F7	F8	FC1	FC2	FC3
31	FC4	FC5	FC6	FCZ	FP1	FP2	FPZ	FT7	FT8	FZ
41	O1	O2	OZ	P1	P2	P3	P4	P5	P6	P7
51	P8	PO1	PO2	PO7	PO8	POZ	PZ	T7	T8	TP7
61	TP8									

We summarize the results in terms of posterior estimates of the spectral densities for the individual channels for the alcoholic and control subjects, and also look at summary measures related to the coherence across channels for a specific frequency band. More specifically, we focus on computing frequency-band collapsed measures at the beta band (16–31 Hz), as this frequency band has been found to provide important neurophysiological information (Ferrarelli et al., 2019; Li et al., 2021). Frequency-band collapsed measures can be computed as inte-

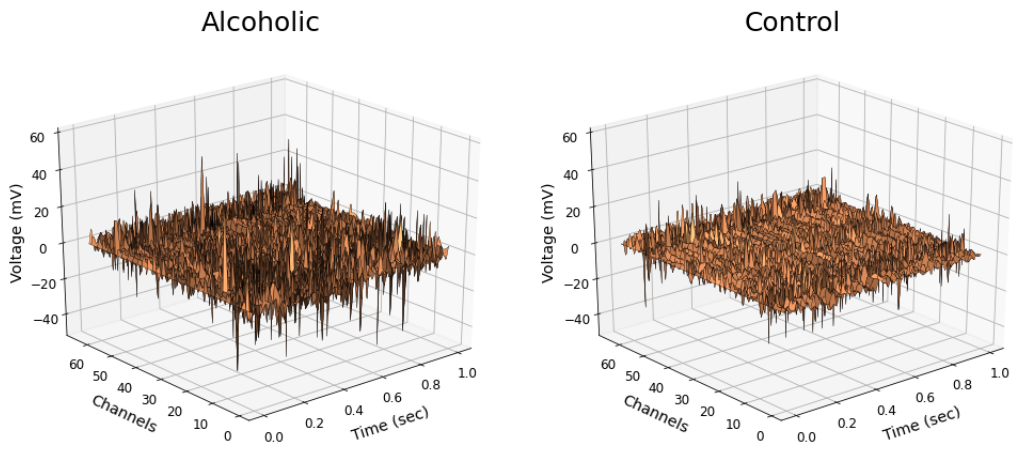

Figure 2.12: Detrended 61-channel EEG data of one alcoholic subject and one control subject exposed to a single stimuli, sampled at 256 Hz for 1 second. The data has been detrended by 1-st order differences.

grals of the power spectra (Li et al., 2021). The beta-band collapsed squared coherence between channels j and l is defined as $\rho_{jl}^{2,\beta} = \left|\boldsymbol{f}_{jl}^{\beta}\right|^2 / \left\{\boldsymbol{f}_{jj}^{\beta}\boldsymbol{f}_{ll}^{\beta}\right\}$, where $\boldsymbol{f}^{\beta} = \int_{16}^{31} \boldsymbol{f}(\nu)\, d\nu$ defines beta-band collapsed spectral matrix. Figure 2.13 shows the estimated beta-band log spectral densities of 61 channels for alcoholic and control subjects under the same stimulus. Figure 2.14 and 2.15 presents the corresponding estimated beta-band squared coherences. We see that, in comparison to the control subject, the alcoholic subject shows a smaller number of non-zero coherences among channels over beta band. Figure 2.16 presents the estimated top 20 largest beta-band squared coherences for each of the groups. We see that most of the actively connected channel pairs differ between groups. For instance, the estimated beta-band squared coherence of the channel pair, C1-C3, is relatively high in the alcoholic subject, but essentially zero in the control subject. In contrast, channel pairs such as AF1-AF2 and C4-FC4 are estimated to have large coherences in the control subject, while showing zero coherence in the alcoholic subject. We note that even though many of the large coherence values occur between nearby locations, some non-negligible coherence values occur be-

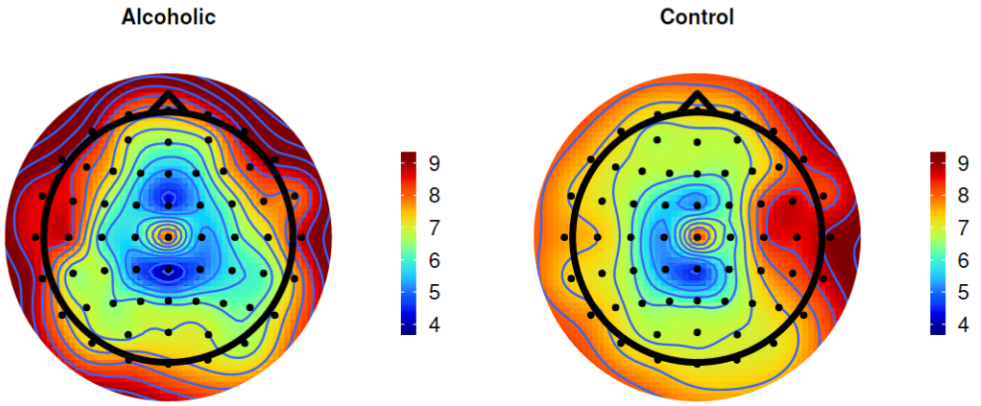

Figure 2.13: Result of EEG analysis: posterior estimates of log beta-band collapsed spectral densities for alcoholic (left) and control (right) subjects.

tween channels that are not very close to each other, particularly for the control subject.

This analysis illustrates that our proposed spectral method leads to fast spectral inference that can offer helpful insight to neuroscientists that need to analyze large-dimensional brain signals recorded in experimental settings that involve multiple trials and/or subjects.

2.5 Conclusions and Future Work

The proposed novel modeling and inference framework provides accurate and computationally fast approximate Bayesian spectral inference for multivariate time series. We showed that the discounted regularized horseshoe prior leads to accurate inference of spectral densities and squared coherences in several simulation settings involving multivariate time series. Furthermore, the proposed inference scheme, that utilizes stochastic gradient variational Bayes, is highly scalable and efficient, providing solid computational support for large-dimensional spectral analysis.

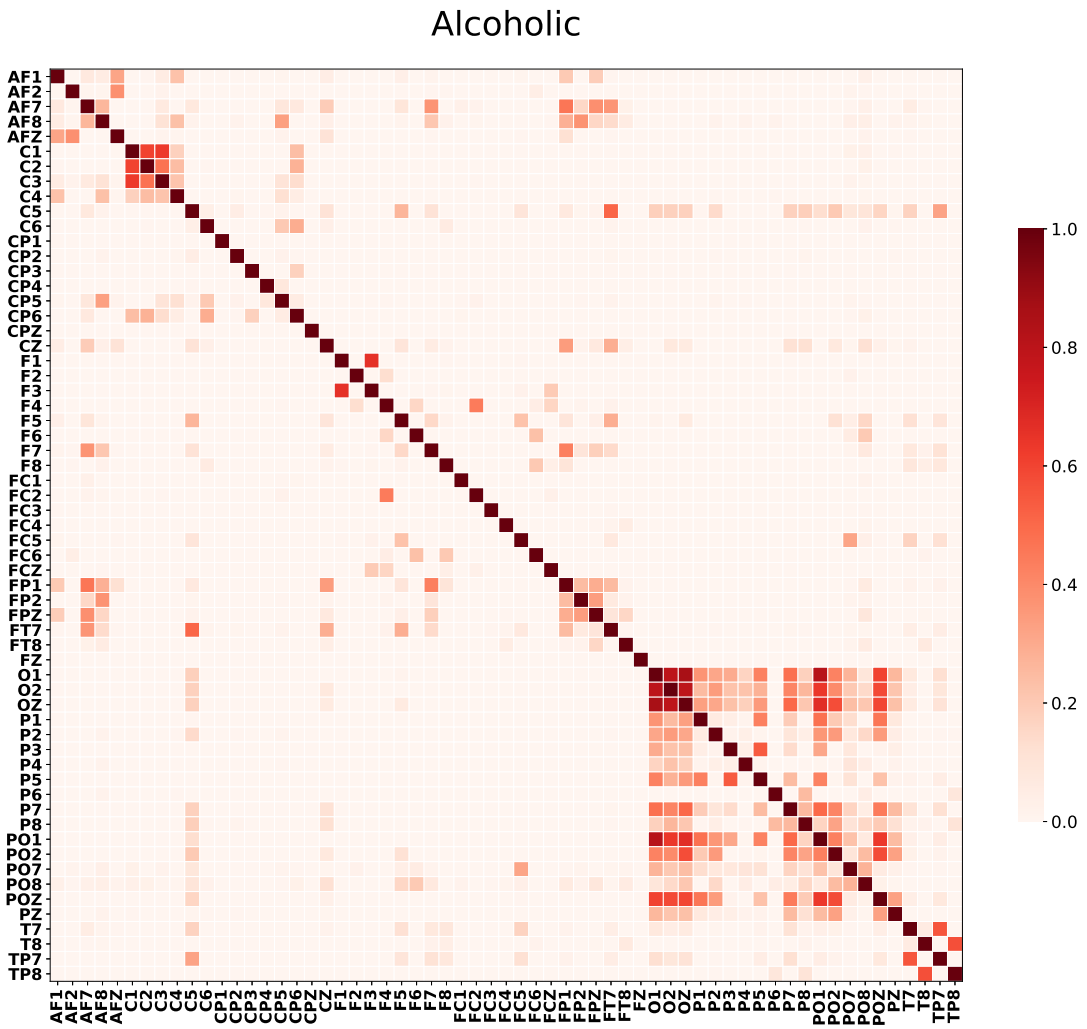

Figure 2.14: Result of EEG analysis: posterior estimates of beta-band collapsed squared coherences for alcoholic subject.

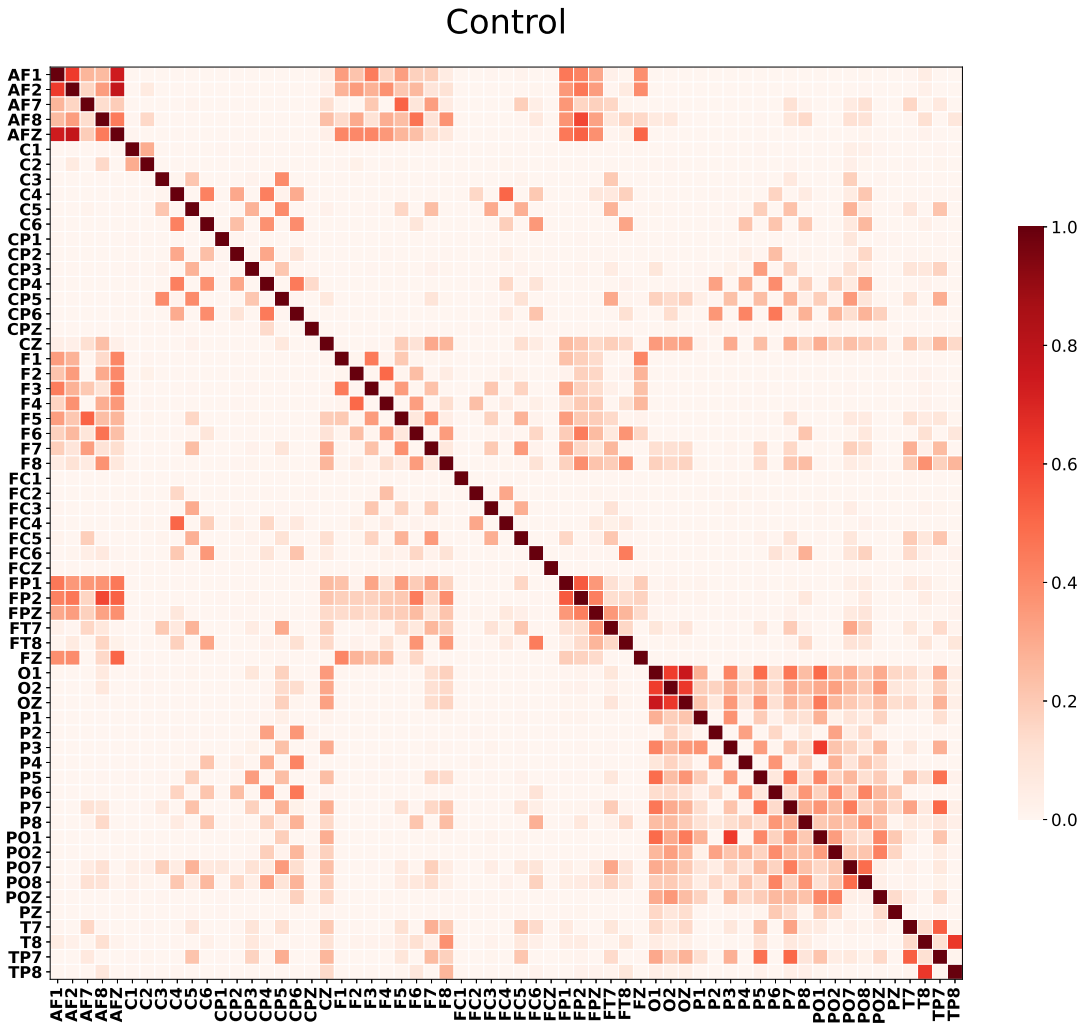

Figure 2.15: Result of EEG analysis: posterior estimates of beta-band collapsed squared coherences for control subject.

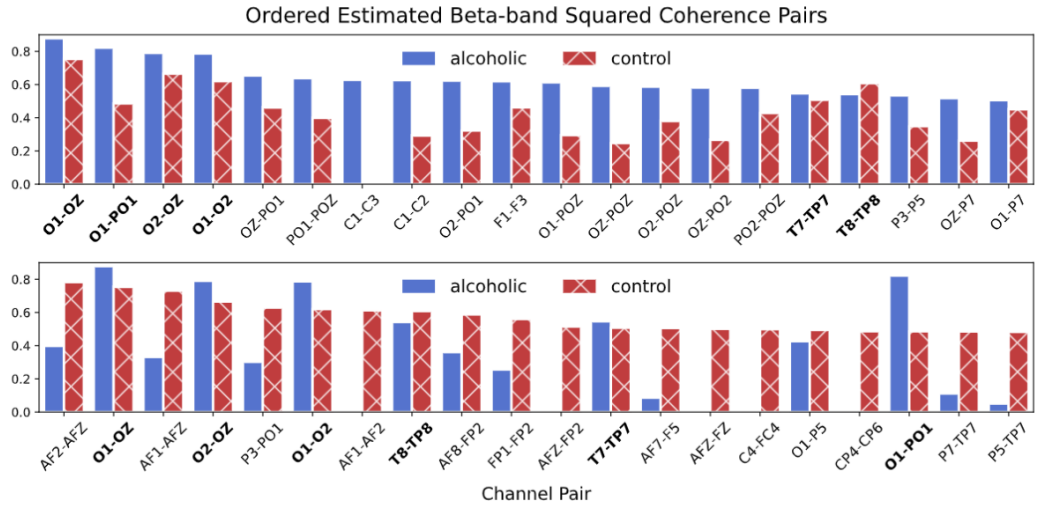

Figure 2.16: Top: top-20 largest estimated beta-band collapsed squared coherence pairs in alcoholic group and their contrasts in control group. Bottom: top-20 largest estimated beta-band collapsed squared coherence pairs in control group and their contrasts in alcoholic group. There are only limited common pairs (bold) existing in both top-20 groups.

Future work involves extending the current prior setting. Note that, we have discussed the normal-gamma prior by Griffin and Brown (2010) in Section 2.3.4. Other possible extensions include the modifications of the normal-gamma prior by Huber and Feldkircher (2019), as well as exploring the benefits of introducing additional sparsification procedures similar to those of Huber et al. (2021), where a parameter sparsification step is proposed after obtaining the posterior estimates using global-local shrinkage priors to further reduce storage load and improve performance.

More relevant extensions in terms of increasing the applicability of the proposed modeling framework and related inferential procedures include considering multivariate spectral analysis for non-stationary time series which we have done in Chapter 3, as well as hierarchical model formulations that allow us to jointly analyze multi-trial data.

Chapter 3

Spectral Analysis of Multivariate Nonstationary Time Series

In this chapter, we extend the proposed framework in Chapter 2 to scalable Bayesian spectral analysis for multivariate nonstationary time series whose power spectrum evolve over time. The remainder of this chapter is organized as follows. Section 3.1 specifies the model and priors used for spectral analysis of nonstationary multivariate time series. Section 3.2 describes the proposed stochastic gradient variational Bayes posterior inference scheme as well as the model selection metric. Section 3.3 reports results of extensive simulation studies that illustrate the accuracy and scalability of the proposed approach in comparison to the current alternatives. Section 3.4 applies the proposed method to the analysis of multi-location wind profiles from the Iowa Environmental Mesonet (IEM) database (Todey et al., 2002) and a multi-channel EEG data (Prado et al., 2001).

3.1 Model Specification

We consider time-varying spectral analysis of a locally stationary P-dimensional time series defined through a Cramér representation (Priestley, 1965). Formally, consider a $\mathbb{R}^P$-valued time series of length T, $\{\boldsymbol{X}_t : t = 1, ..., T\}$, of the form:

$$\boldsymbol{X}_t = \int_{-1/2}^{1/2} \boldsymbol{A}(t/T, \omega)\, e^{2\pi\, \mathrm{i}\, t\omega} d\boldsymbol{Z}(\omega),$$

where $\boldsymbol{A}(u, \omega)$ is a complex-valued time-varying transfer function of scaled time $u \in [0, 1]$ and frequency $\omega \in \mathbb{R}$ such that $A(u, \omega)$ is a nonsingular $P \times P$ complex valued matrix with $\boldsymbol{A}(u, \omega) = \boldsymbol{A}(u, \omega + 1)$ and $\boldsymbol{A}(u, \omega) = \boldsymbol{A}(u, -\omega)^*$, where * denotes the conjugate transpose. $\boldsymbol{Z}(\omega)$ is a Hermitian P-dimensional orthogonal increment zero-mean process with a unit variance. We assume that the cumulants of $d\boldsymbol{Z}$ exist and are bounded for all orders in order to assure that the spectral estimators are well behaved (Guo and Dai, 2006; Li and Krafty, 2019).

Our goal is to efficiently estimate, with high accuracy and low time consumption, the time-varying spectrum

$$\boldsymbol{f}(u, \omega) = \boldsymbol{A}(u, \omega)\boldsymbol{A}(u, \omega)^*, \quad u \in [0, 1],\, \omega \in \mathbb{R},$$

which is a $P \times P$ positive definite Hermitian matrix for any u and ω. Following Li and Krafty (2019) and Li et al. (2021) we further impose some regularity conditions and assume that for every u, each component of $\boldsymbol{f}(u, \cdot)$ has a square-integrable first derivative as a function of frequency and that for every ω, each component of $\boldsymbol{f}(\cdot, \omega)$ is assumed to be continuous as a function of scaled time at all but a possible finite number of points. This allows components of spectrum to evolve not only continuously, but also abruptly in time, which is more flexible in comparison to the ones proposed by Dahlhaus (2000) and Guo and Dai (2006)

where the time-varying spectrum is required to be continuous in both time and frequency domains.

3.1.1 Locally Stationary Approximation

A locally stationary time series $\{\boldsymbol{X}_t : t = 1, ..., T\}$ can be be well approximated by a piecewise stationary process (Adak, 1998; Guo and Dai, 2006; Li and Krafty, 2019; Li et al., 2021). Given a partition of the time series into L segments defined by partition points $\boldsymbol{\xi} = (\xi_0, ..., \xi_L)$ with $\xi_0 = 0$ and $\xi_L = T$ such that $\boldsymbol{X}_t$ is approximately stationary within the segments $\{t : \xi_{\ell-1} < t \leq \xi_\ell\}$ for $\ell = 1, ..., L$, then

$$\boldsymbol{X}_t \approx \sum_{\ell=1}^{L} \int_{-1/2}^{1/2} \boldsymbol{A}(u_\ell, \omega)\, I_{(\xi_{\ell-1}, \xi_\ell]}(t)\, e^{2\pi\, \mathrm{i}\, t\omega} d\boldsymbol{Z}(\omega),$$

where $I_\Omega(t)$ is the indicator function such that $I_\Omega(t) = 1$ if $t \in \Omega$ and 0 otherwise and $u_\ell = \frac{\xi_{\ell-1}+\xi_\ell}{2T}$ is the scaled midpoint of the ℓ-th segment. For any $t \in (\xi_{\ell-1},\, \xi_\ell]$, let $\boldsymbol{A}_\ell(\omega) = \boldsymbol{A}(u_\ell, \omega)$. Then, within each ℓ-th segment, the time series is approximately second-order stationary with local power spectrum $\boldsymbol{f}_\ell(\omega) = \boldsymbol{A}_\ell(\omega)\boldsymbol{A}_\ell(\omega)^*$ that stays the same for any $t \in (\xi_{\ell-1},\, \xi_\ell]$.

Given the partition $\boldsymbol{\xi}$, the local discrete Fourier transform (DFT) at frequency $\omega_{\ell n}$ within segment ℓ can be defined as

$$\boldsymbol{Y}_{\ell n} = T_\ell^{-1/2} \sum_{t=\xi_{\ell-1}+1}^{\xi_\ell} \boldsymbol{X}_t \exp\left(-2\pi\, \mathrm{i}\, \omega_{\ell n} t\right),$$

$$\ell = 1, \ldots, L,\, n = 1, \ldots, N_\ell,$$

where T_ℓ denotes the number of time points in the ℓ-th segment, $\omega_{\ell n} = n/T_\ell$, $n = 1, ..., N_\ell$ are Fourier frequencies with $N_\ell = \lfloor (T_\ell - 1)/2 \rfloor$. If $\boldsymbol{X}_t$ is a zero-mean process, under the multivariate extension of the Whittle approximation (Whittle, 1957), $\boldsymbol{Y}_{\ell n}$ are approximately independent zero-mean complex multivariate

Gaussian random variables with covariance matrices $\boldsymbol{f}_\ell(\omega_{\ell n})$. The sum of local log-Whittle likelihoods of $\boldsymbol{Y}_{\ell n}$ leads to an approximation of the total log-likelihood given by:

$$\mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{f}_{1:L}, \boldsymbol{\xi}, L) \approx \sum_{\ell=1}^{L} \sum_{n=1}^{N_\ell} \left\{ \log |\boldsymbol{f}_\ell(\omega_{\ell n})|^{-1} - \boldsymbol{Y}_{\ell n}^* \, \boldsymbol{f}_\ell^{-1}(\omega_{\ell n}) \, \boldsymbol{Y}_{\ell n} \right\}. \tag{3.1}$$

Here $\boldsymbol{Y} = \{\boldsymbol{Y}_{\ell n} : \ell = 1, ..., L, \, n = 1, ..., N_\ell\}$ represents the collection of all local discrete Fourier transforms and $\boldsymbol{f}_{1:L} = \{\boldsymbol{f}_\ell(\cdot) : \ell = 1, ..., L\}$ is the collection of all local power spectra that need to be estimated. Note that, several previous approaches (Rosen et al., 2012; Zhang, 2016; Li and Krafty, 2019; Li et al., 2021) treat both the locations that define the partition $\boldsymbol{\xi}$ and the number of components in the partition, L, as random variables, hence $\boldsymbol{\xi}$ and L are estimated along with other model parameters, which makes the computational cost extremely expensive. Instead, we fix L as a constant, and take $\boldsymbol{\xi}$ as the known equally-spaced partition (Davis et al., 2006; Guo and Dai, 2006). In the ℓ-th partition component, the number of time points, $T_\ell = \lfloor T/L \rfloor$ where $\lfloor \cdot \rfloor$ denotes the round towards floor, are the same for all $\ell = 1, ..., L-1$. In the last element of the partition, $T_L = T - \lfloor T/L \rfloor \times (L-1)$ to satisfy either of two situations when T can, or cannot, be fully divided by L. Accordingly, the partition points are given by $\xi_\ell = \ell \lfloor T/L \rfloor$ for $\ell = 1, ..., L-1$. On the one hand, L should be sufficiently large to ensure that the slow varying model can lead to smooth estimates of the power spectra over time, and the abrupt changing model is given a rich set of change-point candidates to consider. On the other hand, T_ℓ should also be chosen such that there are enough observations in each partition. In practice, we have found out by sensitivity analysis that having at least 100 observations in each partition generally works well. Hence, we set $T_\ell = 100$, and $L = \lfloor T/100 \rfloor$. In reality, the partition can also be prespecified for the specific application and hand

and the scientific questions of interest. The proposed priors on splines coefficients over partitions, including the time varying double gamma shrinkage prior defined in Section 3.1.2, and the abrupt changing prior in Section 3.1.2, help preserve high model accuracy for both slow varying and piecewise stationary analysis. The customized variational Bayes inference scheme in Section 3.2.1 results in fast and scalable posterior inference. The model selection procedure proposed in Section 3.2.2 automatically picks the winner prior. Extensive simulation studies in Section 3.3 show that the inferences are accurate and time efficient, and that our method is superior to competing methods.

3.1.2 Prior Specification

To guarantee positive-definiteness and flexible smoothing of $\boldsymbol{f}$, we use a Cholesky factorization of the inverse of the power spectrum and then model the components of this factorization via smoothing splines (Rosen et al., 2009, 2012; Zhang, 2016; Li and Krafty, 2019). By modified Cholesky decomposition, the local power spectrum can be represented as

$$\boldsymbol{f}_\ell^{-1}(\omega) = \boldsymbol{\Theta}_\ell(\omega)\boldsymbol{\Psi}_\ell(\omega)^{-1}\boldsymbol{\Theta}_\ell(\omega)^*, \quad \ell = 1, \ldots, L, \tag{3.2}$$

where $\boldsymbol{\Psi}_\ell(\omega)$ is a positive diagonal matrix and $\boldsymbol{\Theta}_\ell(\omega)$ is a complex-valued $P \times P$ lower triangular matrix with ones on the diagonal. Let $\psi_{\ell jj}(\omega)$ be the j-th diagonal element of $\boldsymbol{\Psi}_\ell(\omega)$ for $j = 1, ..., P$. Let $\theta_{\ell jk}(\omega)$ be the jk element in $\boldsymbol{\Theta}_\ell(\omega)$ for $j > k$, $k = 1, ..., P$. Every $\psi_{\ell jj}(\omega)$, $\Re\{\theta_{\ell jk}(\omega)\}$, and $\Im\{\theta_{\ell jk}(\omega)\}$ can be modeled by smoothing splines with Demmler-Reinsch bases Eubank (1999) as of (Rosen

et al., 2012; Zhang, 2016; Li and Krafty, 2019; Li et al., 2021)

$$\begin{aligned}
\Re\left\{\theta_{\ell jk}(\omega)\right\} &= \alpha_{\ell jk0} + \alpha_{\ell jk1}\,\omega + \sum_{s=1}^{S-1} \alpha_{\ell jk(s+1)}\,\sqrt{2}\cos(\pi s\omega), \\
\Im\left\{\theta_{\ell jk}(\omega)\right\} &= \beta_{\ell jk0} + \beta_{\ell jk1}\,\omega + \sum_{s=1}^{S} \beta_{\ell jk(s+1)}\,\sqrt{2}\cos(\pi s\omega), \\
\log\left\{\psi_{\ell jj}(\omega)\right\} &= \gamma_{\ell jj0} + \gamma_{\ell jj1}\,\omega + \sum_{s=1}^{S-1} \gamma_{\ell jj(s+1)}\,\sqrt{2}\cos(\pi s\omega).
\end{aligned} \tag{3.3}$$

S is is a pre-specified constant determining the number of basis functions to be included and therefore the model flexibility. We follow the previous studies and set $S = 10$ (Rosen et al., 2012; Zhang, 2016; Li and Krafty, 2019; Li et al., 2021) to accounts for sufficiently large total variance of the full smoothing spline when $T \leq 4000$ (Krafty et al., 2017). One later variant of (3.3) changes the cosine basis on $\Im\left\{\theta_{\ell jk}(\omega)\right\}$ into sine basis to account for $\Im\left\{\theta_{\ell jk}(\omega)\right\}$ being an odd function (Zhang, 2016; Li and Krafty, 2019; Li et al., 2021). In our simulation studies and a number of data analyses, we have found out that both forms provide quite similar estimates.

For notation brevity, let $\boldsymbol{\gamma}_{\ell jj} = \{\gamma_{\ell jj0}, ..., \gamma_{\ell jjS}\}'$, $\boldsymbol{\alpha}_{\ell jk} = \{\alpha_{\ell jk0}, ..., \alpha_{\ell jkS}\}'$, and $\boldsymbol{\beta}_{\ell jk} = \{\beta_{\ell jk0}, ..., \beta_{\ell jkS}\}'$. Let $\boldsymbol{B}_{\ell n} = \{1, \omega_{\ell n}, \sqrt{2}\cos(\pi\omega_{\ell n}), ..., \sqrt{2}\cos(\pi(S-1)\omega_{\ell n})\}$ be the spline basis vector at Fourier frequency $\omega_{\ell n}$ in ℓ-th partition. Then (3.3) can be written as:

$$\begin{gathered}
\Re\left\{\theta_{\ell jk}(\omega_{\ell n})\right\} = \boldsymbol{B}_{\ell n}\boldsymbol{\alpha}_{\ell jk}, \quad \Im\left\{\theta_{\ell jk}(\omega_{\ell n})\right\} = \boldsymbol{B}_{\ell n}\boldsymbol{\beta}_{\ell jk}, \\
\log\left\{\psi_{\ell jj}(\omega_{\ell n})\right\} = \boldsymbol{B}_{\ell n}\boldsymbol{\gamma}_{\ell jj}.
\end{gathered} \tag{3.4}$$

Taking (3.2) and (3.4) into (3.1) while fixing $\boldsymbol{\xi}$ and L, we have that the total

log-likelihood can be rewritten as

$$\mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{\alpha}_{\ell jk}, \boldsymbol{\beta}_{\ell jk}, \boldsymbol{\gamma}_{\ell jj}) \propto$$
$$-\sum_{\ell=1}^{L}\sum_{n=1}^{N_\ell}\sum_{j=1}^{P}\left\{\boldsymbol{B}_{\ell n}\boldsymbol{\gamma}_{\ell jj} + \frac{|Y_{\ell nj} - \sum_{k=1}^{j-1}(\boldsymbol{B}_{\ell n}\boldsymbol{\alpha}_{\ell jk} + \mathrm{i}\,\boldsymbol{B}_{\ell n}\boldsymbol{\beta}_{\ell jk})\,Y_{\ell nk}|^2}{\exp\{\boldsymbol{B}_{\ell n}\boldsymbol{\gamma}_{\ell jj}\}}\right\}, \tag{3.5}$$

where $Y_{\ell nj}$ is the j-th element (dimension) of $\boldsymbol{Y}_{\ell n}$. Estimating $\boldsymbol{f}$ is now equivalent to estimating $\boldsymbol{\alpha}_{\ell jk}$, $\boldsymbol{\beta}_{\ell jk}$, and $\boldsymbol{\gamma}_{\ell jj}$ for all $\ell = 1, ..., L$, $j > k = 1, ..., P$. We now consider two types of priors on the model parameters. These prior classes lead to two types of models for spectral analysis of multivariate non-stationary time series data, one model that allows for slow-varying changes of the spectral density matrix over time, and another model that considers abrupt changes of the spectral density matrix.

Slow varying prior

We consider a prior structure on the Cholesky components that allows some of the components to remain static, while others can vary over time. In addition, it also induces sparsity. To construct posterior estimates that satisfy all above properties, we proposed using time-varying double gamma shrinkage priors (Bitto and Frühwirth-Schnatter, 2019) on $\boldsymbol{\alpha}_{\ell jk}$, $\boldsymbol{\beta}_{\ell jk}$, and $\boldsymbol{\gamma}_{\ell jj}$. Note that Bitto and Frühwirth-Schnatter (2019) developed such priors for time domain time-varying parameter models. Here we use them in the frequency domain. Formally, for every $\ell = 1, ..., L$, $j > k$, $k = 1, ..., P$, and $s = 0, ..., S$, we define the slow varying prior (SV-Prior) distributions as follows:

- Priors on the parameters that model the real parts of the off diagonal ele-

ments in the Cholesky decomposition:

$$\begin{aligned}
\alpha_{\ell jks} = \alpha_{jks} + \sqrt{\delta^2_{\alpha jks}}\,\tilde{\alpha}_{\ell jks}, \quad & \tilde{\alpha}_{\ell jks} \sim N\left(\tilde{\alpha}_{(\ell-1)jks},\, 1\right), \\
\alpha_{jks} \sim N\left(0,\, \tau^2_{\alpha jks}\right), \quad & \tau^2_{\alpha jks} \sim Ga\left(a_\tau,\, a_\tau \lambda^2_{\alpha jk}/2\right), \qquad \lambda^2_{\alpha jk} \sim Ga(e_1,\, e_2), \\
\delta^2_{\alpha jks} \sim Ga\left(1/2,\, 1/(2\zeta^2_{\alpha jk})\right), \quad & \zeta^2_{\alpha jk} \sim Ga\left(a_\zeta,\, a_\zeta \kappa^2_{\alpha jk}/2\right), \qquad \kappa^2_{\alpha jk} \sim Ga\left(d_1,\, d_2\right),
\end{aligned}$$

- Priors on the parameters that model the imaginary parts of the off diagonal elements in the Cholesky decomposition:

$$\begin{aligned}
\beta_{\ell jks} = \beta_{jks} + \sqrt{\delta^2_{\beta jks}}\,\tilde{\beta}_{\ell jks}, \quad & \tilde{\beta}_{\ell jks} \sim N\left(\tilde{\beta}_{(\ell-1)jks}, 1\right), \\
\beta_{jks} \sim N\left(0, \tau^2_{\beta jks}\right), \quad & \tau^2_{\beta jks} \sim Ga\left(a_\tau, a_\tau \lambda^2_{\beta jk}/2\right), \qquad \lambda^2_{\beta jk} \sim Ga(e_1,\, e_2), \\
\delta^2_{\beta jks} \sim Ga\left(1/2,\, 1/(2\zeta^2_{\beta jk})\right), \quad & \zeta^2_{\beta jk} \sim Ga\left(a_\zeta,\, a_\zeta \kappa^2_{\beta jk}/2\right), \qquad \kappa^2_{\beta jk} \sim Ga\left(d_1,\, d_2\right),
\end{aligned}$$

- Priors on the parameters that model the real diagonal elements in the Cholesky decomposition:

$$\begin{aligned}
\gamma_{\ell jjs} = \gamma_{jjs} + \sqrt{\delta^2_{\gamma jjs}}\,\tilde{\gamma}_{\ell jjs}, \quad & \tilde{\gamma}_{\ell jjs} \sim N\left(\tilde{\gamma}_{(\ell-1)jjs},\, 1\right), \\
\gamma_{jjs} \sim N\left(0,\, \tau^2_{\gamma jjs}\right), \quad & \tau^2_{\gamma jjs} \sim Ga\left(a_\tau,\, a_\tau \lambda^2_{\gamma jj}/2\right), \qquad \lambda^2_{\gamma jj} \sim Ga(e_1,\, e_2), \\
\delta^2_{\gamma jjs} \sim Ga\left(1/2,\, 1/(2\zeta^2_{\gamma jj})\right), \quad & \zeta^2_{\gamma jj} \sim Ga\left(a_\zeta,\, a_\zeta \kappa^2_{\gamma jj}/2\right), \qquad \kappa^2_{\gamma jj} \sim Ga\left(d_1,\, d_2\right).
\end{aligned}$$

Log-transformations are further proposed on all the $\delta^2_\cdot$, $\tau^2_\cdot$, $\lambda^2_\cdot$, $\zeta^2_\cdot$, and $\kappa^2_\cdot$ to guarantee that their final posterior estimates are positive values. Finally, we set $a_\tau = a_\zeta = 1$ as hyperparameters (Bayesian Lasso). We set $e_1 = e_2 = d_1 = d_2 = 0.001$; values of e_1, e_2, d_1 and d_2 in [0.1, 0.0001] lead to similar results. The priors on initial parameters $\tilde{\alpha}_{0jks}$, $\tilde{\beta}_{0jks}$ and $\tilde{\gamma}_{0jks}$ are chosen to be $\mathrm{N}(0,\, \sigma_0^2)$. The choice of σ_0^2 is robust for $1 \leq \sigma_0^2 \leq 10^4$, hence, we simply set $\sigma_0^2 = 1$.

We now highlight some important features of the proposed prior structure. Since the prior structure on $\boldsymbol{\alpha}_{\ell jk}$, $\boldsymbol{\beta}_{\ell jk}$, and $\boldsymbol{\gamma}_{\ell jj}$ is the same, we solely explain the main features of the prior on $\boldsymbol{\alpha}_{\ell jk}$ for illustrative purposes. First, each $\alpha_{\ell jsk}$ is decomposed into the sum of one static term, denoted as α_{jks}, that serves as a baseline remaining the same over all time partitions ℓ, and a time-varying term, denoted as $\hat{\alpha}_{\ell ks}$, that allows us to incorporate changes over time partitions. Such

parameterization allows us to handle static and time-varying coefficients separately in order to achieve posterior inference. Second, double gamma (or normal gamma, accordingly) shrinkage priors are imposed on both, the baseline α_{jks} and $\delta^2_{\alpha jks}$ (the variation of $\hat{\alpha}_{\ell ks}$) to allow for different evolving patterns on $\alpha_{\ell jks}$. Hence, for the Cholesky components: if both α_{jks} and $\delta^2_{\alpha jks}$ (or $\delta^2_{\alpha jks}$ only) escape from the shrinkage, $\alpha_{\ell jks}$ will change over ℓ, causing the corresponding Cholesky component to change over time; if α_{jks} escapes from the shrinkage but $\delta^2_{\alpha jks}$ shrinks towards zero, $\alpha_{\ell jks}$ will be static but non-zero; if both α_{jks} and $\delta^2_{\alpha jks}$ shrink towards zero, $\alpha_{\ell jks}$ will also shrink towards zero in both time and frequency domains. Such prior structure enables the model to capture various evolving patterns of the power spectrum.

Note that, Bitto and Frühwirth-Schnatter (2019) have provided two options for choosing a_τ and a_ζ. One is to treat them as model parameters and impose an exponential prior on them. The other is to fix $a_\tau = a_\zeta = 1$, which reduces the double gamma to a special case, the Bayesian Lasso prior (Park and Casella, 2008). Since we have found out that these two options result in very similar power spectrum estimates and using fixed value simplifies the model, we apply Bayesian Lasso by default.

Abrupt changing prior

To take into account the possible abrupt changing behaviour of the power spectrum for nonstationary time series, we also develop the following prior for change-point detection. For every $\ell = 1, ..., L$, $j > k$, $k = 1, ..., P$, and $s = 0, ..., S$,

the abrupt changing prior (AC-Prior) is defined as follows:

$$\alpha_{\ell jks} = \alpha_{jks} + \sum_{i=1}^{L} \tilde{\alpha}_{\ell jks}\, \mathbb{I}_{\{\ell > i\}}, \quad \beta_{\ell jks} = \beta_{jks} + \sum_{i=1}^{L} \tilde{\beta}_{\ell jks}\, \mathbb{I}_{\{\ell > i\}},$$
$$\gamma_{\ell jjs} = \gamma_{jjs} + \sum_{i=1}^{L} \tilde{\gamma}_{\ell jjs} \mathbb{I}_{\{\ell > i\}},$$

with $\mathbb{I}_{\{\ell > i\}} = 1$ if $\ell > i$ and $\mathbb{I}_{\{\ell > i\}} = 0$ otherwise,

$$\tilde{\alpha}_{\ell jks},\, \tilde{\beta}_{\ell jks},\, \tilde{\gamma}_{\ell jjs} \sim N\left(0, \frac{\tau^2 \lambda_\ell^2}{1 + \tau^2 \lambda_\ell^2}\right),$$

with $\lambda_\ell \sim \mathrm{C}^+(0, 1)$, $\tau \sim \mathrm{C}^+(0, 0.01)$, and finally,

$$\begin{aligned}
\alpha_{jks} &\sim N\left(0,\, \tau^2_{\alpha jks}\right), \quad \tau^2_{\alpha jks} \sim Ga\left(a_\tau,\, a_\tau \lambda^2_{\alpha jk}/2\right), \quad \lambda^2_{\alpha jk} \sim Ga(e_1,\, e_2), \\
\beta_{jks} &\sim N\left(0,\, \tau^2_{\beta jks}\right), \quad \tau^2_{\beta jks} \sim Ga\left(a_\tau,\, a_\tau \lambda^2_{\beta jk}/2\right), \quad \lambda^2_{\beta jk} \sim Ga(e_1,\, e_2), \\
\gamma_{jjs} &\sim N\left(0,\, \tau^2_{\gamma jjs}\right), \quad \tau^2_{\gamma jjs} \sim Ga\left(a_\tau,\, a_\tau \lambda^2_{\gamma jj}/2\right), \quad \lambda^2_{\gamma jj} \sim Ga(e_1,\, e_2),
\end{aligned}$$

with $a_\tau = a_\zeta = 1$, and $e_1 = e_2 = d_1 = d_2 = 0.001$ as in Section 3.1.2. The log-transformations are further proposed on all $\tau^2_\cdot$ and $\lambda^2_\cdot$ to guarantee that their posterior estimates are positive.

The prior formulation, different from the slow varying prior defined in Section 3.1.2, allows spline coefficients to be piecewise constant and change abruptly at certain partitions if needed. The piecewise linear functions are used such that $\alpha_{\ell jks}$, $\beta_{\ell jks}$, and $\gamma_{\ell jjs}$ remain piecewise constant over ℓ. Considering that the piecewise stationary processes usually have only few change points spread out over time, a regularized horseshoe prior (Piironen and Vehtari, 2017) that belongs to the global-local shrinkage family is imposed on the piecewise parameters $\tilde{\alpha}_{\ell jks}$, $\tilde{\beta}_{\ell jks}$, and $\tilde{\gamma}_{\ell jjs}$. The global parameter τ controls the shrinkage effect for all partitions, while the local parameters λ_l allows few individual partitions to escape from

the shrinkage if the partition contains a change point such that there is enough information leading to the change of values of the power spectrum, consequently, the spline coefficients. Note that, if ξ_ℓ is not a change point, all the corresponding piecewise coefficients $\tilde{\alpha}_{\ell jks}$, $\tilde{\beta}_{\ell jks}$, and $\tilde{\gamma}_{\ell jjs}$ should shrink towards zero simultaneously. Hence, for each ℓ, $\tilde{\alpha}_{\ell jks}$, $\tilde{\beta}_{\ell jks}$, and $\tilde{\gamma}_{\ell jjs}$ share the same local parameter λ_ℓ to achieve shrinkage effect in a group manner. Finally, the Bayesian lasso prior defined in Section 3.1.2 is still proposed on the baseline α_{jks}, β_{jks}, and γ_{jjs} to allow both zero and non-zero baseline estimates.

3.2 Posterior Computation

3.2.1 Variational Bayes Posterior

In this section, we describe our approach to adapt variational Bayes techniques to obtain posterior inference of the proposed nonstationary multivariate spectral model. We take advantage of the power of modern computational resources, such as graphic processing units (GPUs), in order to provide a flexible, scalable and computationally efficient inference scheme for high-dimensional time series. Our posterior inference scheme is based on the stochastic gradient variational Bayes (SGVB) approach (Kingma and Welling, 2013). To allow a more compact presentation (Li et al., 2021), we define $\boldsymbol{\Xi}$ as the collection of all parameters, including coefficients of basis functions, $\left\{\alpha_{jks}\,,\,\tilde{\alpha}_{\ell jks}\,,\,\beta_{jks}\,,\,\tilde{\beta}_{\ell jks}\,,\,\gamma_{jjs}\,,\,\tilde{\gamma}_{\ell jjs}\right\}$, and other prior parameters, such as $\{\log\tau_{\cdot}^2\,,\,\log\lambda_{\cdot}^2\,,\,\log\delta_{\cdot}^2\,,\,\log\zeta_{\cdot}^2\,,\,\log\kappa_{\cdot}^2\}$, for $\ell = 1, ..., L$, $j > k = 1, ..., P$, and $s = 1, ..., S$. The joint log density of frequency-domain observations $\boldsymbol{Y}$ and model parameters $\boldsymbol{\Xi}$ can be written as:

$$\log p(\boldsymbol{\Xi}, \boldsymbol{Y}) = \mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{\Xi}) + \log \pi(\boldsymbol{\Xi}), \tag{3.6}$$

where $\mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{\Xi})$ is defined in (3.5), and $\pi(\boldsymbol{\Xi})$ is the joint prior defined by either the SV-Prior in Section 3.1.2 or the AC-Prior in Section 3.1.2. As mentioned in Section 2.2, to mitigate the VI uncertainty underestimation without losing run-time efficiency, we consider the Gaussian variational approximation with a factor covariance structure of Ong et al. (2018). Given a m_1-dimensional $\boldsymbol{\Xi}$, its surrogate posterior is defined as $q_{\boldsymbol{\Phi}}(\boldsymbol{\Xi}) = N(\boldsymbol{\Xi} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + I_{m_1}\boldsymbol{\sigma}^2$ and $\mathbf{B}$ is a $m_1 \times m_2$ full rank matrix with $m_2 << m_1$. The reparameterization trick for the factor covariance structure can then be conducted by calculating $\boldsymbol{\Xi} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$ with $\boldsymbol{\xi} \sim \mathbf{N}(\mathbf{0},\, I_{m_2})$ and $\boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0},\, I_{m_1})$, where I_{m_1}, I_{m_2} denote the m_1, m_2-dimensional identity matrices. $\odot$ denotes element-wise product. The log-transformation is further taken on $\boldsymbol{\sigma}^2$ to guarantee the positivity. Accordingly, $\boldsymbol{\Phi} = \{\boldsymbol{\mu}\,, \log \boldsymbol{\sigma}^2, \boldsymbol{B}\}$ are surrogate parameters, and the ELBO between $p(\boldsymbol{\Xi} \mid \boldsymbol{Y})$ and $q_{\boldsymbol{\Phi}}(\boldsymbol{Y})$ becomes:

$$\mathcal{L}\left(p, q_{\boldsymbol{\Phi}}\right) = \mathrm{E}_{\boldsymbol{\xi},\boldsymbol{\epsilon}}\left[\log p(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \boldsymbol{Y}) - \log q_{\boldsymbol{\Phi}}(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon})\right]. \tag{3.7}$$

With a single sample $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\epsilon}^{(1)}$, the SGVB gradient estimator of (3.7) is given by:

$$\nabla_{\boldsymbol{\Phi}}\mathcal{L}\left(p, q_{\boldsymbol{\Phi}}\right) \simeq \nabla_{\boldsymbol{\Phi}}\left[\log p(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi}^{(1)} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}^{(1)}, \boldsymbol{Y}) - \log q_{\boldsymbol{\Phi}}(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi}^{(1)} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}^{(1)})\right]. \tag{3.8}$$

We set $m_2 = 4$ as suggested in Ong et al. (2018). We have empirically found that (3.8) led to a fast convergence practically.

In practice, we propose the locally stationary variational Bayes posterior inference scheme, as illustrated in Algorithm 2, to stabilize the inference process. In particular, Phase 1 uses (3.6) as the objective function to obtain a point approximation, $\hat{\boldsymbol{\Xi}} = \arg\max_{\boldsymbol{\Xi}} \log p(\boldsymbol{\Xi}, \boldsymbol{Y})$. Phase 2 fixes $\boldsymbol{\mu} = \hat{\boldsymbol{\Xi}}$ and only updates $\log \boldsymbol{\sigma}^2$ by SGVB to maximize (3.7). Finally, both previously updated $\boldsymbol{\mu}$ and $\log \boldsymbol{\sigma}^2$ are

Algorithm 2: Locally Stationary Variational Bayes Posterior Inference.

Input : Local DFT transformed observations of P-dimensional time series, $\boldsymbol{Y}$.

Output: Variational posteriors $q_{\boldsymbol{\Phi}}(\boldsymbol{\Xi})$, and power spectrum estimates.

Params: $\boldsymbol{\Phi} = \{\boldsymbol{\mu}, \log \boldsymbol{\sigma}^2, \boldsymbol{B}\}$.

Phase 1–Point Approximation: maximizes (3.6) via $\boldsymbol{\Xi}$, **while** *iterating* **do**
  Take the gradients of $\boldsymbol{\Xi}$ from (3.6);
  Update $\boldsymbol{\Xi}$ via `Adam` optimizer;
end
Fix $\boldsymbol{\mu}$ equal to the last updated $\boldsymbol{\Xi}$ in Phase 1.
Phase 2–Uncertainty Quant: maximize (3.7) via $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$, **while** *iterating* **do**
  Draw $\boldsymbol{\xi}^{(1)} \sim N(\mathbf{0}, I_{m_1})$, $\boldsymbol{\epsilon}^{(1)} \sim N(\mathbf{0}, I_{m_2})$ and compute the SGVB gradients of $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ using (3.8);
  Update $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ via `Adam` optimizer;
end
Phase 3–Fine-tuning (Optional): **while** *iterating* **do**
  Draw $\boldsymbol{\xi}^{(1)} \sim N(\mathbf{0}, I_{m_1})$, $\boldsymbol{\epsilon}^{(1)} \sim N(\mathbf{0}, I_{m_2})$ and compute the SGVB gradients of $\boldsymbol{\mu}$, $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ using (3.8);
  Update $\boldsymbol{\mu}$, $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ via `Adam` optimizer;
end
Draw posterior samples of $\boldsymbol{\Xi}$ from the learnt $q_{\boldsymbol{\Phi}}(\boldsymbol{\Xi})$ and compute power spectrum estimates.

fine-tuned at Phase 3 by SGVB to maximize (3.7). Results from our extensive simulation studies below show that our proposed TPVB achieves high numerical stability. We built the model and implemented the experiments in `Python 3.7` with `Tensorflow-Probability` and `Tensorflow` (Abadi et al., 2015) packages. All gradients are computed via auto-differentiation modules in `Tensorflow`. At each iteration, we use `Adam` (Kingma and Ba, 2017) to update model parameters.

3.2.2 Model Selection

In this subsection, we discuss how to choose the proper prior during modeling. We have developed the SV-Prior in Section 3.1.2, and the AC-Prior in Section

3.1.2 to tackle both slow varying and abrupt changing nonstationary processes. In practice, we propose to fit the data using both models individually. Then, we use the deviance information criterion (DIC; Spiegelhalter et al., 2014) to select the superior model. In this context, DIC is defined as:

$$\text{DIC} = D(\overline{\boldsymbol{\Xi}}) + 2p_D$$

where $\overline{\boldsymbol{\Xi}}$ is the posterior expectation of $\boldsymbol{\Xi}$ that is obtained by Algorithm 2 using either SV-Prior or AC-Prior. $D(\boldsymbol{\Xi}) = -2\log\mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{\Xi}) + C$ is the likelihood of $\boldsymbol{Y}$ defined in (3.5) adding a constant C that cancels out in all calculations that compare different models, and which therefore does not need to be known. $p_D = \overline{D(\boldsymbol{\Xi})} - D(\overline{\boldsymbol{\Xi}})$ where $\overline{D(\boldsymbol{\Xi})}$ can be approximated by averaging over the posterior samples of $D(\boldsymbol{\Xi})$ that is calculated by the variational Bayes posterior samples of $\boldsymbol{\Xi}$ that are obtained through Algorithm 2 in Section 3.2.1. Each model runs completely independently, thus the procedure can be conducted in parallel.

3.3 Simulation Studies

In this section, we show the accuracy and efficiency of the proposed method through simulations. During the variational posterior inference process, the learning rate is used as a hyper-parameter to control the rate at which the algorithm updates the parameter estimates. The learning rates in `Adam` optimizer are set to be 0.001, 0.05 and 0.001, respectively for each phase, to obtain gradient descent updates for trainable parameters. We decide the number of iterations by monitoring the values of the objective function, and stop the training when a typical convergent pattern is reached. That leads to approximately 12000, 500, and 500 iterations required for each of the 3 phases. The simulation experiments are ex-

ecuted on a x64-based PC with a 2.60-GHz Intel® Core™ i7-9750H CPU and a Nvidia GeForce GTX 1660 Ti GPU card.

3.3.1 Evaluation Metrics

We consider two scenarios: slow varying processes and abrupt changing piecewise stationary processes. In each scenario, we first show that our approach is quite suitable for high-dimensional ($P \geq 24$) analysis. There is a paucity of literature that studies the performance of the current existing approaches on non-stationary spectral analysis of high-dimensional time series when $P \geq 24$ due to the fact that most of current approaches are not scalable for a large P. Li et al. (2021) have showed that their Bayesian factor model improved the accuracy of power spectrum estimates when $P = 24, 48$, in comparison to the rolling-window method from Shumway and Stoffer (2011). We show in this section that our method further improves the accuracy with a more affordable amount of runtime in comparsion to Li et al. (2021). The mean integrated squared error (MISE) (Li et al., 2021) is used to investigate the performance of an estimator of the high-dimensional spectral matrix $\boldsymbol{f}(u, \omega)$, which is calculated by

$$\text{MISE} = [T(N+1)]^{-1} \sum_{t=1}^{T} \sum_{n=0}^{N} \left\| \hat{\boldsymbol{f}}\left(t/T, \omega_n\right) - \boldsymbol{f}\left(t/T, \omega_n\right) \right\|_F^2,$$

where $\|\cdot\|_F$ is the matrix Frobenius norm. $\{\omega_n : n = 1, ..., N\}$ is the prespecified equally spaced frequency grid. $\hat{\boldsymbol{f}}$ is the obtained posterior estimate of $\boldsymbol{f}$. Furthermore, the performance of diagonal and off-diagonal estimators are investigated

separately by

$$\text{MISE}_d = [T(N+1)]^{-1} \sum_{t=1}^{T} \sum_{n=0}^{N} \left\| \text{diag}[\hat{\boldsymbol{f}}\,(t/T, \omega_n)] - \text{diag}[\boldsymbol{f}\,(t/T, \omega_n)] \right\|^2,$$
$$\text{MISE}_o = [T(N+1)]^{-1} \sum_{t=1}^{T} \sum_{n=0}^{N} \left\| \hat{\boldsymbol{f}}\,(t/T, \omega_n) - \boldsymbol{f}\,(t/T, \omega_n) \right\|_{F*}^2,$$

where $\|A\|_{F*} = \sqrt{\sum_{i=1}^{P} \sum_{j \neq i} A_{ij}^2}$. We use MISE, MISE_d and MISE_o to jointly evaluate the performance of candidate methods.

In addition, we show that the proposed method is also competitive in low-dimensional analysis (when $P = 2, 3$) by comparing it with several other current existing methods that have been studied well in low-dimensional analysis, but have insufficient scalability in high dimensional cases. More specifically, we follow the simulation studies by Li and Krafty (2019) and compare our method with the adaptive Bayesian smoothing spline by Li and Krafty (2019), the piecewise vector autoregressive method by Davis et al. (2006), and the nonparametric smoothing spline ANOVA method by Guo and Dai (2006). Considering that there are only a few terms in the power spectrum with $P = 2, 3$, every diagonal and off-diagonal estimates can be fully evaluated separately. There is no need to integrate them together. Hence, instead of using MISE, the average squared errors (ASE) of the estimated individual spectra, $f_{pp}(u, \omega)$ for $p = 1, ..., P$, and the estimated squared coherences, $\rho_{jk}^2(u, \omega) = |f_{jk}(u, \omega)|^2 / \{f_{jj}(u, \omega) f_{kk}(u, \omega)\}$ for $j < k$, $k = 1, ..., P$, are used to evaluate the performance of the candidate methods when $P = 2, 3$ (Li and Krafty, 2019). For example, the ASE for $\hat{f}_{11}$ is calculated as

$$\text{ASE}[\hat{f}_{11}] = [T(N+1)]^{-1} \sum_{t=1}^{T} \sum_{n=0}^{N} \left[\hat{f}_{11}\,(t/T, \omega_n) - f_{11}\,(t/T, \omega_n) \right]^2.$$

To have a fair comparison, we use the same equally spaced frequency grid $N = 50$

as Li and Krafty (2019) do.

3.3.2 Slow Varying Processes

We first consider a P-dimensional process with a slow varying power spectrum. We simulate 50 time series of length $T = 2048$ independently from

$$\boldsymbol{X}_t = \Phi_t \boldsymbol{X}_{t-1} + \boldsymbol{w}_t, \tag{3.9}$$

where $\boldsymbol{w}_t \overset{iid}{\sim} N(\boldsymbol{0}, \Omega)$ with $\Omega = I_{P/3} \otimes \Omega^0$, and Ω^0 is a 3×3 matrix that has 1's on the diagonal and 0.5 off the diagonal. $\otimes$ denotes the Kronecker product. $I_{P/3}$ is a $P/3 \times P/3$ identity matrix. $\Phi_t = I_{P/2} \otimes \Phi_t^0$ with

$$\Phi_t^0 = \begin{pmatrix} \phi_1^0(t) & 0.1 \\ 0 & \phi_2^0(t) \end{pmatrix},$$

and $\phi_1^0(t) = -0.6 + 1.2t/T$, $\phi_2^0(t) = 0.4 - 0.8t/T$ for $t = 1, ..., T$. The true $P \times P$ time varying power spectrum is $\boldsymbol{f}(t/T, \omega) = \Phi^{-1}(t/T, \omega)\, \Omega\, \Phi^{-1}(t/T, \omega)^*$, where $\Phi(t/T, \omega) = I - \Phi_t \exp(-2\pi \mathrm{i}\, \omega)$ (Priestley, 1981). We set $T_\ell = 100$ such that there are in total 20 partitions and the 20 partition points $\boldsymbol{\xi}$ are fixed to be the 100-th, 200-th, ..., and 2000-th time points. Figure 3.1 displays some of the true log power spectra components, coherences and their estimates obtained by the proposed method.

Through the MISE criteria defined in 3.3.1, Table 3.1 compares our approach with another high-dimensional approach, the Bayesian factor model of Li et al. (2021) with its default tuning parameters having a total of 10,000 MCMC iterations with 2000 burn-in iterations and the numbers of factors to be 16. It can be seen that the proposed method outperforms the factor model as the former

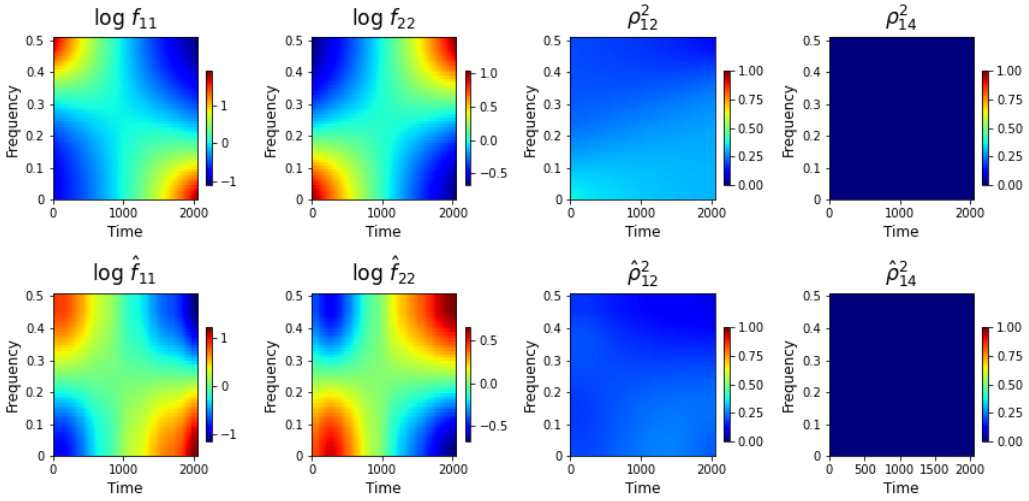

Figure 3.1: True time varying log power spectrum, $\log f_{11}(u, \omega)$, $\log f_{22}(u, \omega)$ and coherences, $\rho^2_{1,2}(u, \omega)$, $\rho^2_{1,4}(u, \omega)$ (top two rows), and their estimates (bottom two rows) via the proposed method from the slow varying process (3.9).

provides more accurate estimates that gain smaller MISEs. As for time efficiency, for instance, when $P = 24$, the Bayesian factor model of Li et al. (2021) takes on average 8350 seconds ($\approx$ 139 mins, over 2 hrs) to obtain posterior inference based on 10,000 MCMC iterations using its published code with default tuning parameters. We note that this implementation does not support GPU computation. In comparison, the average runtime for one run of our method is 1056 seconds ($\approx$ 18 mins) on CPU, and only 225 seconds ($\approx$ 4 mins) with GPU acceleration, which greatly saved computational time.

Additionally, to show that the proposed method is also competitive (high accuracy, low runtime) in low-dimensional cases, we follow the simulation study by Li and Krafty (2019) and compare it with the adaptive Bayesian smoothing spline (MultiSpect) by Li and Krafty (2019), the piecewise vector autoregressive method (Auto-PARM) by Davis et al. (2006), and the nonparametric smoothing spline ANOVA method (SmoothANOVA) by Guo and Dai (2006). In particular,

Table 3.1: Simulation results for the slow varying process (3.9) based on 50 repetitions: means (standard deviations) of the MISE, MISE_o, and MISE_d given by the proposed method (SVAC) and the Bayesian factor method by Li et al. (2021) (FactorSpect).

P	Model	MISE	MISE_d	MISE_o
12	SVAC	**1.35**(0.10)	**0.90**(0.06)	**0.45**(0.03)
	FactorSpect	1.85 (0.06)	1.02 (0.04)	0.83 (0.02)
24	SVAC	**2.84**(0.23)	**1.81**(0.19)	**1.03**(0.07)
	FactorSpect	4.55 (0.09)	2.54 (0.08)	2.01 (0.07)
48	SVAC	**6.25**(0.52)	**3.56**(0.33)	**2.69**(0.25)
	FactorSpect	9.55 (0.11)	5.03 (0.09)	4.52 (0.08)

we simulate 250 2-dimensional datasets generated as

$$\boldsymbol{X}_t = \tilde{\boldsymbol{w}}_t + \Phi_{1t}\tilde{\boldsymbol{w}}_{t-1} + \Phi_{2t}\tilde{\boldsymbol{w}}_{t-2}, \quad t = 1, \ldots, 1024, \tag{3.10}$$

where

$$\Phi_{1t} = \begin{pmatrix} \phi_1(t) & -1 \\ -1 & \phi_2(t) \end{pmatrix}, \quad \Phi_{2t} = \begin{pmatrix} 0.5 & 0 \\ 0 & -1.2 \end{pmatrix},$$

with $\phi_1(t) = 1.122[1-1.781\sin(\pi t/2048)]$, $\phi_2(t) = 1.122[1-1.781\cos(0.8\pi t/2048)]$. $\tilde{\boldsymbol{w}}_t$ are independent zero-mean bivariate Gaussian random variables whose components have unit variance and pairwise correlation 0.2. Table 3.2 shows that our method provide more accurate estimates over two out of three spectral targets. On average, one run of our method takes only approximately 1 min (< 2 mins), which is quite superior comparing to the runtime of MultiSpect (≈ 63 mins) with its default tuning parameters (10,000 MCMC iterations with a burn–in of 2000). GPU acceleration in this case is not necessary since $P = 2$ is rather small.

Table 3.2: Simulation results for the slow varying process (3.10) based on 250 repetitions: means (standard deviations) of the ASE $\times 10^2$ of spectral estimates obtained through the proposed method (SVAC), adaptive Bayesian smoothing spline (MultiSpect), piecewise vector autoregressive modeling (Auto-PARM), and smoothing spline ANOVA (SmoothANOVA).

	$\mathbf{f}_{11}$	$\mathbf{f}_{22}$	ρ^2_{21}
`SVAC`	**36.2**(19.2)	**38.4**(10.1)	1.3 (0.7)
`MultiSpect`	49.2 (19.3)	51.4 (26.3)	**1.1**(0.2)
`Auto-PARM`	90.6 (26.6)	85.5 (34.5)	7.0 (3.2)
`SmoothANOVA`	51.8 (20.3)	61.2 (22.5)	1.6 (1.3)

3.3.3 Piecewise Stationary Processes

In this subsection, we illustrate the performance of our method while modeling the piecewise stationary processes. We first consider a P-dimensional process that possesses a abrupt changing power spectrum. We simulate 50 time series of length $T = 2048$ independently from

$$\boldsymbol{X}_t = \begin{cases} \boldsymbol{w}_t + \Phi_{11}\boldsymbol{w}_{t-1} + \Phi_{12}\boldsymbol{w}_{t-2} & \text{if } 0 < t \leq 888, \\ \boldsymbol{w}_t + \Phi_{21}\boldsymbol{w}_{t-1} + \Phi_{22}\boldsymbol{w}_{t-2} & \text{if } 888 < t \leq 2048, \end{cases} \tag{3.11}$$

$\boldsymbol{w}_t \overset{iid}{\sim} N(\mathbf{0}, \Omega)$ with $\Omega = I_{P/3} \otimes \Omega^0$, and Ω^0 is a 3×3 matrix that has 1's on the diagonal and 0.5 off the diagonal. $\Phi_{11} = I_{P/3} \otimes \Phi^0_{11}$, $\Phi_{21} = I_{P/3} \otimes \Phi^0_{21}$,

$\Phi_{12} = I_{P/3} \otimes \Phi_{12}^0$, and $\Phi_{22} = I_{P/3} \otimes \Phi_{22}^0$ with

$$\Phi_{11}^0 = \begin{pmatrix} 0.6 & 0 & 0 \\ 0.2 & -0.5 & 0 \\ 0.1 & 0.3 & 0.4 \end{pmatrix}, \quad \Phi_{21}^0 = \begin{pmatrix} 0.6 & 0 & 0 \\ 0.2 & 0.5 & 0 \\ -0.1 & -0.3 & 0.4 \end{pmatrix}$$

$$\Phi_{12}^0 = \Phi_{22}^0 = \begin{pmatrix} 0.3 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The true power spectrum is $\boldsymbol{f}(t/T, \omega) = \Phi(t/T, \omega)\, \Omega\, \Phi(t/T, \omega)^*$ (Priestley, 1981), where $\Phi(t/T, \omega) = I + \Phi_{11} \exp(-2\pi \mathrm{i}\, \omega) + \Phi_{12} \exp(-4\pi \mathrm{i}\, \omega)$ for $t \in (0, 888]$, and $\Phi(t/T, \omega) = I + \Phi_{21} \exp(-2\pi \mathrm{i}\, \omega) + \Phi_{22} \exp(-4\pi \mathrm{i}\, \omega)$ for $t \in (888, 2048]$. We set $T_\ell = 100$ such that there are in total 20 partitions and the 20 partition points $\boldsymbol{\xi}$ are fixed to be the 100-th, 200-th, ..., and 2000-th time points. The true change point 888 is a blindly selected value such that a general situation, where the change point does not belong to the partition points $\boldsymbol{\xi}$, is considered. Figure 3.2 displays some of the true power spectrum, coherences and their estimates obtained by the proposed method. It can be seen from Table 3.3 that the proposed method outperforms the factor model as the former provides more accurate estimates that lead to smaller MISE and MISE$_o$. The run time performance is similar to Section 3.3.2. When $P = 24$, the Bayesian factor model takes on average 8,014 seconds ($\approx$ 133 mins, over 2 hrs) to run a inference process using its published code with default tuning parameters (10,000 MCMC iterations with a burin-in of 2000 and number of factors to be 16) on CPU and does not support GPU computation. In comparison, the average runtime for one run of our method is 1,005 seconds ($\approx$ 17 mins) on CPU, and only 213 seconds ($\approx$ 4 mins) with GPU acceleration.

Again, to clarify that the proposed method is also competitive in low-dimensional

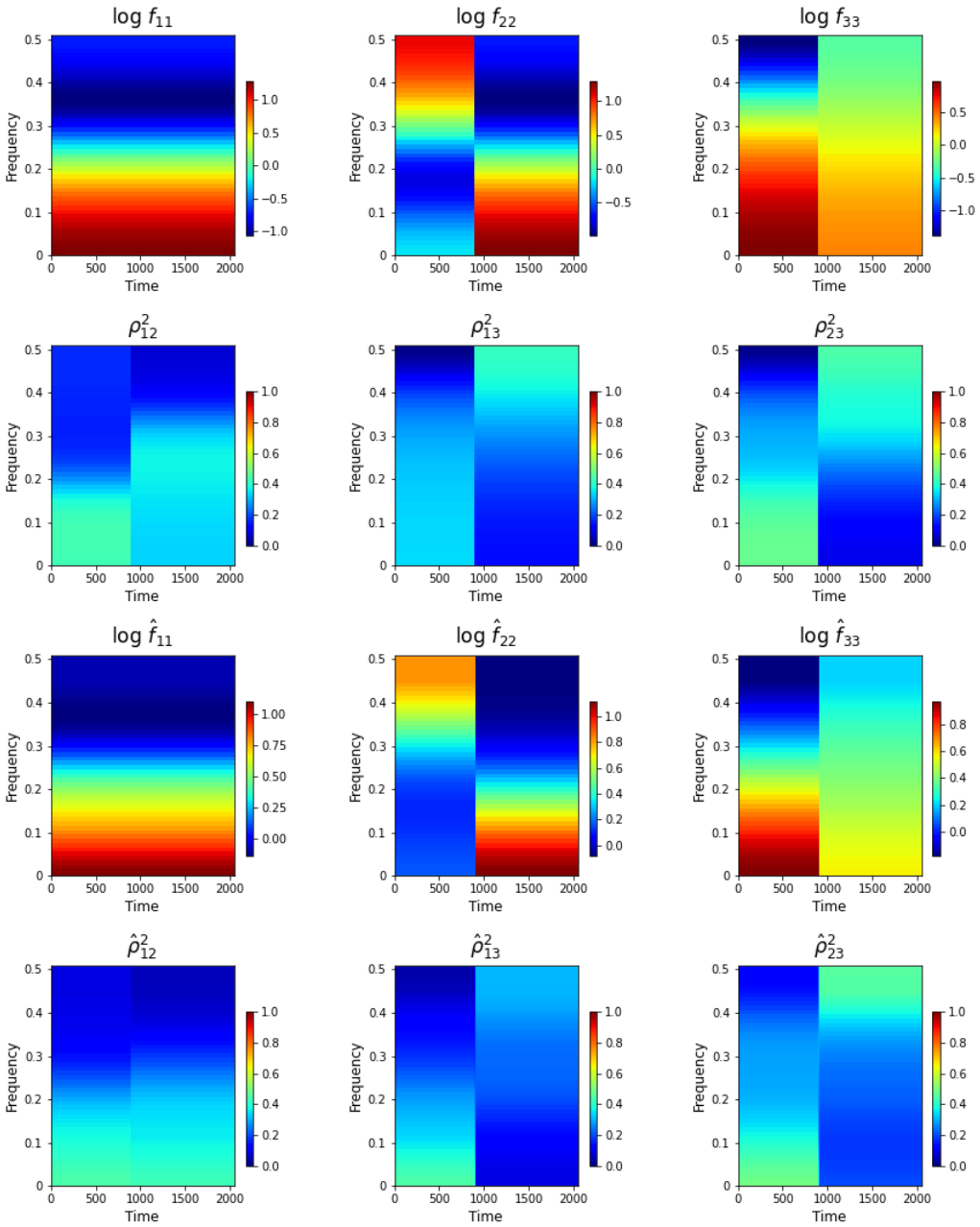

Figure 3.2: True time varying log power spectrum, $\log f_{11}(u,\omega)$, $\log f_{22}(u,\omega)$, $\log f_{33}(u,\omega)$ and coherences, $\rho^2_{1,2}(u,\omega)$, $\rho^2_{1,3}(u,\omega)$, $\rho^2_{2,3}(u,\omega)$ (top two rows), and their estimates (bottom two rows) via the proposed method from the piecewise stationary process (3.11).

Table 3.3: Simulation results for the piecewise stationary process (3.11) based on 50 repetitions: means (standard deviations) of the MISE, MISE_o, and MISE_d given by the proposed method (SVAC) and the Bayesian factor method by Li et al. (2021) (FactorSpect).

P	Model	MISE	MISE_d	MISE_o
12	SVAC	**1.53**(0.22)	0.63 (0.12)	**0.92**(0.15)
	FactorSpect	1.80 (0.08)	0.63 (0.04)	1.17 (0.06)
24	SVAC	**3.32**(0.26)	**1.40**(0.13)	**1.92**(0.18)
	FactorSpect	4.58 (0.12)	1.43 (0.06)	3.15 (0.10)
48	SVAC	**6.90**(0.32)	**2.63**(0.12)	**4.27**(0.21)
	FactorSpect	9.52 (0.16)	2.71 (0.09)	6.91 (0.14)

cases, we consider the piecewise stationary simulation process by Li and Krafty (2019) and simulate 250 repetitions of a 3-dimensional process

$$\boldsymbol{X}_t = \begin{cases} \tilde{\boldsymbol{w}}_t + \Phi_{11}\tilde{\boldsymbol{w}}_{t-1} + \Phi_{12}\tilde{\boldsymbol{w}}_{t-2} & \text{if } 0 < t \leq 300, \\ \tilde{\boldsymbol{w}} + \Phi_{21}\tilde{\boldsymbol{w}}_{t-1} + \Phi_{22}\tilde{\boldsymbol{w}}_{t-2} & \text{if } 300 < t \leq 600, \end{cases} \tag{3.12}$$

where

$$\Phi_{11} = \begin{pmatrix} 0.6 & 0 & 0 \\ 0.2 & -0.5 & 0 \\ 0.1 & 0.3 & 0.4 \end{pmatrix}, \Phi_{21} = \begin{pmatrix} 0.6 & 0 & 0 \\ 0.2 & 0.5 & 0 \\ -0.1 & -0.3 & 0.4 \end{pmatrix},$$

and $\Phi_{12} = \Phi_{22} = \text{diag}(0.3, 0.3, 0)$. The $\tilde{\boldsymbol{w}}_t$s are independent zero-mean trivariate Gaussian random variable whose components have unit variance and pairwise correlation 0.5. Table 3.4 shows that our method provide more accurate estimates over five out of six spectral targets. On average, one run of our method in this study takes only approximately 1 min ($<$ 2 mins) comparing to the run time of MultiSpect ($\approx$ 36 mins) with its default tuning parameters (10,000 MCMC

Table 3.4: Simulation results for the piecewise stationary process (3.12) based on 250 repetitions: means (standard deviations) of the ASE $\times 10^2$ of spectral estimates obtained through the proposed method (SVAC), adaptive Bayesian smoothing spline (MultiSpect), piecewise vector autoregressive modeling (Auto-PARM), and smoothing spline ANOVA (SmoothANOVA).

	$\mathbf{f}_{11}$	$\mathbf{f}_{22}$	$\mathbf{f}_{33}$	ρ^2_{21}	ρ^2_{31}	ρ^2_{32}
SVAC	**8.2** (6.4)	**18.1** (9.5)	**12.9** (2.5)	1.4 (0.4)	**1.6** (0.4)	**2.1**(0.9)
MultiSpect	25.1 (28.6)	22.8 (20.5)	13.8 (13.3)	1.4 (0.8)	2.1 (0.9)	2.2 (1.1)
Auto-PARM	37.2 (33.2)	25.4 (16.2)	14.3 (7.7)	6.2 (2.4)	3.8 (1.1)	9.0 (1.3)
SmoothANOVA	39.8 (10.6)	93.4 (16.2)	27.3 (6.4)	9.6 (0.5)	12.8(0.9)	13.4(1.2)

iterations with a burn–in of 2000). GPU acceleration is not necessary since both $P = 3$ and $T = 600$ are relatively small.

Note that, the proposed DIC-based model selection procedure successfully select the superior model prior for all above cases. Table 3.5 shows that when the slow varying processes (3.9) and (3.10) are given, the proposed slow varying prior is chosen in model selection. When modeling the piecewise stationary processes (3.11) and (3.12), the proposed abrupt changing prior is favorably selected accordingly.

Table 3.5: The percentage of datasets for which the proposed DIC-based model selection procedure in Section 3.2.2 selects the model with either the slow varying prior (SV-Prior) defined in Section 3.1.2 or the abrupt changing prior (AC-Prior) defined in Section 3.1.2.

Process	SV-Prior	AC-Prior
(3.9)	100%	0%
(3.10)	100%	0%
(3.11)	0%	100%
(3.12)	0%	100%

In summary, we have shown with extensive simulation studies that the pro-

posed method are accurate and time efficient, and is superior compared to the competing methods in modeling time series with time varying power spectrum under both high and low dimensionalities.

3.4 Data Analyses

3.4.1 Analysis of IEM Wind Profile in Northern California

In this section, we illustrate the use of our method for spectral analysis of wind profile data from the Iowa State University Environmental Mesonet (IEM) Automated Surface System (ASOS) database (Todey et al., 2002; Mannarano, 1998). This is a publicly available repository of automated airport weather observations and general basic weather reports from the National Weather Service (NWS), the Federal Aviation Administration (FAA) and the Department of Defense (DOE). We consider wind speed (measured in knots) data taken from sites in Northern California. Our goal is to infer the time varying power spectra of wind profiles in order to provide insight into the temporal dynamics of each location and the relationships across different locations over a particular period of time. Specifically, we consider the mean wind speed every 4 hours, starting from 06/01/2020 12:00 am to 05/31/2021 11:59 pm. We take the first order differences for each time series to remove any local trends in the data. To enhance numeric stability, all series are further normalized to have zero mean and unit variance. There are 11 sites that have no records on wind speed for days and so we remove those sites from the analysis. Finally, we end up having 26 wind profiles, each with length $T = 2186$, as model inputs. The names and locations of the corresponding 26 sites are shown in Figure 3.3.

We consider both types of spectral models, one that assumes a smooth chang-

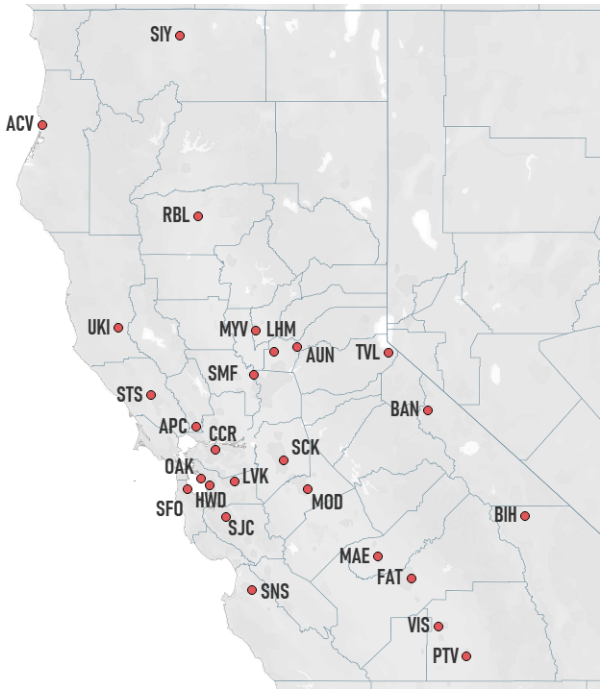

Figure 3.3: Map of 26 wind sites in Northern California.

ing structure and one that assumes a piecewise locally stationary structure. During variational Bayes posterior inference procedure, the learning rates are set to be 0.0005, 0.05 and 0.001, with the number of iterations 10000, 500, and 500 respectively for each of the three phases in Algorithm 2. The DIC selects the model with the piecewise locally stationary structure as the winning model. Based on the posterior inference we see that model infers two change points, one in 11/08/2020 and another in 02/26/2021. For illustrative purposes, we first visualize the complete time varying power spectra estimates for 6 sites: SFO (San Francisco), OAK (Oakland), HWD (Hayward), SJC (San Jose), MAE (Madera) and FAT (Fresno). Note that SFO, OAK, HWD and SJC are closely located in Bay area, while MAE and FAT are in neighboring counties that located in the south middle part of Northern California. Figure 3.4 displays the posterior estimates of the time varying spectral densities for these 6 sites. It can be seen that the estimated spectral

densities have strong signals at the daily frequency 4/24 (approximately 0.17).This indicates a daily quasi-periodicity since the data is collected every 4 hours. There is also a second peak at the half-day frequency (approximately 0.34) from June to November in 2020. We also not that the daily quasi-periodicicty has weaker power in late fall winter months, roughly from November to February, and later increases again in power starting in March 2021.

Figure 3.5 shows the posterior estimates of the time varying pairwise squared coherences among the 6 selected sites. Site pairs that are close to each other show stronger coherences in comparison to the site pairs that are far away from each other. For instance, the coherence estimates are also strong around the daily frequency (4/24, approximately 0.17) among OAK, HWD, SFO, SJC, and between MAE and FAT, while the coherences between pairs such as SJC versus FAT, OAK versus FAT, or SFO versus MAE, are estimated to be not as strong. One can observe similar evolving patterns of daily signals in coherence estimates as in spectral density estimates, which are generally strong in summer and fall, then decrease in winter and later come back in the next spring. Previous studies have justified that the wind profile and precipitation are closely related. For instance, Johansson and Chen (2003) has concluded that precipitation is statistically significantly associated with wind speed. Our analysis provides another insight on the connection between wind profile and precipitation: the estimated daily quasi-periodicity decreasing power behavior in winter for wind profiles in Northern California matches the annual precipitation change pattern in California. According to OEHHA (2018) (OEHHA), California on average receives 75 percent of its annual precipitation from November through March, with 50 percent occurring from December through February.

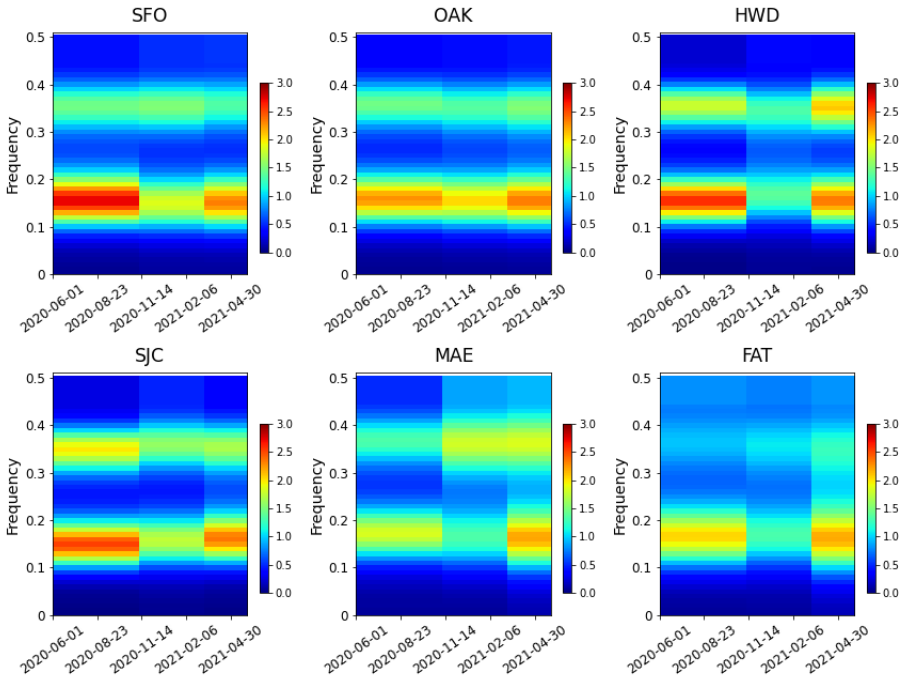

Figure 3.4: Posterior inference of the time varying spectral densities for SFO, OAK, HWD, SJC, MAE, and FAT wind profiles given by the proposed method.

3.4.2 Analysis of Multi-Channel EEG Data

The multi-channel EEG data analyzed here is part of a larger dataset named `Ictal19` which records 19 EEG channels from a patient during the electroconvulsive therapy (ECT) against major depression. Figure 3.6 presents the schematic representation of the channel locations. West et al. (1999); Prado et al. (2001) analyzed this dataset using univariate dynamic regression models separately for each channel. Zhao and Prado (2020) analyzed 9 channels of this dataset jointly using multivariate dynamic partial autocorrelation coefficients (PARCOR) models. The original recordings of 26,000 observations for each channel were downsampled every sixth observation from the highest amplitude portion of the seizure, leading to a set of series of 3600 observations (corresponding to 83.72 seconds) per channel (Prado et al., 2001). We conducted a multivariate spectral analysis on the 19-

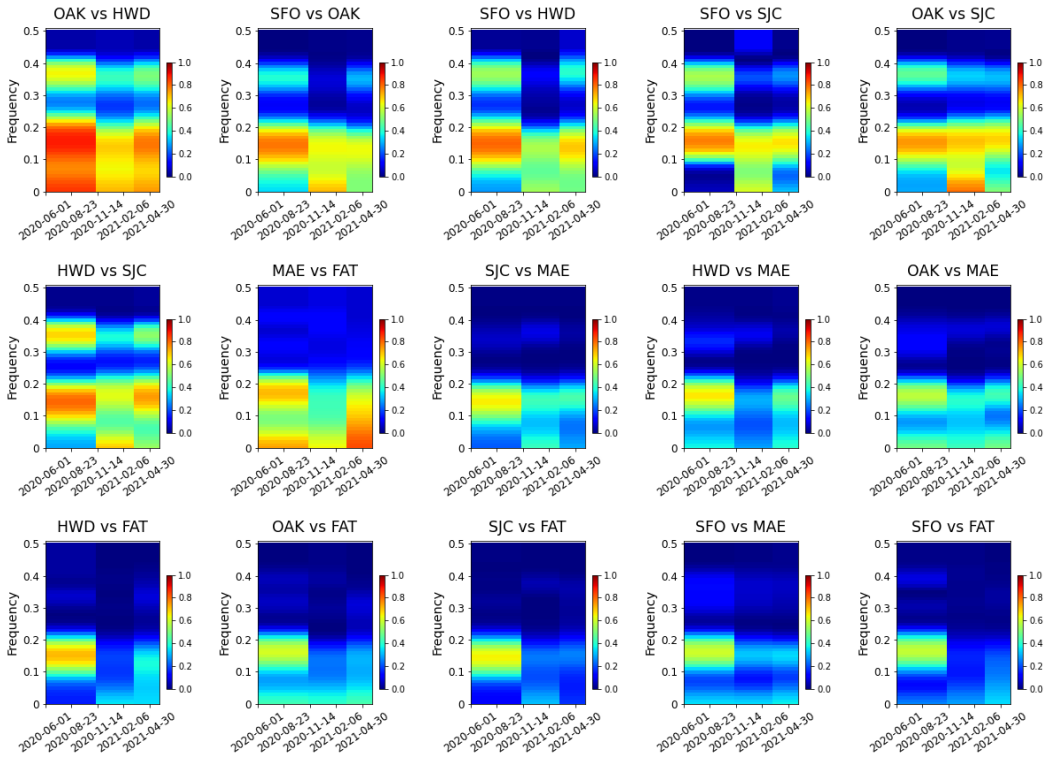

Figure 3.5: Posterior inference of the time varying pairwise squared coherences among SFO, OAK, HWD, SJC, MAE, and FAT given by the proposed method.

channel EEG data using the proposed method. During variational Bayes posterior inference procedure, the learning rates and iterations were set to be the same as in Section 3.4.1. The DIC procedure selects the model with the slow-varying prior structure as the winning model. By aid of GPU accelerations, the run time to search for the optimal model and obtain posterior inference is under 10 minutes in total.

We summarize the results in terms of posterior estimates of the spectral densities for the individual channels and also look at summary measures related to the coherence across channels for a specific frequency band. More specifically, we focus on computing frequency-band collapsed measures at the theta band (4–8 Hz), as this frequency band has been found to provide important neurophysiolog-

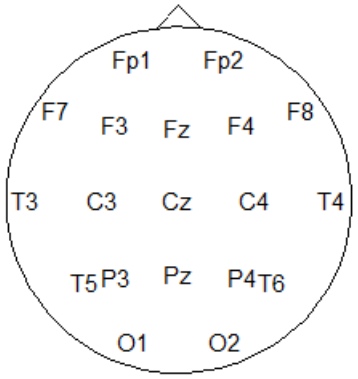

Figure 3.6: Schematic representation of the 19-channel EEG locations.

ical information (West et al., 1999; Prado et al., 2001). Frequency-band collapsed measures can be computed as integrals of the power spectra (Li et al., 2021). The theta-band collapsed squared coherence between channels j and l is defined as $\rho_{jl}^{2,\theta} = \left|\boldsymbol{f}_{jl}^{\theta}\right|^2 / \left\{\boldsymbol{f}_{jj}^{\theta}\boldsymbol{f}_{ll}^{\theta}\right\}$, where $\boldsymbol{f}^{\theta} = \int_4^8 \boldsymbol{f}(\omega)\, d\omega$ defines theta-band collapsed spectral matrix. Figure 3.7 shows the estimated time-varying spectral densities of 19 channels over the whole period of the time. Correspondingly, Figure 3.8 presents the estimated theta-band collapsed spectral densities on the schematic representation of the channel locations at the beginning (1 second), middle (40 seconds), and end (80 seconds) of the time series. It can be seen that the spectral densities are dominated by frequency components in the lower frequency band. Each EEG channel shows a similar decrease in the dominant frequency over time, starting around 5 Hz and ending around approximately 3 Hz. Some channels, such as FP1, F4, F8 and T4, present weaker power spectra at these frequencies than the remaining channels and are consistent with a decrease in seizure activity as the seizure disappears towards the end of the process. Similar behaviors were also found by West et al. (1999); Zhao and Prado (2020).

We can also obtain inferences about the change of the connectivity between

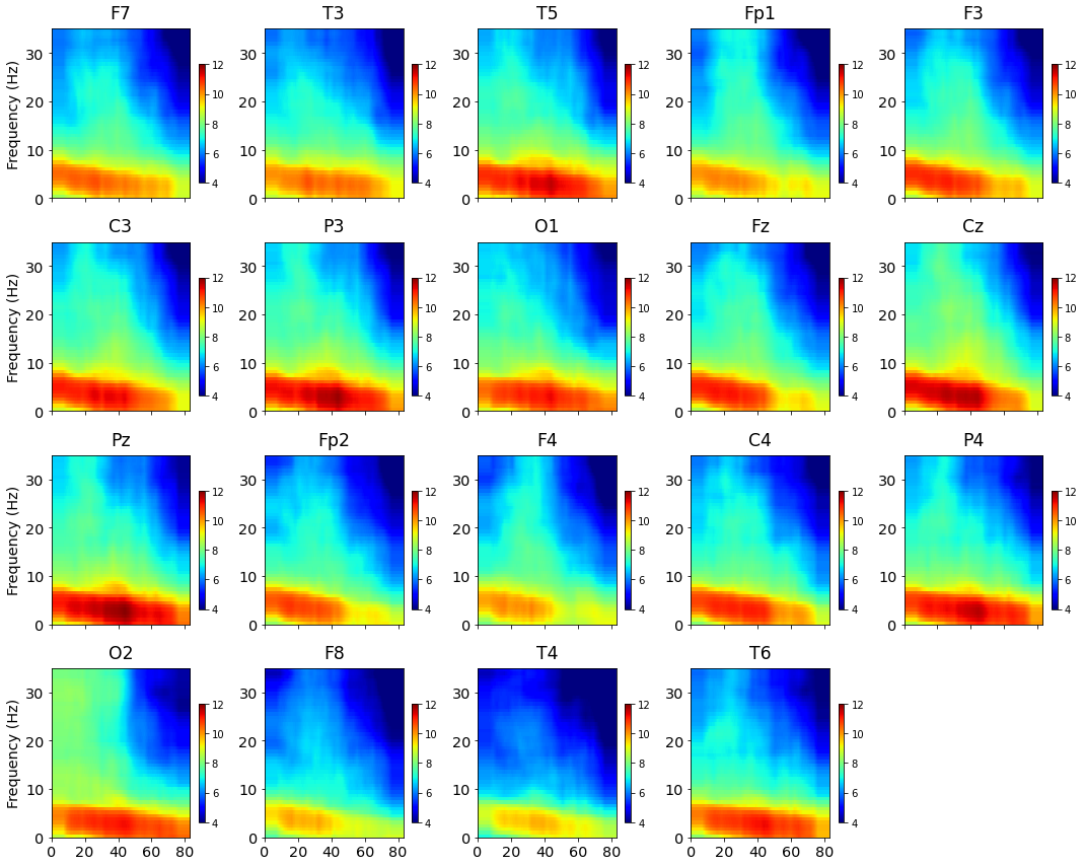

Figure 3.7: Posterior estimates of time varying spectral densities over time.

channels over time. For instance, Figure 3.9 displays the estimated theta-band collapsed squared coherences on the schematic representation of the channel locations (see Figure 3.6) between Cz and the rest of the channels at the beginning (1 second), middle (40 second), and end (80 second) of the entire process. It can be seen that Cz preserves relatively high theta-band coherences with most its neighboring channels, such as F3, Fz, C3, C4, P3, P4 and Pz, throughout the time. One exception is that, even though F4 is also close to Cz, the theta-band coherence between Cz and F4 shows a notable decrease from $t = 1$ to $t = 40$ and remains low until the end, which is consistent with the find of Zhao and Prado (2020). For those channels that are relatively far from Cz, their theta-band coherences with

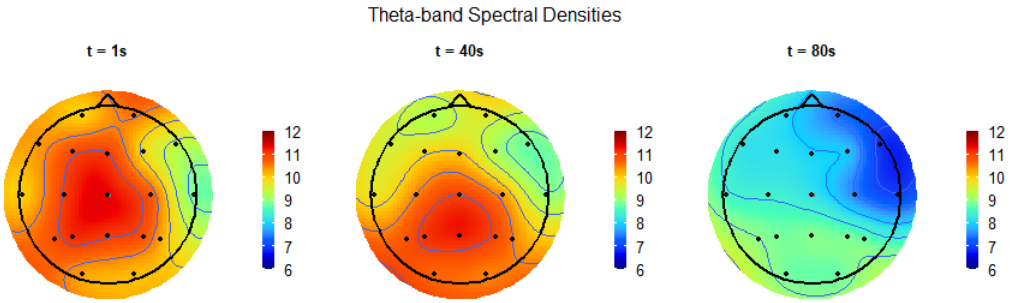

Figure 3.8: Posterior estimates of theta-band (4-8 Hz) collapsed spectral densities at time 1 second (left), 40 second (middle), and 80 second (right).

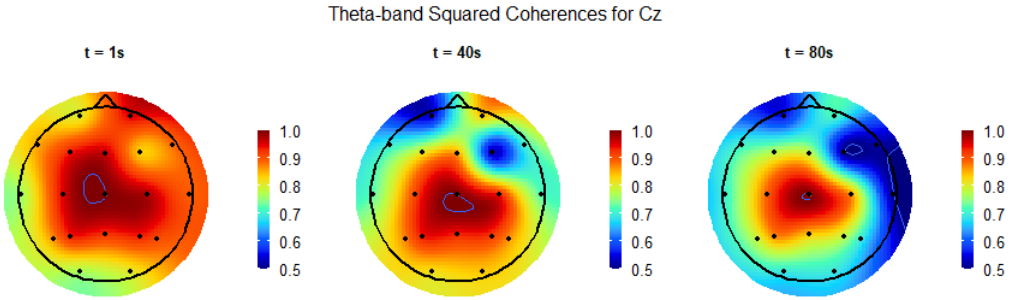

Figure 3.9: Posterior estimates of theta-band (4-8 Hz) collapsed squared coherences between Cz and the rest of the channels at time 1 second (left), 40 second (middle), and 80 second (right).

Cz all experience a remarkable decrease from the beginning to the end.

3.5 Conclusions and Future Work

The proposed novel modeling and inference framework provides accurate and computationally fast approximate Bayesian spectral inference for nonstationary multivariate time series. We showed the proposed time-varying double gamma shrinkage prior and the abrupt changing prior lead to accurate inference of the power spectrum in several simulation settings involving both slow varying and abrupt changing multivariate time series. The proposed inference scheme, that

utilizes stochastic gradient variational Bayes, is highly scalable and efficient, providing solid computational support for large-dimensional spectral analysis. The proposed model selection procedure is capable of picking up the right model. Extensive simulation studies in Section 3.3 show that the inferences are accurate and time efficient, and that our method is superior compared to the competing methods in various situations. The data analysis in Section 3.4 illustrates that our method is capable of discovering the time varying evolving patterns of power spectrum in the data.

To further improve the method, we will study whether the low-rank Bayesian factor model (Li et al., 2021) could be incorporated into our inference framework and further improved by utilizing the proposed shrinkage prior on its model parameters. So far, the proposed method relies on the modified complex Cholesky decomposition on the $P \times P$ spectral matrix, which still suffers from the "curse of dimensionality" since the number of Cholesky components that need to be estimated grows quadratically as dimension P increases. Such drawback will cause exhaustive run time and storage complexity if one wants to apply the analysis on a even larger scale when $P >> 100$. As a remedy, the factor model by Li et al. (2021) has its potential to be more feasible even when $P >> 100$ in that its low-rank factorization on the spectral matrix allows the number of model parameters to grow only linearly as P increases. We can replace the Cholesky decomposition with the low-rank factorization, impose the time varying shrinkage prior on the factor coefficients and conduct posterior inference using the proposed variational Bayes scheme. One of the challenges in the nonstationary setting is the selection of the number of factors as this can also be time-varying.

Another future work is to extend the proposed method to conduct efficient partition adjustment within current modeling procedure such that the equally

spaced partition can be further adjusted for the model with abrupt changing prior. Consequently, the adjusted partitions will no longer be restricted to be equally spaced, adding more flexibility into the detection of change points. We expect that the ideal approach could support parallel computation to be scalable and time-efficient. For instance, one can conduct a grid search in parallel over the time points around the change point estimate given by the model and formulate a selection criteria to evaluate whether the current estimate should be updated adaptively.

Chapter 4

Quantile Spectral Analysis of Multivariate Stationary Time Series

In this chapter, we develop a novel scalable Bayesian approach for the quantile spectral analysis of multivariate stationary time series based on a low-rank factorization of the target copula spectral density kernel (CSDK) matrix and a customized stochastic gradient variational inference (SGVI) scheme for posterior computation. The remainder of the chapter is organized as follows. Section 4.1 specifies the model and priors used for quantile spectral analysis of multivariate stationary time series. Section 4.2 describes the proposed posterior inference scheme adapted from the stochastic gradient variational Bayes. Section 4.3 reports results of extensive simulation studies that illustrate the accuracy and scalability of the proposed approach in comparison to the current alternatives. Section 4.4 applies the proposed method to the analysis of the San Lorenzo river flow data in California (U.S. Geological Survey, 2016) and the multivariate S&P 500 stock

data (Nugent, 2017).

4.1 Model Specification

Let $\{\boldsymbol{X}_t : t = 1, ..., n\}$ be a length-n P-dimensional zero-mean strictly stationary time series, with $\boldsymbol{X}_t = (X_{t,1}, ..., X_{t,P})'$ where $X_{t,p}$ denotes the p-th component of $\boldsymbol{X}_t$ for $p = 1, ..., P$. Let $\boldsymbol{\tau} = (\tau_1, ..., \tau_Q)'$ be a length-Q quantile vector where $\tau_i \in (0, 1)$ for $i = 1, ..., Q$. To measure the serial and cross-dependency structure of $\boldsymbol{X}_t$, the matrix of quantile cross-covariance kernels at lag h for quantiles τ_i and τ_j (Baruník and Kley, 2019) has entries (p_1, p_2) given by

$$\gamma_h^{p_1,p_2}(\tau_i, \tau_j) = \mathrm{Cov}\left(\mathbb{I}\left\{F_{p_1}(X_{t+h,p_1}) \leq \tau_i\right\}, \mathbb{I}\left\{F_{p_2}(X_{t,p_2}) \leq \tau_j\right\}\right), \tag{4.1}$$

for $p_1, p_2 = 1, ..., P$, and $i, j = 1, ..., Q$. F_p denotes the marginal distribution function of $X_{t,p}$, and $\mathbb{I}$ denotes the indicator function. The matrix of quantile cross-spectral density kernels (CSDKs) is then defined as

$$f_{\tau_i,\tau_j}^{p_1,p_2}(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma_h^{p_1,p_2}(\tau_i, \tau_j) \exp(-\mathrm{i}\,\omega h), \tag{4.2}$$

with $\omega \in (0, \pi]$. The quantile coherence, given by

$$\rho_{\tau_i,\tau_j}^{p_1,p_2}(\omega) = \frac{\mid f_{\tau_i,\tau_j}^{p_1,p_2}(\omega) \mid}{\left[f_{\tau_i,\tau_i}^{p_1,p_1}(\omega) \times f_{\tau_j,\tau_j}^{p_2,p_2}(\omega)\right]^{\frac{1}{2}}},$$

is defined as the absolute value of the quantile coherence kernel of Baruník and Kley (2019) for X_{t,p_1} under quantile τ_i and X_{t,p_2} under quantile τ_j to measure the dynamic dependence of the two processes.

Estimation of the quantile cross-spectral density matrix can be obtained by computing the so called ranked-based copula cross-periodograms (Baruník and

Kley, 2019), defined as the collection

$$I^{p_1,p_2}_{\tau_i,\tau_j}(\omega_k) = Y^{p_1}_{\tau_i}(\omega_k)Y^{p_2}_{\tau_j}(-\omega_k), \tag{4.3}$$

for $\omega_k = 2\pi k/n$, $k = 1, \ldots, K$, $K = [(n-1)/2]$, with $Y^p_\tau(\omega_k)$ denoting the discrete Fourier transform (DFT) of the clipped time series or CT (Hong, 2000; Zhang, 2019) for $\{X_{t,p} : t = 1, ..., n\}$ under quantile τ given by

$$Y^p_\tau(\omega_k) = (\sqrt{2\pi n})^{-1} \sum_{t=1}^{n} \mathbb{I}_{(0,\tau]}(n^{-1}R^n_{t,p}) \exp(-\mathrm{i}\,\omega_k t), \tag{4.4}$$

for $p = 1, ..., P$. Here $R^n_{t,p}$ denotes the rank of $X_{t,p}$ among $X_{1,p}, ..., X_{n,p}$.

Now, for each $\tau \in [0, 1]$, denote $\boldsymbol{Y}_\tau(\omega) = (Y^1_\tau(\omega), ..., Y^P_\tau(\omega))'$. Then, under some standard mixing conditions given in Proposition 3.2 of Baruník and Kley (2019), we obtain asymptotic normality for $\boldsymbol{Y}_\tau(\omega)$, i.e., $\boldsymbol{Y}_\tau(\omega)$ converges to a zero-mean P-dimensional circular complex Gaussian distribution with covariance matrix having entries $f^{p_1,p_2}_{\tau,\tau}(\omega)$ as n goes to infinity, and furthermore, $\boldsymbol{Y}_\tau(\omega)$ are asymptotically independent for distinct ωs.

For each component p of $\boldsymbol{X}_t$, we can define $\boldsymbol{Y}^p(\omega) = (Y^p_{\tau_1}(\omega), \ldots, Y^p_{\tau_Q}(\omega))'$ similarly. Note that, $\boldsymbol{Y}^p(\omega_k)$ is used to estimate the quantile cross-spectrum for univariate $X_{t,p}$. In order to estimate (4.2) for multivariate $\boldsymbol{X}_t$, we combine $\boldsymbol{Y}^p(\omega_k)$ for all p, into a joint PQ-dimensional asymptotically normal CT-based DFT vector $\boldsymbol{Y}_k = (\boldsymbol{Y}^1(\omega_k)', \ldots, \boldsymbol{Y}^P(\omega_k)')'$, whose $[i + Q(p-1)]$-th component is $Y^p_{\tau_i}(\omega_k)$ as defined in (4.4). Then, the asymptotic $PQ \times PQ$ complex-valued covariance matrix of $\boldsymbol{Y}_k$, denoted as $\boldsymbol{f}(\omega_k)$, has its entry $([i + Q(p_1 - 1)], [j + Q(p_2 - 1)])$ corresponding to the quantile cross-spectral density kernel, $f^{p_1,p_2}_{\tau_i,\tau_j}(\omega_k)$, defined in (4.2). $\boldsymbol{Y}_k$'s are approximately independent multivariate complex Gaussian random vectors with mean zero and covariance matrix $\boldsymbol{f}(\omega_k)$, allowing us to use the

multivariate extension of the Whittle approximation (Whittle, 1957). We use this approximation below after utilizing a factor model representation of $\boldsymbol{Y}_k$ for dimension reduction.

4.1.1 Factor Model

As explained above, the $PQ \times PQ$ quantile spectral matrix function $\boldsymbol{f}$ has its entry $([i + Q(p_1 - 1)], [j + Q(p_2 - 1)])$ corresponding to $f^{p_1,p_2}_{\tau_i,\tau_j}(\omega)$ defined in (4.2), providing a full characterization of the quantile cross-spectral association between X_{t,p_1} and X_{t,p_2} under quantile pair (τ_i, τ_j) for every $p_1, p_2 = 1, ..., P$ and $i, j = 1, ..., Q$. However, the size of $\mathbf{f}$ grows quadratically on both P and Q, making the inference procedure rather challenging due to the computational burden. Then, to achieve model scalability, we utilize the factor model representation of Li et al. (2021) to write $\boldsymbol{Y}_k$ as

$$\boldsymbol{Y}_k = \boldsymbol{\Lambda}(\omega_k)\boldsymbol{D}_k + \boldsymbol{E}_k, \tag{4.5}$$

where $\boldsymbol{D}_k$ is Q_F-dimensional and approximately independently distributed from a complex multivariate normal distribution with mean zero and covariance matrix equal to the identity $\boldsymbol{I}_{Q_F}$ of dimension $Q_F \times Q_F$, with $Q_F << PQ$. $\boldsymbol{\Lambda}(\omega)$ is a $PQ \times Q_F$ complex-valued loading factor matrix that is a function of the frequency ω. $\boldsymbol{E}_k$ is independently distributed from a complex multivariate normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}_E = \text{diag}\left\{\sigma^2_{E,1}, ..., \sigma^2_{E,PQ}\right\}$. This leads to a parsimonious representation of $\boldsymbol{f}(\omega_k)$ given by

$$\boldsymbol{f}(\omega_k) = \boldsymbol{\Lambda}(\omega_k)\boldsymbol{\Lambda}(\omega_k)^* + \boldsymbol{\Sigma}_E, \tag{4.6}$$

where $*$ denotes conjugate transpose. Note that (4.6) reduces the model complexity of $\boldsymbol{f}$ from $\mathcal{O}(P^2Q^2)$ to $\mathcal{O}(PQQ_F)$. Estimating $\boldsymbol{f}$ is equivalent to estimating

$\boldsymbol{\Lambda}(\cdot)$ and $\boldsymbol{\Sigma}_E$.

The conditional Whittle log-likelihood implied by (4.5) can be written as

$$\mathcal{L}\left(\boldsymbol{Y} \mid \boldsymbol{\Lambda}, \boldsymbol{D}, \boldsymbol{\Sigma}_E\right) \approx \sum_{k=1}^{K} \sum_{p=1}^{PQ} \left\{ \log[\sigma_{E,p}^2]^{-1} - \sigma_{E,p}^{-2} \left| Y_{k,p} - \boldsymbol{\Lambda}_p(\omega_k) \boldsymbol{D}_k \right|^2 \right\}, \tag{4.7}$$

where $\boldsymbol{Y}$ represents $\{\boldsymbol{Y}_k : k = 1, ..., K\}$. $\boldsymbol{\Lambda}$ and $\boldsymbol{D}$ represent $\{\Lambda(\omega_k) : k = 1, ..., K\}$ and $\{\boldsymbol{D}_k : k = 1, ..., K\}$. $Y_{k,p}$ is the p-th element of $\boldsymbol{Y}_k$, $\sigma_{E,p}^2$ is the p-th diagonal element of $\boldsymbol{\Sigma}_E$, and $\boldsymbol{\Lambda}_p(\omega_k)$ is the p-th row of $\boldsymbol{\Lambda}(\omega_k)$. Note that (4.7) allows us to estimate $\boldsymbol{\Lambda}(\omega)$ and $\boldsymbol{\Sigma}_E$ from $\boldsymbol{Y}$.

4.1.2 Prior Specification

To achieve flexible smoothing of $\boldsymbol{f}(\omega)$, we model the $PQ \times Q_F$ complex-valued loading matrix $\boldsymbol{\Lambda}(\omega)$ by smoothing splines (Li et al., 2021; Rosen and Stoffer, 2007) with Demmler-Reinsch bases, i.e.,

$$\begin{aligned} \Re\left\{\Lambda_{pq}(\omega)\right\} &= \alpha_{pq0} + \sum_{s=1}^{S-1} \alpha_{pqs} \sqrt{2} \cos(2\pi s \omega), \\ \Im\left\{\Lambda_{pq}(\omega)\right\} &= \sum_{s=1}^{S} \beta_{pqs} \sqrt{2} \sin(2\pi s \omega), \end{aligned} \tag{4.8}$$

where $\Lambda_{pq}(\omega)$ is (p, q)-th element of $\boldsymbol{\Lambda}(\omega)$ for $p = 1, ..., PQ$ and $q = 1, ..., Q_F$. Here $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote, respectively, the operators that extract the real and imaginary parts of a complex-valued quantity. The α_{pqs}'s and β_{pqs}'s in the equations above are the spline coefficients that control the shape of $\boldsymbol{\Lambda}(\omega)$. For notation brevity, let $\boldsymbol{\alpha}_{pq} = (\alpha_{pq0}, ..., \alpha_{pq(S-1)})'$, and $\boldsymbol{\beta}_{pq} = (\beta_{pq1}, ..., \beta_{pqS})'$. Furthermore, let $\boldsymbol{\alpha}_p = (\boldsymbol{\alpha}_{p1}, ..., \boldsymbol{\alpha}_{pQ_F})$ be the $S \times Q_F$ matrix formed by binding $\boldsymbol{\alpha}_{pq}$ for $q = 1, ..., Q_F$ columnwise and $\boldsymbol{\beta}_p = (\boldsymbol{\beta}_{p1}, ..., \boldsymbol{\beta}_{pQ})$ be the $S \times Q_F$ matrix formed by binding $\boldsymbol{\beta}_{pq}$ for $q = 1, ..., Q_F$ columnwise. Let $\boldsymbol{B}_k^{(re)} = \Big(1, \sqrt{2}\cos(2\pi\omega_k), ..., \sqrt{2}\cos(2\pi(S-1)\omega_k)\Big)$,

and $\boldsymbol{B}_k^{(im)} = \left(\sqrt{2}\sin(2\pi\omega_k), ..., \sqrt{2}\sin(2\pi S\omega_k)\right)$ be the spline basis row vectors at Fourier frequency ω_k. Then the p-th row of $\boldsymbol{\Lambda}(\omega)$, denoted as $\Lambda_p(\omega_k)$, at Fourier frequency ω_k can be written in terms of its real and imaginary parts as follows:

$$\Re\left\{\Lambda_p(\omega_k)\right\} = \boldsymbol{B}_k^{(re)}\boldsymbol{\alpha}_p, \quad \Im\left\{\Lambda_p(\omega_k)\right\} = \boldsymbol{B}_k^{(im)}\boldsymbol{\beta}_p. \tag{4.9}$$

Taking (4.9) into (4.7), the Whittle log-likelihood can be rewritten as

$$\begin{gathered}\mathcal{L}\left(\boldsymbol{Y} \mid \boldsymbol{D}, \boldsymbol{\Sigma}_E, \boldsymbol{\alpha}_p, \boldsymbol{\beta}_p : p = 1, ..., PQ\right) \approx \\ \sum_{k=1}^{K}\sum_{p=1}^{PQ}\left\{\log[\sigma_{E,p}^2]^{-1} - \sigma_{E,p}^{-2}\left|Y_{k,p} - (\boldsymbol{B}_k^{(re)}\boldsymbol{\alpha}_p + \mathrm{i}\,\boldsymbol{B}_k^{(im)}\boldsymbol{\beta}_p)\boldsymbol{D}_k\right|^2\right\}.\end{gathered} \tag{4.10}$$

Priors on spline coefficients are needed to obtain posterior inferences. An ideal prior setting should allow the selection of enough basis terms to achieve a good fit, while also avoiding overfitting for each component. To achieve this, our prior setting is based upon the so-called regularized horseshoe prior (Piironen and Vehtari, 2017). This is a global-local shrinkage prior with a global parameter that provides shrinkage towards zero for all the components sharing this parameter, and local, or component-specific parameters that allow some of individual components to escape from the shrinkage. Formally, we define the full prior distributions on the whole model parameters as follows. First, we impose no shrinkage on $\sigma_{E,p}^2$ and assume that $\sigma_{E,p}^2 \sim N(0, 10^2)$ for all p. This is based on the idea that the individual spectral densities will have at the very least a baseline basis representation that is non-zero (e.g., the spectral density of a white noise process is a linear constant function over the frequencies), and possibly a more sophisticated structure that can be captured by the regularized horseshoe prior on the spline coefficients of the loading factor matrix as explained below. We have found that a prior variance of 10 for $\sigma_{E,p}^2$, (or any value within [10, 10^5] that leads to a relatively vague prior) for

these terms typically works well in practice. Then, for $s = 0, ..., S$, the regularized horseshoe prior is imposed on α_{pqs} and β_{pqs}. Specifically, we set

$$\alpha_{pqs} \mid \tau_{pq}, \lambda_{pqs,(re)} \sim N\left(0, \frac{c^2\tau_{pq}^2\lambda_{pqs,(re)}^2}{c^2 + \tau_{pq}^2\lambda_{pqs,(re)}^2}\right),$$

$$\beta_{pqs} \mid \tau_{pq}, \lambda_{pqs,(im)} \sim N\left(0, \frac{c^2\tau_{pq}^2\lambda_{pqs,(im)}^2}{c^2 + \tau_{pq}^2\lambda_{pqs,(im)}^2}\right).$$

We fix c as a constant and discuss the selection of c below. Note that, this setting is different from the shrinkage prior setting by (Li et al., 2021) where α_{pqs} and β_{pqs} have individual global parameters. In this setting, the global parameters τ_{pq}'s control the overall roughness of the spectral density matrix for each $\Lambda_{pq}(\omega)$ as a function of the frequency. τ_{pq} imposes the joint regularization over α_{pqs} and β_{pqs} which allows the spline coefficients of the real and imaginary parts of the loading factor components to have similar smoothness levels. We show in the simulation studies that the joint regularization improves model accuracy. We assume $\tau_{pq} \sim \mathrm{C}^+(0, c_\tau)$, where C^+ denotes Half-Cauchy distribution and c_τ is a fixed constant. We also have local parameters $\lambda_{pqs,(re)}$'s and $\lambda_{pqs,(im)}$'s which adjust the roughness as a function of the frequency according to individual Demmler-Reinsch basis with $\lambda_{pqs,(re)}, \lambda_{pqs,(im)} \sim \mathrm{C}^+(0, c_s)$. Here, c_s is a pre-specified "discount effect" function varying with s. The choices of c_τ and c_s are described below.

We have found that the posterior inference is robust to the choice of hyper-parameters, with $c \in (1, 10^5)$ and $c_\tau \in (10^{-5}, 10^{-1})$ providing indistinguishable estimates in all empirical examples considered. Thus, we simply set $c = 10$ and $c_\tau = 0.01$ by default. The choice of c_τ matches the suggestion of Piironen and Vehtari (2017) that $c_\tau < 1$ achieves good shrinkage performance. In addition to the regularization obtained from the regularized horseshoe prior above, we further penalize the appearance of basis according to the value of s by introducing a

heuristic discount effect function c_s. Coefficients of basis terms with larger s will be more likely to be shrunk towards zero, unless the need to include them in the model is strongly supported by the data. In this way, the proposed prior will automatically preserve the informative bases and will shrink towards zero those that are not needed to avoid overfitting. Inspired by the shape of the sigmoid function $\text{Sig}(x) = [1+\exp(-x)]^{-1}$, we choose the form of c_s as $c_s = \text{Sig}(-s+S/2)$ such that c_s is a monotone decreasing function of s and is bounded between 0 and 1. This choice is justified through experimentation via simulation studies and real data analyses.

We refer to the prior structure above as the discounted regularized horseshoe (DRH), and the DRH-FM model to the proposed Bayesian factor model with smoothing splines for multivariate quantile spectral analysis that assumes the DRH prior. Log-transformations are further proposed on all τ_{pq}, $\lambda_{pqs,(re)}$, $\lambda_{pqs,(im)}$ and $\sigma^2_{E,p}$ to satisfy positivity constraints. Extensive simulation studies in Section 4.3 show that the proposed DRH-FM model outperforms competing methods.

4.2 Posterior Computation

4.2.1 Variational Bayes Posterior

In this section, we describe our approach to adapting variational Bayes techniques to obtain posterior inference of the proposed stationary multivariate quantile spectral model. Our posterior inference scheme is based on the stochastic gradient variational Bayes (SGVB) approach (Kingma and Welling, 2013). The proposed inference scheme for high-dimensional time series is scalable and computationally efficient by taking advantage of the power of modern computational resources, such as graphic processing units (GPUs).

To allow a more compact presentation, we define $\boldsymbol{\Xi}$ as the collection of all parameters in the DRH-FM model. The joint log density of frequency-domain observations $\boldsymbol{Y}$ and model parameters $\boldsymbol{\Xi}$ can be written as:

$$\log p(\boldsymbol{\Xi}, \boldsymbol{Y}) = \mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{\Xi}) + \log \pi(\boldsymbol{\Xi}), \tag{4.11}$$

where $\mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{\Xi})$ is defined in (4.10), and $\pi(\boldsymbol{\Xi})$ is the joint DRH prior defined in Section 4.1.2. As discussed in Section 3.2, we consider the Gaussian variational approximation with a factor covariance structure of Ong et al. (2018). Given a m_1-dimensional $\boldsymbol{\Xi}$, its surrogate posterior is defined as $q_{\boldsymbol{\Phi}}(\boldsymbol{\Xi}) = N(\boldsymbol{\Xi} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + I_{m_1}\boldsymbol{\sigma}^2$ and $\mathbf{B}$ is a $m_1 \times m_2$ full rank matrix with $m_2 << m_1$. The reparameterization trick for the factor covariance structure can then be conducted by calculating $\boldsymbol{\Xi} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$ with $\boldsymbol{\xi} \sim \mathbf{N}(\mathbf{0},\, I_{m_2})$ and $\boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0},\, I_{m_1})$, where I_{m_1}, I_{m_2} denote the m_1, m_2-dimensional identity matrices. $\odot$ denotes element-wise product. The log-transformation is further taken on $\boldsymbol{\sigma}^2$ to guarantee the positivity. Accordingly, $\boldsymbol{\Phi} = \{\boldsymbol{\mu}\,, \log \boldsymbol{\sigma}^2, \boldsymbol{B}\}$ are surrogate parameters, and the ELBO between $p(\boldsymbol{\Xi} \mid \boldsymbol{Y})$ and $q_{\boldsymbol{\Phi}}(\boldsymbol{Y})$ becomes:

$$\mathcal{L}\left(p, q_{\boldsymbol{\Phi}}\right) = \mathrm{E}_{\boldsymbol{\xi}, \boldsymbol{\epsilon}}\left[\log p(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \boldsymbol{Y}) - \log q_{\boldsymbol{\Phi}}(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon})\right]. \tag{4.12}$$

With a single sample $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\epsilon}^{(1)}$, the SGVB gradient estimator of (4.12) is given by:

$$\nabla_{\boldsymbol{\Phi}} \mathcal{L}\left(p, q_{\boldsymbol{\Phi}}\right) \simeq \nabla_{\boldsymbol{\Phi}} \left[\log p(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi}^{(1)} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}^{(1)}, \boldsymbol{Y}) - \log q_{\boldsymbol{\Phi}}(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi}^{(1)} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}^{(1)})\right]. \tag{4.13}$$

We set $m_2 = 4$ as suggested in Ong et al. (2018). We have empirically found that (4.13) leads to a fast convergence.

We summarized the variational Bayes posterior inference scheme for the pro-

Algorithm 3: Variational Bayes Inference for Quantile Spectral Factor Model

Input : $\boldsymbol{Y}$, the PQ-dimensional CT-based DFT observations.
Output: Variational posteriors $q_{\boldsymbol{\Phi}}(\boldsymbol{\Xi})$, and power spectrum estimates.
Params: $\boldsymbol{\Phi} = \{\boldsymbol{\mu}, \log \boldsymbol{\sigma}^2, \boldsymbol{B}\}$.

Phase 1–Point Approximation: **while** *iterating* **do**
  Take the gradients of $\boldsymbol{\Xi}$ from (4.11) ;
  Update $\boldsymbol{\Xi}$ via `Adam` optimizer;
  Adjust $\boldsymbol{\Xi}$ via Algorithm 4 ;
end
Fix $\boldsymbol{\mu}$ equal to the last updated $\boldsymbol{\Xi}$ in Phase 1.
Phase 2–Uncertainty Quant: **while** *iterating* **do**
  Draw $\boldsymbol{\xi}^{(1)} \sim N(\boldsymbol{0}, I_{m_1})$, $\boldsymbol{\epsilon}^{(1)} \sim N(\boldsymbol{0}, I_{m_2})$ and compute the SGVB gradients of $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ using (4.13);
  Update $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ via `Adam` optimizer;
end
Phase 3–Fine-tuning (Optional): **while** *iterating* **do**
  Draw $\boldsymbol{\xi}^{(1)} \sim N(\boldsymbol{0}, I_{m_1})$, $\boldsymbol{\epsilon}^{(1)} \sim N(\boldsymbol{0}, I_{m_2})$ and compute the SGVB gradients of $\boldsymbol{\mu}$, $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ using (4.13);
  Update $\boldsymbol{\mu}$, $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ via `Adam` optimizer;
end
Draw posterior samples of $\boldsymbol{\Xi}$ from the learnt $q_{\boldsymbol{\Phi}}(\boldsymbol{\Xi})$ and compute quantile power spectrum estimates via Algorithm 5.

posed Bayesian factor model in Algorithm 3. In particular, Phase 1 uses (4.11) as the objective function to obtain a point approximation, $\hat{\boldsymbol{\Xi}} = \arg\max_{\boldsymbol{\Xi}} \log p(\boldsymbol{\Xi}, \boldsymbol{Y})$. Phase 2 fixes $\boldsymbol{\mu} = \hat{\boldsymbol{\Xi}}$ and only updates $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ by SGVB to maximize (4.12). Finally, the previously updated $\boldsymbol{\mu}$, $\log \boldsymbol{\sigma}^2$ and $\boldsymbol{B}$ are fine-tuned at Phase 3 by SGVB to maximize (4.12). Results from our extensive simulation studies below show that our proposed TPVB achieves high numerical stability. We built the model and implemented the experiments in `Python 3.7` by means of the `Tensorflow-Probability` and `Tensorflow` (Abadi et al., 2015) packages. All gradients are computed via auto-differentiation modules in `Tensorflow`. At each iteration, we use `Adam` (Kingma and Ba, 2017) to update model parameters.

It is important to mention that (4.6) provides means to recover $\boldsymbol{f}(\omega)$ from

Algorithm 4: Parameter Ajustement during Posterior Learning

Input : Model parameters $\boldsymbol{\Xi}$.
Output: Adjusted $\boldsymbol{\Xi}$.

Pick $\boldsymbol{D}_k$, $k = 1, ..., K$, from $\boldsymbol{\Xi}$;
Calculate $\overline{\Re(\boldsymbol{D}_k)} = \frac{1}{K}\sum_{k=1}^{K}\Re(\boldsymbol{D}_k)$, and $\overline{\Im(\boldsymbol{D}_k)} = \frac{1}{K}\sum_{k=1}^{K}\Im(\boldsymbol{D}_k)$;
Calculate $\hat{\Sigma}_{\Re(\boldsymbol{D}_k)} = \frac{1}{K}\left(\Re(\boldsymbol{D}_k) - \overline{\Re(\boldsymbol{D}_k)}\right)\left(\Re(\boldsymbol{D}_k) - \overline{\Re(\boldsymbol{D}_k)}\right)^T$, and similarly calculate $\hat{\Sigma}_{\Im(\boldsymbol{D}_k)} = \frac{1}{K}\left(\Im(\boldsymbol{D}_k) - \overline{\Im(\boldsymbol{D}_k)}\right)\left(\Im(\boldsymbol{D}_k) - \overline{\Im(\boldsymbol{D}_k)}\right)^T$;
Update $\Re(\boldsymbol{D}_k) = \frac{1}{\sqrt{2}}\hat{\Sigma}^{-\frac{1}{2}}_{\Re(\boldsymbol{D}_k)}\left(\Re(\boldsymbol{D}_k) - \overline{\Re(\boldsymbol{D}_k)}\right)$ for $k = 1, ..., K$;
Update $\Im(\boldsymbol{D}_k) = \frac{1}{\sqrt{2}}\hat{\Sigma}^{-\frac{1}{2}}_{\Im(\boldsymbol{D}_k)}\left(\Im(\boldsymbol{D}_k) - \overline{\Im(\boldsymbol{D}_k)}\right)$ for $k = 1, ..., K$.

Algorithm 5: Posterior Computation of the Quantile Spectral Estimates

Input : Obtained variational posterior $q_{\boldsymbol{\Phi}}$ from Algorithm 3; Posterior sample size N_{post}.
Output: Power spectrum posterior samples $f^{(i)}(\omega)$ for $i = 1, ..., N_{post}$.

for *i from 1 to* N_{post} *(parallel)* **do**
  Sample $\Sigma_E^{(i)} = \text{diag}\left((\sigma^2_{E,1})^{(i)}, ..., (\sigma^2_{E,P})^{(i)}\right)$;
  Sample $\boldsymbol{\alpha}_{\cdot}^{(i)}$, $\boldsymbol{\beta}_{\cdot}^{(i)}$ from $q_{\boldsymbol{\Phi}}$ and calculate $\Lambda^{(i)}(\omega)$ via (4.9) ;
  Calculate $f^{(i)}(\omega) = \Lambda^{(i)}(\omega)\,\Lambda^{(i)}(\omega)^* + \Sigma_E^{(i)}$.
end

the estimated $\boldsymbol{\Lambda}(\omega)$ and $\boldsymbol{\Sigma}_E$. To let (4.6) hold, the covariance matrix of $\boldsymbol{D_k}$, $\text{Cov}\,[\boldsymbol{D}_k]$ for $k = 1, ..., K$, should equal to $\boldsymbol{I}_{Q_F}$, which implies that $\text{Cov}\,[\Re(\boldsymbol{D}_k)] = \text{Cov}\,[\Im(\boldsymbol{D}_k)] = \frac{1}{2}\boldsymbol{I}_{Q_F}$ in the context of the complex normality (Goodman, 1963). Algorithm 4 standardizes $\boldsymbol{D}_k$'s in each iteration to guarantee that the $\boldsymbol{D}_k$'s preserve the identity covariance matrix throughout the posterior inference procedure. In the end, the detailed steps on obtaining the posterior samples of the quantile power spectrum estimates after variational posterior learning are included in Algorithm 5. Empirical results in Section 4.3 illustrate the effectiveness of the proposed posterior inference scheme.

4.2.2 Selection of the Number of Factors

In practice, the number of factors Q_F for the factor model defined in Section 4.1.1 is expected to be relatively small to save computational time. Here we formulate the selection of Q_F as a model selection problem. Given a certain range of values for Q_F, we use the deviance information criterion (DIC; Spiegelhalter et al. (2014)) to select the superior model. In this context, DIC is defined as:

$$\text{DIC} = D(\overline{\boldsymbol{\Xi}}_{Q_F}) + 2p_{Q_F}$$

where $\boldsymbol{\Xi}_{\boldsymbol{Q_F}}$ specially denotes the model parameter $\boldsymbol{\Xi}$ in the proposed factor model having the number of factors to be Q_F. $\overline{\boldsymbol{\Xi}}_{Q_F}$ is the posterior expectation of $\boldsymbol{\Xi}_{\boldsymbol{Q_F}}$ that is obtained by Algorithm 3 in Section 4.2.1. $D(\boldsymbol{\Xi}) = -2\log \mathcal{L}(\boldsymbol{Y} \mid \boldsymbol{\Xi}) + C$ is the likelihood of $\boldsymbol{Y}$ defined in (4.10) adding a constant C that cancels out in all calculations that compare different models, and which therefore does not need to be known. $p_{Q_F} = \overline{D(\boldsymbol{\Xi}_{Q_F})} - D(\overline{\boldsymbol{\Xi}}_{Q_F})$ where $\overline{D(\boldsymbol{\Xi}_{Q_F})}$ can be approximated by averaging over the posterior samples of $D(\boldsymbol{\Xi}_{Q_F})$ that is calculated by the variational Bayes posterior samples of $\boldsymbol{\Xi}_{Q_F}$ that are obtained through Algorithm 3. During DIC selection, each model runs completely independently, thus the procedure can be efficiently conducted in parallel.

4.3 Simulation Studies

In this section, we show the accuracy and efficiency of the proposed method through simulations. During the variational posterior inference process for the simulation studies below, the learning rate is used as a hyper-parameter to control the rate at which the algorithm updates the parameter estimates. The learning rates in `Adam` optimizer were set to be 0.001, 0.05 and 0.05, respectively for each

phase, to obtain gradient descent updates for trainable parameters. The number of iterations was chosen by monitoring the values of the objective function, and stopping the training when a typical convergent pattern was reached. That led to approximately 8000, 500, and 500 iterations required for each of the 3 phases. The simulation experiments were executed on a x64-based PC with a 2.60-GHz Intel® Core™ i7-9750H CPU and a Nvidia GeForce GTX 1660 Ti GPU card.

4.3.1 Simulation Study 1

In this first simulation study, we show that our proposed approach outperforms currently available methods for univariate quantile spectral analysis. Dette et al. (2015) and Kley et al. (2016) utilized the smoothed rank-based Laplace periodogram (RLP) defined by either quantile regression (QR) or clipped time series (CT) to estimate the copula spectral density kernels. The `quantspec` package (Kley, 2014; Kley et al., 2020) in `R` has implemented both smoothed RLP approaches, serving as an ideal benchmark for evaluating our proposed method. We consider simulated data from three univariate processes studied in Kley et al. (2016) and Zhang (2019): a $QAR(1)$ process given by $x_t = 0.1\Phi^{-1}(\nu_t) + 1.9(\nu_t - 0.5)x_{t-1}$ (Koenker and Xiao, 2006) where the ν_ts are i.i.d. standard uniform random variables and $\Phi(\cdot)$ denotes the cdf of the standard normal distribution; an $AR(2)$ process given by $x_t = -0.36x_{t-2} + \epsilon_t$ (Li, 2012) where the ϵ_ts are of i.i.d. standard uniform random variables; and an $ARCH(1)$ process given by $x_t = (1/1.9 + 0.9x_{t-1}^2)^{0.5}\epsilon_t$ (Lee and Rao, 2012) where the ϵ_ts are independent standard uniform random variables. We simulate 100 time series of length $T = 1024$ from each of the 3 processes, and choose the quantile vector $\boldsymbol{\tau} = (0.1, 0.5, 0.9)'$ such that $\tau_1 = 0.1$, $\tau_2 = 0.5$, and $\tau_3 = 0.9$.

Considering that both QR- and CT-based smoothed RLP approaches pro-

Table 4.1: Simulation results for $QAR(1)$, $AR(2)$, and $ARCH(1)$ processes based on 100 repetitions: means (standard deviations) of the RMSE $\times 10^3$ of quantile spectral estimates obtained through the proposed factor model with DRH prior (FM-DRH), the baseline factor model with the original prior (FM-BASE) of Li et al. (2021), the Bayesian smoothing spline model (SPEC) of Zhang (2019), and the CT-based smoothed RLP (CT-SRLP).

Process	Model	f_{τ_1,τ_1}	f_{τ_1,τ_2}	f_{τ_1,τ_3}	f_{τ_2,τ_2}	f_{τ_2,τ_3}	f_{τ_3,τ_2}
QAR	FM-DRH	3.15	3.29	1.39	6.01	3.84	3.28
		(0.34)	(0.59)	(0.24)	(1.44)	(0.76)	(1.09)
	FM-BASE	3.53	4.45	2.69	7.47	4.11	3.36
		(0.38)	(0.63)	(0.31)	(1.50)	(0.63)	(0.95)
	SPEC	3.41	3.98	2.01	6.89	4.11	3.38
		(0.37)	(0.51)	(0.25)	(1.46)	(0.60)	(0.91)
	CT-SRLP	3.47	5.75	3.39	9.79	5.74	3.74
		(0.54)	(0.63)	(0.43)	(1.74)	(0.67)	(0.87)
AR	FM-DRH	4.13	3.66	2.57	6.94	2.79	3.36
		(0.82)	(0.79)	(0.76)	(2.92)	(0.73)	(0.52)
	FM-BASE	4.27	4.92	2.87	7.30	4.68	3.45
		(0.98)	(0.73)	(0.63)	(2.96)	(0.73)	(0.58)
	SPEC	4.22	4.88	2.85	7.14	3.83	3.40
		(0.90)	(0.75)	(0.61)	(2.85)	(0.74)	(0.55)
	CT-SRLP	3.31	5.54	3.54	9.98	5.84	3.39
		(0.59)	(0.72)	(0.45)	(2.26)	(0.76)	(0.64)
ARCH	FM-DRH	3.62	3.09	2.81	6.53	2.77	3.14
		(0.77)	(0.82)	(0.56)	(1.75)	(0.68)	(0.86)
	FM-BASE	4.05	3.87	2.82	7.58	3.01	3.31
		(0.97)	(0.70)	(0.52)	(1.59)	(0.65)	(0.80)
	SPEC	4.12	3.79	2.83	7.33	2.94	3.30
		(0.96)	(0.66)	(0.54)	(1.26)	(0.61)	(0.80)
	CT-SRLP	4.13	5.63	3.89	9.80	5.85	4.07
		(1.09)	(0.71)	(0.60)	(1.57)	(0.76)	(1.14)

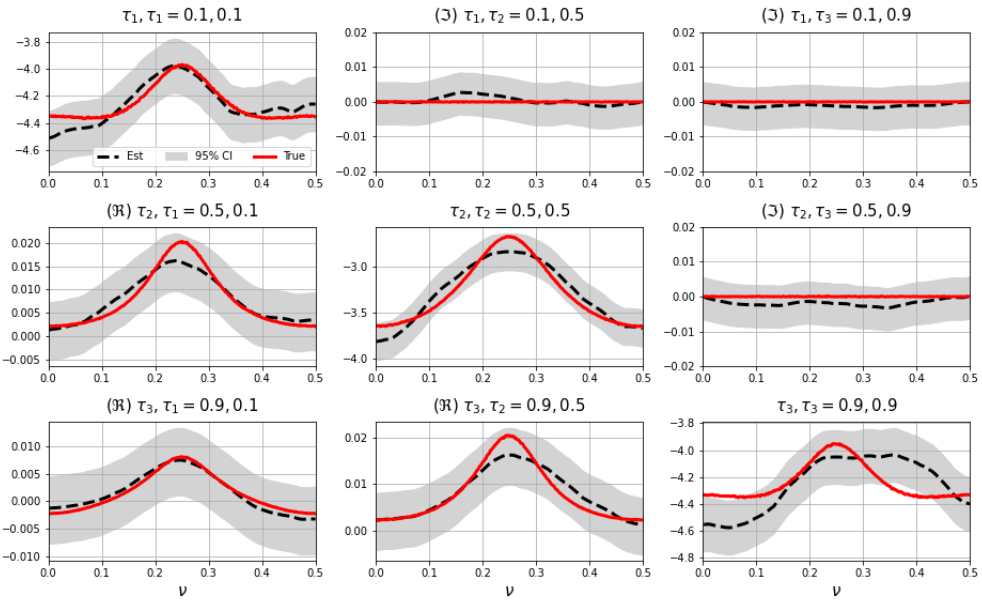

Figure 4.1: FM-DRH posterior inference of the quantile spectral estimates for the $AR(2)$ process. Diagonal plots correspond to $\log f_{\tau_i,\tau_j}$ for $i = j$. Lower and upper diagonal plots correspond to $\Re(f_{\tau_i,\tau_j})$ and $\Im(f_{\tau_i,\tau_j})$ for $i \neq j$. $i, j = 1, ..., 3$. Grey regions are 95% posterior intervals. Dashed lines are posterior mean estimates. Solid lines are true values.

vide quite similar estimates, but the QR-based smoothed RLP is more time consuming and our method also uses CT-based smoothed RLP approach, we discuss comparisons only with this latter approach. Here, we use CT-SRLP to denote the CT-based smoothed RLP approach. More specifically, we compare our proposed factor model with DRH priors (FM-DRH) with the following approaches: a baseline factor model (FM-BASE) with the prior structure of Li et al. (2021) that does not include the joint regularization over the spline coefficients of the real and imaginary parts of the loading factor components; the quantile spectral Bayesian smoothing spline approach (SPEC) of Zhang (2019) that models the modified complex Cholesky decomposition of the inverse of the quantile spectral matrix via smoothing splines and imposes normal conjugate priors on spline coefficients; and the CT-based smoothed RLP approach (CT-

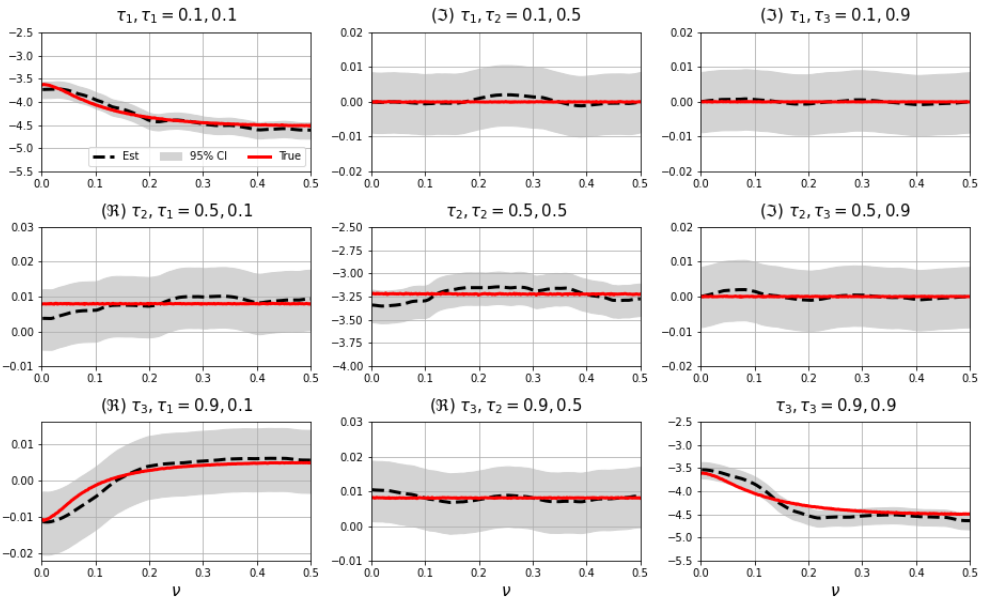

Figure 4.2: FM-DRH posterior inference of the quantile spectral estimates for the $ARCH(1)$ process. Diagonal plots correspond to $\log f_{\tau_i,\tau_j}$ for $i = j$. Lower and upper diagonal plots correspond to $\Re(f_{\tau_i,\tau_j})$ and $\Im(f_{\tau_i,\tau_j})$ for $i \neq j$. $i, j = 1, ..., 3$. Grey regions are 95% posterior intervals. Dashed lines are posterior mean estimates. Solid lines are true values.

SRLP) implemented in `quantspec`. We compute the CT-SRLP using a kernel of order 4, $W(u) := \frac{15}{32}\frac{1}{\pi}\left(7(u/\pi)^4 - 10(u/\pi)^2 + 3\right)\mathbb{I}_{[0,\pi]}(|u|)$, and common bandwidth $b_n = 0.4n^{-1/4}$, to guarantee the positive definitness of the estimated power spectrum and achieve better coverage probabilities for the confidence intervals constructed by this method (Zhang, 2019; Kley et al., 2016). In this case, $P = 1$ and we analyze a set of $Q = 3$ quantiles. Since these are relatively small, there is no need to conduct dimension reduction in this example. Accordingly, we set $Q_F = P \times Q = 3$ for FM-DRH and FM-BASE. Furthermore, we set the number of spline basis to $S = 20$ in FM-DRH, FM-BASE and SPEC for fair comparisons.

The squared root of mean squared error (RMSE) of the power spectra for each pair of (τ_i, τ_j) for $i, j = 1, ..., 3$ in each process is computed to measure the performance of the estimates for each individual power spectrum, which is defined

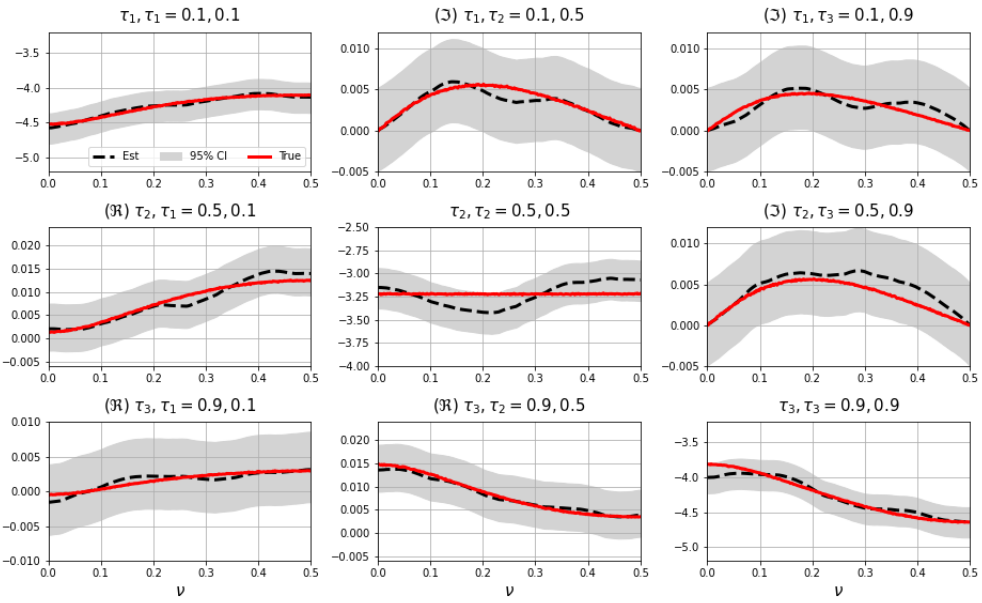

Figure 4.3: FM-DRH posterior inference of the quantile spectral estimates for the $QAR(1)$ process. Diagonal plots correspond to $\log f_{\tau_i,\tau_j}$ for $i = j$. Lower and upper diagonal plots correspond to $\Re(f_{\tau_i,\tau_j})$ and $\Im(f_{\tau_i,\tau_j})$ for $i \neq j$. $i, j = 1, ..., 3$. Grey regions are 95% posterior intervals. Dashed lines are posterior mean estimates. Solid lines are true values.

as

$$\text{RMSE}\left[\hat{f}^{p,p}_{\tau_i,\tau_j}\right] = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left|\hat{f}^{p,p}_{\tau_i,\tau_j}(\omega_l) - f^{p,p}_{\tau_i,\tau_j}(\omega_l)\right|^2},$$

where $\hat{f}^{p,p}_{\tau_i,\tau_j}(\omega)$ is the estimate of $f^{p,p}_{\tau_i,\tau_j}(\omega)$ for each process. The RMSE is evaluated at a prespecified equally spaced frequency grid $\omega_l = l/512$, $l = 1, ..., L$ with $L = 256$. A smaller the RMSEs indicates a better model accuracy. It can be seen from Table 4.1 that the proposed FM-DRH outperforms the competing methods for all the simulated processes and all the quantile combinations, indicating that the shrinkage induced by the DHR prior results in more accurate estimation. Figure 4.3-4.2 display the corresponding quantile spectral estimates for the three processes given by FM-DRH. Within each repetition, the run time of FM-DRH is less than 2 minutes for all three processes.

4.3.2 Simulation Study 2

In this study, we illustrate the performance of the proposed FM-DRH method in high dimensional analysis to evaluate its scalability. We still set the length-Q quantile vector to be $\boldsymbol{\tau} = (0.1, 0.5, 0.9)'$ ($Q = 3$), and consider a P-dimensional $QVAR(1)$ process defined by

$$\boldsymbol{X}_t = 0.1\Phi^{-1}(\boldsymbol{\nu}_t) + \Psi_1(\boldsymbol{\nu}_t)\boldsymbol{X}_{t-1}, \quad t = 1, \ldots, 1024, \tag{4.14}$$

where $\boldsymbol{\nu}_t = \{\nu_{t,p} : p = 1, ..., P\}$ is a sequence of i.i.d. P-dimensional standard independent uniform random variables with $\Phi^{-1}(\cdot)$ being the inverse cdf of the standard normal. $\Psi_1(\boldsymbol{\nu}_t) = \text{blockdiag}\left[\Psi_1^0(\nu_{t,1}, \nu_{t,2}), \Psi_1^0(\nu_{t,3}, \nu_{t,4}), ..., \Psi_1^0(\nu_{t,P-1}, \nu_{t,P})\right]$ where

$$\Psi_1^0(u_1, u_2) = \begin{bmatrix} 1.9(u_1 - 0.5) & u_1 - 0.5 \\ u_2 - 0.5 & 0.4(u_2 + 0.2) \end{bmatrix}.$$

Additionally, we also consider a P-dimensional $VAR(2)$ process from

$$\boldsymbol{X}_t = \Psi_1\boldsymbol{X}_{t-1} + \Psi_2\boldsymbol{X}_{t-2} + \boldsymbol{\epsilon_t}, \quad t = 1, \ldots, 1024, \tag{4.15}$$

where $\boldsymbol{\epsilon}_t \overset{iid}{\sim} N(\boldsymbol{0}, \Omega)$ with $\Omega = I_{P/2} \otimes \Omega^0$, and Ω^0 is a 2×2 matrix that has 1's on the diagonal and 0.5 off the diagonal. $\otimes$ denotes the Kronecker product. $I_{P/2}$ is a $P/2 \times P/2$ identity matrix. $\Psi_1 = I_{P/2} \otimes \Psi_1^0$, and $\Psi_2 = I_{P/2} \otimes \Psi_2^0$ with

$$\Psi_1^0 = \begin{bmatrix} 0.5 & 0 \\ 0 & -0.36 \end{bmatrix}, \quad \Psi_2^0 = \begin{bmatrix} 0.1 & -0.3 \\ 0 & 0.2 \end{bmatrix}.$$

We consider $P = 12, 24, 48$ and simulate 50 datasets for each process under each P. Note that since the quantile vector $\boldsymbol{\tau}$ has length 3, a P-dimensional process ends up having a corresponding $3P \times 3P$ cross quantile density function,

which can be large as P increases (e.g., $P = 48$ results in a 144×144 dimensional spectral matrix). In such situation, the RMSEs for all the individual quantile spectral components are too many. Instead, we utilize the squared root of mean integrated squared error (RMISE) to investigate the performance of an estimator of the high-dimensional cross quantile spectral matrix $\boldsymbol{f}(\omega)$, which is calculated by

$$\text{RMISE} = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left\|\hat{\boldsymbol{f}}\left(\omega_l\right) - \boldsymbol{f}\left(\omega_l\right)\right\|_F^2},$$

where $\|\cdot\|_F$ is the matrix Frobenius norm. Once again, $\{\omega_l : l = 1, ..., L\}$ is a pre-specified equally-spaced frequency grid with $\omega_l = l/512$, $l = 1, ..., L$ and $L = 256$. $\hat{\boldsymbol{f}}$ denotes the obtained posterior estimate of $\boldsymbol{f}$ from a given method. Furthermore, the performance of diagonal and off-diagonal estimators are investigated separately by

$$\text{RMISE}_d = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left\|\text{diag}\left[\hat{\boldsymbol{f}}\left(\omega_l\right)\right] - \text{diag}\left[\boldsymbol{f}\left(\omega_l\right)\right]\right\|},$$
$$\text{RMISE}_o = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left\|\hat{\boldsymbol{f}}\left(\omega_l\right) - \boldsymbol{f}\left(\omega_l\right)\right\|_{F*}^2},$$

where $\|A\|_{F*}^2 = \sum_{i=1}^{P}\sum_{j\neq i} A_{ij}^2$. The RMISE, RMISE_d and RMISE_o are jointly used to evaluate the performance of the proposed FM-DRH model in comparsion to the baseline approaches. Note that, the model of Zhang (2019) is orginally used for univariate analysis. By using the multivariate CT-based DFT vectors, $\boldsymbol{Y}_k$'s defined in Section 4.1, as model inputs, we adapt to multivariate analysis also. The number of spline basis is fixed to be 20 in FM-DRH, FM-BASE and SPEC for fair comparisons.

It can be seen from Table 4.2 and Table 4.3 that the proposed FM-DRH outperforms the other approaches as the former provides more accurate estimates

Table 4.2: Simulation results for the $QVAR(1)$ processe (4.14) based on 50 repetitions: means (standard deviations) of RMISE, RMISE_d and RMISE_o of quantile spectral estimates given by the proposed factor model with DRH prior (FM-DRH), the baseline factor model with its original prior (FM-BASE) of Li et al. (2021), the Bayesian smoothing spline model (SPEC) of Zhang (2019), and the CT-based smoothed RLP (CT-SRLP) of Baruník and Kley (2019). Q_F denotes the rank of the factor loading matrices in factor models. $\text{DIC}_{\%}$ denotes the percentage of times that DIC selects between $Q_F = 10$ and $Q_F = 15$ for FM-DRH.

			FM-DRH		FM-BASE			
Process	$P(\times Q)$		Q_F=10	Q_F=15	Q_F=10	Q_F=15	SPEC	CT-SRLP
	12(×3)	RMISE	0.15	0.13	0.21	0.19	0.17	0.22
			(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
		RMISE_d	0.10	0.08	0.14	0.13	0.12	0.15
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
		RMISE_o	0.11	0.11	0.15	0.14	0.12	0.16
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
		$\text{DIC}_{\%}$	8.00	92.0	−−	−−	−−	−−
	24(×3)	RMISE	0.32	0.28	0.43	0.41	0.38	0.43
			(0.06)	(0.06)	(0.07)	(0.06)	(0.05)	(0.05)
		RMISE_d	0.17	0.14	0.28	0.27	0.23	0.27
			(0.03)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)
		RMISE_o	0.27	0.25	0.33	0.31	0.30	0.33
			(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)
		$\text{DIC}_{\%}$	2.00	98.0	−−	−−	−−	−−
QVAR	48(×3)	RMISE	0.66	0.61	0.91	0.87	−−	0.89
			(0.09)	(0.08)	(0.13)	(0.12)	(−−)	(0.09)
		RMISE_d	0.32	0.28	0.52	0.50	−−	0.53
			(0.05)	(0.05)	(0.07)	(0.06)	(−−)	(0.07)
		RMISE_o	0.58	0.53	0.75	0.72	−−	0.72
			(0.07)	(0.07)	(0.11)	(0.11)	(−−)	(0.08)
		$\text{DIC}_{\%}$	0.00	100	−−	−−	−−	−−

Table 4.3: Simulation results for the $VAR(2)$ processe (4.15) based on 50 repetitions: means (standard deviations) of RMISE, RMISE_d and RMISE_o of quantile spectral estimates given by the proposed factor model with DRH prior (FM-DRH), the baseline factor model with its original prior (FM-BASE) of Li et al. (2021), the Bayesian smoothing spline model (SPEC) of Zhang (2019), and the CT-based smoothed RLP (CT-SRLP) of Baruník and Kley (2019). Q_F denotes the rank of the factor loading matrices in factor models. $\text{DIC}_{\%}$ denotes the percentage of times that DIC selects between $Q_F = 10$ and $Q_F = 15$ for FM-DRH.

			FM-DRH		FM-BASE			
Process	$P(\times Q)$		Q_F=10	Q_F=15	Q_F=10	Q_F=15	SPEC	CT-SRLP
	12(×3)	RMISE	0.22	0.21	0.28	0.26	0.25	0.34
			(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
		RMISE_d	0.14	0.12	0.18	0.16	0.14	0.20
			(0.03)	(0.02)	(0.04)	(0.04)	(0.03)	(0.03)
		RMISE_o	0.16	0.15	0.22	0.21	0.21	0.28
			(0.02)	(0.02)	(0.04)	(0.04)	(0.03)	(0.03)
		$\text{DIC}_{\%}$	6.00	94.0	−−	−−	−−	−−
	24(×3)	RMISE	0.42	0.39	0.54	0.51	0.50	0.59
			(0.09)	(0.08)	(0.13)	(0.12)	(0.11)	(0.12)
		RMISE_d	0.21	0.20	0.31	0.28	0.27	0.32
			(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.08)
		RMISE_o	0.37	0.34	0.46	0.44	0.42	0.50
			(0.09)	(0.07)	(0.11)	(0.11)	(0.07)	(0.12)
		$\text{DIC}_{\%}$	4.00	96.0	−−	−−	−−	−−
VAR	48(×3)	RMISE	0.83	0.76	1,01	0.92	−−	0.93
			(0.12)	(0.15)	(0.18)	(0.14)	(−−)	(0.18)
		RMISE_d	0.45	0.40	0.69	0.61	−−	0.61
			(0.08)	(0.08)	(0.08)	(0.08)	(−−)	(0.06)
		RMISE_o	0.70	0.65	0.74	0.68	−−	0.70
			(0.09)	(0.08)	(0.09)	(0.08)	(−−)	(0.08)
		$\text{DIC}_{\%}$	0.00	100	−−	−−	−−	−−

that lead to smaller RMISEs in all the situations, indicating that the FM-DRH shrinkage prior is still effective in high-dimensional modeling. Note that, as P is large, the benefit of scalability gained by using the factor model becomes quite essential. For instance, when $P = 48$, the Bayesian smoothing splines approach of Zhang (2019), which represents the power spectrum using the complex modified Cholesky decomposition, is not computationally feasible as our computer does not have enough memory to store all the model parameters. In comparison, the factor model is still scalable due to the parsimonious low-rank representation. Note that, for illustrative purposes, here we only compare $Q_F = 10$ with $Q_F = 15$ to clarify the effect of Q_F on model estimates. In practice, a wide range of Q_F could be considered for DIC selection (see Section 4.4). Figure 4.4-4.6 displays some of the true $QVAR(1)$ cross quantile spectra and their corresponding estimates along with 95% posterior intervals obtained by FM-DRH. The run time of FM-DRH within each repetition is less than 7 minutes when $P = 12$, 10 minutes when $P = 24$, and 25 minutes when $P = 48$, for both aforementioned multivariate processes.

4.4 Data Analyses

4.4.1 San Lorenzo River Flow Data Analysis

Analyzing the statistical properties of water flow levels and their relationship with environmental variables is important to better understand droughts and floods (Lombard, 2016; NOAA National Centers for Environmental Information, 2021). In this section we consider quantile spectral analysis of anomalies of monthly precipitation (in inches) in Santa Cruz CA, and anomalies of monthly log water flow (with water flow measured in cubic feet per second) at the Big Tree (BT) water gauge of the San Lorenzo river in Santa Cruz, CA from Jan-

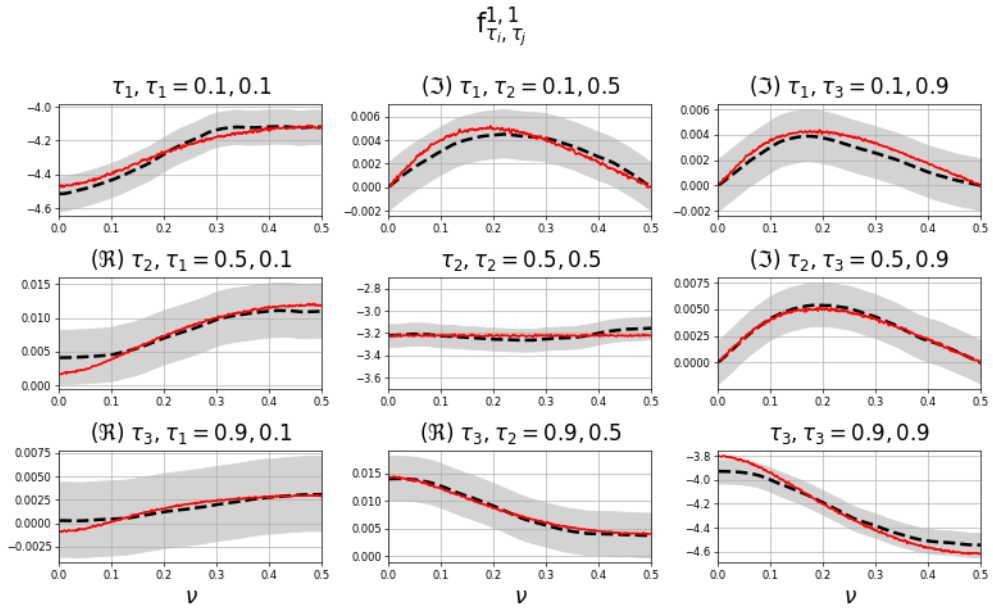

Figure 4.4: FM-DRH posterior inference of the quantile spectral estimates of $f^{1,1}_{\tau_i,\tau_j}$, $i, j = 1, ..., 3$, for process (4.14). Diagonal plots correspond to $\log f^{1,1}_{\tau_i,\tau_j}$ for $i = j$. Lower and upper diagonal plots correspond to $\Re(f^{1,1}_{\tau_i,\tau_j})$ and $\Im(f^{1,1}_{\tau_i,\tau_j})$ for $i \neq j$. Grey regions are 95% posterior intervals. Dashed lines are posterior mean estimates. Solid lines are true values.

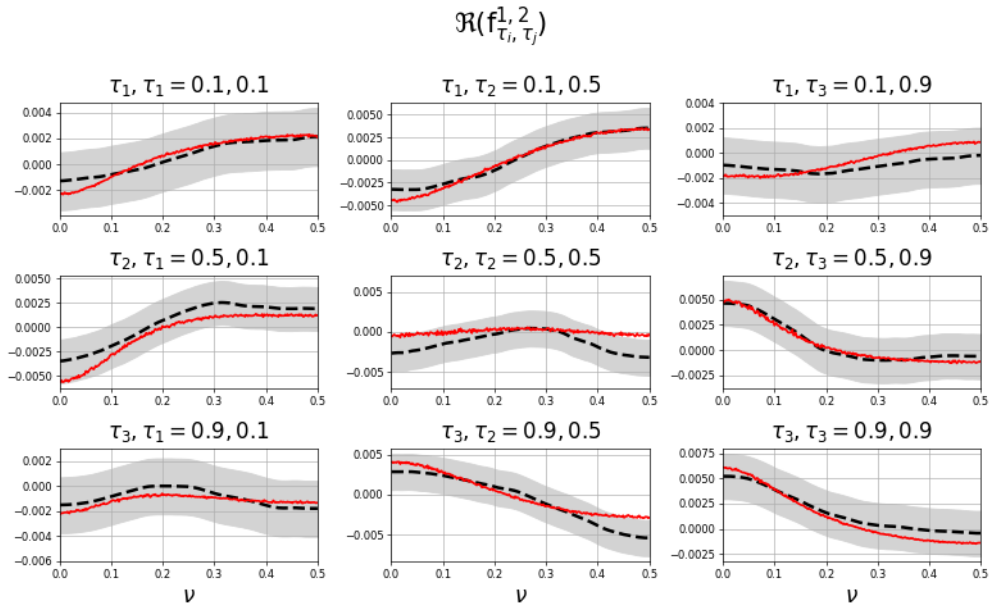

Figure 4.5: FM-DRH posterior inference of the quantile spectral estimates of $\Re(f^{1,2}_{\tau_i,\tau_j})$, $i, j = 1, ..., 3$, for process (4.14). Grey regions are 95% posterior intervals. Dashed lines are posterior mean estimates. Solid lines are true values.

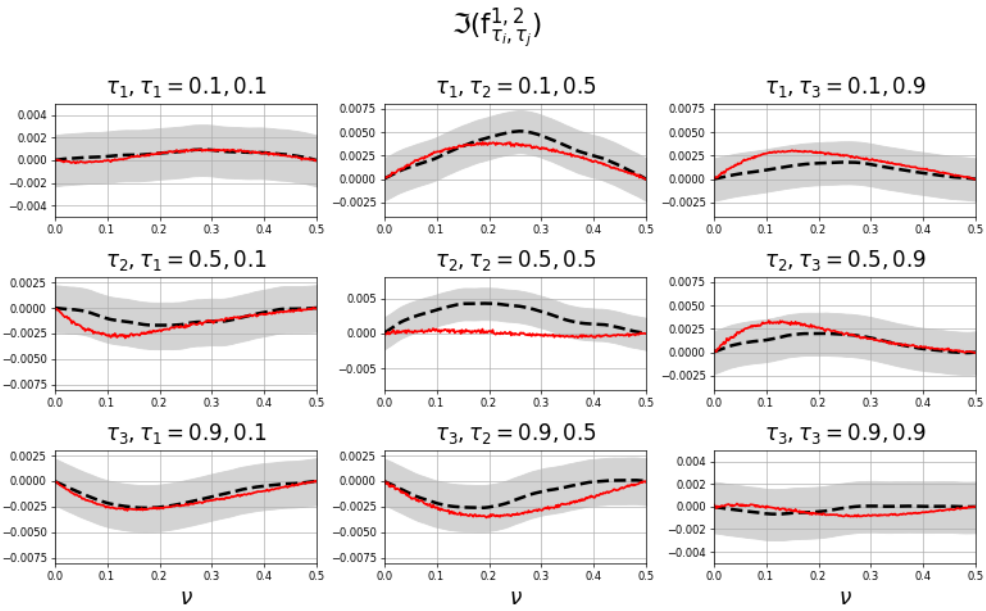

Figure 4.6: FM-DRH posterior inference of the quantile spectral estimates of $\Im(f^{1,2}_{\tau_i,\tau_j})$, $i, j = 1, ..., 3$, for process (4.14). Grey regions are 95% posterior intervals. Dashed lines are posterior mean estimates. Solid lines are true values.

uary 1937 through December (2020) (U.S. Geological Survey, 2016), resulting in 1008 observations for each time series. An analysis of of the observed log water flow and precipitation time series (not their anomalies) using dynamic quantile models was presented in Barata (2021). More specifically, Barata (2021) infer the time-varying 0.15 quantiles both monthly water flow and precipitation, and the 0.85 quantile for temperature. The main goal of this time-domain dynamic quantile approach was to infer the instantaneous quantiles for each time series while incoporating potential trend and seasonal components and accounting for the correlation between the 3 time series. Here we consider a frequency-domain quantile spectral analysis that aims to study the relationships between the log water flow and the precipitation anomalies at different quantile levels. We consider the anomalies data rather than the original observed time series in order to remove the annual seasonal behavior in both time series.

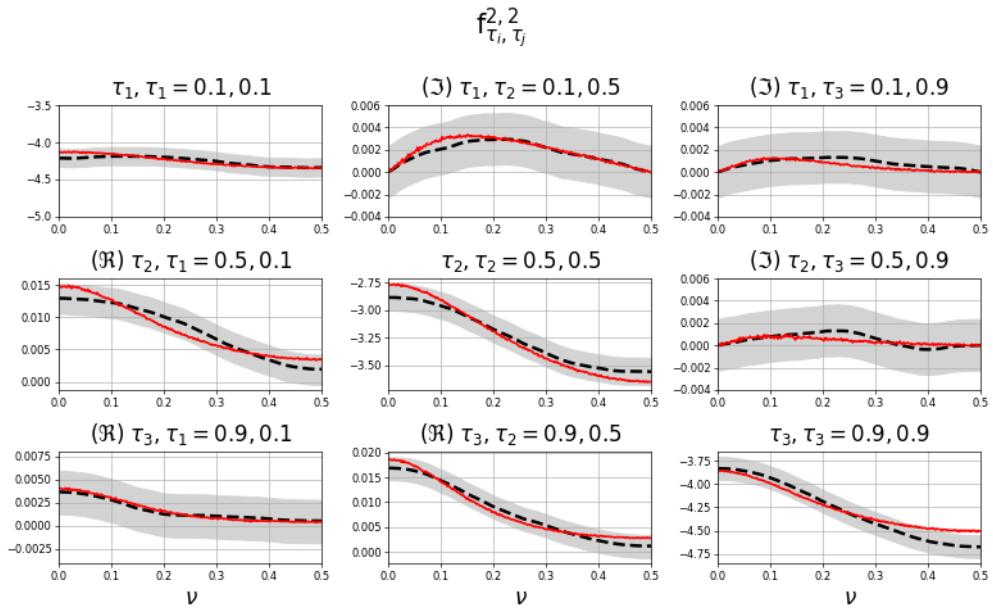

Figure 4.7: FM-DRH posterior inference of the quantile spectral estimates of $f^{2,2}_{\tau_i,\tau_j}$, $i, j = 1, ..., 3$, for process (4.14). Diagonal plots correspond to $\log f^{2,2}_{\tau_i,\tau_j}$ for $i = j$. Lower and upper diagonal plots correspond to $\Re(f^{2,2}_{\tau_i,\tau_j})$ and $\Im(f^{2,2}_{\tau_i,\tau_j})$ for $i \neq j$. Grey regions are 95% posterior intervals. Dashed lines are posterior mean estimates. Solid lines are true values.

We jointly model the 0.15, 0.5 and 0.85 quantiles of the log BT flow anomalies and the precipitation anomalies series using the proposed method. We use the same number of spline basis, $S = 20$, as in Section 4.3. The number of factors is set to be 6 based on DIC selection (see Figure 4.12). Figure 4.8 shows the log spectral density estimates at the 3 quantiles. We see that all three estimated quantile spectral densities of log BT flow anomalies display a similar pattern with higher power towards the low frequencies and lower power towards the high frequencies. The power is also higher at the median spectral density. Instead, the inferred quantile spectral densities at the 0.5 quantile for the precipitation anomalies displays a very different pattern than those inferred for the 0.15 and 0.85 quantiles. The 0.15 and 0.85 quantiles show higher power around frequency 1/12 (0.08) consistent with a periodic yearly pattern and decreased power towards

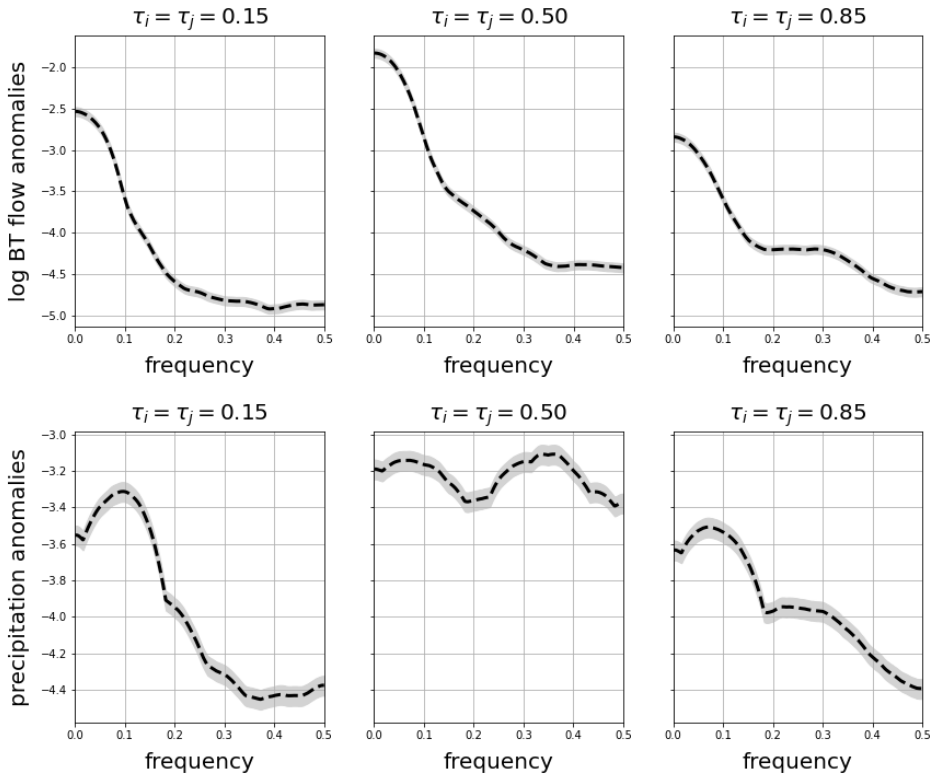

Figure 4.8: Estimated log spectral densities of the log BTflow anomalies (top) and precipitation anomalies (bottom) under 0.15, 0.5, and 0.85 quantiles. Dashed lines are posterior mean estimates. Grey regions are 95% posterior intervals.

higher frequency, while this behavior is not present at the median quantile. This is indicative that considering anomalies in order to remove the annual cycle from the precipitation series works only at the 0.5 quantile but does not fully remove this seasonal pattern at higher or lower quantiles.

Figure 4.9 displays the cross-quantile squared coherence estimates between the log BT flow anomalies and the precipitation anomalies at quantile pairs $(0.15, 0.15)$, $(0.15, 0.85)$, $(0.85, 0.15)$ and $(0.85, 0.85)$, respectively for each time series. We see that the higher coherence appears at the low frequencies for all the quantile combinations considered. The coherences for the $(0.15, 0.15)$ and $(0.85, 0.85)$ quantile pairs are larger than those for the $(0.15, 0.85)$ and $(0.85, 0.15)$ quantile pairs, showing, as expected (Johannis et al., 2016), that below-normal BT flow has a relatively high coherence to below-normal precipitation (signals of droughts) and above-normal BT flow and above-normal precipitation (signals of floods) also show a relatively high coherence. Moreover, when frequency is greater than 0.2, the coherence for at the $(0.85, 0.85)$ quantile pair is larger than the coherence at the $(0.15, 0.15)$ quantile pair, indicating that the mid-range and short-range memories for the dependency between BT flow and precipitation at floods levels are stronger than those at droughts levels. From these findings, we see how quantile spectral analysis can be helpful in discovering the relationship between the water flow and precipitation and how this relationship changes at different quantiles.

4.4.2 Stock Data Analysis

In this section, we apply the proposed method to conduct quantile spectral analysis on S&P 500 stock market index data Nugent (2017). The complete dataset in (Nugent, 2017) consists of daily stock prices from 2013 to 2018 for all

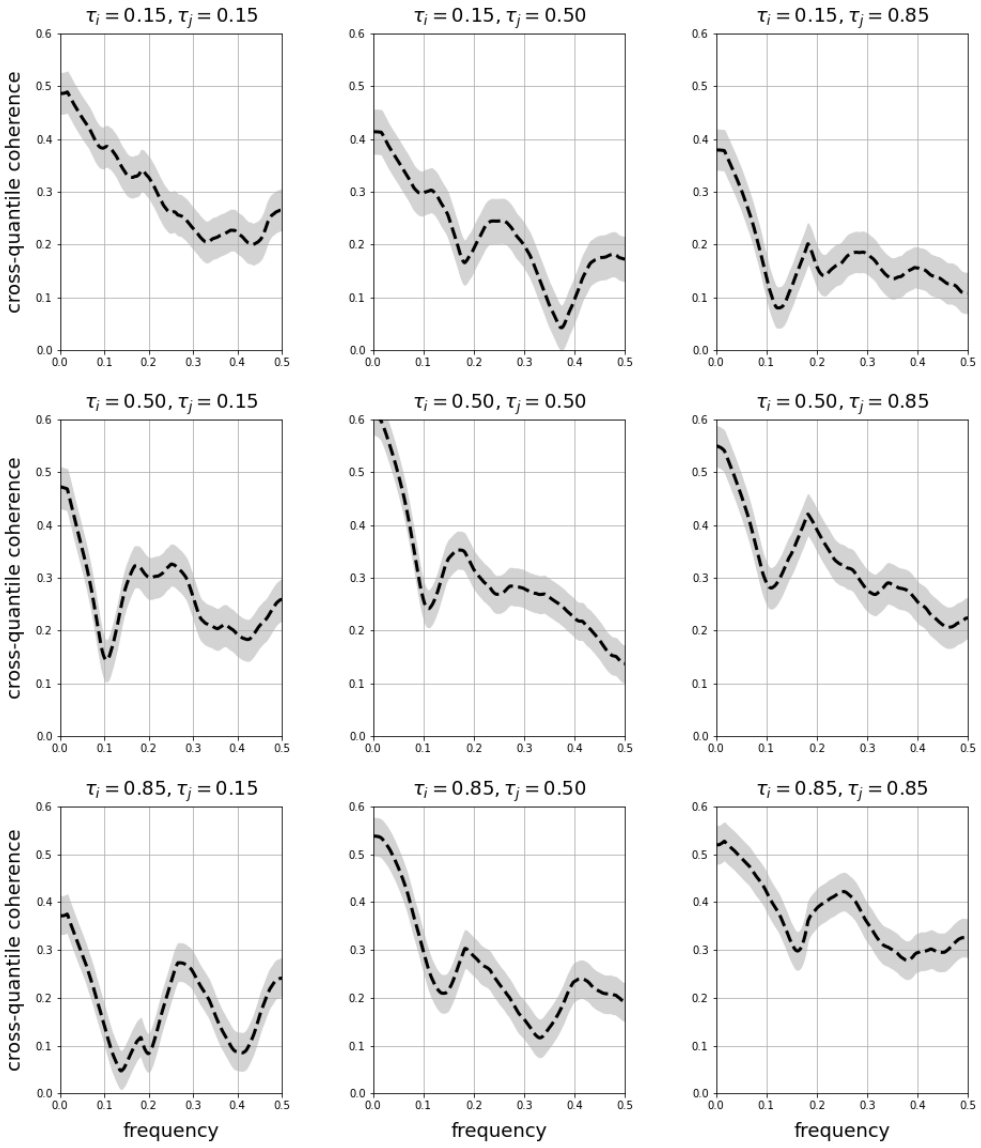

Figure 4.9: Estimated cross-quantile coherences between log BT flow anomalies at τ_i quantile and precipitation anomalies at τ_j quantile for $i, j = 1, 2, 3$. Dashed lines are posterior mean estimates. Grey regions are 95% posterior intervals.

companies found on the 2018 S&P 500 index. The 11 S&P sectors, or Global Industry Classification Standard (GICS), organize companies based on their primary business activities. In this study, we only analyze data from companies belonging to two sectors, the financials and utilities sectors in order to explore the spectral quantile behavior of the indexes from any two companies within one of the two sectors and any two companies in different sectors. It is expected that two companies within the same sector will have more similar spectral behavior than companies from two different sectors, however, it is not clear that this is the case for all the companies and also it is not clear that the spectral behavior remains consistent across different quantiles.

Here we select companies according to GICS sub-industries. More specifically, all 12 companies (LNT, AEP, D, DUK, ED, EIX, ETR, FE, PPL, PEG, SO, WEC) from the electric utilities industry are chosen for the utilities sector. For the finance sector, 6 companies (BAC, C, CMA, JPM, USB, WFC) from the diversified banks, 3 companies (AXP, COF, DFS) from the consumer finance, and 4 companies (SCHW, GS, MS, RJF) from the investment banking & brokerage are included. That results in totally 25 companies in the analysis. Note that we choose these companies for illustrative purposes, one can also substitute other companies in the analysis. We follow a common transformation and take the first differences of the natural logarithm of the original prices (Chen, 2012; López-Oriona and Vilar, 2021b) to obtain the so-called change in price, which represents the volatility in each of the series.

This type of data shows heavy tails and skewed density shapes (López-Oriona and Vilar, 2021b), which makes quantile spectral analysis better suited to analyze these data than standard spectral analysis. We consider quantiles 0.1, 0.5 and 0.9 for all series which is a common choice of the quantiles (López-Oriona and Vilar,

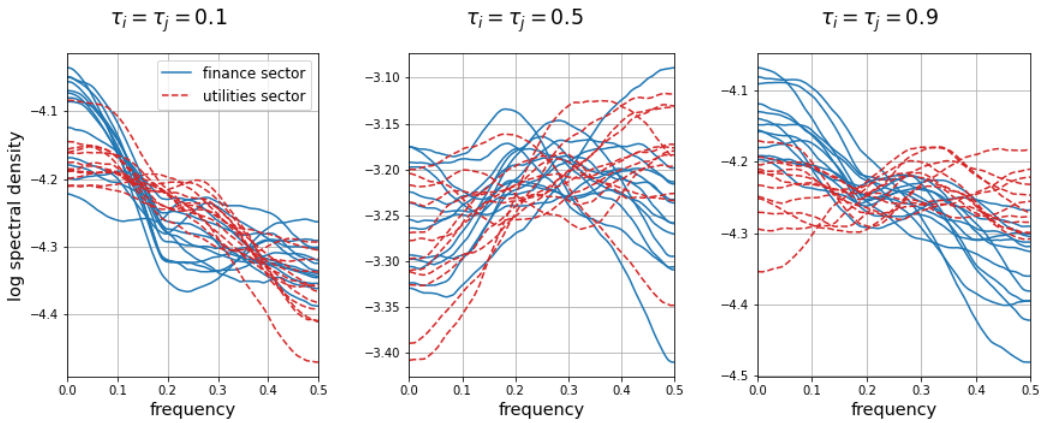

Figure 4.10: Estimated log spectral densities under quantile 0.1 (left), 0.5 (middle) and 0.9 (right). Blue solid lines denote companies from the finance sector. Red dashed lines correspond to the utilities sector.

2021b). We use the same number of spline basis, $S = 20$, as in Section 4.4.1. The number of factors is chosen to be 15 according to the proposed model selection procedure (see Figure 4.12). Here we show some of the inference results of the analysis obtained from the proposed method. Figures 4.10 present the estimated log spectral densities under each quantile. It can be seen that the spectral density estimates vary by quantile. For instance, under 0.1 and 0.5 quantiles, the log spectral densities estimates from two sectors are hard to distinguish. However, under 0.9 quantile, the different quantile power spectrum of the financials sector companies in comparison to the companies from the utilities sector are more distinguishable. We see that the financials sector group has the estimated log quantile spectral densities decreasing over frequency. In comparison, the estimated log quantile spectral densities from the utilities sector group are much flatter. Figure 4.11 compares the estimated cross-series cross-quantile coherences between two companies that come from the same sector with those from disparate sectors. It is noticeable that the coherences in most quantile paires, except (0.1, 0.9) and (0.9, 0.1) pairs, are relatively strong among companies that within the

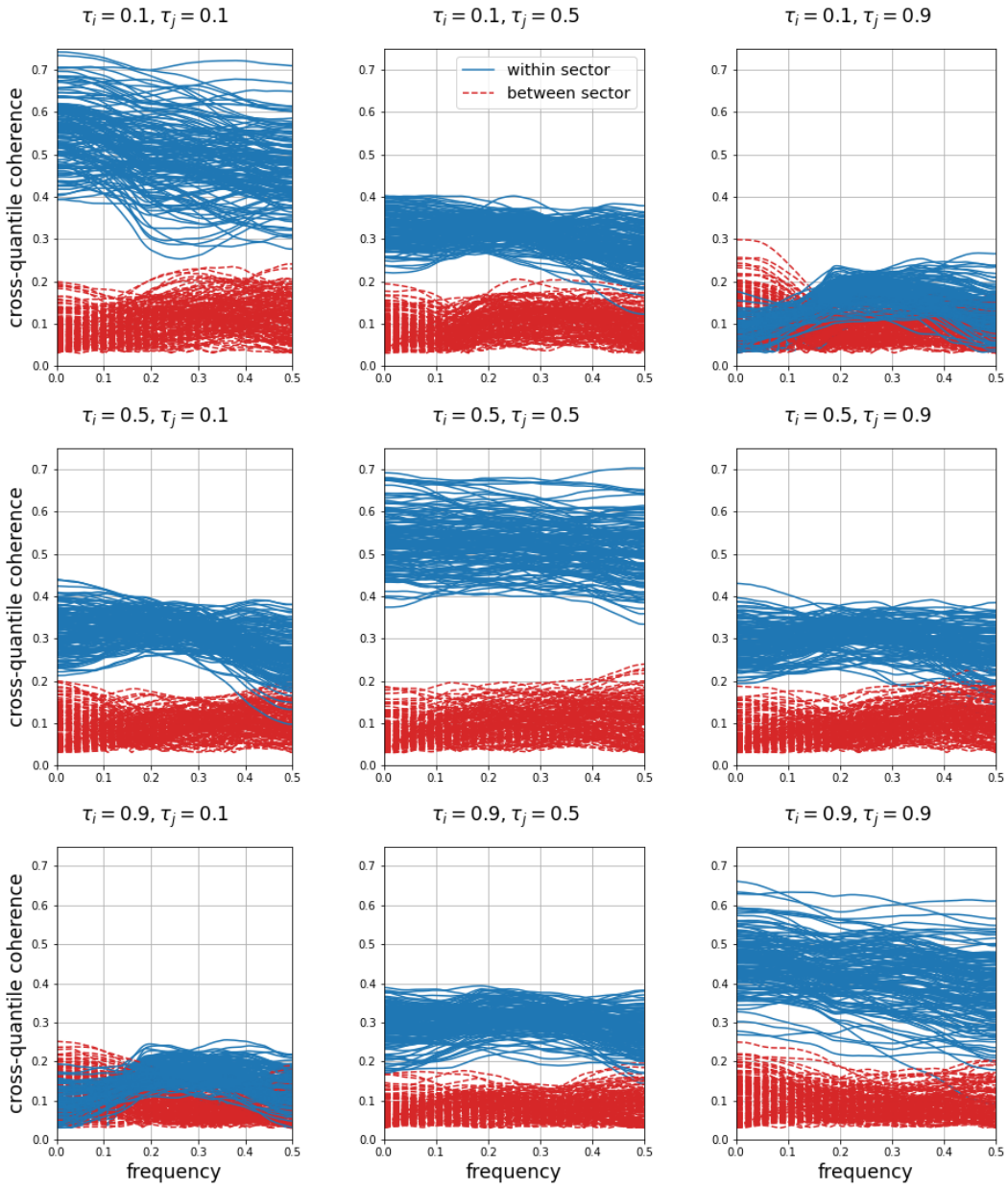

Figure 4.11: Estimated cross-quantile coherences for selective (τ_i, τ_j) quantile pairs. Blue solid lines are coherences between two companies within the same sector. Red dashed lines are coherences between two companies pertaining to different sectors.

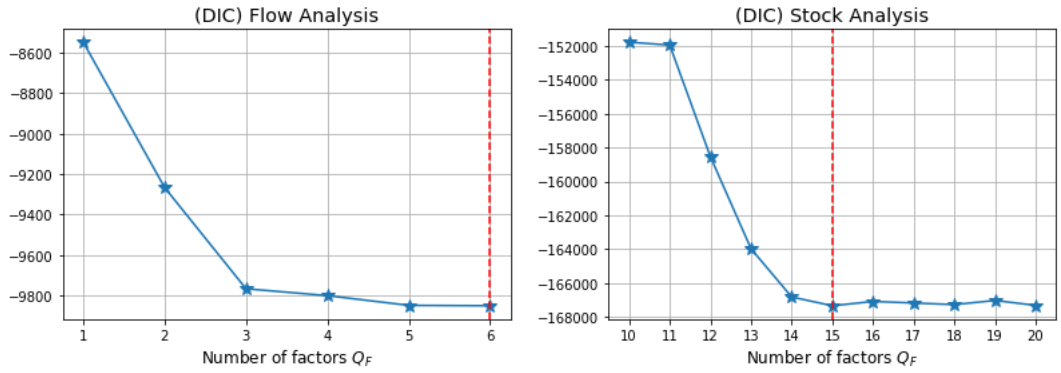

Figure 4.12: : Model selection for the number of factors. Left: San Lorenzo river flow data. Right: S&P 500 stock data

same sector, while being obviously weaker between any two companies pertaining to different sectors. Our inferences reflect the fact that financials and utilities sectors present strongly different economic behaviours, thus having quite different patterns of their quantile power spectrum, while companies from the same sector preserve relatively similar quantile spectral information.

There are several real-world applications based on the quantile spectral analysis. López-Oriona and Vilar (2021b) developed a dissimilarity metric based on the estimated quantile spectral densities to cluster multivariate time series. López-Oriona and Vilar (2021a) described multivariate time series as quantile spectral density estimates and developed methodologies for outlier detection of multivariate time series. The quantile power spectrum given by the proposed method include a rich set of frequency-domain information for multivariate time series, which is quite insightful and can be utilized as useful features to potentially improve the performances of the downstream applications.

4.5 Conclusions and Future Work

The proposed novel modeling and inference framework provides accurate and computationally fast approximate Bayesian quantile spectral inference for multivariate stationary time series. We showed that the discounted regularized horseshoe prior proposed on the low-rank factor model leads to accurate inference of quantile spectral densities and squared coherences in several simulation settings involving multivariate time series, such that the low-rank decomposition of the full spectral matrix not only greatly reduces model complexity but also preserves desirable accuracy. The proposed inference scheme, that utilizes stochastic gradient variational Bayes, is highly scalable and efficient, providing solid computational support for large-dimensional quantile spectral analysis.

In the future, more options on the prior can be explored. Possible extensions include considering different options of spline basis (e.g., B-spline, P-spline) to satisfy particular needs. For instance, a spectra lying within certain frequency bands can be handled by B-spline basis with selective knots. Also, other shrinkage prior could also be compared, such as the normal-gamma prior by Griffin and Brown (2010) and its modifications by Huber and Feldkircher (2019), as well as exploring the benefits of introducing additional sparsification procedures similar to those of Huber et al. (2021), where a parameter sparsification step is proposed after obtaining the posterior estimates using global-local shrinkage priors to further reduce storage load and improve performance.

More relevant extensions in terms of increasing the applicability of the proposed modeling framework and related inferential procedures include considering multivariate quantile spectral analysis for non-stationary time series, as well as hierarchical model formulations that allow us to jointly analyze multi-trial data. For instance, Birr et al. (2017) has developed theories and methodologies of quantile

spectral analysis for locally stationary time series. However, their work is restricted to the univariate time series, which could be extended to the multivariate time series in the future.

Chapter 5

Discussion

We have developed several new Bayesian models and methods for spectral analysis of multivariate time series: a spectral approach for multivariate stationary time series in Chapter 2 and extensions to multivariate nonstationary time series in Chapter 3, and the quantile-based spectral analysis for stationary multivariate time series in Chapter 4. We now summarize the key achievements and the future steps of our work.

Our model for spectral analysis of multivariate stationary time series developed in Chapter 2 assumes a DRH prior on the spline coefficients that define the components of the Cholesky decomposition of the inverse spectral density matrix, allowing us to obtain smooth estimates of the multivariate spectral matrix while avoiding overfitting. A TPVB algorithm is proposed for fast and accurate approximate posterior inference. Then, in Chapter 3 we expanded this approach leading to non-stationary multivariate spectral analysis by means of a locally stationary basis representation of the time-varying spectra. We developed and implemented two models, a model with a slow-varying prior (SV-Prior) that provides smooth estimates of the spectral density matrix over time, and a model with an abruptly changing prior (AC-Prior) to handle situations in which there are change points

in the spectral behavior over time. The SV-Prior provides estimates of the power spectrum that evolve smoothly over time. A model selection procedure based on DIC and a customized stochastic gradient variational Bayes algorithm are incorporated to obtain posterior inference efficiently. Finally, we developed an approach to jointly estimate the matrix of quantile cross-spectral density kernels under multiple quantiles. In this model we used a low-rank factorization of the spectral density matrix for dimension reduction and assumed a DRH prior on the model parameters. Once again, we developed and implemented a variational Bayes approach for approximate posterior inference. We illustrated the accuracy and scalability of the proposed methods via extensive simulation studies. We also showed the applicability of these approaches in the analysis of real datasets. More specifically, analyzed EEG datasets, IEM wind profiles, San Lorenzo river flow measurements, and an S&P 500 stock market index dataset.

Several future expansions on the work presented in this dissertation could be considered. For instance, a study comparing the properties of the new DRH prior with other existing shrinkage priors, such as the modifications of the normal-gamma prior by Huber and Feldkircher (2019) and the sparsification procedure by Huber et al. (2021), would be of interest. The model setting can also be further extended by providing more options of spline basis, such as P-splines of (Lang and Brezger, 2004) and B-splines (De Boor, 1972) other than the cosine basis. Another possible place for further improvement is to develop a partition adjustment procedure for choosing the partitions in the non-stationary models. This is particularly important in the locally stationary model since the equally-spaced partition only allows the model to select change point estimates among the pre-specified partition points. For example, by means of a model selection criterion, time points near the current change point estimate could be evaluated and used

to update the estimate adaptively. Future studies are needed to formulate this idea and conduct the adjustment of the change point estimate effectively. Another relevant extension in terms of increasing the applicability of the proposed modeling framework is the development of the hierarchical model formulations that allow for joint analysis of data with hierarchical structure, such as multi-trial EEG data. In quantile spectral analysis, one avenue that remains pending is the development of a method for nonstationary multivariate time series. A possible way is to extend the definition of the univariate locally strictly stationary processes with copula spectral density kernels in Birr et al. (2017) to multivariate time series. Finally, it would be interesting to further explore the impact of our work in other applications. The spectral estimates can be used as features for machine learning tasks to recognize useful patterns among subjects that are represented by time series. For instance, López-Oriona and Vilar (2021b) developed a dissimilarity metric based on the estimated quantile spectral densities to cluster multivariate time series. López-Oriona and Vilar (2021a) described multivariate time series as quantile spectral density estimates and developed methodologies for outlier detection of multivariate time series. We can follow their experiments to highlight improvements on the performances of the clustering and outlier detection algorithms by aid of our estimates.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. URL: `https://www.tensorflow.org/`. software available from tensorflow.org.

Adak, S., 1998. Time-dependent spectral analysis of nonstationary time series. Journal of the American Statistical Association 93, 1488–1501.

Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle, in: Selected papers of hirotugu akaike. Springer, pp. 199–213.

Amari, S.I., 1998. Natural gradient works efficiently in learning. Neural Computation 10, 251–276.

Andrieu, C., Thoms, J., 2008. A tutorial on adaptive mcmc. Statistics and Computing 18, 343–373.

Baccalá, L.A., Sameshima, K., 2001. Partial directed coherence: a new concept in neural structure determination. Biological Cybernetics 84, 463–474.

Barata, R.A., 2021. Flexible Dynamic Quantile Linear Models. Ph.D. thesis. University of California, Santa Cruz.

Baruník, J., Kley, T., 2019. Quantile coherency: A general measure for dependence between cyclical economic variables. The Econometrics Journal 22, 131–152.

Birr, S., Volgushev, S., Kley, T., Dette, H., Hallin, M., 2017. Quantile spectral analysis for locally stationary time series. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79, 1619–1643.

Bishop, C.M., 2006. Pattern recognition and machine learning. Springer. chapter 10. pp. 461 – 522.

Bitto, A., Frühwirth-Schnatter, S., 2019. Achieving shrinkage in a time-varying parameter model framework. Journal of Econometrics 210, 75–97.

Blei, D.M., Kucukelbir, A., McAuliffe, J.D., 2017. Variational inference: A review for statisticians. Journal of the American Statistical Association 112, 859–877.

Campbell, J.Y., Grossman, S.J., Wang, J., 1993. Trading volume and serial correlation in stock returns. The Quarterly Journal of Economics 108, 905–939.

Chen, S.S., 2012. Revisiting the empirical linkages between stock returns and trading volume. Journal of Banking & Finance 36, 1781–1788.

Chipman, H.A., Kolaczyk, E.D., McCulloch, R.E., 1997. Adaptive Bayesian wavelet shrinkage. Journal of the American Statistical Association 92, 1413–1421.

Choudhuri, N., Ghosal, S., Roy, A., 2004. Bayesian estimation of the spectral density of a time series. Journal of the American Statistical Association 99, 1050–1059.

Dahlhaus, R., 2000. A likelihood approximation for locally stationary processes. The Annals of Statistics 28, 1762–1794.

Dai, M., Guo, W., 2004. Multivariate spectral analysis using Cholesky decomposition. Biometrika 91, 629–643.

Davis, R.A., Lee, T.C.M., Rodriguez-Yam, G.A., 2006. Structural break estimation for nonstationary time series models. Journal of the American Statistical Association 101, 223–239.

De Boor, C., 1972. On calculating with b-splines. Journal of Approximation theory 6, 50–62.

Dette, H., Hallin, M., Kley, T., Volgushev, S., 2015. Of copulas, quantiles, ranks and spectra: An L_1-approach to spectral analysis. Bernoulli 21, 781–831.

Dickey, D.A., 2015. Stationarity issues in time series models. SAS Users Group International 30.

Domke, J., 2019. Provable gradient variance guarantees for black-box variational inference. Advances in Neural Information Processing Systems 32.

Dua, D., Graff, C., 2017. UCI machine learning repository, School of Information and Computer Sciences, University of California, Irvine. http://archive.ics.uci.edu/ml. Available online.

Eichler, M., Dahlhaus, R., Sandkühler, J., 2003. Partial correlation analysis for the identification of synaptic connections. Biological Cybernetics 89, 289–302.

Eubank, R.L., 1999. Nonparametric regression and spline smoothing. CRC Press.

Ferrarelli, F., Kaskie, R.E., Graziano, B., Reis, C.C., Casali, A.G., 2019. Abnormalities in the evoked frontal oscillatory activity of first-episode psychosis: a TMS/EEG study. Schizophrenia Research 206, 436–439.

Fox, C.W., Roberts, S.J., 2012. A tutorial on variational Bayesian inference. Artificial Intelligence Review 38, 85–95.

Garcia, I., Huo, S., Prado, R., Bravo, L., 2020. Dynamic Bayesian temporal modeling and forecasting of short-term wind measurements. Renewable Energy 161, 55–64. doi:10.1016/j.renene.2020.05.182.

Gebka, B., Wohar, M.E., 2013. International herding: Does it differ across sectors? Journal of International Financial Markets, Institutions and Money 23, 55–84.

Gelfand, A.E., Dey, D.K., 1994. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society: Series B (Methodological) 56, 501–514.

Goodman, N.R., 1963. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). The Annals of mathematical statistics 34, 152–177.

Griffin, J.E., Brown, P.J., 2010. Inference with normal-gamma prior distributions in regression problems. Bayesian Analysis 5, 171–188.

Guo, W., Dai, M., 2006. Multivariate time-dependent spectral analysis using Cholesky decomposition. Statistica Sinica , 825–845.

Hagemann, A., 2013. Robust spectral analysis. arXiv preprint arXiv:1111.1965v2 .

Herzmann, D., Wolt, J., . Iowa state university iowa environmental mesonet. URL: https://mesonet.agron.iastate.edu/ASOS/. accessed on 10/05/2020.

Hoffman, M.D., Blei, D.M., 2015. Structured stochastic variational inference, in: Artificial Intelligence and Statistics, pp. 361–369.

Hoffman, M.D., Blei, D.M., Wang, C., Paisley, J., 2013. Stochastic variational inference. Journal of Machine Learning Research 14.

Hong, Y., 2000. Generalized spectral tests for serial dependence. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62, 557–574.

Hosseini, R., Yang, K., Chen, A., Patra, S., 2021. A flexible forecasting model for production systems. arXiv preprint arXiv:2105.01098 .

Huber, F., Feldkircher, M., 2019. Adaptive shrinkage in Bayesian vector autoregressive models. Journal of Business & Economic Statistics 37, 27–39.

Huber, F., Koop, G., Onorante, L., 2021. Inducing sparsity and shrinkage in time-varying parameter models. Journal of Business & Economic Statistics 39, 669–683.

Johannis, M., Flint, L.E., Dettinger, M.D., Flint, A.L., Ochoa, R., 2016. The role of snowpack, rainfall, and reservoirs in buffering California against drought effects. Technical Report. US Geological Survey.

Johansson, B., Chen, D., 2003. The influence of wind and topography on precipitation distribution in sweden: Statistical analysis and modelling. International Journal of Climatology: A Journal of the Royal Meteorological Society 23, 1523–1535.

Karpoff, J.M., 1987. The relation between price changes and trading volume: A survey. Journal of Financial and quantitative Analysis 22, 109–126.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. arXiv no. 1412.6980 .

Kingma, D.P., Welling, M., 2013. Auto-encoding variational Bayes. arXiv no. 1312.6114 .

Kley, T., 2014. Quantile-based spectral analysis in an object-oriented framework and a reference implementation in r: The quantspec package. arXiv preprint arXiv:1408.6755 .

Kley, T., Birr, S., Kley, M.T., 2020. Package 'quantspec' .

Kley, T., Volgushev, S., Dette, H., Hallin, M., 2016. Quantile spectral processes: Asymptotic analysis and inference. Bernoulli 22, 1770–1807.

Koenker, R., Xiao, Z., 2006. Quantile autoregression. Journal of the American statistical association 101, 980–990.

Krafty, R.T., 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., 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.

Lang, S., Brezger, A., 2004. Bayesian p-splines. Journal of computational and graphical statistics 13, 183–212.

Lee, J., Rao, S.S., 2012. The quantile spectral density and comparison based tests for nonlinear time series. arXiv preprint arXiv:1112.2759 .

Lee, L., Lawrence, D., Price, M., 2006. Analysis of water-level response to rainfall and implications for recharge pathways in the chalk aquifer, se england. Journal of hydrology 330, 604–620.

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of experimental social psychology 49, 764–766.

Li, T.H., 2008. Laplace periodogram for time series analysis. Journal of the American Statistical Association 103, 757–768.

Li, T.H., 2012. Quantile periodograms. Journal of the American Statistical Association 107, 765–776.

Li, Z., Krafty, R.T., 2019. Adaptive Bayesian time–frequency analysis of multivariate time series. Journal of the American Statistical Association 114, 453–465.

Li, Z., Rosen, O., Ferrarelli, F., Krafty, R.T., 2021. Adaptive Bayesian spectral analysis of high-dimensional nonstationary time series. Journal of Computational and Graphical Statistics 30, 794–807.

Liang, F., 2009. Improving samc using smoothing methods: theory and applications to Bayesian model selection problems. The Annals of Statistics 37, 2626–2654.

Liang, F., Liu, C., Carroll, R.J., 2007. Stochastic approximation in Monte Carlo computation. Journal of the American Statistical Association 102, 305–320.

Lombard, P.J., 2016. Flood-and drought-related natural hazards activities of the US Geological Survey in New England. Technical Report. US Geological Survey.

López-Oriona, Á., Vilar, J.A., 2021a. Outlier detection for multivariate time series: A functional data approach. Knowledge-Based Systems 233, 107527.

López-Oriona, Á., Vilar, J.A., 2021b. Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series. Expert Systems with Applications 185, 115677.

Mannarano, D., 1998. Automated Surface Observing System (ASOS) User's Guide. NOAA, DoD, FAA, USNavy. URL: `https://www.weather.gov/media/asos/aum-toc.pdf`.

Meier, A., Kirch, C., Meyer, R., 2020. Bayesian nonparametric analysis of multivariate time series: a matrix gamma process approach. Journal of Multivariate Analysis 175, 104560.

Neal, R.M., 2011. MCMC using Hamiltonian dynamics. Handbook of Markov Chain Monte Carlo 2, 2.

NOAA National Centers for Environmental Information, 2021. Streamflow drought indicators. https://www.ncdc.noaa.gov/monitoring-references/dyk/streamflow-indicators. (Accessed 20 May 2022).

Nugent, C., 2017. S&P 500 stock data: Historical stock data for all current S&P 500 companies. Kaggle, https://www.kaggle.com/datasets/camnugent/sandp500. (Accessed 20 April 2022).

OEHHA, 2018. Precipitation. Office of Environmental Health Hazard Assessment. URL: `https://oehha.ca.gov/media/epic/downloads/09precipitation_19dec2018.pdf`.

Ombao, H., Von Sachs, R., Guo, W., 2005. Slex analysis of multivariate nonstationary time series. Journal of the American Statistical Association 100, 519–531.

Ong, V.M.H., Nott, D.J., Smith, M.S., 2018. Gaussian variational approximation with a factor covariance structure. Journal of Computational and Graphical Statistics 27, 465–478.

Park, T., Casella, G., 2008. The bayesian lasso. Journal of the American Statistical Association 103, 681–686.

Park, T., Eckley, I.A., Ombao, H.C., 2014. Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. IEEE Transactions on Signal Processing 62, 5240–5250.

Piironen, J., Vehtari, A., 2017. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics 11, 5018–5051.

Polson, N.G., Scott, J.G., 2010. Shrink globally, act locally: Sparse Bayesian regularization and prediction. Bayesian Statistics 9, 105.

Prado, R., West, M., Krystal, A.D., 2001. Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag–lead structure. Journal of the Royal Statistical Society: Series C (Applied Statistics) 50, 95–109.

Priestley, M.B., 1965. Evolutionary spectra and non-stationary processes. Journal of the Royal Statistical Society: Series B (Methodological) 27, 204–229.

Priestley, M.B., 1981. Spectral analysis and time series (Vols. 1 and 2). London: Academic Press.

Robbins, H., Monro, S., 1951. A stochastic approximation method. The Annals of Mathematical Statistics 22, 400–407.

Rosen, O., Stoffer, D.S., 2007. Automatic estimation of multivariate spectra via smoothing splines. Biometrika 94, 335–345.

Rosen, O., Stoffer, D.S., Wood, S., 2009. Local spectral analysis via a Bayesian mixture of smoothing splines. Journal of the American Statistical Association 104, 249–262.

Rosen, O., Wood, S., Stoffer, D.S., 2012. Adaptspec: Adaptive spectral estimation for nonstationary time series. Journal of the American Statistical Association 107, 1575–1589.

Ruiz, F.R., AUEB, M.T.R., Blei, D., 2016. The generalized reparameterization gradient, in: Advances in Neural Information Processing Systems, pp. 460–468.

Sanderson, J., Fryzlewicz, P., Jones, M., 2010. Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. Biometrika 97, 435–446.

Schwarz, G., 1978. Estimating the dimension of a model. The annals of statistics , 461–464.

Shumway, R.H., Stoffer, D.S., 2011. Time series analysis and its applications. New York: Springer.

Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory 6, 174.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A., 2014. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B: Statistical Methodology 76, 485–493.

Todey, D.P., Herzmann, D.E., Takle, E.S., 2002. The Iowa environmental mesonet—combining observing systems into a single network, in: Sixth Symposium on Integrated Observing Systems.

U.S. Geological Survey, 2016. National water information system data available onthe world wide web (usgs water data for the nation). https://waterdata.usgs.gov/nwis/. (Accessed 20 May 2022).

West, M., Prado, R., Krystal, A.D., 1999. Evaluation and comparison of eeg traces: Latent structure in nonstationary time series. Journal of the American Statistical Association 94, 375–387.

Whittle, P., 1957. Curve and periodogram smoothing. Journal of the Royal Statistical Society: Series B (Methodological) 19, 38–47.

Wood, S.A., Jiang, W., Tanner, M., 2002. Bayesian mixture of splines for spatially adaptive nonparametric regression. Biometrika 89, 513–528.

Xie, Y., 2018. A Bayesian model for spectral density estimation. Open Access Theses & Dissertations, University of Texas at El Paso.

Xu, M., Quiroz, M., Kohn, R., Sisson, S.A., 2019. Variance reduction properties of the reparameterization trick, in: The 22nd International Conference on Artificial Intelligence and Statistics, PMLR. pp. 2711–2720.

Zhang, S., 2016. Adaptive spectral estimation for nonstationary multivariate time series. Computational Statistics & Data Analysis 103, 330–349.

Zhang, S., 2019. Bayesian copula spectral analysis for stationary time series. Computational Statistics & Data Analysis 133, 166–179.

Zhang, X.L., Begleiter, H., Porjesz, B., Wang, W., Litke, A., 1995. Event related potentials during object recognition tasks. Brain Research Bulletin 38, 531–538.

Zhao, W., Prado, R., 2020. Efficient bayesian parcor approaches for dynamic modeling of multivariate time series. Journal of Time Series Analysis 41, 759–784.