UNIVERSITY OF CALIFORNIA
SANTA CRUZ

BAYESIAN MIXTURE MODELS FOR SPECTRAL DENSITY ESTIMATION

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

DOCTOR OF PHILOSOPHY

in

STATISTICS AND APPLIED MATHEMATICS

by

Annalisa Cadonna

March 2017

The Dissertation of Annalisa Cadonna
is approved:

Professor Raquel Prado, Co-chair

Professor Athanasios Kottas, Co-chair

Professor Bruno Sanso

Professor Juhee Lee

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

Bayesian Mixture Models for Spectral Density Estimation

by

Annalisa Cadonna

We introduce a novel Bayesian modeling approach to spectral density estimation for multiple time series. Considering first the case of non-stationary time series, the log-periodogram of each series is modeled as a mixture of Gaussian distributions with frequency-dependent weights and mean functions. The implied model for the log-spectral density is a mixture of linear mean functions with frequency-dependent weights. The mixture weights are built through successive differences of a logit-normal distribution function with frequency-dependent parameters. Building from the construction for a single log-spectral density, we develop a hierarchical extension for multiple stationary time series. Specifically, we set the mean functions to be common to all log-spectral densities and model time series specific mixtures through the parameters of the logit-normal distribution. In addition to accommodating flexible spectral density shapes, a practically important feature of the proposed formulation is that it allows for ready posterior simulation through a Gibbs sampler with closed form full conditional distributions for all model parameters. We then extend the model to multiple locally stationary time series, a particular class

of non-stationary time series, making it suitable for the analysis of time series with spectral characteristics that vary slowly with time. The modeling approach is illustrated with different types of simulated datasets, and used for spectral analysis of multichannel electroencephalographic recordings (EEGs), which provides a key motivating application for the proposed methodology.

Acknowledgments

First and foremost, I wish to thank my advisors, Professor Raquel Prado and Professor Athanasios Kottas, for their support, encouragement and guidance. A sincere thank you for caring about both my academic and personal life during the past four years. Thank you also for tolerating my lack of organization, and for believing things would change any time I showed up to a meeting with a notebook. I would like to thank the other members of my dissertation committee: Professor Bruno Sansó and Professor Juhee Lee, for the helpful comments and insightful suggestions. I also want to thank Professor Hernando Ombao and the Space-Time group at UC Irvine, for the constant flood of ideas and for the constructive feedback on my research.

I am deeply grateful to the friends and AMS colleagues that I met during my Ph.D studies: you guys made a small hippie town feel like home. I am thankful to Chelsea for making me feel welcome in California, for being the first friend in my new life abroad; a profound thanks goes to your family for all the support, warmth, and holiday meals. Thank you Sean for reminding me that whatever happens, the earth is still spinning and the sun is still shining, and Vassilis for bringing the right amount of pessimism in my everyday life. Thanks to Paulo and Aya Gomes for taking care of me, to Nicole B. for trips, chips with hummus and nights on the couch, to Nick for teaching me that fish can be interesting and for his big enthusiasm for the small things, and to Brenton

for being the best frenemy ever. Thank you Nicole M. for your contagious motivation and Brenda for your big empathy. I would like to thank Pedro for being my AMS reference point from day one and for letting me always count on correct homework solutions when I was too busy or too lazy, and Georgi for all the times at the whiteboard discussing (pure) math and proofs. Thank you Sara for showing up and making AMS a happier place. A special mention to the foosball table in 356 and to Yuan's entertaining activities.

Thank you to Lambrate is my Playground for making me wake up everyday laughing and for always being there.

Grazie alla mia famiglia per il supporto incondizionato, per aver appoggiato ogni mia decisione, ogni mia scelta ed ogni mio spostamento e per avermi sempre incoraggiato senza riserve. Vi prometto ufficialmente che questa è la mia ultima tesi.

Chapter 1

Introduction

The problem of modeling multiple time series in the spectral domain arises naturally in fields where information about frequency behavior is relevant and several signals are recorded concurrently, as in neuroscience, econometrics, and geoscience. In these fields, there is growing interest in different types of inference based on a collection of related time series. For example, multichannel electroencephalography records measurements of electrical potential fluctuations at multiple locations on the scalp of a human subject. Identifying which locations lead to electrical brain signals with similar spectral densities is particularly meaningful, as it provides insights about the physiological state of the subject and about the spatial structure of cortical brain activity under certain experimental or clinical conditions. Section 1.1 provides some background and definitions that are key to understand spectral analysis and its

importance. In Section 1.2 we describe the key motivating application for the proposed methodology, specifically the analysis of multi-channel electroencephalographic recordings (EEGs). Finally, in Section 1.3 the main objectives and contribution of this PhD dissertation are outlined.

1.1 Time series in the frequency domain

Consider a real valued discrete time series $\{X_t\}_{t=0,1,...}$. A time series is said to be strictly stationary if, for any $t_1, t_2, .., t_n$, and for any k, the joint probability distribution of $\{X_{t_1}, ..., X_{t_n}\}$ is identical to the joint probability distribution of $\{X_{t_1+k}, ..., X_{t_n+k}\}$. A time series is said to be weakly stationary, or simply stationary, if its first two moments do not vary with time. The expected value of a stationary time series is a constant $\mu = \mathbb{E}[X_t]$, for $t = 1, 2, ...$, and the autocovariance function depends only on the lag k, that is $\gamma(k) = \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)]$.

A stationary process can be represented as

$$X_t = \int_{-\pi}^{\pi} \exp(it\omega) dZ(\omega) \tag{1.1}$$

where $Z(\omega)$ a process whose increments for any two distinct values are uncorrelated, i.e., for two distinct frequencies ω and ω', $dZ(\omega) = \{Z(\omega + d\omega) - Z(\omega)\}$ and $dZ(\omega') = \{Z(\omega + d\omega') - Z(\omega')\}$ are uncorrelated. In addition, for each ω, $Z(\omega)$ is such that $\mathbb{E}(|Z(\omega)|^2) = dF(\omega)$. Equation (1.1) provides the spectral

representation of the process $X(t)$. The integral on the right hand side, for a fixed realization of the process, is a Fourier-Stieltjes integral. If $f(\omega) = F'(\omega)$ exists for every ω, then $dF(\omega) = f(\omega)d\omega$, where $f(\omega)$ is the spectral density of the process. The spectral representation states that, virtually, any stationary process can be represented as the limit of the sum of sine and cosine functions with random uncorrelated coefficients. This is crucial for the physical interpretation of the spectral density as distribution of the energy of the process over the frequency range. For a more complete explanation of the spectral representation of stationary processes, refer to Priestley (1982).

Consider a zero-mean stationary time series $\{X_t : t = 1, 2, ...\}$, with absolutely summable autocovariance function $\gamma(\cdot)$. The spectral density function can be expressed as the Fourier transform of the autocovariance function:

$$f(\omega) = \sum_{k=-\infty}^{+\infty} \gamma(k) \exp(-ik\omega), \quad \text{for} - \pi \leq \omega \leq \pi,$$

where $\gamma(k) = \mathrm{E}(X_{t+k}X_t)$ denotes the autocovariance function. The relationship between the spectral density function and the autocovariance function provides a link between the time-domain and the frequency-domain analyses. Given n observation $X_1, ..., X_n$, the priodogram, denoted with $I_n(\omega)$, is defined for $\omega \in [-\pi, \pi]$ by

$$I_n(\omega) = \left| \sum_{t=1}^{n} X_t \exp(-it\omega) \right|^2 /n. \tag{1.2}$$

Although $I_n(\omega)$ is defined for all $\omega \in [-\pi, \pi]$, it is computed at the Fourier

frequencies $\omega_j = 2\pi j/n$, for $j = 0, \ldots, \lfloor n/2 \rfloor$, where $\lfloor n/2 \rfloor$ is the largest integer not greater than $n/2$. Because of the symmetry of the periodogram, there are only $\lfloor n/2 \rfloor + 1$ effective observations. The periodogram is, for each ω, an asymptotically unbiased estimator of the spectral density $f(\omega)$. However, the periodogram is not a consistent estimator of the spectral density $f(\omega)$, in the sense that $\mathrm{Var}(I_n(\omega))$ does not tend to zero as $n \to \infty$. Moreover, as a function of ω, the periodogram has an erratic and fluctuating form, due to the fact that $\mathrm{Cov}(I_n(\omega), I_n(\omega'))$ decreases as n increases. One way to reduce the variance of the periodogram is to include less terms in (1.2). Improved estimators have been obtained by smoothing the periodogram (equivalent to weighting less the terms corresponding to the tails of the autocovariance function) through windowing methods (e.g., Parzen, 1962). However, truncating or smoothing the periodogram inevitably introduces bias, that adds to the bias due to the finite sample size n. For proofs of the asymptotic properties of the periodogram and smoothed periodogram, see Priestley (1982).

Let us consider a zero-mean Gaussian time series. Whittle (1957) derived an approximation of the Gaussian likelihood. He proved that, for large sample size n, the periodogram observations $I_n(\omega_j)$, for $\omega_j \neq 0, \pi$, are independent exponentially distributed with mean the spectral density $f(\omega)$. For $\omega_j = 0, \pi$, the priodogram observations $I_n(\omega_j)$ follow a gamma distribution with mean $f(\omega)$ and shape parameter $1/2$. The main advantage of the Whittle likelihood with

respect to the true likelihood is that the spectral density appears explicitly and not through the autocovariance function. Most model-based approaches to spectral density estimation are built from the Whittle likelihood approximation to the periodogram distribution. Moreover, it has been shown that the asymptotic Whittle approximation to the likelihood remains valid for non-Gaussian (Hannan, 1973; Naito et al., 2010) and even non-linear time series (Shao and Wu, 2007). In practice, for finite sample size, the Whittle likelihood for non-Gaussian models presents a significan efficiency loss compared to the true likelihood, as studied in Contreras-Cristan et al. (2006).

Using the Whittle's likelihood, the estimation problem can be cast in a regression framework with observations given by the log-periodogram ordinates and regression function defined by the log-spectral density. Specifically, $\log(I_n(\omega_j)) = \log(f(\omega_j)) + \epsilon_j$, for $j = 1, \ldots, N$, where the ϵ_j follow a log-exponential distribution with scale parameter 1. In this context, frequentist estimation approaches include approximating the distribution of the ϵ_j with a normal distribution and fitting a smoothing spline to the log-periodogram (Wahba, 1980), and maximizing the Whittle likelihood with a roughness penalty term (Pawitan and O'Sullivan, 1994). Regarding Bayesian modeling approaches: Carter and Kohn (1997) approximate the distribution of the ϵ_j with a mixture of normal distributions and assign a smoothing prior to $\log(f(\omega))$; Choudhuri et al. (2004) develop an approach based on Bernstein

polynomial priors (Petrone, 1999) for the spectral density; Rosen and Stoffer (2007) express the log-spectral density as $\log(f(\omega)) = \alpha_0 + \alpha_1\omega + h(\omega)$, with a Gaussian process prior on $h(\omega)$; and Pensky et al. (2007) propose Bayesian wavelet-based smoothing of the log-periodogram. More recently, Macaro and Prado (2014) extended Choudhuri et al. (2004) to consider spectral decompositions of multiple time series in designed factorial experiments.

1.2 Motivating application: multichannel electroencephalographic recordings (EEGs)

Spectral densities can appropriately summarize characteristics of brain signals recorded in various experimental or clinical settings, as documented in the literature. For instance, certain spectral characteristics of electroencephalographic recordings (EEGs) recorded from patients who received electroconvulsive therapy (ECT) as a treatment for major depression have been associated with the clinical efficacy of such treatment (Krystal et al., 1999). Moreover, identifying which locations lead to electrical brain signals with similar spectral densities is particularly meaningful, as it provides insights about the physiological state of the subject and about the cortical brain activity under certain experimental or clinical conditions. For example, in the area of monitoring and detection of mental fatigue, prior EEG studies have suggested an association

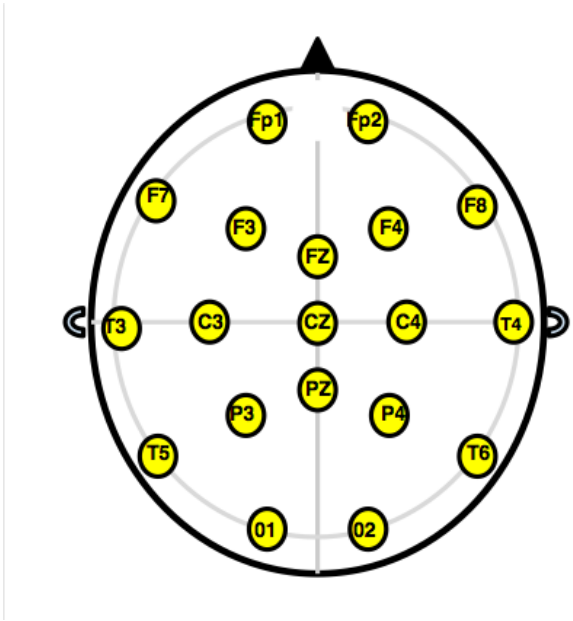

Figure 1.1: Schematic representation of the EEG electrodes placed on a subject's scalp.

of fatigue with an increase in the theta (4-8 Hz) band power observed in the estimated spectral of signals recorded in channels located in mid-line frontal scalp areas (Trejo et al., 2007).

Figure 1.1 shows a schematic representation of EEG electrodes placed on a subject's scalp; in this case, the number of channels is 19. Figure 1.2 shows the EEGs recording from the C_z channel, for a subject undergoing ECT treatment. Usually only a small segment of the EEG observations is considered, because the entire series is non-stationary. In Chapter 3, we will analyze the three segments between vertical lines in Figure 1.2 separately. In Chapter 4 we consider the time series in its entirety, through an extension of the hierarchical model to non-stationary time series.

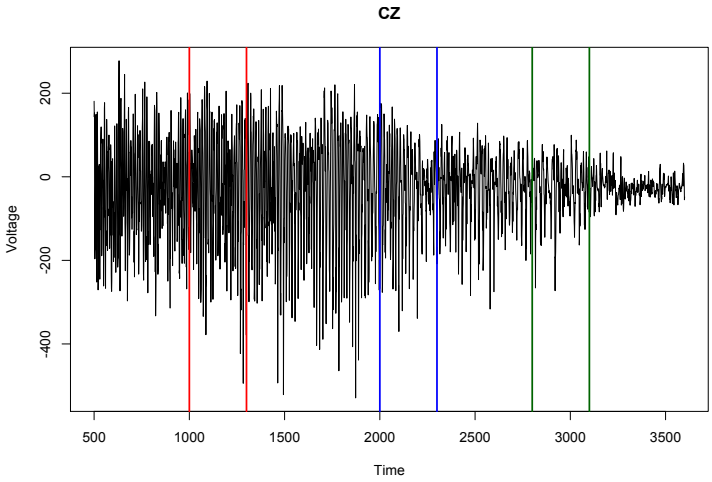

Figure 1.2: EEG signal for the C_z channel for a subject undergoing ECT treatment.

It is clear that eveloping and implementing flexible methods for spectral analysis of multiple time series is crucial in neuroscience. Available parametric time-domain approaches based on state-space models are often used in practice because of their computational tractability, but such approaches are often restrictive. In particular, they require the selection of a specific parametric model to describe the data, whereas in real-world applications that involve the analysis of complex time series, it is uncommon to know in advance the form of the data-generating process. Moreover, many existing approaches, especially those in the frequency-domain, cannot be easily extended to settings that involve multiple signals, multiple subjects and/or multiple trials. In fact, most time-domain analyses can be extended through hierarchical modeling. This is much harder to do in the frequency-domain, especially due to a huge increase in the computational complexity

1.3 Research objectives

In the following paragraphs, we summarize the main research contributions of this Ph.D. thesis.

- We propose a new approach to modeling spectral densities. Specifically, motivated by results in Jiang and Tanner (1999a) and Norets (2010), we replace the Whittle likelihood for the log-periodogram distribution with a mixture of Gaussian distributions with frequency-dependent mixture weights and mean parameters. This structure implies that the log-spectral density is modeled as a smooth mixture of the Gaussian mean functions. Moreover, casting the spectral density in a mixture framework is key for the implementation of efficient Markov Chain Monte Carlo algorithms.

- We introduce a novel formulation for the mixture weights and propose a Bayesian hierarchical model for multiple time series in the spectral domain. The mixture weights are built by consecutive differences of a logit-normal distribution function with frequency-dependent parameters. Such construction presents a critical computational advantage, as we can introduce normally distributed auxiliary random variables and draw from well established posterior simulation methods for mixture models. Moreover, the new proposed weights are the starting point for the hier-

archical extension of our model to multiple time series. We propose a flexible Bayesian modeling approach for multiple time series that leads to full posterior inference of the multiple spectral densities and also allows us to identify groups of time series with similar spectral characteristics. The proposed model is more parsimonious than the fully Bayesian model-based spectral estimation approaches mentioned in Section 1.1, leading to more efficient posterior simulation. Therefore, the methodology can be used to analyze temporal datasets that consist of a relatively large number of related time series. Moreover, a nice feature of our modeling approach is that it allows to group sets of time series that share similar spectral characteristics as illustrated in Chapter 3, through simulations and real data examples.

- The last contribution of this dissertation is the extension of the model to non-stationary time series. Often, in fact, the stationarity assumption is not satisfied, especially when we deal with long time series. Our model formulation is key for this extension, in that we can make the weights and the means of the mixture depend both on frequency and on time. The computational advantages are still valid and we can extend the model for one non-stationary time series to multiple non-stationary time series, in the same fashion as we did for multiple stationary time series.

The outline of the dissertation is as follows. We begin in Chapter 2 by

presenting the new modeling idea for spectral density estimation through a local mixture model on the log-periodogram observations. First, we use logistic weights to model the dependency on the frequency in the weights. We then introduce a novel form of the weights, based on differences of cumulative distribution functions that depend on the frequency. In Chapter 3, the model is extended to multiple time series. Specifically, a Bayesian hierarchical model is used to borrow strength across related time series. Spectral analysis for non-stationary time series is the focus in Chapter 4; the model can be expanded to include information about time and provide an estimate of a log-spectral density that varies with time. In all the chapters we provide results and estimates for significant synthetic datasets and from the EEG dataset described in Section 1.2. Chapter 5 concludes with a summary and some final remarks. Proofs of Lemmas and Theorems are provided in Appendix A. Technical details on posterior simulation methods can be found in Appendix B.

Chapter 2

Mixture modeling approach to spectral density estimaton

2.1 Introduction

In this chapter, we present a new Bayesian approach to spectral density estimation. Specifically, we approximate the distribution of the log-periodogram observations under the Whittle with a mixture of normal distributions, with frequency-dependent mean functions and weights. The mixture of normal distributions implies a model for the log-spectral density that is a mixture of the mean functions, which enables a wide range of shapes. First, we study the performance of our model using logistic weights. Then, we propose a new formulation for the weights, based on successive differences of a cumulative

distribution function with finite support, whose location parameter depends on the frequency. The proposed weights are advantageous in that they facilitate extensions to hierarchical modeling in a multiple time series setting. The mixture representation of the log-spectral density is also key for the implementation of a Gibbs sampling Markov Chain Monte Carlo (MCMC) algorithm. In the case of the proposed weights, all the full posterior distributions are available in closed form.

The outline of the chapter is as follows. The modeling approach is introduced in section 2.2. In Section 2.3 we describe the logistic weights and introduce the new weights based on the differences of cumulative distribution functions. Section 2.4 contains the main theoretical results. Our method, for logistic weights, is illustrated with synthetic and real data sets in Section 2.5. Results, using the new proposed weights, are presented in Section 2.6. Section 2.7 contains further discussion about the advantages of the proposed modeling approach.

2.2 The modeling approach

Let n be number of observations from the time series and $N = \lfloor n/2 \rfloor - 1$ the number of observations of the periodogram. Following common practice, we exclude the observations at $\omega_j = 0, \pi$, resulting in $N = \lfloor n/2 \rfloor - 1$ for the sample size Let $I_n(\omega_j)$, with $\omega_j = 2\pi j/n$ (for $j = 1, ..., N$), be the observations of the

periodogram of a zero-mean stationary time series. We define the translated log-periodogram, at the Fourier frequency $\omega_j = 2\pi j/n$, as $y_j = \log(I_n(\omega_j)) + \gamma$, for $j = 1, ..., N$, where γ is the Euler-Mascheroni constant. The translation constant is such that, under the Whittle approximation, the expected value of the log-periodogram is the log-spectral density. At the Fourier frequencies, for a large enough sample size n, we can consider the y_j as independent. Moreover, under the Whittle likelihood approximation, the y_j have the following distribution:

$$q_Y(y) = \exp\{y - \gamma - \log(f(\omega)) - \exp(y - \gamma - \log(f(\omega)))\}, \quad y \in \mathbb{R}. \quad (2.1)$$

Therefore, $\mathrm{E}[y_j] = \log(f(\omega_j))$ and $\mathrm{Var}[y_j] = \pi^2/6$. Notice that the distribution in (2.1) is in the exponential family, and $-y_j$ are Gumbel distributed with scale parameter 1 and location parameter defined additively through $\log(f(\omega))$ and γ, such that the mean is $-\log(f(\omega))$. The Whittle approximation has been widely used in the literature because the spectral density appears explicitly in the approximate likelihood rather than through the covariance function. Although (2.1) is a standard distribution, the spectral density enters the likelihood in a non-standard fashion through the mean parameter.

We propose to replace the distribution in (2.1) (derived from the Whittle likelihood) with a structured mixture of Gaussian distributions, defined through frequency-dependent mixture weights and Gaussian mean functions. In this way, we obtain a model for the log-spectral density that is a mixture

of the Gaussian mean functions. More specifically, we approximate the distribution of y_j with a mixture of normal distributions, with means that depend linearly on ω_j and frequency-dependent weights:

$$y_j|\boldsymbol{\theta} \overset{ind.}{\sim} \sum_{k=1}^{K} g_k(\omega_j;\boldsymbol{\xi})\,\mathrm{N}(y_j \mid \alpha_k + \beta_k\omega_j, \sigma^2), \qquad j = 1,\ldots,N, \tag{2.2}$$

where $g_k(\omega_j;\boldsymbol{\xi})$ denotes the k-th mixture weight, and $\boldsymbol{\xi}$ is the vector of the weight parameters. The weight parameters, as well as the number of weight parameters, vary depending on the specific form of the weights and will be fully specified in each case. The vector $\boldsymbol{\theta}$ collects all model parameters, specifically, the weight parameters $\boldsymbol{\xi}$, the intercept and slope parameters of the K mixture components means, that is, $\boldsymbol{\alpha} = \{\alpha_k : k = 1,...,K\}$ and $\boldsymbol{\beta} = \{\beta_k : k = 1,...,K\}$, and the common variance parameter σ^2.

From the approximation in (2.2), taking the expected value, we obtain that the model for the log-spectral density is given by

$$\log(f(\omega)) = \sum_{k=1}^{K} g_k(\omega;\boldsymbol{\xi})\,\{\alpha_k + \beta_k\omega\}, \qquad \omega \in (0,\pi), \tag{2.3}$$

that is, the log-spectral density admits a representation as a mixture of linear functions with component specific intercept and slope parameters, and with frequency-dependent weights that allow for local adjustment, and thus flexible spectral density shapes.

In addition to the appealing interpretation of the implied spectral density model, further theoretical justification for the approximation in (2.2) can be

provided by means of results in the L_p norm for the spectral density. Such justifications can be found in section 2.4. The Gaussian mixture model in (2.2), and the implied model for the log-spectral density in (2.3), are motivated by the theoretical results of Jiang and Tanner (1999a) and Norets (2010). Jiang and Tanner (1999a) show that an exponential response distribution, involving a regression function on a finite support, can be approximated by a mixture of exponential distributions with means that depend on the covariates and with covariate-dependent mixture weights. More directly related to our model, Norets (2010) presents approximation properties of finite local mixtures of normal regressions as flexible models for conditional densities. The work in Norets (2010) focuses on the joint distribution of the response and covariates, showing that, under certain conditions, the joint distribution can be approximated in Kullback-Leibler divergence by different specifications of local finite mixtures of normals in which means, variances, and weights can depend on the covariates. Here, we consider fixed covariate values defined by the Fourier frequencies. We have explored both methods and found that the mixture of normal distributions is computationally more efficient.

2.3 Local mixture weights

The key property underlying the approximation results in Jiang and Tanner (1999a) and Norets (2010) is that the covariate-dependent (frequency-

dependent in our context) mixture weights are such that, for some values of the weight parameters, they approximate a set indicator functions on a fine partition of the finite support. Under this condition, it can be proved that, as the number of components increases, the approximation in (2.2) tends to (2.1) in the sense of the Kullback-Liebler divergence. Moreover, under a smoothness assumption for the log-spectral density – assuming that $\log(f(\omega))$ and its first and second derivatives are continuous and bounded – further theoretical justification for the approximation in (2.3) can be provided by means of results in the L_p norm for the log-spectral density. The theoretical results for both formulations of the weights are enunciated and discussed in Section 2.4.

2.3.1 Logistic weights

Logistic weights are defined, for $k = 1, ..., K$, as

$$g_k(\omega; \boldsymbol{\xi}) \equiv g_k(\omega_j; \lambda, \boldsymbol{\zeta}, \boldsymbol{\phi}) = \frac{\{\exp(\zeta_k + \phi_k \omega_j)/\lambda\}}{\sum_{i=1}^{K} \exp\{(\zeta_i + \phi_i \omega_j)/\lambda\}}. \tag{2.4}$$

In this case, the vector of weight parameters $\boldsymbol{\xi}$ becomes $\boldsymbol{\xi} = (\{\zeta_k : k = 1, ..., K\}, \{\phi_k : k = 1, ..., K\}, \lambda)$. Hence, there are $2K + 1$ mixture weight parameters. The logistic weights partition the support with soft boundaries. The parameter λ controls the smoothness of the transition between the subsets of $(0, \pi)$ induced by the logistic weights. The larger the value of λ, the smoother is the corresponding estimate of the spectral density.

Here, $\boldsymbol{\theta}$ collects all model parameters, in particular, it includes the parameters for the logistic weights, $\zeta = \{\zeta_k : k = 1, ..., K\}$, $\phi = \{\phi_k : k = 1, ..., K\}$ and λ, the intercept and slope parameters for the means of the normal mixture components, $\alpha = \{\alpha_k : k = 1, ..., K\}$ and $\beta = \{\beta_k : k = 1, ..., K\}$, and the common variance parameter σ^2.

Prior specification

For practical purposes, confirmed from empirical investigation with several data sets including the ones of Section 3, we have observed that in general a relatively small number of mixture components suffices to capture different spectral density shapes, with inference results being robust to the choice of K. Naturally, the modeling approach can be generalized with random K, albeit at the expense of a more computationally challenging approach to inference. An arguably more interesting extension involves a nonparametric formulation for the mixture model which can provide model-based estimation of the effective number of mixture components.

To complete the full Bayesian model, we assume prior independence among and between the parameters of each mixture component. Specifically, we use a normal prior distribution with mean μ_α and variance σ^2_α for the α_k, and a normal prior with mean μ_β and variance σ^2_β for the β_k. For the common variance parameter, σ^2, we use an inverse-gamma prior, and for the smoothness

parameter, λ, a gamma prior. We place standard normal priors on the ζ_k and ϕ_k. This choice is motivated by the fact that the variances of ζ_k and ϕ_k and λ cannot be estimated simultaneously. The prior of λ expresses our belief on the degree of smoothness of the spectral density. As demonstrated with the data examples of Section 3, the smoothness parameter can be learned from the data. The prior on the intercept parameters α_k summarizes information about the spectral density value near $\omega = 0$. Moreover, the prior on the slope parameters β_k can be used to express beliefs about the shape of the spectral density. For instance, for multimodal spectral densities, we expect some β_k to be positive and some negative, whereas for unimodal spectral densities, we expect all β_k to have the same sign.

Posterior simulation

Using logistic weights, the model in (2.2) can be expanded in hierarchical form by introducing configuration variables $(r_1, \ldots, r_N)$, where each r_j, $j = 1, \ldots, N$, has a discrete distribution with values in $\{1, \ldots, K\}$:

$$y_j \mid r_j, \alpha, \beta, \sigma^2 \overset{ind}{\sim} \mathrm{N}(y_j | \alpha_{r_j} + \beta_{r_j}\omega_j, \sigma^2), \qquad j = 1, \ldots, N$$

$$r_j \mid \zeta, \phi, \lambda \overset{ind}{\sim} \sum_{k=1}^{K} g_k(\omega_j; \lambda, \zeta, \phi)\delta_k(r_j), \qquad j = 1, \ldots, N$$

where $\delta_k(r_j) = 1$ if and only if $r_j = k$. The equivalent model is key in sampling from the joint posterior distribution of the parameters. We develop a Markov chain Monte Carlo algorithm to simulate from the joint posterior distribu-

tion of the parameters, which is based almost exclusively on Gibbs sampling steps. The full conditional distribution for each r_j is a discrete distribution on $\{1, \ldots, K\}$ with updated probabilities. We have conjugate full conditional distributions for the (α_k, β_k), and for σ^2. Updating the parameters for the logistic weights is more challenging, and we use a data augmentation step based on auxiliary Pólya-Gamma variables (Polson et al., 2013). Technical details about how we can use auxiliary Pólya-Gamma variables to obtain closed form full conditional distributions for $\{\zeta_k : k = 1, ..., K\}$ and $\{\phi_k : k = 1, ..., K\}$ can be found in Appendix B. Sampling λ requires a Metropolis-Hastings step. Details of the posterior simulation algorithm are provided in Appendix B.

Note that the variance of the prior spectral density increases with ω. To minimize this effect, when fitting the model we normalize the support $(0, \pi)$ of the spectral density to the unit interval; the results can be reported on the original scale through straightforward transformation.

2.3.2 Local weights based on differences of cumulative distribution functions

We propose a novel specification for the mixture weights, as an alternative to the logistic weights. The proposed weights are built by consecutive differences of a distribution function on $(0, \pi)$ with frequency-dependent parameters. We use a logit-normal distribution on $(0, \pi)$, but any distribution

with bounded support can be considered. The idea is reminiscent of the Bernstein polynomial prior in Petrone (1999), in that the weights are built through differences of cumulative distribution functions.

The weights, for $k = 1, .., K$ are defined as follows:

$$g_k(\omega; \boldsymbol{\xi}) \equiv g_k(\mu(\omega), \tau) = \int_{\pi(k-1)/K}^{\pi k/K} f_Y(y \mid \mu(\omega), \tau)\, \mathrm{d}y, \tag{2.5}$$

where $f_Y(y \mid \mu(\omega), \tau)$ is the density of a logit-normal$(\mu(\omega), \tau)$ distribution on $(0, \pi)$. A random variable Y is said to have logit-normal(μ, τ) distribution on $(0, \pi)$ if $\log(Y/(\pi - Y))$ is normal distributed with mean μ and precision parameter τ. Hence, at each frequency, we have a different set of weights which however evolve smoothly with the frequency. If $\mu(\omega)$ is a monotonic function in ω, the weights define a partition on the support $(0, \pi)$. We use a parametric form for $\mu(\omega)$ to determine the location of the modes of the weights. In fact, the mode for the k-th weight does not depend on τ, and it is the value ω_k such that $\mu(\omega_k) = \{\log(k/(K-k)) + \log((k-1)/(K-k+1))\}/2$. We use a linear function $\mu(\omega) = \zeta + \phi\omega$, but another monotonic function of the frequency can be chosen. The parameter τ is a smoothness parameter, with smaller values of τ leading to smoother spectral densities. Hence, the parameters of the logit-normal distribution are interpretable and play a clear role in the shape of the weights.

The formulation for the mixture weights in (3.1), including the choice of the logit-normal distribution function, facilitates the implementation of a Markov

chain Monte Carlo (MCMC) algorithm for posterior simulation. In particular, we can augment model (2.2), using continuous auxiliary variables. For each y_j, $j = 1, ..., N$, we introduce auxiliary variable r_j, which is normally distributed with mean $\mu(\omega_j) = \zeta + \phi\omega_j$ and precision parameter τ. Then, the augmented model can be written as:

$$y_j \mid r_j, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2 \overset{ind.}{\sim} \sum_{k=1}^{K} \mathrm{N}(y_j \mid \alpha_k + \beta_k\omega_j, \sigma^2)\, \mathbb{I}\left\{\frac{k-1}{K} < \frac{\exp(r_j)}{1+\exp(r_j)} \leq \frac{k}{K}\right\},$$

$$r_j \mid \boldsymbol{\xi} \overset{ind.}{\sim} \mathrm{N}(r_j \mid \zeta + \phi\omega_j, 1/\tau),$$

where $\boldsymbol{\xi} = (\zeta, \phi, \tau)$. The full Bayesian model for a single spectral density would be completed with priors for σ^2 and for the elements of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$. This structure allows for a straightforward implementation of a Gibbs sampling algorithm with full conditional distributions available in closed form for all model parameters. The full conditional distributions are available in Appendix B.

2.4 Theoretical results

We want to prove that a smooth mixture of linear functions, $h_K(\omega) = \sum_{k=1}^{K} g_k(\omega; \boldsymbol{\xi})\{\alpha_k + \beta_k\omega\}$, can be used to approximate a smooth function $h(\cdot)$ on $\Omega = (0, \pi)$.

In the following, $\|f(\cdot)\|_p = \left(\int_0^\pi |f(\omega)|^p d\sigma(\omega)\right)^{\frac{1}{p}}$, where σ is a probability measure on $\Omega = (0, \pi)$, absolutely continuous with respect to Lebesgue measure;

χ_B is the indicator function on the subset $B \subseteq \Omega$: $\chi_B(\omega) = 1$ when $\omega \in B$, $\chi_B(\omega) = 0$ when $\omega \notin B$. Let define the sequence of partitions of Ω given by $\{Q_k^K, k = 1, \ldots, \}$, where $Q_k^K = [(k-1)\pi/K, k\pi/K)$ for $k = 1, \ldots, K-1$, and $Q_K^K = [(K-1)\pi/K, \pi]$. This sequence of partitions has the property that, for every K and k, being ω_1 and ω_2 two points in Q_k^K, $|\omega_1 - \omega_2| \leq \pi/K$.

Here, we show that the approximation above is valid in the case of the local weights introduced in Section 2.3. Lemma 2.1, the proof of which can be found in Jiang and Tanner (1999a), shows that logistic weights converge, for some values of the parameters, to a set of indicator functions on a partition of Ω, specifically K intervals of equal length. Lemma 2.2 states a similar results for the weights based on differences of cumulative distribution functions, with the difference that the partition consist on a smaller number of intervals. Theorem 2.1 and 2.2 state the convergence results in the L_p norm of our approximation to the true log-spectral density, as the number of components K goes to infinity, both n the case of logistic weights and of the new weights, respectively. Note that we enunciate the results for $\Omega = (0, \pi)$, but they are valid for any bounded interval on the real line.

Lemma 2.1. *For logistic weights, as defined in (2.4) we have that, for all K, for each $\epsilon > 0$, there exist $\zeta^{(\epsilon)}, \phi^{(\epsilon)}, \lambda^{(\epsilon)}$ such that*

$\sup_{1 \leq k \leq K} ||g_k(\cdot; \lambda^{(\epsilon)}, \zeta^{(\epsilon)}, \phi^{(\epsilon)}) - \chi_{Q_k^K}||_p < \epsilon$.

The following result establishes that the distance in the L_p norm (denoted

by $\|\cdot\|_p$) between the target log-spectral density, h, and the mixture model $h_K(\omega) = \sum_{k=1}^{K} g_k(\omega; \lambda, \zeta, \phi)(\alpha_k + \beta_k \omega)$ is bounded by a constant which is inversely proportional to the square of the number of mixture components K.

Theorem 2.1. *Let $h(\omega) \in W^{\infty}_{2,K_0}$, where W^{∞}_{2,K_0} denotes the Sobolev space of continuous functions on $\Omega = (0, \pi)$ bounded by K_0, with the first two derivatives continuous and bounded by K_0. Then, $\inf_{h_K} \|h_K - h(\cdot)\|_p \leq \pi^2 K_0 / 2K^2$, where $h_K(\omega) = \sum_{k=1}^{K} g_k(\omega; \lambda, \boldsymbol{\zeta}, \boldsymbol{\phi})(\alpha_k + \beta_k \omega)$.*

If Y is distributed as a logit-normal$(\mu(\omega), \tau)$ on $(0, \pi)$, then $Y = \pi \exp(X)/(1+\exp(X))$, where X follows a normal distribution with mean $\mu(\omega)$ and variance $1/\tau$. Hence, the weights defined in (3.1), can be rewritten as

$$g_k(\mu(\omega), \tau) = \frac{1}{\sqrt{2\pi}} \int_{(b_{k-1}-\mu(\omega))\sqrt{\tau}}^{(b_k-\mu(\omega))\sqrt{\tau}} \exp(-x^2/2)\, \mathrm{d}x, \tag{2.6}$$

with $b_k = \log\{k/(K-k)\}$, $b_{k-1} = \log\{(k-1)/(K-k+1)\}$, and $\mu(\omega) = \zeta + \phi\omega$.

Lemma 2.2. *Let $g_k(\mu(\omega), \tau)$ be the k-th weight as defined in (2.6). Then, there exist values for ζ and ϕ, and integers k_1 and k_2, with $k_2 > k_1$, such that, for $k = k_1 + 1, ..., k_2$, $\lim_{\tau \to \infty} ||g_k - \chi_{Q_k}||_p = 0$, for any $p \in \mathbb{N}$. Moreover, for $1 < k \leq k_1$ or $k_2 < k \leq K$, $\lim_{\tau \to \infty} ||g_k||_p = 0$, for any $p \in \mathbb{N}$.*

According to Lemma 2.2 the proposed local mixture weights approximate the set of indicator functions on the partition $\{Q_{k_1+1}, ..., Q_{k_2}\}$, for any fixed K, k_1 and k_2, with $k_2 > k_1$. The following result establishes that the distance

in the L_p norm between the target log-spectral density, h, and the proposed mixture model h_K is bounded by a constant that is inversely proportional to the square of $K^* = k_2 - k_1 < K$.

Theorem 2.2. *Let $h \in W^{\infty}_{2,K_0}$, that is, the Sobolev space of continuous functions bounded by K_0, with the first two derivatives continuous and bounded by K_0. Then,* $\inf_{h_K} \|h_K - h\|_p \leq K_0/(2K^{*2})$.

The proof of Lemmas 2.1 and 2.2 and Theorems 2.1 2.2 can be found in Appendix A.

The above results show that we can choose values of the parameters $\boldsymbol{\theta}$ such that the L_p distance between the true log-spectral density and the proposed mixture is bounded by a constant. This constant, in the case of logistic weights, is proportional to K_0/K^2, where K_0 is related to the smoothness of the true spectral density. Hence, the distance decreases quadratically with the number of components. If we have prior knowledge on the smoothness of the log-spectral density, we can use it to fix K. For the new proposed weights, the constant is proportional to K_0/K^{*2}, where K_0 is the same as above, and $K^* < K$ is the number of effective components determined by $\boldsymbol{\theta}$. Hence, in general, we need a larger K for the new proposed weights to obtain the same degree of approximation as in the logistic weights. However, the computational advantages make this issue secondary and the new formulation of the weights is key for the extension to multiple spectral densities.

2.5 Numerical illustration: Logistic weights mixture model

2.5.1 Synthetic datasets

We evaluate the performance of our method by analyzing simulated datasets. Focusing first on monotonic spectral densities, we simulated data from two autoregressive processes of order one. Specifically, we considered $x_{1,t} = 0.7x_{1,t-1} + \epsilon_{1,t}$, with $\epsilon_{1,t} \sim \mathrm{N}(0,1)$ and $x_{2,t} = -0.9x_{2,t-1} + \epsilon_{2,t}$, with $\epsilon_{2,t} \sim \mathrm{N}(0,1)$, and simulated 400 observations from each of these processes. Here, we only show the results with the number of components fixed at $K = 10$, however, results were essentially the same with larger K. For the α_k and β_k, we use zero-mean normal priors with variances $\sigma_\alpha^2 = \sigma_\beta^2 = 100$. An inverse-gamma prior with shape parameter 3 and scale 6 is placed on σ^2. For λ, we use a gamma prior with shape parameter 2 and mean 0.2, which places almost all its mass in $(0,1)$. The estimated log-spectral densities and the corresponding 95% posterior intervals are reported in Figure 2.1, together with the true log-spectral densities and the periodogram observations. The spectral densities for these processes are respectively monotonically decreasing and increasing. Notice that our model successfully captures their behavior. The posterior density for λ for each case is plotted in Figure 2.3.

To evaluate model performance for more complex spectral densities, we

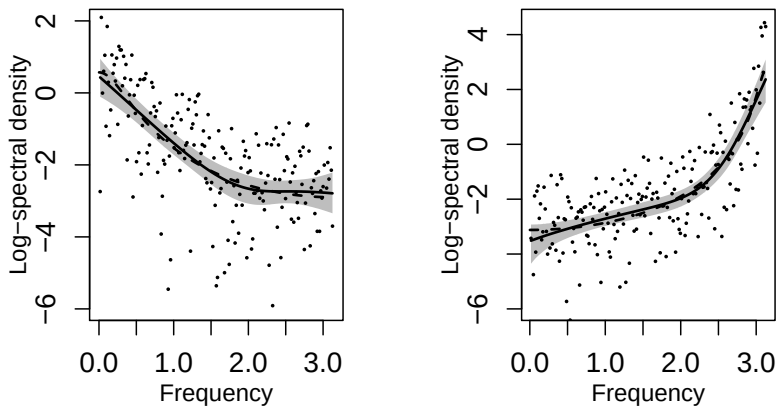

Figure 2.1: Synthetic data from a single AR(1) process. Posterior mean (solid dark line), 95% credible interval (shaded area), true log-spectral density (dashed line) and log-periodogram (dots) for data simulated from autoregressive processes of order one, with parameters 0.7 (left) and -0.9 (right), under the model with logistic weights.

simulated data from the sum of two autoregressive processes and a white noise term. Specifically, let $x_{3,t} = 0.9x_{3,t-1}+\epsilon_{3,t}$, with $\epsilon_{3,t} \sim \mathrm{N}(0,1)$, $x_{4,t} = 0.9x_{4,t-1}-0.9x_{4,t-2}+\epsilon_{4,t}$, with $\epsilon_{4,t} \sim \mathrm{N}(0,1)$ and $x_{5,t} = -0.8x_{5,t-1} - 0.8x_{5,t-2} + \epsilon_{5,t}$, with $\epsilon_{5,t} \sim \mathrm{N}(0,1)$. We construct $z_{1,t} = x_{3,t} + x_{5,t} + \nu_{1,t}$, where $\nu_{1,t} \sim \mathrm{N}(0,1)$, and $z_{2,t} = x_{4,t} + x_{5,t} + \nu_{2,t}$, where $\nu_{2,t} \sim \mathrm{N}(0,1)$. We simulated 400 observations from each of these two processes. The spectral density of $z_{1,t}$ is decreasing for low frequencies and has a peak around $\omega = 1.1$; the spectral density of $z_{2,t}$ is bimodal, with two peaks around $\omega = 1.1$ and $\omega = 2$. The estimated log-spectral densities for the two processes along with 95% posterior intervals are reported in Figure 2.2. Our model does well in capturing the shape of the log-

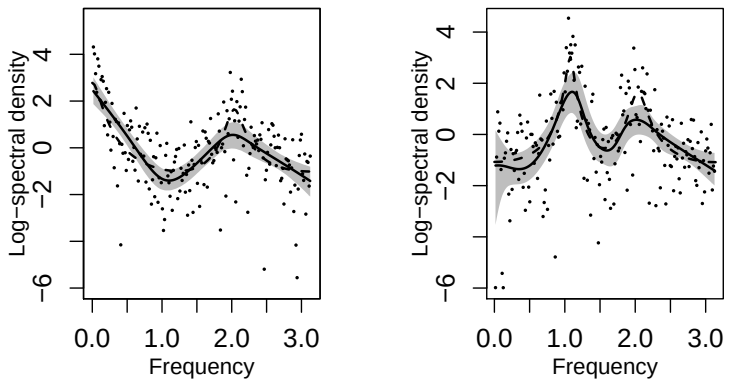

Figure 2.2: Synthetic data from sum of AR processes and white noise. Posterior mean (solid dark line), 95% credible interval (shaded area), true log-spectral density (dashed line) and log-periodogram (dots) for the sum of autoregressive processes and white noise, under the model with logistic weights.

spectral densities, and it successfully identifies the peaks. The corresponding posterior densities for λ are plotted in Figure 2.3. In both cases, the posterior distribution of λ is supported by smaller values than for the autoregressive processes of order one, in agreement with the fact that the spectral densities of $x_{1,t}$ and $x_{2,t}$ are smoother than those of $z_{1,t}$ and $z_{2,t}$.

2.5.2 Electroncephalogram data

We consider four times series that correspond to portions of electroncephalograms taken from a larger dataset. The original time series were recorded at 19 locations over the scalp of a patient who received electroconvulsive therapy. Further details and data analysis can be found in Krystal et al.

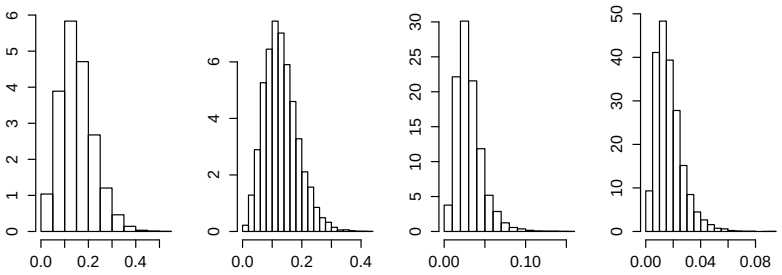

Figure 2.3: Synthetic data. Posterior density for λ for data simulated from (from the left) autoregressive processes of order one, with parameter 0.7 and -0.9, and for the sum of autoregressive processes and white noise, under the model with logistic weights.

(1999). The time series were recorded in four left channels, two of which are in the frontal region of the scalp (F7 and F3), one is in the temporal region (T5), and one is in the parietal region (P3). For each time series, we have 299 observations, obtained by subsampling the electroencephalogram signal every sixth observation from a mid-seizure section. The original sampling rate was 256 Hz. The priors were the as in Section 2.5.1, save for the prior for λ that here is given by a gamma distribution with shape parameter 3 and mean 0.1, which places almost all its mass in $(0, 0.5)$. We choose this prior because we expect at least one pronounced peak in the spectral density, reflecting brain activity in at least one frequency band, and we thus want to avoid oversmoothing. The number of components is fixed to $K = 10$. Also in this case, the results were robust with respect to the number of components.

Figure 2.7 shows the posterior mean estimates and 95% posterior credible intervals for the spectral densities together with the logged spectral peri-

odogram. The behavior of the spectral density is similar for the two frontal channels, shown in the first row of Figure 2.7, and for the two temporal ones, on the second raw. All the channels show a peak around 3.4 Hz. that is slightly shifted to the left in T5 and P3. These results are consistent with previous analyses that indicate that the observed quasi-periodicity is dominated by activity in the delta frequency range (1-5 Hz.). We also note that although there are important similarities across the spectral densities, each density has its own features with channels F3, T5 and P3 sharing more similarities and F7 being different from the rest.

2.6 Numerical illustration: Mixture model with local weights based on differences of CDFs

We evaluate the performance of our method by analyzing the same synthetic datasets we presented in Section 2.5.1. For the α_k and β_k, we use zero-mean normal priors with variances $\sigma_\alpha^2 = \sigma_\beta^2 = 100$. An inverse-gamma prior with shape parameter 3 and scale 6 is placed on σ^2. For ζ and ϕ we use two independent zero mean normal priors with variances $\sigma_\zeta^2 = \sigma_\phi^2 = 100$. The estimated log-spectral densities and the corresponding 95% posterior intervals are reported in Figures 2.5 and 2.6. For the autoregressive processes of order one, we show the results obtained with $K = 10$. In fact, the monotonic shapes

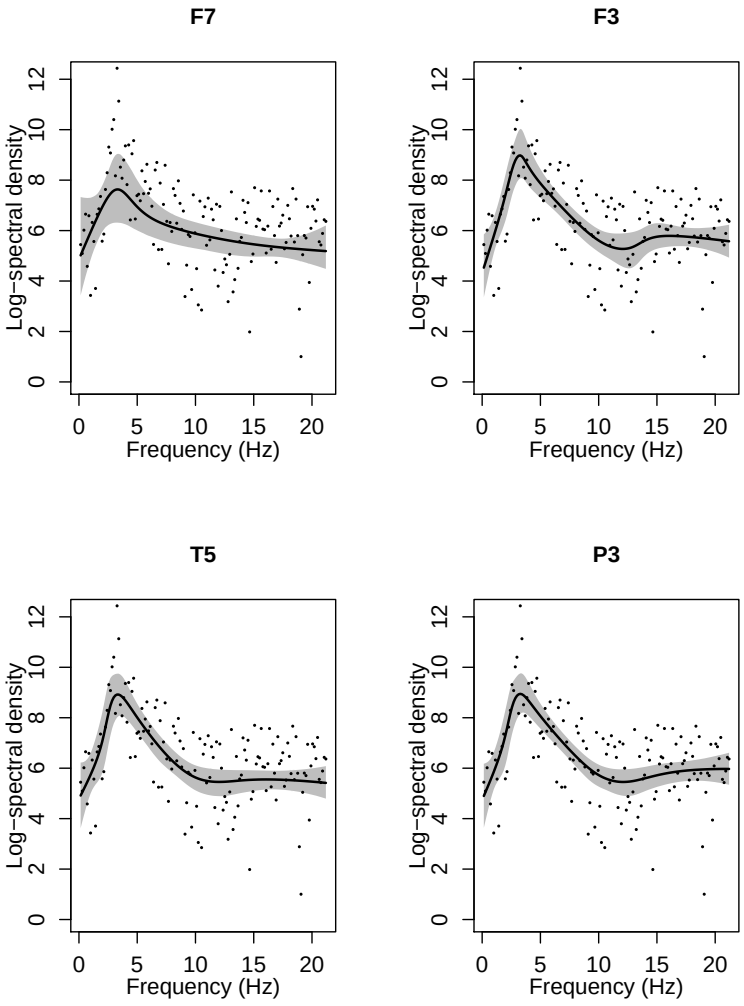

Figure 2.4: EEG data. Posterior mean (solid dark line), 95% credible interval (shaded area), and log-periodogram (dots) for the electroencephalogram datasets recorded in four channels, under the model with logistic weights.

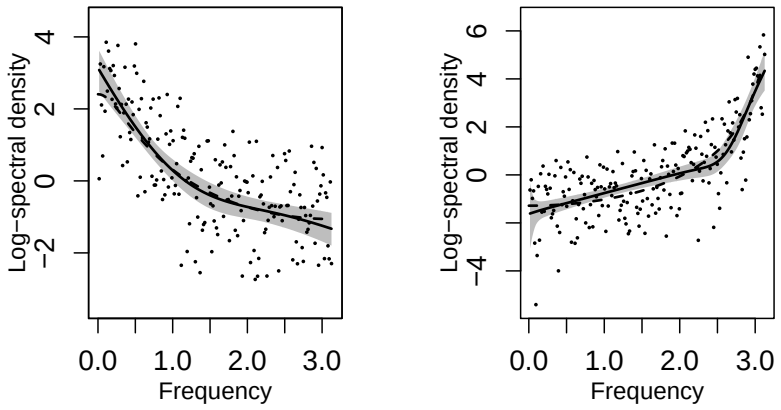

Figure 2.5: Synthetic data from a single AR(1) process. Posterior mean (solid dark line), 95% credible interval (shaded area), true log-spectral density (dashed line) and log-periodogram (dots) for data simulated from autoregressive processes of order one, with parameters 0.7 (left) and -0.9 (right), under the model with weights based on differences of CDFs.

do not require a larger number of components than those used in the case of logistic weights. For the sum of autoregressive processes, we need a larger number of components than in the case of logistic weights. This is due to the fact that we have only three weight parameters, hence we need more components to capture the bimodal shape of the spectral densities. Specifically, we show results obtained for $K = 50$. The interval bands in the case of the bimodal spectral densities are larger than the ones obtained with the logistic weights. This is due to the fact that a larger number of effective components are needed.

Finally, we use the model with the weights based on differences of CDFs

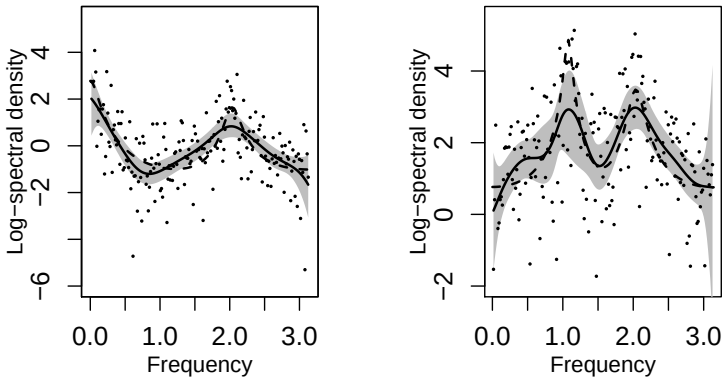

Figure 2.6: Synthetic data from sum of AR processes and white noise. Posterior mean (solid dark line), 95% credible interval (shaded area), true log-spectral density (dashed line) and log-periodogram (dots) for the sum of autoregressive processes and white noise, under the model with weights based on differences of CDFs.

to estimate the spectral densities of the EEG data. The dataset is the one analyzed in Section 2.5.2. Again, the number of components is fixed to $K = 50$. Also in this case, the results were robust with respect to the number of components. Figure 2.7 shows the posterior mean estimates and 95% posterior credible intervals for the spectral densities together with the logged spectral periodogram. The estimate obtained under the model with weights based on differences of CDFs appears to be smoother than the estimates obtained under the model with logistic weights. In fact, the number of effective weights, meaning weights that are not identical to zero on the support, is larger in this case. The weights based on differences of CDFs distributions create a

soft partition of the support with more intervals than those in the partition induced by the logistic weights.

2.7 Discussion

We have presented a Bayesian approach to single spectral density estimation that builds from an approximation to the Whittle log-likelihood through a mixture of normal distributions. We have obtained good estimates with simulated datasets. In the real data example, we have shown that the spectral densities corresponding to different channels in the same subject can present different characteristics. There is the need to develop a model that is easy to extend to multiple related time series. Note that the logistic weights are specified through a $(2K+1)$-dimensional vector $\boldsymbol{\xi}$. Therefore, the number of parameters for models that consider these weights increases linearly with the number of components K. Moreover, the denominator that arises from the structure implied by the logistic weights complicates posterior simulation. For the logistic weights, we need a data augmentation step, based on auxiliary Pólya-gamma variables (Polson et al., 2013), which requires N latent variables for each $k = 1, ..., K$, thus increasing considerably the computational cost even for a single spectral density. Alternatively, our proposed formulation for the mixture weights in (3.1) provides key computational advantages, as the weights are fully specified through three parameters for a single spectral

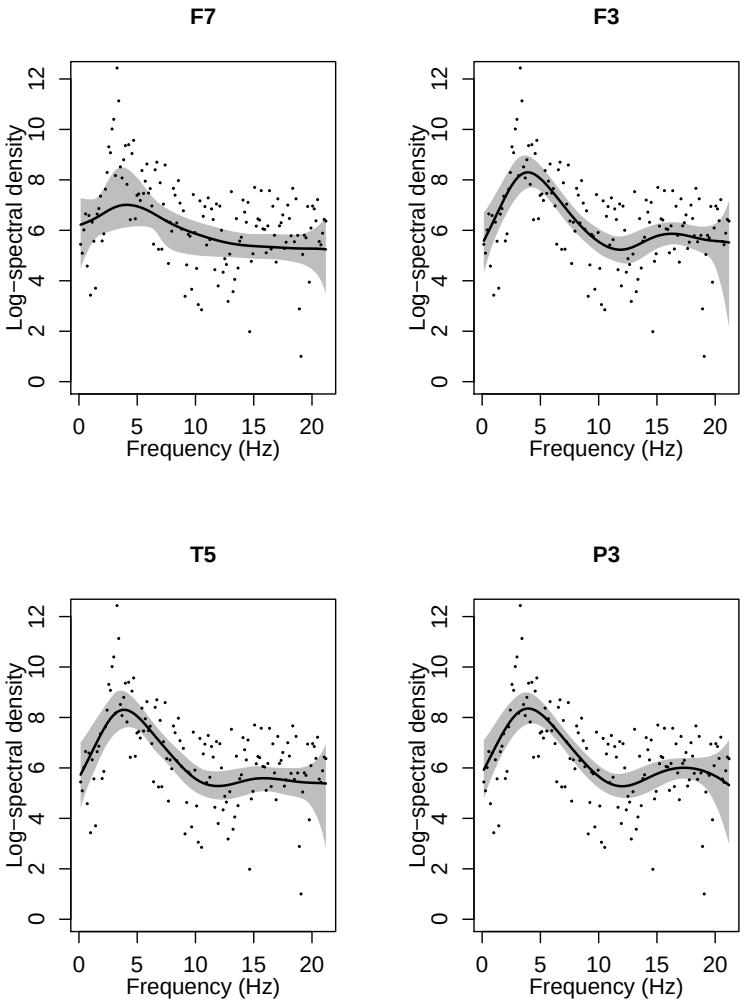

Figure 2.7: EEG data. Posterior mean (solid dark line), 95% credible interval (shaded area), and log-periodogram (dots) for the electroencephalogram datasets recorded in four channels, under the model with weights based on differences of CDFs.

density, leading to more efficient posterior simulation.

The new formulation of the weight is key for the extension of the model to multiple related time series. In fact, we can make the weight parameters series specific. We develop a Bayesian hierarchical model for spectral density estimation of multiple related time series in Chapter 3.

Chapter 3

Hierarchical model for multiple time series

3.1 Introduction

In this chapter, we extend the modeling approach for spectral density estimation, presented in Chapter 2, to multiple time series. Building from the construction for a single spectral density, we develop a hierarchical extension for multiple spectral densities. Specifically, we set the mean functions to be common to all spectral densities and make the weights specific to the time series through the parameters of the logit-normal distribution. In addition to accommodating flexible spectral density shapes, a practically important feature of the proposed formulation is that it allows for ready posterior simulation

through a Gibbs sampler with closed form full conditional distributions for all model parameters.

It is worth emphasizing that we are considering multiple – not multivariate – time series. For a description of methods for multivariate time series in the spectral domain, refer, for example, to Shumway and Stoffer (2011).

The outline of the Chapter is as follows. In Section 3.2, we describe the Bayesian hierarchical model for multiple time series. In Section 3.3, we present results from an extensive simulation study. In Section 3.4, we apply the proposed model to data from multichannel electroencephalographic recordings. Finally, Section 3.5 concludes with a summary and discussion of the advantages of our approach.

3.2 Bayesian Hierarchical Model

The model for a single time series, presented in Section 2.2, was developed with a hierarchical extension in mind. Let us consider M related time series, which, without loss of generality, are assumed to have the same number of observations n. For example, assume that M is the number of channels located over a subject's scalp for which we have electroencephalographic recordings. For each time series, we have N observations from the (translated) log-periodogram, which we denote as y_{mj}, where the first index indicates the time series ($m = 1, ..., M$) and the second indicates the Fourier frequency

$(j = 1, ..., N)$.

Now, for each m and each j we approximate the distribution of the y_{mj} with a smooth mixture of Gaussian distributions, as described in the previous section. We take the mean parameters of the Gaussian mixture components, that is, (α_k, β_k), for $k = 1, ..., K$, to be common among time series. This translates into a set of K linear basis functions for the log-spectral density approximation which are common to all time series. On the other hand, we let the parameters that specify the weights be time series specific, that is, we use the form in (3.1) with parameters $\boldsymbol{\xi}_m = (\zeta_m, \phi_m, \tau_m)$, for $m = 1, ..., M$. Specifically, the weights, for $k = 1, .., K$ are defined as follows:

$$g_k(\omega; \boldsymbol{\xi}_m) \equiv g_k(\mu_m(\omega), \tau_m) = \int_{\pi(k-1)/K}^{\pi k/K} f_Y(y \mid \mu_m(\omega), \tau_m)\, \mathrm{d}y, \tag{3.1}$$

where $f_Y(y \mid \mu_m(\omega), \tau_m)$ is the density of a logit-normal$(\mu_m(\omega), \tau_m)$ distribution on $(0, \pi)$. Here, $\mu_m(\omega) = \zeta_m + \phi_m \omega$.

For each time series, the weights select the linear functions to approximate the corresponding log-spectral density. We use M distinct smoothness parameters τ_m to allow different levels of smoothness across the spectral densities.

Hence, extending (2.2), the observation stage for the hierarchical model on the M time series can be written as

$$y_{mj} \mid \boldsymbol{\theta} \overset{ind.}{\sim} \sum_{k=1}^{K} g_k(\mu_m(\omega_j), \tau_m)\, \mathrm{N}(y_{mj} \mid \alpha_k + \beta_k \omega_j, \sigma^2), \tag{3.2}$$

where the k-th weight at the m-th location is defined as in (3.1) in terms of in-

crements of a logit-normal distribution function, with mean function, $\mu_m(\omega_j) = \zeta_m + \phi_m\omega_j$, and precision parameter, τ_m, that are time series specific. Again, $\boldsymbol{\theta}$ collects all model parameters: the intercept and slope parameters of the K mixture components means, $\boldsymbol{\alpha} = \{\alpha_k : k = 1, ..., K\}$ and $\boldsymbol{\beta} = \{\beta_k : k = 1, ..., K\}$, the common variance parameter σ^2, and the mixture weights parameters, $\boldsymbol{\zeta} = \{\zeta_m : m = 1, ..., M\}$, $\boldsymbol{\phi} = \{\phi_m : m = 1, ..., M\}$, and $\boldsymbol{\tau} = \{\tau_m : m = 1, ..., M\}$. Posterior simulation is implemented using the augmented version of the model based on MN normally distributed auxiliary variables, r_{mj}, for $m = 1, ..., M$ and $j = 1, ..., N$. In particular,

$$y_{mj} \mid r_{mj}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2 \overset{ind.}{\sim} \sum_{k=1}^{K} \mathrm{N}(y_{mj} \mid \alpha_k + \beta_k\omega_j, \sigma^2)\mathbb{I}\left\{\frac{k-1}{K} < \frac{\exp(r_{mj})}{1+\exp(r_{mj})} \leq \frac{k}{K}\right\}$$

$$r_{mj} \mid \zeta_m, \phi_m, \tau_m \overset{ind.}{\sim} \mathrm{N}(r_{mj} \mid \zeta_m + \phi_m\omega_j, 1/\tau_m).$$

Technical details on the Gibbs sampler used to implement the hierarchical model are given in Appendix B.

The full Bayesian model is completed with priors for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ^2, and a hierarchical prior for the (ζ_m, ϕ_m) and τ_m, for $m = 1, ..., M$. The weight parameters are assumed a priori independent of the Gaussian mixture component parameters. We assume $\sigma^2 \sim$ inv-gamma$(n_{\sigma^2}, d_{\sigma^2})$, that is, an inverse gamma prior (with mean $d_{\sigma^2}/(n_{\sigma^2} - 1)$, and $n_{\sigma^2} > 1$), $\alpha_k \sim \mathrm{N}(\mu_{0\alpha}, \sigma^2_\alpha)$, and

$\beta_k \sim \mathrm{N}(\mu_{0\beta}, \sigma^2_\beta)$, for $k = 1, ..., K$. The hierarchical prior is given by

$$\begin{aligned}(\zeta_m, \phi_m) \mid \boldsymbol{\mu}_w, \Sigma_w &\overset{ind.}{\sim} \mathrm{N}(\boldsymbol{\mu}_w, \Sigma_w), \quad m = 1, ..., M, \\ \tau_m \mid d_\tau &\overset{ind.}{\sim} \mathrm{gamma}(n_\tau, d_\tau), \quad m = 1, ..., M,\end{aligned}$$

where $\mathrm{gamma}(n, d)$ denotes the gamma distribution with mean n/d. To borrow strength across the time series, we place a bivariate normal prior on $\boldsymbol{\mu}_w$, and an inverse Wishart prior on the covariance matrix Σ_w. For the τ_m, we fix the shape parameter, n_τ, and place a gamma prior on the rate parameter, d_τ.

The prior on the intercept parameters, α_k, summarizes information about the spectral density value near $\omega = 0$, while the prior on the slope parameters, β_k, can be used to express beliefs about the shape of the spectral density. For instance, for multimodal spectral densities, we expect some selected β_k to be positive and some negative, whereas for unimodal spectral densities, we expect all the selected β_k to have the same sign. The parameters ζ_m and ϕ_m, for $m = 1, ..., M$, determine the location of the modes of the weights corresponding to the m-the spectral density, while the τ_m are smoothness parameters with smaller values favoring smoother spectral densities. Given the model structure that involves common parameters for the mixture components, inferences for the (ζ_m, ϕ_m) are useful in identifying groups of time series with similar spectral characteristics. This is demonstrated with the data illustrations of Sections 3.3 and 3.4.

In this work, the number of mixture components, K, is fixed. The modeling

approach can be generalized to a random K, albeit at the expense of a more computationally challenging posterior simulation method. The values of the weight parameters determine how many effective components are used by the model.

The time series specific weights will select a subset of components to approximate the specific spectral density. Hence, we expect that the necessary number of mixture components K will be significantly larger than in the univariate case when the underlying spectral densities of the multiple time series differ substantially. When the spectral characteristics do not vary among locations, the number of necessary mixture components K will not increase.

If we have prior knowledge on the smoothness of the log-spectral densities, we can use it in the specification of K. Based on extensive empirical investigation with several data sets, including the ones of Section 3.3, we have observed that, in general, a relatively small number of mixture components suffices to capture different spectral density shapes, with inference results being robust to the choice of K.

Since the linear bases are common to all the time series, whereas the weights are time series specific, we look at the posterior distribution of the weight parameters. This allows us to identify groups of spectral densities with similar characteristics. Notice that our model does not allow two spectral densities to be identical, hence we cannot do clustering of identical spectral densities.

3.3 Simulation study

In order to assess the performance of the proposed spectral models, we designed three different data generating mechanisms that represent three hypothetical scenarios involving multiple related time series. In each scenario, we have $M = 15$ time series. Moreover, we consider replicates, meaning that more than one time series is generated from the same underlying process. For each time series, we simulated $n = 300$ time points, leading to $N = 149$ observations from the log-periodogram. In addition to posterior estimates and credible intervals for the spectral densities, we investigate the posterior distribution of the weight parameters, ζ_m, ϕ_m and τ_m, for $m = 1, ..., M$, which can be useful in identifying similar spectral characteristics across multiple time series.

3.3.1 Total Variation Distance

To quantify the discrepancy between two spectral densities, we use the concept of total variation distance (TVD) for normalized spectral densities. The total variation is a distance measure for probability distributions and it has been used to quantify the distance between two spectral densities, after normalization (e.g., Euan et al., 2015). In particular, the total variation distance between two normalized spectral densities $f^*(\omega) = f(\omega)/\int_\Omega f(\omega)\mathrm{d}\omega$ and $g^*(\omega) = g(\omega)/\int_\Omega g(\omega)\mathrm{d}\omega$, where $\Omega = (0, \pi)$, is defined as $\mathrm{TVD}(f^*, g^*) =$

$1 - \int_\Omega \min\{f^*(\omega), g^*(\omega)\}\mathrm{d}\omega$. This is equivalent to half of the L_1 distance between f^* and g^*, i.e., $\mathrm{TVD}(f^*, g^*) = ||f^* - g^*||_1/2$. We use the TVD as a measure of discrepancy between spectral densities because it is symmetric and bounded between 0 and 1, with the value of 1 corresponding to the largest possible distance between the normalized spectral densities. Moreover, we have the following result. Let f_k be a sequence of functions and f be a function defined on $(0, \pi)$. Let $L_p(0, \pi)$ denote L_p convergence on $(0, \pi)$, and let TVD denote convergence in the total variation distance.

Lemma 3.1. *If* $\log f_k \overset{L_p(0,\pi)}{\longrightarrow} \log f$, *then* $f_k / \int_0^\pi f_k(\omega)\mathrm{d}\omega \overset{TVD}{\longrightarrow} f / \int_0^\pi f(\omega)\mathrm{d}\omega$.

The proof of the above lemma is in appendix A.

Under a Bayesian modeling approach we have a posterior distribution for the TVD of any two given spectral densities. We use the posterior distributions of the TVDs to compare the inferred spectral densities of multiple time series, as illustrated in the analyses of simulated and real data in Chapter 3.

3.3.2 First scenario

The goal of this simulated scenario is to evaluate the performance of our model for time series with monotonic spectral densities, and also to test if the model is able to recognize white noise. In order to compare our posterior estimates to the true spectral densities, we simulated data from processes with spectral densities available in analytical form. We considered three underlying

generating processes, with five replicates in each case, leading to a total of $M = 15$ time series. The first five time series were generated from an autoregressive process of order one, or AR(1) process, with parameter 0.9. The next five time series (labeled from 6-10) were generated from an AR(1) process with parameter 0.5. Finally, the last five time series were generated from pure white noise, or equivalently an AR(1) process with parameter 0. Hence, the underlying spectral densities for the first two groups are monotonic decreasing. The spectral density corresponding to the first five time series has a larger slope and is less noisy, while the one corresponding to the second group has smaller slope and more variability in the periodogram realizations. The spectral density for the last five time series is a constant at one, that corresponds to the variance of the white noise.

We fixed the number of mixture components to $K = 30$; similar results were obtained with a larger value of K. We assumed α_k and β_k to be independent normally distributed centered at zero with variance 1000 such that the linear basis can have a wide range of motion. For the common variance parameters, we used an inverse gamma prior with mean 3 and variance 9. For the smoothness parameters τ_m, $m = 1, ..., M$, we fixed the shape parameter to 30 and placed a gamma$(3, 20)$ on the rate parameter. This results in a marginal prior distribution for each τ_m that supports a large interval on the positive real line. Moreover, since each time series has its own smoothness parameter,

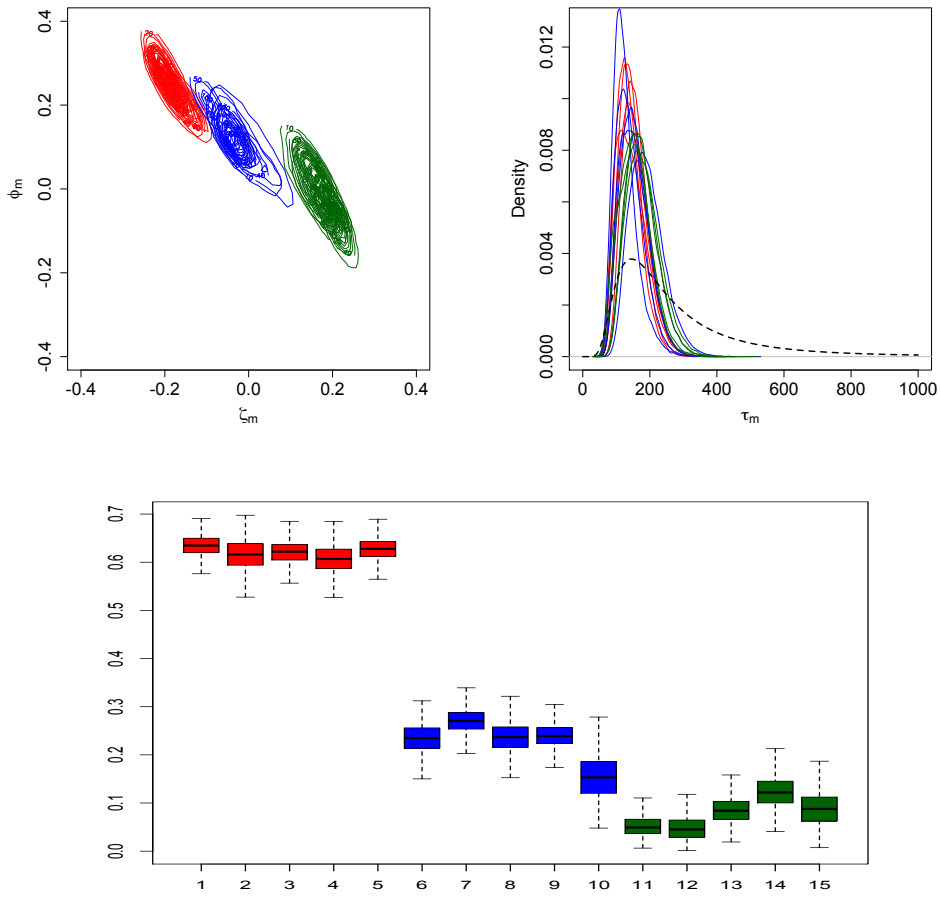

Figure 3.1: First simulation scenario. Joint posterior densities for (ζ_m, ϕ_m) (top left panel), marginal prior density (dashed line) and posterior densities for τ_m (top right panel), for $m = 1, ..., M$, and boxplots of posterior samples for the total variation distance of each estimated spectral density from the true white noise spectral density (bottom panel).

we can have different levels of smoothness for different spectral densities. The hyper-prior on the mean parameter was centered at 0 and had variance 10, while the Inverse Wishart distribution parameters were chosen in a way that the marginal distributions for the diagonal elements were inv-gamma$(3, 3)$, and the implied prior distribution on the correlation between ζ_m and ϕ_m was diffuse

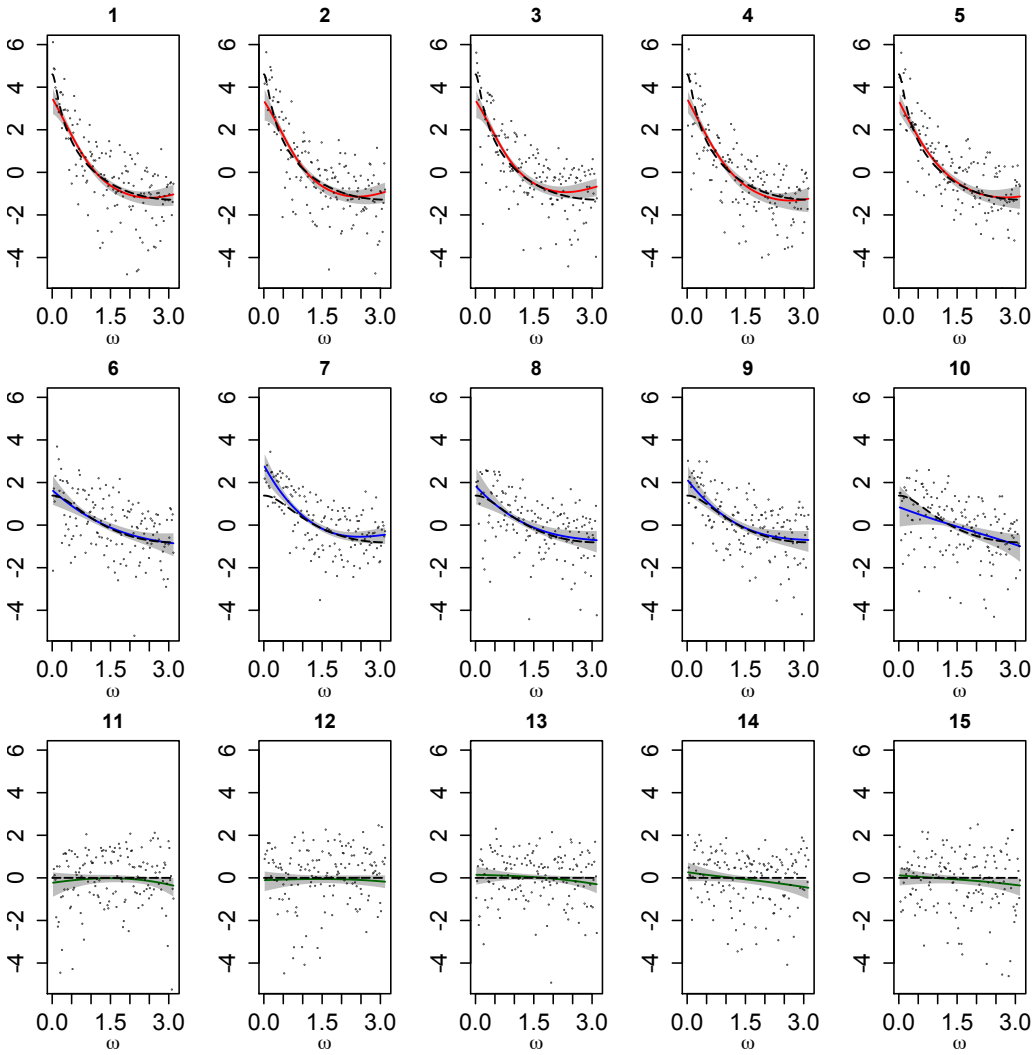

Figure 3.2: First simulation scenario. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for each log-spectral density. Each panel includes also the true log-spectral density (dashed line) and the log-periodogram (dots).

on $(0, 1)$.

Figure 3.1 shows the joint posterior densities for (ζ_m, ϕ_m) (top left panel) and the prior and posterior densities for τ_m (top right panel) for $m = 1, ...M$. The color red corresponds to the first five time series, the blue to the time series

from sixth to tenth, and the green one to the last five time series. Clearly, the joint posterior distribution of (ζ_m, ϕ_m) allows us to accurately identify the three groups. In addition, we notice that there is a pattern in the posterior distribution: the steeper the slope of the spectral density (i.e., the larger the AR coefficient), the larger the value of ζ_m/ϕ_m, which determines the shape of the posterior spectral density estimates. The posterior distributions of the τ_m parameters that determine the smoothness of the spectral densities do not show a clear distinction among the three groups. Figure 3.1 (bottom panel) shows the posterior distributions of the total variation distances with respect to the true white noise spectral density. As expected, the distances for the time series in the third group are the smallest. In addition, the TVD results support the grouping identified through the posterior distribution of the (ζ_m, ϕ_m).

Figure 3.2 shows the true log-spectral densities, as well as the corresponding posterior mean estimates and 95% credible intervals. The model adequately captures the different log-spectral density shapes and is successful in discerning noisy processes with corresponding monotonic spectral densities from pure white noise processes.

3.3.3 Second scenario

The first scenario dealt with monotonic spectral densities. Here, we test model performance in the case of multiple unimodal spectral densities. A uni-

modal spectral density shows a single major peak at a particular frequency. For example, processes with corresponding unimodal spectral densities are second order quasi-periodic autoregressive processes with one dominating frequency. We generated a set of $M = 15$ time series from two different AR(2) processes. The first 8 time series were simulated from an AR(2) process which characteristic polynomial has complex conjugate roots. The modulus of the roots is 0.95 and the argument $\omega = 2.07$, while the last 7 time series were simulated from an AR(2) process with the same modulus of 0.95 but with argument $\omega = 1.08$. Hence, the time series contain essentially the same amount of information (the modulus was 0.95 in both groups) and have a single quasi-periodic component, with dominating frequency $\omega = 2.07$ for the first group, and $\omega = 1.08$ for the second group.

We applied again the model with $K = 30$ components, and with the same prior specification used in the first scenario for all parameters, except for the hyperparameter that controls the smoothness of the estimates. Since we expect less smooth spectral densities than the first scenario, we fix the shape parameter of the gamma prior on τ_m to 60 for all m, and place a gamma$(10, 300)$ hyperprior on the rate parameter. This results in a marginal prior distribution for the τ_m that has support on relatively large values.

Figure 3.3 shows the joint posterior densities for (ζ_m, ϕ_m) (left panel) and the posterior densities for τ_m (right panel), together with the prior marginal

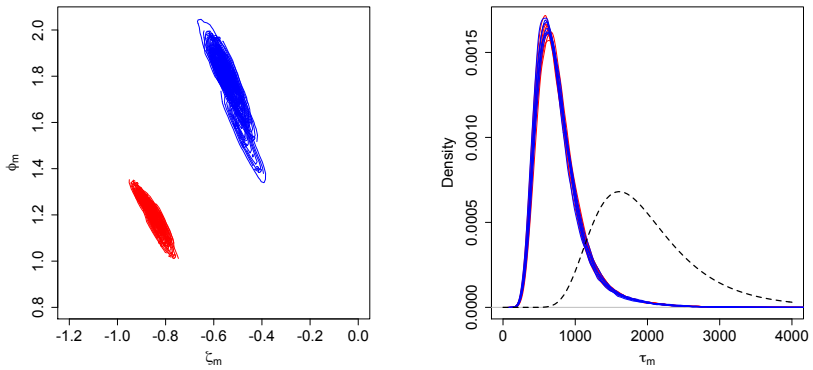

Figure 3.3: Second simulation scenario. Joint posterior densities for (ζ_m, ϕ_m) (left panel) and for τ_m (right panel), for $m = 1, ..., M$. The right panel includes also the marginal prior density (dashed line) for the τ_m.

density for τ_m, for $m = 1, ..., M$. The color red corresponds to the first eight time series and the blue to the last seven time series. Since parameters (ζ_m, ϕ_m) determine the location of the peak for each time series, the posterior densities of (ζ_m, ϕ_m) show a clear separation of the parameters relative to the two groups. The posterior densities of the τ_m parameters are similar for all the time series, as expected, since the peak has the same amplitude. Figure 3.4 shows the posterior mean estimates and 95% credible intervals for the log-spectral densities. The log-periodograms and true log-spectral densities are also shown. Our model adequately captures the distinct log-spectral density shapes, and successfully identifies the peaks of the quasi-periodic components for the two types of processes.

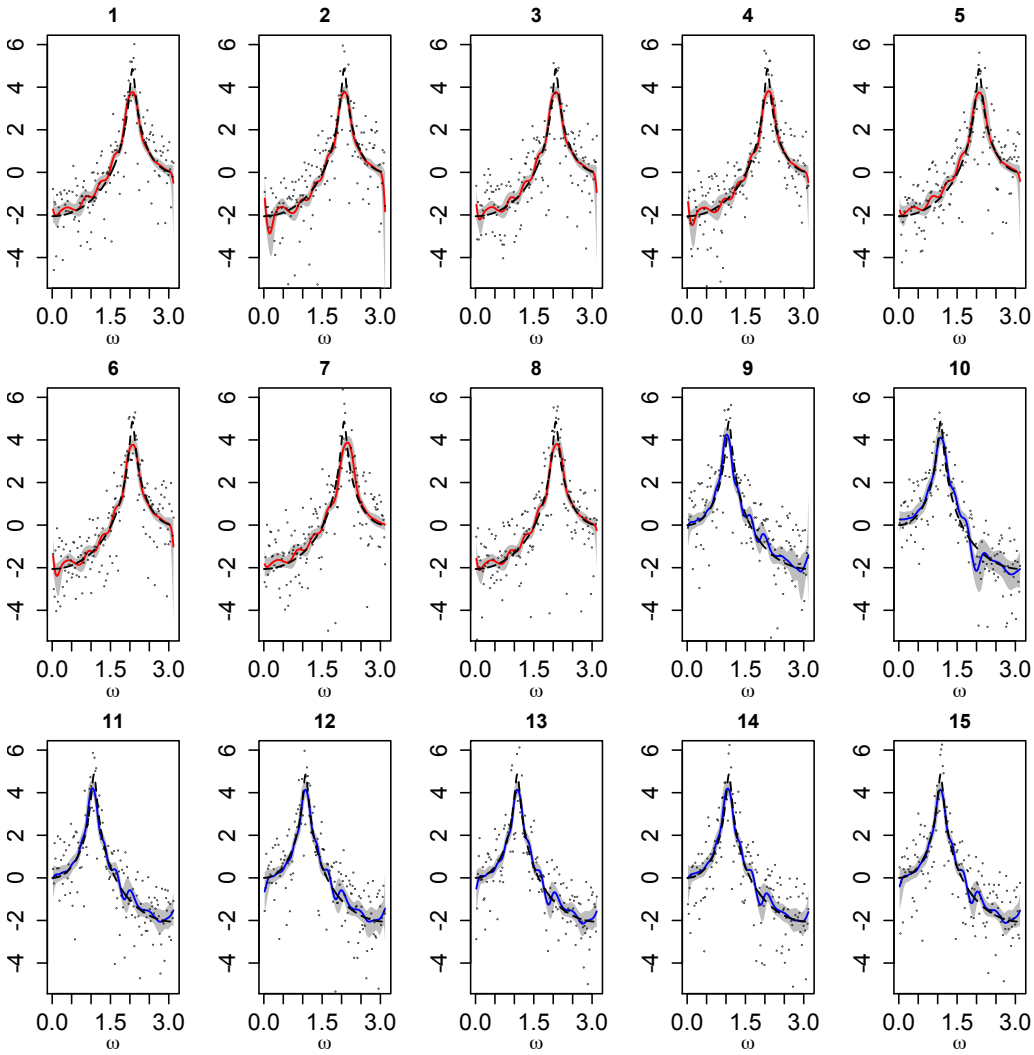

Figure 3.4: Second simulation scenario. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for each log-spectral density. Each panel includes also the true log-spectral density (dashed line) and the log-periodogram (dots).

3.3.4 Third scenario

In this scenario, all $M = 15$ simulated time series share an underlying first order autoregressive component, and some of them present an additional second order autoregressive component. Specifically, the first five time series

were simulated from an AR(1) with parameter 0.9. The next five time series were simulated from a sum of two autoregressive processes, an AR(1) and an AR(2). The AR(1) process has parameter 0.9, as in the previous set of time series, while the AR(2) process was assumed to be quasi-periodic, with modulus 0.83 and argument $\omega = 1.54$. The last 5 time series were again simulated from a sum of an AR(1) process and an AR(2) process. The AR(1) process has parameter 0.9 as before, whereas the AR(2) was a quasi-periodic process with modulus 0.97 and argument $\omega = 1.54$. In the second and third groups, the spectral densities show an initial decreasing shape, and a peak corresponding to the argument $\omega = 1.54$. While the argument is the same, the modulus is larger in the third group, hence the peak is more pronounced.

We applied the model with $K = 30$ mixture components, using the same prior specification with the second scenario, because we expected similar smoothness for the spectral densities. Figure 3.5 shows the joint posterior densities for (ζ_m, ϕ_m) (top left panel) and the posterior densities for τ_m (top right panel), for $m = 1, ...M$. The color red identifies the first five time series, the blue the time series from sixth to tenth, and the green the last five time series. The posterior distributions for (ζ_m, ϕ_m) cluster into two groups, the time series corresponding to the AR(1) process, and the time series corresponding to the sum of AR(1) and AR(2) processes. However, as expected, it is hard to differentiate between the two groups of time series generated from the sum of AR(1)

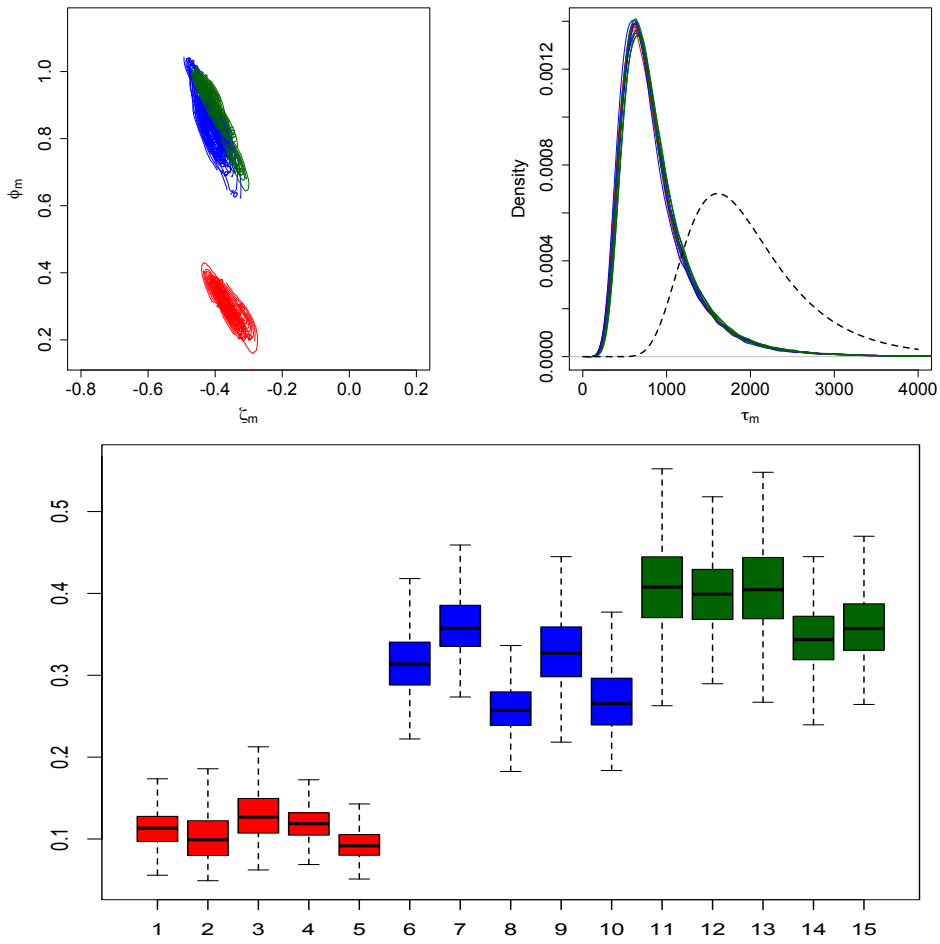

Figure 3.5: Third simulation scenario. Joint posterior densities for (ζ_m, ϕ_m) (top left panel), marginal prior density (dashed line) and posterior densities for τ_m (top right panel), for $m = 1, ..., M$, and boxplots of posterior samples for the total variation distance of each estimated spectral density from the AR(1) spectral density (bottom panel).

and AR(2) processes, because they share the same periodicities, with only the moduli being different. The boxplots in Figure 3.5 summarize the posterior distributions of the total variation distances between the estimates and the spectral density of an AR(1) model with parameter 0.9, which corresponds to the true spectral density for the first set of five time series. As expected,

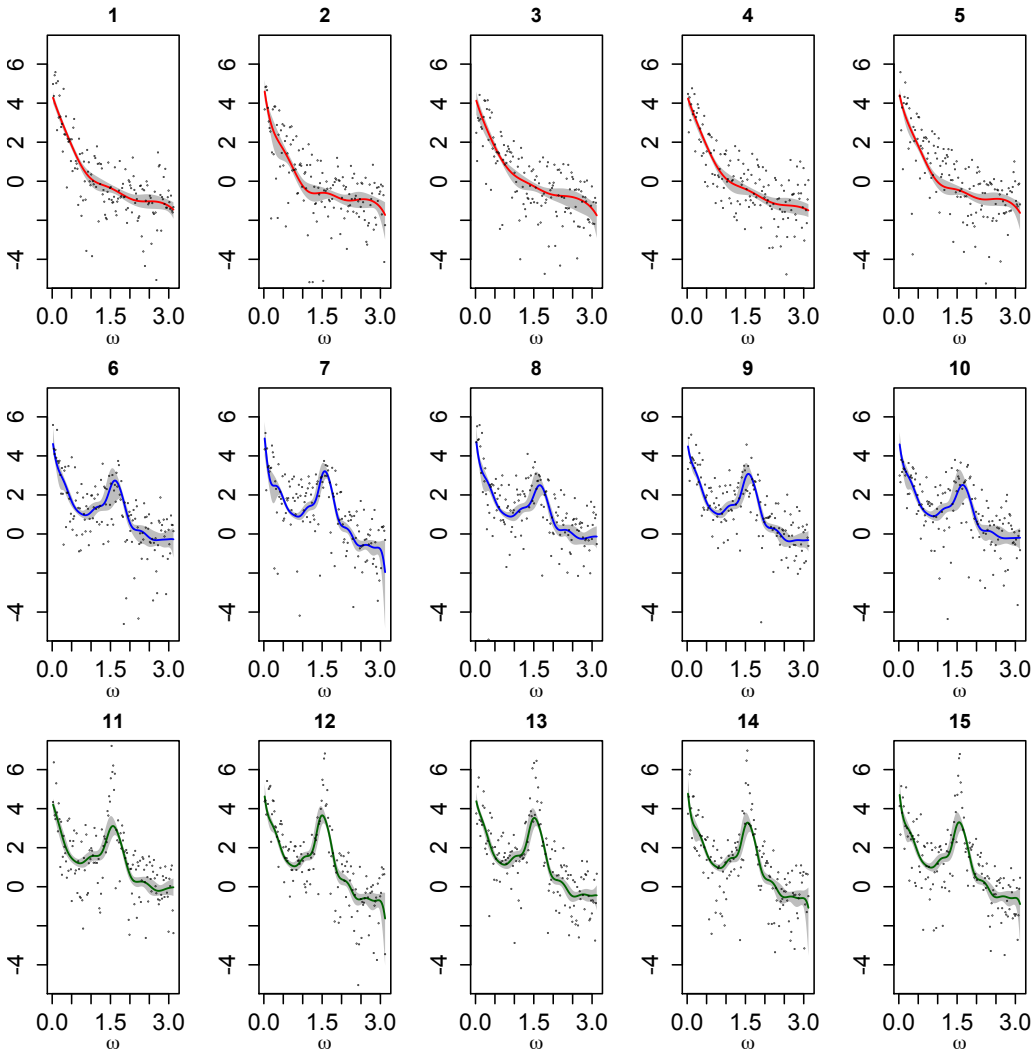

Figure 3.6: Third simulation scenario. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for each log-spectral density. Each panel includes also the log-periodogram (dots).

the posterior distribution of the total variation distance for the first five time series is concentrated around smaller values. Also as expected, there is no clear distinction between the second and the third group. Figure 3.6 displays the posterior mean estimates and 95% credible intervals for the log-spectral densities. As with the previous simulation examples, the model successfully

recovers the different spectral density shapes, and identifies the peak of the quasi-periodic component for the last ten time series.

3.4 Application: Electroencephalogram data

Multichannel electroencephalographic recordings (EEGs) arise from simultaneous measurements of electrical fluctuations induced by neuronal activity in the brain, using electrodes placed at multiple sites on a subject's scalp. One application area in which electroencephalographic recordings have proved very useful is the study of brain seizures induced by electroconvulsive therapy (ECT) as a treatment for major depression. The time series studied here are part of a more extensive study. Further details and data analyses can be found in West et al. (1999) and Krystal et al. (1999). EEGs were recorded at 19 locations over the scalp of one subject that received electroconvulsive therapy. The original sampling rate was 256 Hz.

3.4.1 ECT data 1

We consider first 300 observations from a mid-seizure portion, after subsampling the electroencephalogram signals every sixth observation. We refer to this dataset as ECT data 1.

We applied our model to these 19 time series, using $K = 50$ mixture com-

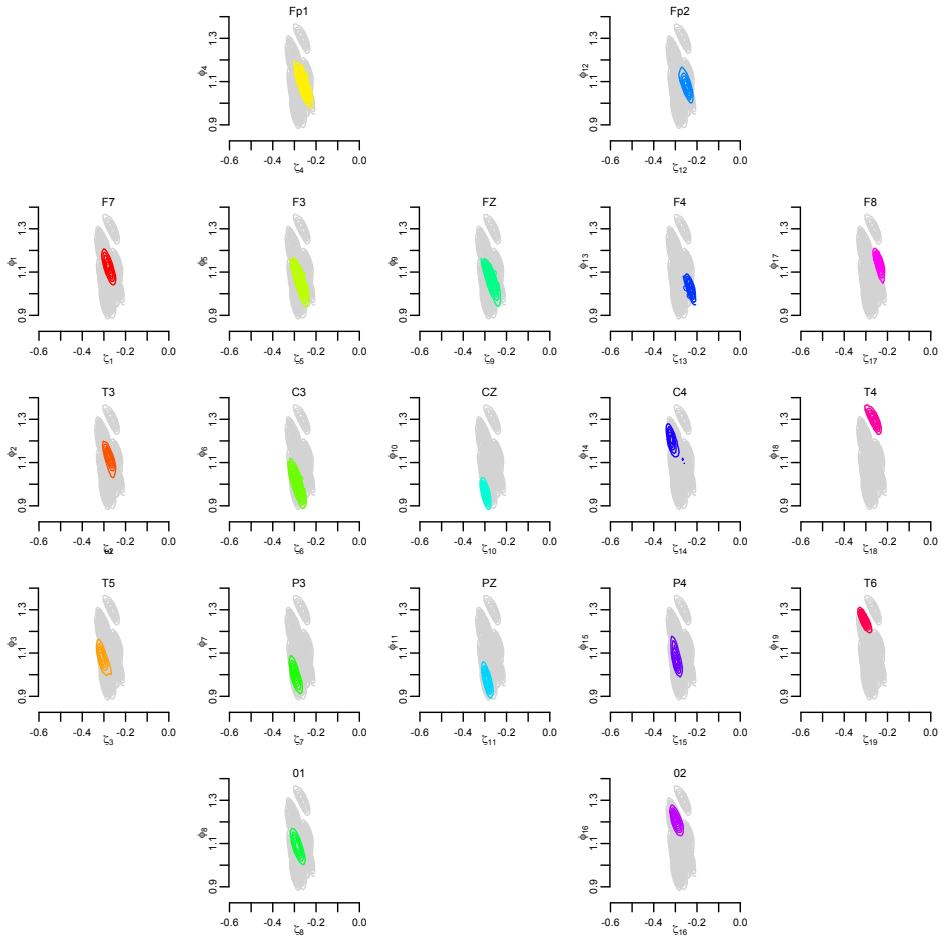

Figure 3.7: ECT data 1. Joint posterior densities for (ζ_m, ϕ_m), $m = 1, ..., 19$.

ponents. Similar results were obtained using a larger number of components. The priors on the parameters were defined as in the second and third simulated scenarios above. Figure 3.7 shows the joint posterior densities for (ζ_m, ϕ_m), for the 19 channels. The configuration of the plots shown in the figure aims to provide a schematic representation of the physical location of the electrodes over the subject's scalp. For example, the first row of the plots represent the front-

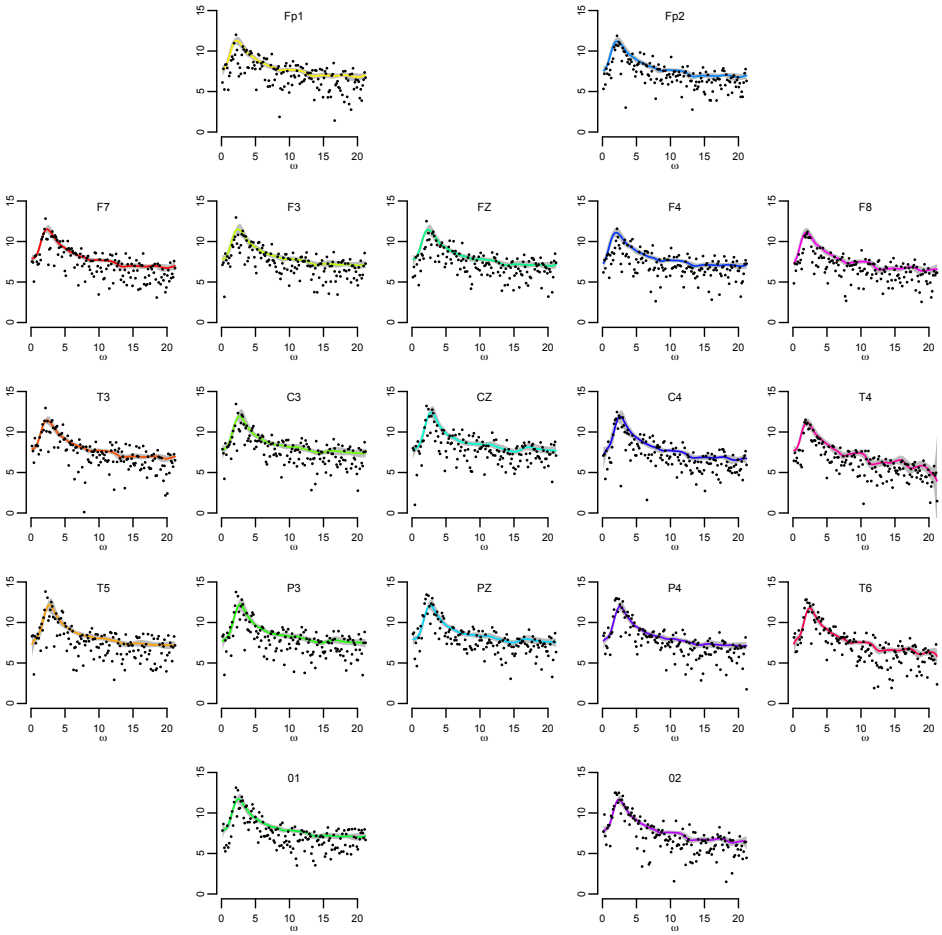

Figure 3.8: ECT data 1. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral densities corresponding to the 19 channels. Each panel includes also the log-periodogram (dots) from the specific channel.

most electrodes on the patient's scalp (F_{p_1} and F_{p_2}) viewed from above. All the posterior distributions are plotted in grey for reference, and the relevant one is in color. Overall, there is no clear distinction of the posterior distributions among the various channels. However, in certain regions of the brain the posterior distributions of the (ζ_m, ϕ_m) are concentrated around values similar to the those obtained from locations in that same region (e.g., channels C_z, P_z, P_3

and C_3). On the other hand, some channels that are next to each other show differences in their posterior distributions (for example C_z and C_4). Figure 3.8 shows the posterior mean estimates and the corresponding 95% posterior credible intervals for the spectral densities along with the log-periodograms. All the channels show a peak around 3.3-3.5 Hz for these series taken from the central portion of the EEG signals. These results are consistent with previous analyses which indicate that the observed quasi-periodicity is dominated by activity in the delta frequency range, that is, in the range from 1 to 5 Hz (West et al., 1999; Prado et al., 2001). The peak is slightly shifted to the left in the temporal channels with respect to the frontal channels. This aspect is also consistent with previous analyses. To quantify the differences among spectral densities, we chose to compare each density to the one in the central channel, C_z, as this channel has been used as a reference channel in previous anayses (Prado et al., 2001). Figure 3.9 shows the posterior distributions of the total variation distances between the spectral density estimates at each channel and that for the reference channel C_z. We can clearly see a correspondence between the posterior distribution of the weight parameters and the spectral density estimates. Figures 3.7 and 3.9 suggest that channels P_3, P_z, C_3, are the ones that share the most similar spectral features with channel C_z.

Here, we consider the total variation distance between the estimated spectral densities for each channel and the one for channel C_z. In order to do a

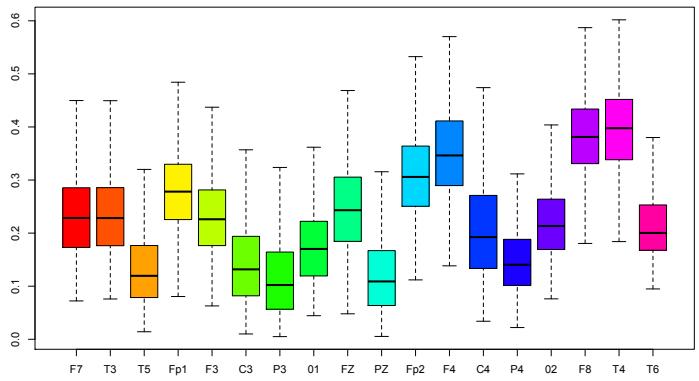

Figure 3.9: ECT data 1. Boxplots of posterior samples for the total variation distances between the spectral densities for each channel and the spectral density of the reference channel C_z.

formal clustering of the spectral densities using the total variation distance, we would need to cluster the spectral densities using a clustering algorithm, like the one in Euan et al. (2015).

3.4.2 ECT data 2

The analysis above shows that, although there are some differences across the time series recorded at different locations for the same time period, all the locations share similar features with respect to the location of the peak in their estimated log-spectral densities. We now show that our method can effectively capture differences in the spectral content of EEG time series that were recorded during different time periods over the course of the ECT induced seizure. To this end, we use the same dataset described above, but analyze

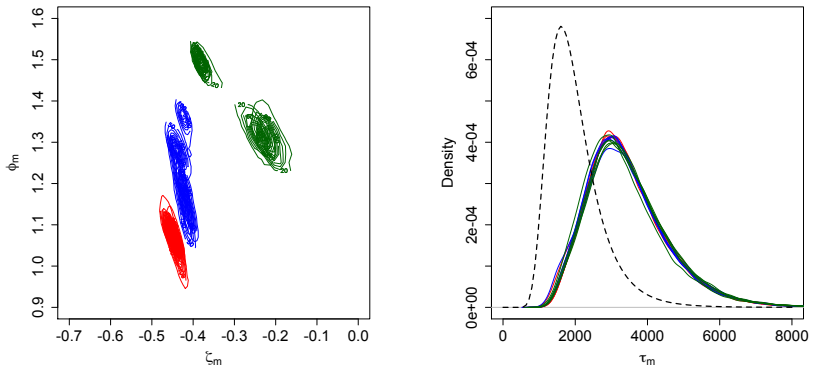

Figure 3.10: ECT data 2. Joint posterior densities for (ζ_m, ϕ_m) (left panel) and τ_m (right panel), for $m = 1, ..., 15$. The right panel includes also the marginal prior density (dashed line) for the τ_m.

time series recorded only in 5 channels, specifically, channels C_3, F_z, C_z, P_z and C_4, at 3 different temporal intervals (we refer to this dataset as ECT data 2). The first temporal interval corresponds to the beginning of the seizure, the second one is the interval considered in the previous analysis which corresponds to a mid-seizure period, while the third one was recorded later in time, when the seizure was fading. We emphasize that this is only an illustrative example to study if our method is able to capture different spectral characteristics in multiple EEGs. This is not the ideal model for this more general data structure, as we are not taking into account the fact that we have three different time periods. We analyze the 15 EEGs corresponding to 5 channels for three different time periods, using the model with $K = 50$ mixture components and the same prior specification described above. Figure 3.10 shows the joint posterior densities for (ζ_m, ϕ_m) and τ_m, for the 15 time series. The 5 series in

the first time period (plotted in red color) are essentially indistinguishable in terms of the distributions of (ζ_m, ϕ_m), while the series that correspond to mid (blue color) and later (green color) portions of the induced seizure display more variability. Figure 3.11 shows the posterior mean estimates of the log-spectral densities and the corresponding 95% posterior credible intervals along with the log-periodograms. In this case, there is a clear distinction in the posterior distributions of the time series corresponding to different time periods. In fact, the peak in the log-spectral density is more pronounced for those series that correspond to the beginning of the seizure. The peak shifts to the left and its power decreases in the successive time periods. In particular, in the last time period, the power of the peaks is the lowest and the variability in the log-periodogram observations and the estimated log-spectral densities is larger. There is also an increase of spectral variability over the time periods. This example shows that the frequency content differs among the different time intervals and suggests that the stationarity assumption is not satisfied for the entire time series. Our findings are consistent with previous analyses of these data, using non-stationary time-varying AR models (West et al., 1999; Prado et al., 2001).

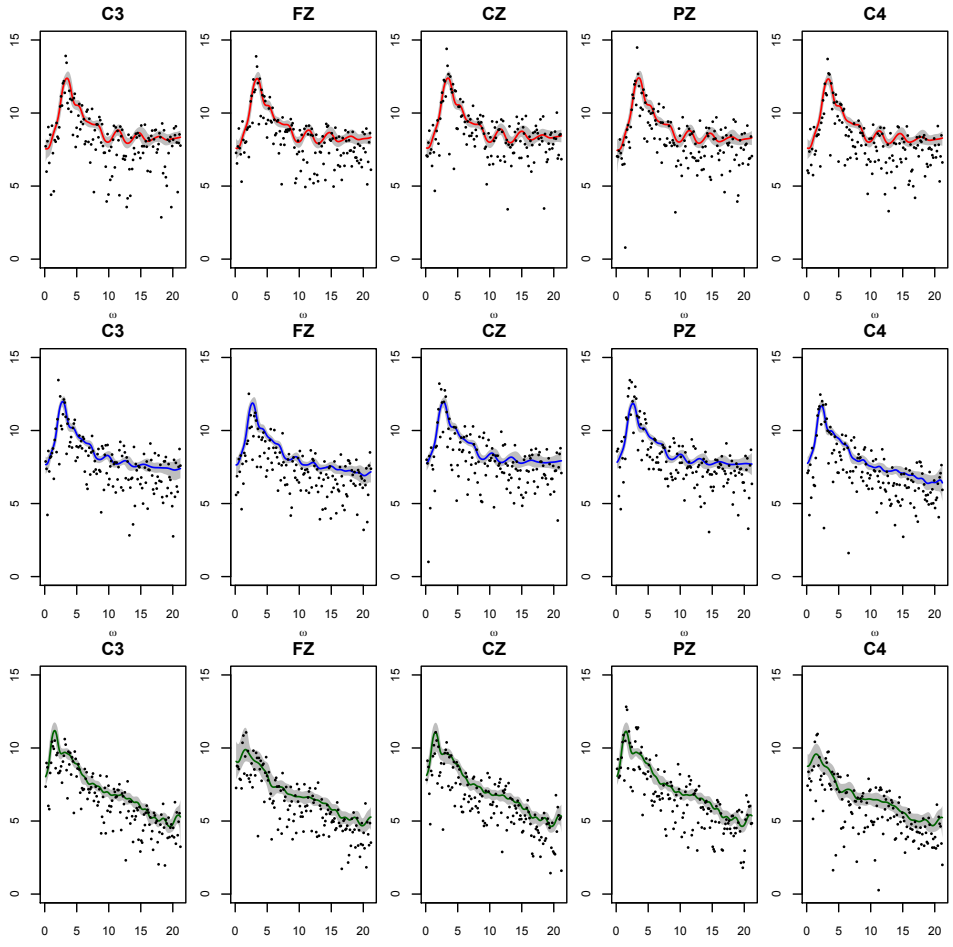

Figure 3.11: ECT data 2. Log-periodograms (dots), posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral densities corresponding to the 15 time series obtained from 5 channels for 3 time periods: beginning of the seizure (top row), mid-seizure (middle row), and end of the seizure (bottom row).

3.5 Discussion

In this Chapter, we have developed a new methodology for the analysis and estimation of multiple time series in the spectral domain. We note again that the methodology is developed for multiple, not multivariate, time series. To our knowledge, there are no Bayesian methods in the literature that deal jointly

and efficiently with multiple time series in the spectral domain. Methods for multivariate time series analysis are available, but often have the drawback of high computational cost, and are applicable in practice to a limited number of dimensions (rarely higher that 2-3). Our approach is based on modeling the distribution of the log-periodogram through a mixture of Gaussian distributions with frequency-dependent weights and mean functions, which implies a flexible mixture model for the corresponding log-spectral density. The main idea for a single unidimensional time series was presented in Chapter 2. Here, the mixture weights are built through differences of a distribution function, resulting in a substantially more parsimonious specification than logistic mixture weights. This is a fundamental feature of the proposed model, as it naturally leads to a hierarchical extension that allows us to efficiently consider multiple time series and borrow strength across them. As an additional advantage, casting the spectral density estimation problem in a mixture modeling framework allows for relatively straightforward implementation of a Gibbs sampler for inference.

The proposed modeling approach is parsimonious without sacrificing flexibility. Through simulation studies, we have demonstrated the ability of the model to uncover both monotonic and multimodal spectral density shapes, as well as white noise. We also applied the methodology to multichannel electroencephalographic recordings, obtaining results that are in agreement with

neuroscientists' understanding. As the last ECT example shows, the frequency content is different in the different time intervals. Ideally we would like to have a model that allows us to infer time-varying spectral characteristics in multiple time series. Classical spectral analysis is based on the assumption of weak stationarity, that is, the observations are assumed to come from a process with constant mean and covariance function that depends only on the lag between observations. The stationarity assumption is often not satisfied, especially when we need to analyze long time series, and the covariance properties vary over time. This is equivalent to say that the distribution of power over frequency changes as time evolves. Chapter 4 will focus on expanding the methodology in such a way that the evolution of the spectral content over time can also be included, with the goal of estimating time-varying spectral densities.

Chapter 4

Spectral density estimation for multiple non-stationary time series

4.1 Introduction

The classical spectral analysis is based on the assumption of (weak) stationarity, that is, the generating process is assumed to have mean and autocovariance function that do not vary with respect to time. The stationarity assumption is often valid only as an approximation to the non-stationary nature of the process. Moreover, this assumption is not reasonable in many cases, especially when we look at long time series, coming from a variety of

fields such as geophysics, econometrics, speech recognition and neuroscience. For this reason, spectral analysis is often done on short time intervals, for which the assumption of stationarity is acceptable. For example, electroencephalographic signals (EEGs) oscillate with greater amplitude and at higher frequency during an epileptic seizure than before or after such seizure (Ombao et al., 2001). In Chapter 3, we have analyzed three different temporal intervals in EEGs recorded at five channels for a patient subject to electroconvulsive therapy. The first temporal interval corresponded to EEGs recorded during the beginning of the electrically induced seizure, the second one to a mid-seizure period, while the third one was recorded later in time, when the seizure was fading. We emphasized that the spectral densities were different, with the peak becoming lower and shifting to the left with time. This is consistent with results from previous analyses of the same data (e.g., West et al. (1999); Prado et al. (2001)).

When the stationarity assumption is valid, the spectral density is interpreted as the the distribution of the energy of the process over the frequency range. When the stationary assumption is not satisfied, the distribution of power over frequencies changes as time evolves. The key idea for extending the spectral analysis to non-stationary time series is to define a spectral density that evolves over time, while keeping a similar interpretation locally as the spectral density in the stationary case.

In this chapter, we extend the methods developed in Chapters 2 and 3 to modeling a single and multiple non-stationary time series, respectively. In Section 4.2, we introduce the concept of evolutionary spectrum as developed in Priestley (1965) and the extension to the Whittle likelihood introduced by Dahlhaus (1997), and we present an overview of applied papers that used these theoretical results. In Section 4.4, we present the extension of our model to non-stationary time series, also based on the Dahlhaus's generalization of the Whittle likelihood, and we show results for one simulated dataset. The approach is extended to multiple non-stationary time series in Section 4.4, and results are shown for synthetic and real datasets. Specifically, we focus on the multichannel EEG dataset analyzed in Chapter 3.

4.2 Background

During the past two to three decades, the focus has been on modeling non-stationary time series. Time domain approaches consider models with parameters that vary over time. Common approaches include piecewise autoregressive models and time-varying autoregressive moving average models. For a review of these methods, see Yang et al. (2016). Here, we focus on frequency domain approaches.

Priestley (1965) introduces the idea of evolutionary spectrum. He extends the spectral representation in (1.1) to oscillatory processes, that is, processes

that evolve smoothly over time. Let us first define an oscillatory function as $\phi_t(\omega) = A_t(\omega)\exp(i\theta(\omega)t)$, where $A_t(\omega) = \int_{-\infty}^{\infty}\exp\{itu\}\mathrm{d}K_\omega(u)$ and $|\mathrm{d}K_\omega(u)|$ has a maximum in $u = 0$. Then, an oscillatory process is defined as $X(t) = \int_{-\pi}^{\pi}\phi_t(\omega)\mathrm{d}Z(\omega)$ where $Z(\omega)$ is a process whose increments are uncorrelated, with $\mathbb{E}[|\mathrm{d}Z(\omega)|^2] = \mathrm{d}\mu(\omega)$, and μ is a measure that is absolute continuous with respect to Lebesgue measure. The evolutionary spectrum is defined as $\mathrm{d}H(\omega) = |A_t(\omega)|^2\mathrm{d}\mu(\omega)$. When $X(t)$ is stationary and $\theta(\omega) = \omega$, then $\mathrm{d}H(\omega)$ reduces to the spectral density for a stationary process.

Another step in the development of a spectral theory for non-stationary processes was made by Dahlhaus (1997), which introduced the concept of locally stationary processes. Locally stationary processes have the property that the spectral characteristics change slowly over time, and they can be locally approximated with stationary processes. We denote a locally stationary time series of length T as $\{X_{1,T}, X_{2,T}, ..., X_{T,T}\}$, where the double index indicates the observation and the total number of observations. Notice that increasing the number of observations, T, implies that we allow for more local information in the sense of infill asymptotics, rather than in the sense of increasing domain asymptotics (i.e., information about the future). We introduce the rescaled time $u = t/T \in (0, 1)$.

A zero-mean process $X_{t,T}$ is called locally stationary with transfer function

$A^0_{t,T}(\omega)$ if there exist a representation

$$X_{t,T} = \int_{-\pi}^{\pi} A^0_{t,T}(\omega) \exp(it\omega) \mathrm{d}Z(\omega),$$

for $t = 1, ..., T$, where

(i) $Z(\omega)$ is process on $[-\pi, \pi]$ with $\overline{Z(\omega)} = Z(-\omega)$ and the cumulant of k-th order $\mathrm{cum}\{dZ(\omega_1), .., dZ(\omega_k).\} = \eta(\sum_{j=1}^{k})v_k(\omega_1, ..., \omega_{k-1})d\omega_1...d\omega_k$. Here, $v_1 = 0$, $v_2 = 1$, $|v_k(\omega_1, ..., \omega_{k-1})| \leq c_k$, where c_k a constant that depends on k, and $\eta(\omega) = \sum_{j=-\infty}^{\infty}$, where $\delta(\omega + 2\pi j)$ is the periodic 2π extension of the Dirac delta function.

(ii) There exist a constant $L > 0$ and a 2π-periodic smooth function $A : (0, 1] \times \mathbb{R} \to \mathbb{C}$ with $A(u, -\omega) = A(u, \omega)$ such that $\sup_{t,\omega} |A^0_{t,T}(\omega) - A(t/T, \omega)| \leq L/T$, for all T. $A(u, t)$ is assumed to be continuous in u.

The idea is essentially that for each fixed T, we have a local interval of stationarity about each time point. Moreover, the smoothness of A as a function of u controls the change of $A^0_{t,T}(\omega)$, as a sequence in t, such that it is allowed to change only smoothly over time. The time-varying spectrum of the Dahlhaus' locally stationary process, at time $u = t/T \in (0, 1)$ and frequency $\omega \in (-\pi, \pi)$ is $f(u, \omega) = |A(u, \omega)|^2$. Dahlhaus (1997) provides various examples of processes in this category. In particular, it is worth mentioning that autoregressive moving average processes with time varying parameters are locally stationary processes. For a time-varying autoregressive process, we have

that $A^0_{t,T}(\omega) \approx A(t/T, \omega)$; for a time-varying moving average process, we have that $A^0_{t,T}(\omega) = A(t/T, \omega)$.

The estimation theory of locally stationary processes, introduced by Dahlhaus (1997), is mostly based on the generalization of the Whittle likelihood function for stationary processes. The periodogram is replaced by local periodograms over possibly overlapping stationary data segments. Let n be the number of observations in each segment. Let $h : \mathbb{R} \to \mathbb{R}$ denote a data taper, with $h(x) = 0$ for $x \notin [0, 1)$. Then, for n even, define

$$d_n(u, \omega) = \sum_{s=0}^{n-1} h(s/n) X_{\lfloor uT \rfloor - n/2 + s + 1, T} \exp(-i\omega s)$$

$$H_{k,n}(\omega) = \sum_{s=0}^{n-1} h(s/n)^k \exp(-i\omega s)$$

$$I_n(u, \omega) = \frac{1}{2\pi H_{2,n}(0)} |d_n(u, \omega)|^2$$

and therefore

$$I_n(u, \omega) = \frac{|\sum_{s=0}^{n-1} h(s/n) X_{\lfloor uT \rfloor + n/Q + s + 1, T} \exp(-i\omega s)|^2}{2\pi \sum_{s=0}^{n-1} h(s/n)^2}.$$

$I_n(u, \omega)$ is the local periodogram of a segment of length n and midpoint $\lfloor uT \rfloor$. Q is the shift from segment to segment, where $Q = n$ means non overlapping segments. We calculate the periodogram for segments with midpoints $\bar{t}_i = Q(i-1) + n/2$, for $i = 1, ..., S$. Here S is the number of segments (and midpoints). The length of the time series is $T = Q(S-1)+n$. We can calculate the periodogram for each segment and model the periodogram observations at

the Fourier frequencies through the Whittle likelihood on each segment. This approximation of the likelihood is called blocked Whittle likelihood and is given by

$$\prod_{i=1}^{S} \prod_{j=1}^{\lfloor n/2 \rfloor - 1} \frac{1}{f(u_i, \omega_j)} \exp\left\{ -\frac{I_n(u_i, \omega_j)}{f(u_i, \omega_j)} \right\} \tag{4.1}$$

where $\bar{u}_i = \bar{t}_i / T$ and $f(\bar{u}_i, \omega_j)$ is the spectrum for segment i, at frequency ω_j.

Local stationarity plays a key role among the methodologies for analyzing non-stationary data. Ombao et al. (2001, 2002, 2005) used Dahlhaus's definition of local stationarity concept to develop the Smooth Localized complex Exponential (SLEX) basis functions, which are orthogonal and localized in both time and frequency domains, and applied them to analyze bivariate and multivariate epileptic electroencephalograms. Fryzlewicz et al. (2003) also used Dahlhaus's definition of local stationarity to fit finance and environmental time series using wavelets. A Bayesian approach is used in Rosen et al. (2009, 2012), which extends the splines method for one time series mentioned in Section 1.1. In Rosen et al. (2009), the time series is divided into equally long segments and the global log-spectrum in each segment is modeled through a local mixture log-spectra. Each log-spectrum is modeled through smoothing splines. In Rosen et al. (2012), the length of the segments is not fixed and each segment has it is own log-spectrum, modeled, again, using smoothing splines.

4.3 Modeling approach for non-stationary time series

In Chapter 2, we used a mixture of Gaussian distributions with frequency dependent weights and frequency dependent means to approximate the distribution of the log-periodogram observations. The Gaussian mixture model on the log-periodogram observations induces a model for the log-spectral density, specifically a mixture of linear (with respect to the frequency) basis, with the same weights as in the Gaussian mixture. We have pointed out how the frequency can be interpreted as a covariate and the log-periodogram observations as the response variable.

Since the frequency has a role of a covariate, a natural generalization of our model to a time-varying model is the introduction of a time covariate. This can be achieved by making the mixture weights and the mixture means dependent on both time and frequency. A key observation, at this point, is that we need more than one observation in the time domain to calculate the periodogram and we need enough observations to gain meaningful information about the frequency behavior. It follows naturally, as we have seen in Section 4.2, to divide the time series in segments that are assumed to be stationary. For each segment we calculate the log-periodogram. The smoothness of the evolution from one segment to the next one is naturally modeled by the local

weights in the Gaussian mixture.

For now, let us consider the dimension of length of the time series to be fixed to T. Recall that we denote the non-stationary time series with $\{X_{1,T}, ..., X_{T,T}\}$, where the double subscript indicates the specific observation and the total number of observations. Moreover, we consider the rescaled time $u = t/T$, for $t = 1, ..., T$. Increasing the number of observations T, then, implies that we have time points that are closer to each other. We segment the time series into S non overlapping segments. The i-th segment contains n_i observations from the time series, for $i = 1, ..., S$, and $\sum_{i=1}^{S} n_i = T$. Equivalently, we segment $[0, 1]$ in S time intervals, L_i, for $i = 1, ..., S$, in such a way that $\cup_{i=1}^{S} L_i = [0, 1]$. Observation $x_{t,T}$ is associated to the i-th time interval if $t/T \in L_i$, for $t = 1, ..., T$. Since the i-th time interval contains n_i observations, we have $N_i = \lfloor n_i/2 \rfloor - 1$ effective observations of the periodogram for the i-th interval. We denote by $\boldsymbol{y}_i = (y_{i1}, ..., y_{iN_i})$ the (translated) log-periodogram observations for segment i, and we denote the mid point of each interval as $\bar{u}_i$, for $i = 1, ...S$. Hence, if $t/T \in L_i$, then the (translated) log-periodogram associated with $X_{t,T}$ is $\boldsymbol{y}_i$, for $i = 1, ..., S$.

Now, we can extend the model presented in Chapter 2 by making the weights and the Gaussian means dependent both on the frequency and on the segment midpoint.

The mixture model for the (translated) log-periodogram observation j and

time period i becomes:

$$y_{ij}|\boldsymbol{\theta} \overset{ind.}{\sim} \sum_{k=1}^{K} g_k(\mu(\bar{u}_i, \omega_{ij}), \tau) N(y_{ij}|\alpha_k + \beta_k \omega_{ij} + \gamma_k \bar{u}_i + \delta_k \omega_{ij} \bar{u}_i, \sigma^2), \quad (4.2)$$

where $\omega_{ij} = 2\pi j/n_i$, for $j = 1, ..., N_i$ and $i = 1, ..., S$. The weights are defined as follows:

$$g_k(\mu(\bar{u}_i, \omega), \tau) = \int_{\pi(k-1)/K}^{\pi k/K} f_Y(y \mid \mu(\bar{u}_i, \omega), \tau) \, \mathrm{d}y, \quad (4.3)$$

where $f_Y(y \mid \mu(\bar{u}_i, \omega), \tau)$ is the density of a logit-normal$(\mu(\bar{u}_i, \omega), \tau)$ distribution on $(0, \pi)$. Here, $\mu(\bar{u}_i, \omega) = \zeta + \phi\omega + \eta\bar{u}_i + \psi\omega\bar{u}_i$, and the precision parameter τ plays again the role of a smoothness parameter.

The vector $\boldsymbol{\theta}$ in 4.2 collects all model parameters, in particular, it includes the parameters for the weights, $(\zeta, \phi, \eta, \psi)$ and τ, and all the parameters of the normal mixture components, $\boldsymbol{\alpha} = \{\alpha_k : k = 1, ..., K\}$, $\boldsymbol{\beta} = \{\beta_k : k = 1, ..., K\}$, $\boldsymbol{\gamma} = \{\gamma_k : k = 1, ..., K\}$, $\boldsymbol{\delta} = \{\delta_k : k = 1, ..., K\}$, and the common variance parameter σ^2.

In each mixture mean, together with the intercept, α_k, and the parameter associated with the frequency, β_k, we include the midpoint with associated parameter γ_k and we also include an interaction term between frequency and time, with parameter δ_k. In this way, for each midpoint (segment), we have a different intercept and a different slope (with respect to frequency). In the mixture weights, we make the first parameter of the logit-normal distribution dependent on the frequency and the midpoints, in the same fashion as

we did for the mixture means. The common variance parameter σ^2 and the smoothness parameter τ do not depend on time.

If we fix the time period, we have K basis functions that are linear in ω, that is $(\alpha_k+\gamma_k\bar{u}_i)+(\beta_k+\delta_k\bar{u}_i)\omega$. Moreover, the location parameter of the logit-normal distribution is linear in ω, specifically $\mu(\bar{u}_i,\omega)=(\zeta+\eta\bar{u}_i)+(\phi+\psi\bar{u}_i)\omega$. For each segment, we recover the model we presented in Chapter 2, with parameters $(\alpha_k+\gamma_k\bar{u}_i)$ and $(\beta_k+\delta_k\bar{u}_i)$ for the mixture means, and $(\zeta+\eta\bar{u}_i)$, $(\phi+\psi\bar{u}_i)$ and τ for the mixture weights.

Since all the observations that belong to the same time interval have the same associated log-periodogram, we have a model for the log-spectral density that is continuous in ω and it is piecewise constant in t. The induced model for the log-spectral density is:

$$\log f_S(u,\omega) \approx \log \widehat{f_S(u,\omega)} = \sum_{i=1}^{S}\sum_{k=1}^{K} g_k(\mu(\bar{u}_i,\omega),\tau)(\alpha_k+\beta_k\omega+\gamma_k\bar{u}_i+\delta_k\omega\bar{u}_i)\mathbb{I}(u\in L_i).$$

The weights were defined in (4.3).

Based on the blocked Whittle likelihood in 4.1, we can rewrite the model using $\sum_{i=1^S} N_i$ normally distributed auxiliary variables, r_{ij}, for $i=1,...,S$ and $j=1,...,N_i$.

$$\begin{aligned} y_{ij} \mid r_{ij},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\sigma^2 &\overset{ind.}{\sim} \textstyle\sum_{k=1}^{K} \mathrm{N}(y_{ij} \mid \alpha_k+\beta_k\omega_{ij}+\gamma_k\bar{u}_i+\delta_k\omega_{ij}\bar{u}_i,\sigma^2) \\ &\qquad \mathbb{I}\left\{\frac{k-1}{K} < \frac{\exp(r_{ij})}{1+\exp(r_{ij})} \leq \frac{k}{K}\right\} \\ r_{ij} \mid \zeta,\phi,\eta,\psi,\tau &\overset{ind.}{\sim} \mathrm{N}(r_{ij} \mid \zeta+\phi\omega_{ij}+\eta\bar{u}_i+\psi\omega_{ij}\bar{u}_i, 1/\tau). \end{aligned}$$

Posterior simulation is implemented using the augmented version of the model. For practical and theoretical purposes, as explained in Section 4.3.1, we use the same number of observation for each segment, that is $n_i = n = T/S$, for $i = 1, ..., S$. Technical details on the Gibbs sampler used to sample from the joint posterior distribution of the parameters are given in Appendix B.

4.3.1 Idea for asymptotic theory

Let us consider the case in which each segment of the time series has the same number of observations, that is, $n_i = n = T/S$. The intervals are $L_1 = [0, 1/S]$ and $L_i = ((i-1)/S, i/S)$ for $i = 2, ..., S$. Each time interval L_i has length $1/S$. The midpoint of each segment is $\bar{u}_i = (i-1)/S + 1/(2S)$, for $i = 1, ..., S$. As before, we assume stationarity for each segment. In the following, we fix the number of observations per segment. As the total number of observations T increases, that is we have a higher sample rate, we consider a larger number of segments S. Moreover, the length of each $L_i = 1/S$ decreases. We define our estimator of the evolutionary log-spectral density as a bivariate continuous function, both with respect to frequency and (rescaled) time. The model for the evolutionary spectral density is:

$$\log f(u, \omega) \approx \log \widehat{f(u, \omega)} = \sum_{k=1}^{K} g_k(\mu(u, \omega), \tau)(\alpha_k + \beta_k \omega + \gamma_k u + \delta_k \omega u),$$

where the weights are defined as:

$$g_k(\mu(u,\omega),\tau) = \int_{\pi(k-1)/K}^{\pi k/K} f_Y(y \mid \mu(u,\omega),\tau)\,\mathrm{d}y,$$

where $f_Y(y \mid \mu(u,\omega),\tau)$ is the density of a logit-normal$(\mu(u,\omega),\tau)$ distribution on $(0,\pi)$, with $\mu(u,\omega) = \zeta + \phi\omega + \eta u + \psi\omega u$.

For a finite T and a number S of segments, we built a model for the spectral density that is continuous in frequency and piecewise constant in (rescaled) time, that is:

$$\log f_S(u,\omega) \approx \log \widehat{f_S(u,\omega)} = \sum_{i=1}^{S}\sum_{k=1}^{K} g_k(\mu(\bar{u}_i,\omega),\tau)(\alpha_k + \beta_k\omega + \gamma\bar{u}_i + \delta\omega\bar{u}_i)\mathbb{I}(u \in L_i),$$

with weights given in 4.3.

If we look at $\log \widehat{f_S(u,\omega)}$ as a function of the (rescaled) time u, at a fixed frequency ω, we have a piecewise constant function on $[0,1]$ with values $\sum_{k=1}^{K} g_k(\mu(\bar{u}_i,\omega),\tau)(\alpha_k+\beta_k\omega+\gamma\bar{u}_i+\delta\omega\bar{u}_i)$ for $u \in L_i$. Moreover $\cup_{i=1}^{S} L_i = [0,1]$. As the number of observations increases, that is $T \to \infty$, we can increase the number of segments, that is $S \to \infty$, keeping the number of observations in each segment constant and large enough to provide an adequate estimate of the spectral density. It is clear that as the number of segments S increases, the piecewise constant function $\log \widehat{f_S(u,\omega)}$ converges pointwise to the continuous function $\log \widehat{f(u,\omega)}$. Moreover, since $|\log \widehat{f_S(u,\omega)}| \leq \max_u \log \widehat{f(u,\omega)}$, by the dominated convergence theorem, we have L_p convergence, for any $p \geq 1$.

4.3.2 Simulation study

We are interested in the performance of our model in the case of locally stationary processes, for which the spectral density varies slowly over time. A class of zero-mean locally stationary processes is, for example, the class of time varying autoregressive processes. For this reason, we perform a simulation study using data from a time varying autoregressive process of order two. We are then able to compare our estimate of the time-varying spectral density to the truth. The simulation scenario is taken from Ombao et al. (2001) and is also used in Rosen et al. (2009, 2012). We use a total of $T = 4800$ observations, divided into $S = 16$ segments with $n = 300$ observations each. This means that, for each segment, we have $N = 149$ effective observations from the translated log-periodogram. This is consistent with the settings used in Chapter 2 and Chapter 3, in which we had 300 observations for the time series.

The observations are generated from the following time varying autoregressive process of order two:

$$x_t = a_t x_{t-1} - 0.81 x_{t-2} + \epsilon_t, \quad t = 1, ..., T, \tag{4.4}$$

where $\epsilon_t \overset{iid}{\sim} N(0, 1)$ and $T = 4800$. The first parameter of the autoregressive process varies with time, $a_t = 0.8(1 - 0.5\cos(\pi t/T))$. The second parameter does not vary with time and it is fixed at -0.81. This implies that the modulus

of the process, 0.9, does not vary with time; in other words, the amount of information contained in the time series does not vary with time. The argument, on the other hand, varies with time and the peak moves to the left and becomes more accentuated as time evolves. Figure 4.1 shows the data (left panel), the argument of the process as a function of time (center panel), and the true continuous bivariate spectral density as a function of time and frequency (right panel). Time is scaled to the unit interval. In the right plot, the darker the shades the higher is the power.

We fixed the number of mixture components to $K = 50$; similar results were obtained with a larger value of K. We assumed α_k , β_k, γ_k and δ_k to be independent normally distributed centered at zero with variance 1000. For the common variance parameter, we used an inverse gamma prior with mean 3 and variance 9. For the weight parameters, we assumed ζ, ϕ, η and ψ to be independent normally distributed centered at zero with variance 100. We placed a gamma$(10, 0.01)$ prior on the smoothness parameters τ, to support a large interval on the positive real line. We run an MCMC Gibbs sampler algorithm with a total of 100000 iterations, of which 20000 were taken as burn-in period.

Figure 4.2 shows the joint posterior densities for (ζ, ϕ) (left panel), (η, ψ) (center panel) and the histogram of the posterior density for τ (right panel), together with the prior (solid line). The joint posterior distribution of (η, ψ)

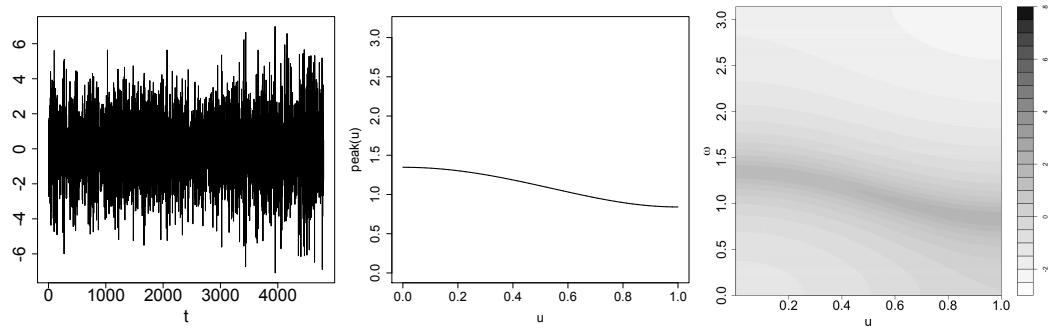

Figure 4.1: Process realization (left panel), the argument of the process as a function of time (center panel), and the true continuous bivariate spectral density as a function of time and frequency (right panel).

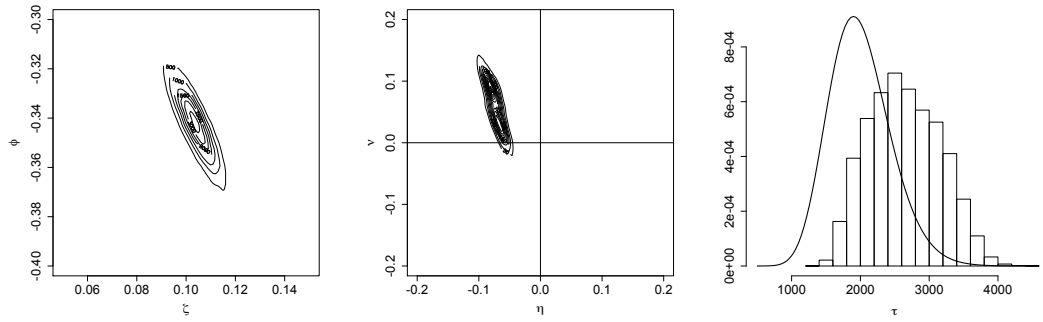

Figure 4.2: Slowly varying autoregressive process. Joint posterior densities for (ζ, ϕ) (left panel), (η, ψ) (center panel) and the prior and posterior density for τ (right panel).

takes values different from zero. This suggests that the partitioning of the (frequency) support varies with time, and both the time alone and the interaction between time and frequency are significant.

We compare the analytically available spectral density of the time varying autoregressive process with the posterior estimate obtained with our method. For a fair comparison, we averaged the analytical log-spectral density over the same intervals we used in the estimation. Figure 4.3 shows the log-spectral

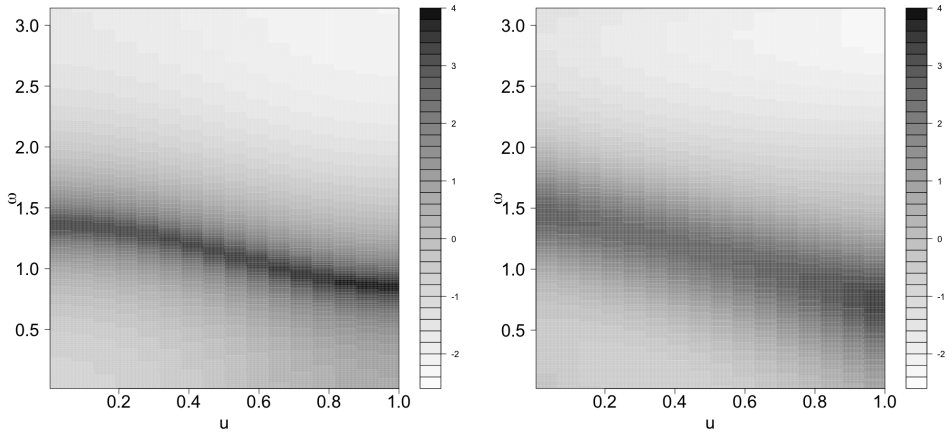

Figure 4.3: Slowly varying autoregressive process. True log-spectral density, averaged on each time period (left panel) and posterior mean estimate of the log-spectral density obtained with our model (right panel).

density, averaged on each time period (left panel) and the posterior mean estimate of the log-spectral density obtained with our model (right panel). The posterior mean estimate captures the time varying shape of the log-spectral density.

Figure 4.4 shows, for each time period, the average log-spectral density, as well as the corresponding posterior mean estimates and 95% credible intervals. The model adequately captures the different log-spectral density shapes and is successful in discerning the evolution of the shape with time.

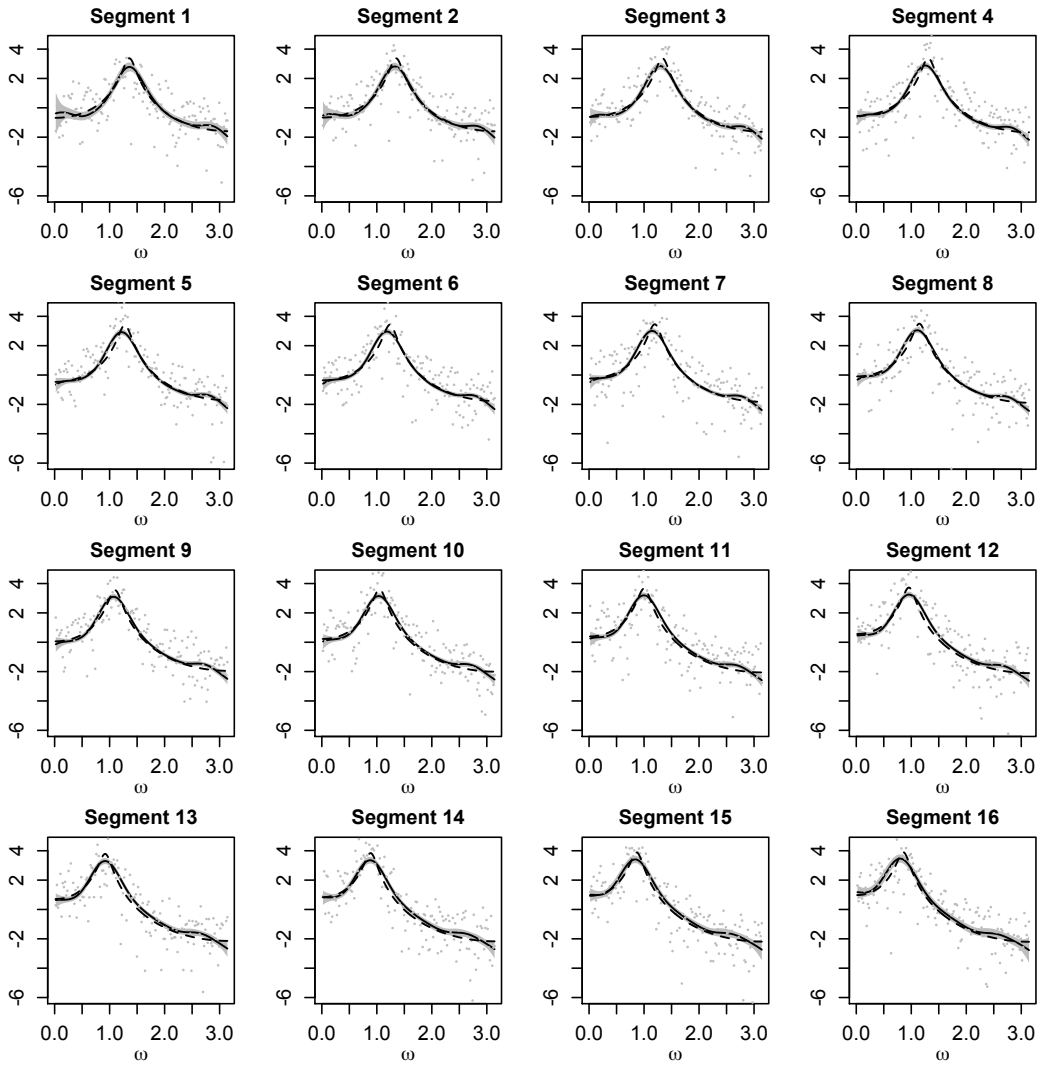

Figure 4.4: Slowly varying autoregressive process. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral density at each time period. Each panel includes also the average log-spectral density (dashed line) for each period and the log-periodogram (dots).

4.4 Modeling for multiple non-stationary time series

The key feature of the model in (4.2) is that it can be extended in a hierarchical fashion to multiple related non-stationary time series. Consider

M related non-stationary time series, which, without loss of generality, are assumed to have the same number of observations n. Following the approach of Section 4.3, we divide each time series in S time intervals, each one containing $n = T/S$ observations. For each time series, we calculate S local periodograms, each one with N observations. In the following, y_{mij} is one observation from a (translated) log-periodogram, specifically, the first index indicates the time series ($m = 1, ..., M$), the second index indicates the time interval ($i = 1, ..., S$) and the third index indicates the Fourier frequency ($j = 1, ..., n_i$).

Now, for each m, i and j we approximate the distribution of the y_{mij} with a smooth mixture of Gaussian distributions. We take the mean parameters of the Gaussian mixture components, that is, $(\alpha_k, \beta_k, \gamma_k, \delta_k)$, for $k = 1, ..., K$, to be common among time series. This translates into a set of K functions for the log-spectral density approximation which are common to all time series. On the other hand, we let the parameters that specify the weights be time series specific, that is, $\boldsymbol{\xi}_m = (\zeta_m, \phi_m, \eta_m, \psi_m, \tau_m)$, for $m = 1, ..., M$. For each time series, the weights select the mean functions to approximate the corresponding log-spectral density. Since the spectral densities are related, similar mean functions can be selected for more than one location, allowing grouping of spectral densities. We use M distinct smoothness parameters τ_m to allow different levels of smoothness across the spectral densities. However, we do not make the smoothness parameters τ_m vary with time.

Hence, extending (4.2), the observation stage for the hierarchical model on the M time series can be written as

$$y_{mij} \mid \boldsymbol{\theta} \overset{ind.}{\sim} \sum_{k=1}^{K} g_k(\mu_m(\bar{u}_i, \omega_{ij}), \tau_m)\, \mathrm{N}(y_{mij} \mid \alpha_k + \beta_k \omega_{ij} + \gamma_k \bar{u}_i + \delta_k \bar{u}_i \omega_{ij}, \sigma^2), \tag{4.5}$$

where the k-th weight at the m-th location is defined, again, in terms of increments of a logit-normal distribution function, with mean function, $\mu_m(\bar{u}_i, (\omega) = \zeta_m + \phi_m \omega + \eta_m \bar{u}_i + \psi_m \omega \bar{u}_i$, and precision parameter, τ_m, that are time series specific. Again, $\boldsymbol{\theta}$ collects all model parameters: the parameters of the K mixture components means, $\boldsymbol{\alpha} = \{\alpha_k : k = 1, ..., K\}$, $\boldsymbol{\beta} = \{\beta_k : k = 1, ..., K\}$, $\boldsymbol{\gamma} = \{\gamma_k : k = 1, ..., K\}$ and $\boldsymbol{\delta} = \{\delta_k : k = 1, ..., K\}$, the common variance parameter σ^2, and the mixture weights parameters, $\boldsymbol{\zeta} = \{\zeta_m : m = 1, ..., M\}$, $\boldsymbol{\phi} = \{\phi_m : m = 1, ..., M\}$, $\boldsymbol{\eta} = \{\eta_m : m = 1, ..., M\}$, $\boldsymbol{\psi} = \{\psi_m : m = 1, ..., M\}$, and $\boldsymbol{\tau} = \{\tau_m : m = 1, ..., M\}$.

Posterior simulation is implemented using the augmented version of the model based on $M \sum_{i=1}^{S} N_i$ normally distributed auxiliary variables, r_{mij}, for $m = 1, ..., M$, $i = 1, ..., S$ and $j = 1, ..., N_i$. In particular,

$$\begin{aligned} y_{mij} \mid r_{mij}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \sigma^2 &\overset{ind.}{\sim} \textstyle\sum_{k=1}^{K} \mathrm{N}(y_{mij} \mid \alpha_k + \beta_k \omega_{ij} + \gamma_k \bar{u}_i + \delta_k \omega_{ij} \bar{u}_i, \sigma^2) \\ &\qquad \mathbb{I}\left\{(k-1)/K < \frac{\exp(r_{mij})}{1 + \exp(r_{mij})} \leq k/K\right\} \\ r_{mij} \mid \zeta_m, \phi_m, \eta_m, \psi_m, \tau_m &\overset{ind.}{\sim} \mathrm{N}(r_{mij} \mid \zeta_m + \phi_m \omega_{ij} + \eta_m \bar{u}_i + \psi_m \omega_{ij} \bar{u}_i, 1/\tau_m). \end{aligned}$$

In our examples, we use $n_i = n$ for $i = 1, ..., S$. Technical details on the Gibbs sampler used to implement the hierarchical model are given in Appendix B.

The full Bayesian model is completed with priors for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ and σ^2, and a hierarchical prior for the $(\zeta_m, \phi_m, \eta_m, \psi_m)$ and τ_m, for $m = 1, ..., M$. The weight parameters are assumed a priori independent of the Gaussian mixture component parameters. We assume $\sigma^2 \sim$ inv-gamma$(n_{\sigma^2}, d_{\sigma^2})$, that is, an inverse gamma prior (with mean $d_{\sigma^2}/(n_{\sigma^2} - 1)$, and $n_{\sigma^2} > 1$), $\alpha_k \sim \mathrm{N}(\mu_{0\alpha}, \sigma^2_\alpha)$, $\beta_k \sim \mathrm{N}(\mu_{0\beta}, \sigma^2_\beta)$, $\gamma_k \sim \mathrm{N}(\mu_{0\gamma}, \sigma^2_\gamma)$ and $\delta_k \sim \mathrm{N}(\mu_{0\delta}, \sigma^2_\delta)$, for $k = 1, ..., K$. The hierarchical prior is given by

$$\begin{aligned}(\zeta_m, \phi_m, \eta_m, \psi_m) \mid \boldsymbol{\mu}_w, \Sigma_w &\overset{ind.}{\sim} \mathrm{N}(\boldsymbol{\mu}_w, \Sigma_w), \quad m = 1, ..., M, \\ \tau_m \mid d_\tau &\overset{ind.}{\sim} \text{gamma}(n_\tau, d_\tau), \quad m = 1, ..., M,\end{aligned}$$

where gamma(n, d) denotes the gamma distribution with mean n/d. To borrow strength across the time series, we place a four dimensional normal prior on $\boldsymbol{\mu}_w$, and a four dimensional inverse Wishart hyper-prior on the covariance matrix Σ_w. For the τ_m, we fix the shape parameter, n_τ, and place a gamma prior on the rate parameter, d_τ.

4.5 Simulation study

In order to assess the performance of the proposed spectral model for related non-stationary time series, we consider a generating mechanism that represents a hypothetical scenario involving multiple related time series with slowly varying spectral density. In order to compare our posterior estimates to the true spectral densities, we simulated data from processes with spectral den-

sities available in analytical form. Moreover, we consider replicates, meaning that more than one time series is generated from the same underlying process. We considered two underlying generating processes, with five replicates in each case, leading to a total of $M = 10$ time series. For each time series, we simulated 4800 time points, and we divided the time series in $S = 16$ segments, leading to $N = 149$ observations from the log-periodogram for segment. In addition to posterior estimates and credible intervals for the spectral densities, we investigate the posterior distribution of the weight parameters, ζ_m, ϕ_m, η_m, ψ_m and τ_m, for $m = 1, ..., M$, which can be useful in identifying similar time varying spectral characteristics across multiple time series.

Both the underlying generating processes are time varying autoregressive processes of order two, or TVAR(2) process, with the second parameter that does not vary with time set to -0.81, which corresponds to a modulus of 0.9. The first parameter varies with time and is $a_1(t) = 0.8(1 - 0.5\cos(\pi t/T))$ for the first generating process, and $a_2(t) = 0.9(1 - 0.99\cos(\pi t/T))$ for the second one. This means that the argument (peak of the spectral density) varies with time for both processes, but varies faster in the second one. Figure 4.5 shows the arguments of the two generating processes as a function of time (left panel), and the true continuous bivariate spectral densities as a function of time and frequency (central and right panels). The time is scaled to the unit interval; the darker the grey scale the higher is the power.

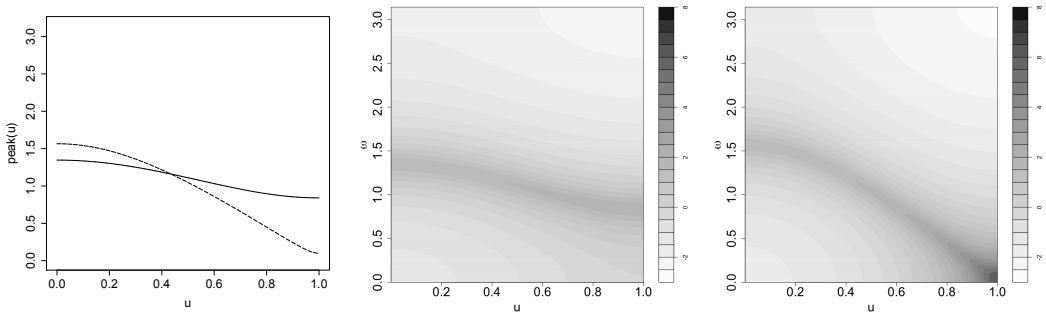

Figure 4.5: Multiple time series with varying argument and constant modulus. Arguments of the two generating processes as a function of time (left panel), where the solid line denotes the argument for the first process and the dashed for the second process, and the true continuous bivariate spectral densities as a function of time and frequency, for the first process (central panel) and the second process (right panel).

We fixed the number of mixture components to $K = 50$; similar results were obtained with a larger value of K. We assumed α_k, β_k, γ_k and δ_k to be independent normally distributed centered at zero with variance 1000 such that the linear basis can have a wide range of motion. For the common variance parameters, we used an inverse gamma prior with mean 3 and variance 9. For the smoothness parameters τ_m, $m = 1, ..., M$, we fixed the shape parameter to 30 and placed a gamma$(3, 20)$ on the rate parameter. This results in a marginal prior distribution for each τ_m that supports a large interval on the positive real line. Moreover, since each time series has its own smoothness parameter, we can have different levels of smoothness for different spectral densities. The hyper-prior on the mean vector $\boldsymbol{\mu}_w$ was centered at 0 and had a diagonal covariance matrix, with diagonal elements equal to 10, while the inverse Wishart distribution parameters were chosen in a way that the marginal

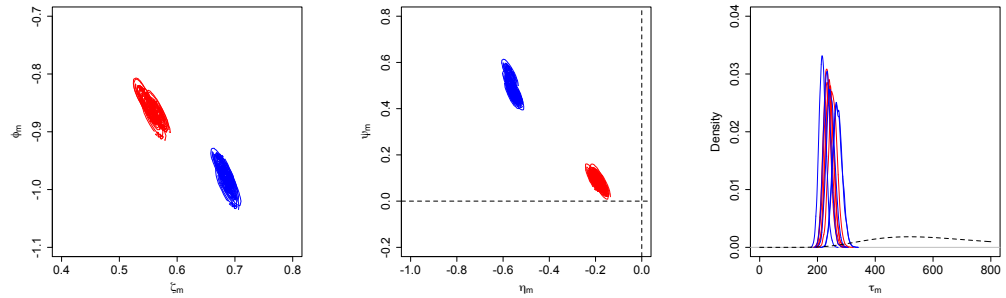

Figure 4.6: Multiple time series with varying argument and constant modulus. Joint posterior densities for (ζ_m, ϕ_m) (left panel), (η_m, ψ_m) (center panel) and the prior marginal and posterior densities for the τ_m (right panel).

distributions for the diagonal elements were inv-gamma$(3, 3)$, and the implied prior distribution on the correlation between parameters was diffuse on $(0, 1)$. We run an MCMC Gibbs sampler algorithm with a total of 100000 iterations, of which 20000 were burn-in.

Figure 4.6 shows the joint posterior densities for the (ζ_m, ϕ_m) parameters (left panel), (η_m, ψ_m) parameters (center panel), and the prior marginal distribution and posterior densities for τ_m (right panel) for $m = 1, ...M$. The color red corresponds to the first five time series, and the blue color to the time series from sixth to tenth. Clearly, the joint posterior distribution of (ζ_m, ϕ_m) and (η_m, ψ_m) allows us to accurately identify the two groups. The posterior distributions of the τ_m parameters, which determine the smoothness of the spectral densities, do not show a clear distinction among the two groups.

We compare the analytical spectral densities with the posterior estimate obtained with our method. For a fair comparison, we averaged the analytical

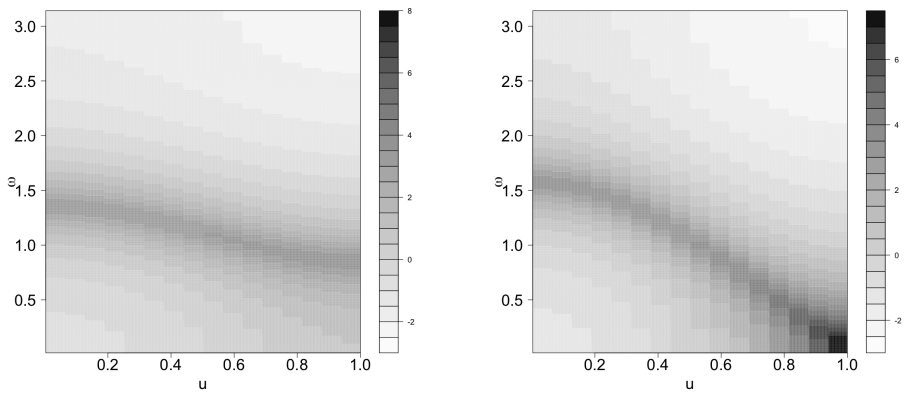

Figure 4.7: Multiple time series with varying argument and constant modulus. Log-spectral densities, averaged over each time period, for the first generating process (left panel) and for the second generating process (right panel).

log-spectral densities over the same intervals we used in the estimation. Figure 4.7 shows the log-spectral density, averaged on each time period, for the first generating process (left panel) and for the second generating process (right panel). Figure 4.8 shows posterior mean estimates of the log-spectral densities obtained with our model (first group in the left column, second group in the right column). The model adequately captures the time varying shape of the log-spectral density.

To further investigate the performance of our model, we plot the spectral density for each time period for two time series, one for each group. Figure 4.9 and Figure 4.10 show the posterior mean estimates and 95% credible intervals for the third and eighth time series, respectively. We also plot the average log-spectral density for each segment. The model captures the changes over time of the spectral characteristics of the underlying processes.

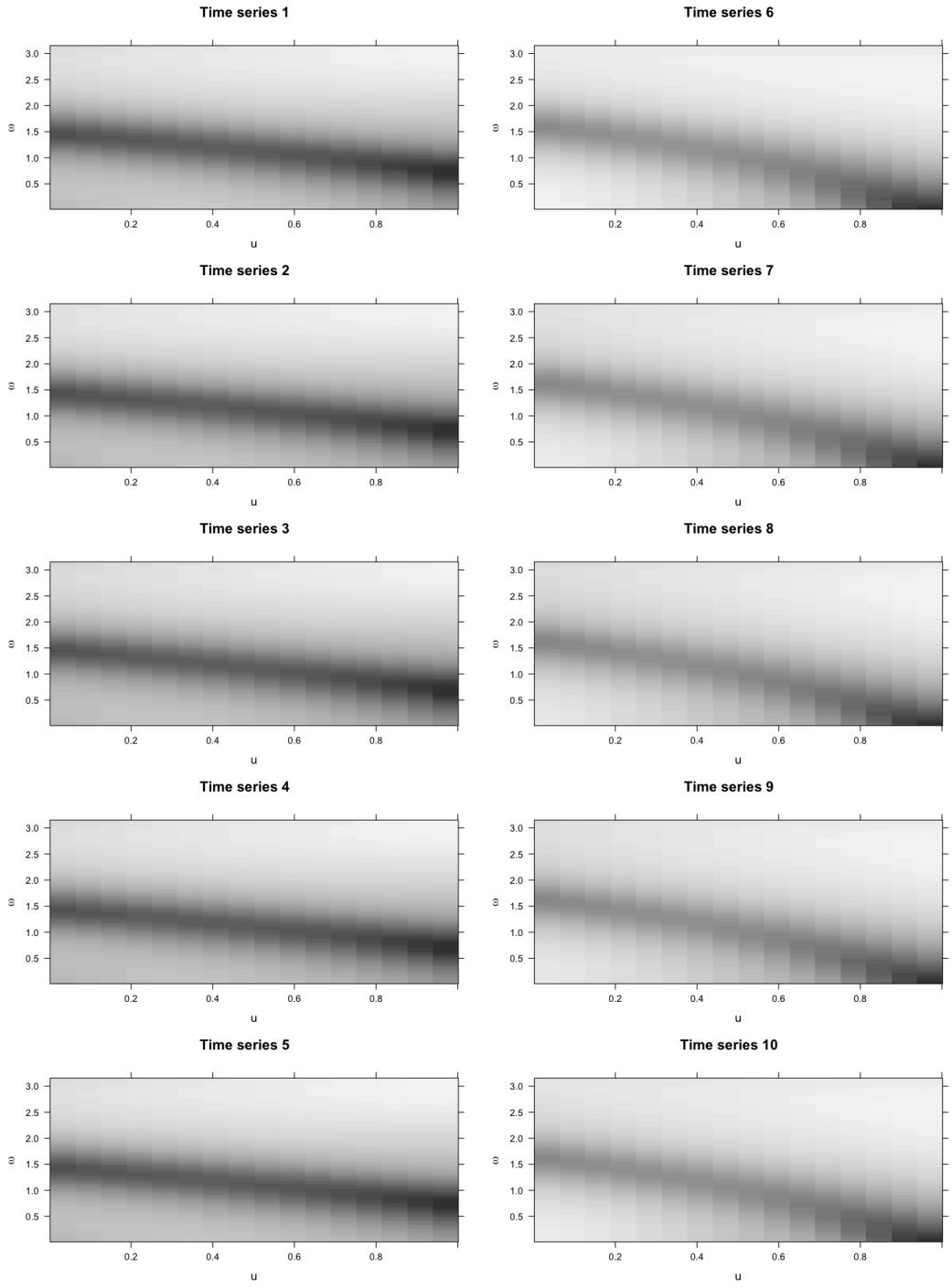

Figure 4.8: Multiple time series with varying argument and constant modulus. Posterior mean estimates of the log-spectral density obtained with our model (first group in the left column, second group in the right column).

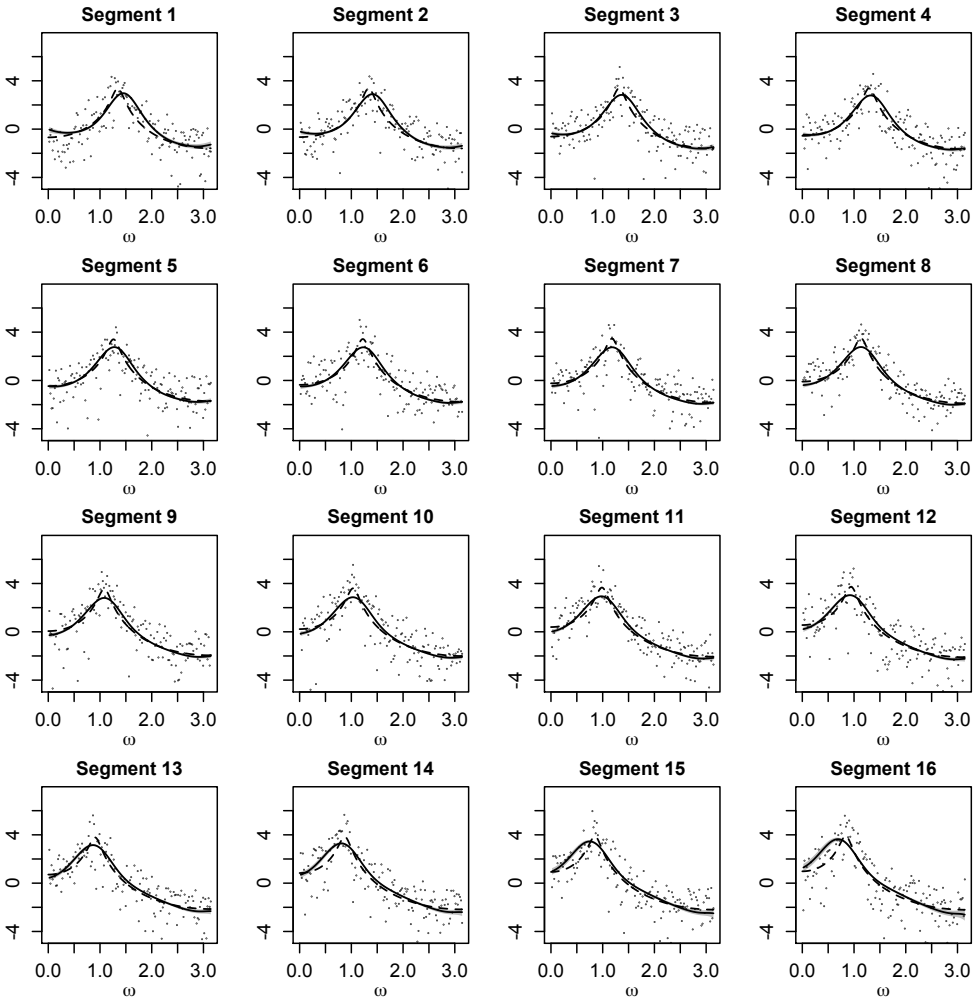

Figure 4.9: Multiple time series with varying argument and constant modulus. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral density at each time period. Each panel includes also the average log-spectral density (dashed line) for each period and the log-periodogram (dots), for the third time series.

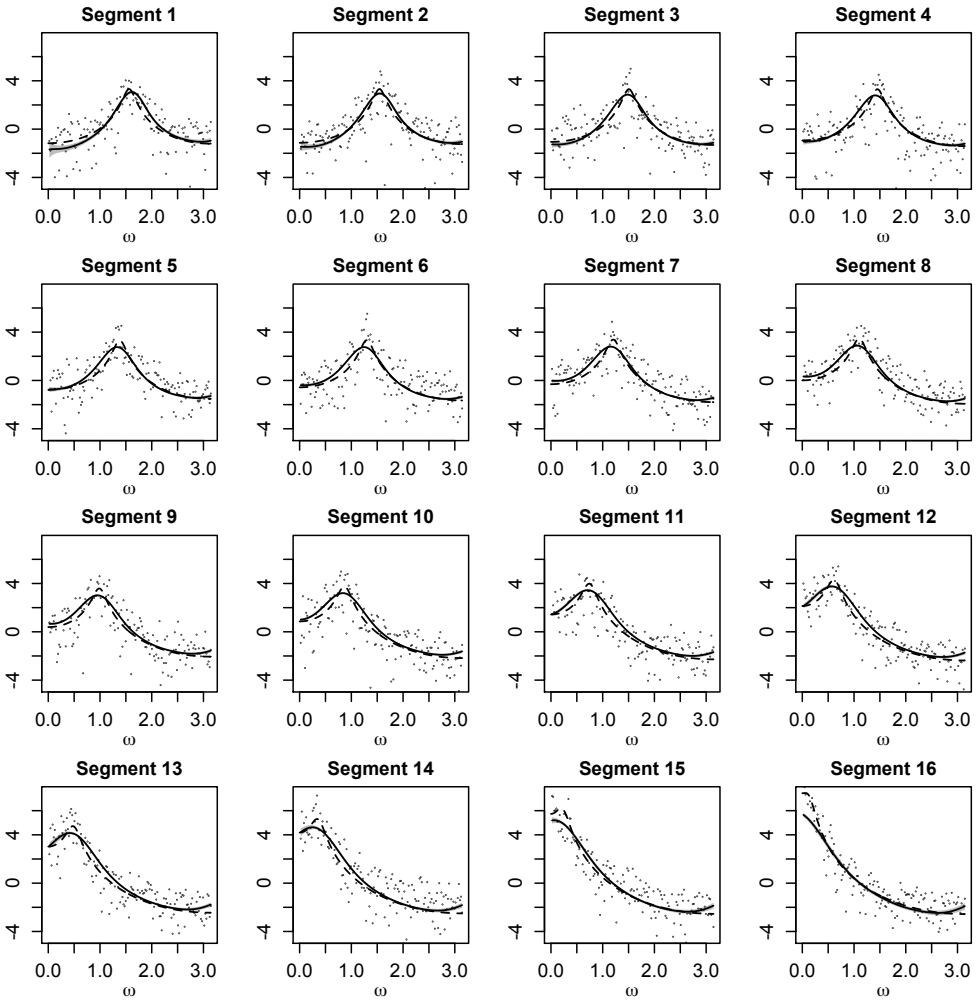

Figure 4.10: Multiple time series with varying argument and constant modulus. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral density at each time period. Each panel includes also the average log-spectral density (dashed line) for each period and the log-periodogram (dots), for the eighth time series.

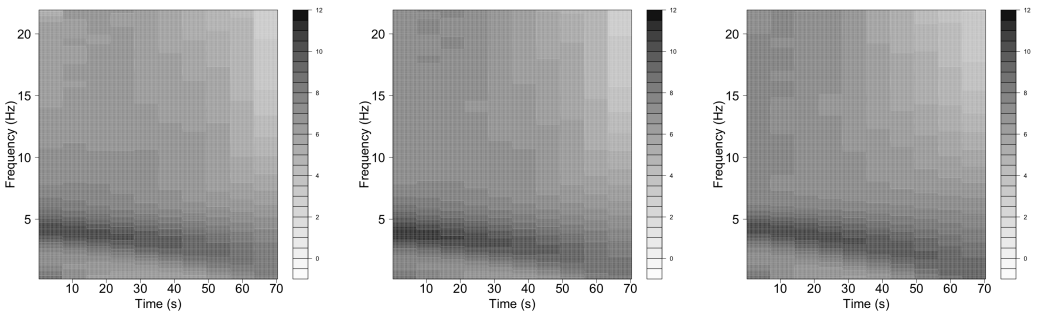

Figure 4.11: EEG data. Posterior mean estimates of the evolutionary spectral density as a function of time (s) and frequency (Hz)for channel Fp_1 (left panel), channel C_z(center panel), and channel O_2 (right panel).

4.6 Electroencephalogram data

We have seen in section 3.4 that the EEG recordings of a subject who received ECT therapy are not stationary in their entirety. Here, we consider 19 EEG time series from the same subject, with 3000 observations each. We label this dataset as ECT data 2. These time series include the observations analyzed in Section 3.4, ECT data 2. The prior specification is the same as in Section 4.5. We divide the series in $S = 10$ time periods, each one with 300 observations.

Figure 4.11 shows the posterior mean estimates of the evolutionary spectral density as a function of time and frequency for channel Fp_1, channel C_z and channel O_2. Figures 4.12, 4.13 and 4.14 show the posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral density at time periods three, six and ten respectively.

As we expect, the peak of the spectral density is more pronounced at the

beginning of the seizure. The peak becomes less pronounced and shifts towards left as the seizure fades. Moreover, there is an increase in variability with time. These results are consistent with the finding in Section 3.4, and previous studies (West et al., 1999; Prado et al., 2001). Figure 4.15 shows the joint posterior densities for (ζ_m, ϕ_m), for the 19 channels. Figure 4.16 shows the joint posterior densities for (η_m, ψ_m), for the 19 channels. The posterior distribution for the (η_m, ψ_m) is concentrated on values different from zero, confirming the evolution over time. Overall, there is no clear distinction of the posterior distributions among the various channels. However, in certain regions of the brain the posterior distributions of the (ζ_m, ϕ_m) are concentrated around values similar to the those obtained from locations in that same region (e.g., channels C_z, P_z, P_3 and C_3). The posterior distribution of the parameters for channels F_8 and T_4 is concentrated around different values than the posterior distribution corresponding to the other channels. This is consistent with previous studies.

4.7 Discussion

In this Chapter, we have extended our methodology to the analysis of non-stationary time series, specifically locally stationary time series. Here, the mixture weights and means depend both on frequency and on time. In this way we are able to estimate a time varying log-spectral density. Since we need more than one observation in the time domain to calculate the log-

periodogram, and enough observations to obtain an adequate estimate, what we obtain is a piecewise estimate of the log-spetral density. Asymptotically, as the number of observations tends to infinity in the sense of infill asymptotics, we have an estimate of the log-spectral density that is continuous in both time and frequency. As done in Chapter 3 for stationary time series, we can extend the model from one time series to multiple time series, in a hierarchical fashion. Hence, we have proposed a Bayesian hierarchical model for non-stationary time series. The model preserves the computational advantages of the models presented in Chapter 2 and Chapter 3, allowing for relatively straightforward implementation of a MCMC Gibbs sampler algorithm for sampling from the posterior distribution of the parameters.

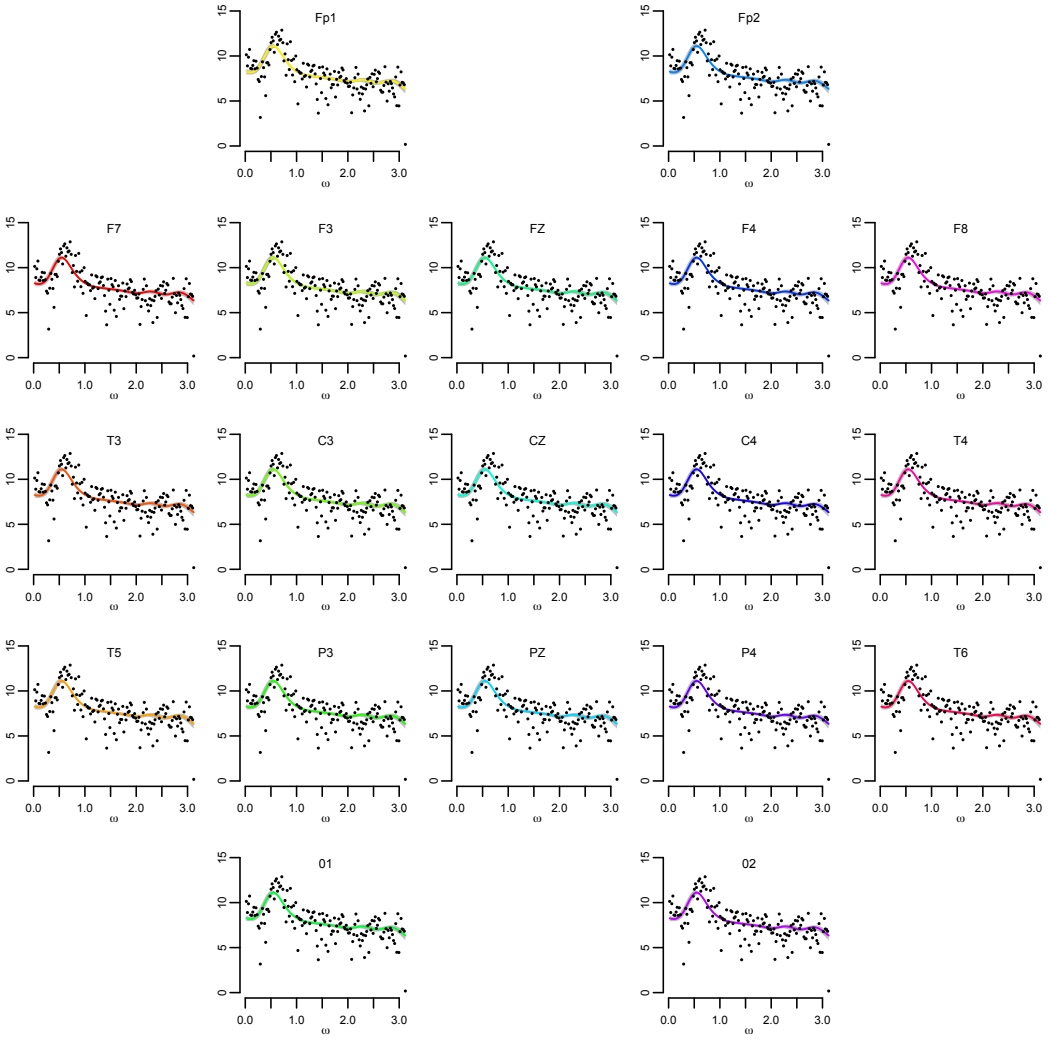

Figure 4.12: EEG dataset, third time interval. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral densities corresponding to the 19 channels. Each panel includes also the log-periodogram (dots) from the specific channel.

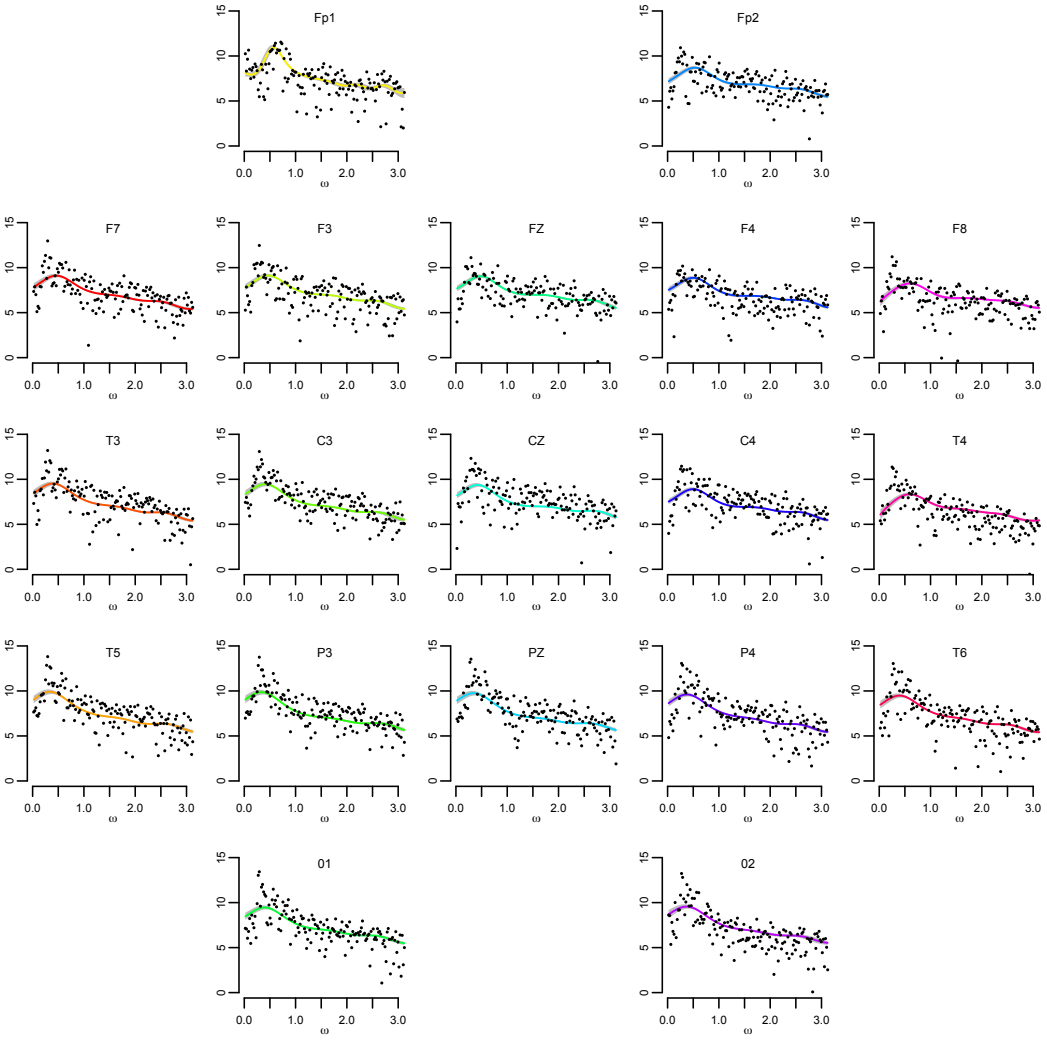

Figure 4.13: EEG data, sixth time interval. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral densities corresponding to the 19 channels. Each panel includes also the log-periodogram (dots) from the specific channel.

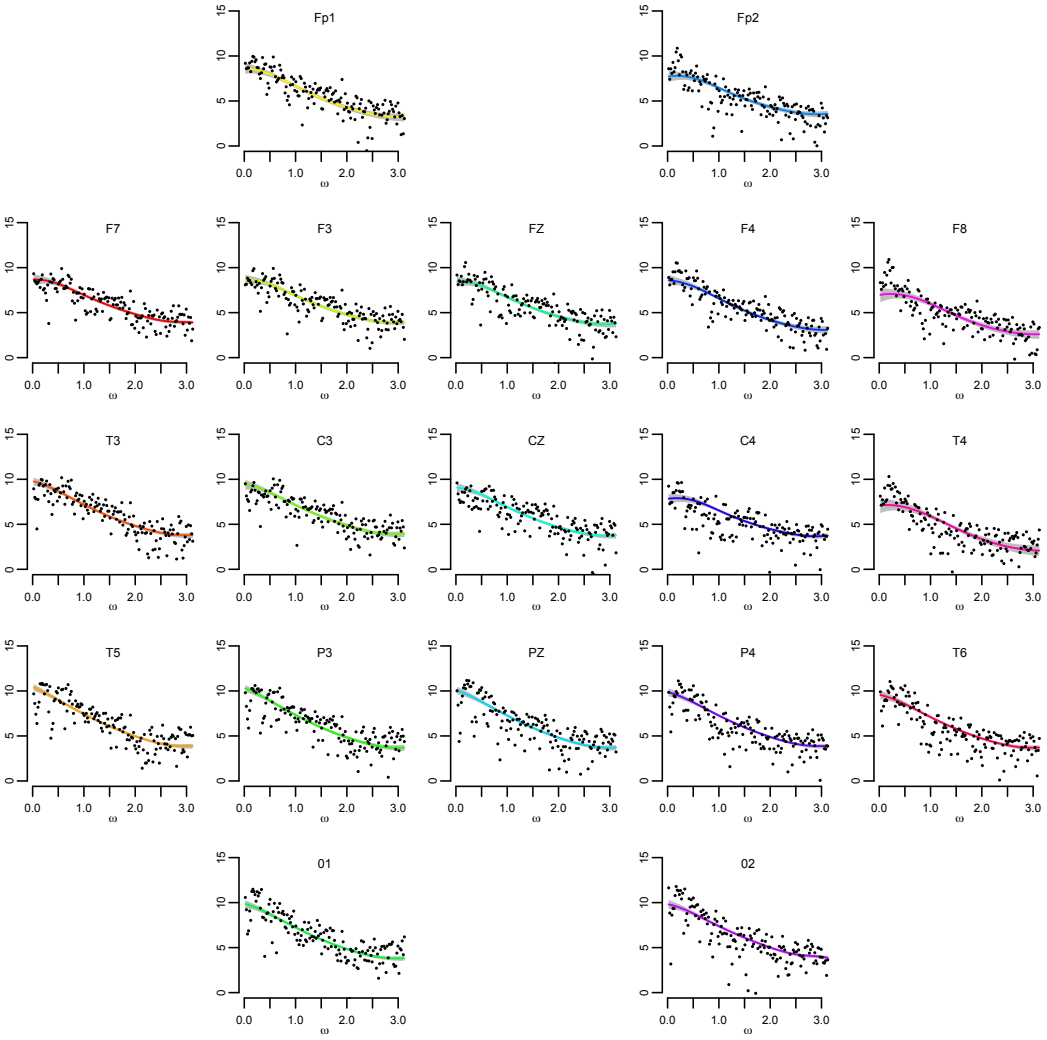

Figure 4.14: EEG data, tenth time interval. Posterior mean estimates (solid lines) and 95% credible intervals (shaded regions) for the log-spectral densities corresponding to the 19 channels. Each panel includes also the log-periodogram (dots) from the specific channel.

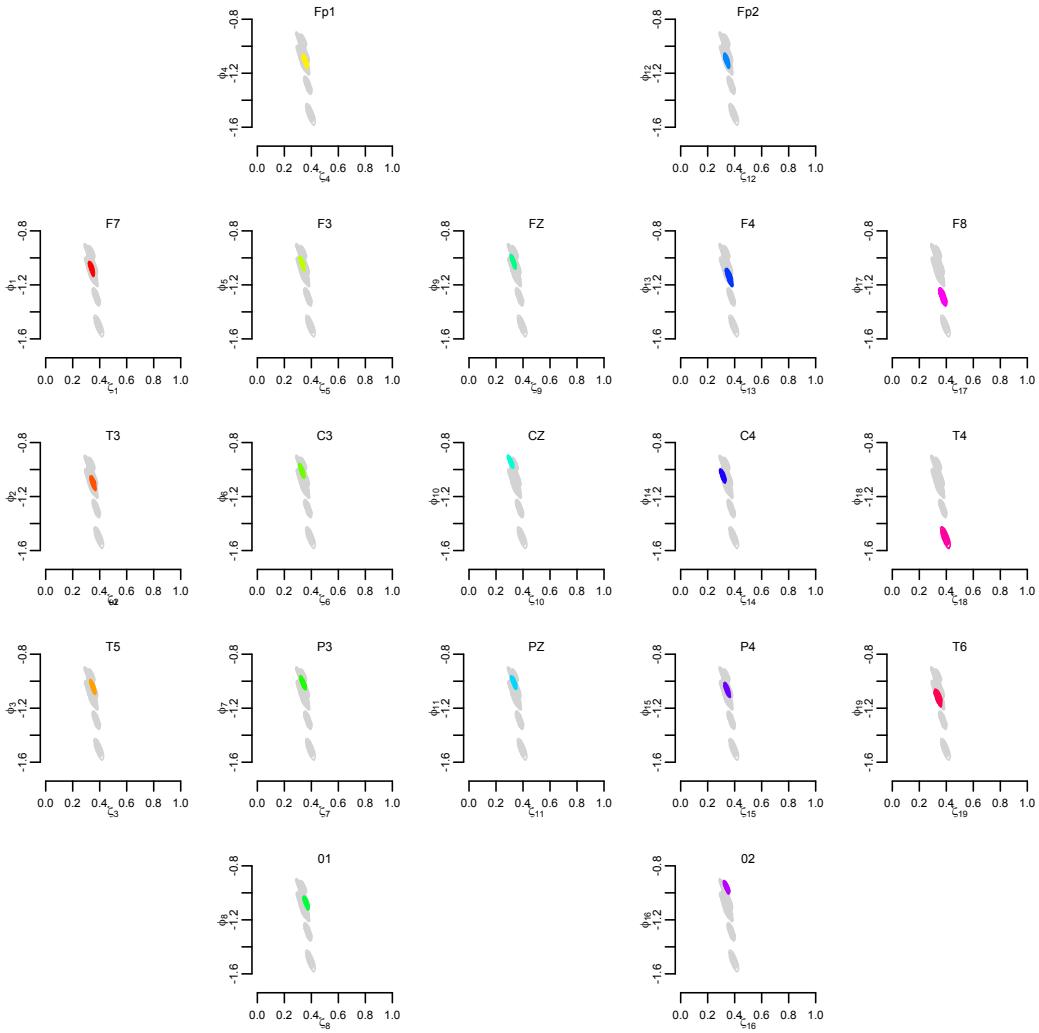

Figure 4.15: EEG data. Joint posterior densities for (ζ_m, ϕ_m), $m = 1, ..., 19$.

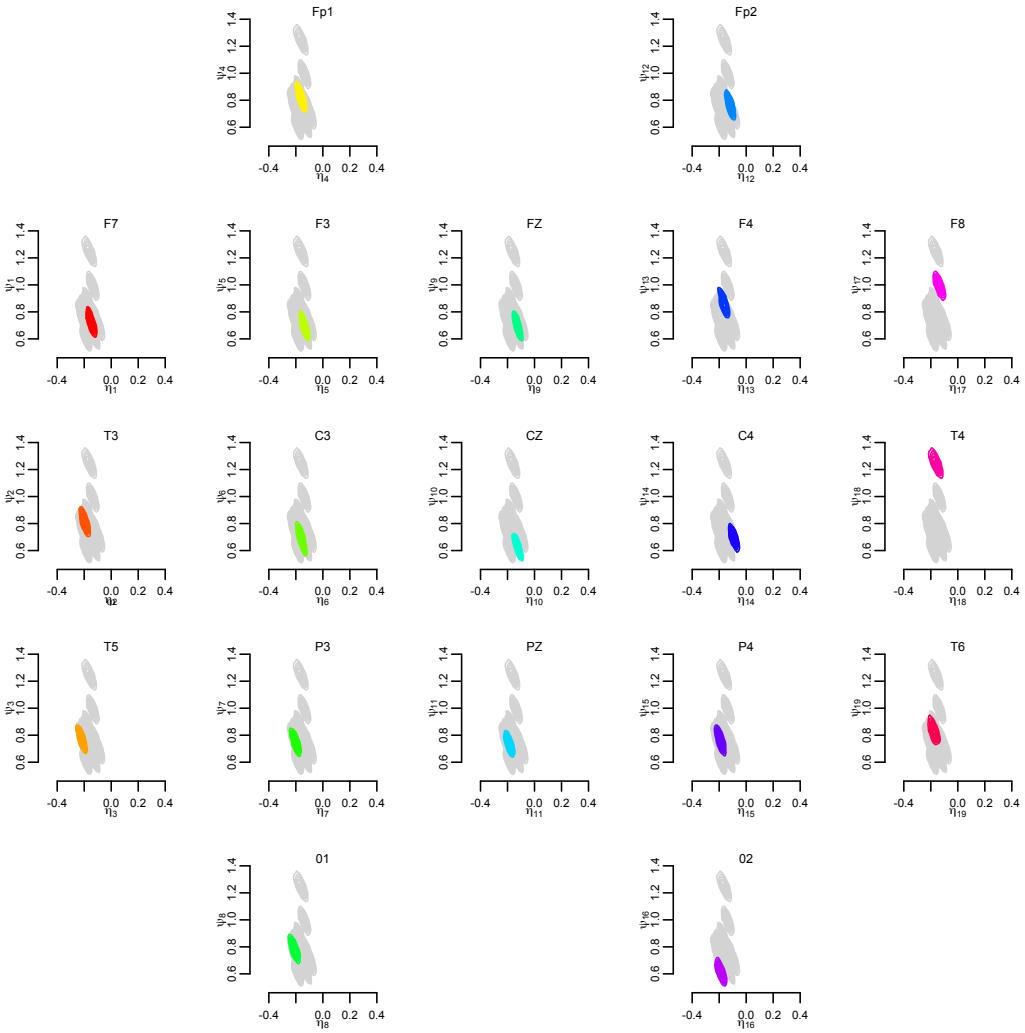

Figure 4.16: EEG data. Joint posterior densities for (η_m, ψ_m), $m = 1, ..., 19$.

Chapter 5

Conclusions

We have presented a new Bayesian approach to spectral density estimation. Our approach is based on modeling the distribution of the log-periodogram through a mixture of Gaussian distributions with frequency-dependent weights and mean functions, which implies a flexible mixture model for the corresponding log-spectral density. Starting with the model for one time series, we proposed a Bayesian hierarchical model for the analysis and estimation of multiple time series in the spectral domain. We have then expanded the model to take into consideration the variation in time of spectral characteristics. A very interesting feature of the time-varying spectral density is that it can be use to detect changes in frequency behavior. For example it allows us to study how the spectral features change with time and it can be used to monitor fatigue, for example, astronauts or drivers.

Our model presents computational advantages in that, for all of the cases mentioned above, it is possible to implement a Markov Chain Monte Carlo algorithm with closed form full conditional distributions. The fact that the model is easy to implement is an advantage in those areas where we need a quick analysis and inference, as in neuroscience. The considered application is from neuroscience, but the approach has a broad range of applications, including econometrics and geoscience.

A key feature of our model is that it allows to group spectral densities based on the posterior distribution of the parameters. Extensions in this sense are discrete nonparametric priors that allow the parameters to be identical, such as a Dirichlet Process prior. This would allow us to cluster time series with identical spectral densities, hence identical underlying generating process. Being able to cluster identical time series is not a relevant problem in neuroscience data because spectral characteristics vary among channels, even if slightly, but it could be relevant for other applications.

We did not include a spatial component to our model, because EEGs do not have a good spatial resolution. However, extensions include adding a spatial structure, for example through a Gaussian process prior on the weight parameters. Moreover, in the presence of covariates we could incorporate them in the mixture weights and/or means.

For certain types of non-Gaussian time series, it has been shown in Contreras-

Cristan et al. (2006) that the Whittle likelihood does not provide a satisfactory approximation for small sample sizes. An interesting direction for future study involves the applicability of the modeling approach for spectral densities of general non-Gaussian time series.

Appendix A. Theoretical results

A.1 Proof of Theorem 2.1

Proof. The proof is along the lines of the one in Jiang and Tanner (1999b). We start with proving that $\left\|\sum_{k=1}^{K}\chi_{Q_k^K}(\cdot)\hat{h}_k(\cdot)-h(\cdot)\right\| \leq \pi^2 K_0/2K^2$, where, for each $\omega \in Q_k^K$, $\hat{h}_k(\omega) \equiv \hat{\alpha}_k + \hat{\beta}_k\omega$. For $\omega_k^* \in Q_k^K$, $\hat{\alpha}_k = h(\omega_k^*) - \omega_k^* h'(\omega)$ and $\hat{\beta}_k = h'(\omega)$. We have that $\left\|\sum_{k=1}^{K}\chi_{Q_k^K}\hat{h}_k(\cdot)-h(\cdot)\right\|_p = \left\|\sum_{k=1}^{K}\chi_{Q_k^K}(\hat{h}_k-h(\cdot))\right\|_p \leq \sup_{1\leq k\leq K}\left\|\hat{h}_k(\cdot)-h(\cdot)\right\|_\infty \left\|1\sum_{k=1}^{K}\chi_{Q_k^K(\cdot)}\right\|_p = \sup_{1\leq k\leq K}\left\|\hat{h}_k(\cdot)-h(\cdot)\right\|_\infty$. Considering a second order expansion of $h(\omega)$ around ω_k^* and the definition of $\hat{h}(\omega)$, we have that $|\hat{h}_k(\omega)-h(\omega)| \leq |(\omega-\omega_k^*)^2 h''(\omega_k^*)|/2 \leq \pi^2 K_0/2K^2$. Therefore, $\left\|\sum_{k=1}^{K}\chi_{Q_k^K(\cdot)}\hat{h}_k(\cdot)-h(\cdot)\right\| \leq \pi^2 K_0/2K^2$.

Now, $\left\|\sum_{k=1}^{K}g_k(\cdot;\lambda,\zeta,\phi)\hat{h}_k(\cdot)-h(\cdot)\right\|_p \leq \left\|\sum_{k=1}^{K}(g_k(\cdot;\lambda,\zeta,\phi)-\chi_{Q_k^K}(\cdot))\hat{h}_k(\cdot)\right\|_p + \left\|\sum_{k=1}^{K}\chi_{Q_k^K}(\cdot)\hat{h}_k(\cdot)-h(\cdot)\right\|_p \leq \sum_{k=1}^{K}\left\|(g_k(\cdot;\lambda,\zeta,\phi)-\chi_{Q_k^K}(\cdot))\right\|_p \left\|\hat{h}_k(\cdot)\right\|_\infty + \pi^2 K_0/2K^2 \leq K\epsilon(1+\pi)K_0 + \pi^2 K_0/2K^2$. The last step follows from the Lemma, and from the fact that $|\hat{h}_k| \leq |h(\omega_k^*)| + |h'(\omega)(\omega-\omega_k^*)| \leq K_0 + \pi K_0$. Letting ϵ tend to zero, we obtain $\inf_{h_K}\|h_K - h(\cdot)\|_p \leq \pi^2 K_0/2K^2$. □

A.2 Proof of Lemma 2.2

Proof. Based on the form of the mixture weights in 3.1, for any fixed ω, and for any $k = 1, ..., K$,we have

$$\lim_{\tau\to\infty} g_k(\mu(\omega), \tau) = \begin{cases} 1 & \log\left(\frac{k-1}{K-k+1}\right) \leq \mu(\omega) < \log\left(\frac{k}{K-k}\right) \\ 0 & o.w. \end{cases}$$

and thus

$$\lim_{\tau\to\infty} g_k(\mu(\omega), \tau) = \begin{cases} 1 & \frac{k-1}{K} \leq \frac{\exp(\mu(\omega))}{1+\exp(\mu(\omega))} < \frac{k}{K} \\ 0 & o.w. \end{cases}$$

We can find values of ζ and $\phi > 0$, and integers k_1 and k_2, with $k_2 > k_1$, such that $\exp(\mu(0))/\{1 + \exp(\mu(0))\} = k_1/K$ and $\exp(\mu(1))/\{1 + \exp(\mu(1))\} = k_2/K$, and such that we can build a linear approximation of the logistic function $\exp(\mu(\omega))/\{1 + \exp(\mu(\omega))\}$, given by $(k_1/K)+\{(k_2-k_1)/K\}\omega$. Therefore, the partition induced on Ω is $\{(0, 1/(k_2 - k_1)), [1/(k_2 - k_1), 2/(k_2 - k_1)), ..., [(k_2-k_1-1)/(k_2-k_1), 1)\}$. From the limiting result above, for $k = k_1+1, ..., k_2$, we have $g_k(\mu(\omega), \tau) \to \chi_{Q_k}(\omega)$, almost surely, as $\tau \to \infty$. In addition, for $0 < k \leq k_1$ or $k_2 < k \leq K$, $g_k(\mu(\omega), \tau) \to 0$, almost surely, as $\tau \to \infty$. Moreover, for $k = k_1 + 1, ..., k_2$, $|g_k(\mu(\omega), \tau) - \chi_{Q_k}(\omega)|^p \leq 1$, for $\omega \in (0, 1)$, and for $0 < k \leq k_1$ or $k_2 < k \leq K$, $|g_k(\mu(\omega), \tau)|^p \leq 1$, for $\omega \in (0, 1)$. Hence, from the dominated convergence theorem, for $k = k_1+1, ..., k_2$, $\lim_{\tau\to\infty} ||g_k-\chi_{Q_k}||_p = 0$, for any $p \in \mathbb{N}$. Finally, for $1 < k \leq k_1$ or $k_2 < k \leq K$, we have that

$\lim_{\tau\to\infty} ||g_k||_p = 0$, for any $p \in \mathbb{N}$. □

A.3 Proof of Theorem 2.2

Proof. We start by proving that, for fixed K, k_1 and k_2, with $k_2 > k_1$, any $h \in W^{\infty}_{2,K_0}$ can be approximated by a piecewise linear function on the partition $\{Q_{k_1+1}, ..., Q_{k_2}\}$, with the L_p distance bounded by a constant that depends on $K^* = k_2 - k_1$. For each interval Q_k, $k = k_1 + 1, ..., k_2$, consider a point $\omega_k^* \in Q_k$ and the linear approximation based on the first-order Taylor series expansion: $\hat{h}_k(\omega) = \hat{\alpha}_k + \hat{\beta}_k\omega$, for $\omega \in Q_k$, where $\hat{\alpha}_k = h(\omega_k^*) - \omega_k^* h'(\omega_k^*)$ and $\hat{\beta}_k = h'(\omega_k^*)$; here, $h'(\omega_k^*)$ denotes the first derivative of $h(\omega)$ evaluated at ω_k^*, with similar notation used below for the second derivative. We have $\left\| \left\{ \sum_{k=k_1+1}^{k_2} \chi_{Q_k} \hat{h}_k \right\} - h \right\|_p$ $= \left\| \sum_{k=k_1+1}^{k_2} \chi_{Q_k} \left\{ \hat{h}_k - h \right\} \right\|_p \leq \sup_{k_1+1 \leq k \leq k_2} \left\| \hat{h}_k - h \right\|_{\infty}$, where $\|\|_{\infty}$ denotes the L_∞ norm. Now, for each interval Q_k, we consider the second-order expansion of h around the same $\omega_k^* \in Q_k$. Note that the partition $\{Q_{k_1+1}, ..., Q_{k_2}\}$ satisfies the property that, for any k, and for any ω_1 and ω_2 in Q_k, $|\omega_1 - \omega_2| \leq 1/K^*$. Using this property and the fact that the second derivative of h is bounded by K_0, we obtain $|\hat{h}_k(\omega) - h(\omega)| \leq |0.5(\omega - \omega_k^*)^2 h''(\omega_k^*)| \leq K_0/(2K^{*2})$. Therefore, $\left\| \left\{ \sum_{k=k_1+1}^{k_2} \chi_{Q_k} \hat{h}_k \right\} - h \right\|_p \leq K_0/(2K^{*2})$. Using the triangular inequality, we can write

$$\left\| \left\{ \sum_{k=k_1+1}^{k_2} g_k \hat{h}_k \right\} - h \right\|_p \leq \left\| \sum_{k=k_1+1}^{k_2} \{g_k - \chi_{Q_k}\} \hat{h}_k \right\|_p + \left\| \left\{ \sum_{k=k_1+1}^{k_2} \chi_{Q_k} \hat{h}_k \right\} - h \right\|_p .$$

Based on the previous result, the second term is bounded by $K_0/(2K^{*2})$. For the first term, $\left\|\sum_{k=k_1+1}^{k_2}\{g_k-\chi_{Q_k}\}\hat{h}_k\right\|_p \leq \sum_{k=k_1+1}^{k_2}\|g_k-\chi_{Q_k}\|_p\left\|\hat{h}_k\right\|_\infty$. Using Lemma 1 and the fact that $|\hat{h}_k(\omega)| \leq |h(\omega_k^*)| + |h'(\omega_k^*)(\omega-\omega_k^*)| \leq 2K_0$, we have that the first term is bounded by $2\epsilon K^* K_0$, for any $\epsilon > 0$ given sufficiently large τ. Finally, $\left\|\left\{\sum_{k=1}^{K} g_k\hat{h}_k\right\} - h\right\|_p \leq 2\epsilon K^* K_0 + \{K_0/(2K^{*2})\}$, and letting ϵ tend to zero, we obtain the result. □

A.4 Proof of Lemma 3.1

Proof. If $\log f_k \xrightarrow{L_p(0,\pi)} \log f$, then $f_k \xrightarrow{L_p(0,\pi)} f$, for any $1 \leq p < \infty$, because the exponential transformation preserves the L_p convergence on a set of finite measure. Assume, without loss of generality, that $\int_0^\pi f(\omega)\mathrm{d}\omega \neq 0$. We need to prove that $\int_0^\pi f_k(\omega)\mathrm{d}\omega \to \int_0^\pi f(\omega)\mathrm{d}\omega$. We have that

$$\left|\int_0^\pi f(\omega)\mathrm{d}\omega\right| - \int_0^\pi |f_k(\omega)|\mathrm{d}\omega \leq \int_0^\pi |f(\omega)-f_k(\omega)|\mathrm{d}\omega = ||f-f_k||_{L_1}.$$

The last term tends to zero based on Holder's inequality. Recall that, if we have a sequence of constants c_k, such that $c_k \to c$ and a sequence of functions f_k, such that $f_k \xrightarrow{L_p(0,\pi)} f$, then $c_k f_k \xrightarrow{L_p(0,\pi)} cf$. Setting $c_k^{-1} = \int_0^\pi f_k(\omega)\mathrm{d}\omega$ and $c^{-1} = \int_0^\pi f(\omega)\mathrm{d}\omega$, we obtain $f_k/\int_0^\pi f_k(\omega)\mathrm{d}\omega \xrightarrow{L_p(0,\pi)} f/\int_0^\pi f(\omega)\mathrm{d}\omega$. Again, from Holder's inequality, L_p convergence implies L_1 convergence, which is equivalent to convergence in the total variation distance. □

Appendix B. MCMC details

B.1 MCMC algorithm for one time series with logistic weights

We generate samples from the posterior distribution of the model in (**??**), using an MCMC algorithm with Gibbs sampling steps and one Metropolis step for λ. The MCMC algorithm comprises the following updates:

– For $j = 1, \ldots, N$, sample r_j from a discrete distribution on $\{1, \ldots, K\}$ with probabilities proportional to $g_k(\omega_j; \lambda, \zeta, \phi)\mathrm{N}(y_j \mid \alpha_k + \beta_k\omega_j, \sigma^2)$, for $k = 1, ..., K$.

– For $k = 1, \ldots, K$, sample $(\alpha_k, \beta_k)'$ from a bivariate normal distribution with covariance matrix $\Sigma^* = (\sigma^{-2}\sum_{\{j:r_j=k\}} z_j z_j' + \Sigma_0^{-1})^{-1}$ and mean $\mu^* = \Sigma^*(\Sigma_0^{-1}\mu_0 + \sum_{\{j:r_j=k\}} y_j z_j)$. Here, $z_j = (1, \omega_j)'$, $\mu_0 = (\mu_\alpha, \mu_\beta)'$ is the prior mean, and Σ_0 the diagonal prior covariance matrix with diagonal elements σ_α^2 and σ_β^2.

– Sample σ^2 from an inverse-gamma with parameters $n_0 + N/2$ and $d_0 + \{\sum_{j=1}^{N}(y_j - (\alpha_{r_j} + \beta_{r_j}\omega_j))^2\}/2$, where n_0 and d_0 are the parameters of the inverse-gamma prior.

– In order to sample $(\zeta_k, \phi_k)'$, for $k = 1, \ldots, K$, we use a data augmentation technique. For each k, we introduce N Pólya-Gamma random variables, q_j, $j = 1, ..., N$, as defined in Polson et al. (2013). For $k = 1, ..., K$, let $\psi_{kj} = (\zeta_k + \phi_k\omega_j)/\lambda$ and $c_{kj} = \log(\sum_{\{i \neq k\}} \exp\{(\zeta_i + \phi_i\omega_j)/\lambda\})$. Moreover, let n_k be the number of r_j such that $r_j = k$. Then, the q_j are drawn from a Pólya-Gamma distribution with parameters 1 and $\psi_{kj} - c_{kj}$. And, given the q_j, we sample $(\zeta_k, \phi_k)'$ from a bivariate normal with covariance matrix $\Sigma_k^* = \left(\mathbb{I} + \sum_{j=1}^{N} \lambda^{-2} z_j z_j' q_j\right)^{-1}$, where $\mathbb{I}$ is the identity matrix, and mean $\mu_k^* = \Sigma_k^*(\sum_{\{j:r_j=k\}} \lambda^{-1} z_j c_{kj} q_j + \sum_{\{j:r_j=k\}} 0.5\lambda^{-1} z_j - \sum_{\{j:r_j \neq k\}} 0.5\lambda^{-1} z_j)$.

– Update λ with a Gaussian random-walk Metropolis step on $\log(\lambda)$.

B.2 MCMC algorithm for Bayesian hierarchical model for multiple time series

Here, we present the details of the Gibbs sampler that can be used for posterior simulation from the hierarchical model developed in Section 3.2.

The full conditional distribution for each configuration variable r_{mj}, $m = 1, \ldots M$, $j = 1, \ldots, N$, is a piecewise Gaussian distributed on $[\log((k-1)/(K-$

$k+1)), \log(k/(K-k))]$ with weights

$$w_k = \frac{g_k(\mu_m(\omega_j), \tau_m)\mathrm{N}(y_{mj} \mid \alpha_k + \beta_k\omega_j)}{\sum_{i=1}^{K} g_i(\mu_m(\omega_j), \tau_m)\mathrm{N}(y_{mj} \mid \alpha_i + \beta_i\omega_j)},$$

for $k = 1, \ldots, K$.

We sample (α_k, β_k) jointly, for $k = 1, \ldots, K$. Let $\mu = (\mu_\alpha, \mu_\beta)'$ and Σ_0 the diagonal matrix that has σ_α^2 and σ_β^2 as diagonal terms. The full conditional distribution for (α_k, β_k) is a bivariate normal with covariance matrix $\Sigma^* = \sigma^2 \left(\sum_{m,j: \frac{k-1}{K} < \frac{\exp(r_{mj})}{1+\exp(r_{mj})} \leq \frac{k}{K}} z_j z_j' + \Sigma_0^{-1} \right)^{-1}$ and mean vector $\mu^* = \Sigma^* \left(\Sigma_0^{-1}\mu_0 + \sum_{m,j: \frac{k-1}{K} < \frac{\exp(r_{mj})}{1+\exp(r_{mj})} \leq \frac{k}{K}} y_{mj} z_j \right)$, where $z_j = (1, \omega_j)'$.

We sample (ζ_m, ϕ_m) jointly, for $m = 1, ..., M$. The full conditional distribution is a bivariate normal with covariance matrix $\Sigma_w^* = \left(\sigma^{-2} \sum_{j=1}^{N} q_j q_j' + \Sigma_w^{-1} \right)^{-1}$ and mean $\mu_w^* = \Sigma_w^* \left(\sum_{j=1}^{N} r_{mj} q_j + \Sigma_w^{-1}\mu_w \right)$, where $q_j = (1, \omega_j)'$.

The full conditional distribution for the common variance parameter σ^2 follows an inverse-gamma distribution with updated parameters n^* and d^*, where $n^* = n_{\sigma^2} + NM/2$ and $d^* = d_{\sigma^2} + 2^{-1} \sum_{m=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{K} (y_{mj} - (\alpha_k + \beta_k\omega_j))^2) \mathbb{I}\left(k - 1/K < \exp(r_{mj})/(1+\exp(r_{mj})) \leq k/K\right)$.

The full conditional for τ_m, $m = 1, ...M$ is gamma with parameters $n_\tau + N/2$ and $d_\tau + \sum_{j=1}^{N} (r_{mj} - (\zeta_m + \phi_m\omega_j))^2/2$.

The full conditional for d_τ is a gamma with parameters $a_{d_\tau} + Mn_\tau$ and $b_{d_\tau} + M \sum_{m=1}^{M} \tau_m$, where a_{d_τ} and a_{d_τ} are the parameters of the hyperprior.

The full conditional for μ_w is a bivariate normal with covariance matrix $\Sigma_0^* = (\Sigma_{00} + M\Sigma_w)^{-1}$, and mean $\mu_0^* = \Sigma_0^* [\Sigma_{00}^{-1}\mu_{00} + \Sigma_w^{-1} \sum_{m=1}^{M} (\zeta_m, \phi_m)']$, where

μ_{00} is the hyperprior mean and Σ_{00} the hyperprior covariance matrix.

The full conditional for Σ_w is an inverse Wishart with $\nu_0 + M$ degrees of freedom and scale matrix $\Psi + \sum_{m=1}^{M}[(\zeta_m, \phi_m)' - \mu_w][(\zeta_m, \phi_m)' - \mu_w]'$, where ν_0 are the hyperprior degrees of freedom and Ψ is the hyperprior scale matrix.

B.3 MCMC algorithm for Bayesian hierarchical model for multiple non-stationary time series

Here, we present the details of the Gibbs sampler that can be used for posterior simulation from the hierarchical model developed in Section 4.4.

The full conditional distribution for each configuration variable r_{mij}, $m = 1, \dots M$, $i = 1, \dots S$, $j = 1, \dots, N$, is a piecewise Gaussian distributed on $[\log((k-1)/(K-k+1)), \log(k/(K-k))]$ with weights

$$w_k = \frac{g_k(\mu_m(\bar{u}_i, \omega_j), \tau_m)\mathrm{N}(y_{mij} \mid \alpha_k + \beta_k\omega_j + \gamma_k\bar{u}_i + \delta_k\bar{u}_i\omega_j)}{\sum_{l=1}^{K} g_l(\mu_m(\bar{u}_i, \omega_j), \tau_m)\mathrm{N}(y_{mij} \mid \alpha_l + \beta_l\omega_j + +\gamma_l\bar{u}_i\delta_l\bar{u}_i\omega_j)},$$

for $k = 1, \dots, K$.

We sample $(\alpha_k, \beta_k, \gamma_k, \delta_k)$ jointly, for $k = 1, \dots, K$. Let $\mu = (\mu_\alpha, \mu_\beta, \mu_\gamma, \mu_\delta)'$ and Σ_0 the diagonal matrix that has σ_α^2, σ_β^2, σ_γ^2 and σ_δ^2 as diagonal terms. The full conditional distribution for $(\alpha_k, \beta_k, \gamma_k, \delta_k)$ is a four dimensional normal with covariance matrix $\Sigma^* = \sigma^2\left(\sum_{m,j:\frac{k-1}{K} < \frac{\exp(r_{mij})}{1+\exp(r_{mij})} \leq \frac{k}{K}} z_{ij}z_{ij}' + \Sigma_0^{-1}\right)^{-1}$ and mean vector $\mu^* = \Sigma^*\left(\Sigma_0^{-1}\mu_0 + \sum_{m,i,j:\frac{k-1}{K} < \frac{\exp(r_{mij})}{1+\exp(r_{mij})} \leq \frac{k}{K}} y_{mij}z_{ij}\right)$, where $z_{ij} = (1, \omega_j, \bar{u}_i, \bar{u}_i\omega_j)'$.

We sample $(\zeta_m, \phi_m, \eta_m, \psi_m)$ jointly, for $m = 1, ..., M$. The full conditional distribution is a four dimensional normal with covariance matrix $\Sigma_w^* = \left(\sigma^{-2} \sum_{i=1}^{S} \sum_{j=1}^{N} q_{ij} q_{ij}' + \Sigma_w^{-1}\right)^{-1}$ and mean $\mu_w^* = \Sigma_w^* \left(\sum_{i=1}^{S} \sum_{j=1}^{N} r_{mj} q_{ij} + \Sigma_w^{-1} \mu_w\right)$, where $q_{ij} = (1, \omega_j, \bar{u}_i, \bar{u}_i \omega_j)'$.

The full conditional distribution for the common variance parameter σ^2 follows an inverse-gamma distribution with updated parameters n^* and d^*, where $n^* = n_{\sigma^2} + NMS/2$ and $d^* = d_{\sigma^2} + 2^{-1} \sum_{m=1}^{M} \sum_{i=1}^{S} \sum_{j=1}^{N} \sum_{k=1}^{K} (y_{mij} - (\alpha_k + \beta_k \omega_j + \gamma_k \bar{u}_i + \delta_k \bar{u}_i \omega_j))^2) \mathbb{I}\left(k - 1/K < \exp(r_{mij})/(1 + \exp(r_{mij})) \leq k/K\right)$.

The full conditional for τ_m, $m = 1, ...M$ is gamma with parameters $n_\tau + N/2$ and $d_\tau + \sum_{i=1}^{S} \sum_{j=1}^{N} (r_{mij} - (\zeta_m + \phi_m \omega_j + \eta_m \bar{u}_i + \psi_m \bar{u}_i \omega_j))^2/2$.

The full conditional for d_τ is a gamma with parameters $a_{d_\tau} + M n_\tau$ and $b_{d_\tau} + M \sum_{m=1}^{M} \tau_m$, where a_{d_τ} and a_{d_τ} are the parameters of the hyperprior.

The full conditional for μ_w is a four dimensional normal with covariance matrix $\Sigma_0^* = (\Sigma_{00} + M\Sigma_w)^{-1}$, and mean $\mu_0^* = \Sigma_0^* [\Sigma_{00}^{-1} \mu_{00} + \Sigma_w^{-1} \sum_{m=1}^{M} (\zeta_m, \phi_m, \eta_m, \psi_m)']$, where μ_{00} is the hyperprior mean and Σ_{00} the hyperprior covariance matrix.

The full conditional for Σ_w is an inverse Wishart with $\nu_0 + M$ degrees of freedom and scale matrix $\Psi + \sum_{m=1}^{M} [(\zeta_m, \phi_m, \eta_m, \psi_m)' - \mu_w][(\zeta_m, \phi_m, \eta_m, \psi_m)' - \mu_w]'$, where ν_0 are the hyperprior degrees of freedom and Ψ is the hyperprior scale matrix.

Bibliography

Carter, C. K. and Kohn, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society, Series B*, 59(1):255–268.

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

Contreras-Cristan, A., Gutierrez-Pena, E., and Walker, S. G. (2006). A note on Whittle's likelihood. *Communications in Statistics - Simulation and Computation*, 35(4):857–875.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1):1–37.

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

Fryzlewicz, P., Van Bellegem, S., and von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, 55(4):737–764.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability*, 10(1):130–145.

Jiang, W. and Tanner, M. A. (1999a). Hierarchical mixtures-of-experts for exponential family regression models: Approximation and maximum likelihood estimation. *The Annals of Statistics*, 27:987–1011.

Jiang, W. and Tanner, M. A. (1999b). On the approximation rate of hierarchical mixtures-of-experts for generalized linear models. *Neural computation*, 11(5):1183–1198.

Krystal, A., Prado, R., and West, M. (1999). New methods of time series analysis of non-stationary EEG data: eigenstructure decompositions of autoregressions. *Clinical Neurophysiology*, 110:2197–2206.

Macaro, C. and Prado, R. (2014). Spectral decompositions of multiple time series: a Bayesian non-parametric approach. *Psychometrika*, 79:105–129.

Naito, T., Asai, K., Amano, T., and Taniguchi, M. (2010). Local whittle likelihood estimators and tests for non-gaussian stationary processes. *Statistical Inference for Stochastic Processes*, 13(3):163–174.

Norets, A. (2010). Approximation of conditional densities by smooth mixtures of regressions. *The Annals of Statistics*, 38:1733–1766.

Ombao, H., Raz, J., von Sachs, R., and Guo, W. (2002). The SLEX model of a non-stationary random process. *Annals of the Institute of Statistical Mathematics*, 54(1):171–200.

Ombao, H., Raz, J., von Sachs, R., and Malow, B. (2001). Automatic statistical analysis of bivariate non-stationary time series. *Journal of the American Statistical Association*, 96:543–560.

Ombao, H., von Sachs, R., and Guo, W. (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, 100(470):519–531.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.

Pawitan, Y. and O'Sullivan, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association*, 89(426):600–610.

Pensky, M., Vidakovic, B., and DeCanditiis, D. (2007). Bayesian decision theoretic scale-adaptive estimation of a log-spectral density. *Statistica Sinica*, 17:635–666.

Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26:373–393.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349.

Prado, R., West, M., and Krystal, A. (2001). Multi-channel EEG 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. (1982). *Spectral analysis and time series*. Number v. 1-2 in Probability and mathematical statistics. Academic Press.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(2):204–237.

Rosen, O. and Stoffer, D. (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika*, 94:335–345.

Rosen, O., Stoffer, D. S., and Wood, S. (2009). Local spectral analysis via a bayesian mixture of smoothing splines. *Journal of the American Statistical Association*, 104(485):249–262.

Rosen, O., Wood, S., and Stoffer, D. S. (2012). Adaptspec: Adaptive spectral

estimation for nonstationary time series. *Journal of the American Statistical Association*, 107(500):1575–1589.

Shao, X. and Wu, W. B. (2007). Local whittle estimation of fractional integration for nonlinear processes. *Econometric Theory*, 23(5):899–929.

Shumway, R. H. and Stoffer, D. S. (2011). *Time series analysis and its applications: with R examples*. Springer.

Trejo, L. J., Knuth, K., Prado, R., Rosipal, R., Kubitz, K., Kochavi, R., Matthews, B., and Zhang, Y. (2007). EEG-based estimation of mental fatigue: Convergent evidence for a three- state model. In Schmorrow, D. and Reeves, L., editors, *Augmented Cognition, HCII 2007, LNAI 4565*, pages 201–211, New York. Springer LNCS.

Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, 75(369):122–132.

West, M., Prado, R., and Krystal, A. (1999). Evaluation and comparison of EEG traces: Latent structure in nonstationary time series. *Journal of the American Statistical Association*, 94:1083–1095.

Whittle, P. (1957). Curve and periodogram smoothing. *Journal of the Royal Statistical Society, Series B*, 19:38–63.

Yang, W.-H., Holan, S. H., and Wikle, C. K. (2016). Bayesian lattice filters for

time-varying autoregression and time-frequency analysis. *Bayesian Anal.*, 11(4):977–1003.