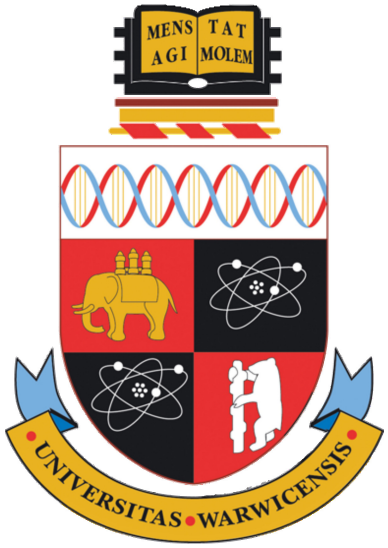

Bayesian Analysis of Nonstationary Periodic Time Series

by

Beniamino Hadj-Amar

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

September 2019

THE UNIVERSITY OF WARWICK

Sky’s the Limit

in memory of Kento, Adriana and Gabriele

Contents

Acknowledgments

It has been a great journey. Time has gone by so quickly and I realize this is often the case when you put passion, love and dedication into trying to achieve your goals. Still, I would never have imagined that pursuing a PhD in Statistics would have ended up being such an enjoyable and rewarding experience.

A special thought is dedicated to Bärbel Finkenstädt, my supervisor, who has looked after me over these last four years. She has taught me how to be a good researcher by constantly providing valuable feedback and guiding me along the right path during this doctoral experience. She provided light and inspiration for me, in the good and the bad moments. She has been patient and passionate. I am extremely proud to have spent all this time working with such an outstanding academic.

I want to thank Eva Riccomagno who five years ago recommended that I begin my adventure in England, and without whom this experience abroad would probably not have been possible. I wish to highlight Paul Jenkins for his exceptional mentoring at the University of Warwick and the insightful and valuable discussions we had regarding Bayesian statistics and Monte Carlo methods. I am very grateful to Mark Fiecas who introduced me to statistical spectral analysis and gave me the opportunity for a research visit at the University of Minnesota. Last, but definitely not the least, I am really thankful to my friends Jack Jewson, Karla Monterrubio, Joe Meagher and Arthur King. They have played an important role in making my graduate life so enjoyable and have always been there for me when I needed them.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. Chapter 2 formed the manuscript for "Bayesian Model Search for Nonstationary Periodic Time Series" which has been published in *Journal of the American Statistical Association*, with Prof. Bärbel Finkenstädt, Dr. Mark Fiecas, Prof. Francis Lévi and Dr. Robert Huckstepp. Chapter 3 is in preparation for submission. All experimental data discussed in this research have been kindly provided by the labs of Prof. Francis Lévi (Warwick Medical School and INSERM U935, Villejuif, France) and Dr. Robert Huckstepp (Life Sciences, University of Warwick).

Abstract

Identifying the periodicities present in a cyclical process allows us to gain knowledge about the sources of variability that drive that phenomenon. For instance, respiratory traces obtained from a plethysmograph used on rodents in experimental sleep apnea research reveal many sudden changes in their periodic features as the rat spontaneously changes its breathing pattern during its sleep-wake activities. Similarly, human temperature, as measured by a wearable sensing device over several days at relatively high temporal resolution (e.g. 5 minutes), may be subject to a different periodic behaviour during the night when the individual transitions between different stages of sleep. While the theory and methods for analyzing the periodicities of time series data are relatively well-developed for the case of stationary time series, the task of modelling time series that undergo regime shifts in periodicity, amplitude and phase remains challenging because the timing of the changes and the relevant periodicities are usually unknown (both in value and number).

This thesis introduces new methodologies for the automated analysis of non-stationary periodic time series. In the first part of this research, we present a novel Bayesian approach for analyzing time series data that exhibit regime shifts in periodicity, amplitude and phase, where we approximate the time series using a piecewise oscillatory model with unknown periodicities, and our goal is to estimate the change-points while simultaneously identifying the changing periodicities in the data. Our proposed methodology is based on a trans-dimensional Markov chain Monte Carlo (MCMC) algorithm that simultaneously updates the change-points

and the periodicities relevant to any segment between them. We show that the proposed methodology successfully identifies time changing oscillatory behaviour in two applications which are relevant to e-Health and sleep research, namely the occurrence of ultradian oscillations in human skin temperature during the time of night rest, and the characterization of instances of sleep apnea in plethysmographic respiratory traces.

In addition to detecting temporal changes, it may also be of interest to recognize the recurrence of a relevant periodic pattern. In the second half of this thesis, we consider periodic phenomena, whose behaviour switches over time, as realizations of a hidden Markov model where the number of states is unknown along with the relevant periodicities, the role of which varies over the different states. Flexibility on the number of states is achieved by using Bayesian nonparametric techniques that address the stochastic switching dynamics of the time series via a hierarchical Dirichlet process that captures the temporal mode persistence of the hidden states. The variable dimensionality regarding the number of periodicities that characterizes the different regimes is addressed by developing an appropriate trans-dimensional MCMC sampler. We illustrate the use of our proposed approach in a case study relevant to respiratory research, namely the detection of recurring instances of sleep apnea in human respiratory traces.

Chapter 1

Introduction

One major purpose of time series analysis is to build statistical models that provide accurate understanding of the underlying structure that gives rise to the observed data. A classic approach for the analysis of time series is via *time domain* analysis which builds statistical models with respect to distinctive temporal characteristics. An alternative strategy to view and examine time series is through *frequency domain* (or *spectral*) analysis, where the aim is to recognize and investigate the periodic or sinusoidal components that are responsible for the variation in the data. Finding the periodicities present in an oscillatory process allows us to get insight into the sources of variability that drive that phenomenon. For example, we analyzed several respiratory traces obtained from rodents in experimental sleep apnea research and observed that they exhibit many abrupt changes in their periodic components as the rat naturally switches its breathing patterns in the course of its sleep-wake activities [Han et al., 2002; Nakamura et al., 2003]. An example of a breathing trace obtained from a freely behaving rat via an unrestrained whole-body *plethysmograph* is shown in Figure 1.1, where we recall that a plethysmograph is a device used for recording and measuring the volume of air displaced between normal inhalation and

exhalation when extra effort is not applied. One main goal of this study is indeed to provide a statistical methodology that can be applied to analyze large breathing data sets resulting from *in vivo* plethysmograph studies in rodents to characterize the occurrence of sleep apnea under different experimental conditions. Sleep apnea is the temporary interruption of breathing during sleep and is linked to many diseases, including cardiovascular events, cancer and diabetes [Lanfranchi et al., 1999; Nieto et al., 2012; Harsch et al., 2004], and detecting (and, eventually, forecasting) all apnea instances that occur while sleeping is in fact one of the primary interests for healthcare providers, experimental researchers and clinicians working in sleep apnea research. We therefore aim to provide them with useful statistical tools to aid the understanding of the pathological implications of this status.

The development of this thesis was also motivated by the recent progress of information and communication technologies in the health care system. To address the issue of personalized medical treatment according to the circadian timing system of the patient, referred to as *chronotherapy* [Lévi and Schibler, 2007], a non-invasive mobile e-Health platform pioneered by the French project PiCADo (Komarzynski et al. 2018) is used to record and teletransmit skin surface temperature as well as physical activity data [Huang et al., 2018] from a wearable sensing device located on the chest. Figure 1.1 displays an example of 4 days of 5-min skin temperature recording for a healthy individual. This dense physiological signal seems to exhibit periodicities whose role changes over time in a more or less abrupt manner. A natural question to ask is whether in addition to the circadian (24 hours) periodicity that can be expected, this person may be subject to an additional ultradian periodic behaviour during the night when the individual transitions between different sleep stages [Carskadon et al., 2005; Komarzynski et al., 2018].

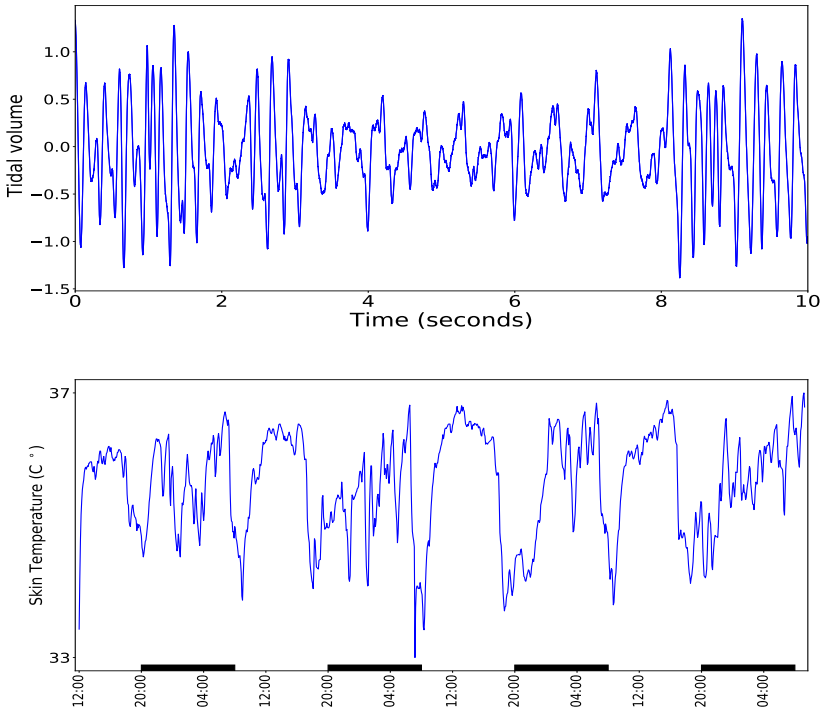

Figure 1.1: (Top) Plethysmographic breathing trace of a rodent characterized by an alternation of sniffing and normal breathing. The time series displayed contains 20,000 observations where the signal was sampled at 2000 Hz. (Bottom) Skin surface temperature of a healthy subject recorded every 5 minutes over 4 days. Block bars on the time axis mark the evening and night intervals from 20.00 to 8.00.

1.1 From Time Domain to Frequency Domain

We define a time series to be a collection of random variables $\{Y_t\}_{t\in\mathbb{Z}}$ indexed by the sequential order in time. The behaviour of such time series processes are usually assumed to be weakly stationary. Generally speaking, weakly stationary processes are characterized by the property that their statistical features do not change over time. More formally, the definition of weakly stationarity is obtained by restricting attention to those properties that depend only on the first and second-order moments of Y_t [Priestley, 1981; Brockwell et al., 1991] which are summarized by the mean

function $\mu(t) = \mathbb{E}(Y_t)$ and the autocovariance function $\gamma(h)$, respectively, where the autocovariance function is defined as

$$\gamma(h) = \mathrm{Cov}(Y_t, Y_{t+h}) = \mathbb{E}\Big[(Y_t - \mathbb{E}(Y_t))(Y_{t+h} - \mathbb{E}(Y_{t+h}))\Big], \quad h \in \mathbb{Z}, \tag{1.1}$$

and we assume that $\mathbb{E}(Y_t^2) < \infty$, i.e. the second moment of Y_t exists and is finite for all t. Note that $\gamma(h)$ is an even function, namely $\gamma(h) = \gamma(-h)$. Then, a time series process is said to be *weakly* stationary if the mean function $\mu(t)$ is constant and does not dependent on time t, and the autocovariance function $\gamma(h)$ is independent of t for each h. A stronger definition of stationarity is given by *strict* stationarity, where it is assumed that the time series $(Y_1, \ldots, Y_n)$ and the time-shifted time series $(Y_{1+h}, \ldots, Y_{n+h})$ have the same joint distribution, for all integers h and $n > 0$. Throughout the thesis, we shall use the term *stationary* to indicate weak stationarity.

The autocovariance function defined in Equation (1.1) is a measure of the degree of dependence between values of a time series at distinct time lags. However, in many cases the observed time series are composed of many periodic components, and it may be of interest to determine how many and which frequencies drive the variation in the data. Indeed, recognizing the periodicities that occur in a periodic time series allow us to get an understanding about the causes of variability that drive that time series. While the autocovariance function may be used to identify some simple periodic behaviour, spectral analysis provides a more powerful tool to decompose the variability of a time series according to the underlying periodicities.

1.2 The Spectral Density Function

Let us consider a zero-mean stationary process $\{Y_t\}_{t\in\mathbb{Z}}$ with autocovariance function $\gamma(h)$. Let us also assume that the autocovariance function is absolutely summable, namely it satisfies $\sum_{h\in\mathbb{Z}}|\gamma(h)|<\infty$. Then, the *spectral density* function $f(\omega)$ - or *power spectrum* - of the process $\{Y_t\}$ is defined as the Fourier transform of the autocovariance function [Priestley, 1981]

$$f(\omega)=\sum_{h\in\mathbb{Z}}\gamma(h)\,\exp(-i2\pi\omega h),\quad -0.5<\omega<0.5, \tag{1.2}$$

where $\exp(i\omega)=\cos(\omega)+i\sin(\omega)$ and $i=\sqrt{-1}$. The spectral density measures how much of the variation in the process is driven by the oscillations at each frequency ω. It is an even function of the frequencies and has period one, that is $f(\omega)=f(1-\omega)$. Hence, we usually analyze and visualize the spectral density for frequencies in the interval $(0,0.5)$. Moreover, the autocovariance function can be expressed as [Brockwell et al., 1991]

$$\gamma(h)=\int_{-0.5}^{0.5}\exp(i2\pi\omega h)f(\omega)\,d\omega. \tag{1.3}$$

Equations (1.2) and (1.3) suggest that time and frequency domain are linked to each other via a duality between the spectral density and the autocovariance function. Indeed, this unique relationship between $\gamma(h)$ and $f(\omega)$ guarantees that no information is lost when considering second-order features of Y_t via the spectral density $f(\omega)$ instead of the autocovariance function $\gamma(h)$. Note also that the variance of the process Y_t (expressed as the autocovariance evaluated at lag $h=0$) is equal to the integrated spectral density, i.e. $\mathrm{Var}(Y_t)=\gamma(0)=\int_{-0.5}^{0.5}f(\omega)\,d\omega$, and hence the spec-

tral density provides a variance decomposition of a time series process according to the contribution of each frequency.

1.2.1 Spectral Estimation

Power spectral estimation can be achieved first by mapping the data to the frequency domain via the discrete Fourier transform (DFT) and then obtaining a statistic called the *periodogram* [Brillinger, 1981], which provides an unbiased estimator of the spectral density function, but is not consistent since its variance does not decrease as the sample size gets larger.

1.2.1.1 The Periodogram

Suppose that Y_t is a zero-mean stationary time series with spectral density $f(\omega)$. Given realization $\boldsymbol{y} = (y_1, \ldots, y_n)'$, the DFT of the data is defined as

$$d(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t \exp(-i2\pi\omega_j t), \tag{1.4}$$

for $j = 0, \ldots, n-1$, where the $\omega_j = \dfrac{j}{n}$ are referred to as the *Fourier frequencies* and the $d(\omega_j)$ are the corresponding *Fourier coefficients*. The periodogram of the data at frequency ω_j is obtained by taking the square of the absolute value of the Fourier coefficient, namely

$$I(\omega_j) = |d(\omega_j)|^2 = \frac{1}{n}\left| \sum_{t=1}^{n} y_t \exp(-i2\pi\omega_j t) \right|^2. \tag{1.5}$$

Asymptotic properties of the periodogram are well studied and summarized, for example, in Priestley [1981] and Percival et al. [1993]. Here, we recall that for a large sample size and under some regularity conditions, the periodogram ordinates

$I(\omega_j)$ are approximately uncorrelated at distinct frequencies and are asymptotically distributed as

$$I(\omega_j) \xrightarrow{D} f(\omega_j)\,\varepsilon_j, \tag{1.6}$$

where the ε_j are exponentially distributed random variables with mean one, and we assume $\omega_j \neq \{0, 0.5\}$. If the time series process is also Gaussian, then the above statement is valid for any sample size [Shumway and Stoffer, 2017]. The periodogram is an unbiased estimator for the spectral density, since it can be shown that $\mathbb{E}[I(\omega_j)] = f(\omega_j)$. However, it is not a consistent estimator as it can also be shown that the variance $\mathrm{Var}[I(\omega_j)] = f^2(\omega_j)$ does not vanish as the sample size increases [Priestley, 1981].

1.2.1.2 Smooth Spectral Estimates

Approaches to obtain consistent estimates of the spectral density were proposed by several authors and can be broadly grouped in two categories: parametric and nonparametric.

Nonparametric methods were first developed by Bartlett [1950] and Welch [1967]. They proposed to split the time series into a fixed number of segments and obtain an estimate of the spectral density by averaging the periodogram ordinates across those segments. Other procedures based on spectral windows were investigated by Blackman and Tukey [1958] and Parzen [1961]. Generally, a windowed spectral estimate can be expressed as

$$\widehat{f_b}(\omega_j) = \sum_{k=-b}^{k=b} W_b(k)\, I\left(\frac{j+k}{n}\right), \tag{1.7}$$

where $I(\cdot)$ is the periodogram defined in Equation (1.5), $2b + 1$ is the window width (or *bandwidth*), and $W_b(k)$ are symmetric nonnegative weights that satisfy

$\sum_{k=-b}^{k=b} W_b(k) = 1$. The basic idea behind spectral windowing is that we can derive a consistent estimator of the spectral density $f(\omega)$ by applying weighted local smoothing of the periodogram estimates in a small frequency interval containing ω. However, the performance of these methods is strongly affected by the correct tuning of the bandwidth which is indeed crucial for balancing the trade-off between variance reduction and periodogram resolution [Percival et al., 1993].

Equation (1.6) suggests that estimating the spectral density can be seen as a nonparametric regression problem with the observations being the log-periodogram ordinates, the unknown regression curve being the logarithm of the spectral density and the innovations having exponential distributions. Under this framework, Cogburn and Davis [1974] proposed fitting a smoothing spline to the log periodogram. This work was improved by Wahba [1980] who placed a smoothness prior on the logarithm of the spectral density and developed a data-driven criterion for estimating the smoothing parameter. Carter and Kohn [1997] extended this approach to a Bayesian setting where posterior inference is carried out efficiently using Markov chain Monte Carlo (MCMC) methods and the estimation of the spectral density is made computationally tractable by using the Whittle approximation to the Gaussian likelihood. In this regard, Whittle [1953, 1957] showed that for large n, the likelihood of the data $\boldsymbol{y}$ given the spectral density $\boldsymbol{f} = (f(\omega_0), \ldots, f(\omega_{n-1}))$ can be approximated as

$$\mathcal{L}_W(\boldsymbol{y} \,|\, \boldsymbol{f}) = (2\pi)^{-n/2} \prod_{j=0}^{n-1} \exp\left\{ -\frac{1}{2}\Big[\log f(\omega_j) + I(\omega_j)/f(\omega_j) \Big] \right\}, \tag{1.8}$$

where $\mathcal{L}_W$ is referred to as the *Whittle's likelihood*. It should be noted that $\mathcal{L}_W$ is only a large sample approximation to the true likelihood and its plausibility depends on the data being satisfactorily close to a stationary regime and on the sam-

ple size being large enough especially when data appear non-Gaussian. Bayesian procedures for nonparametric spectral analysis that makes use of the Whittle likelihood were also developed by Choudhuri et al. [2004], who placed a Bernstein polynomial prior [Petrone and Wasserman, 2002] on the normalized spectral density. Bayesian nonparametric approaches for the estimation of the spectral density that are based on the true likelihood, without resorting to Whittle's approximation, were explored by Rousseau et al. [2012] and Kruijer et al. [2013]. They focused on studying the theoretical properties of their approach - asymptotic, consistency and rates of convergence - for a general class of priors and apply their results to the family of fractionally exponential priors [Robinson, 1991], the latter being a spectral representation of a process which separates the long-range behavior and the short-memory part of a time series. Practical implementation and computational aspects of such a modeling approach are discussed in Chopin et al. [2013], where they propose to draw from the approximate posterior, namely the prior times the approximate likelihood, via a sequential Monte Carlo sampler [Del Moral et al., 2006] and then to recover the exact posterior through importance sampling [Robert and Casella, 2013].

Parametric techniques for spectral estimation are mostly based on autoregressive (AR) models for which an analytical formulation of the power spectrum is available in closed form [Parzen, 1974]. We note that an AR process of order p can be expressed as

$$Y_t = \sum_{j=1}^{p} \psi_j Y_{t-j} + \varepsilon_t \tag{1.9}$$

where ε_t are independent zero-mean errors with variance σ^2, and $\psi_1, \ldots, \psi_p$ are the autoregressive coefficients. The general form of the spectral density is then given

by

$$f(\omega) = \frac{\sigma^2}{\left|\phi\left(\exp\left(-i2\pi\omega\right)\right)\right|^2}, \tag{1.10}$$

where $\phi(z) = 1 - \sum_{j=1}^{p} \psi_j z^j$ and we assume a stationary solution of Equation (1.10), i.e. we require that the autoregressive polynomial satisfies $\phi(z) \neq 0$, for all $|z| = 1$ [Brockwell et al., 2002]. Parameter estimation can be carried out via maximum likelihood estimation or by using the Yule-Walker equations (see e.g. Shumway and Stoffer [2017]). Model selection to determine the order of the AR model is usually achieved by employing information criteria such as Akaike's information criterion (AIC) [Akaike, 1974] or the Bayesian information criterion (BIC) [Schwarz et al., 1978]. These parametric methods however tend to provide biased estimates of the spectral density function when the AR approximation to the underlying process is poor and their performance is very much affected by the correct choice of the order of the AR model. Nevertheless, a parametric technique will generally outperform a nonparametric approach when the parametric model is correct [Rosen et al., 2012].

We note that the approaches mentioned above (both parametric and nonparametric) are formulated under the assumption that a power spectrum is continuous across frequencies. Although these methods can be well suited for time series processes with smooth underlying spectrum with few or no peaks, they are severely challenged in detecting pronounced peakedness, possibly at nearby frequencies, as can be expected to occur for the type of time series that we wish to analyze in this thesis.

1.3 Modeling Periodic Signals

In the signal processing literature it is common to model periodic data as a linear combination of a finite number of sinusoids plus noise [Priestley, 1981]. Let us assume that the underlying data generating process Y_t is characterized by m relevant distinct frequencies $\omega_1, \ldots, \omega_m$. Let us also define A_l and ϕ_l as the *amplitude* and *phase shift* of the corresponding frequency ω_l, for $l = 1, \ldots, m$. Then, we can write the following periodic model

$$Y_t = \sum_{l=1}^{m} \Big(A_l \cos(2\pi\omega_l t + \phi_l) \Big) + \varepsilon_t, \tag{1.11}$$

where we might assume that the ε_t are independent zero-mean Gaussian errors with variance σ^2.

There have been several developments for estimating the parameters of such an oscillatory model. Rife and Boorstyn [1976] and Stoica et al. [1989] addressed the problem of estimating the parameters, i.e. frequencies, phases and amplitudes, of the sinusoidal signals under the assumption of a known number of sinusoids m, where inference is based on maximum likelihood. These approaches, however, require very long time series and a large separation in the frequencies that drive the process, which will not always be the case in practice [Djuric, 1996; Andrieu and Doucet, 1999]. Quinn [1989], Yau and Bresler [1993] and Zhang and Wong [1993] tackled the problem of model selection on the number of sinusoidal components by employing AIC and the minimum description length (MDL) principle [Rissanen, 1978], the latter being an information criterion based on the concept that the best-fitting model is the one that enables maximum *compression* of the data. Djuric [1996] showed that both AIC and MDL tend to estimate a wrong number of

sinusoids when the sample size is small and the signal-to-noise ratio is low.

Bayesian approaches to modelling stationary oscillatory signals were explored for the first time by Bretthorst [1988, 1990] with applications to nuclear magnetic resonance spectroscopy. Dou and Hodgson [1995, 1996] presented a Bayesian approach that uses a Gibbs sampler to identify multiple frequencies that drive the signal. Their method required the number of frequencies to be fixed in advance, and model selection was achieved by choosing the most probable model based on the estimation of the parameters for all possible models. Bayesian model selection for stationary oscillatory signals based on posterior model probabilities were also investigated by Djuric [1996]. Andrieu and Doucet [1999] introduced a reversible-jump MCMC method [Green, 1995] that is able to jointly address model selection and parameter estimation for an unknown number of stationary sinusoidal signals and avoids the computationally expensive numerical optimization of Dou and Hodgson [1995, 1996] by sampling the frequencies one-at-time via Metropolis-Hastings (M-H) steps. To the best of our knowledge, currently there is no extension of this methodology to analyze nonstationary oscillatory signals.

1.4 Spectral Analysis of Nonstationary Time Series

Many natural phenomena may be nonstationary in the sense that they exhibit changes in their cyclic or periodic behaviour. Approaches to spectral analysis of nonstationary processes were first developed by [Priestley, 1965] who introduced the concept of *evolutionary spectra*, namely spectral density functions which are time dependent as well as localized in the frequency domain. A formal statistical modeling framework for a specific class of nonstationary time series data, called *locally stationary* time series, was presented by Dahlhaus et al. [1997]. Intuitively, a process

is locally stationary if we can construct an interval around each time point in such a way that the time series can be considered to be stationary in that interval. Locally stationary processes can hence be well approximated by *piecewise stationary* processes and several authors proposed to approximate the time-varying spectra of locally stationary time series through the piecewise constant spectra of the corresponding stationary segments [Ombao et al., 2001; Davis et al., 2006; Rosen et al., 2012]. Let $\{ Y_t : t = 1, \ldots, n \}$ be a time series process whose behaviour changes at m time points $s_1, \ldots, s_m$, and let $I_j = [s_{j-1}, s_j)$ denotes the sub-interval corresponding to the j^{th} segment. The process Y_t is said to be piecewise stationary if can be expressed [Adak, 1998] as

$$Y_t = \sum_{j=1}^{m+1} Y_t^{(j)} \, \mathbb{1}_{[t \in I_j]}, \tag{1.12}$$

where $Y_t^{(j)}$ are independent and stationary processes with spectral density $f^j(\omega)$, $\mathbb{1}_{[\cdot]}$ denotes the indicator function, and we set $s_0 = 1$ and $s_{m+1} = n$. Under this framework, a major challenge is to determine the number $m + 1$ of stationary segments and their corresponding partitions which are identified by the change-point locations $s_1, \ldots, s_m$.

[Ombao et al., 2001] proposed a nonparametric method to estimate the time-varying spectrum of nonstationary processes based on the smooth localized complex exponential (SLEX) basis library. This procedure automatically divides the time series (in the time domain) into a hierarchy of dyadic subintervals and provides a windowed estimate of the power spectra in the corresponding blocks. A frequentist parametric approach based on fitting piecewise AR processes was developed by Davis et al. [2006] where both the number and locations of the segments and the orders of the corresponding AR process are assumed unknown. They suggest an MDL

criterion for comparing different segmented AR model fits to the data, which results in an optimization problem that is addressed using a genetic algorithm. Rosen et al. [2012] introduced a Bayesian approach to model the log of the time-varying spectral density by splitting the time series into an unknown but finite number of segments of variable lengths, and to estimate the time-varying spectral density using a fixed number of smoothing splines. For a given partition of the time series, the likelihood function is approximated via a product of local Whittle likelihoods and posterior inference relies on a reversible-jump MCMC algorithm. The methodology is based on the assumption that, conditional on the position and number of partitions, the time series are piecewise stationary, and the underlying spectral density for each partition is smooth over frequencies. However, exploratory analyses of the time series in our case studies revealed spectral densities with very sharp peaks, often at several nearby frequencies, thus invalidating the assumption that the spectral density is smooth over frequencies. In addition, the frequency location of these sharp peaks appears to change over time.

This thesis introduces new approaches to the automated analysis of nonstationary periodic time series. Chapter 2 presents a novel Bayesian methodology for analyzing nonstationary time series data that exhibit regime shifts in periodicity, amplitude and phase, where we approximate the time series using a piecewise oscillatory model with unknown periodicities, and our goal is to estimate the change-points while simultaneously identifying the potentially changing periodicities in the data. Bayesian inference is performed by developing a reversible jump MCMC based algorithm for sampling from the posterior in a manner that can simultaneously estimate both the number and location of the approximately stationary segments, as well as the number, frequency and magnitude of the sinusoids within each segment. Our methodology can be seen as a novel and non-trivial ex-

tension of the statistical methodology introduced in [Andrieu and Doucet, 1999] to the nonstationary setting. We show that the proposed approach successfully identifies time changing oscillatory behaviour in two case studies, where it detects the occurrence of ultradian oscillations in human skin surface temperature during the time of night rest, and can be used to characterize instances of sleep apnea in plethysmographic respiratory traces. Furthermore, to develop a more general framework, it is of interest to identify and model the recurrence of a relevant cyclical pattern in a probabilistic way. In Chapter 3, we consider periodic phenomena whose behaviour switches dynamically over time as realizations of a hidden Markov model (HMM). The number of states is assumed unknown along with their relevant periodicities, which may vary over the different states as each state is characterized by different spectral properties. Flexibility on the number of states is achieved following Fox et al. [2011], namely by using Bayesian nonparametric methods that address the stochastic switching dynamics of the time series via a hierarchical Dirichlet process that captures the temporal mode persistence of the hidden states. The variable dimensionality regarding the number of periodicities that characterizes the different states can also be addressed by developing a reversible-jump MCMC sampler. We illustrate the use of our proposed methodology in an application relevant to respiratory research and sleep medicine, namely the detection of recurring instances of sleep apnea in human respiratory traces. We conclude and discuss our current findings in Chapter 4.

Chapter 2

Bayesian Model Search for Nonstationary Periodic Time Series

We develop a novel Bayesian methodology for modelling oscillatory data that show regime shifts in periodicity, amplitude and phase. We assume that, conditional on the position and number of change-points, the time series can be approximated by a piecewise changing sinusoidal regression model. The timing and number of changes are unknown, along with the number and values of relevant periodicities in each segment. We develop a reversible jump MCMC technique that jointly explores the parameter space of the change-points and sub-models for all segments. The chapter is organized as follows. Section 2.1 and 2.2 present the model, the prior specifications and the general structure of our Bayesian approach. Sections 2.3 and 2.4 provide a detailed explanation of our sampling scheme and simulation studies to demonstrate the performance of our proposed procedure. In Section 2.5 we illustrate the use of our methodology in two data-rich scenarios related to sleep, circadian rhythm and e-Health research, namely the identification of the spectral properties of experimental breathing traces arising in sleep apnea research, and

the analysis of human temperature data measured over several days by a wearable sensor. Further details about the sampling scheme are provided in Section 2.A. This chapter formed the manuscript for "Bayesian Model Search for Nonstationary Periodic Time Series" which has been published in *Journal of the American Statistical Association*. Julia code that implements the methodology can be found at https://github.com/Beniamino92/AutoNOM.

2.1 The Model

Consider a time series realization $y_1, \dots, y_n$ whose periodic behaviour may change at k unknown time-points $\boldsymbol{s}_{(k)} = (s_1, \dots, s_k)'$ where k is also unknown. Assume that in each sub-interval $\mathrm{I}_j = [s_{j-1}, s_j)$ there are m_j relevant frequencies $\boldsymbol{\omega}_j = (\omega_{j,1}, \dots, \omega_{j,m_j})'$, for $j = 1, 2, \dots, k+1$. Setting $s_0 = 1$ and $s_{k+1} = n$, we can write the following sinusoidal model (Andrieu and Doucet, 1999)

$$y_t = \sum_{j=1}^{k+1} f(t, \boldsymbol{\beta}_j, \boldsymbol{\omega}_j) \mathbb{1}_{[t \in I_j]} + \varepsilon_t, \tag{2.1}$$

where

$$f(t, \boldsymbol{\beta}_j, \boldsymbol{\omega}_j) = \alpha_j + \mu_j t + \sum_{l=1}^{m_j} \left(\beta_{j,l}^{(1)} \cos(2\pi\omega_{j,l} t) + \beta_{j,l}^{(2)} \sin(2\pi\omega_{j,l} t) \right), \tag{2.2}$$

$\boldsymbol{\beta}_j = (\alpha_j, \mu_j, \boldsymbol{\beta}_{j,1}', \dots, \boldsymbol{\beta}_{j,m_j}')'$, $\boldsymbol{\beta}_{j,l} = (\beta_{j,l}^{(1)}, \beta_{j,l}^{(2)})'$, $\mathbb{1}_{[\cdot]}$ denotes the indicator function, and μ_j and α_j may, if needed, account for a linear trend within each segment. For simplicity we assume independent zero-mean Gaussian errors with regime-specific variances

$$\varepsilon_t \sim \mathcal{N}(0, \sigma_j^2), \quad \text{for } t \in \mathrm{I}_j \text{ and } j = 1, \dots, k+1, \tag{2.3}$$

noting that, in principle, the methodology can in principle be extended to the non-Gaussian case.

The dimension of the model is given by the number of change-points k and the number of frequency components in each regime denoted by $\boldsymbol{m}_{(k)} = (m_1, \ldots, m_{k+1})'$. Furthermore, let $\boldsymbol{\beta}_{(k)} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_{k+1})'$, $\boldsymbol{\omega}_{(k)} = (\boldsymbol{\omega}'_1, \ldots, \boldsymbol{\omega}'_{k+1})'$, $\boldsymbol{\sigma}^2_{(k)} = (\sigma^2_1, \ldots, \sigma^2_{k+1})'$ and $\boldsymbol{\theta}_{(k)} = (\boldsymbol{\beta}'_{(k)}, \boldsymbol{\omega}'_{(k)}, \boldsymbol{\sigma}^{2'}_{(k)})'$. Using Equation (2.1), the likelihood of $(k, \boldsymbol{m}_{(k)}, \boldsymbol{s}_{(k)}, \boldsymbol{\theta}_{(k)})$ given the data $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is

$$\mathscr{L}(k, \boldsymbol{m}_{(k)}, \boldsymbol{s}_{(k)}, \boldsymbol{\theta}_{(k)}, |\boldsymbol{y}) = \prod_{j=1}^{k+1} \mathscr{L}(m_j, \boldsymbol{\theta}_j | \boldsymbol{y}_j), \quad \boldsymbol{y}_j = (y_t : t \in \mathrm{I}_j), \tag{2.4}$$

where

$$\mathscr{L}(m_j, \boldsymbol{\theta}_j | \boldsymbol{y}_j) = (2\pi\sigma_j^2)^{-n_j/2} \exp\left[-\frac{1}{2\sigma_j^2} \sum_{t \in \mathrm{I}_j} \left\{ y_t - \boldsymbol{x}_t(\boldsymbol{\omega}_j)' \boldsymbol{\beta}_j \right\}^2 \right], \tag{2.5}$$

$\boldsymbol{\theta}_j = (\boldsymbol{\beta}'_j, \boldsymbol{\omega}'_j, \sigma^{2'}_j)'$ is the vector of parameters, n_j the number of observations of the jth segment, and the vector of basis functions $\boldsymbol{x}_t(\boldsymbol{\omega}_j)$ is defined as

$$\boldsymbol{x}_t(\boldsymbol{\omega}_j) = (1, t, \cos(2\pi\omega_{j,1}t), \sin(2\pi\omega_{j,1}t), \ldots, \cos(2\pi\omega_{j,m_j}t), \sin(2\pi\omega_{j,m_j}t))'. \tag{2.6}$$

2.2 Bayesian Inference

Given some pre-fixed maximal numbers of change-points, $k_{\max}$, and frequencies per regime, $m_{\max}$, inference is achieved by assuming that the true model is unknown but comes from a finite class of models where each model $\mathcal{M}_k$, with k change-points, is parameterized by the vector

$$(\boldsymbol{m}_{(k)}, \boldsymbol{s}_{(k)}, \boldsymbol{\theta}_{(k)}) \in \boldsymbol{\Pi}_k, \quad \boldsymbol{\Pi}_k \in \boldsymbol{\Pi}.$$

Let $\boldsymbol{S}_k = \left\{ \boldsymbol{s}_{(k)} \in [1,\, n]^k \;:\; \mathbb{1}_{[1<s_1<\cdots<s_k<n]} \right\}$ and $\boldsymbol{\Omega}_{m_j} = (0, 0.5)^{m_j}$ denote, respectively, the sample space for the locations of change-points and the frequencies of the jth segment. The overall parameter space can be written as a finite union of subspaces

$$\boldsymbol{\Pi} = \bigcup_{k=0}^{k_{\max}} \left\{ k \right\} \times \boldsymbol{\Pi}_k, \qquad \text{and} \quad \boldsymbol{\Pi}_k = \prod_{j=1}^{k+1} \{m_j\} \times \boldsymbol{S}_k \times \bigcup_{m_j=1}^{m_{\max}} \left\{ \mathbb{R}^{2m_j+2} \times \boldsymbol{\Omega}_{m_j} \times \mathbb{R}^{+} \right\}.$$

Bayesian inference on k, $\boldsymbol{m}_{(k)}$, $\boldsymbol{s}_{(k)}$ and $\boldsymbol{\theta}_{(k)}$ may be achieved through the following factorization of the joint posterior distribution

$$\pi(k,\, \boldsymbol{m}_{(k)},\, \boldsymbol{s}_{(k)},\, \boldsymbol{\theta}_{(k)} \,|\, \boldsymbol{y}) = \pi(k \,|\, \boldsymbol{y})\, \pi(\boldsymbol{m}_{(k)} \,|\, k,\, \boldsymbol{y})\, \pi(\boldsymbol{s}_{(k)} \,|\, \boldsymbol{m}_{(k)},\, k,\, \boldsymbol{y})\, \pi(\boldsymbol{\theta}_{(k)} \,|\, \boldsymbol{s}_{(k)},\, k,\, \boldsymbol{m}_{(k)},\, \boldsymbol{y}),$$

where we use $\pi(\cdot)$ as generic notation for probability density or mass function, whichever is appropriate. Sampling from it poses a multiple model selection problem, namely of the number of change-points and number of frequencies in each regime, which can be addressed by constructing a reversible-jump MCMC algorithm Green [1995]. The algorithm in its basic structure iterates between the following two moves:

1. **Segment model move:** Given a partition of the data at k locations $\boldsymbol{s}_{(k)}$, inference on the parameters $\boldsymbol{m}_{(k)}$ and $\boldsymbol{\theta}_{(k)}$ is based on the conditional posterior

$$\pi(\boldsymbol{m}_{(k)},\, \boldsymbol{\theta}_{(k)} \,|\, k,\, \boldsymbol{s}_{(k)},\, \boldsymbol{y}) = \prod_{j=1}^{k+1} \pi(m_j,\, \boldsymbol{\theta}_j \,|\, k,\, \boldsymbol{s}_{(k)},\, \boldsymbol{y}_j).$$

   A reversible-jump MCMC algorithm is performed in parallel on each of the $k+1$ segments, where at each iteration the number of sinusoids m_j, the linear coefficients $\boldsymbol{\beta}_j$, the frequencies $\boldsymbol{\omega}_j$ and the residual variances σ_j^2 are sampled independently in each segment, for $j = 1, \ldots, k+1$. Notice that at this stage

the algorithm will explore subspaces of variable dimensionality regarding the number of frequencies per segment, while the change-point model remains fixed.

2. **Change-point model move:** This step performs a reversible-jump MCMC algorithm for change-point model search where the number k and locations of change-points $\boldsymbol{s}_{(k)}$ are sampled, along with the linear coefficients, number of frequencies and their values as well as the residual variances for any segments affected by the move.

2.2.1 Prior Specifications

Our prior specifications assume independent Poisson distributions for the number of change-points k and the frequencies in each segment m_j, conditioned on $k \leq k_{\max}$ and $1 \leq m_j \leq m_{max}$, respectively. For example, if X is Poisson distributed with mean λ, then the truncated Poisson distribution at $x_{\max}$ has probability mass function

$$\pi\left(X = x \,|\, X \leq x_{\max}\right) = \frac{\lambda^x}{x! \sum_{h=0}^{x_{\max}} \frac{\lambda^h}{h!}}, \qquad x = 0, \ldots, x_{\max}.$$

This prior is used by several authors. For example, Green [1995] assumed this prior to model the number of change-points in a Poisson point process, and Denison et al. [1998] for the number of knots when fitting smoothing splines. Richardson and Green [1997] utilized a Poisson prior on the unknown number of Gaussian mixtures, and similarly Andrieu and Doucet [1999] on the number of sine waves in a regression model. The prior mean λ is an hyper-parameter fixed in advance that can potentially differ between the prior on k and the one on m_j. It can be interpreted as the expected number of change-points and the expected number of sinusoids in

each regime, respectively.

Given k, a prior distribution for the positions of the change-points $\boldsymbol{s}_{(k)}$ can be chosen as in Green [1995]

$$\pi(\boldsymbol{s}_{(k)} \mid k) = \frac{(2k+1)!}{(n-1)^{2k+1}} \prod_{j=0}^{k} \left(s_{j+1} - s_j\right) \mathbb{1}_{[s_0 < s_1 < \cdots < s_k < n]}, \qquad s_0 = 1, s_{k+1} = n. \quad (2.7)$$

This implies that the location of the changes $s_1, \ldots, s_k$ are distributed as the even-numbered order statistics from $2k+1$ points uniformly distributed on $[1, n]$. To show this, consider $X_1, \ldots, X_{2k+1} \sim \text{Uniform}(1, n)$, where the joint distribution of the order statistics is given by

$$\pi_{1,\ldots,2k+1}(x_1, \ldots, x_{2k+1}) = \frac{(2k+1)!}{(n-1)^{2k+1}} \mathbb{1}_{[1 < x_1 < \cdots < x_{2k+1} < n]}. \quad (2.8)$$

The interpretation of the latter is straightforward: there are $(2k+1)!$ ways to order the realisations of $X_1, \ldots, X_{2k+1}$, and the product of their densities is $\frac{1}{(n-1)^{2k+1}}$. Hence, we can marginalise the joint density in Equation (2.8) with respect to the even numbered $x_2, \ldots, x_{2k}$ as

$$\begin{aligned}
\pi_{2,\ldots,2k}(x_2, \ldots, x_{2k}) &= \int_1^{x_2} \int_{x_2}^{x_4} \cdots \int_{x_{2k-2}}^{x_{2k}} \int_{x_{2k}}^{n} \frac{(2k+1)!}{(n-1)^{(2k+1)}} dx_{2k+1} dx_{2k-1} \ldots dx_3 dx_1 \\
&= \frac{(2k+1)!}{(n-1)^{(2k+1)}} (n - x_{2k})(x_{2k} - x_{2k-2}) \cdots (x_4 - x_2)(x_2 - 1)\, \mathbb{1}_{[1 < x_2 < \cdots < x_{2k} < n]}.
\end{aligned}$$

Finally, we can set $s_1 = x_2$, $s_2 = x_4$, $\ldots$, $s_k = x_{2k}$ and obtain the prior density for the locations of the change-points $\pi(\boldsymbol{s}_{(k)} \mid k)$ given in Equation (2.7).

Conditional on k and $\boldsymbol{m}_{(k)}$, we choose a uniform prior for the frequencies $\omega_{j,l} \sim \text{Uniform}(0, \phi_\omega)$, $\quad l = 1, \ldots, m_j$, and $j = 1, \ldots, k+1$, where $0 < \phi_\omega < 0.5$. The value of ϕ_ω can be chosen to be informative in the sense that it may reflect prior information about the significant frequencies that drive the variation in the

data, for example by choosing ϕ_ω in the low frequencies range ($0 < \phi_\omega < 0.1$). Analogous to a Bayesian regression [Gelman et al., 2014; Bishop, 2006], a zero-mean isotropic Gaussian prior is assumed for the coefficients of the j^{th} segment, $\boldsymbol{\beta}_j \sim \mathcal{N}_{2m_j+2}(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I})$, $j = 1, \ldots, k+1$ where the prior variance σ_β^2 is fixed at a relatively large value (e.g. in our case 10^2). The prior on the residual variance σ_j^2 of state j is specified as Inverse-Gamma $(\frac{\xi_0}{2}, \frac{\tau_0}{2})$, where η_0 and ν_0 are fixed at small values. Notice that when these two hyper-parameters are set to zero we obtain the non-informative Jeffrey's prior [Gelman et al., 2014].

In ordinary parametric problems it is often considered a safe choice to use improper priors [Robert, 2007]. However, in view of the Jeffreys-Lindley's paradox [Lindley, 1957; Jeffreys, 1998] when performing model selection, it is known that improper priors may cause several problems. For example, in the context of Bayesian models for normal mixtures, as the priors become less informative, the posterior probability for the number of mixture components tends to concentrate on low values and often to have a mode on the smallest model, as discussed in Jennison [1997] and Jasra et al. [2005]. Assuming prior distributions that are proper, but reasonably flat over the range of parameters values that could possibly arise - such as $\boldsymbol{\beta}_j \sim \mathcal{N}_{2m_j+2}(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I})$ - mitigate this issue, but only partially, as the results will then depend on the arbitrary degree - e.g. σ_β^2 - of vagueness utilized. In fact, by fixing $\sigma_\beta^2 = 10^6$ in our simulation studies, we always find a posterior mode that corresponds to the smallest model, i.e. a model with no change-points and one relevant frequency. However, a weakly informative prior obtained by fixing $\sigma_\beta^2 = 10^2$ appears to perform satisfactorily, for all our experiments. Alternatively, the model could be parameterized in such a way that the priors do not depend on the scale of the observations. For example, a popular Bayesian model for linear regression is to assume a hierarchical parameterization for the linear coefficients $\boldsymbol{\beta}_j$ and the

residual variance σ_j^2 of the form $\pi(\boldsymbol{\beta}_j, \sigma_j^2) = \pi(\sigma_j^2)\pi(\boldsymbol{\beta}_j \mid \sigma_j^2)$ [Gelman et al., 2014]. The prior on the residual variance $\pi(\sigma_j^2)$ may be specified as before (i.e Inverse-Gamma $(\frac{\xi_0}{2}, \frac{\tau_0}{2})$), whereas the prior on the basis function coefficients can be chosen as $\pi(\boldsymbol{\beta}_j \mid \sigma_j^2) \sim \mathcal{N}_{2m_j+2}(\mathbf{0}, \sigma_j^2 \boldsymbol{V}_0)$. Following this hierarchical structure, the prior variance of $\boldsymbol{\beta_j}$ is tied to σ_j^2, which is the sampling variance of the observations $\boldsymbol{y}$ and thus prior beliefs about $\boldsymbol{\beta}_j$ are calibrated by the scale of the measurements $\boldsymbol{y}$. However, we still need to specify $\boldsymbol{V}_0$ which can be chosen of the form $\sigma_\beta^2 \boldsymbol{I}$, where σ_β^2 may be fixed, as discussed before, at a relatively large value.

2.3 Sampling Scheme for Nonstationary Periodic Processes

Here we provide the sampling scheme associated with the nonstationary periodic processes that we wish to model. An outline of the overall procedure is as follows. Start with an initial configuration of number of change-points k, along with their locations $\boldsymbol{s}_{(k)}$; this yields a partition of the data $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k+1})'$. Initialize the number of frequencies in each regime $\boldsymbol{m}_{(k)}$ and their values $\boldsymbol{\omega}_{(k)}$, along with the coefficients $\boldsymbol{\beta}_{(k)}$ and residual variances $\boldsymbol{\sigma}^2_{(k)}$. At each iteration of the algorithm a segment model and a change-point model move are estimated. A random choice with probabilities (2.9) based on the current number of parameters will determine whether to attempt a birth, death or a within-model move. In particular, let z denote the current number of parameters, i.e. change-points k in the change-point model or frequencies m_j in the jth segment model; then, the dimension may increase by one (*birth step*) with probability b_z, decrease by one (*death step*) with probability d_z or

remain unchanged (*within step*) with probability $\mu_z = 1 - b_z - d_z$, where

$$b_z = c \min\left\{1, \frac{\pi(z+1)}{\pi(z)}\right\}, \qquad d_{z+1} = c \min\left\{1, \frac{\pi(z)}{\pi(z+1)}\right\}, \tag{2.9}$$

for some constant $c \in [0, \frac{1}{2}]$, and $\pi(z)$ is the prior probability of the model including z. Reversibility of the Markov chain is guaranteed for move types that involve a change in dimensionality as $b_z \pi(z) = d_{z+1} \pi(z+1)$. Here we chose $c = 0.4$ but other values are legitimate as long as c is not larger than 0.5, to assure that the sum of the probabilities does not exceed 1 for some values of z. Naturally, $b_{k=k_{\max}} = b_{m=m_{\max}} = 0$ and $d_{k=0} = d_{m=1} = 0$. The pseudocode of the overall algorithm that describes an iteration of the sampler is given in Algorithm 1. We next describe the specific procedures needed to update the moves. More details are provided at the end of the chapter, in Section 2.A.

Algorithm 1

1. For each segment $j = 1, \ldots, k+1$, perform a segment model move (Section 2.3.1)

   Draw $U \sim \text{Uniform}(0, 1)$
   if $U \leq b_{m_j} \rightarrow$ *birth-step*
   else if $b_{m_j} \leq U \leq d_{m_j} \rightarrow$ *death-step*
   else $\rightarrow$ *within-step*

2. Perform a change-point model move (Section 2.3.2):

   Draw $U \sim \text{Uniform}(0, 1)$
   if $U \leq b_k \rightarrow$ *birth-step*
   else if $b_k \leq U \leq d_k \rightarrow$ *death-step*
   else $\rightarrow$ *within-step*

2.3.1 Updating a Segment Model

Given the number of change-points k and their locations $\boldsymbol{s}_{(k)}$, a segment model move is performed independently and in parallel on each of the $k+1$ partitions. Hence, throughout this subsection the subscript relating to the jth segment may be dropped and a segment of interest is denoted by $\boldsymbol{y} = (y_a, \ldots, y_b)'$, which contains n observations. Assume that the current number of frequencies is set at m; then, an independent random choice is made between attempting a birth, death or within-model step, with probabilities given in (2.9). An outline of these moves is as follows (further details are provided in Section 2.A.1).

Within-Model Move: Conditioned on the number of frequencies m, we sample the vector of frequencies $\boldsymbol{\omega}$ following Andrieu and Doucet [1999], i.e. by sampling the frequencies one-at-time using a mixture of M-H steps, with target distribution

$$\pi(\boldsymbol{\omega} \,|\, \boldsymbol{\beta}, \sigma^2, m, \boldsymbol{y}) \propto \exp\left[-\frac{1}{2\sigma^2} \sum_{t=a}^{b} \left\{ y_t - \boldsymbol{x}_t(\boldsymbol{\omega})' \boldsymbol{\beta} \right\}^2 \right] \mathbb{1}_{[\boldsymbol{\omega} \in \boldsymbol{\Omega}_m]}. \tag{2.10}$$

In particular, the proposal distribution is a combination of a Normal random walk centred around the current frequency and a sample from values of the Discrete Fourier transform of $\boldsymbol{y}$. Next, the corresponding vector of linear coefficients $\boldsymbol{\beta}$ is updated in a Gibbs step from its conjugate Gaussian posterior

$$\boldsymbol{\beta} \,\big|\, \boldsymbol{\omega}, \sigma^2, m, \boldsymbol{y} \sim \boldsymbol{\mathcal{N}}_{2m+2}(\hat{\boldsymbol{\beta}}, \boldsymbol{V}_\beta), \tag{2.11}$$

where

$$\begin{aligned} \boldsymbol{V}_\beta &= \left(\sigma_\beta^{-2}\,\boldsymbol{I} + \sigma^{-2}\boldsymbol{X}(\boldsymbol{\omega})'\boldsymbol{X}(\boldsymbol{\omega})\right)^{-1}, \\ \hat{\boldsymbol{\beta}} &= \boldsymbol{V}_\beta\,(\sigma^{-2}\boldsymbol{X}(\boldsymbol{\omega})'\boldsymbol{y}), \end{aligned} \tag{2.12}$$

and we denote with $\boldsymbol{X}(\boldsymbol{\omega})$ the design matrix with rows given by $\boldsymbol{x}_t(\boldsymbol{\omega})$ (Equation 2.6), for $t = a, \ldots, b$. Finally, the residual variance σ^2 is then updated in a Gibbs step from

$$\sigma^2 \,|\, \boldsymbol{\omega},\ \boldsymbol{\beta},\ m,\ \boldsymbol{y} \ \sim \text{Inverse-Gamma}\left(\frac{n+\nu_0}{2},\ \frac{\gamma_0 + \sum_{t=a}^{b}\left\{y_t - \boldsymbol{x}_t(\boldsymbol{\omega})'\boldsymbol{\beta}\right\}^2}{2}\right). \tag{2.13}$$

Between-Model Moves: For this type of move, the number of frequencies is either proposed to increase by one (birth) or decrease by one (death). If a birth move is proposed, we have that $m^p = m^c + 1$, where current and proposed values are denoted by the superscripts c and p, respectively. The proposed vector of frequencies is constructed by proposing an additional frequency to include in the current vector. Conditional on the frequencies, the corresponding vector of linear coefficients and the residual variance are sampled as in the within-model move. If a death move is proposed, we have that $m^p = m^c - 1$. Hence, one of the current frequencies is randomly chosen to be removed. The proposed corresponding vector of linear coefficients is drawn, along with the residual variance. For both moves, the updates are jointly accepted or rejected in a M-H step.

We note that we may use a reversible-jump MCMC sampler on a somewhat reduced model, namely by analytically integrating the linear basis coefficients $\boldsymbol{\beta}$ out of the segment model. This may be lead to a slightly more accurate and faster algorithm than the corresponding samplers operating on the full model [Han and

Carlin, 2001]. However, posterior samples are not provided for any parameters no longer appearing within the sampling scheme. In particular, no information about the amplitude and phase of any frequency can be retrieved by such a reduced sampler. Hence, we have not investigated further this option, since posterior estimates of those parameters are necessary for characterizing the spectral properties of the time series we wish to analyze.

2.3.2 Updating the Change-Point Model

This part of the algorithm identifies the number and locations of change-points. Suppose the number of change-points is currently set to some value k, then according to the probabilities given in (2.9) a random decision is made between adding, removing or moving a change-point. The rules for updating these types of moves are described below and more details are given in Section 2.A.2.

Within-Model Move: An existing change-point is proposed to be relocated with probability $\frac{1}{k}$, obtaining say s_j^c. The update for the selected change-point is proposed from a mixture of a Normal random walk centred on the current change-point s_j^c and a sample from a uniform distribution on the interval $[s_{j-1}^c+\psi_s, s_{j+1}^c-\psi_s]$. Here, we introduced ψ_s as a fixed minimum time between change-points avoiding change-points being too close to each other. Rosen et al. [2012] used a similar scheme, but on a discrete-scale. The number of frequencies and their values are kept fixed, and, conditional on the relocation, the linear coefficients for the segments affected by the relocation are sampled. These updates are jointly accepted or rejected in a M-H step and the residual variances are updated in a Gibbs step.

Between-Model Moves: For this type of move, the number of change-points may either increase (birth) or decrease (death) by one. If a birth move is

proposed, we have that $k^p = k^c + 1$. The new proposed change-point is drawn uniformly on $f(\boldsymbol{s}^c_{(k^c)}, \psi_s)$, the support of $\boldsymbol{s}^c_{(k^c)}$ given the constraints imposed by ψ_s, i.e. $f(\boldsymbol{s}^c_{(k^c)}, \psi_s) = [1 + \psi_s,\ s^c_1 - \psi_s] \cup [s^c_1 + \psi_s,\ s^c_2 - \psi_s] \cup \ldots \cup [s^c_{k^c} + \psi_s,\ n - \psi_s]$. The latter involves splitting an existing segment. The number of frequencies and their values in the proposed segments are selected from the current states. Two residual variances for the new proposed segments are then constructed from the current single residual variance. Finally, two new vectors of linear parameters are sampled. If a death move is proposed, we have that $k^p = k^c - 1$. Hence, a candidate change-point to be removed is selected from the vector of existing change-points, with probability $\frac{1}{k^c}$. The latter involves merging two existing partitions. The number of frequencies and their values in the proposed segments are selected from the current states. A single residual variance is constructed from the current variances relative to the segments affected by the relocation. Finally, a new vector of linear coefficient is drawn. For both type of moves, these updates are jointly accepted or rejected.

2.4 Simulation Studies

We carry out simulation studies to explore the performance of our method, which will be referred to as AutoNOM (Automatic Nonstationary Oscillatory Modelling). In Section 2.4.1 we illustrate the performance of our methodology when the simulated data are generated from the proposed model. Section 2.4.2 deals with scenarios when the model is misspecified relative to the generating process. Our results are compared with two state-of-the-art existing methods.

2.4.1 Illustrative Example

In this simulation example, we generate a time series consisting of $n = 900$ data points from model (2.1) with $k = 2$ change-points located at positions $\boldsymbol{s}_{(2)} = (300, 650)$, and fixed number of frequencies per regime $\boldsymbol{m}_{(2)} = (3, 1, 2)$. Further details of the parameterization are given in Table 2.1.

Figure 2.1 (top panel) shows a realization from this model. The prior means λ_ω and λ_s, say, on the number of frequencies and change-points, respectively, were set to 2, to reflect a fair degree of prior information on their numbers. We discuss in Section 2.4.1.1 that AutoNOM was relatively insensitive to these prior specifications, for this example. The maximum number of change-points $k_{\max}$ was set to 15, and the maximum number of frequencies per regime $m_{\max}$ was set to 10. Furthermore, we fixed $\psi_s = 20$ and $\phi_\omega = 0.25$ (see Appendix 2.A) for the uniform distribution for sampling the frequencies. The full estimation algorithm was ran for 20,000 updates, 5,000 of which are discarded as burn-in period. The estimation took 390 seconds with a (serial) program written in Julia 0.62 on a Intel® Core™ i7-4790S Processor 16 GB RAM. The results, summarized in Table 2.4 clearly show that a model with two change-points has the highest estimated posterior probability (left panel) and that AutoNOM correctly identifies the right number of significant frequencies in each regime (right panel).

Figure 2.1 (middle panel) shows the estimated posterior distribution for the location of the change-points, conditioned on three segments. The posterior means of the change-point locations are $\hat{\mathbb{E}}(s_1 \,|\, k = 2, \boldsymbol{y}) = 298.7$ and $\hat{\mathbb{E}}(s_2 \,|\, k = 2, \boldsymbol{y}) = 650.1$. Figure 2.1 (bottom panel) shows that the estimated posterior distributions are an excellent match to the true frequencies.

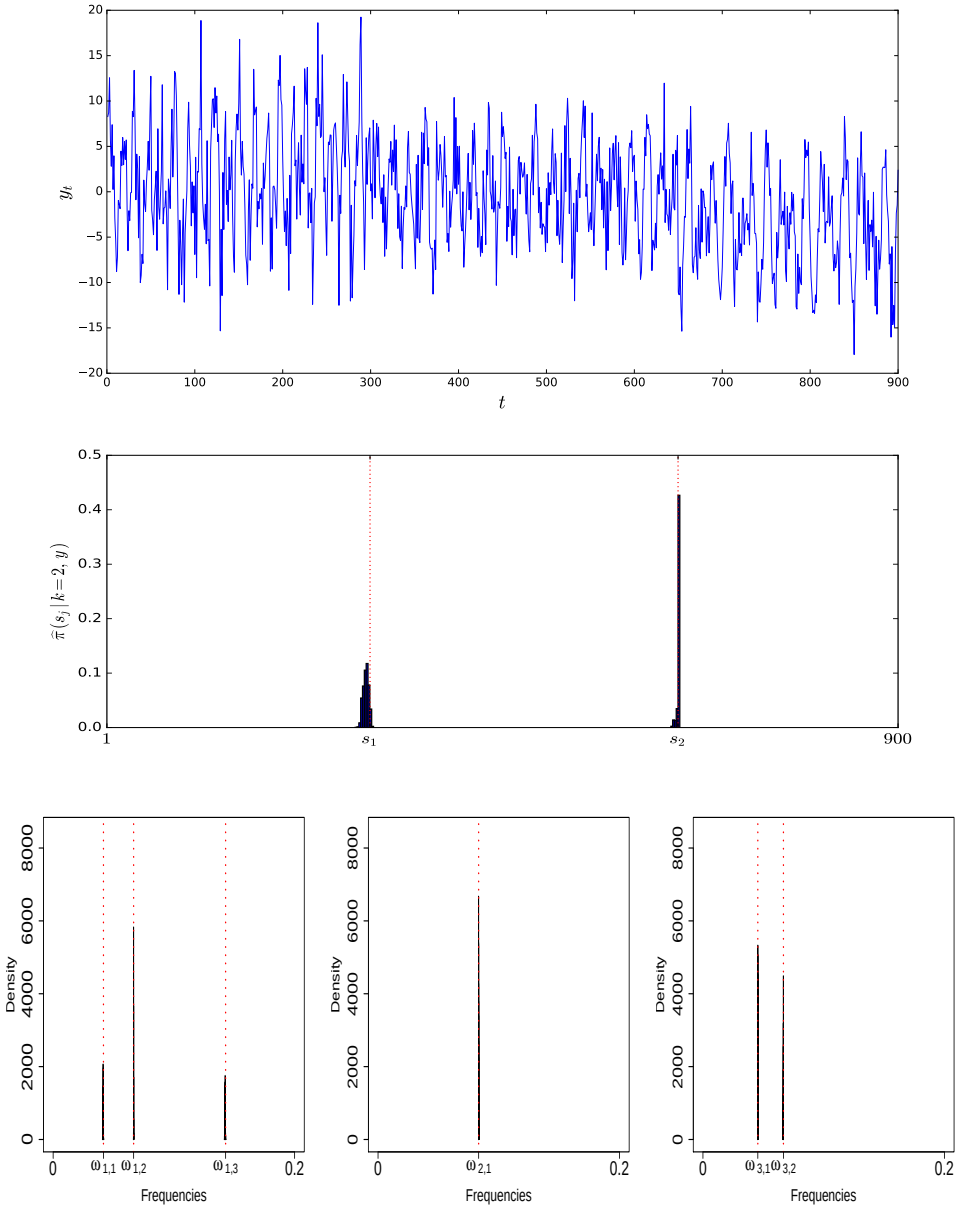

Figure 2.1: Illustrative example. (Top) Simulated time series. (Middle) Estimated posterior distribution for the location of the change-points, conditioned on $k = 2$. The dotted vertical lines represent true location of change-points. (Bottom) Estimated posterior distribution of the frequencies for each different segment, conditioned on $k = 2$, $m_1 = 3$, $m_2 = 1$ and $m_3 = 2$. The dotted vertical lines represent true values of the frequencies.

Table 2.1: Illustrative example. Parameter values for simulation from model (2.1). The value of the intercept was set to zero for every segment.

Frequencies		Linear coefficients		Trends and variances	
$\omega_{1,1}$	1/24	$\boldsymbol{\beta}_{1,1}$	(2.0, 3.0)	μ_1	.010
$\omega_{1,2}$	1/15	$\boldsymbol{\beta}_{1,2}$	(4.0, 5.0)	μ_2	.000
$\omega_{1,3}$	1/7	$\boldsymbol{\beta}_{1,3}$	(1.0, 2.5)	μ_3	-.005
$\omega_{2,1}$	1/12	$\boldsymbol{\beta}_{2,1}$	(4.0, 3.0)	σ_1^2	4.0^2
$\omega_{3,1}$	1/22	$\boldsymbol{\beta}_{3,1}$	(2.5, 4.0)	σ_2^2	3.5^2
$\omega_{3,2}$	1/15	$\boldsymbol{\beta}_{3,2}$	(4.0, 2.0)	σ_3^2	2.8^2

2.4.1.1 Sensitivity Analysis

To investigate the influence of the prior means λ_ω and λ_s we simulate 10 realizations from the same model and run our estimation algorithm for combinations of values for λ_ω and λ_s, ranging from 0.1 to 10.0. Table 2.2 shows the average posterior probability of choosing the correct model, i.e. $\hat{\pi}(k = 2, m_1 = 3, m_2 = 1, m_3 = 2 \,|\, \boldsymbol{y})$. Table 2.3 displays the average mean squared error

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^{n} \left\{ \hat{f}_t - f_t \right\}^2,$$

to asses the distance between the true underlying signal f_t and the estimated signal $\hat{f}_t$. The latter is obtained by averaging across models of differing number of components, in contrast to model selection. Specifically, if we run our procedure for S iterations, then the estimated signal $\hat{f}_t$ is defined as

$$\hat{f}_t = \frac{1}{S} \sum_{s=1}^{S} \sum_{j=1}^{k^{(s)}+1} f(t, \boldsymbol{\beta}_j^{(s)}, \boldsymbol{\omega}_j^{(s)}) \mathbb{1}_{[t \in I_j^{(s)}]}, \quad t = 1, \ldots, n, \tag{2.14}$$

where the superscript (s) denotes the s^{th} sample of the Markov chain. Both analyses suggest that, for this example, the choice of the prior means λ_ω and λ_s has hardly noticeable impact on the results. However, our experience is that small values for

these hyper-parameters are preferable as they prevent the algorithm from overfitting and seems to be more robust to model misspecification.

Table 2.2: Sensitivity analysis of illustrative example. Average probability of chosing the correct model from 10 replications with varying λ_ω and λ_s

	$\lambda_s = .1$	$\lambda_s = .2$	$\lambda_s = 0.5$	$\lambda_s = 1.0$	$\lambda_s = 2.0$	$\lambda_s = 5.0$	$\lambda_s = 10.0$
$\lambda_\omega = .1$	.98	.97	.97	.99	.93	.91	1.0
$\lambda_\omega = .2$	.99	.99	.99	.99	.99	.99	.98
$\lambda_\omega = .5$	.99	1.0	.98	.95	.98	1.0	.99
$\lambda_\omega = 1.0$	.99	.94	.93	.99	.99	.99	.99
$\lambda_\omega = 2.0$	.99	.99	.99	.99	.99	.99	.99
$\lambda_\omega = 5.0$	.97	.96	.96	.97	.97	.93	.98
$\lambda_\omega = 10.0$	.94	.95	.95	.91	.93	.95	.85

Table 2.3: Sensitivity analysis of illustrative example. Average MSE from 10 replications with varying λ_ω and λ_s

	$\lambda_s = .1$	$\lambda_s = .2$	$\lambda_s = .5$	$\lambda_s = 1.0$	$\lambda_s = 2.0$	$\lambda_s = 5.0$	$\lambda_s = 10.0$
$\lambda_\omega = .1$	.349	.435	.370	.416	.446	.470	.319
$\lambda_\omega = .2$	.354	.360	.378	.361	.370	.404	.394
$\lambda_\omega = .5$	.400	.347	.378	.447	.430	.346	.364
$\lambda_\omega = 1.0$	.302	.392	.400	.382	.321	.400	.337
$\lambda_\omega = 2.0$	.307	.369	.404	.340	.407	.329	.391
$\lambda_\omega = 5.0$	.360	.387	.362	.396	.346	.350	.324
$\lambda_\omega = 10.0$	.355	.386	.340	.387	.420	.393	.428

2.4.1.2 Detecting Spectral Peaks and Comparison with Existing Methods

We simulate a time series from the same simulation model as in Section 2.4.1 with the only difference that the residual variances were set equal to one for all segments and thus are smaller than above. The performance of AutoNOM is compared with two existing methods, namely the Bayesian adaptive spectral estimation for nonstationary time series proposed by Rosen et al. [2012], referred to as Adapt-

Table 2.4: Illustrative example. (left panel) posterior probabilities for number of change-points; (right panel) posterior probabilities for number of frequencies in each regime, conditioned on $k = 2$.

k	$\hat{\pi}(k \mid \boldsymbol{y})$
0	.00
1	.02
2	.97
3	.01
4	.00

m	$\hat{\pi}(m_1 \mid k = 2, \boldsymbol{y})$	$\hat{\pi}(m_2 \mid k = 2, \boldsymbol{y})$	$\hat{\pi}(m_3 \mid k = 2, \boldsymbol{y})$
1	.00	.99	.00
2	.00	.01	.98
3	.98	.00	.02
4	.02	.00	.00
5	.00	.00	.00

SPEC, and the frequentist piecewise vector autoregressive method of Davis et al. [2006], referred to as AutoPARM. Specifically, we explore the performances of these methodologies in identifying the number and location of change-points, and the number and location of frequency peaks in each estimated segment. AdaptSPEC requires the user to specify in advance the number of basis function J used for smoothing the periodogram in the segments. We run AdaptSPEC for two different specifications, namely $J = 7$ and $J = 15$ basis functions. The model is fitted with a total of 15,000 iterations, 5,000 of which are discarded as burn-in, by using the R package provided by the authors. Posterior samples of peak frequencies are obtained by considering the modes of the spectrum per MCMC iteration. AutoPARM is performed with default tuning parameters. We note that Davis et al. [2006] do not discuss computation of confidence intervals for frequencies.

The modal number of change-points for AdaptSPEC is 2 for both $J = 7$ and $J = 15$, with posterior probability $\widehat{\pi}(k = 2 \mid \boldsymbol{y})$ of 76% and 88%, respectively; the modal number of change-points for AutoNOM is 2 and AutoPARM identifies 2 change-points as well. Conditioned on the modal number of change-points, Table 2.5 displays the estimated location of changes (left panel) and frequency peaks (right panel) for the different compared methods, where we report the standard deviation for the estimate obtained from the empirical distribution of the posterior

samples. Similarly, we show in Figure 2.2 the estimated location of the frequency peaks and their 95% credible intervals, for each of the three identified segments; dotted vertical lines represents the true location of the frequency peaks. Results for AutoNOM are conditioned on the modal number of frequencies per regime.

It becomes clear that the detection of periodicities by AdaptSPEC is affected by the specification of the number of spline basis functions used for the smoothing, where increasing the number of basis function yields a better performance for AdaptSPEC. The example also shows that smoothing by splines may lead to peaks in the periodogram to be over-smoothed and neighbouring close peaks to be merged. AutoPARM seems to also suffer from the latter problem.

Table 2.5: Illustrative example with unitary residual variances. Estimated change-points locations (left panel) and frequency peaks (right panel) for AutoNOM (AN), AdaptSPEC (AS $J = 7, 15$) and AutoPARM (AP); posterior standard deviations are also reported for Bayesian methods.

	s_1	s_2	$\omega_{1,1}$	$\omega_{1,2}$	$\omega_{1,3}$	$\omega_{2,1}$	$\omega_{3,1}$	$\omega_{3,2}$
True	300	650	.042	.067	.143	.083	.046	.067
AN	300.87 (.97)	650.46 (.31)	.042 (.0004)	.067 (.0002)	.142 (.0005)	.083 (.0002)	.045 (.0003)	.067 (.0003)
AS J7	319.65 (15.36)	628.51 (37.35)	.083 (.0030)	-	-	.082 (.0070)	.057 (.0010)	-
AS J15	298.5 (1.21)	647.01 (.14)	.057 (.0030)	.141 (.0020)	-	.088 (.0050)	.056 (.0030)	-
AP	299	648	.060	.140	-	.080	.055	-

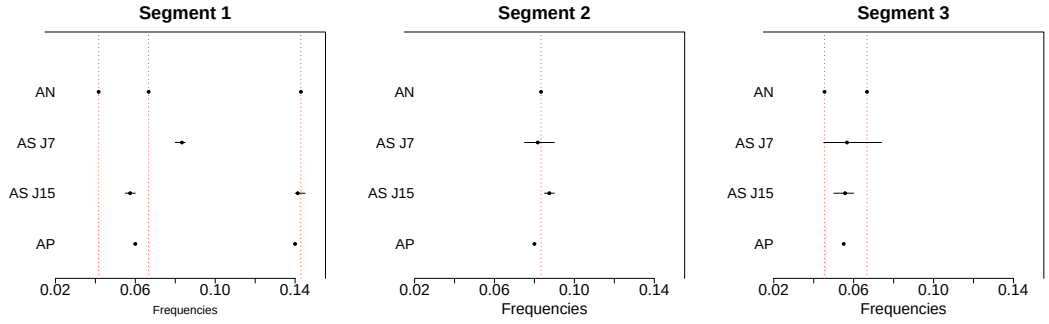

Figure 2.2: Illustrative example with unitary residual variances. Estimated frequency peaks for AutoNOM (AN), AdaptSPEC (AS, $J = 7, 15$) and AutoPARM (AP); 95% credible intervals (horizontal lines) are also reported for Bayesian methods. Dotted vertical lines are true locations of the frequency peaks.

When we increased the residual variance to the high levels set originally, AdaptSPEC failed to detect any change-points for both $J = 7$ and $J = 15$, with posterior probability $\widehat{\pi}(k = 0 \,|\, \boldsymbol{y})$ of 69% and 93%, respectively, while AutoPARM found 7 change-points and thus severely overestimates their number. Our conclusion from this comparison is that although AdaptSPEC and AutoPARM may be well suited for time series processes with smooth time-varying spectra with few or no peaks, both methods are severely challenged in detecting changes in spectra that exhibit pronounced peakedness, possibly at nearby frequencies, as can be expected to occur in reality for the type of time series that we wish to analyze.

2.4.2 Misspecified Model

We investigate the performance of our proposed method for identifying spectral peaks when the model is misspecified relative to the generating process. In particular, we explored simulation studies under three different settings. In the first two scenarios we generated data from two types of autoregressive (AR) processes, namely a piecewise AR process and a slowly varying AR process. We compare the performance of our procedure with AutoPARM and AdaptSPEC. In the third

setting we assumed that the innovations are t-distributed, and therefore violate the Gaussianity assumption of ε_t in Equation (2.3). For all models, our estimation algorithm was run for 20,000 iterations, 5,000 of which were used as burn in, and the hyperparameters were chosen as $\phi_s = 40$, $\lambda_\omega = 0.05$ and $\lambda_s = 0.01$.

2.4.2.1 Piecewise Autoregressive Process

Although modeling a time series as a linear combination of a finite number of sinusoids plus noise is common in the signal processing literature, such line-spectrum based models are rare in the statistics literature. In fact, it is commonly assumed that the power spectrum is continuous across frequencies. We investigate the performance of the proposed procedure when analyzing data generated from a piecewise AR process whose local spectral density functions show sharp peaks. Specifically, a realization is simulated from

$$y_t = \begin{cases} 1.9\, y_{t-1} - .975\, y_{t-2} + \varepsilon_t^{(1)} & \text{for } 1 \leq t \leq 250 \\ 1.9\, y_{t-1} - .991\, y_{t-2} + \varepsilon_t^{(2)} & \text{for } 251 \leq t \leq 400 \\ -1.35\, y_{t-1} - .37\, y_{t-2} + .36\, y_{t-3} + \varepsilon_t^{(3)} & \text{for } 401 \leq t \leq 550, \end{cases} \tag{2.15}$$

where $\varepsilon_t^{(1)} \overset{iid}{\sim} \mathcal{N}(0, 0.25)$ and $\varepsilon_t^{(i)} \overset{iid}{\sim} \mathcal{N}(0, 1)$ for $i = 2, 3$. Figure 2.3 (top panel) shows a realization from model (2.15). After applying our methodology AutoNOM, the posterior probability of two change-points is 97.93% and the posterior means of the change-point locations are $\hat{\mathbb{E}}(s_1 \,|\, k = 2, \boldsymbol{y}) = 251.19$ and $\hat{\mathbb{E}}(s_2 \,|\, k = 2, \boldsymbol{y}) = 401.56$. The estimated location of the frequency peaks for our proposed procedure in comparison to AdaptSPEC and AutoPARM and the true values are shown in Figure (2.3) (bottom panels). It is evident that the proposed and existing methodologies successfully identify the true location of the frequency peaks in each segment, with

AdaptSPEC showing less precision.

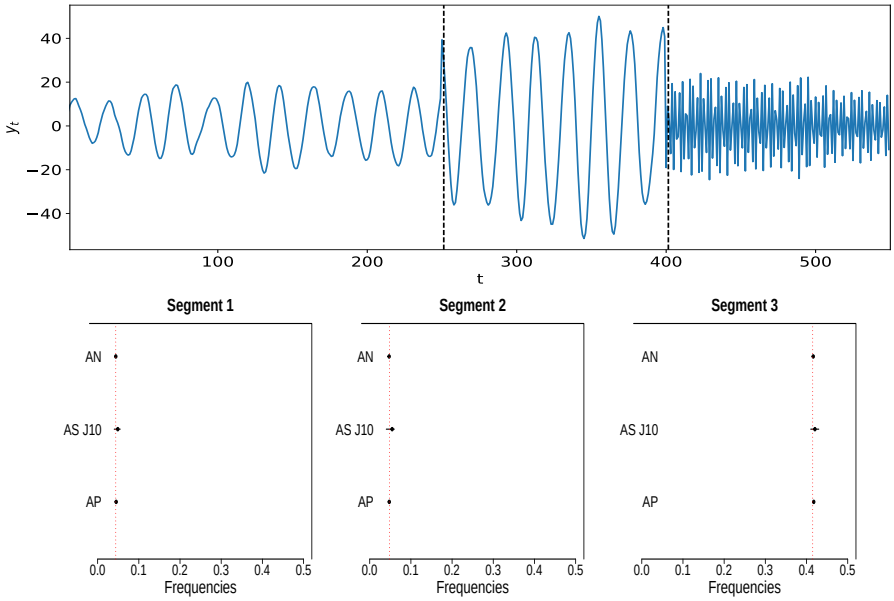

Figure 2.3: Piecewise AR process. (Top) A realization from model (2.15). Vertical dotted lines are the estimated locations of the change-points. (Bottom) Estimated frequency peaks for AutoNOM (AN), AdaptSPEC (AS $J = 10$) and AutoPARM (AP); 95% credible intervals (horizontal lines) are also reported for Bayesian methods. Dotted vertical lines are true locations of the frequency peaks.

2.4.2.2 Slowly Varying Autoregressive Process

In this section, we analyze an AR process whose continuous spectral density is changing slowly over time. We note though that this scenario is a large departure from the assumptions of our model. In particular, we consider the same slowly varying AR(2) process investigated by Ombao et al. [2001] and Davis et al. [2006], namely

$$y_t = a_t\, y_{t-1} - .81\, y_{t-2} + \varepsilon_t, \quad t = 1, \ldots, 1031, \tag{2.16}$$

where $a_t = .8\,[1 - .5\cos(\pi t/1031)]$ and $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$. Notice that the parameter a_t is changing gradually over time whereas the coefficient associated with the second lag remains constant. A realization from model (2.16) is shown in Figure

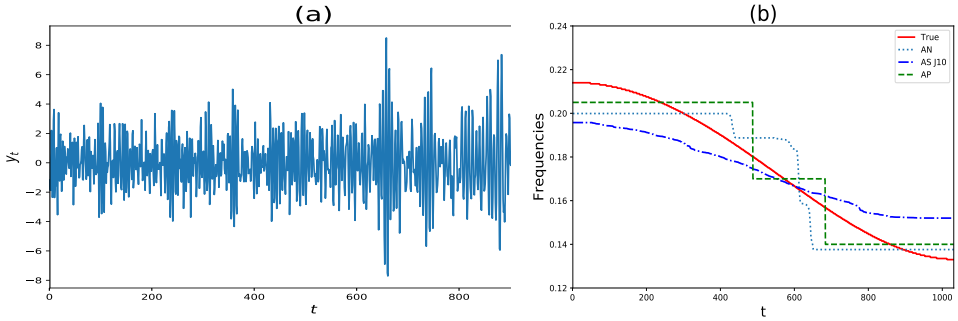

Figure 2.4: Slowly varying AR(2) process. (a) A realization from model (2.16). (b) True time varying frequency peak (solid line) and estimated time varying frequency peak for AutoNOM (AN), AdaptSPEC (AS $J = 10$) and AutoPARM (AP).

2.4 (a) and the corresponding time varying frequency peak is displayed in Figure 2.4 (b) as a solid line. Figure 2.4 (b) also shows the estimated time varying frequency peak for AutoNOM, AdaptSPEC and AutoPARM. For AutoNOM and AdaptSpec, the time changing frequency peak has been averaged across the MCMC samples, giving a smoother estimate (especially for AdaptSPEC) than the one obtained by AutoPARM. For each method, we compute the residual sum of squares $RSS = \sum_{t=1}^{1031}(\omega_t - \hat{\omega}_t)^2$ between the true time changing frequency peak ω_t and its estimate $\hat{\omega}_t$. The RSS in this example was 0.111, 0.174, 0.085 for AutoNOM, AdaptSPEC and AutoPARM, respectively. It is clear that even in this scenario where the data generating model was very different from the underlying assumptions of our model, our approach seems to outperform AdaptSPEC and remains competitive with AutoPARM in estimating the time varying frequency peak.

2.4.2.3 Non-Gaussian Time Series

We investigate the performance of our approach in the scenario when the innovations are t-distributed. We simulate a time series from the same simulation model presented in Section 2.4.1, where errors were generated from a t-distribution with 2, 3, and 2 degrees of freedom for the sequence of three segments, respectively.

The degrees of freedom were chosen low such that the corresponding distributions show heavy tails. A realization of this time series is shown in Figure 2.5. Our proposed methodology correctly identifies the 2 change-points, as the estimated posterior probability $\hat{\pi}(k = 2 \,|\, \boldsymbol{y})$ is 0.99. The posterior means of the change-point locations are $\hat{\mathbb{E}}(s_1 \,|\, k = 2, \boldsymbol{y}) = 303.6$ and $\hat{\mathbb{E}}(s_2 \,|\, k = 2, \boldsymbol{y}) = 650.5$, showing an excellent match to the true values $\boldsymbol{s}_{(2)} = (300, 650)$. Furthermore, the posterior mode of the number of frequencies in each segment is $\hat{\boldsymbol{m}}_{(2)} = (3, 1, 2)$, which is a correct estimate of $\boldsymbol{m}_{(2)} = (3, 1, 2)$. We also display in Figure 2.5 the estimated signal (using Equation (2.14)) as a dotted line. We can conclude that, although our model assumes Gaussianity, AutoNOM seems to perform well even in the case where the oscillatory underlying process is t-distributed with heavy tails.

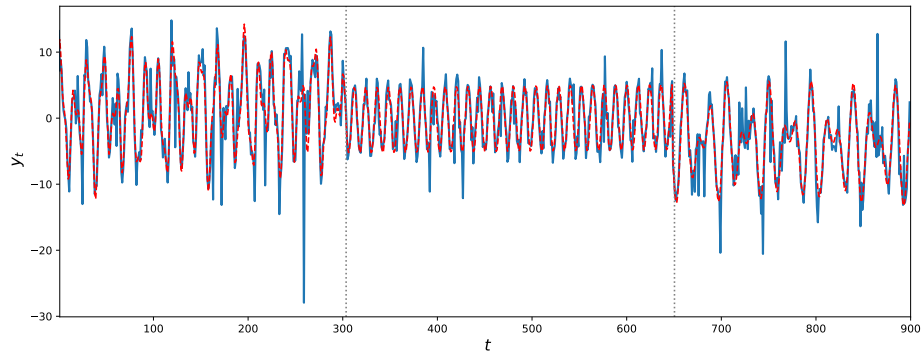

Figure 2.5: Illustrative example with t-distributed residual variances. Simulated time series (solid line) and estimated signal (dotted line). The dotted vertical lines represent the estimated location of the change-points.

2.5 Case Studies

The development of our methodology was motivated by the following two case studies where dense physiological signals were observed which exhibit unknown periodicities whose role may change over time in a more or less abrupt manner and

where their detection is of relevance to the health and well-being of the subject.

2.5.1 Analysis of Human Skin Temperature

The development of information and communication technologies, in particular widespread internet access and availability of mobile phones and tablets, allows considering new developments in the health care system. To address the issue of personalized medical treatment according to the circadian timing system of the patient, referred to as *chronotherapy* [Lévi and Schibler, 2007], a novel and validated non-invasive mobile e-Health platform pioneered by the French project PiCADo (Komarzynski et al. 2018) is used to record and teletransmit skin surface temperature as well as physical activity data [Huang et al., 2018] from an upper chest e-sensor. Figure 2.6 (a) shows an example of 4 days of 5-minutes temperature recording for a healthy individual. The circadian rhythms in core and skin surface temperature are usually 8-12 hours out of phase, with respective maxima occurring near 16:00 at day time, and near 2:00 at night [Kräuchi and Wirz-Justice, 1994]. The early night drop in core body temperature, which is critical for triggering the onset of sleep [Van Someren, 2006], results from the vasodilation of the skin vessels and associated rise in skin surface temperature [Kräuchi, 2002]. Under the assumption of stationarity Komarzynski et al. [2018] analyzed the skin temperature time series identifying both strong 12 hours (circahemidian) and 24 hours (circadian) rhythms.

Here, we applied our methodology to the skin-temperature time series shown in Figure 2.6 (a) for 300,000 iterations, discarding the first 100,000 updates as burn-in. The maximum number of change-points $k_{\max}$ was set to 10, whereas the maximum number of frequencies per regime $m_{\max}$ was set to 5. The estimated number

of change-points had a mode at 7, with $\hat{\pi}(k = 7 \mid \boldsymbol{y}) = 0.97$ and their estimated posterior distributions are shown in Figure 2.6 (c). Inspecting them alongside the physical activity data we can see that the change points mainly correspond to the start and end points of the prolonged rest periods at nights showing that skin temperature alternates between day activity and night rest including sleep. Figure 2.7 shows the estimated posterior distribution of the frequencies for the sleep segments (2, 4, 6, 8) along with the square root of the estimated power of the corresponding frequencies, where the power of each is frequency $\omega_{j,l}$ is summarized by the sum of squares of the corresponding linear coefficients, i.e. $I(\omega_{j,l}) = \beta_{j,l}^{(1)^2} + \beta_{j,l}^{(2)^2}$ [Shumway and Stoffer, 2005]. Figure 2.6 (b) shows the piecewise fitted signal, along with a 95 % credible interval obtained from the 2.5 and 97.5. Cycles of approximately 3 hours appear in segments 2, 4 and 6; cycles that range approximately 1-1.5 hour appear in segments 2, 4, 8 and cycles of around 2 hours appear in segments 4 and 6 while some longer periods identified in segments 4, 6, 8 indicate the presence of a trend.

Stages of sleep are characterized by ultradian oscillations between rapid eye movements (REM) and non-rapid eye movements (non-REM). The biological functionality that regulates the alternations between these two types of sleep is not yet much understood [Altevogt et al., 2006]. However, several physiological changes that occur over night differ between REM and non-REM phases, such as heart rate, brain activity, muscle tone and body temperature (Berlad et al., 1993; Pace-Schott and Hobson, 2002). The body cycles between REM and non-REM sleep stages with an average length that ranges approximately between 70 to 120 minutes, and there are usually four to six of these sleep cycles each night (Carskadon et al., 2005; Shneerson, 2009). Our analysis was able to use skin temperature data alone to detect periods of sleep throughout the day and identify oscillatory behaviour during

the night, whose frequencies are compatible with ultradian oscillations between REM and different non-REM sleep stages.

2.5.1.1 Comparison with Existing Methods

Here, we provide a comparison with the current state-of-the-art methods, AutoPARM and AdaptSPEC. AdaptSPEC was fitted with $J = 12$ basis functions and the results shown below are conditioned on the modal number of segments, whereas AutoPARM is performed with default tuning parameters. The estimated logarithm of the time-varying spectrum of the skin temperature time series is displayed in Figure 2.8, for both AutoPARM (top panel) and AdaptSPEC (bottom panel). The elements of the time-varying spectrum are functions of frequency and time, and the locations of the change-points are identified visually by inspecting the abrupt changes in power over the time axis. Broadly speaking, both AutoPARM and AdaptSPEC identify five segments and show some general agreement with each other in estimating change-points and local spectra. Both methodologies, which are based on continuous spectrum models, seem to smooth the local spectra at low frequencies in a considerable way. The only frequency peak is estimated by AutoPARM in the third segment and corresponds to a cycle of approximately 1.2 hour, which finds analogies with the spectral properties of Segment 4 estimated by our proposed approach. In particular, AutoPARM identifies the spectrum of an AR(2) process with autoregressive parameters (1.55, -0.69) in that segment. However, and most importantly, both existing methods clearly fail to detect either circadian or ultradian rhythmicity which were elicited by our method (see Figure 2.6 and Figure 2.7) and are to be expected as body temperature is known to be a circadian biomarker [Kräuchi and Wirz-Justice, 1994].

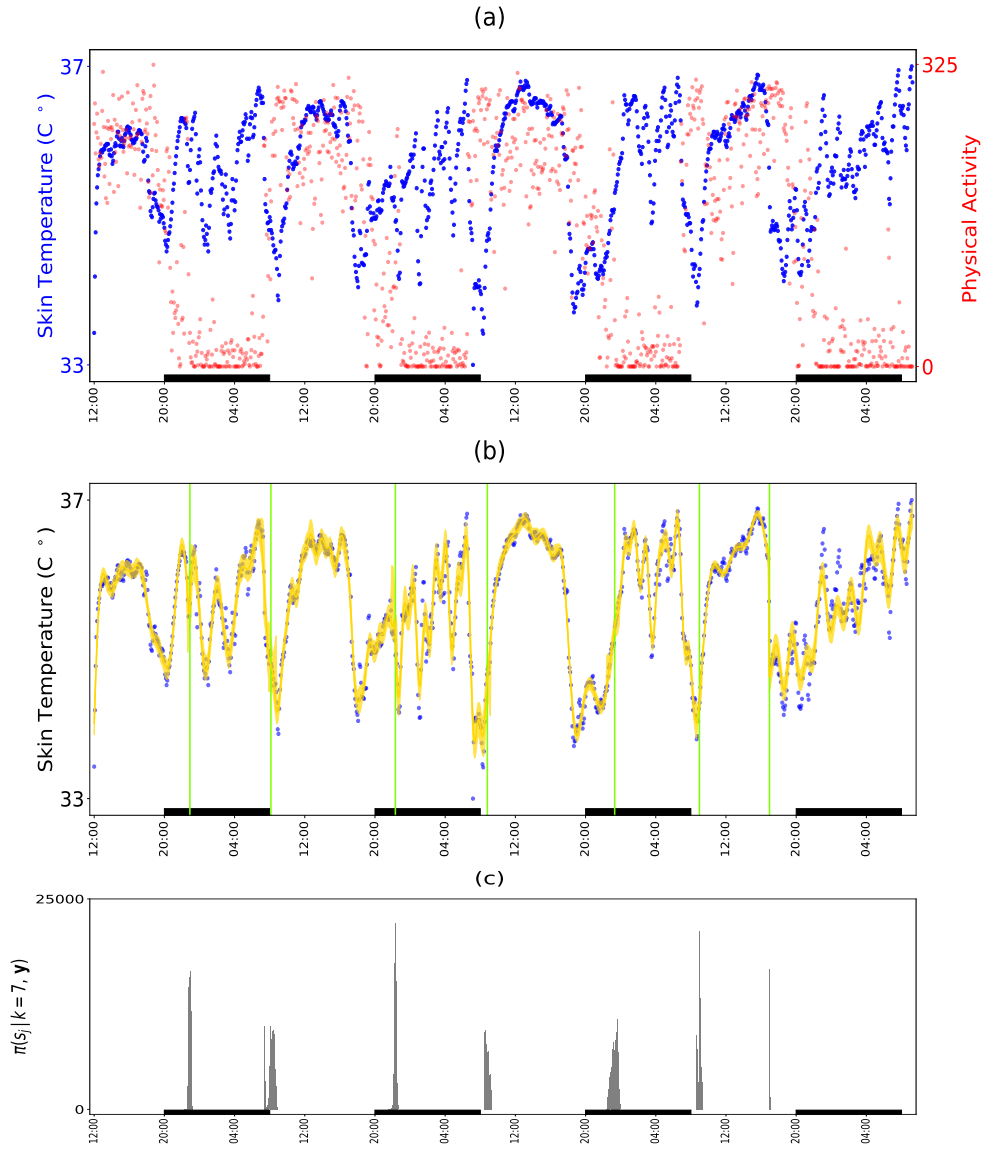

Figure 2.6: Analysis of skin temperature of a healthy subject. Panel (a) are the time series of skin temperature and corresponding physical activity. Panel (b) is the estimated signal (solid line) along with its 95% credible interval; vertical lines are the estimated locations of the change-points. Panel(c) is the estimated posterior density histogram of the locations of the changes, conditioned on $k = 7$ change-points. Rectangles on the time axis of each plot correspond to periods from 20.00 to 8.00. The variation in skin temperature finds analogies with the rest-activity pattern that alternates between day activity and night rest.

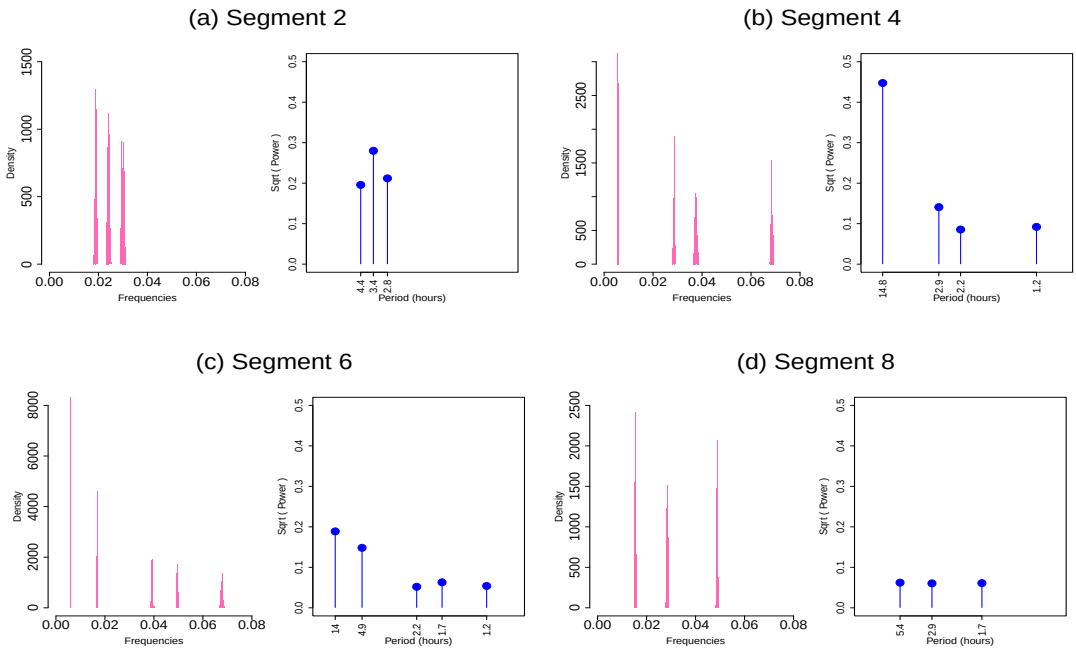

Figure 2.7: Spectral properties for segments corresponding to night rest. Estimated posterior distribution of the frequencies along with square root of the estimated power of the corresponding frequencies. The results are conditional on the modal number of frequencies per segment.

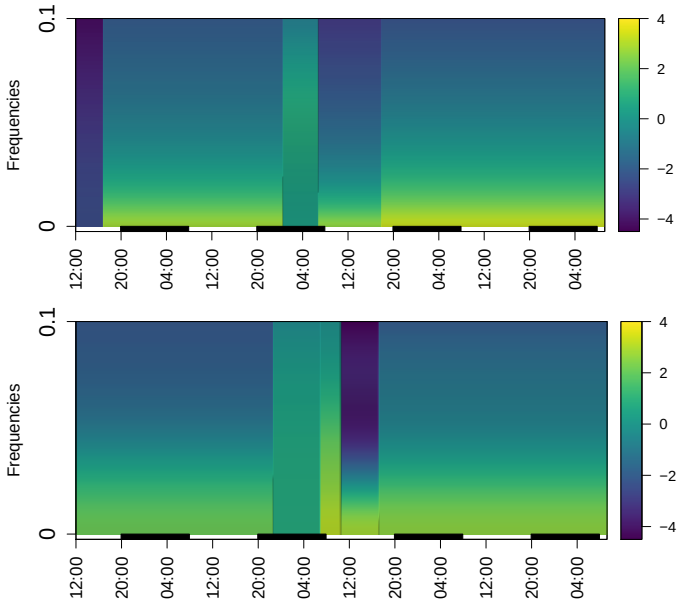

Figure 2.8: Estimated time-varying log spectrum for the skin temperature time series. (Top) AutoPARM. (Bottom) AdaptSPEC. Rectangles on the time axis of each plot correspond to periods from 20.00 to 8.00.

2.5.1.2 Phase Shift

We notice that in the framework of analyzing circadian biomarker data, such as body temperature, a change in acrophase may be of interest to the clinician as this may be indicative of a disruption of the bodyclock. Our proposed methodology can indeed be used to investigate phase since the sinusoidal function $f(t, \boldsymbol{\beta}_j, \boldsymbol{\omega}_j)$ that characterizes the j^{th} segment (see Equation 2.2)) can be re-written using trigonometric identities[1] as

$$f(t, \boldsymbol{B}_j, \boldsymbol{\omega}_j, \boldsymbol{\tau}_j) = \alpha_j + \mu_j t + \sum_{l=1}^{m_j} \Big(B_{j,l} \cos(2\pi\omega_{j,l} t + \tau_{j,l}) \Big),$$

where $\boldsymbol{B}_j = (B_{j,1}, \ldots, B_{j,m_j})$ and $\boldsymbol{\tau}_j = (\tau_{j,1}, \ldots, \tau_{j,m_j})$. With this notation, $B_{j,l}$ is the amplitude of the frequency $\omega_{j,l}$ and $\tau_{j,l}$ is the phase shift of the corresponding frequency. The phase $\tau_{j,l}$ of a frequency of interest $\omega_{j,l}$ can be estimated in terms of the coefficients $\beta_{j,l}^{(1)}$ and $\beta_{j,l}^{(2)}$ using the following equality

$$\tau_{j,l} = \arctan\left(-\frac{\beta_{j,l}^{(2)}}{\beta_{j,l}^{(1)}} \right), \qquad -\pi \leq \tau_{j,l} \leq \pi,$$

where credible intervals can be easily obtained from the empirical percentiles of the posterior sample.

[1] $\cos(a \pm b) = \cos(a)\cos(b) \mp \sin(a)\sin(b)$

2.5.2 Characterizing Instances of Sleep Apnea in Rodents

Sleep apnea is the temporary (≥ 2 breaths) interruption of breathing during sleep. Moderate or severe (≥ 15 events per hour) sleep apnea, occurs in about 50 % of men and 25 % of women over the age of 40 [Heinzer et al., 2015], with 91% of people with sleep apnea being undiagnosed [Tan et al., 2016]. Sleep apnea is linked to many diseases. Patients with sleep apnea are at increased risk of: cardiovascular events [Lanfranchi et al., 1999], cancer [Nieto et al., 2012], liver disease [Sundaram et al., 2016], diabetes [Harsch et al., 2004], metabolic syndrome [Parish et al., 2007], cognitive decline [Osorio et al., 2016], and increased risk of dementia in the elderly [Lal et al., 2012]. The motivation of this research is to provide a statistical methodology that can be applied to analyze large breathing data sets resulting from *in vivo* plethysmograph studies in rats to characterize the occurrence of sleep apnea under different experimental conditions. If this could be attained, a concrete aid to the understanding of the pathological implications of this status could be provided to clinicians and experimental biologists.

An unrestrained whole-body plethysmograph is used to produce a breathing trace from freely behaving rats for periods of up to 3 hours. Plethysmographs were made using an 2L air-tight box connected to a pressure transducer, with an air pump and outlet valve producing a flow rate of 2L/min. Airflow pressure signals were amplified using Neurolog system (Digitimer) connected to a 1401 interface and acquired on a computer using *Spike2* software (CED). Apneas are subclassified as post-sigh apneas, if the preceding breath was at least 25% above the average amplitude of prior breaths, or spontaneous apneas, if there was no manifestation of a previous sigh [Davis and ODonnell, 2013]. Airflow traces from the plethysmograph are shown in Figure 2.9 (left panels) and consist of three time series, which will be

referred to as (a), (b) and (c). They correspond to different actions for this rat: (a) an alternation of sniffing and normal breathing; (b) spontaneous apnea followed by normal breathing; (c) normal breathing followed by a sigh, and a post-sigh apnea. We note that these actions were classified by eye by an experienced experimental researcher. Each time series contains 20,000 observations where the signal was sampled at 2000 Hz so that we have 2000 observations per second.

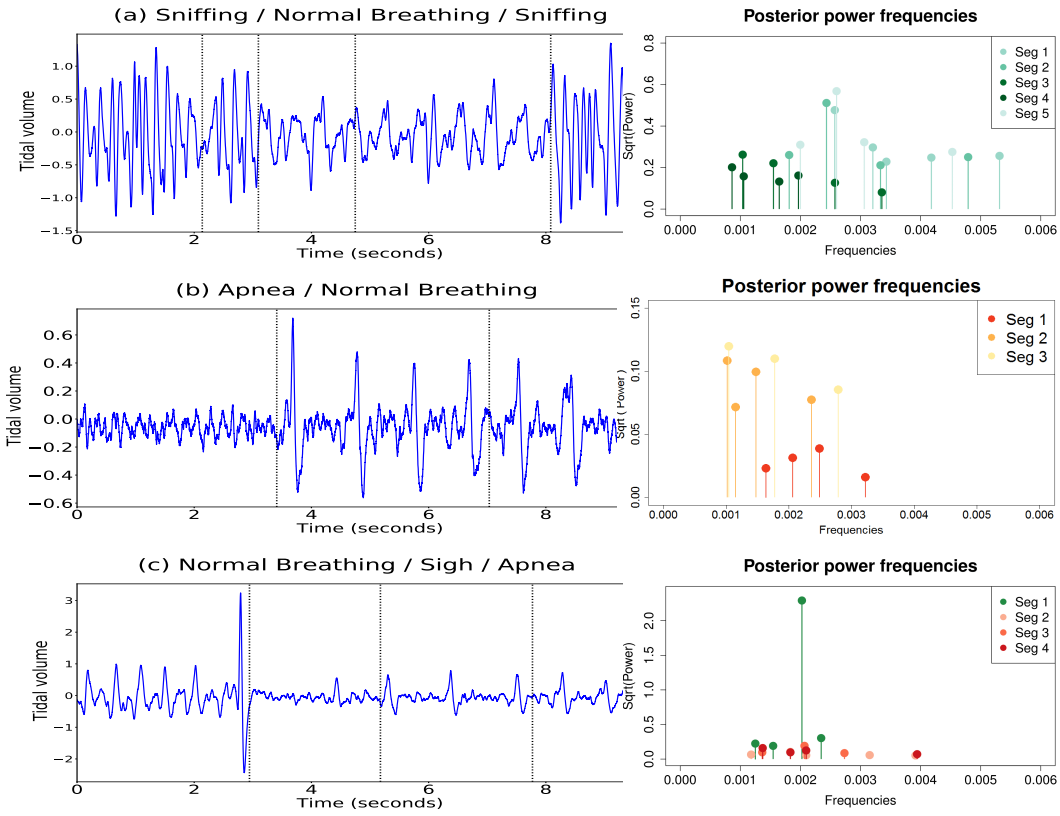

Figure 2.9: Plots of the respiratory traces of a rat (left panels) and corresponding estimated posterior power (right panels). Panel (a) is characterised by an alternation of sniffing and normal breathing. Panel (b) is a plot of the trace of a spontaneous apnea, followed by normal breathing. Panel (c) shows normal breathing followed by a sigh, and a post-sigh apnea. Dotted vertical lines correspond to the estimated locations of the change-points.

Our procedure allows us to set an upper bound, ϕ_ω, (Appendix 2.A) for the uniform interval where the new frequencies are sampled. As the periodogram ordinates for these data were approximately zero for all frequencies larger than 0.01, we decided accordingly to set $\phi_\omega = 0.01$. The locations of the changes (vertical lines) are displayed in Figure 2.9 (left panels). The posterior power of the frequencies, for each time series, is shown in Figure 2.9 (right panels). These results are conditional on the modal number of change-points and the modal number of frequencies per segment.

Table 2.6: Spectral properties of respiratory traces of a rat. Periodicities (in seconds) corresponding to the first two largest values of the estimated power, for each time series (a), (b) and (c).

	(a)		**(b)**		**(c)**	
	Period	Power	Period	Power	Period	Power
Segment 1	.19	.2272	.20	.0015	.25	5.2540
	.15	.0883	.25	.0010	.21	.0915
Segment 2	.20	.2613	.49	.0117	.27	.0107
	.28	.0686	.34	.0099	.42	.0043
Segment 3	.48	.0692	.48	.0145	.24	.0365
	.32	.0491	.28	.0122	.37	.0095
Segment 4	.58	.0400	-	-	.36	.0253
	.47	.0251			.24	.0155
Segment 5	.19	.3231	-	-	-	-
	.16	.1044				

For each data set, we summarise in Table 2.6 the spectral properties of each partition by displaying the periodicities corresponding to the first two largest values of the estimated power. When the rat is sniffing, (a), the air flow trace oscillates with a dominant period of approximately 0.2 seconds, namely 5 cycles per second. Normal breathing, (a) and (b), is characterised by lower frequencies and lower magnitude than sniffing, by oscillating with a dominant period of around 0.5 seconds, namely around 2 cycles every second. Apneas, (b) and (c), appear to be characterised by higher frequencies than normal breathing but with a lower power, with dominant periods of around 0.25 and 0.35 seconds. Notice that in the first partition

of (c), the highest value of the power corresponds to the frequency responsible for a sigh before apnea. Moreover, our methodology identifies different frequencies that explain the variation between the third and fourth partition of (c), leading to the hypothesis that there might be a time changing spectrum during the occurrence of an apnea instance. A comparison of our results with the results from AutoPARM and AdaptSPEC is provided in the Supplementary Material, Section 4.2.

2.5.2.1 Comparison with Existing Methods

The estimated time-varying spectral properties for three plethysmographic respiratory traces of the rat are displayed in Figure 2.10, for both AutoPARM (center panels) and AdaptSPEC (right panels). AutoPARM appears to identify fairly well changes in actions of this rat, such as (a) the alternation between sniffing and normal breathing, (b) the change from normal breathing to a spontaneous apnea, and (c) different actions of apnea, sigh and post-sigh apnea. We note that these actions were classified by eye by an experienced experimental researcher. AdaptSPEC detects changes in a satisfactory way for (a) and (c) but does not detect which distinct frequencies drive the oscillations in these data in particular as all peaks corresponding to low frequencies are smoothed. Notice that the periodogram ordinates for these time series were approximately zero for all frequencies larger than 0.01. In addition, AdaptSPEC is not able to detect any changes from normal breathing to a spontaneous apnea since it identifies only one segment in (b). It seems that the AR building block of AutoPARM can better model peaked structures compared to the smoothing spline nature of AdaptSPEC. Generally, our method find a larger number of change-points which are associated with changes in the spectrum, as seen in Figure 2.10 (left panels). For example, in (c) AutoNOM identifies different frequencies that explain the variation between Segment 3 and Segment 4, leading to

the hypothesis that there might be a time changing spectrum during the occurrence of an apnea instance.

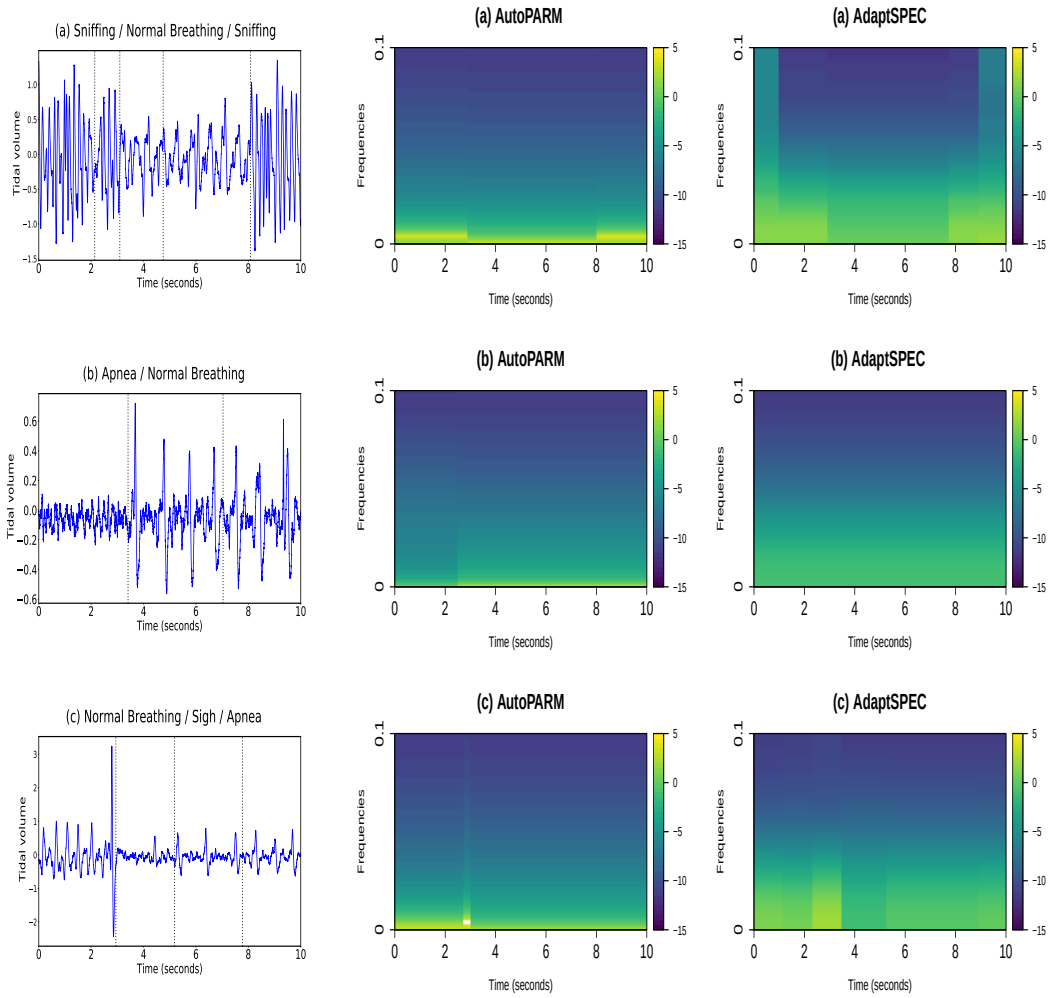

Figure 2.10: (Left) Plots of respiratory traces of a rat along with estimated change-points locations by AutoNOM. (a) is characterised by an alternation of sniffing and normal breathing. (b) is a plot of the trace of a spontaneous apnea, followed by normal breathing. (c) shows normal breathing followed by a sigh, and a post-sigh apnea. (Center) AutoPARM estimated time-varying log spectra for three different respiratory traces of a rat. (Right) AdaptSPEC estimated time-varying log spectra for three different respiratory traces of a rat.

2.A Details of the Sampling Scheme

In this section we provide further details about the sampling scheme for nonstationary periodic processes explained in Section 2.3. Section 2.A.1 describes a segment model move and Section 2.A.2 shows a change-point model move. In Section 2.A.3 we provide more details about Jacobian evaluations which are required to compute the (trans-dimensional) acceptance probabilities for both segment and change-point model moves.

2.A.1 Updating a Segment Model

2.A.1.1 Within-Model Move

Sampling $\boldsymbol{\omega}$: To obtain samples from the conditional posterior distribution $\pi(\boldsymbol{\omega} \,|\, \boldsymbol{\beta}, \sigma^2, m, \boldsymbol{y})$ (see Equation (2.10)), we draw the frequencies one-at-time using a mixture of M-H steps. In order to explore the parameter space efficiently, we design a mixture distribution $q(\omega_l^p \,|\, \omega_l^c)$, so that

$$q(\omega_l^p \,|\, \omega_l^c) = \delta_\omega \, q_1(\omega_l^p \,|\, \omega_l^c) + (1 - \delta_\omega) \, q_2(\omega_l^p \,|\, \omega_l^c), \qquad l = 1, \ldots, m, \quad (2.17)$$

where q_1 is defined in Equation (2.18) below, q_2 is the density of a univariate Normal $\mathcal{N}(\omega_l^c, \sigma_\omega^2)$, δ_ω is a positive real number such that $0 \leq \delta_\omega \leq 1$, and c and p refer to current and proposed values, respectively. According to Equation (2.17) we carry out with probability δ_ω a M-H step with proposal distribution $q_1(\omega_l^p \,|\, \omega_l^c)$,

$$q_1(\omega_l^p \,|\, \omega_l^c) \propto \sum_{h=0}^{\tilde{n}-1} I_h \, \mathbb{1}_{\left[h/n \,\leq\, \omega_l^p \,<\, (h+1)/n\right]}, \quad (2.18)$$

where $\tilde{n} = \lfloor n/2 \rfloor$ and I_h is the value of the squared modulus of the Discrete Fourier transform of the observations $\boldsymbol{y}$ evaluated at frequency h/n

$$I_h = \Big| \sum_{j=a}^{b} y_j \, \exp\Big(- i\, 2\pi \, \frac{h}{n} \Big) \Big|^2 .$$

In this way frequencies are proposed from regions in parameter space with high posterior density, yielding a Markov chain which converges quickly to its invariant distribution. The proposal distribution $q_1 (\omega_l^p \,|\, \omega_l^c)$ is independent of the current state ω_l^c. The acceptance probability for this move is

$$\alpha = \min \left\{ 1, \frac{\pi (\boldsymbol{\omega}^p \,|\, \boldsymbol{\beta},\, \sigma^2,\, m,\, \boldsymbol{y})}{\pi (\boldsymbol{\omega}^c \,|\, \boldsymbol{\beta},\, \sigma^2,\, m,\, \boldsymbol{y})} \times \frac{q_1 (\omega_l^c)}{q_1 (\omega_l^p)} \right\},$$

where $\boldsymbol{\omega}^p = (\omega_1^c, \ldots, \omega_{l-1}^c, \omega_l^p, \omega_{l+1}^c, \ldots, \omega_m^c)'$. On the other hand, with probability 1 - δ_ω, we perform a random walk M-H step with proposal distribution $q_2 (\omega_l^p \,|\, \omega_l^c)$, whose density is a univariate Normal distribution with mean ω_l^c and variance σ_ω^2, i.e. $\omega_l^p \,|\, \omega_l^c \sim \mathcal{N}(\omega_l^c, \sigma_\omega^2)$. This perturbation around the current value ω_l^c allows a local exploration of the conditional posterior distribution. The acceptance probability for this move is

$$\alpha = \min \left\{ 1, \frac{\pi (\boldsymbol{\omega}^p \,|\, \boldsymbol{\beta},\, \sigma^2,\, m,\, \boldsymbol{y})}{\pi (\boldsymbol{\omega}^c \,|\, \boldsymbol{\beta},\, \sigma^2,\, m,\, \boldsymbol{y})} \right\}.$$

Setting δ_ω to a relative low value integrates a fairly high acceptance rate with a quick exploration of the parameter space. For our experiments, we set $\sigma_\omega^2 = (1/50n)^2$ and $\delta_\omega = 0.2$.

2.A.1.2 Between-Model Moves

The number of frequencies on a segment is proposed to either increase or decrease by one. Let $\boldsymbol{\theta}^c = (\boldsymbol{\beta}^{c\prime}, \boldsymbol{\omega}^{c\prime}, \sigma^{2c\prime})'$ and assume the Markov chain is currently at $(m^c, \boldsymbol{\theta}^c)$. We propose a move to $(m^p, \boldsymbol{\theta}^p)$ by drawing from a proposal density $q(m^p, \boldsymbol{\theta}^p \,|\, m^c, \boldsymbol{\theta}^c)$ and accepting this update with probability

$$\alpha = \min\left\{1, \frac{\mathscr{L}(m^p, \boldsymbol{\theta}^p \,|\, \boldsymbol{y})}{\mathscr{L}(m^c, \boldsymbol{\theta}^c \,|\, \boldsymbol{y})} \times \frac{\pi(m^p)\,\pi(\boldsymbol{\theta}^p \,|\, m^p)}{\pi(m^c)\,\pi(\boldsymbol{\theta}^c \,|\, m^c)} \times \frac{q(m^c, \boldsymbol{\theta}^c \,|\, m^p, \boldsymbol{\theta}^p)}{q(m^p, \boldsymbol{\theta}^p \,|\, m^c, \boldsymbol{\theta}^c)}\right\}, \tag{2.19}$$

where we notice (and show in Section 2.A.3.1) that the Jacobian which takes account for the change of dimension is equal to one. The proposed state $(m^p, \boldsymbol{\theta}^p)$ is drawn by first drawing m^p, followed by $\boldsymbol{\omega}^p$, $\boldsymbol{\beta}^p$ and σ^{2p}. In fact, we can rewrite the proposal density as

$$\begin{aligned} q(m^p, \boldsymbol{\theta}^p \,|\, m^c, \boldsymbol{\theta}^c) &= q(m^p \,|\, m^c) \times q(\boldsymbol{\theta}^p \,|\, m^p, m^c, \boldsymbol{\theta}^c) \\ &= q(m^p \,|\, m^c) \times q(\boldsymbol{\omega}^p \,|\, m^p, m^c, \boldsymbol{\theta}^c) \\ &\quad \times q(\boldsymbol{\beta}^p \,|\, \boldsymbol{\omega}^p, m^p, m^c, \boldsymbol{\theta}^c) \\ &\quad \times q(\sigma^{2p} \,|\, \boldsymbol{\beta}^p, \boldsymbol{\omega}^p, m^p, m^c, \boldsymbol{\theta}^c). \end{aligned} \tag{2.20}$$

Birth move: If a birth move is proposed, we have that $m^p = m^c + 1$. The proposed frequency vector $\boldsymbol{\omega}^p$ is constructed as

$$\boldsymbol{\omega}^p = (\omega_1^c, \ldots, \omega_{m^c}^c, \omega_{m^p}^p)',$$

namely by keeping the current vector of frequencies and proposing an additional frequency $\omega_{m^p}^p$. Alternatively to Andrieu and Doucet [1999], we choose to sample $\omega_{m^p}^p$ uniformly on the interval $(0, \phi_\omega)$, where we recall that $0 < \phi_\omega < 0.5$. Addi-

tionally, for computational and/or modelling reasons, we would like not to sample frequencies that are too close to each other. Hence, we choose to draw a candidate value $\omega^p_{m^p}$ uniformly from the union of intervals of the form $[\omega^c_l + \psi_\omega, \omega^c_{l+1} - \psi_\omega]$, for $l = 0, \ldots, m_c$ and denoting $\omega^c_0 = 0$ and $\omega^c_{m^c+1} = \phi_\omega$. Here, ψ_ω is a fixed minimum distance between frequencies, which is chosen larger than $\frac{1}{n}$; in fact, when the separation of two frequencies is less than the *Nyquist step* [Priestley, 1981], i.e. $|\omega_l - \omega_{l+1}| < \frac{1}{n}$, the two frequencies are indistinguishable [Dou and Hodgson, 1995]. Moreover, we sort the proposed vector of frequencies $\boldsymbol{\omega}^p$ to ensure identifiability and perform practical estimation, as suggested in Andrieu and Doucet [1999]. For proposed $\boldsymbol{\omega}^p$ and given σ^{2c}, the proposed vector of linear coefficients $\boldsymbol{\beta}^p$ is drawn from its conjugate Gaussian posterior (as in Equation 2.11). Finally, the residual variance σ^{2p} is sampled directly from its posterior conditional distribution $\pi(\sigma^{2p} \,|\, \boldsymbol{\omega}^p, \boldsymbol{\beta}^p, m^p, \boldsymbol{y})$ (see Equation (2.13)). The proposed state $(m^p, \boldsymbol{\theta}^p)$ is accepted or rejected in a Metropolis-Hastings step with probability

$$\alpha = \min\left\{1, \frac{\mathscr{L}(\boldsymbol{\theta}^p, m^p \,|\, \boldsymbol{y})}{\mathscr{L}(\boldsymbol{\theta}^c, m^c \,|\, \boldsymbol{y})} \times \frac{\pi(m^p)\,\pi(\boldsymbol{\theta}^p \,|\, m^p)}{\pi(m^c)\,\pi(\boldsymbol{\theta}^c \,|\, m^c)} \times \frac{d_{m^p} \cdot \left(\frac{1}{m^p}\right) \cdot q(\boldsymbol{\beta}^c) \cdot q(\sigma^{2c})}{b_{m^c} \cdot q(\omega^p_{m^p}) \cdot q(\boldsymbol{\beta}^p) \cdot q(\sigma^{2p})}\right\},$$

where the likelihood function is given in Equation (2.5), $\pi(m)$ is the density of the Poisson distribution truncated at $m_{\max}$, b_{m^c} and d_{m^p} are defined in Equation (2.9), $q(\boldsymbol{\beta}^c)$ and $q(\boldsymbol{\beta}^p)$ are the Normal densities $\boldsymbol{\mathcal{N}}_{2m^c+2}(\hat{\boldsymbol{\beta}}^c, \boldsymbol{V}_{\beta^c})$ and $\boldsymbol{\mathcal{N}}_{2m^p+2}(\hat{\boldsymbol{\beta}}^p, \boldsymbol{V}_{\beta^p})$, respectively (Equation 2.11); $q(\sigma^2_p)$ and $q(\sigma^2_c)$ are the Inverse-Gamma proposal densities defined in Equation (2.13).

Death move: If a death move is proposed, then $m^p = m^c - 1$. A vector of frequencies $\boldsymbol{\omega}^p$ is constructed by randomly selecting with probability $\frac{1}{m^c}$ one of the current frequencies as the candidate frequency for removal. Given $\boldsymbol{\omega}^p$ and σ^{2c}, a vector of linear coefficients $\boldsymbol{\beta}^p$ is drawn from its Gaussian posterior conditional

distribution. Finally, conditioned on $\boldsymbol{\omega}^p$ and $\boldsymbol{\beta}^p$, the residual variance σ^{2p} is drawn from its posterior Inverse-Gamma distribution. It is straightforward to see that the acceptance probability for the death step has the same form as the birth step, with the proper change of labelling of the variables, and the ratio terms inverted.

2.A.2 Updating a Change-Point Model

2.A.2.1 Within-Model Move

Let $\boldsymbol{s}^c_{(k)} = (s^c_1, \ldots, s^c_k)'$ be the current vector of change-points locations, $\boldsymbol{m}^c_{(k)} = (m^c_1, \ldots, m^c_{k+1})'$ be the current vector of number of frequencies, $\boldsymbol{\omega}^c_{(k)} = (\boldsymbol{\omega}^{c'}_1, \ldots, \boldsymbol{\omega}^{c'}_{k+1})'$ be the current vector of frequencies. Let $\boldsymbol{\beta}^c_{(k)} = (\boldsymbol{\beta}^{c'}_1, \ldots, \boldsymbol{\beta}^{c'}_{k+1})'$ and $\boldsymbol{\sigma}^2_{(k)} = (\sigma^{2c}_1, \ldots, \sigma^{2c}_{k+1})'$ be the current vectors of linear coefficients and residual variances, respectively.

Let us also define $\boldsymbol{\theta}^c_{(k)} = (\boldsymbol{\beta}^{c'}_{(k)}, \boldsymbol{\omega}^{c'}_{(k)}, \boldsymbol{\sigma}^{2c'}_{(k)})'$. Following Green [1995], a change-point, s^c_j say, is randomly selected with probability $\frac{1}{k}$ from the existing set of change-points. In order to explore the parameter space in an efficient way and similar to above we construct a mixture distribution $q(s^p_j \mid s^c_j)$, as

$$q(s^p_j \mid s^c_j) = \delta_s \, q_1(s^p_j \mid s^c_j) + (1 - \delta_s) \, q_2(s^p_j \mid s^c_j) \tag{2.21}$$

where q_1 is the density of a Uniform $[s^c_{j-1} + \psi, s^c_{j+1} - \psi]$, q_2 is the density of a univariate Normal $\mathcal{N}(s^c_j, \sigma^2_s)$ and δ_s is a positive real number such that $0 \leq \delta_s \leq 1$. We propose with probability δ_s a candidate value s^p_j from the above uniform distribution where ψ is a fixed minimum time between change-points avoiding change-points being too close to each other. On the other hand, with probability $(1 - \delta_s)$, s^p_j arises as a Normal random walk proposal centered at the current change-point s^c_j.

The proposed vector of change-points locations is denoted by

$$\boldsymbol{s}^{p}_{(k)} = (s^{c}_{1}, \ldots, s^{c}_{j-1}, s^{p}_{j}, s^{c}_{j+1}, \ldots, s^{c}_{k})',$$

and hence the proposed value s^{p}_{j} induces a new proposed data partition on $[s^{c}_{j-1}, s^{c}_{j+1}]$ corresponding to $[s^{c}_{j-1}, s^{p}_{j})$ and $[s^{p}_{j}, s^{c}_{j+1})$. We denote the vectors of observations belonging to these two proposed segments as $\boldsymbol{y}^{p}_{j}$ and $\boldsymbol{y}^{p}_{j+1}$, which include n^{p}_{j} and n^{p}_{j+1} observations, respectively. Given $\boldsymbol{s}^{p}_{(k)}$, the proposed number of frequencies m^{p}_{j}, m^{p}_{j+1} are set equal to the current ones m^{c}_{j}, m^{c}_{j+1}, so that $\boldsymbol{m}^{p}_{(k)} = \boldsymbol{m}^{c}_{(k)}$. Similarly, the proposed pair of frequency vectors $\boldsymbol{\omega}^{p}_{j}$, $\boldsymbol{\omega}^{p}_{j+1}$ is chosen equal to the current pair $\boldsymbol{\omega}^{c}_{j}$, $\boldsymbol{\omega}^{c}_{j+1}$ in the corresponding segments, i.e. $\boldsymbol{\omega}^{p}_{(k)} = \boldsymbol{\omega}^{c}_{(k)}$. The proposed vectors $\boldsymbol{\beta}^{p}_{j}$, $\boldsymbol{\beta}^{p}_{j+1}$ are sampled from their Gaussian posterior conditional distributions $\pi(\boldsymbol{\beta}^{p}_{j} \,|\, \boldsymbol{\omega}^{p}_{j}, \sigma^{2c}_{j}, m^{p}_{j}, \boldsymbol{y}^{p}_{j})$ and $\pi(\boldsymbol{\beta}^{p}_{j+1} \,|\, \boldsymbol{\omega}^{p}_{j+1}, \sigma^{2c}_{j+1}, m^{p}_{j+1}, \boldsymbol{y}^{p}_{j+1})$. This move is accepted in a M-H step with probability

$$\alpha = \min\left\{1, \frac{\mathscr{L}(k, \boldsymbol{m}^{p}_{(k)}, \boldsymbol{s}^{p}_{(k)}, \boldsymbol{\theta}^{p}_{(k)} \,|\, \boldsymbol{y})}{\mathscr{L}(k, \boldsymbol{m}^{c}_{(k)}, \boldsymbol{s}^{c}_{(k)}, \boldsymbol{\theta}^{c}_{(k)} \,|\, \boldsymbol{y})} \times \frac{\pi(\boldsymbol{s}^{p}_{(k)} \,|\, k)\, \pi(\boldsymbol{\theta}^{p}_{(k)} \,|\, \boldsymbol{m}^{p}_{(k)}, k)}{\pi(\boldsymbol{s}^{c}_{(k)} \,|\, k)\, \pi(\boldsymbol{\theta}^{c}_{(k)} \,|\, \boldsymbol{m}^{c}_{(k)}, k)} \times \frac{\prod_{h=j}^{j+1} q(\boldsymbol{\beta}^{c}_{h})}{\prod_{h=j}^{j+1} q(\boldsymbol{\beta}^{p}_{h})}\right\},$$

where the likelihood is specified in Equation (2.4) and $q(\boldsymbol{\beta}^{c}_{h})$ and $q(\boldsymbol{\beta}^{p}_{h})$ are the multivariate Normal densities $\boldsymbol{\mathcal{N}}_{2m^{c}_{h}+2}(\hat{\boldsymbol{\beta}}^{c}_{h}, \boldsymbol{V}_{\beta^{c}_{h}})$ and $\boldsymbol{\mathcal{N}}_{2m^{p}_{h}+2}(\hat{\boldsymbol{\beta}}^{p}_{h}, \boldsymbol{V}_{\beta^{p}_{h}})$, respectively (Equation 2.11), for $h = \{j, j+1\}$. Note that the likelihood ratio and the prior ratio differ from one only for the two segments affected by the move of the change-points. Next, the residual variances σ^{2p}_{j}, σ^{2p}_{j+1} are drawn from their posterior conditional distributions $\pi(\sigma^{2p}_{j} \,|\, \boldsymbol{\omega}^{p}_{j}, \boldsymbol{\beta}^{p}_{j}, m^{p}_{j}, \boldsymbol{y}^{p}_{j})$, $\pi(\sigma^{2p}_{j+1} \,|\, \boldsymbol{\omega}^{p}_{j+1}, \boldsymbol{\beta}^{p}_{j+1}, m^{p}_{j+1}, \boldsymbol{y}^{p}_{j+1})$ in a Gibbs step.

2.A.2.2 Between-Model Moves

Let $\boldsymbol{\xi}^{c}_{(k^c)} = (\boldsymbol{s}^{c'}_{(k^c)}, \boldsymbol{m}^{c'}_{(k^c)}, \boldsymbol{\theta}^{c'}_{(k^c)})'$ and assume the Markov chain is at $(k^c, \boldsymbol{\xi}^{c}_{(k^c)})$. We propose a move to $(k^p, \boldsymbol{\xi}^{p}_{(k^p)})$ by first drawing k^p, followed by sampling the change-point locations $\boldsymbol{s}^{p}_{(k^p)}$. The latter involves either merging two segments (death) or splitting a segment (birth). The number of frequencies and their values in the proposed segments are selected from the current state. We draw $\boldsymbol{\beta}^{p}_{(k^p)}$ and jointly update the entire state $(k^p, \boldsymbol{\xi}^{p}_{(k^p)})$. Hence, we propose a move to $(k^p, \boldsymbol{\xi}^{p}_{(k^p)})$ by drawing from a proposal density of the form

$$
\begin{aligned}
q(k^p, \boldsymbol{\xi}^{p}_{(k^p)} \,|\, k^c, \boldsymbol{\xi}^{c}_{(k^c)}) &= q(k^p \,|\, k^c) \times q(\boldsymbol{\xi}^{p}_{(k^p)} \,|\, k^p, k^c, \boldsymbol{\xi}^{c}_{(k^c)}) \\
&= q(k^p \,|\, k^c) \times q(\boldsymbol{s}^{p}_{(k^p)} \,|\, k^p, k^c, \boldsymbol{\xi}^{c}_{(k^c)}) \\
&\quad \times q(\boldsymbol{m}^{p}_{(k^p)}, \boldsymbol{\omega}^{p}_{(k^p)} \,|\, \boldsymbol{s}^{p}_{(k^p)}, k^p, k^c, \boldsymbol{\xi}^{c}_{(k^c)}) \\
&\quad \times q(\boldsymbol{\sigma}^{2p}_{(k^p)} \,|\, \boldsymbol{m}^{p}_{(k^p)}, \boldsymbol{\omega}^{p}_{(k^p)}, \boldsymbol{s}^{p}_{(k^p)}, k^p, k^c, \boldsymbol{\xi}^{c}_{(k^c)}) \\
&\quad \times q(\boldsymbol{\beta}^{p}_{(k^p)} \,|\, \boldsymbol{\sigma}^{2p}_{(k^p)}, \boldsymbol{m}^{p}_{(k^p)}, \boldsymbol{\omega}^{p}_{(k^p)}, \boldsymbol{s}^{p}_{(k^p)}, k^p, k^c, \boldsymbol{\xi}^{c}_{(k^c)}).
\end{aligned}
$$

Birth move: If a birth move is proposed, we have that $k^p = k^c + 1$. We draw a new change-point uniformly on $f(\boldsymbol{s}^{c}_{(k^c)}, \psi_s)$, the support of $\boldsymbol{s}^{c}_{(k^c)}$ given the constraints imposed by ψ_s, i.e. $f(\boldsymbol{s}^{c}_{(k^c)}, \psi_s) = [1 + \psi_s, s^c_1 - \psi_s] \cup [s^c_1 + \psi_s, s^c_2 - \psi_s] \cup \ldots \cup [s^c_{k^c} + \psi_s, n - \psi_s]$. Hence, the new proposed location $\tilde{s}_j$ is sampled from a Uniform$\{f(\boldsymbol{s}_{(k^c)}, \psi_s)\}$, where the proposal density is given by

$$
q(\boldsymbol{s}^{p}_{(k^p)} \,|\, k^p, k^c, \boldsymbol{\xi}^{c}_{(k^c)}) = \frac{1}{(n - 2\psi_s(k^c + 1) - 1)}. \tag{2.22}
$$

As the proposed location $\tilde{s}_j$ will lie within an existing interval $(s_j^c,\ s_{j+1}^c)$ with probability one, we can define the proposed change-points location vector as

$$\boldsymbol{s}^{p}_{(k^p)} = (s_1^c, \ldots, s_j^c, \tilde{s}_j, s_{j+1}^c, \ldots, s_{k^c}^c)'.$$

The number of frequencies m_j^p, m_{j+1}^p corresponding to the two newly proposed segments $[s_j^c, \tilde{s}_j)$ and $[\tilde{s}_j, s_{j+1}^c)$ are set equal to the current number of frequencies on the whole segment $(s_j^c,\ s_{j+1}^c)$. Therefore, we can construct the proposed vector of the number of frequencies $\boldsymbol{m}^{p}_{(k^p)}$ and the proposed vector of frequencies $\boldsymbol{\omega}^{p}_{(k^p)}$ as

$$\begin{aligned}
\boldsymbol{m}^{p}_{(k^p)} &= (m_1^c, \ldots, m_{j-1}^c, m_j^c, m_j^c, m_{j+1}^c, \ldots, m_{k^c+1}^c)', \\
\boldsymbol{\omega}^{p}_{(k^p)} &= (\boldsymbol{\omega}_1^{c'}, \ldots, \boldsymbol{\omega}_{j-1}^{c'}, \boldsymbol{\omega}_j^{c'}, \boldsymbol{\omega}_j^{c'}, \boldsymbol{\omega}_{j+1}^{c'}, \ldots, \boldsymbol{\omega}_{k^c+1}^{c'})'.
\end{aligned}$$

The proposed vector of residual variances $\boldsymbol{\sigma}^{2p}_{(k^p)}$ is

$$\boldsymbol{\sigma}^{2p}_{(k^p)} = (\sigma_1^{2c}, \ldots, \sigma_{j-1}^{2c}, \sigma_j^{2p}, \sigma_{j+1}^{2p}, \sigma_{j+1}^{2c}, \ldots, \sigma_{k^c+1}^{2c})',$$

where the residual variances $\sigma_j^{2p}, \sigma_{j+1}^{2p}$ for the split partition are constructed following Green [1995], namely as a perturbation of the current variance σ_j^{2c}. Specifically, we draw $u \sim \text{Uniform}\,(0, 1)$ and let $\sigma_j^{2p}, \sigma_{j+1}^{2p}$ be deterministic transformations of σ_j^{2c}, i.e

$$\sigma_j^{2p} = \frac{u}{1-u}\sigma_j^{2c}, \qquad \sigma_{j+1}^{2p} = \frac{1-u}{u}\sigma_j^{2c}. \tag{2.23}$$

Finally, the proposed vector of linear coefficients $\boldsymbol{\beta}^{p}_{(k^p)}$ is

$$\boldsymbol{\beta}^{p}_{(k^p)} = (\boldsymbol{\beta}_1^{c'}, \ldots, \boldsymbol{\beta}_{j-1}^{c'}, \boldsymbol{\beta}_j^{p'}, \boldsymbol{\beta}_{j+1}^{p'}, \boldsymbol{\beta}_{j+1}^{c'}, \ldots, \boldsymbol{\beta}_{k^c+1}^{c'})',$$

where the vectors $\boldsymbol{\beta}_j^p, \boldsymbol{\beta}_{j+1}^p$ are drawn from their posterior conditional distribution. The proposed move to the state $(k^p, \boldsymbol{\xi}_{(k^p)}^p)$ is accepted with probability

$$\alpha = \min\left\{1, \frac{\mathscr{L}(k^p, \boldsymbol{\xi}_{(k^p)}^p \,|\, \boldsymbol{y})}{\mathscr{L}(k^c, \boldsymbol{\xi}_{(k^c)}^c \,|\, \boldsymbol{y})} \times \frac{\pi(k^p)\,\pi(\boldsymbol{\xi}_{(k^p)}^p \,|\, k^p)}{\pi(k^c)\,\pi(\boldsymbol{\xi}_{(k^c)}^c \,|\, k^c)} \times \frac{d_{k^p} \cdot \frac{1}{k^p} \cdot \frac{1}{2} \cdot q(\boldsymbol{\beta}_j^c)}{b_{k^c} \cdot q(\boldsymbol{s}_{(k^p)}^p) \cdot \prod_{h=j}^{j+1} q(\boldsymbol{\beta}_h^p)} \times \boldsymbol{J}_{\sigma^2}\right\},$$

where the likelihood function is provided in Equation (2.4), $q(\boldsymbol{s}_{(k^p)}^p)$ is the uniform density defined in Equation (2.22); $q(\boldsymbol{\beta}_h^c)$, $q(\boldsymbol{\beta}_h^p)$ are the multivariate Normal proposal densities as in Equation (2.11), and the Jacobian $\boldsymbol{J}_{\sigma^2}$ is

$$\boldsymbol{J}_{\sigma^2} = \left|\frac{\partial(\sigma_j^{2p}, \sigma_{j+1}^{2p})}{\partial(\sigma_j^{2c}, u)}\right| = 2\left(\sqrt{\sigma_j^{2p}} + \sqrt{\sigma_{j+1}^{2p}}\right)^2.$$

The evaluation of the Jacobian is provided in Section 2.A.3.2 and the numerator of the proposal ratio is better understood by looking at the details of the death step, which are given below.

Death Move: If a death step is proposed, then $k^p = k^c - 1$. A candidate change-point s_j^c to be removed is sampled uniformly from the vector of existing change-points; that is, we propose to remove s_j^c with probability $\frac{1}{k^c}$. Then, the proposed vector of change-points locations $\boldsymbol{s}_{(k)}^p$ is defined as

$$\boldsymbol{s}_{(k)}^p = (s_1^c, \ldots, s_{j-1}^c, s_{j+1}^c, \ldots, s_{k^c}^c)'.$$

The number m_j^p and the vector of relevant frequencies $\boldsymbol{\omega}_j^p$ of the newly merged segment $[s_{j-1}^c, s_{j+1}^c)$ are selected by drawing an index at random from $\{j, j+1\}$, obtaining say j^*, and setting the proposed parameters equal to the current ones relative to the selected index. That is, we set $m_j^p = m_{j^*}^c$ and $\boldsymbol{\omega}_j^p = \boldsymbol{\omega}_{j^*}^c$. Hence, the proposed vectors of number of frequencies $\boldsymbol{m}_{(k^p)}^p$ and their values $\boldsymbol{\omega}_{(k^p)}^p$ are constructed as

follows

$$\boldsymbol{m}^{p}_{(k^p)} = (\, m^{c}_{1}, \ldots, m^{c}_{j-1}, m^{p}_{j}, m^{c}_{j+2}, \ldots, m^{c}_{k^c+1}\,)^{\prime},$$
$$\boldsymbol{\omega}^{p}_{(k^p)} = (\, \boldsymbol{\omega}^{c^{\prime}}_{1}, \ldots, \boldsymbol{\omega}^{c^{\prime}}_{j-1}, \boldsymbol{\omega}^{p^{\prime}}_{j}, \boldsymbol{\omega}^{c^{\prime}}_{j+2}, \ldots, \boldsymbol{\omega}^{c^{\prime}}_{k^c+1})^{\prime}.$$

The residual variance σ^{2p}_{j} of the newly merged segment is obtained by inverting the transformation of Equation (2.23). Specifically, we construct $\sigma^{2p}_{j} = \sqrt{\sigma^{2c}_{j}\,\sigma^{2c}_{j+1}}$ and set the proposed vector of residual variances $\boldsymbol{\sigma}^{2p}_{(k^p)}$ as

$$\boldsymbol{\sigma}^{2p}_{(k^p)} = (\, \sigma^{2c}_{1}, \ldots, \sigma^{2c}_{j-1}, \sigma^{2p}_{j}, \sigma^{2c}_{j+2}, \ldots, \sigma^{2c}_{k^c+1}\,)^{\prime}.$$

The proposed vector of linear coefficients $\boldsymbol{\beta}^{p}_{(k^p)}$ is

$$\boldsymbol{\beta}^{p}_{(k^p)} = (\boldsymbol{\beta}^{c^{\prime}}_{1}, \ldots, \boldsymbol{\beta}^{c^{\prime}}_{j-1}, \boldsymbol{\beta}^{p^{\prime}}_{j}, \boldsymbol{\beta}^{c^{\prime}}_{j+2}, \ldots, \boldsymbol{\beta}^{c^{\prime}}_{k^c+1})^{\prime},$$

where the vector of coefficients $\boldsymbol{\beta}^{p}_{j}$ is drawn from its Gaussian posterior conditional distribution. The acceptance probability for the death step has the same form of the birth step, with the proper change of labelling of the variables, and the ratio terms inverted.

2.A.3 Jacobian Evaluation

2.A.3.1 Segment Model Move

The Jacobian term involved in Equation (2.19) is equal to one. In fact, in a similar fashion as of Godsill [2001], the proposal is made directly in the new parameter space rather than via the dimension matching random variables introduced by Green [1995]. Since equivalent results were stated but not proved by Godsill [2001] or

Rosen et al. [2012] among the others, we believe is worth making a brief digression and show that the Jacobian term is unity for a simple example.

Example: Assume the current dimension of the model is $m^c = 3$ with corresponding value of the parameters $\boldsymbol{\theta}^c = (x_1^c, x_2^c, x_3^c)$. We propose a birth move to a model with $m^p = 4$ and $\boldsymbol{\theta}^p = (x_1^p, x_2^p, x_3^p, x_4^p)$; the first three entries of the proposed parameter are set equal to the ones of the current values, whereas the fourth entry is drawn from a proposal distribution $q(u)$. Specifically,

$$\boldsymbol{\theta}^p = g(\boldsymbol{\theta}^c, u),$$

where

$$g(\boldsymbol{\theta}^c, u) := \begin{cases} g_1: & x_1^p = x_1^c, \\ g_2: & x_2^p = x_2^c, \\ g_3: & x_3^p = x_3^c, \\ g_4: & x_4^p = u, \qquad u \sim q(u) \end{cases}$$

Then, the Jacobian term is given by

$$\frac{\partial \boldsymbol{\theta}^p}{\partial (\boldsymbol{\theta}^c, u)} = \begin{vmatrix} \frac{\partial g_1}{\partial x_1^c} & \frac{\partial g_1}{\partial x_2^c} & \frac{\partial g_1}{\partial x_3^c} & \frac{\partial g_1}{\partial u} \\ \frac{\partial g_2}{\partial x_1^c} & \frac{\partial g_2}{\partial x_2^c} & \frac{\partial g_2}{\partial x_3^c} & \frac{\partial g_2}{\partial u} \\ \frac{\partial g_3}{\partial x_1^c} & \frac{\partial g_3}{\partial x_2^c} & \frac{\partial g_3}{\partial x_3^c} & \frac{\partial g_3}{\partial u} \\ \frac{\partial g_4}{\partial x_1^c} & \frac{\partial g_4}{\partial x_2^c} & \frac{\partial g_4}{\partial x_3^c} & \frac{\partial g_4}{\partial u} \end{vmatrix} = 1,$$

as the matrix of partial derivatives is the identity.

If we propose a death move to a model with $m^p = 2$ and $\boldsymbol{\theta}^p = (x_1^p, x_2^p)$, the proposed vector of parameter is constructed by deleting at random one of the

entries of the current parameter vector. Then, it is easily noticeable that for this type of move the Jacobian term is again equal to one.

2.A.3.2 Change-Point Model Move

Now we provide the details of the Jacobian evaluation for the birth step of the change-point model move. Notice that except for the residual variances $\sigma_j^{2c}, \sigma_j^{2p}$ and σ_{j+1}^{2p}, the matrices of partial derivatives corresponding to all the other parameters are identity matrices, and their determinants are equal to one. Hence, as the determinant of a block diagonal matrix is equal to the product of the determinants of the diagonal blocks, we focus only on the block whose determinant differs from one.

For simplicity of notation, let $x = \sigma_j^{2c}$, $y_1 = \sigma_j^{2p}$, and $y_2 = \sigma_{j+1}^{2p}$. We propose a move from x to $\boldsymbol{y} = (y_1, y_2)$ via a transformation $g(x, u)$, where we draw $u \sim \text{Uniform}(0, 1)$ and set

$$g(x,u) := \begin{cases} g_1: \quad y_1 = \dfrac{u}{1-u}\, x, \\ \\ g_2: \quad y_2 = \dfrac{1-u}{u}\, x. \end{cases} \tag{2.24}$$

Then, the Jacobian $\boldsymbol{J}$ is given by

$$\boldsymbol{J} := \frac{\partial \boldsymbol{y}}{\partial\, g(x,u)} = \begin{vmatrix} \dfrac{\partial\, g_1}{\partial\, x} & \dfrac{\partial\, g_1}{\partial\, u} \\ \dfrac{\partial\, g_2}{\partial\, x} & \dfrac{\partial\, g_2}{\partial\, u} \end{vmatrix} = \begin{vmatrix} \dfrac{u}{1-u} & \dfrac{x}{(1-u)^2} \\ \dfrac{1-u}{u} & -\dfrac{x}{u^2} \end{vmatrix} = \frac{2x}{u\,(1-u)}. \tag{2.25}$$

Moreover, if we use that

$$x = u^2 x + (1-u)^2 x + 2x(1-u)u,$$

we can rewrite Equation (2.25) as

$$\begin{aligned}\frac{2x}{u\,(1-u)} &= \frac{2\left(u^2x + (1-u)^2x + 2x(1-u)u\right)}{u\,(1-u)} \\ &= 2\left(\frac{u}{1-u}x + \frac{1-u}{u}x + 2x\right) \\ &= 2\left(y_1 + y_2 + 2\sqrt{y_1}\sqrt{y_2}\right) \qquad \text{by using (2.24)}, \\ &= \left(\sqrt{y_1} + \sqrt{y_2}\right)^2\end{aligned} \quad , \tag{2.26}$$

as required.

For the death step, the residual variance of the newly merged segment is constructed by inverting the transformation of Equation (2.24). If we let $y_1 = \sigma_j^{2c}$, $y_2 = \sigma_{j+1}^{2c}$, then the proposed residual variance $x = \sigma_j^{2p}$ is constructed as

$$x = \sqrt{y_1 y_2}.$$

In fact,

$$\begin{cases} y_1 &= \dfrac{u}{1-u}\,x \\ y_2 &= \dfrac{1-u}{u}\,x \end{cases} \rightarrow \begin{cases} x = y_1\dfrac{1-u}{u} \\ \sqrt{\dfrac{y_2}{y_1}} = \dfrac{1-u}{u} \end{cases} \rightarrow \begin{cases} x &= \sqrt{y_1\,y_2} \\ u &= \dfrac{1}{\left(1+\sqrt{\frac{y_2}{y_1}}\right)} \end{cases} \tag{2.27}$$

Notice that the Jacobian of the inverse transformation is the inverse of the Jacobian. Hence, the Jacobian term for the death step is obtaining by inverting the result obtained in Equation (2.26).

Chapter 3

A Spectral Hidden Markov Model for Nonstationary Oscillatory Processes

In the previous chapter we presented a Bayesian approach for analyzing nonstationary processes that exhibit time-varying oscillatory behaviour to detect changepoints and simultaneously identify the potentially changing periodicities in the data. However, in addition to detect temporal changes, it may also be of interest to recognize and model the recurrence of a relevant periodic behaviour in a probabilistic way. For example, we found that human respiratory traces measured in the context of experimental sleep apnea research exhibit abrupt changes in their periodic behaviour, where several patterns, such as instances of apnea, recur over time. Indeed, identifying (and ultimately predicting) all instances of apnea that occur during sleep is one of the primary interests for healthcare providers, clinicians and experimental biologists working in sleep apnea research.

Our methodology is built on the framework of a hidden Markov model

(HMM) [Rabiner et al., 1989] where we assume the number of states to be unknown, since we do not always have complete prior information about the number of different regimes of periodic behaviours that occur in a time series. The number of frequencies that characterizes each periodic regime is also assumed unknown, as the variability in each different state may be driven by a different number of relevant periodicities.

An HMM is a stochastic process model based on a discrete latent state sequence whose transition probabilities follow a Markovian structure [Rabiner, 1989; Ephraim and Merhav, 2002]. Conditioned on this state sequence, the observations are assumed to be conditionally independent, and generated from a family of probability distributions, which hereafter we refer to as the *emission* distributions. The HMM has been successfully utilized in many applications, such as speech recognition [Rabiner, 1989; Juang and Rabiner, 1991; Jelinek, 1997], digit recognition [Raviv, 1967; Rabiner et al., 1989], biological and physiological data [Krogh et al., 1994, 2001; Huang et al., 2018; Langrock et al., 2013; Yaghouby and Sunderam, 2015], and finance [Bhar and Hamori, 2004]. A disadvantage of the classical parametric HMM setting is that it generally requires the number of states to be fixed in advance, where model selection is usually carried out by employing Akaike's information criterion [Akaike, 1974] or the Bayesian information criterion [Schwarz et al., 1978]. Alternatively, in a nonparametric approach, the hierarchical Dirichlet process (HDP) presented by Teh et al. [2006] may be used as a building block for an HMM with an unspecified number of states. Specifically, the HDP-HMM approach places a Dirichlet process (DP) prior on the Markovian transition probabilities of the system, while allowing the atoms associated with the state-specific conditional DPs to be shared between each other, yielding an HMM with a countably infinite number of states. However, the approach proposed by Teh et al. [2006] does not

adequately represent the temporal persistence of the regimes, resulting in unrealistic rapid switching and redundancy of states. This issue was addressed by Fox et al. [2011] who presented a *sticky* HDP-HMM, which includes a self-transition parameter that better captures the temporal mode persistence of the hidden state sequence.

In this chapter, we build upon the work by Fox et al. [2011] to propose a novel Bayesian methodology for modelling periodic phenomena whose behaviour switches dynamically over time as realizations of an HMM. The number of states is assumed unknown along with their relevant periodicities, which may vary over the different states as each state is characterized by different spectral properties. Flexibility on the number of states is achieved by assuming a sticky HDP-HMM that penalises rapid switching dynamics of the underlying process as in Fox et al. [2011]. Inference for variable dimensionality regarding the number of periodicities that characterizes the emission distributions of each regime is attained by developing an appropriate form of reversible-jump MCMC sampler [Green, 1995]. The rest of the chapter is organized as follows. Section 3.1 presents the model and the general framework of our Bayesian approach. Section 3.2 and 3.3 provide the inference scheme and simulation studies to illustrate the performance of the proposed method. In Section 3.4, we illustrate the use of our approach to detect instances of apnea in human breathing traces (data provided from Dr. Robert Huckstepp's lab, Warwick Life Sciences).

3.1 The Model

As before, let $\boldsymbol{y} = (y_1, \ldots, y_T)'$ be a realization of a time series whose oscillatory behaviour may switch dynamically over time. In addition, we introduce $\boldsymbol{z} =$

$(z_1, \ldots, z_T)'$ to denote the hidden discrete-valued states of a Markov chain that characterizes the different periodic regimes, where z_t represents the state of the Markov chain at time t. Any observation y_t given the state z_t, is assumed to be conditionally independent of the observations and states at other time steps [Rabiner, 1989]. A flexible nonparametric approach is obtained by assuming that the state space is unbounded, i.e. has infinitely many states as in [Beal et al., 2002; Teh et al., 2006]. Thus, the Markovian structure on the state sequence $\boldsymbol{z}$ is given by

$$z_t \,|\, z_{t-1}, (\boldsymbol{\pi}_j)_{j=1}^{\infty} \sim \boldsymbol{\pi}_{z_{t-1}}, \quad t = 1, \ldots, T, \tag{3.1}$$

where $\boldsymbol{\pi}_j = (\pi_{j1}, \pi_{j2}, \ldots)$ represents the (countably infinite) state-specific vector of transition probabilities, and in particular $\pi_{jk} = p(z_t = k \,|\, z_{t-1} = j)$, where $p(\cdot)$ is used as a generic notation for probability density or mass function, whichever appropriate. We assume the state at the first time step follows an initial transition distribution $\boldsymbol{\pi}_0 = (\pi_{01}, \pi_{02}, \ldots)$, namely $z_0 \sim \boldsymbol{\pi}_0$.

3.1.1 Oscillatory Emissions

Assume that each state j represents a periodic regime that is characterized by d_j relevant periodicities whose frequencies are denoted by $\boldsymbol{\omega}_j = (\omega_{j1}, \ldots, \omega_{jd_j})'$, recalling that periodicity is the inverse of frequency. Let $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{j1}', \ldots, \boldsymbol{\beta}_{jd_j}')'$ be the vector of linear coefficients that can be associated with the amplitude and phase corresponding to each frequency ω_{jl} that characterizes the oscillatory behaviour of state j, where $\boldsymbol{\beta}_{jl} = (\beta_{jl}^{(1)}, \beta_{jl}^{(2)})'$ and $l = 1, \ldots, d_j$. Furthermore, let us define $\boldsymbol{\theta}_j = (d_j, \boldsymbol{\omega}_j', \boldsymbol{\beta}_j', \sigma_j^2)'$, where σ_j^2 accounts for a state-specific variance. Then, each observation is assumed to be generated from the following emission distribution

$$y_t \,|\, z_t = j, (\boldsymbol{\theta}_j)_{j=1}^{\infty} \sim \mathcal{N}\left(f_{tj}, \sigma_j^2\right), \quad t = 1, \ldots, T, \tag{3.2}$$

where the mean function f_{tj} for state j at time t is [Andrieu and Doucet, 1999; Hadj-Amar et al., 2019] specified to be oscillatory

$$f_{tj} = \boldsymbol{x}_t(\boldsymbol{\omega}_j)' \boldsymbol{\beta}_j, \tag{3.3}$$

and the vector of basis functions $\boldsymbol{x}_t(\boldsymbol{\omega}_j)$ is defined as

$$\boldsymbol{x}_t(\boldsymbol{\omega}_j) = (\cos(2\pi\omega_{j1}t), \sin(2\pi\omega_{j1}t), \ldots, \cos(2\pi\omega_{jd_j}t), \sin(2\pi\omega_{jd_j}t))'.$$

The dimension of each oscillatory function depends on the unknown number d_j of periodicities relevant to that specific regime. Given a pre-fixed maximal number of periodicities per regime, $d_{\max}$, the parameter space $\boldsymbol{\Theta}_j$ for the vector of emission parameters $\boldsymbol{\theta_j}$ can be written as

$$\boldsymbol{\Theta}_j = \bigcup_{d_j=1}^{d_{\max}} \left\{d_j\right\} \times \left\{\boldsymbol{\Omega}_{d_j} \times \mathbb{R}^{2d_j} \times \mathbb{R}^{+}\right\},$$

where $\boldsymbol{\Omega}_{d_j} = (0, 0.5)^{d_j}$ denotes the sample space for the frequencies of the jth regime.

We notice that in the previous chapter we used the sinusoidal modelling approach specified in Equation (3.3) for oscillatory data that show regime shifts in periodicity, amplitude and phase, where we assumed that, conditional on an (unknown) number of change-points and their (unknown) positions, the time series process can be approximated by a sequence of segments, each with time-varying sinusoidal models of the general form given in Equation (3.3). However, the ap-

proach proposed in this chapter is combined with an HMM framework, which can furthermore address the occurrence and recurrence of periodic regimes over time by assigning probabilities with which these transitions occur between regimes.

3.1.2 Bayesian Nonparametric Framework for Unbounded Markov States

In contrast to classic methods that assume a parametric prior on the number of states or use model selection techniques to determine the number of regimes in an HMM, here we follow Beal et al. [2002]; Teh et al. [2006] and Fox et al. [2011], and assume the number of states to be unknown. We therefore do not need to pre-specify the number of hidden states which provides a more flexible modelling framework. The DP [Ferguson, 1973] may be used in frameworks where an element of the model is a discrete random variable of unknown cardinality [Hjort et al., 2010]. The underlying structure of an unbounded HMM (i.e. where the number of possible states is unknown) provides such a setting and can be seen as an infinite mixture model, where the mixing proportions are modelled as DPs [Beal et al., 2002; Rasmussen and Ghahramani, 2002; Teh et al., 2006].

In particular, the current state z_t indexes a specific transition distribution $\boldsymbol{\pi}_{z_t}$ over the positive integers, whose probabilities are the mixing proportions for the choice of the next state z_{t+1}. To allow the same set of next states to be reachable from each of the current states, we introduce a set of state-specific DPs, whose atoms are shared between each other [Teh et al., 2006]. Furthermore, we follow Fox et al. [2011] and penalize rapid switching dynamics between the states or redundant modes by increasing the expected probability of self-transitions. In particular, the

state-specific transition distribution $\boldsymbol{\pi}_j$ is modelled with the following HDP

$$\boldsymbol{\pi}_j \,\big|\, \eta,\, \kappa,\, \boldsymbol{\alpha} \sim \mathrm{DP}\left(\eta + \kappa,\, \frac{\eta\, \boldsymbol{\alpha} + \kappa\, \delta_j}{\eta + \kappa} \right), \tag{3.4}$$

where

$$\boldsymbol{\alpha} \,\big|\, \gamma \sim \mathrm{GEM}\,(\,\gamma\,). \tag{3.5}$$

The sequence $\boldsymbol{\alpha} = (\,\alpha_k\,)_{k=1}^{\infty}$ can be seen as a *global* probability distribution over the positive integers that ties together the transition distributions $\boldsymbol{\pi}_j$ and guarantees that they have the same support. We denote by GEM (γ) [1] the *stick-breaking construction* [Sethuraman, 1994; Pitman, 2002] of $\boldsymbol{\alpha}$ as

$$\alpha_k = \nu_k \prod_{l=1}^{k-1} (1 - \nu_l), \tag{3.6}$$

where

$$\nu_k \,|\, \gamma \sim \mathrm{Beta}\,(\,1,\, \gamma\,), \tag{3.7}$$

for $k = 1, 2, \ldots,$ and γ is a positive real number that controls the expected value of the number of elements in $\boldsymbol{\alpha}$ with significant probability mass. In fact, Equations (3.6) and (3.7) can be motivated by the equivalent process where a stick of length one is split into lengths specified by the weights α_k, where the k^{th} proportion is a random fraction ν_k of the remaining stick after the preceding $(\,k - 1\,)$ proportions have been constructed. Notice that this construction ensures that the sequence $\boldsymbol{\alpha}$ satisfies $\sum_{k=1}^{\infty} \alpha_k = 1$ with probability one.

Conditional on $\boldsymbol{\alpha}$, the hierarchical structure given in Equation (3.4) indicates that each state-specific transition distribution $\boldsymbol{\pi}_j$ is distributed according to a DP

[1] GEM is an abbreviation for Griffiths, Engen and McCloskey, see Ignatov [1982]; Perman et al. [1992]; Pitman [1996] for background.

with *concentration* parameter $\eta + \kappa$ and *base* distribution that is itself a DP, namely $(\eta\,\boldsymbol{\alpha} + \kappa\,\delta_j)/(\eta + \kappa)$. Here, η is a positive real number that controls the variability of the $\boldsymbol{\pi}_j$'s around $\boldsymbol{\alpha}$, while κ is a positive real number that inflates the expected probability of a self-transition [Fox et al., 2007], and δ_j denotes a unit-mass measure concentrated at j. By setting $\kappa = 0$ in Equation (3.4), we obtain the original HDP-HMM framework proposed by Teh et al. [2006]. Their approach might inadequately model the temporal persistence of the hidden states, yielding an unrealistically rapid alternation between different (and often redundant) regimes. Instead, we use the *sticky* version formulated in Fox et al. [2007] which allows to increase the temporal state persistence of the system since the expected probabilities of self-transitions are inflated by an amount proportional to κ, i.e.

$$\mathbb{E}\left[\,\pi_{jk}\,|\,\eta,\,\kappa,\,\boldsymbol{\alpha}\right] = \frac{\eta}{\eta+\kappa}\,\alpha_k + \frac{\kappa}{\eta+\kappa}\,\delta\,(j,\,k),$$

where $\delta\,(j,\,k) = 1$ if $k = j$ and zero otherwise.

3.2 Inference

Our inference scheme is formulated in a full Bayesian framework. Section 3.2.1 presents a reversible jump MCMC based algorithm to obtain posterior samples of the emission parameters $\boldsymbol{\theta}_j$, where a trans-dimensional MCMC sampler is developed to explore subspaces of variable dimensionality regarding the number of periodicities that characterize state j. In Section 3.2.2 we introduce the *Chinese restaurant franchise with loyal custumers* [Fox et al., 2011], a metaphor that provides the building blocks to perform Bayesian nonparametric inference for updating the HMM parameters. The resulting Gibbs sampler is provided in Section 3.2.2.1.

3.2.1 Emission Parameters

Conditional on the state sequence $\boldsymbol{z}$, the observations $\boldsymbol{y}$ are implicitly partitioned in a finite number of states, where each state refers to at least one segment of the time series. When a type of periodic behaviour recurs over time, the corresponding state is necessarily related to more than one segment of the data. Let $\boldsymbol{y}_j^* = (\boldsymbol{y}_{j1}^{'}, \boldsymbol{y}_{j2}^{'}, \ldots, \boldsymbol{y}_{jR_j}^{'})^{'}$ be the vector of (non-adjacent) segments that are assigned to state j, where $\boldsymbol{y}_{jr}$ denotes the r^{th} segment of the time series for which $z_t = j$ and R_j is the total number of segments assigned to that state. Then, the likelihood of the emission parameter $\boldsymbol{\theta}_j$ given the observations in $\boldsymbol{y}_j^*$ is

$$\mathscr{L}(\boldsymbol{\theta}_j \,|\, \boldsymbol{y}_j^*) = (2\pi\sigma_j^2)^{-T_j^*/2} \exp\left[-\frac{1}{2\sigma_j^2} \sum_{t \in I_j^*} \left\{ y_t - \boldsymbol{x}_t (\boldsymbol{\omega}_j)^{'} \boldsymbol{\beta}_j \right\}^2 \right], \tag{3.8}$$

where I_j^* and T_j^* denote the set of time points and number of observations, respectively, associated with $\boldsymbol{y}_j^*$.

Following Hadj-Amar et al. [2019], the prior specifications of our Bayesian framework assume independent Poisson distributions for the number of frequencies d_j for each state j, constrained on $1 \leq d_j \leq d_{max}$. Conditional on d_j, we choose a uniform prior for the frequencies $\omega_{j,l} \sim \text{Uniform}(0, \phi_\omega)$, $l = 1, \ldots, d_j$, where $0 < \phi_\omega < 0.5$. The value of ϕ_ω can be chosen to reflect prior information about the significant frequencies that are responsible for the overall variation in the data. Analogous to a Bayesian regression [Bishop, 2006], a zero-mean isotropic Gaussian prior is assumed for the coefficients of the j^{th} regime, $\boldsymbol{\beta}_j \sim \mathcal{N}_{2d_j}(\boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I})$, where the prior variance σ_β^2 is fixed at a large value (e.g 10^2). The prior on the residual variance σ_j^2 of state j is specified as Inverse-Gamma $(\frac{\xi_0}{2}, \frac{\tau_0}{2})$, where ξ_0 and τ_0 are fixed at small values, noticing that when $\xi_0 = \tau_0 = 0$ we obtain Jeffreys'

uninformative prior [Bernardo and Smith, 2009].

Bayesian inference on $\boldsymbol{\theta}_j$ is based on the following factorization of the joint posterior distribution

$$p(\boldsymbol{\theta}_j \mid \mathbf{y}_j^*) = p(d_j \mid \mathbf{y}_j^*)\, p(\boldsymbol{\omega}_j \mid d_j, \mathbf{y}_j^*)\, p(\boldsymbol{\beta}_j \mid \boldsymbol{\omega}_j, d_j, \mathbf{y}_j^*)\, p(\sigma_j^2 \mid \boldsymbol{\beta}_j, \boldsymbol{\omega}_j, d_j, \mathbf{y}_j^*). \tag{3.9}$$

Sampling from the posterior distribution in (3.9) poses a model selection problem regarding the number of periodicities thus requiring an inference algorithm that is able to explore subspaces of variable dimensionality. This will be addressed by the reversible-jump sampler introduced in the following section.

3.2.1.1 Reversible-Jump Sampler

Here we provide the details for drawing $\boldsymbol{\theta}_j$ from the posterior distribution $p(\boldsymbol{\theta}_j \mid \mathbf{y}_j^*)$ given in Equation (3.9). Our methodology follows Andrieu and Doucet [1999] and Hadj-Amar et al. [2019] and relies on the principles of reversible-jump MCMC introduced in Green [1995]. Notice that, conditional on the state sequence $\boldsymbol{z}$, the emission parameters $\boldsymbol{\theta}_j$ can be updated independently and in parallel for each of the current states. Hence, for the rest of this subsection and for ease of notation, we drop the subscript corresponding to the j^{th} state.

At each iteration of the algorithm, a random choice with probabilities given in (3.10) based on the current number of frequencies d will dictate whether to add a frequency (*birth step*) with probability b_d, remove a frequency (*death step*) with probability r_d, or update the frequencies (*within step*) with probability $\mu_d = 1 - b_d - r_d$, where

$$b_d = c \min\left\{1, \frac{p(d+1)}{p(d)}\right\}, \qquad r_{d+1} = c \min\left\{1, \frac{p(d)}{p(d+1)}\right\}, \tag{3.10}$$

for some constant $c \in [0, \frac{1}{2}]$ and $p(d)$ is the prior probability. We fixed $c = 0.4$ but other values are admissible as long as c is not larger than 0.5 to guarantee that the sum of the probabilities does not exceed 1 for some values of c. Naturally, $b_{d_{\max}} = r_1 = 0$. An outline of these moves is as follows (further details are provided in Appendix 3.B).

Within-Model Move: Conditional on the number of frequencies d, we sample the vector of frequencies $\boldsymbol{\omega}$ following a similar procedure to the one of Andrieu and Doucet [1999] and Hadj-Amar et al. [2019]. That is, we update the frequencies one-at-time using a mixture of Metropolis-Hastings (M-H) steps, with target distribution

$$p(\boldsymbol{\omega} \,|\, \boldsymbol{\beta}, \sigma^2, d, \boldsymbol{y}^*) \propto \exp\left[-\frac{1}{2\sigma^2} \sum_{t \in I^*} \left\{ y_t - \boldsymbol{x}_t(\boldsymbol{\omega})' \boldsymbol{\beta} \right\}^2 \right] \mathbb{1}_{[\boldsymbol{\omega} \in \boldsymbol{\Omega}_d]}. \tag{3.11}$$

Specifically, the proposal distribution is a combination of a Normal random walk centred around the current frequency and a draw from the periodogram of $\hat{\boldsymbol{y}}$, where $\hat{\boldsymbol{y}}$ denotes a segment of data randomly chosen from $\boldsymbol{y}^*$ with probability proportional to the number of observations belonging to that segment. Naturally, when a state does not recur over time, i.e. when a state refers to only one segment of the time series, that segment is chosen with probability one. Next, the corresponding vector of linear coefficients $\boldsymbol{\beta}$ is updated in a Gibbs step from

$$\boldsymbol{\beta} \,\big|\, \boldsymbol{\omega}, \sigma^2, d, \boldsymbol{y}^* \sim \boldsymbol{N}_{2d}(\hat{\boldsymbol{\beta}}, \boldsymbol{V}_\beta), \tag{3.12}$$

where

$$\begin{aligned} \boldsymbol{V}_\beta &= \left(\sigma_\beta^{-2}\,\boldsymbol{I} + \sigma^{-2}\boldsymbol{X}^*(\boldsymbol{\omega})'\boldsymbol{X}^*(\boldsymbol{\omega})\right)^{-1}, \\ \hat{\boldsymbol{\beta}} &= \boldsymbol{V}_\beta\,(\sigma^{-2}\boldsymbol{X}^*(\boldsymbol{\omega})'\boldsymbol{y}^*), \end{aligned} \tag{3.13}$$

and we denote with $\boldsymbol{X}^*(\boldsymbol{\omega})$ the design matrix with rows given by $\boldsymbol{x}_t(\boldsymbol{\omega})$ (Equation 3.1.1), for $t \in I^*$. Finally, the residual variance σ^2 is updated in a Gibbs step directly from

$$\sigma^2 \,\big|\, \boldsymbol{\beta},\, \boldsymbol{\omega},\, \boldsymbol{y}^* \sim \text{Inverse-Gamma}\left(\frac{T+\xi_0}{2},\, \frac{\tau_0 + \sum_{t\in I^*}\left\{y_t - \boldsymbol{x}_t(\boldsymbol{\omega})'\boldsymbol{\beta}\right\}^2}{2}\right). \tag{3.14}$$

Trans-Dimensional Moves: For these type of moves, the number of periodicities is either proposed to increase by one (birth) or decrease by one (death) [Green, 1995]. If a birth move is proposed, we have that $d^p = d^c + 1$, where current and proposed values are denoted by the superscripts c and p, respectively. The proposed vector of frequencies is constructed by proposing an additional frequency to be included in the current vector. If a death move is chosen, we have that $d^p = d^c - 1$ and one of the current periodicities is randomly selected to be removed. Conditional on the frequencies, the corresponding vector of linear coefficients and the residual variance are sampled as in the within-model move. For both moves, the updates are jointly accepted or rejected in a M-H step.

3.2.2 HMM Parameters

Here we explain how to perform posterior inference about the probability distribution $\boldsymbol{\alpha}$, the transition probabilities $\boldsymbol{\pi}_j$ and the state sequence $\boldsymbol{z}$. As described in Section 3.1.2, flexibility on the number of states is achieved by assuming an HDP that

penalises rapid switching dynamics of the hidden states. The *Chinese restaurant franchise with loyal customers* presented by Fox et al. [2011], which extends the *Chinese restaurant franchise* introduced by Teh et al. [2006], is a metaphor that can be used to express the generative process behind such an HDP and provides a general framework for performing inference. We provide a more detailed introduction to DP and HDP in Appendix A at the end of the thesis, and a detailed explanation of the Chinese restaurant franchise with loyal customers in Appendix 3.A of this chapter. A high level summary of the metaphor is as follows: in a *Chinese restaurant franchise* the analogy of a *Chinese restaurant process* [Aldous, 1985] is extended to a set of restaurants, where an infinite global menu of dishes is shared across these restaurants. The process of seating customers at tables happens in a similar way as for the Chinese restaurant process, but is restaurant-specific. The process of choosing dishes at a specific table happens franchise-wide, namely the dishes are selected with probability proportional to the number of tables (in the entire franchise) that have previously served that dish. However, in the Chinese restaurant franchise with loyal customers, each restaurant in the franchise has a speciality dish which may keep many generations of customers eating in the same restaurant.

The global probability distribution $\boldsymbol{\alpha}$ can be seen as a collection of ratings for dishes served in the global menu, the values of the dishes correspond to the hidden states $\boldsymbol{z}$, whereas transition probabilities $\boldsymbol{\pi}_j$ represent restaurant-specific probability distributions over unique dishes. As explained in Appendix 3.A.2, table counts $\bar{m}_{jk}$ of considered dishes are sufficient statistics for updating the collection of dish ratings $\boldsymbol{\alpha}$, where $\bar{m}_{jk}$ denotes how many of the tables in restaurant j considered dish k. The sampling of $\bar{m}_{jk}$ is additionally simplified when introducing table counts m_{jk} of served dishes and override variables o_{jt}. The meaning and definition of *served* and *considered* dishes as well as the role of the newly introduced auxiliary variables

$m_{jk}, \bar{m}_{jk}$ and o_{jt} are given in Appendix 3.A.1.

3.2.2.1 Gibbs Sampler

Kivinen et al. [2007] and Fox et al. [2011] consider a Gibbs sampler which uses finite approximations to the DP to allow sampling in blocks of the state sequence $\boldsymbol{z}$. In particular, conditioned on observations $\boldsymbol{y}$, transition probabilities $\boldsymbol{\pi}_j$ and emission parameters $\boldsymbol{\theta}_j$, the hidden states $\boldsymbol{z}$ are sampled using a variant of the HMM forward-backward procedure (see Appendix 3.C.1) presented in Rabiner [1989]. In order to use this scheme, we must truncate the countably infinite transition distributions $\boldsymbol{\pi}_j$ (and global menu $\boldsymbol{\alpha}$), and this is achieved using the following approximation to a DP [Ishwaran and Zarepour, 2002]

$$\text{GEM}_{K_{\max}}(\gamma) := \text{Dir}(\gamma/K_{\max}, \ldots, \gamma/K_{\max}), \tag{3.15}$$

where the truncation level $K_{\max}$ is a number that exceeds the total number of expected HMM states, and $\text{Dir}(\cdot)$ denote the Dirichlet distribution. Conditioned on the state sequence $\boldsymbol{z}$ and collection of dish ratings $\boldsymbol{\alpha}$, we sample m_{jk}, o_{jt} and $\bar{m}_{jk}$ as described in Appendix 3.C.2. Dish ratings $\boldsymbol{\alpha}$ and transition distributions $\boldsymbol{\pi}_j$ are then updated from the following posterior distributions

$$\begin{aligned} \boldsymbol{\alpha} \,|\, \bar{\boldsymbol{m}}, \gamma &\sim \text{Dir}(\gamma/K_{\max} + \bar{m}_{\cdot 1}, \ldots, \gamma/K_{\max} + \bar{m}_{\cdot K_{\max}}) \\ \boldsymbol{\pi}_j \,|\, \boldsymbol{z},\, \boldsymbol{\alpha},\, \eta,\, \kappa &\sim \text{Dir}(\eta\,\alpha_1 + n_{j1}, \ldots, \eta\,\alpha_j + \kappa + n_{jj}, \ldots, \eta\,\alpha_{K_{\max}} + n_{jK_{\max}}), \end{aligned} \tag{3.16}$$

for each state $j = 1, \ldots, K_{max}$. Here, $\bar{\boldsymbol{m}}$ is the vector of table counts of considered dishes for the whole franchise, and marginal counts are described with dots, so that $\bar{m}_{\cdot k} = \sum_{j=1}^{K_{\max}} \bar{m}_{jk}$ is the number of tables in the whole franchise considering dish k. We denote with n_{jk} the number of Markov chain transitions from state j to state k in

the hidden sequence z. Then, given the state sequence z and transition probabilities $\boldsymbol{\pi}_j$, we draw the emission parameters $\boldsymbol{\theta}_j$ for each of the currently instantiated state as described in Section 3.2.1, where each reversible-jump MCMC update is run for several iterations. We also need to update emission parameters for states which are not instantiated (namely, those states among $\{1, \ldots, K_{\max}\}$ that are not represented during a particular iteration of the sampler), and hence we draw the corresponding emission parameters from their priors. Finally, we sample the hyperparameters γ, η and κ in a Gibbs step (see Appendix 3.C.3).

For the HDP-HMM, different procedures have been applied for sampling the hidden state sequence z. Teh et al. [2006] originally introduced an approach based on a Gibbs sampler which has been shown to suffer from slow mixing behaviour due to strong correlations that is frequently observed in the data at nearby time points. Van Gael et al. [2008] presented a *beam sampling* algorithm that combines a slice sampler [Neal et al., 2003] with dynamic programming. This allows to constrain the number of reachable states at each MCMC iteration to a finite number, where the entire hidden sequence z is drawn in block using a form of forward-backward filtering scheme. However, Fox et al. [2011] showed that applications of the beam sampler to the HDP-HMM resulted in slower mixing rates compared to the forward-backward procedure that we use in our truncated model. Recently, Tripuraneni et al. [2015] developed a particle Gibbs MCMC algorithm [Andrieu et al., 2010] which uses an efficient proposal and makes use of ancestor sampling to enhance the mixing rate.

3.2.2.2 Label Switching

The proposed approach may suffer from *label switching* (see e.g. Redner and Walker [1984]; Stephens [2000]; Jasra et al. [2005]) since the likelihood is invari-

ant under permutations of labelling of the mixture components, for both hidden state labels $\{1, \ldots, K_{max}\}$ and frequency labels $\{1, \ldots, d_{max}\}$ in each state. The label switching problem occurs when using Bayesian mixture models and needs to be addressed in order to draw meaningful inference about the posterior model parameters. In our multiple model search, the frequencies (and their corresponding linear coefficients) are identified by keeping them in ascending order for every iteration of the sampler. Posterior samples of the model parameters corresponding to different hidden states are post-processed (after the full estimation run) using the relabelling algorithm developed by Stephens [2000]. The basic idea behind this algorithm is to find permutations of the MCMC samples in such a way that the Kullback-Leibler (KL) divergence [Kullback and Leibler, 1951] between the 'true' distribution on clusterings, say $P(\boldsymbol{\theta})$, and a matrix of classification probabilities, say $\boldsymbol{Q}$, is minimized. The KL distance is given by $d(\boldsymbol{Q}, P(\boldsymbol{\theta}))_{KL} = \sum_t \sum_j p_{tj}(\boldsymbol{\theta}) \log \frac{p_{tj}(\boldsymbol{\theta})}{q_{tj}}$, where $p_{tj}(\boldsymbol{\theta}) = p(z_t = j \,|\, z_{t-1}, \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta})$ is part of the MCMC output obtained as in Appendix 3.C.1, and q_{tj} is the probability that observation t is assigned to class j. The algorithm iterates between estimating $\boldsymbol{Q}$ and the most likely permutation of the hidden labels for each MCMC iteration. We chose the strategy of Stephens [2000] since it has been shown to perform very efficiently in terms of finding the correct relabelling (see e.g. Rodriguez and Walker [2014]). However, it may be computationally quite intensive in memory since it requires the storage of a matrix of probabilities of dimension $N \times T \times K_{max}$, where N is the number of MCMC samples. Furthermore, at each iterative step, the algorithm requires to go over $K_{max}!$ permutations of the labels for each MCMC iteration, which might significantly slow down the computation when using large values of K_{max}. Related approaches to the label switching issue include pivotal reordering algorithms [Marin et al., 2005], label invariant loss functions [Celeux et al., 2000; Hurn et al., 2003] and equivalent classes represen-

tative methods [Papastamoulis and Iliopoulos, 2010], where an overview of these strategies can be found in Rodriguez and Walker [2014].

3.3 Simulation Studies

This section presents results of simulation studies to explore the performance of our proposed methodology in two different settings. In the first scenario the data are generated from the model described in Section 3.1 and thus this simulation study provides a "sanity" check that the algorithm is indeed retrieving the correct pre-fixed parameters. We also investigate signal extraction for the case that the innovations come from a heavy-tailed t-distribution instead of a Gaussian. Our second study deals with artificial data from an HMM whose emission distributions are characterized by oscillatory dynamics generated by state-specific autoregressive (AR) time series models.

3.3.1 Illustrative Example

We generated a time series consisting of $T = 1450$ data points from a three-state HMM with the following transition probability matrix showing high probabilities of self transition along the diagonal

$$\boldsymbol{\pi} = \begin{bmatrix} 0.9900 & 0.0097 & 0.0003 \\ 0.0001 & 0.9900 & 0.0099 \\ 0.0097 & 0.0003 & 0.9900 \end{bmatrix},$$

and Gaussian oscillatory emissions as specified in Equation (3.2), where the parameters of each of the three regimes are given in Table 3.1. A realization from this

model is displayed in Figure 3.1. The prior mean on the number of frequencies d_j is set equal to 1 and we place a Gamma $(1, 0.01)$ prior on the concentration parameters γ and $(\eta + \kappa)$, and a Beta $(100, 1)$ prior on the self-transition proportion ρ as in Fox et al. [2011]. The maximum number of periodicities per regime $d_{\max}$ is set to 5, while the truncation level $K_{\max}$ for the DP approximation is set equal to 7. Also, we set $\phi_\omega = 0.25$ as a threshold for the uniform prior. The proposed estimation algorithm is run for 15,000 iterations, 3,000 of which are discarded as burn-in. At each iteration, for each instantiated set of emission parameters, 2 reversible-jump MCMC updates were performed. The full estimation algorithm took 31 minutes with a program written in Julia 0.62 on an Intel® Core™ i7 2.2 GHz Processor 8 GB RAM. For our experiments, we used the R package *label.switching* of Papastamoulis [2016] to post-process the MCMC output with the relabelling algorithm of Stephens [2000].

Table 3.1: Illustrative Example. Frequencies $\boldsymbol{\omega}_j$ and linear coefficients $\boldsymbol{\beta}_j$ for the three different regimes. The number of periodicities d_j in each regime is 1, 1 and 2, respectively. The innovations σ_j^2 are set to $(0.4)^3$, $(0.08)^2$ and $(0.3)^2$, respectively.

Frequencies		Linear Coefficients	
ω_{11}	$1/25$	$\boldsymbol{\beta}_{11} =$	$(0.8, 0.8)'$
ω_{21}	$1/19$	$\boldsymbol{\beta}_{21} =$	$(0.2, 0.2)'$
ω_{31}	$1/12$	$\boldsymbol{\beta}_{31} =$	$(1.0, 1.0)'$
ω_{32}	$1/8$	$\boldsymbol{\beta}_{32} =$	$(1.0, 1.0)'$

Table 3.2 (left panel) shows that our estimation algorithm successfully detects the correct number of states in the sense that a model with $k = 3$ regimes has the highest posterior probability. In addition, our approach correctly identifies the right number of frequencies in each regime, as shown in Table 3.2 (right panel). Table 3.3 displays the estimated posterior mean and standard deviation of the frequencies along with the square root of the power of the corresponding fre-

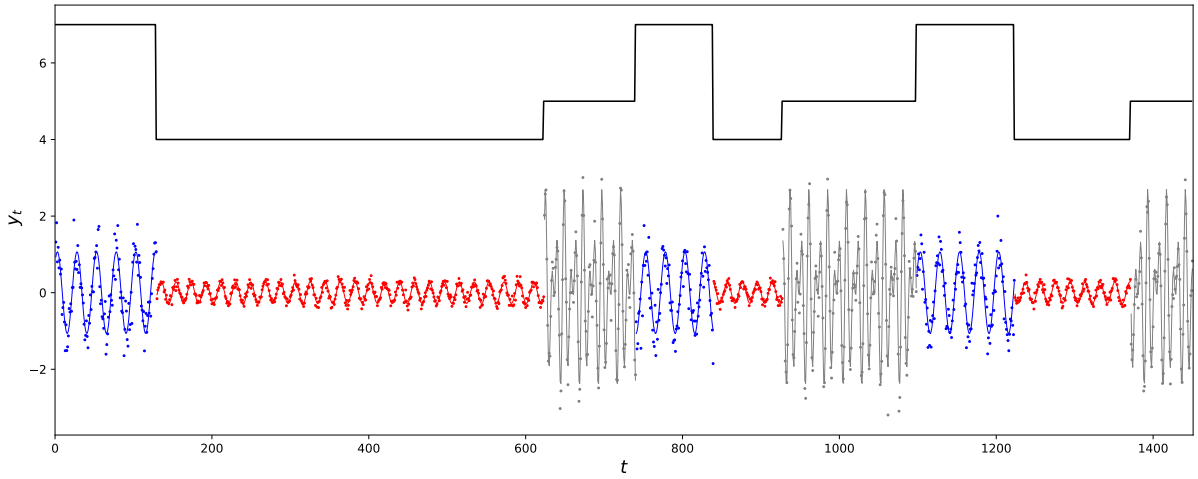

Figure 3.1: Illustrative Example. Dots represent the simulated time series, where the different colors corresponds to (true) different regimes. The state-specific estimated oscillatory mean function is displayed as a solid curve, and the estimated state sequence as a piecewise horizontal line at the top part of the graph.

Table 3.2: Illustrative example. (left panel) posterior probabilities for number of distinct states k; (right panel) posterior probabilities for number of frequencies in each state, conditioned on $k = 3$.

k	$\hat{\pi}(k \mid \boldsymbol{y})$
1	0.00
2	0.00
3	0.99
4	0.01
5	0.00
6	0.00
7	0.00

m	$\hat{\pi}(d_1 \mid k = 3, \boldsymbol{y})$	$\hat{\pi}(d_2 \mid k = 3, \boldsymbol{y})$	$\hat{\pi}(d_3 \mid k = 3, \boldsymbol{y})$
1	0.99	1.00	0.01
2	0.01	0.00	0.99
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.00	0.00	0.00

Table 3.3: Illustrative Example. Estimated posterior mean (and standard deviation) of frequencies and square root of the power of the corresponding frequencies.

	ω_{11}	ω_{21}	ω_{31}	ω_{32}
True	0.0400	0.0526	0.0833	0.1250
Estimated	0.0399 ($8.8 \cdot 10^{-6}$)	0.0526 ($6.3 \cdot 10^{-6}$)	0.0833 ($9.6 \cdot 10^{-6}$)	0.1249 ($9.4 \cdot 10^{-6}$)
	A_{11}	A_{21}	A_{31}	A_{32}
True	1.131	0.283	1.414	1.414
Estimated	1.069 (0.029)	0.281 (0.004)	1.380 (0.022)	1.367 (0.022)

quencies, where the results are conditional on three estimated states and the modal number of frequencies within each state. Here, the power of each frequency ω_{jl} is summarized by the amplitude $A_{jl} = \sqrt{\beta_{jl}^{(1)^2} + \beta_{jl}^{(2)^2}}$, namely the square root of the sum of squares of the corresponding linear coefficients (see e.g. Shumway and Stoffer [2017]). Our proposed method seems to provide a good match between true and estimated values for both frequencies and their power, for this example. We also show in Figure 3.1 the state-specific estimated signal (Equation (3.3)), and the estimated state sequence using the method of Stephens [2000] (as a piecewise horizontal line). The rows of the estimated transition probability matrix were $\hat{\boldsymbol{\pi}}_1 = (0.9921, 0.0073, 0.0006)$, $\hat{\boldsymbol{\pi}}_2 = (0.0005, 0.9956, 0.0040)$ and $\hat{\boldsymbol{\pi}}_3 = (0.0051, 0.0006, 0.9942)$. The high probabilities along the diagonal reflect the estimated posterior mean of the self transition parameter $\hat{\rho} = 0.9860$, which is indeed centered around the true probability of self-transition.

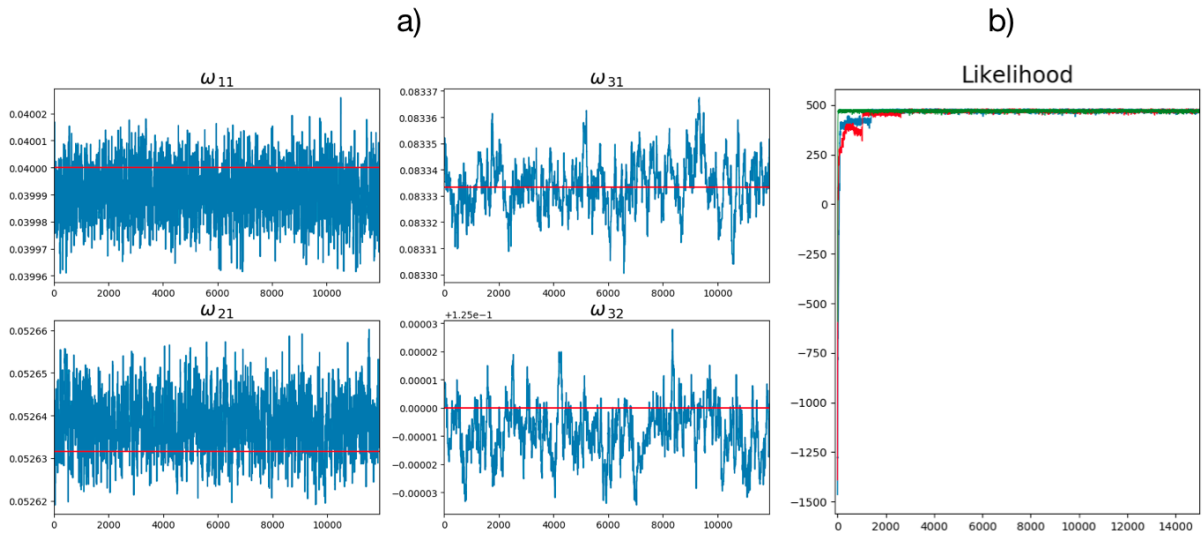

Figure 3.2: Illustrative Example. (a) Trace plots (after burn-in) for posterior sample of frequencies, conditional on modal number of states and number of frequencies in each state; red lines correspond to true values of the frequencies. (b) Trace plots (including burn-in) of the likelihood for three Markov chains initialized at different starting values.

Diagnostics for verifying convergence were performed in several ways. For example, we observed that the MCMC samples of the likelihood of the HMM reached a stable regime, while initializing the Markov chains from overdispersed starting values (see Figure 3.2 (b)). This diagnostic might be very useful, for example, in determining the burn-in period. However, we note that it does not guarantee convergence since steady values of the log likelihood might be the result of a Markov chain being stuck in some local mode of the target posterior distribution. The likelihood of an HMM with Gaussian emissions can be expressed as

$$\mathcal{L}(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta} \,|\, \boldsymbol{y}) = p(z_1 \,|\, \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta}) \mathcal{N}(y_1 \,;\, f_{1\,z_1},\, \sigma^2_{z_1}) \prod_{t=2}^{T} p\,(z_t \,|\, z_{t-1},\, \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta}\,) \mathcal{N}(y_t \,;\, f_{t\,z_t},\, \sigma^2_{z_t}),$$

where $\mathcal{N}\,(\,y_t;\, f_{jt},\, \sigma^2_j)$ denotes the density of a Gaussian distribution with mean $f_{jt} = \boldsymbol{x}_t\,(\boldsymbol{\omega}_j)^{'}\,\boldsymbol{\beta}_j$ (as in Equation 3.3) and variance σ^2_j, evaluated at y_t. Conditioned on the modal number of states, we also validated convergence for the state-specific

emission parameters by analyzing trace plots and running averages of the corresponding MCMC samples, with acceptable results as each trace reached a stable regime. As an example, we show in Figure 3.2 (a) trace plots (after burn-in) for the posterior values of the frequencies.

Signal Extraction with Non-Gaussian Innovations: In many scientific experiments it may be of interest extracting the underlying signal that generates the observed time series and HMMs can be used to this end [Chung et al., 1991; Krishnamurthy et al., 1993; Krishnamurthy and Chung, 2007]. Here, we study the performance of our proposed approach in estimating the time-varying oscillatory signal f_{jt} (Equation 3.3) when the Gaussian assumption of ε_t in Equation (3.2) is violated. In particular, we generated 20 time serie, each consisting of 1024 observations from the same simulation setting introduced above, where the innovations are now simulated from heavy-tailed $t-$distributions with 2, 3, 2 degrees of freedom for state 1,2,3, respectively. The linear basis coefficients are chosen to be $\boldsymbol{\beta}_{11} = (3,2)^{'}$, $\boldsymbol{\beta}_{21} = (1.2,4.0)^{'}$, $\boldsymbol{\beta}_{31} = (1.0,5.0)^{'}$, $\boldsymbol{\beta}_{32} = (4.0,3.0)^{'}$. As a measure of performance we compute the mean squared error MSE $= \frac{1}{1024}\sum_{t=1}^{1024}(f_{tz_t} - \hat{f}_{tz_t})^2$ between true and estimated signal and compare the proposed approach with the method of Hadj-Amar et al. [2019] referred to as AutoNOM (Automatic Nonstationary Oscillatory Modelling), which we believe is the state-of-the-art in extracting the signal of nonstationary periodic processes. Our proposed estimation algorithm was run with the same parameterization as above while AutoNOM was performed for 15,000 updates, 3,000 of which were discarded as burn-in, where we fixed 15 maximum number of change points and 5 maximum number of frequencies per segment. The prior means for the number of change-points and frequencies per segment are fixed at 2 and 1, respectively, and the minimum distance between change-

points is set at 10. For both methodologies, the estimated signal is obtained by averaging across MCMC iterations.

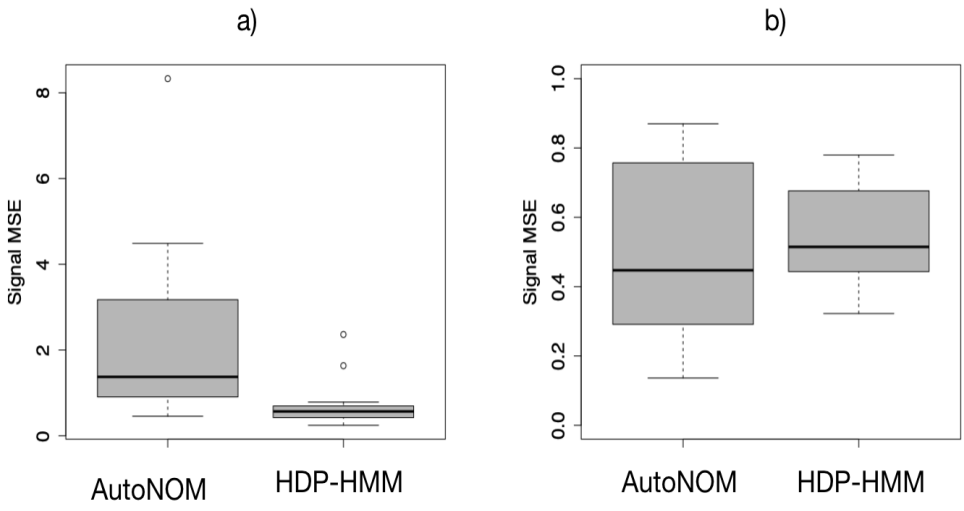

Figure 3.3: Signal extraction with non-Gaussian innovations. Boxplots of the MSE values for AutoNOM and our oscillatory HDP-HMM when (a) the data exhibit recurrent patterns (b) the data do not exhibit recurrent patterns.

Figure 3.3 (a) presents boxplots of the MSE values for AutoNOM and our proposed approach. It becomes clear that the estimates of the signal obtained using our proposed methodology are superior to those obtained using AutoNOM. However, this result is not surprising as the two approaches make different assumptions. In particular, AutoNOM does not assume recurrence of a periodic behaviour and hence needs to estimate the regime-specific modeling parameters each time it detects a new segment, while our oscillatory HDP-HMM has the advantage of using the same set of parameters whenever a particular periodic pattern recurs in the time series. Hence, we also compared the performance of the two approaches in extracting the signal (under non-Gaussian innovations) in a scenario where the time series do not exhibit recurrence. Specifically, we generated 10 time series manifesting two change-points (where the oscillatory behaviour corresponding to the three different

partitions are parameterized as above) and computed the MSE between true and estimated signal as we did in the previous scenario. The corresponding boxplots displayed in Figure 3.3 (b) show that the two approaches seem to perform in similar way, with AutoNOM being slightly more accurate than our oscillatory HDP-HMM. We conclude that both methodologies have their own strengths. Our proposed procedure is superior to AutoNOM in the sense that the additional HMM provides a framework for modelling and explicitly quantifying the switching dynamics and connectivity between different states. On the other hand, AutoNOM is better suited to scenarios where there are nonstationarities arising from singular change-points and the observed oscillatory processes evolve without exhibiting recurrent patterns.

3.3.2 Markov Switching Autoregressive Process

We now investigate the performance of our approach in detecting time-changing periodicities in a scenario where the data generating process shows large departures from our modelling assumptions. The HMM hypothesis which assumes conditionally independent observations given the hidden state sequence, such as the one formulated in Equation (3.2), may sometimes be inadequate in expressing the temporal dependencies occurring in some phenomena. A different class of HMMs that relaxes this assumption is given by the *Markov switching autoregressive process*, also referred to as the AR-HMM [Juang and Rabiner, 1985; Albert and Chib, 1993; Frühwirth-Schnatter, 2006], where an AR process is associated with each state. This model is used to design autoregressive dynamics for the emission distributions while allowing the state transition mechanisms to follow a discrete state Markov chain.

We generated $T = 900$ observations from an AR-HMM with two hidden

states and autoregressive order fixed at $p = 2$, that is

$$\begin{aligned} z_t &\sim \boldsymbol{\pi}_{z_{t-1}}, \\ y_t &= \sum_{l=1}^{p} \psi_l^{(z_t)} y_{t-l} + \varepsilon^{(z_t)}, \end{aligned} \tag{3.17}$$

where $\boldsymbol{\pi}_1 = (0.99, 0.01)$ and $\boldsymbol{\pi}_2 = (0.01, 0.99)$. The AR parameterization $\boldsymbol{\psi}^{(1)} = (1.91, -0.991)$ and $\boldsymbol{\psi}^{(2)} = (1.71, -0.995)$ is chosen in such a way that the state-specific spectral density functions display a pronounced peakedness. Furthermore, $\varepsilon_t^{(1)} \overset{iid}{\sim} \mathcal{N}(0, 0.1^2)$ and $\varepsilon_t^{(2)} \overset{iid}{\sim} \mathcal{N}(0, 0.05^2)$. A realization from this model is shown in Figure 3.4 (top) as a blue solid line. Our proposed estimation algorithm was run for 15,000 iterations 5,000 of which are used as burn-in. At each iteration, we performed 2 reversible-jump MCMC updates for each instantiated set of emission parameters. The rate of the Poisson prior for the number of periodicities is fixed at 10^{-1} and the corresponding truncation level $d_{\max}$ was fixed to 3. The maximum number of states $K_{\max}$ was set equal to 10 whereas the rest of the hyperparameters are specified as in Section 3.3.1. Our procedure seems to overestimate the number of states, as a model with 8 regimes had the highest posterior probability $\hat{\pi}(k = 8 \,|\, \boldsymbol{y}) = 97\%$. However, this is not entirely unexpected as a visual inspection of the realization displayed in Figure 3.4 (top) suggest more than two distinct spectral patterns in the sense that the phases, amplitudes and frequencies appear to vary stochastically within a regime. Figure 3.4 (bottom) shows the estimated time-varying frequency peak along with a 95% credible interval obtained from the posterior sample. The estimate was determined by first selecting the dominant frequency (i.e. the frequency with the highest posterior power) corresponding to each observation and then averaging the frequency estimates over MCMC iterations. While our approach identifies a larger number of states when the data were generated from

an AR-HMM we note that the data generating process are very different from the assumptions of our model and the proposed procedure still provides a reasonable summary of the underlying time changing spectral properties observed in the data. Furthermore, by setting the truncation level K_{max} equal to 2, we retrieve the true transition probability matrix that generates the switching dynamics between the two different autoregressive patterns, as the vectors of transition probabilities obtained using our estimation algorithm are $\hat{\boldsymbol{\pi}}_1 = (0.99, 0.01)$ and $\hat{\boldsymbol{\pi}}_2 = (0.98, 0.02)$.

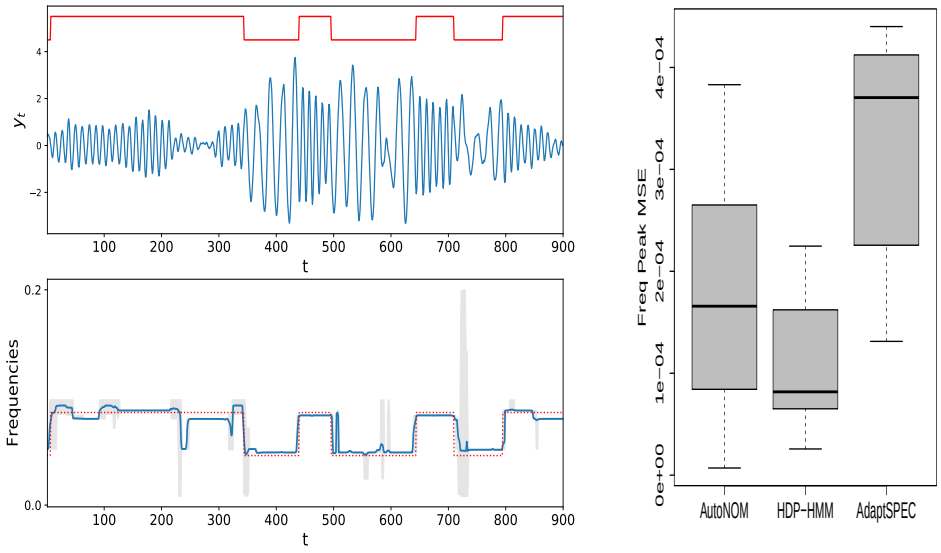

Figure 3.4: (Top) A realization from model (3.17), where the piecewise horizontal line represents the true state sequence. (Bottom) True time varying frequency peak (dotted red line) and the estimate provided by our proposed approach (solid blue line) where we highlight a 95% credible interval obtained from the posterior sample. (Right) Boxplots of the MSE values for AutoNOM, our oscillatory HDP-HMM and AdaptSPEC.

In addition, we simulated 10 time series from model (3.17) and computed the mean squared error MSE $= \frac{1}{900}\sum_{t=1}^{900}(\omega_t - \hat{\omega}_t)$ between the true time-varying frequency peak ω_t and its estimate $\hat{\omega}_t$ for the proposed approach, AutoNOM and the procedure of Rosen et al. [2012], referred to as AdaptSPEC (Adaptive Spectral Estimation). For both AutoNOM and AdaptSPEC, we ran the algorithm for 15,000

MCMC iterations (5,000 of which were used as burn-in), fixed the maximum number of change-points at 15 and set the minimum distance between change-points to 30. The number of spline basis functions for AdaptSPEC is set to 10. AutoNOM is performed using a Poisson prior with rate 10^{-1} for both number of frequencies and number of change-points. Boxplots of the MSE values for the three different methodologies are displayed in Figure 3.4 (right), showing that our oscillatory HDP-HMM seems to outperform the other two approaches in detecting the time-varying frequency peak, for this example. However, our procedure finds some very short sequences (such as in Figure 3.4 (bottom) for $t \approx 200, 500, 700$) demonstrating that the sticky parameter might not always be adequate enough in capturing the correct temporal mode persistence of the latent state sequence. AutoNOM and AdaptSPEC are less prone to this problem as both methodologies are able to specify a minimum time distance between change-points; though, we acknowledge that this constraint might not be optimal when the observed data exhibit relatively rapid changes. We also notice that, not surprisingly, the estimates of the time-varying frequency peak obtained using AutoNOM and our oscillatory HDP-HMM, which are based on a line-spectrum model, are both superior than the ones obtained via the smoothing spline basis of AdaptSPEC, which is built upon a continuous-spectrum setting; this is consistent with the findings in Hadj-Amar et al. [2019]. However, it is important to keep in mind that, while AutoNOM and AdaptSPEC allow to retrospectively analyse the spectral changing properties of a process from an exploratory angle, unlike our spectral HDP-HMM they do not quantify a probabilistic mechanism for the recurrence of periodic dynamic patterns.

3.4 A Case Study: Identifying Recurrence of Sleep Apnea in Humans

Sleep apnea is a chronic respiratory disorder characterized by recurrent episodes of temporary cessation of breathing during sleep [Heinzer et al., 2015]. There are two main types of sleep apnea in humans: Obstructive Sleep Apnea/Hypopnea Syndrome (OSAHS), a disorder caused by complete or partial obstruction of the upper airways whilst still having the drive to breathe from the central nervous system, and Central Sleep Apnea (CSA), a cessation of breathing caused by reduced impulses from the central nervous system, suppressing the neuronal drive to active respiratory muscles to breathe [Banno and Kryger, 2007; Albert et al., 2008]. Instances of OSAHS and CSA can be furthermore subclassified based on the degree of reduction in airflow to the lungs whereby apneas are classified as a reduction of airflow by 90% and hypopneas require a reduction in airflow by at least 30% (with a reduction of blood oxygen levels by at least 3%) [Berry et al., 2017]. OSAHS and CSA both negatively affect several organ systems, such as the heart and kidneys in the long term. They are also associated with increased likelihood of hypertension, stroke, cardiovascular diseases, daytime sleepiness, depression and a diminished quality of life [Teran-Santos et al., 1999; Peker et al., 2002; Young et al., 2002; Yaggi et al., 2005; Dewan et al., 2015]. Hence, detecting apneic and hyponeic events during sleep is one of the primary interests of researchers and clinicians working within the field of sleep medicine as well as healthcare providers,

The airflow trace shown in Figure 3.5 was collected over a time span of 5.5 minutes of continuous breathing and measured via a facemask attached to a pressure sensor. The data are sampled at rate of 4Hz, namely 4 observations per second, for

a total of 1314 data points. During this time, simulated apneic and hyponeic events occurred; apneas appear in the first and second minute and around the start of the fifth minute, where there are two instances of hypopneas in the first half of the second minute and at the start of the fourth minute as marked in Figure 3.5. Note these events were classified by eye by an experienced experimental researcher.

We fitted our oscillatory HDP-HMM to the time series displayed in Figure 3.5 for 100,000 iterations, 60,000 of which were discarded as burn-in, where at each iteration, we carried out 10 reversible-jump MCMC updates for each instantiated set of emission parameters. The truncation level $K_{\max}$ was set to 10, whereas the maximum number of frequencies per state $d_{\max}$ was fixed to 3. The rate of the Poisson prior for the number of frequencies is set equal to 10^{-2} and the prior on the self-transition proportion ρ is specified as Beta $(10^3, 1)$, and the rest of the parameterization is chosen as in Section 3.3.1. The posterior distribution over the number of states had a mode at 7, with posterior probability $\hat{\pi}(k = 7 \mid \boldsymbol{y}) = 77\%$, as reported in Table 3.4. Figure 3.5 shows the fitted signal (yellow line) along with a 95% credible interval obtained from the posterior sample and the estimated hidden state sequence (piecewise horizontal line), where we highlight our model estimate for apnea state (red) and hypopnea state (blue) while reporting the ground truth at the top of the plot. Conditional on the modal number of regimes, the number of periodicities belonging to apnea and hypopnea had a posterior mode at 2 and 3, respectively. Conditional on the modal number of frequencies, Table 3.5 displays the posterior mean and standard deviation of periodicities (in seconds) and powers that characterize the two states classified as apnea and hypopnea, showing that apnea instances seem to be characterized by larger periods and lower amplitude than hypopnea.

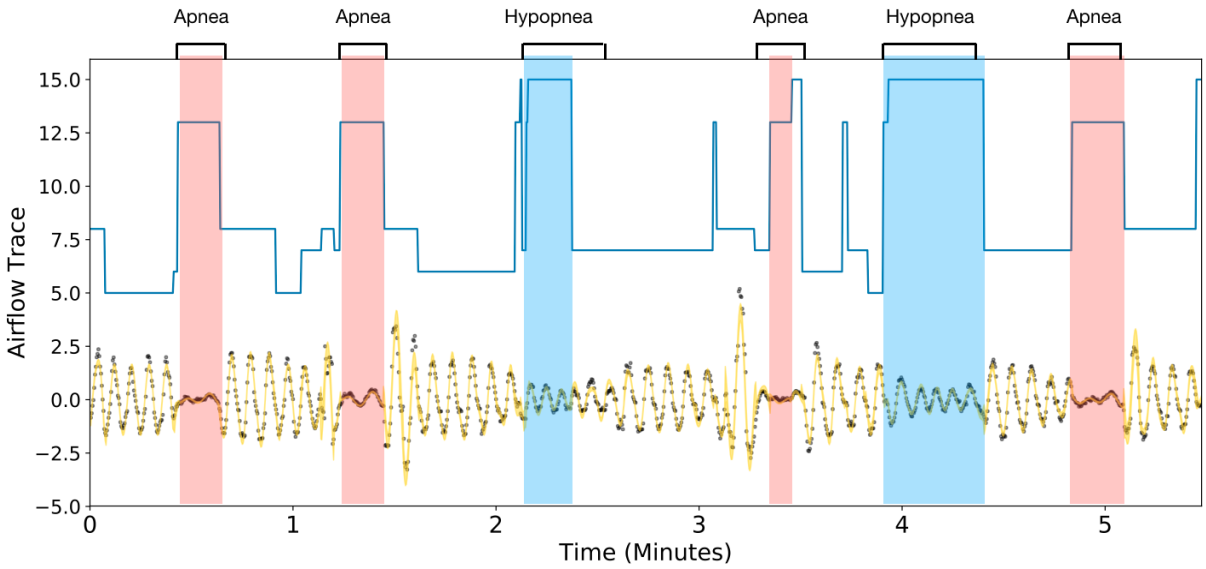

Figure 3.5: Case Study. Dots represent the airflow trace collected over a period of five and half minutes of continuous breathing. The estimated signal (solid line) is shown along with its 95% credible interval. The piecewise horizontal line corresponds to the estimated state sequence where we highlight the states corresponding to estimated apnea (red) and hypopnea (blue), while reporting the ground truth at the top of the plot.

Table 3.4: Case Study. Posterior distribution for number of distinct states k

k	1	2	3	4	5	6	7	8	9	10
$\hat{\pi}(k \mid \boldsymbol{y})$	0.00	0.00	0.00	0.00	0.00	0.21	0.77	0.02	0.00	0.00

Table 3.5: Case study. Posterior mean and standard deviation of periodicities (in seconds) and powers that characterize the two states classified as apnea and hypopnea.

Apnea		**Hypopnea**	
Period	Power	Period	Power
9.862	0.1581	6.084	0.370
$(2.20 \cdot 10^{-2})$	$(1.05 \cdot 10^{-2})$	$(6.95 \cdot 10^{-5})$	$(2.2 \cdot 10^{-2})$
6.823	0.2161	5.252	0.683
$(1.54 \cdot 10^{-5})$	$(1.06 \cdot 10^{-2})$	$(3.48 \cdot 10^{-5})$	$(2.1 \cdot 10^{-2})$
		3.984	0.178
-	-	$(7.910 \cdot 10^{-5})$	$(1.8 \cdot 10^{-2})$

Our estimation algorithm detected all known apnea and hypopnea instances. In order for them to qualify as a clinically relevant obstructive event they must have a minimum length of 10 seconds [Berry et al., 2017]. Thus, we only highlight the clinically relevant instances in Figure 3.5, discarding sequences of duration less than 10 seconds. We also detected a *post sigh* apnea (after the third minute) which is a normal phenomenon to observe in a breathing trace and hence should not count as a disordered breathing event. Again, such an event after a sigh can be identified as a sigh is characterized by an amplitude which is always higher than any other respiratory event and hence can be easily detected. Subtracting the number of sighs from the total number of apneas/hypopneas results in a measure of all apneas of interest without the confounding data from post sigh apneas. A common score to indicate the severity of sleep apnea is given by the Apnea-Hypopnea Index (AHI) which consists of the number of apneas and hypopneas per hour of sleep [Ruehland et al., 2009]. Our proposed approach seems to provide a realistic estimate of the total number of apnea and hypopnea instances recurring in this case study.

3.A Chinese Restaurant Franchise with Loyal Customers

The *Chinese restaurant franchise with loyal customers* [Fox et al., 2011], which extends the *Chinese restaurant franchise* introduced by Teh et al. [2006], is a metaphor which aids in expressing the generative process behind an HDP such as the one formulated here and provides a general framework for performing inference. In a Chinese restaurant franchise the analogy of a *Chinese restaurant process* [Aldous,

1985] is extended to a set of restaurants, where an infinite global menu of dishes is shared across J of such restaurants. The process of seating customers at tables happens in a similar way as for the Chinese restaurant process, but is restaurant-specific. The process of choosing dishes at a specific table happens franchise-wide, namely the dishes are selected with probability proportional to the number of tables (in the entire franchise) that have previously served that dish. However, in the Chinese restaurant franchise with loyal customers, each restaurant in the franchise has a speciality dish which may keep many generations of eating in the same restaurant.

More formally, define $y_{j1}, \ldots, y_{jN_j}$ to be the set of customers in restaurant j, where N_j is the number of customers in restaurant j and each customer is pre-allocated to a specific restaurant designated by that customer's group j. Let us also define indicator random variables t_{ji} and k_{jt}, such that t_{ji} indicates the table assignment for customer i in restaurant j, and k_{jt} the dish assignment for table t in restaurant j. In the Chinese restaurant franchise with loyal customers, customer i in restaurant j chooses a table via $t_{ji} \sim \tilde{\boldsymbol{\pi}}_j$, where $\tilde{\boldsymbol{\pi}}_j \sim \text{GEM}(\eta + \kappa)$, and η and κ are as in Section 3.1.2. Each table is assigned a dish via $k_{jt} \sim (\eta\boldsymbol{\alpha} + \kappa\delta_j)/(\eta + \kappa)$, so that there is more weight on the house speciality dish, namely the dish that has the same index as the restaurant. Despite the fact that this dish can be served in each restaurant, it is more special in the dish's namesake restaurant. Here, $\boldsymbol{\alpha}$ follows a DP with concentration parameters γ, as defined in Equation (3.5), and can be seen as a collection of ratings for the dishes served in the global menu. After marginalizing over $\tilde{\boldsymbol{\pi}}_j$ and $\boldsymbol{\alpha}$, we obtain the following predicting distributions [Teh et al., 2006;

Fox et al., 2011]

$$p\left(t_{ji} \mid t_{j1}, \ldots, t_{ji-1}, \eta, \kappa\right) \propto \sum_{t=1}^{T_j} \tilde{n}_{jt}\, \delta(t_{ji}, t) + (\eta + \kappa)\, \delta\left(t_{ji}, T_j + 1\right), \tag{3.18}$$

$$p\left(k_{jt} \mid \boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_{j-1}, k_{j1}, \ldots, k_{jt-1}, \gamma\right) \propto \sum_{k=1}^{K} m_{\cdot k}\, \delta\left(k_{jt}, k\right) + \gamma\, \delta\left(k_{jt}, K + 1\right), \tag{3.19}$$

where $\tilde{n}_{jt}$ is the number of customers in restaurant j sitting at table t, m_{jk} is the number of tables in restaurant j serving dish k, T_j is the total number of occupied tables in restaurant j, and K is the total number of unique dishes served in the franchise. Marginal counts are described with dots and thus $m_{\cdot k} = \sum_{j=1}^{J} m_{jk}$ is the number of tables in the whole franchise serving dish k. Also we denote with $\boldsymbol{k}_j = (k_{j1}, \ldots, k_{jT_j})'$ the dish assignments for all tables in restaurant j. Equations (3.18) and (3.19) dictate that customer i in restaurant j sits in one among the T_j currently occupied tables, say t_{ji}, with probability proportional to the number of currently seated customers $\tilde{n}_{jt}$, or starts a new table $T_j + 1$ with probability proportional to $\eta + \kappa$. The first customer to sit at a table proceeds to the buffet line and chooses a dish based on the popularity of that dish in the whole franchise, namely chooses an already selected dish k with probability proportional to $m_{\cdot k}$, or orders a new dish $K + 1$ with probability proportional to γ.

Note that in the HMM formulated in Equation (3.1), the value of the state corresponds to the dish index, i.e. $k_{jt_{ji}} = z_{ji} = z_t$, where we follow Fox et al. [2011] and suppose there exist a bijection $f : t \to ji$ of time indexes t to restaurant-customer indexes ji (see Figure 3.6, [Fox, 2009]). This mapping still ensures a Markov structure over the hidden states z_t, which, based on the value of the state

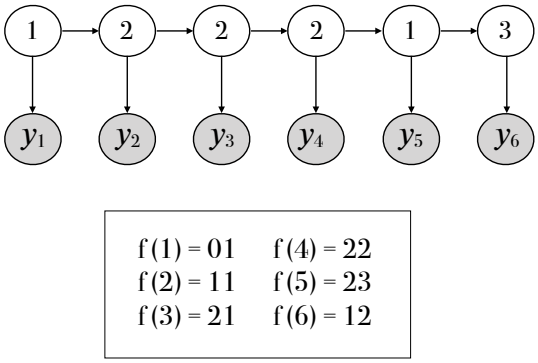

Figure 3.6: Mapping between time indexes t and restaurant-customer indexes ji. This example shows: (top) a directed acyclic graph of an HMM with hidden states $z = (1, 2, 2, 2, 1, 3)$, (bottom) the corresponding bijection $f(t) \rightarrow ij$ of time indexes t to restaurant-customer indexes ji. The Markovian structure given in Equation (3.1) implies that customers y_t are assigned to restaurants determined by z_{t-1}. For example, as $z_3 = 2$, we have that y_4 is seated at restaurant $j = 2$. In addition, y_4 is the second customer to be seated in that restaurant which means that y_4 is assigned a restaurant-customer index y_{22}.

$z_{t-1} = j$, are linked to a restaurant-customer index ji. Moreover, as motivated in Section 3.1.2, we increase the temporal state persistence of the system by inflating probabilities of self-transitions by an amount proportional to κ. The state z_{t-1} determines the restaurant to which the customer y_t is assigned, since $z_t \sim \boldsymbol{\pi}_{z_{t-1}}$, and, in particular, every observations y_t for which $z_{t-1} = j$ are assigned to restaurant j. The increased popularity of the speciality dish of the house, which is determined by the sticky parameter κ, implies that children, say z_{t+1}, are more keen to eat in the same restaurant as their parent, say z_t, and, in turn, more likely to eat the specialty dish of the restaurant. Once a restaurant is instantiated, this family loyalty may keep many generations eating in the same restaurant. We display in Figure 3.7 an example of a Chinese restaurant franchise corresponding to the hidden state sequence $z = (1, 1, 1, 3, 3, 3, 3, 2, 2, 1, 1, 1)$.

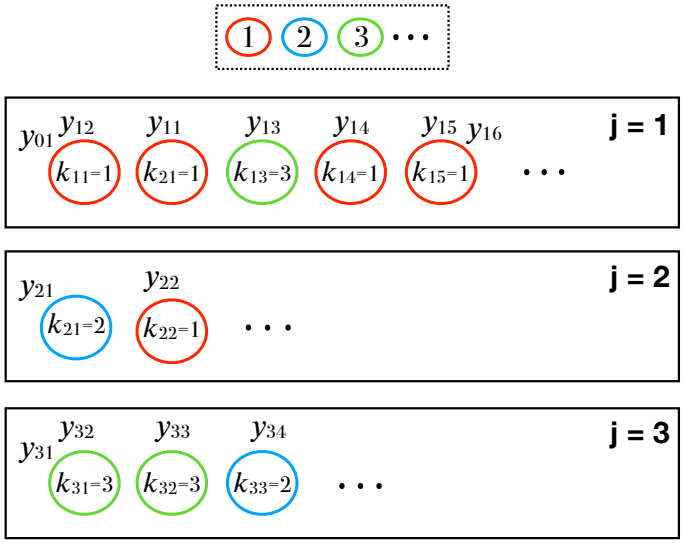

Figure 3.7: Chinese restaurant franchise including $J = 3$ restaurants (rectangles) corresponding to a hidden state sequence $\boldsymbol{z} = (1, 1, 1, 3, 3, 3, 3, 2, 2, 1, 1, 1)$. Conditioned on the value of the states $\boldsymbol{z}$, the customers $\boldsymbol{y}$ are partitioned in both restaurants and dishes, but the table assignments are unknown as multiple tables can serve the same dish. For this example we choose $\boldsymbol{t}_1 = (1, 2, 1, 3, 4, 5, 5)$, $\boldsymbol{t}_2 = (1, 2)$ and $\boldsymbol{t}_3 = (1, 1, 2, 3)$. Custumers ($y_{ji}$) are seated at tables (circles) in the restaurants, where at each table a dish k_{jt} is chosen from a global menu (dotted rectangle on top). This picture is best viewed in colours.

3.A.1 Served and Considered Dishes:

As suggested in [Fox et al., 2011] and briefly introduced in Section 3.2.2, we augment the space and introduce *considered* dishes $\bar{k}_{jt}$ and *override* variables o_{jt} so that we have the following generative process

$$\begin{aligned} \bar{k}_{jt} \,|\, \boldsymbol{\alpha} &\sim \boldsymbol{\alpha} \\ o_{jt} \,|\, \eta, \kappa &\sim \text{Bernoulli}\left(\frac{\kappa}{\eta + \kappa}\right) \\ k_{jt} \,|\, \bar{k}_{jt}, o_{jt} &= \begin{cases} \bar{k}_{jt}, & o_{jt} = 0, \\ j, & o_{jt} = 1. \end{cases} \end{aligned} \tag{3.20}$$

That is, a table first considers a dish $\bar{k}_{jt}$ without taking into account the dish of the house, i.e. $\bar{k}_{jt}$ is chosen from the infinite buffet line according to the ratings provided by $\boldsymbol{\alpha}$. Then, the dish k_{jt} that is actually being served can be the house-speciality dish j, with probability $\rho = \kappa/(\eta + \kappa)$, or the initially considered dish $\bar{k}_{jt}$, with probability $1 - \rho$. The predictive distribution for the considered dishes $\bar{k}_{jt}$ is analogous to Equation (3.19) and is given by

$$p\left(\bar{k}_{jt} \,|\, \bar{\boldsymbol{k}}_1, \bar{\boldsymbol{k}}_2, \ldots, \bar{\boldsymbol{k}}_{j-1}, \bar{k}_{j1}, \ldots, \bar{k}_{jt-1}, \gamma\right) \propto \sum_{k=1}^{\bar{K}} \bar{m}_{\cdot k}\, \delta\left(\bar{k}_{jt}, k\right) + \gamma\, \delta\left(\bar{k}_{jt}, \bar{K} + 1\right), \tag{3.21}$$

where $\bar{m}_{\cdot k}$ is the total number of tables in the franchise that considered dish k, $\bar{K}$ is the number of unique dishes considered in the franchise, and we denote with $\bar{\boldsymbol{k}}_j = (\bar{k}_{j1}, \ldots, \bar{k}_{jT_j})$ the considered dishes in restaurant j. We observe that each served dish had to be considered by at least one table in the franchise. However, there may be some dishes considered that were never served, implying that $\bar{K} \geq K$.

3.A.2 Sufficient Statistics for Updating Dish Ratings

Table counts $\bar{m}_{jk}$ of considered dishes are sufficient statistics [Fox et al., 2007] for updating the collection of dish rating $\boldsymbol{\alpha}$ (see Equation 3.16). In fact, the posterior distribution of $\boldsymbol{\alpha}$ can be written as

$$(\alpha_1, \ldots, \alpha_{\bar{K}}, \tilde{\alpha}) \,\big|\, \bar{\boldsymbol{m}},\ \gamma \ \sim \mathrm{Dir}\,(\bar{m}_{\cdot 1}, \ldots, \bar{m}_{\cdot \bar{K}},\ \gamma), \tag{3.22}$$

where we define $\tilde{\alpha} = \sum_{k=\bar{K}+1}^{\infty} \alpha_k$ and recall that $\bar{K}$ is the number of unique dishes considered in the franchise. Therefore, when the states $\boldsymbol{z}$ are given, it is sufficient to update $\bar{m}_{jk}$, instead of sampling t_{ji} and k_{jt}.

In order to show this result, let $G_0 \sim \mathrm{DP}\,(\gamma, H)$ be the global probability distribution of an HDP having base measure H with support Φ, and concentration parameter γ. We provide an introduction to DP and HDP in Supplementary Material *A* and *B*. Let us also consider a finite partition $\{\phi_1,\ \phi_2, \ldots, \phi_{\bar{K}}, \tilde{\phi}_{\bar{K}}\}$ of Φ, where $\tilde{\phi}_{\bar{K}} = \Phi \setminus \cup_{k=1}^{\bar{K}} \{\phi_k\}$ is the set of all currently unrepresented parameters. The distribution of G_0 over this finite partition may be expressed using the formal definition of DP given by Ferguson [1973] as

$$\begin{aligned}(G_0(\phi_1), \ldots, G_0(\phi_{\bar{K}}),\ G_0(\tilde{\phi}_{\bar{K}})) \,\big|\, \gamma,\ H &\sim \mathrm{Dir}\,(\gamma\, H(\phi_1), \ldots, \gamma\, H(\phi_{\bar{K}}), \gamma\, H(\tilde{\phi}_{\bar{K}})) \\ &\sim \mathrm{Dir}\,(0,\ \ldots, 0,\ \gamma),\end{aligned} \tag{3.23}$$

assuming that H is absolutely continuous with respect to Lebesgue measure, and thus yielding $H(\phi_k) = 0$, for $k = 1, \ldots, \bar{K}$ and $H(\tilde{\phi}_{\bar{K}}) = 1$. A definition of Dirichlet distribution where some of the parameters are allowed to be 0 can be found in Ethier and Griffiths [1993].

We observe that, for each currently instantiated table t, the table-specific considered dish $\bar{\phi}_{jt}$ is drawn from G_0 and is associated, through $\bar{k}_{jt}$, to one among the unique set of dishes $\{\phi_1, \ldots, \phi_{\bar{K}}\}$. Thus, we have $\bar{m}_{\cdot k}$ observations $\bar{\phi}_{jt}$ drawn from G_0 in the franchise assuming value ϕ_k. Then, the posterior of G_0 over the previously defined finite partition is given by

$$\left(G_0(\phi_1), \ldots, G_0(\phi_{\bar{K}}), G_0(\tilde{\phi}_{\bar{K}})\right) \Big| \bar{\boldsymbol{\phi}}, \gamma \sim \mathrm{Dir}\left(\bar{m}_{\cdot 1}, \ldots, \bar{m}_{\cdot \bar{K}}, \gamma\right), \tag{3.24}$$

where we are using the Dirichlet-categorical conjugacy and we denote with $\bar{\boldsymbol{\phi}}$ all realisations $\bar{\phi}_{jt}$ from G_0. Since $(G_0(\phi_1), \ldots, G_0(\phi_{\bar{K}}), G_0(\tilde{\phi}_{\bar{K}}))$ is by definition equal to $(\alpha_1, \ldots, \alpha_{\bar{K}}, \tilde{\boldsymbol{\alpha}})$ we obtain Equation (3.22).

3.B Updating Emission Parameters

3.B.1 Within-Model Move

Updating ω: Samples from the conditional posterior distribution $p(\omega \,|\, \boldsymbol{\beta}, \sigma^2, d, \boldsymbol{y}^*)$ (see Equation (3.11)) are obtained by drawing the frequencies one-at-time using a mixture of M-H steps. We follow Andrieu and Doucet [1999] and Hadj-Amar et al. [2019] and design a mixture proposal distribution of the form

$$q(\omega_l^p \,|\, \omega_l^c) = \xi_\omega \, q_1(\omega_l^p \,|\, \omega_l^c) + (1 - \xi_\omega) \, q_2(\omega_l^p \,|\, \omega_l^c), \qquad l = 1, \ldots, d, \tag{3.25}$$

where q_1 is defined in Equation (3.26) below, q_2 is the density of a Normal $\mathcal{N}(\omega_l^c, \sigma_\omega^2)$, ξ_ω is a positive value such that $0 \leq \xi_\omega \leq 1$, and the superscripts c and p refer to current and proposed values, respectively. Equation (3.25) states that a M-H step

with proposal distribution $q_1(\omega_l^p \mid \omega_l^c)$

$$q_1(\omega_l^p \mid \omega_l^c) \propto \sum_{h=0}^{\tilde{T}-1} I_h \, \mathbb{1}_{[h/T \leq \omega_l^p < (h+1)/T]}, \tag{3.26}$$

is performed with probability ξ_ω. Here $\tilde{T} = \lfloor \hat{T}/2 \rfloor$, $\hat{T}$ is the number of observations in $\hat{\boldsymbol{y}}$, and I_h is the value of the periodogram of $\hat{\boldsymbol{y}}$, i.e the squared modulus of the Discrete Fourier transform evaluated at frequency h/T

$$I_h = \Big| \sum_{t \in I^*} y_t \, \exp\Big(- i 2\pi \frac{h}{T} \Big) \Big|^2,$$

where we recall that $\hat{\boldsymbol{y}}$ is a segment of data that is randomly selected from $\boldsymbol{y}^*$ with probability proportional to the number of observations belonging to that segment. The acceptance probability for this move is

$$\alpha = \min \left\{ 1, \frac{p(\boldsymbol{\omega}^p \mid \boldsymbol{\beta}, \sigma^2, d, \boldsymbol{y}^*)}{p(\boldsymbol{\omega}^c \mid \boldsymbol{\beta}, \sigma^2, d, \boldsymbol{y}^*)} \times \frac{q_1(\omega_l^c)}{q_1(\omega_l^p)} \right\},$$

where $\boldsymbol{\omega}^p = (\omega_1^c, \ldots, \omega_{l-1}^c, \omega_l^p, \omega_{l+1}^c, \ldots, \omega_p^c)'$. With probability 1 - ξ_ω, we carry out a random walk M-H step with proposal distribution $q_2(\omega_l^p \mid \omega_l^c)$, whose density is Normal with mean ω_l^c and variance σ_ω^2, i.e. $\omega_l^p \mid \omega_l^c \sim \mathcal{N}(\omega_l^c, \sigma_\omega^2)$. This move is accepted with probability

$$\alpha = \min \left\{ 1, \frac{p(\boldsymbol{\omega}^p \mid \boldsymbol{\beta}, \sigma^2, d, \boldsymbol{y}^*)}{p(\boldsymbol{\omega}^c \mid \boldsymbol{\beta}, \sigma^2, d, \boldsymbol{y}^*)} \right\}.$$

Low values of ξ_ω yield to high acceptance rate combined with an efficient exploration of the parameter space. We followed Andrieu and Doucet [1999] and set $\sigma_\omega^2 = (1/5T)^2$ and $\xi_\omega = 0.2$.

3.B.2 Trans-Dimensional Moves:

Birth move: If a birth move is attempted, the number of frequencies is proposed to increase by one, namely $d^p = d^c + 1$. The proposed frequency vector $\boldsymbol{\omega}^p$ is constructed as follows

$$\boldsymbol{\omega}^p = (\omega_1^c, \ldots, \omega_{d^c}^c, \omega_{d^p}^p)',$$

where the proposed additional frequency $\omega_{d^p}^p$ is drawn following Hadj-Amar et al. [2019], namely by drawing a candidate value $\omega_{d^p}^p$ uniformly from the union of intervals of the form $[\omega_l^c + \psi_\omega, \omega_{l+1}^c - \psi_\omega]$, for $l = 0, \ldots, d_c$ and denoting $\omega_0^c = 0$ and $\omega_{d^c+1}^c = \phi_\omega$. Here, ψ_ω is a fixed minimum distance between frequencies, which is chosen larger than $\frac{1}{n}$; in fact, when the separation of two frequencies is less than the *Nyquist step* [Priestley, 1981], i.e. $|\omega_l - \omega_{l+1}| < \frac{1}{n}$, the two frequencies are indistinguishable [Dou and Hodgson, 1995]. Also, we sort the proposed vector of frequencies $\boldsymbol{\omega}^p$ to ensure identifiability when performing estimation, as suggested by Andrieu and Doucet [1999]. For proposed $\boldsymbol{\omega}^p$ and given σ^{2c}, the proposed vector of linear coefficients $\boldsymbol{\beta}^p$ is drawn from its Gaussian posterior conditional distribution $p(\boldsymbol{\beta}^p \,|\, \boldsymbol{\omega}^p, \sigma^{2c}, d^p, \boldsymbol{y}^*)$. The proposed state $(d^p, \boldsymbol{\omega}^p, \boldsymbol{\beta}^p)$ is jointly accepted or rejected with probability

$$\alpha = \min\left\{1, \frac{\mathscr{L}(\boldsymbol{\theta}^p \,|\, \boldsymbol{y}^*)}{\mathscr{L}(\boldsymbol{\theta}^c \,|\, \boldsymbol{y}^*)} \times \frac{p(d^p)\, p(\boldsymbol{\omega}^p \,|\, d^p)\, p(\boldsymbol{\beta}^p \,|\, \boldsymbol{\omega}^p, d^p)}{p(d^c)\, p(\boldsymbol{\omega}^c \,|\, d^c)\, p(\boldsymbol{\beta}^c \,|\, \boldsymbol{\omega}^c, d^c)} \times \frac{r_{d^p} \cdot (\frac{1}{d^p}) \cdot q(\boldsymbol{\beta}^c)}{b_{d^c} \cdot q(\omega_{d^p}^p) \cdot q(\boldsymbol{\beta}^p)}\right\}, \tag{3.27}$$

where the likelihood $\mathscr{L}(\cdot \,|\, \boldsymbol{y}^*)$ is provided in Equation (3.8), $p(d)$ is the Poisson density truncated at $d_{\max}$, b_{d^c} and r_{d^p} are the probabilities specified in Equation (3.10), $q(\omega_{d^p}^p)$ is the density of the uniform proposal for sampling the additional frequency, $q(\boldsymbol{\beta}^c)$ and $q(\boldsymbol{\beta}^p)$ are the Normal densities $\boldsymbol{\mathcal{N}}_{2d^c}(\hat{\boldsymbol{\beta}}^c, \boldsymbol{V}_\beta^c)$ and $\boldsymbol{\mathcal{N}}_{2d^p}(\hat{\boldsymbol{\beta}}^p, \boldsymbol{V}_\beta^p)$, respectively (Equation 3.12). Finally, the residual variance σ^2 is

updated in a Gibbs step from Equation (3.14).

Death move: If a death move is attempted, the number of frequencies is proposed to decrease by one, i.e. $d^p = d^c - 1$. The proposed vector of frequencies $\boldsymbol{\omega}^p$ is constructed by choosing with probability $\frac{1}{d^c}$ one of the current frequencies as the candidate frequency to be removed. Conditioned on $\boldsymbol{\omega}^d$ and σ^{2c}, a vector of linear coefficients $\boldsymbol{\beta}^p$ is drawn from its Gaussian posterior conditional distribution. The proposed state $(d^p, \boldsymbol{\omega}^p, \boldsymbol{\beta}^p)$ is jointly accepted or rejected with probability given in Equation (3.27), with the correct adjustment of labelling of the variables, and the terms in the ratio inverted. The residual variance is then updated in a Gibbs step from Equation (3.14).

3.C Updating HMM Parameters

3.C.1 Sampling Hidden State Sequence

Given observations $\boldsymbol{y}$, transition probabilities $\boldsymbol{\pi}$ and emission parameters $\boldsymbol{\theta}$, we sample the state sequence $\boldsymbol{z}$ using the forward-backward procedure introduced by Rabiner [1989] and presented for a Bayesian setting in Fox et al. [2011]. Let us define, recursively, backward messages $b_{t,t-1}(k)$ as

$$b_{T+1,T}(k) = 1, \qquad b_{t,t-1}(k) \propto \sum_{j=1}^{K_{max}} \pi_{kj}\, \mathcal{N}(y_t;\, f_{jt},\, \sigma_j^2)\, b_{t+1,t}(j), \quad t = T, \ldots, 2, \tag{3.28}$$

for each state $k = 1, \ldots, K_{max}$, where we recall that $\pi_{kj} = p(z_t = j \,|\, z_{t-1} = k)$. Here, $\mathcal{N}(y_t;\, f_{jt},\, \sigma_j^2)$ denotes the density of a Gaussian distribution with mean $f_{jt} = \boldsymbol{x}_t(\boldsymbol{\omega}_j)'\boldsymbol{\beta}_j$ (as in Equation 3.3) and variance σ_j^2, evaluated at y_t. Note that $b_{t,t-1}(k) \propto p(y_t, \ldots, y_T \,|\, z_{t-1} = k, \boldsymbol{\pi}, \boldsymbol{\theta})$, namely a backward message passed from

z_t to z_{t-1} is proportional to the probability of the partial observation sequence from t to the end, given the state z_{t-1}. We then observe that we may recursively sample each state z_t conditioned on z_{t-1} since

$$p(\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta}) = p(z_1 \mid \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta}) \prod_{t=2}^{T} p(z_t \mid z_{t-1}, \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta}).$$

The first state z_1 is sampled from the following posterior conditional distribution

$$p(z_1 = k \mid \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta}) \propto \pi_{0k} \mathcal{N}(y_1;\, f_{k\,t},\, \sigma_k^2)\, b_{2,1}(k), \tag{3.29}$$

where we recall that π_{0k} is the initial transition distribution $p(z_1 = k)$. The rest of the sequence $z_2, \ldots, z_T$ are then sampled, recursively, from

$$p(z_t = k \mid z_{t-1} = j,\, \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta}) \propto \pi_{jk} \mathcal{N}(y_t;\, f_{k\,t},\, \sigma_k^2)\, b_{t+1,t}(k).$$

3.C.2 Sampling Table Counts and Override Variables

Conditioned on state sequence $\boldsymbol{z}$ and collection of dish ratings $\boldsymbol{\alpha}$, as well as hyper-parameters η and κ, we sample m_{jk}, o_{jt} and $\bar{m}_{jk}$ as in Fox et al. [2011] from

$$\begin{aligned} p(\boldsymbol{m},\, \boldsymbol{o},\, \bar{\boldsymbol{m}} \mid \boldsymbol{z},\, \boldsymbol{\alpha},\, \eta, \kappa) = p(\boldsymbol{m} \mid \boldsymbol{z},\, \boldsymbol{\alpha},\, \eta,\, \kappa) \times p(\boldsymbol{o} \mid \boldsymbol{m},\, \boldsymbol{z},\, \boldsymbol{\alpha},\, \eta,\, \kappa) \\ \times p(\bar{\boldsymbol{m}} \mid \boldsymbol{o},\, \boldsymbol{m},\, \boldsymbol{z},\, \boldsymbol{\alpha},\, \eta,\, \kappa), \end{aligned}$$

where $\boldsymbol{m}$ and $\bar{\boldsymbol{m}}$ denote the vectors of table counts of served and considered dishes, respectively, and $\boldsymbol{o}$ is the vector of override variables. Hence, we first draw $\boldsymbol{m}$, we then sample $\boldsymbol{o}$, and finally determine $\bar{\boldsymbol{m}}$.

Updating m_{jk} : Let consider sampling from $p(\boldsymbol{m} \mid \boldsymbol{z},\, \boldsymbol{\alpha},\, \eta,\, \kappa)$. Conditional on the value of the states $\boldsymbol{z}$, the customers $\boldsymbol{y}$ are partitioned according to both restau-

rants and dishes, but the table assignments are unknown because multiple tables can be served the same dish (see Figure 3.7). Table assignments may be sampled from the following conditional distribution

$$p\left(t_{ji}=t \,|\, \boldsymbol{t}^{-ji},\, k_{jt}=k,\, \boldsymbol{k}^{-jt},\, \boldsymbol{\alpha},\, \eta,\, \kappa\right) \propto \begin{cases} \tilde{n}_{jt}^{-ji}, & t \in \{1,\ldots,T_j\} \\ \eta\,\alpha_k + \kappa\,\delta(k,j), & t = T_j+1, \end{cases} \tag{3.30}$$

where $\tilde{n}_{jt}^{-ji}$ is the number of customers in restaurant j that sits at table t without including the customer y_{ji}, $\boldsymbol{t}^{-ji}$ are the table assignments for all customers in restaurant j except for y_{ji}, and similarly $\boldsymbol{k}^{-jt}$ denote the dish assignments for all tables without counting table t in restaurant j. Equation (3.30) implies that we can sample table assignments after knowing the dish assignments and also states that a custumer enters the restaurant and chooses an already occupied table with probability proportional to $\tilde{n}_{jt}^{-ji}$, or starts a new table served dish k with probability proportional to $\eta\,\alpha_k + \kappa\,\delta(k,j)$. Note that, when joining a new table, a mass proportional to κ is added if the dish assigned to that table was the house speciality dish. Moreover, the form of Equation (3.30) also implies that a customer table assignment t_{ji}, conditioned on the dish assignment k, follows a DP with concentration parameter $\eta\,\alpha_k + \kappa\,\delta\,(j,k)$. Hence, the inference algorithm is performed by introducing a set of auxiliary random variables $t_{jk}^{(i)}$ which indicate whether or not customer i in restaurant j has joined a new table serving dish k. These variable are sampled in the following way

$$t_{jk}^{(i)} \,\Big|\, n_{jk},\, \alpha_k,\, \eta, \kappa \;\sim \text{Bernoulli}\left(\frac{\eta\,\alpha_k + \kappa\,\delta(k,j)}{\eta\,\alpha_k + \kappa\,\delta(k,j) + i}\right), \qquad i = 1,\ldots,n_{jk}, \tag{3.31}$$

recalling that n_{jk} is the number of transitions from state j to state k. Table counts

m_{jk} are then determined by summing over these auxiliary random variables, i.e.

$$m_{jk} = \sum_{i=1}^{n_{jk}} t_{jk}^{(i)}. \tag{3.32}$$

Updating o_{jt} : Let now consider sampling from $p(\boldsymbol{o} \,|\, \boldsymbol{m}, \boldsymbol{z}, \boldsymbol{\alpha}, \eta, \kappa)$. First, we observe that, when $j \neq k$, there are m_{jk} tables for which $o_{jt} = 0$, since the corresponding considered dishes were not overridden with probability one. On the other hand, when $j = k$, the served dish $k_{jt} = j$, which is the house speciality dish, can arise from either an override decision (i.e. $o_{jt} = 1$) or a considered dish $\bar{k}_{jt} = j$ which has not been overridden (i.e. $o_{jt} = 0$). The resulting conditional distribution is given below

$$p(o_{jt} \,|\, k_{jt} = j, \boldsymbol{\alpha}, \rho) \propto \begin{cases} \rho & o_{jt} = 1, \\ \alpha_j (1 - \rho) & o_{jt} = 0. \end{cases} \tag{3.33}$$

Hence, we may sample m_{jj} Bernoulli random variables from Equation (3.33) or sample $o_{j\cdot} = \sum_t o_{jt}$, i.e. the total number of override dishes in restaurant j, from the following Binomial distribution

$$o_{j\cdot} \sim \text{Binomial}\left(m_{jj}, \frac{\alpha_j (1 - \rho)}{\rho + \alpha_j (1 - \rho)}\right) \tag{3.34}$$

Updating $\bar{m}_{jk}$: Conditioned on table counts m_{jk} of served dishes for all j and k, and override variables o_{jt} for each of these instantiated tables, the number of tables $\bar{m}_{jk}$ that considered dish k in restaurant j is computed deterministically in the following way

$$\bar{m}_{jk} = \begin{cases} m_{jk} & j = k, \\ m_{jj} - o_{j\cdot} & j \neq k. \end{cases} \tag{3.35}$$

noticing that, for house speciality dishes, we subtract the sum $o_{j\cdot}$ of override tables within restaurant j, since $o_{jt} = 1$ holds when table t is served dish j.

3.C.3 Sampling Hyperparameters

We follow Fox et al. [2011] and parameterize the model by $(\eta+\kappa)$ and $\rho = \kappa/(\eta+\kappa)$. The previous parameterization can be restored using the equalities $\eta = (1-\rho)(\eta+\kappa)$ and $\kappa = \rho(\eta+\kappa)$. We place a Beta (c_ρ, d_ρ) prior on the expected value ρ of the override variable o_{jt} and a vague Gamma $(a_{\eta+\kappa}, b_{\eta+\kappa})$ prior on $(\eta+\kappa)$. A Gamma (a_γ, b_γ) prior is placed on the concentration parameter γ. The posterior distribution of these hyperparameters and the concentration parameter γ are given below. Here, we omit the full derivations of these results which follow closely Escobar and West [1995] and Teh et al. [2006] and are provided in details in Fox et al. [2007].

Updating $(\eta+\kappa)$: Let J be the number of instantiated restaurant in the franchise at a given iteration of the sampler. We introduced auxiliary random variables $\boldsymbol{r} = \{r_1, \ldots, r_j\}$, where each $r_j \in [0,1]$, and $\boldsymbol{s} = \{s_1, \ldots, s_J\}$, with each $s_j \in \{0,1\}$. Then, the posterior distribution of $(\eta+\kappa)$, conditioned on these newly introduced set of parameters is given by

$$(\eta+\kappa) \,\big|\, \boldsymbol{r},\, \boldsymbol{s},\, \boldsymbol{y} \;\sim\; \text{Gamma}\left(a_{\eta+\kappa} + m_{\cdot\cdot} - \sum_{j=1}^{J} s_j,\; b_{\eta+\kappa} - \sum_{j=1}^{J} \log r_j\right), \tag{3.36}$$

where the auxiliary variables $\boldsymbol{r}$ and $\boldsymbol{s}$ are updated in a Gibbs step from

$$\begin{aligned} r_j \,\big|\, \eta + \kappa,\, \boldsymbol{r}^{-j},\, \boldsymbol{s},\, \boldsymbol{y} \;&\sim\; \text{Beta}\,(\eta + \kappa + 1, n_{j\cdot}), \\ s_j \,\big|\, \eta + \kappa,\, \boldsymbol{s}^{-j},\, \boldsymbol{r},\, \boldsymbol{y} \;&\sim\; \text{Bernoulli}\left(\frac{n_{j\cdot}}{n_{j\cdot} + \eta + \kappa}\right), \end{aligned} \tag{3.37}$$

and we recall that marginal counts are described with dots and hence $m_{\cdot\cdot}$ is the total number of tables serving dishes in the franchise, and n_{jk} is the number of transitions from state j to k in the state sequence $\boldsymbol{z}$.

Updating γ : The posterior distribution of γ can be updated in a similar way. We introduce auxiliary random variables $\psi \in [0, 1]$ and $\xi \in \{0, 1\}$ and draw γ from the full conditional

$$\gamma \,\big|\, \psi,\, \xi,\, \boldsymbol{y} \;\sim\; \text{Gamma}\Big(a_\gamma + \bar{K} - \xi,\, b_\gamma - \log \psi\Big), \tag{3.38}$$

where the auxiliary variables ψ and ξ are drawn in a Gibbs step from

$$\begin{aligned} \psi \,\big|\, \gamma,\, \xi,\, \boldsymbol{y} \;&\sim\; \text{Beta}\,(\gamma + 1,\, \bar{m}_{\cdot\cdot}), \\ \xi \,\big|\, \gamma,\, \psi\, \boldsymbol{y} \;&\sim\; \text{Bernoulli}\left(\frac{\bar{m}_{\cdot\cdot}}{\bar{m}_{\cdot\cdot} + \gamma}\right), \end{aligned} \tag{3.39}$$

where we recall that $\bar{K}$ is the number of unique considered dishes in the franchise, and $\bar{m}_{j\cdot}$ represents the total number of tables in restaurant j with considered dishes that were not overridden.

Updating ρ: Finally, we sample the self-transition proportion ρ from its conditional posterior distribution which is given by

$$\rho \,|\, \boldsymbol{o} \;\sim\; \text{Beta}\left(\sum_{j=1}^{J} o_{j\cdot} + c_\rho,\, m_{\cdot\cdot} - \sum_{j=1}^{J} o_{j\cdot} + d_\rho\right). \tag{3.40}$$

Chapter 4

Summary and Discussion

This thesis presented novel Bayesian approaches to the automated analysis of non-stationary periodic time series. The development of our methodologies was motivated by our collaborators in the medical life sciences who observe dense physiological signals, here breathing traces and skin temperature, which naturally are subject to oscillatory behaviour with unknown periodicities which may change, often suddenly, over time. In particular, to date no probabilistic modelling approach has been proposed that is able to address the time changing spectral characteristics of the data in a satisfactory way such that it could be conceived to be used, for example, for detecting, and ultimately predicting, instances of apnea.

In Chapter 2, we introduced a novel Bayesian approach for analyzing non-stationary time series data that exhibit oscillatory behaviour. Our approach is based on the assumption that, conditional on the position and number of change-points, the time series can be approximated by a piecewise changing sinusoidal regression model. The timing and number of changes are unknown, along with the number and values of relevant periodicities in each regime. Bayesian inference is performed via a reversible jump MCMC algorithm that can simultaneously estimate both the num-

ber and location of change-points, as well as the number, frequency and magnitude of sinusoids within each segment. Our methodology can be seen as a novel and relevant extension of the work in Andrieu and Doucet [1999] to the nonstationary setting. We illustrated the utility of our methodology in two case studies. First, we analyzed human skin temperature time series data obtained from a wearable device, which exhibited unknown periodicities that changed over time in an abrupt manner. Our proposed methodology identified interesting oscillations whose frequencies are consistent with ultradian oscillations between REM and non-REM sleep stages. Second, we characterized the occurrence of sleep apnea in large breathing data sets resulting from *in vivo* plethysmograph studies on rodents. Our spectral investigation was able to distinguish very sharp peaks, corresponding to different nearby frequencies, that may be associated to the different actions of the rodent including instances of apnea. Although we have not discussed this in detail here, several diagnostics for monitoring convergence were carried out in both simulation and case studies. In particular, we verified that the target posterior distribution reached a stable regime by analyzing the trace plot of the log likelihood across MCMC iterations [Marin and Robert, 2007]. We are aware that assessing convergence only based on this simple tool may sometimes be misleading since stable values of the log likelihood could simply mean that the Markov chain is stuck in some local mode of the posterior distribution. Additionally, conditioned on the modal number of change-points and modal number of frequencies per regime, we have also monitored (within-model) convergence by analyzing the traces and running averages plots for all parameters across MCMC iterations. Comparable results were also obtained when running several chains starting at over-dispersed initial values. We notice that the diagnostic tool recently introduced by Bruce et al. [2018] and Li and Krafty [2019] to assess convergence for reversible jump MCMC samplers appears relevant. In the context

of adaptive spectral analysis of nonstationary time series, they point out that although the number of partitions change across models, a power spectrum is defined at each time point. The power spectrum is modeled with a fixed number of splines, yielding a vector of summary measures of parameters that maintain the same interpretation across models in their samplers. However, our proposed sampler has a further layer of variable dimensionality, as not only the number and locations of the change-points may change from one iteration to the next, but also the number of frequencies in each segment are not fixed throughout the simulation. Finally, although a Gaussian distribution is assumed, it is conceivable that our model can be extended to allow for other error distributions in Equation (2.1). For example, a generalized linear model [McCullagh and Nelder, 1989] may be used to model periodic count data by assuming that the observed data follows a Poisson distribution, i.e. $y_t \sim \text{Poisson}(\mu_t)$. The logarithmic link function of the expected value μ_t of the response variable y_t may be expressed as $\log(\mu_t) = \sum_{j=1}^{k+1} f(t, \boldsymbol{\beta}_j, \boldsymbol{\omega}_j)\mathbb{1}_{[t \in I_j]}$, where the definitions of the variables are the same as for Equation (2.1). Bayesian inference can, in principle, be achieved in a similar way as described in Chapter 2, namely by iterating between segment and change-point model moves, where the formulation of the acceptance probabilities and some proposal distributions need to be modified accordingly. We believe that such an extension would find use in several ranges of applications, for example in studying population cycles in ecology and epidemiology, where the abundance of species are measured as count variables [White and Bennetts, 1996; Bhaskaran et al., 2013; Bramness et al., 2015].

Additionally to recognize temporal changes, it is of interest to identify and model the recurrence of a relevant cyclical pattern in a probabilistic manner. Chapter 3 presented a novel HMM approach which can address the challenges of modelling periodic phenomena whose behaviour switches dynamically over time, and

provides a flexible framework for gaining information from them. The number of states is assumed unknown as well as their relevant periodicities which may differ over the different regimes since each regime is represented by a different periodic pattern. To address flexibility on the number of states, we assume an HDP that penalises rapid changing dynamics of the process and provides effective control over the switching rate as in Fox et al. [2011]. The variable dimensionality with respect to the number of frequencies that specifies the different states is tackled by developing an appropriate reversible-jump MCMC algorithm. We note that we used a similar sinusoidal modelling approach, for cyclical data that show regime shifts in periodicity, amplitude and phase, to the one developed in Chapter 2. However, the statistical methodology proposed in Chapter 3 is built upon an HMM framework, which can additionally model and recognize the occurrence and recurrence of periodic patterns over time by assigning probabilities with which these transitions occur between different states. We also observe that our proposed methodology will revisit a regime only if that regime is not out of phase with earlier visits. Future research may address this matter by first specifying the oscillatory mean function of Equation (3.3) as $f_{tj} = \sum_{l=1}^{d_j}(A_{j,l}\cos(2\pi\omega_{j,l}\,t + \tau_{j,l}))$ and then allow the phase $\tau_{j,l}$ associated with each frequency $\omega_{j,l}$ to vary independently of previous visits to the same state. However, it may not always be correct to a priori assume that a shift in phase for a particular periodicity should be classified with the same state. For example, when analyzing physiological data in the context of cancer chronotherapy [Lévi and Schibler, 2007] or delayed sleep phase syndrome [Sack et al., 2007], detecting a phase shift in the circadian rhythm of a patient might be crucial in determining when and which particular treatment to give.

We illustrated the use of our approach in a case study relevant to respiratory research, where our methodology was able to identify recurring instances of sleep

apnea in human breathing traces. Although here we have focused on the detection of apnea instances, our proposed methodology provides a very flexible and general framework to analyze different breathing patterns. The growth of information and communication technologies permits new advancements in the health care system to facilitate support in the homes of patients in order to proactively enhance their health and well-being. We believe that our proposed HMM approach can aid the iterative feedback between clinical investigations in sleep apnea research and practice with computational, statistical and mathematical analysis and modelling of disease progression and remission, treatment responses and adverse events as well as disease prevention.

Bibliography

Adak, S. [1998], 'Time-dependent spectral analysis of nonstationary time series', *Journal of the American Statistical Association* **93**(444), 1488–1501.

Akaike, H. [1974], 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Albert, J. H. and Chib, S. [1993], 'Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts', *Journal of Business & Economic Statistics* **11**(1), 1–15.

Albert, R. K., Spiro, S. G. and Jett, J. R. [2008], *Clinical Respiratory Medicine*, Elsevier Health Sciences.

Aldous, D. J. [1985], Exchangeability and related topics, *in* 'École d'Été de Probabilités de Saint-Flour XIII1983', Springer, pp. 1–198.

Altevogt, B. M., Colten, H. R. et al. [2006], *Sleep Disorders and Sleep Deprivation: an Unmet Public Health Problem*, Washington (DC): National Academies Press.

Andrieu, C. and Doucet, A. [1999], 'Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC', *IEEE Transactions on Signal Processing* **47**(10), 2667–2676.

Andrieu, C., Doucet, A. and Holenstein, R. [2010], 'Particle markov chain monte carlo methods', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**(3), 269–342.

Banno, K. and Kryger, M. H. [2007], 'Sleep apnea: clinical investigations in humans', *Sleep Medicine* **8**(4), 400–426.

Bartlett, M. S. [1950], 'Periodogram analysis and continuous spectra', *Biometrika* **37**(1/2), 1–16.

Beal, M. J., Ghahramani, Z. and Rasmussen, C. E. [2002], The infinite hidden Markov model, *in* 'Advances in Neural Information Processing Systems', pp. 577–584.

Berlad, I., Shlitner, A., Ben-Haim, S. and Lavie, P. [1993], 'Power spectrum analysis and heart rate variability in Stage 4 and REM sleep: evidence for state-specific changes in autonomic dominance', *Journal of Sleep Research* **2**(2), 88–90.

Bernardo, J. M. and Smith, A. F. [2009], *Bayesian Theory*, Vol. 405, John Wiley & Sons.

Berry, R. B., Brooks, R., Gamaldo, C., Harding, S. M., Lloyd, R. M., Quan, S. F., Troester, M. T. and Vaughn, B. V. [2017], 'Aasm scoring manual updates for 2017 (version 2.4)', *Journal of Clinical Sleep Medicine* **13**(05), 665–666.

Bhar, R. and Hamori, S. [2004], *Hidden Markov models: applications to financial economics*, Vol. 40, Springer Science & Business Media.

Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L. and Armstrong, B. [2013], 'Time series regression studies in environmental epidemiology', *International Journal of Epidemiology* **42**(4), 1187–1195.

Bishop, C. M. [2006], *Pattern Recognition and Machine Learning (Information Science and Statistics)*, New York: Springer-Verlag.

Blackman, R. B. and Tukey, J. W. [1958], 'The measurement of power spectra from the point of view of communications engineeringpart I', *Bell System Technical Journal* **37**(1), 185–282.

Blackwell, D., MacQueen, J. B. et al. [1973], 'Ferguson distributions via Pólya urn schemes', *The Annals of Statistics* **1**(2), 353–355.

Bramness, J. G., Walby, F. A., Morken, G. and Røislien, J. [2015], 'Analyzing seasonal variations in suicide with Fourier Poisson time-series regression: a registry-based study from Norway, 1969–2007', *American Journal of Epidemiology* **182**(3), 244–254.

Bretthorst, G. L. [1988], *Bayesian Spectrum Analysis and Parameter Estimation*, Vol. 48, New York: Springer-Verlag.

Bretthorst, G. L. [1990], 'Bayesian analysis. I. Parameter estimation using quadrature NMR models', *Journal of Magnetic Resonance* **88**(3), 533–551.

Brillinger, D. R. [1981], *Time series: data analysis and theory*, Vol. 36, SIAM.

Brockwell, P. J., Davis, R. A. and Calder, M. V. [2002], *Introduction to Time Series and Forecasting*, Vol. 2, Springer.

Brockwell, P. J., Davis, R. A. and Fienberg, S. E. [1991], *Time Series: Theory and Methods*, Springer Science & Business Media.

Bruce, S. A., Hall, M. H., Buysse, D. J. and Krafty, R. T. [2018], 'Conditional adaptive Bayesian spectral analysis of nonstationary biomedical time series', *Biometrics* **74**(1), 260–269.

Carskadon, M. A., Dement, W. C. et al. [2005], 'Normal human sleep: an overview', *Principles and Practice of Sleep Medicine* **4**, 13–23.

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

Celeux, G., Hurn, M. and Robert, C. P. [2000], 'Computational and inferential difficulties with mixture posterior distributions', *Journal of the American Statistical Association* **95**(451), 957–970.

Chopin, N., Rousseau, J. and Liseo, B. [2013], 'Computational aspects of Bayesian spectral density estimation', *Journal of Computational and Graphical Statistics* **22**(3), 533–557.

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**(468), 1050–1059.

Chung, S.-H., Krishnamurthy, V. and Moore, J. [1991], 'Adaptive processing techniques based on hidden Markov models for characterizing very small channel currents buried in noise and deterministic interferences', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **334**(1271), 357–384.

Cogburn, R. and Davis, H. T. [1974], 'Periodic splines and spectral estimation', *The Annals of Statistics* pp. 1108–1126.

Dahlhaus, R. et al. [1997], 'Fitting time series models to nonstationary processes', *The Annals of Statistics* **25**(1), 1–37.

Davis, E. M. and ODonnell, C. P. [2013], 'Rodent models of sleep apnea', *Respiratory Physiology & Neurobiology* **188**(3), 355–361.

Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. [2006], 'Structural break estimation for nonstationary time series models', *Journal of the American Statistical Association* **101**(473), 223–239.

Del Moral, P., Doucet, A. and Jasra, A. [2006], 'Sequential Monte Carlo samplers', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(3), 411–436.

Denison, D., Mallick, B. and Smith, A. [1998], 'Automatic Bayesian curve fitting', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**(2), 333–350.

Dewan, N. A., Nieto, F. J. and Somers, V. K. [2015], 'Intermittent hypoxemia and osa: implications for comorbidities', *Chest* **147**(1), 266–274.

Djuric, P. M. [1996], 'A model selection rule for sinusoids in white Gaussian noise', *IEEE Transactions on Signal Processing* **44**(7), 1744–1751.

Dou, L. and Hodgson, R. [1995], 'Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation. I', *Inverse Problems* **11**(5), 1069.

Dou, L. and Hodgson, R. [1996], 'Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation: II', *Inverse Problems* **12**(2), 121.

Ephraim, Y. and Merhav, N. [2002], 'Hidden Markov processes', *IEEE Transactions on Information Theory* **48**(6), 1518–1569.

Escobar, M. D. and West, M. [1995], 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association* **90**(430), 577–588.

Ethier, S. N. and Griffiths, R. [1993], 'The transition function of a fleming-viot process', *The Annals of Probability* pp. 1571–1590.

Ferguson, T. S. [1973], 'A Bayesian analysis of some nonparametric problems', *Annals of Statistics* **1**(2), 209–230.

Fox, E. B. [2009], Bayesian nonparametric learning of complex dynamical phenomena, PhD thesis, Massachusetts Institute of Technology.

Fox, E. B., Sudderth, E. B., Jordan, M. I. and Willsky, A. S. [2007], 'The sticky HDP-HMM: Bayesian nonparametric hidden Markov models with persistent states', *Arxiv preprint* .

Fox, E. B., Sudderth, E. B., Jordan, M. I. and Willsky, A. S. [2011], 'A sticky HDP-HMM with application to speaker diarization', *The Annals of Applied Statistics* pp. 1020–1056.

Frühwirth-Schnatter, S. [2006], *Finite Mixture and Markov Switching Models*, Springer Science & Business Media.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. [2014], *Bayesian Data Analysis*, Vol. 2, Boca Raton: CRC Press.

Godsill, S. J. [2001], 'On the relationship between Markov chain Monte Carlo methods for model uncertainty', *Journal of Computational and Graphical Statistics* **10**(2), 230–248.

Green, P. J. [1995], 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika* **82**(4), 711–732.

Hadj-Amar, B., Finkenstädt Rand, B., Fiecas, M., Lévi, F. and Huckstepp, R. [2019], 'Bayesian Model Search for Nonstationary Periodic Time series', *Journal of the American Statistical Association* **0**(0), 1–29.

Han, C. and Carlin, B. P. [2001], 'Markov chain monte carlo methods for computing bayes factors: A comparative review', *Journal of the American Statistical Association* **96**(455), 1122–1132.

Han, F., Subramanian, S., Price, E. R., Nadeau, J. and Strohl, K. P. [2002], 'Periodic breathing in the mouse', *Journal of Applied Physiology* **92**(3), 1133–1140.

Harsch, I. A., Schahin, S. P., Brückner, K., Radespiel-Tröger, M., Fuchs, F. S., Hahn, E. G., Konturek, P. C., Lohmann, T. and Ficker, J. H. [2004], 'The effect of continuous positive airway pressure treatment on insulin sensitivity in patients with obstructive sleep apnoea syndrome and type 2 diabetes', *Respiration* **71**(3), 252–259.

Heinzer, R., Vat, S., Marques-Vidal, P., Marti-Soler, H., Andries, D., Tobback, N., Mooser, V., Preisig, M., Malhotra, A., Waeber, G. et al. [2015], 'Prevalence of sleep-disordered breathing in the general population: the Hypnolaus study', *The Lancet Respiratory Medicine* **3**(4), 310–318.

Hjort, N. L., Holmes, C., Müller, P. and Walker, S. G. [2010], *Bayesian Nonparametrics*, Vol. 28, Cambridge University Press.

Huang, Q., Cohen, D., Komarzynski, S., Li, X.-M., Innominato, P., Lévi, F. and Finkenstädt, B. [2018], 'Hidden Markov models for monitoring circadian

rhythmicity in telemetric activity data', *Journal of The Royal Society Interface* **15**(139), 20170885.

Hurn, M., Justel, A. and Robert, C. P. [2003], 'Estimating mixtures of regressions', *Journal of computational and graphical statistics* **12**(1), 55–79.

Ignatov, T. [1982], 'On a constant arising in the asymptotic theory of symmetric groups, and on Poisson-Dirichlet measures', *Theory of Probability & Its Applications* **27**(1), 136–147.

Ishwaran, H. and Zarepour, M. [2002], 'Exact and approximate sum representations for the Dirichlet process', *Canadian Journal of Statistics* **30**(2), 269–283.

Jasra, A., Holmes, C. C. and Stephens, D. A. [2005], 'Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling', *Statistical Science* pp. 50–67.

Jeffreys, H. [1998], *The theory of probability*, OUP Oxford.

Jelinek, F. [1997], *Statistical Methods for Speech Recognition*, MIT press.

Jennison, C. [1997], 'Discussion of on Bayesian analysis of mixtures with an unknown number of components, by s. richardson and pj green', *J. Roy. Statist. Soc. Ser. B* **59**, 778–779.

Juang, B.-H. and Rabiner, L. [1985], 'Mixture autoregressive hidden Markov models for speech signals', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **33**(6), 1404–1413.

Juang, B. H. and Rabiner, L. R. [1991], 'Hidden Markov models for speech recognition', *Technometrics* **33**(3), 251–272.

Kivinen, J. J., Sudderth, E. B. and Jordan, M. I. [2007], Learning multiscale representations of natural scenes using Dirichlet processes, *in* '2007 IEEE 11th International Conference on Computer Vision', IEEE, pp. 1–8.

Komarzynski, S., Huang, Q., Innominato, P. F., Maurice, M., Arbaud, A., Beau, J., Bouchahda, M., Ulusakarya, A., Beaumatin, N., Breda, G. et al. [2018], 'Relevance of a mobile internet platform for capturing inter-and intrasubject variabilities in circadian coordination during daily routine: pilot study', *Journal of Medical Internet Research* **20**(6), e204.

Kräuchi, K. [2002], 'How is the circadian rhythm of core body temperature regulated?', *Clinical Autonomic Research* **12**(3), 147–149.

Kräuchi, K. and Wirz-Justice, A. [1994], 'Circadian rhythm of heat production, heart rate, and skin and core temperature under unmasking conditions in men', *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **267**(3), R819–R829.

Krishnamurthy, V. and Chung, S.-H. [2007], Signal processing based on hidden Markov models for extracting small channel currents, *in* 'Biological Membrane Ion Channels', Springer, pp. 623–650.

Krishnamurthy, V., Moore, J. B. and Chung, S.-H. [1993], 'Hidden Markov model signal processing in presence of unknown deterministic interferences', *IEEE Transactions on Automatic Control* **38**(1), 146–152.

Krogh, A., Brown, M., Mian, I. S., Sjölander, K. and Haussler, D. [1994], 'Hidden Markov models in computational biology: applications to protein modeling', *Journal of Molecular Biology* **235**(5), 1501–1531.

Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E. L. [2001], 'Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes', *Journal of Molecular Biology* **305**(3), 567–580.

Kruijer, W., Rousseau, J. et al. [2013], 'Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors', *Electronic Journal of Statistics* **7**, 2947–2969.

Kullback, S. and Leibler, R. A. [1951], 'On information and sufficiency', *The annals of mathematical statistics* **22**(1), 79–86.

Lal, C., Strange, C. and Bachman, D. [2012], 'Neurocognitive impairment in obstructive sleep apnea', *Chest* **141**(6), 1601–1610.

Lanfranchi, P. A., Braghiroli, A., Bosimini, E., Mazzuero, G., Colombo, R., Donner, C. F. and Giannuzzi, P. [1999], 'Prognostic value of nocturnal Cheyne-Stokes respiration in chronic heart failure', *Circulation* **99**(11), 1435–1440.

Langrock, R., Swihart, B. J., Caffo, B. S., Punjabi, N. M. and Crainiceanu, C. M. [2013], 'Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms', *Statistics in Medicine* **32**(19), 3342–3356.

Lévi, F. and Schibler, U. [2007], 'Circadian rhythms: mechanisms and therapeutic implications', *Annual Review of Pharmacology and Toxicology* **47**, 593–628.

Li, Z. and Krafty, R. T. [2019], 'Adaptive Bayesian time-frequency analysis of multivariate time series', *Journal of the American Statistical Association* **114**(525), 453–465.

Lindley, D. V. [1957], 'A statistical paradox', *Biometrika* **44**(1/2), 187–192.

Marin, J.-M., Mengersen, K. and Robert, C. P. [2005], 'Bayesian modelling and inference on mixtures of distributions', *Handbook of statistics* **25**, 459–507.

Marin, J.-M. and Robert, C. [2007], *Bayesian core: a practical approach to computational Bayesian statistics*, Berlin: Springer Science & Business Media.

McCullagh, P. and Nelder, J. [1989], *Generalized Linear Models, Second Edition*, Monographs on Statistics and Applied Probability Series, London: Chapman & Hall.

Nakamura, A., Fukuda, Y. and Kuwaki, T. [2003], 'Sleep apnea and effect of chemostimulation on breathing instability in mice', *Journal of Applied Physiology* **94**(2), 525–532.

Neal, R. M. et al. [2003], 'Slice sampling', *The annals of statistics* **31**(3), 705–767.

Nieto, F. J., Peppard, P. E., Young, T., Finn, L., Hla, K. M. and Farré, R. [2012], 'Sleep-disordered breathing and cancer mortality: results from the Wisconsin sleep cohort study', *American Journal of Respiratory and Critical Care Medicine* **186**(2), 190–194.

Ombao, H. C., Raz, J. A., von Sachs, R. and Malow, B. A. [2001], 'Automatic statistical analysis of bivariate nonstationary time series', *Journal of the American Statistical Association* **96**(454), 543–560.

Osorio, R. S., Ducca, E. L., Wohlleber, M. E., Tanzi, E. B., Gumb, T., Twumasi, A., Tweardy, S., Lewis, C., Fischer, E., Koushyk, V. et al. [2016], 'Orexin-A is associated with increases in cerebrospinal fluid phosphorylated-tau in cognitively normal elderly subjects', *Sleep* **39**(6), 1253–1260.

Pace-Schott, E. F. and Hobson, J. A. [2002], 'The neurobiology of sleep: genetics, cellular physiology and subcortical networks', *Nature Reviews Neuroscience* **3**(8), 591.

Papastamoulis, P. [2016], 'label.switching: An R package for dealing with the label switching problem in MCMC outputs', *Journal of Statistical Software, Code Snippets* **69**(1), 1–24.

Papastamoulis, P. and Iliopoulos, G. [2010], 'An artificial allocations based solution to the label switching problem in Bayesian analysis of mixtures of distributions', *Journal of Computational and Graphical Statistics* **19**(2), 313–331.

Parish, J. M., Adam, T. and Facchiano, L. [2007], 'Relationship of metabolic syndrome and obstructive sleep apnea', *Journal of Clinical Sleep Medicine* **3**(5), 467.

Parzen, E. [1961], 'Mathematical considerations in the estimation of spectra', *Technometrics* **3**(2), 167–190.

Parzen, E. [1974], 'Some recent advances in time series modeling', *IEEE Transactions on Automatic Control* **19**(6), 723–730.

Peker, Y., Hedner, J., Norum, J., Kraiczi, H. and Carlson, J. [2002], 'Increased incidence of cardiovascular disease in middle-aged men with obstructive sleep apnea: a 7-year follow-up', *American Journal of Respiratory and Critical Care Medicine* **166**(2), 159–165.

Percival, D. B., Walden, A. T. et al. [1993], *Spectral Analysis for Physical Applications*, Cambridge University Press.

Perman, M., Pitman, J. and Yor, M. [1992], 'Size-biased sampling of Poisson point processes and excursions', *Probability Theory and Related Fields* **92**(1), 21–39.

Petrone, S. and Wasserman, L. [2002], 'Consistency of Bernstein polynomial posteriors', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(1), 79–100.

Pitman, J. [1996], 'Blackwell-Macqueen urn scheme', *Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell* **30**, 245.

Pitman, J. [2002], 'Poisson–Dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition', *Combinatorics, Probability and Computing* **11**(5), 501–514.

Priestley, M. B. [1965], 'Evolutionary spectra and non-stationary processes', *Journal of the Royal Statistical Society: Series B (Methodological)* **27**(2), 204–229.

Priestley, M. B. [1981], *Spectral Analysis and Time Series*, London: Academic Press.

Quinn, B. G. [1989], 'Estimating the number of terms in a sinusoidal regression', *Journal of Time Series Analysis* **10**(1), 71–75.

Rabiner, L. R. [1989], 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proceedings of the IEEE* **77**(2), 257–286.

Rabiner, L. R., Wilpon, J. G. and Soong, F. K. [1989], 'High performance connected digit recognition using hidden Markov models', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **37**(8), 1214–1225.

Rasmussen, C. E. and Ghahramani, Z. [2002], Infinite mixtures of Gaussian process experts, *in* 'Advances in Neural Information Processing Systems', pp. 881–888.

Raviv, J. [1967], 'Decision making in Markov chains applied to the problem of pattern recognition', *IEEE Transactions on Information Theory* **13**(4), 536–551.

Redner, R. A. and Walker, H. F. [1984], 'Mixture densities, maximum likelihood and the EM algorithm', *SIAM review* **26**(2), 195–239.

Richardson, S. and Green, P. J. [1997], 'On Bayesian analysis of mixtures with an unknown number of components (with discussion)', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**(4), 731–792.

Rife, D. C. and Boorstyn, R. R. [1976], 'Multiple tone parameter estimation from discrete-time observations', *Bell System Technical Journal* **55**(9), 1389–1410.

Rissanen, J. [1978], 'Modeling by shortest data description', *Automatica* **14**(5), 465–471.

Robert, C. [2007], *The Bayesian choice: from decision-theoretic foundations to computational implementation*, Springer Science & Business Media.

Robert, C. and Casella, G. [2013], *Monte Carlo statistical methods*, Springer Science & Business Media.

Robinson, P. M. [1991], 'Nonparametric function estimation for long memory time series', *Nonparametric and semiparametric methods in econometrics and statistics* pp. 437–457.

Rodriguez, C. E. and Walker, S. G. [2014], 'Label switching in bayesian mixture models: Deterministic relabeling strategies', *Journal of Computational and Graphical Statistics* **23**(1), 25–45.

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.

Rousseau, J., Chopin, N., Liseo, B. et al. [2012], 'Bayesian nonparametric estimation of the spectral density of a long or intermediate memory Gaussian process', *The Annals of Statistics* **40**(2), 964–995.

Ruehland, W. R., Rochford, P. D., ODonoghue, F. J., Pierce, R. J., Singh, P. and Thornton, A. T. [2009], 'The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index', *Sleep* **32**(2), 150–157.

Sack, R. L., Auckley, D., Auger, R. R., Carskadon, M. A., Wright Jr, K. P., Vitiello, M. V. and Zhdanova, I. V. [2007], 'Circadian rhythm sleep disorders: part ii, advanced sleep phase disorder, delayed sleep phase disorder, free-running disorder, and irregular sleep-wake rhythm', *Sleep* **30**(11), 1484–1501.

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

Sethuraman, J. [1994], 'A constructive definition of Dirichlet priors', *Statistica Sinica* pp. 639–650.

Shneerson, J. M. [2009], *Sleep Medicine: a Guide to Sleep and its Disorders*, Hoboken: John Wiley & Sons.

Shumway, R. H. and Stoffer, D. S. [2005], *Time Series Analysis and Its Applications (Springer Texts in Statistics)*, New York: Springer-Verlag.

Shumway, R. H. and Stoffer, D. S. [2017], *Time series analysis and its applications: with R examples*, Springer.

Stephens, M. [2000], 'Dealing with label switching in mixture models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(4), 795–809.

Stoica, P., Moses, R. L., Friedlander, B. and Soderstrom, T. [1989], 'Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **37**(3), 378–392.

Sundaram, S. S., Halbower, A., Pan, Z., Robbins, K., Capocelli, K. E., Klawitter, J., Shearn, C. T. and Sokol, R. J. [2016], 'Nocturnal hypoxia-induced oxidative stress promotes progression of pediatric non-alcoholic fatty liver disease', *Journal of Hepatology* **65**(3), 560–569.

Tan, A., Cheung, Y. Y., Yin, J., Lim, W.-Y., Tan, L. W. and Lee, C.-H. [2016], 'Prevalence of sleep-disordered breathing in a multiethnic Asian population in Singapore: A community-based study', *Respirology* **21**(5), 943–950.

Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M. [2006], 'Hierarchical Dirichlet Processes', *Journal of the American Statistical Association* **101**(476), 1566–1581.

Teran-Santos, J., Jimenez-Gomez, A., Cordero-Guevara, J. and Burgos-Santander, C. G. [1999], 'The association between sleep apnea and the risk of traffic accidents', *New England Journal of Medicine* **340**(11), 847–851.

Tripuraneni, N., Gu, S. S., Ge, H. and Ghahramani, Z. [2015], Particle gibbs for infinite hidden markov models, *in* 'Advances in Neural Information Processing Systems', pp. 2395–2403.

Van Gael, J., Saatci, Y., Teh, Y. W. and Ghahramani, Z. [2008], Beam sampling for the infinite hidden markov model, *in* 'Proceedings of the 25th international conference on Machine learning', ACM, pp. 1088–1095.

Van Someren, E. J. [2006], 'Mechanisms and functions of coupling between sleep and temperature rhythms', *Progress in Brain Research* **153**, 309–324.

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

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

White, G. C. and Bennetts, R. E. [1996], 'Analysis of frequency count data using the negative binomial distribution', *Ecology* **77**(8), 2549–2557.

Whittle, P. [1953], 'Estimation and information in stationary time series', *Arkiv för Matematik* **2**(5), 423–434.

Whittle, P. [1957], 'Curve and periodogram smoothing', *Journal of the Royal Statistical Society. Series B (Methodological)* **19**(1), 38–63.

Yaggi, H. K., Concato, J., Kernan, W. N., Lichtman, J. H., Brass, L. M. and Mohsenin, V. [2005], 'Obstructive sleep apnea as a risk factor for stroke and death', *New England Journal of Medicine* **353**(19), 2034–2041.

Yaghouby, F. and Sunderam, S. [2015], 'Quasi-supervised scoring of human sleep in polysomnograms using augmented input variables', *Computers in Biology and Medicine* **59**, 54–63.

Yau, S.-F. and Bresler, Y. [1993], 'Maximum likelihood parameter estimation of superimposed signals by dynamic programming', *IEEE Transactions on Signal Processing* **41**(2), 804–820.

Young, T., Peppard, P. E. and Gottlieb, D. J. [2002], 'Epidemiology of obstructive sleep apnea: a population health perspective', *American Journal of Respiratory and Critical Care Medicine* **165**(9), 1217–1239.

Zhang, Q. and Wong, K. M. [1993], 'Information theoretic criteria for the determination of the number of signals in spatially correlated noise', *IEEE Transactions on Signal Processing* **41**(4), 1652–1663.

Appendix A

Dirichlet Process and Hierarchical Dirichlet Process

A.1 Dirichlet Process

The Dirichlet process (DP) is a stochastic process used in Bayesian nonparametric models. Informally, it can be seen as a distribution over probability distributions, meaning that a realization from a DP is itself a probability distribution.

Dirichlet Process: A formal definition of DP is due to Ferguson [1973], who specified a DP via a base distribution H with support Θ and a positive concentration parameter γ. Particularly, given a finite partition $\{A_1, \ldots, A_K\}$ of Θ, G_0 is a DP with base distribution H and concentration parameter γ, denoted by $G_0 \sim \mathrm{DP}(\gamma, H)$, if

$$(G_0(A_1), \ldots, G_0(A_K)) \sim \mathrm{Dirichlet}(\gamma H(A_1), \ldots, \gamma H(A_K)). \tag{A.1}$$

That is, the marginal distribution of G_0 on every finite partition of Θ follows a Dirichlet distribution. The base distribution H can be seen as the mean of the DP,

as $\mathbb{E}[G(A)] = H(A)$ for any measurable set $A \subseteq \Theta$, whereas the concentration parameter γ can be interpreted as an inverse variance of the concentration of mass around the base distribution H, namely $\mathrm{Var}\,[G(A)] = H(A)(1 - H(A))/(\gamma + 1)$.

Stick-Breaking Construction: Another representation of a DP was given by Sethuraman [1994]. Let $\boldsymbol{\alpha} = (\alpha_k)_{k=1}^{\infty}$ be a probability mass function on a countably infinite set, where the discrete probabilities are defined as follows

$$\begin{aligned} \nu_k \,|\, \gamma &\sim \mathrm{Beta}\,(1,\, \gamma), \quad k = 1, 2, \ldots \\ \alpha_k &= \nu_k \prod_{l=1}^{k-1} (1 - \nu_l), \quad k = 1, 2, \ldots \end{aligned} \tag{A.2}$$

This *stick-breaking* construction is commonly denoted by $\boldsymbol{\alpha} \sim \mathrm{GEM}(\gamma)$ (GEM is an abbreviation for Griffiths, Engen and McCloskey [Pitman, 2002]). Here, γ is a positive real number that controls the expected value of the number of elements in $\boldsymbol{\alpha}$ with significant probability mass. In fact, Equation (A.2) can be interpreted in the following way: begin with a stick of unit length and split it at location ν_1; then, assign α_1 to be the length of the remaining stick that has been split and recursively, break the other portion to obtain α_2, α_3 and so on. Notice that this construction ensures that the sequence $\boldsymbol{\alpha}$ satisfies $\sum_{k=1}^{\infty} \alpha_k = 1$.

Sethuraman [1994] showed that, conditional on the probability mass function $\boldsymbol{\alpha}$ constructed as in Equation (A.2), the random distribution G_0

$$G_0 = \sum_{k=1}^{\infty} \alpha_k \delta_{\theta_k}, \qquad \theta_k \,|\, H \sim H, \qquad k = 1, 2, \ldots \tag{A.3}$$

is a draw from a $\mathrm{DP}\,(\gamma, H)$, θ_k are distributed according to H, and δ_θ denote a unit-mass measure concentrated at θ. This representation shows that measures drawn

from a DP are always discrete.

Example: Figure A.1 displays draws from $G_0 \sim \text{DP}(\gamma, H)$ for several values of γ, where the base distribution H is $\mathcal{N}(0, \sqrt{10})$. It becomes clear that the smaller γ, the fewer elements in $\boldsymbol{\alpha}$ have a significant probability mass, resulting in draws from a DP to be concentrated on a fewer atoms. Conversely, increasing the magnitude of γ leads to draws from a DP spread over a larger number of atoms. Furthermore, Figure A.1 shows how the parameter γ influences the concentration of

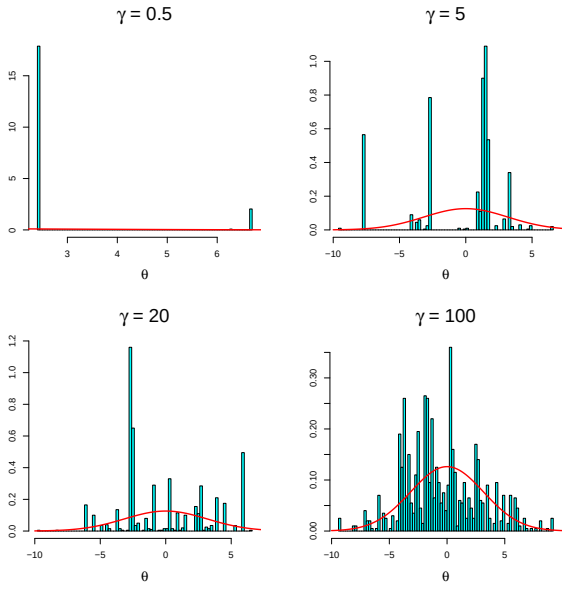

Figure A.1: Example of a DP represented via stick-breaking construction. Histograms of draws from $G_0 \sim \text{DP}(\gamma, H)$ for $\gamma = 0.5, 5, 20$ and 100, where $H \sim \mathcal{N}(0, \sqrt{10})$. The red line is the density of H.

mass around the base distribution H: the greater γ, the closer the draw from the DP appears to be to the base measure H. In fact, as $\gamma \to \infty$, we have that $G(A) \to H(A)$ for any measurable set $A \in \Theta$

Polya Urn Representation: We now consider analyzing the distribution of draws $\theta_i' \sim G_0$, without explicitly examining G_0 directly. In particular, let $\theta_1', \theta_2', \ldots$ be a sequence of independent and identically distributed random variables

distributed according to G_0. Blackwell et al. [1973] showed that the conditional distributions $\theta_i^{'} \,|\, \theta_1^{'}, \ldots, \theta_{i-1}^{'}$ when G_0 is marginalized out is given by

$$\theta_i^{'} \,|\, \theta_1^{'}, \ldots, \theta_{i-1}^{'} \sim \sum_{l=1}^{i-1} \frac{1}{i-1+\gamma} \delta_{\theta_l^{'}} + \frac{\gamma}{i-1+\gamma} H \tag{A.4}$$

This representation, commonly referred to as the *Pólya urn* scheme, can be interpreted with a metaphor in which balls are drawn from an urn model. Here, a ball of a unique color is associated with each atom and these balls are drawn at random. With probability proportional to γ, we create a new atom by drawing from H, and a ball of this new color is placed into the urn. Otherwise, we draw a ball at random from the urn and we put it back together with a further ball having the same color.

Because samples from a DP are always discrete, there is a positive probability that different observations $\theta_i^{'}$ assume identical values from the set of K unique values $\{\theta_1, \ldots, \theta_K\}$, where θ_k is defined as in Equation (A.3). Hence, we can re-write Equation (A.4) as

$$\theta_i^{'} \,|\, \theta_1^{'}, \ldots, \theta_{i-1}^{'} \sim \sum_{k=1}^{K} \frac{N_k}{i-1+\gamma} \delta_{\theta_k} + \frac{\gamma}{i-1+\gamma} H \tag{A.5}$$

where N_k is the number of observations $\theta_{i'}^{'}$ assuming value θ_k for $1 \leq i' < i$.

Chinese Restaurant Process: Another perspective on the DP that highlights its clustering properties is obtained by analyzing the predictive distribution of draws $\theta_1^{'}, \ldots, \theta_N^{'} \sim G_0$. Let z_i be a discrete random variable over the positive integers that selects the unique value θ_k such that $\theta_i^{'} = \theta_{z_i}$ and let $N_k = \sum_{i=1}^{N} \delta(z_i, k)$ be the number of indicator random variables assuming the value of k. Then, Equation (A.5) yields to the following predictive distribution on the indicator random

variables $z_{N+1} \,|\, z_1, \ldots, z_N$

$$\pi\,(\,z_{N+1} = z \,|\, z_1, \ldots, z_N,\, \gamma\,) = \sum_{k=1}^{K} \frac{N_k}{N+\gamma}\,\delta\,(z,\, k) + \frac{\gamma}{N+\gamma}\,\delta\,(z,\, K+1), \tag{A.6}$$

where $\delta\,(z,\, k) = 1$ if $k = z$ and zero otherwise and $K + 1$ is a value that has not previously been seen. This representation, which is commonly referred to as the *Chinese restaurant process*, induces a distribution over partitions of $\{1, \ldots, N\}$. The analogy is as follows. Consider a Chinese restaurant with infinitely many tables, where each table can sit infinitely many costumers and each table is served a unique dish θ_k. A costumer $\theta_i^{'}$ enters the restaurant and decides whether to join a currently occupied table k, with probability proportional to the amount of people sitting at that table, or sit at a previously unoccupied table $K + 1$, with probability proportional to γ.

A.2 Hierarchical Dirichlet Process

Hierarchical approaches are naturally applied in Bayesian nonparametric statistics for problems that involves modelling several groups of data [Hjort et al., 2010]. For example, when the model for each group of data includes a discrete variable of unknown cardinality, and we would like to tie these variables across groups, i.e allow these groups to be *statistically* linked with each other, the hierarchical Dirichlet process introduced by Teh et al. [2006] can be particularly well-suited for the task.

Hierarchical Dirichlet Process: The hierarchical Dirichlet process (HDP) defines a set of *group-specific* probability distributions G_j - where we assume there

are J of such groups - and a *global* probability distribution G_0 as follows

$$\begin{aligned} G_0 \,|\, \gamma,\, H \;&\sim\; \mathrm{DP}\,(\,\gamma,\, H\,) \\ G_j \,|\, \eta,\, G_0 \;&\sim\; \mathrm{DP}\,(\,\eta,\, G_0\,), \qquad j = 1, \ldots, J. \end{aligned} \tag{A.7}$$

The global distribution G_0 is a DP with base distribution H and strength parameter γ, the group-specific probability distributions G_j are DP distributed with base distribution G_0 and concentration parameter η, and the probability distributions G_j are conditionally independent given G_0. Hence, the HDP is based on a recursion where the base distribution G_0, i.e the base distribution for the group-specific DPs G_j, is itself a DP. The locations of the atoms for the group-specific probability distributions G_j are discrete and drawn from the support of G_0. These atoms are also shared among different groups, as each G_j inherits atoms from the same base distribution G_0 and $\mathbb{E}[G_j \,|\, G_0] \to G_0$.

Stick-Breaking Construction: A construction that provide a better intuition about the sharing of atoms across multiple DPs in a HDP, is the stick-breaking representation

$$G_0 = \sum_{k=1}^{\infty} \alpha_k \, \delta_{\theta_k} \qquad \begin{aligned} \boldsymbol{\alpha} \,\big|\, \gamma &\sim \mathrm{GEM}\,(\,\gamma\,) \\ \theta_k \,\big|\, H &\sim H, \quad k = 1, 2, \ldots \end{aligned} \tag{A.8}$$

$$G_j = \sum_{t=1}^{\infty} \tilde{\pi}_{jt} \, \delta_{\theta^*_{jt}} \qquad \begin{aligned} \tilde{\boldsymbol{\pi}}_j \,\big|\, \eta &\sim \mathrm{GEM}\,(\,\eta\,) \quad j = 1, \ldots, J \\ \theta^*_{jt} \,\big|\, G_0 &\sim G_0 \quad t = 1, 2, \ldots \end{aligned} \tag{A.9}$$

This formulation shows how the groups-specific DPs G_j are constructed as a reweighted sum of the atoms from the global DP G_0. That is, the group-specific set of atoms θ^*_{jt} are drawn from the global set of atoms θ_k of G_0, ensuring that there exists non-zero probability that distinct G_j share atoms between each other. Figure (A.2) displays

an example of an HDP with $J = 3$ groups, represented via stick-breaking construction. In this case, $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\pi}}_j$ can be interpreted as probability distributions over the positive integers, namely $\boldsymbol{\alpha} \,|\, \gamma \sim \text{GEM}(\gamma)$ and $\tilde{\boldsymbol{\pi}}_j \,|\, \eta, \boldsymbol{\alpha} \sim \text{DP}(\eta, \boldsymbol{\alpha})$.

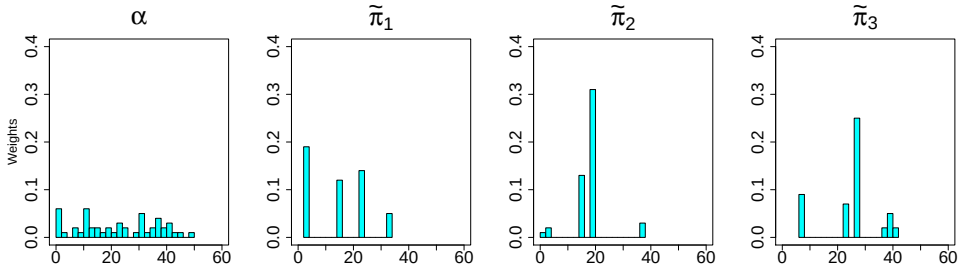

Figure A.2: Example of a HDP with $J = 3$ groups, represented via stick-breaking construction. Left panel shows a draw of the global distribution $\boldsymbol{\alpha}$, whereas the other panels displays draws of the group-specific distributions $\tilde{\boldsymbol{\pi}}_1$, $\tilde{\boldsymbol{\pi}}_2$ and $\tilde{\boldsymbol{\pi}}_3$, conditional on $\boldsymbol{\alpha}$.

Chinese Restaurant Franchise: A related representation of the Chinese restaurant process is the *Chinese restaurant franchise* [Teh et al., 2006] which describes the marginal probabilities obtained by marginalizing out G_0 and G_j in the HDP. In a Chinese restaurant franchise the analogy of a Chinese restaurant process is extended to a set of restaurants, one for each of the group of the HDP. An infinite *global* menu of dishes is shared across the set of restaurants. The process of seating costumers at tables happens in a similar way as for the Chinese restaurant process, but is restaurant specific. The process of choosing dishes at a specific table happens franchise-wide, namely the dishes are selected with probability proportional to the number of tables (in the *entire* franchise) that have previously served that dish.

More formally, define $\theta'_{j1}, \ldots, \theta'_{jN_j}$ to be the set of costumers in restaurant j, where each θ'_{ji} is distributed according to G_j. Each costumer is pre-allocated to a specific restaurant designated by that costumer's group j. Let $\theta^*_{j1}, \theta^*_{j2}, \ldots$ denote the dish served at table t in restaurant j, where θ^*_{jt} is distributed according to G_0. Finally, let $\theta_1, \theta_2, \ldots$ be the global menu of dishes of the franchise, where θ_k is distributed

according to the base measure H. We augment the space and introduce indicator random variables t_{ji} and k_{jt}, such that t_{ji} indicates the table assignment for the i^{th} costumer in the j^{th} restaurant, whereas k_{jt} represents the dish assignment of the t^{th} table in the j^{th} restaurant. The i^{th} costumer in the j^{th} restaurant chooses a table via $t_{ji} \sim \tilde{\boldsymbol{\pi}}_j$, and at each table is assigned a dish via $k_{jt} \sim \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ can be seen as a collection of ratings for the dishes served in the global menu. Additionally, the following equality holds: $\theta'_{ji} = \theta^*_{jt_{ji}} = \theta_{k_{jt_{ji}}}$. Figure A.3 displays an an example of a Chinese restaurant franchise with $J = 2$ restaurants.

The predictive distributions that outline the Chinese restaurant franchise are given as follows

$$p\left(t_{ji} \mid t_{j1}, \ldots, t_{ji-1}, \eta\right) \propto \sum_{t=1}^{T_j} \tilde{n}_{jt}\, \delta(t_{ji}, t) + \eta\, \delta\left(t_{ji}, T_j + 1\right)$$
$$p(k_{jt} \mid \boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_{j-1}, k_{j1}, \ldots, k_{jt-1}, \gamma) \propto \sum_{k=1}^{K} m_{\cdot k}\, \delta\left(k_{jt}, k\right) + \gamma\, \delta\left(k_{jt}, K + 1\right), \tag{A.10}$$

where $\tilde{n}_{jt}$ is the number of costumers in the j^{th} restaurant sitting at table t, m_{jk} is the number of tables in the j^{th} restaurant serving dish k, T_j is the total number of occupied tables in restaurant j, and K is the total number of unique dishes served in the franchise. Marginal counts are described with dots. Thus, $m_{\cdot k} = \sum_{j=1}^{J}$ is the number of tables in the whole franchise serving dish k. Also we denote with $\boldsymbol{k}_j = (k_{j1}, \ldots, k_{jT_j})$ the dish assignments for the tables of the j^{th} restaurant.

Equation (A.10) dictates that the i^{th} costumer entering the j^{th} restaurant sits in one among the T_j currently occupied tables, say t, with probability proportional to the number of currently seated costumers $\tilde{n}_{jt}$, or start a new table $T_j + 1$ with probability proportional to η. The first costumer to sit at a table proceeds to the

buffet line and chooses a dish based on the popularity of that dish in the whole franchise, namely he chooses an already selected dish $\theta_{k_{jt}}$ with probability proportional to $m_{\cdot k}$, or order a new dish θ_{K+1} with probability proportional to γ.

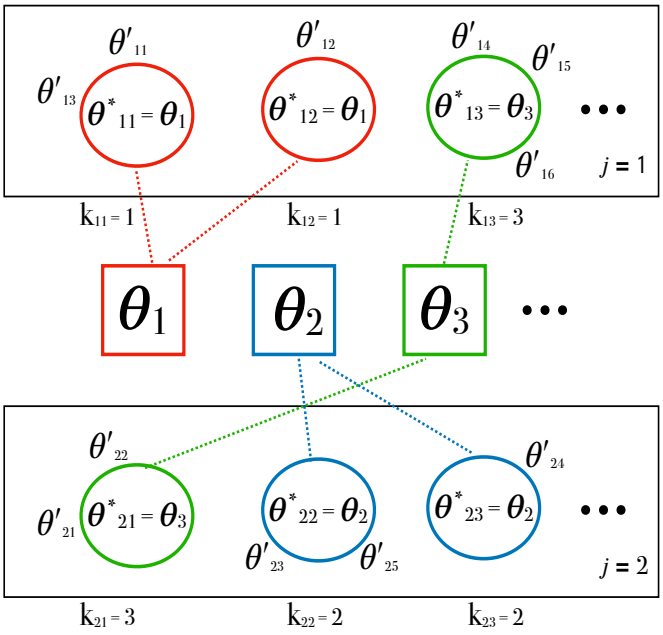

Figure A.3: Chinese restaurant franchise including $J = 2$ restaurants (rectangles). Costumers (θ'_{ji}) are seated at tables (circles) in the restaurants, where at each table a dish (θ^*_{jt}) is chosen from a global menu (θ_k). Notice that in a restaurant multiple tables may serve the same dish. This picture is best viewed in colours.

As depicted in Figure A.3, the Chinese restaurant franchise allows multiple tables in the same restaurant to serve the same dish. The restaurant-specific DPs can be instead re-written in terms of the unique dishes in the following way [Fox et al., 2011]

$$G_j = \sum_{k=1}^{\infty} \pi_{jk}\, \delta_{\theta_k} \qquad \begin{aligned} \theta_k \,|\, H &\sim H && k = 1, 2, \ldots \\ \boldsymbol{\pi}_j \,|\, \eta, \boldsymbol{\alpha} &\sim \mathrm{DP}(\eta, \boldsymbol{\alpha}) && j = 1, \ldots, J \end{aligned} \tag{A.11}$$

where

$$\pi_{jk} = \sum_{t:k_{jt}=k} \tilde{\pi}_{jt}.$$

According to this representation, $\boldsymbol{\pi}_j$ now defines a probability distributions over unique dishes rather than tables.