

# Beyond Whittle: Nonparametric Correction of a Parametric Likelihood with a Focus on Bayesian Time Series Analysis\*

Claudia Kirch<sup>†,‡</sup>, Matthew C. Edwards<sup>§</sup>, Alexander Meier<sup>¶</sup>, and Renate Meyer<sup>||</sup>

**Abstract.** Nonparametric Bayesian inference has seen a rapid growth over the last decade but only few nonparametric Bayesian approaches to time series analysis have been developed. Most existing approaches use Whittle’s likelihood for Bayesian modelling of the spectral density as the main nonparametric characteristic of stationary time series. It is known that the loss of efficiency using Whittle’s likelihood can be substantial. On the other hand, parametric methods are more powerful than nonparametric methods if the observed time series is close to the considered model class but fail if the model is misspecified. Therefore, we suggest a nonparametric correction of a parametric likelihood that takes advantage of the efficiency of parametric models while mitigating sensitivities through a nonparametric amendment. We use a nonparametric Bernstein polynomial prior on the spectral density with weights induced by a Dirichlet process and prove posterior consistency for Gaussian stationary time series. Bayesian posterior computations are implemented via an MH-within-Gibbs sampler and the performance of the nonparametrically corrected likelihood for Gaussian time series is illustrated in a simulation study and in three astronomy applications, including estimating the spectral density of gravitational wave data from the Advanced Laser Interferometer Gravitational-wave Observatory (LIGO).

**Keywords:** Bayesian nonparametrics, frequency domain, laser interferometric gravitational wave data, spectral density, stationary time series.

## 1 Introduction

Statistical models can be broadly classified into *parametric* and *nonparametric* models. Parametric models, indexed by a finite dimensional set of parameters, are focused, easy

---

\*A previous version of the Supplementary material contained proofs of Lemma 1.2 and Theorem 2 that relied on an incorrect deduction. The current version amends the proofs. In addition, an additional assumption for the validity of Lemma 1.2 has been included to require that  $S_n$  are symmetric positive definite matrices with uniformly bounded eigenvalues  $\mu_i(S_n)$ , i.e. there exist constants  $m$  and  $M$  such that  $0 < m \leq \mu_i(S_n) \leq M < \infty$  for  $i = 1, \dots, n$  and all  $n$ . The statement of Theorem 2 in the main text remains unchanged.

<sup>†</sup>Institute of Mathematical Stochastics, Department of Mathematics, Otto-von-Guericke University, Magdeburg, Germany

<sup>‡</sup>Center for Behavioral Brain Sciences (CBBS), Magdeburg, Germany, [claudia.kirch@ovgu.de](mailto:claudia.kirch@ovgu.de)

<sup>§</sup>School of Mathematics, The University of Edinburgh, Edinburgh, United Kingdom, [Matthew.Edwards@ed.ac.uk](mailto:Matthew.Edwards@ed.ac.uk)

<sup>¶</sup>Institute of Mathematical Stochastics, Department of Mathematics, Otto-von-Guericke University, Magdeburg, Germany, [alexander.meier@ovgu.de](mailto:alexander.meier@ovgu.de)

<sup>||</sup>Department of Statistics, The University of Auckland, Auckland, New Zealand, [renate.meyer@auckland.ac.nz](mailto:renate.meyer@auckland.ac.nz)

to analyse and have the big advantage that when correctly specified, they will be very efficient and powerful. However, they can be sensitive to misspecifications and even mild deviations of the data from the assumed parametric model can lead to unreliabilities of inference procedures. Nonparametric models, on the other hand, do not rely on data belonging to any particular family of distributions. As they make fewer assumptions, their applicability is much wider than that of corresponding parametric methods. However, this generally comes at the cost of reduced efficiency compared to parametric models.

Standard time series literature is dominated by parametric models like autoregressive integrated moving average models (Box et al., 2013), autoregressive conditional heteroskedasticity models for time-varying volatility (Engle, 1982; Bollerslev, 1986), state-space (Durbin and Koopman, 2012) and Markov switching models (Bauwens et al., 2000). In particular, Bayesian time series analysis (Steel, 2008) is inherently parametric in that a completely specified likelihood function is needed. Nonetheless, the use of nonparametric techniques has a long tradition in time series analysis. Schuster (1898) introduced the periodogram which may be regarded as the origin of spectral analysis and a classical nonparametric tool for time series analysis (Härdle et al., 1997). An important class of nonparametric methods is based on frequency domain techniques, in particular those based on smoothing the periodogram.

*Frequentist* time series analyses especially use nonparametric methods (Fan and Yao, 2002; Wasserman, 2006) including a variety of bootstrap methods, computer-intensive resampling techniques initially introduced for independent data, that have been tailored to and specifically developed for time series (Härdle et al., 2003; Kreiss and Lahiri, 2012; Kreiss and Paparoditis, 2011). These include a variety of frequency domain bootstrap methods based on the periodogram strongly related to the Whittle likelihood (Hurvich and Zeger, 1987; Franke and Härdle, 1992; Kirch and Politis, 2011; Kirch, 2007; Kim and Nordman, 2013) and found important applications in a variety of disciplines (Hidalgo, 2008; Costa et al., 2013; Emmanoulopoulos et al., 2013).

Despite the fact that *nonparametric Bayesian* inference has been rapidly expanding over the last decade, as reviewed by Hjort et al. (2010), Müller and Mitra (2013), and Walker (2013), only very few nonparametric Bayesian approaches to time series analysis have been developed. With the exception of Rousseau et al. (2012), who use the full Gaussian likelihood, most papers use the Whittle likelihood (Whittle, 1957) for Bayesian modelling of the spectral density as the main nonparametric characteristic of stationary time series. Examples include the papers by Carter and Kohn (1997), Gangopadhyay et al. (1999), Liseo et al. (2001), Choudhuri et al. (2004), Hermansen (2008), Chopin et al. (2013).

The Whittle likelihood is an approximation of the true likelihood, which is exact only for Gaussian white noise. While in a Gaussian setting, the advantage of the Whittle likelihood is mainly of computational nature both in terms of speed and numerical errors, for non-Gaussian time series the full likelihood is no longer accessible, while Whittle's likelihood still often leads to asymptotically correct statistical inference in many situations. As shown in Contreras-Cristán et al. (2006), the loss of efficiency using the Whittle likelihood can be substantial in the Gaussian AR-model for small samples if the autocorrelation of the Gaussian process is high. For this reason in the Gaussian small sample situation, Chopin et al. (2013) use a second importance sampling

step to correct a posterior sample obtained in an initial step based on the Whittle likelihood. This importance sampling step is necessary only for time series of lengths  $< 10^3$ , while for lengths  $\gg 10^4$  the importance sampling correction step becomes quickly too expensive. A different approach based on de-biasing the Whittle likelihood has been suggested in Sykulski et al. (2017) for second-order stationary stochastic processes with the same computational complexity as the original Whittle likelihood. Cadonna et al. (2017) develop a novel Bayesian approach to spectral density estimation for multiple time series by replacing the Whittle likelihood with a structured mixture of Gaussian distributions, defined through frequency-dependent mixture weights and Gaussian mean functions.

The Whittle likelihood has two potential sources of approximation errors: The first one is the assumption of independence between Fourier coefficients which holds only asymptotically but not exactly for a finite time series and the second one is the Gaussianity assumption. In this paper, we restrict our attention to the first problem. Extending the proposed methods to non-Gaussian situations will be a focus of future work. In order to introduce some reasonable dependence between Fourier coefficients Kreiss and Paparoditis (2003) propose a semi-parametric approach for bootstrapping that we adopt in this work for Bayesian inference.

Parametric methods are more powerful than nonparametric methods if the observed time series is close to the considered model class but fail if the model is misspecified. To exploit the advantages of both parametric and nonparametric approaches, two conceptually different lines of research have been pursued in the frequentist literature. The first approach consists of pre-whitening by fitting a parametric model first, followed by a nonparametric kernel density estimate based on the residual time series and subsequent recolouring (Andrews and Monahan (1992)). The second approach uses a nonparametric correction of a parametric likelihood. This idea was first proposed by Hjort and Glad (1995) for i.i.d. observations and by Kreiss and Paparoditis (2003) in terms of the autoregressive-aided periodogram bootstrap within the frequentist bootstrap world of time series analysis. It fits a parametric working model to generate periodogram ordinates that mimic the essential features of the data and the weak dependence structure of the periodogram while a nonparametric correction is used to capture features not represented by the parametric fit. This has been extended in various ways (see Jentsch and Kreiss (2010); Jentsch et al. (2012); Kreiss and Paparoditis (2012)). Its main underlying idea is a nonparametric correction of a parametric likelihood approximation. The parametric model is used as a proxy for rough shape of the autocorrelation structure as well as the dependency structure between periodogram ordinates. Sensitivities with respect to the spectral density are mitigated through a nonparametric amendment. We propose to use a similar nonparametrically corrected likelihood approximation as a pseudo-likelihood in the Bayesian framework to compute the pseudo-posterior distribution of the power spectral density (PSD) and other parameters in time series models. This will yield a pseudo-likelihood that generalises the widely used Whittle likelihood which, as we will show, can be regarded as a special case of a nonparametrically corrected likelihood under the Gaussian i.i.d. working model. Software implementing the methodology is available in the R package `beyondWhittle`, which is available on the Comprehensive R Archive Network (CRAN), see Meier et al. (2017). In this paper, we will focus and illustrate the methodology on *zero mean* stationary time series which

in more general contexts such as regression models with time series errors, time series change point models, and unit root tests will represent the error term. Our proposed Bayesian semiparametric estimation approach can be readily extended to these models.

The main contribution of this paper is a novel semiparametric Bayesian estimator of the spectral density of a stationary time series via a nonparametric correction of a parametric likelihood. The correction adds robustness while taking advantage of the efficiency of the parametric model. The nonparametrically corrected likelihood provides a generalization of the Whittle likelihood which is shown to be a special case when the parametric working model is i.i.d. Gaussian white noise. Making use of a parametric model other than i.i.d. Gaussian white noise has the advantage that sharp peaks are better estimated than with the Whittle likelihood. The new method facilitates the estimation of the spectral density by a reparametrization. It puts the Bernstein-Dirichlet prior not on the spectral density itself but on a pre-whitened version which is in general smoother and thus easier to approximate by Bernstein polynomials. Furthermore, a parameter  $\eta \in [0, 1]$  that quantifies the degree of confidence in the parametric model is introduced and treated as an unknown parameter to be estimated from the data. It is shown in simulations to have ‘oracle’ properties in the sense of aligning the semiparametric spectral density estimate to the parametric model in the well-specified case and to the nonparametric estimate under model misspecification. Last but not least, the pseudo-posterior distribution based on Bernstein polynomial prior and corrected likelihood is proven to be consistent.

The paper is structured as follows: In Section 2, we briefly revisit the Whittle likelihood and demonstrate that it is a nonparametrically corrected likelihood, namely that of a Gaussian i.i.d. working model. Then, we extend this nonparametric correction to a general parametric working model. The corresponding pseudo-likelihood turns out to be equal to the true likelihood if the parametric working model is correctly specified but also still yields asymptotically unbiased periodogram ordinates if it is not correctly specified. In Section 3, we propose a Bayesian nonparametric approach to estimating the spectral density using the pseudo-posterior distribution induced by the corrected likelihood of a fixed parametric model. We describe the Gibbs sampling implementation for sampling from the pseudo-posterior. This nonparametric approach is based on the Bernstein polynomial prior of Petrone (1999) and used to estimate the spectral density via the Whittle likelihood in Choudhuri et al. (2004). We show posterior consistency of this approach and discuss how to incorporate the parametric working model in the Bayesian inference procedure. Section 4 gives results from a simulation study, including case studies of sunspot data, *S. Carinae* variable star light intensities, and gravitational wave data from the sixth science run of the Advanced Laser Interferometric Gravitational Wave Observatory (LIGO). This is followed by discussion in Section 5, which summarises the findings and points to directions for future work. The proofs, the details about the Bayesian autoregressive sampler as well as some additional simulation results are deferred to the supplementary material (Kirch et al., 2018).

## 2 Likelihood approximation for time series

For Gaussian time series, the full likelihood is a function of the spectral density but its evaluation is computationally expensive and prone to numerical errors. To avoid

the full likelihood evaluation, Whittle (1957) defined a pseudo-likelihood, known as the Whittle likelihood, which is only an approximation of the true likelihood but computationally much easier. Chopin et al. (2013) initially sampled from a pseudo-posterior using the Whittle likelihood, then subsequently used importance sampling to obtain samples from the true posterior with a numerically efficient implementation of the full likelihood. However, in the non-Gaussian situation, the full likelihood is not accessible, but statistical procedures based on the Whittle likelihood still yield asymptotically correct results for many but not all statistics.

Another approach is to use parametric models such as AR or MA models with data-dependent order selection. These can also be considered as approximations to the true likelihood due to their universal approximation properties. For the MA model, this comes by way of the famous Wold theorem (see e.g. Paragraph 5.7 in Brockwell and Davis (2009)). Similar results also hold for AR-approximations, which gave rise to AR-sieve bootstrap methods (see Kreiss et al. (2011)) among others. However, these methods tend to produce biased results when the parametric approximation to the underlying time series is poor.

In the following subsection we will revisit the approximate likelihood proposed by Whittle (1957) before introducing a semiparametric approach which extends the Whittle likelihood in a way that combines a parametric approximation with a nonparametric correction of the parametric likelihood. In Section 4 we will see that a Bayesian procedure based on this likelihood inherits the efficiency of the parametric model when correctly specified and the robustness of the nonparametric approach when the parametric model is completely misspecified. Moreover, it even improves upon both when the parametric model is only partially correct.

## 2.1 Whittle likelihood revisited

Assume that  $\{Z_t: t = 0, 1, \dots\}$  is a real zero mean stationary time series with absolutely summable autocovariance function  $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$ . Under these assumptions the spectral density of the time series exists and is given by the Fourier transform (FT) of the autocovariance function

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) \exp(-ik\lambda), \quad 0 \leq \lambda \leq 2\pi.$$

Consequently, there is a one-to-one-correspondence between the autocovariance function and the spectral density, and estimation of the spectral density is amenable to smoothing techniques. The idea behind these smoothing techniques is the following observation, which also gives rise to the so-called Whittle approximation of the likelihood of a time series: Consider the periodogram of  $\mathbf{Z}_n = (Z_1, \dots, Z_n)^T$ ,

$$I_n(\lambda) = I_{n,\lambda}(\mathbf{Z}_n) = \frac{1}{2\pi n} \left| \sum_{t=1}^n Z_t \exp(-it\lambda) \right|^2.$$

The periodogram is given by the squared modulus of the discrete Fourier coefficients, the Fourier transformed time series evaluated at Fourier frequencies  $\lambda_j = \frac{2\pi j}{n}$ , for  $j = 0, \dots, N = \lfloor (n-1)/2 \rfloor$ . It can be obtained by the following transformation: Define

$$\mathbf{c}_j = \sqrt{2}\Re \mathbf{e}_j = \frac{1}{\sqrt{2}}(\mathbf{e}_j + \mathbf{e}_{n-j}), \quad \mathbf{s}_j = \sqrt{2}\Im \mathbf{e}_j = \frac{1}{i\sqrt{2}}(\mathbf{e}_j - \mathbf{e}_{n-j}),$$

where

$$\mathbf{e}_j = n^{-1/2}(e_j, e_j^2, \dots, e_j^n)^T, \quad e_j = \exp(-2\pi i j/n), \quad j = 0, \dots, N$$

and for  $n$  even,  $\mathbf{e}_{n/2}$  is defined analogously. Then,

$$F_n = \begin{cases} (\mathbf{e}_0, \mathbf{c}_1, \mathbf{s}_1, \dots, \mathbf{c}_N, \mathbf{s}_N, \mathbf{e}_{n/2})^T, & n \text{ even,} \\ (\mathbf{e}_0, \mathbf{c}_1, \mathbf{s}_1, \dots, \mathbf{c}_N, \mathbf{s}_N)^T, & n \text{ odd,} \end{cases}$$

is an orthonormal  $n \times n$  matrix (cf. e.g. Brockwell and Davis (2009), paragraph 10.1). Real- and imaginary parts of the discrete Fourier coefficients are collected in the vector

$$\tilde{\mathbf{Z}}_n := (\tilde{Z}_n(0), \dots, \tilde{Z}_n(n-1))^T = F_n \mathbf{Z}_n$$

and the periodogram can be written as

$$\begin{aligned} I_n(\lambda_j) &= \frac{1}{4\pi} \left( \tilde{Z}_n^2(2j) + \tilde{Z}_n^2(2j-1) \right), \quad j = 1, \dots, N, \\ I_n(\lambda_0) &= \frac{1}{2\pi} \tilde{Z}_n^2(0), \quad \text{as well as for } n \text{ even: } I_n(\lambda_{n/2}) = \frac{1}{2\pi} \tilde{Z}_n^2(n-1). \end{aligned} \quad (1)$$

It is well known that the periodograms evaluated at two different Fourier frequencies are asymptotically independent and have an asymptotic exponential distribution with mean equal to the spectral density, a statement that remains true for non-Gaussian and even non-linear time series (Shao and Wu, 2007). Similarly, the Fourier coefficients  $\tilde{\mathbf{Z}}_n$  are asymptotically independent and normally distributed with variances equal to  $2\pi$  times the spectral density at the corresponding frequency. This result gives rise to the following Whittle approximation in the frequency domain

$$p_{\tilde{\mathbf{Z}}_n}^W(\tilde{\mathbf{z}}_n | f) \propto \det(D_n)^{-1/2} \exp\left(-\frac{1}{2} \tilde{\mathbf{z}}_n^T D_n^{-1} \tilde{\mathbf{z}}_n\right) \quad (2)$$

by the likelihood of a Gaussian vector with diagonal covariance matrix

$$\begin{aligned} D_n &:= D_n(f) \\ &:= 2\pi \begin{cases} \text{diag}(f(0), f(\lambda_1), f(\lambda_1), \dots, f(\lambda_N), f(\lambda_N), f(\lambda_{n/2})), & n \text{ even,} \\ \text{diag}(f(0), f(\lambda_1), f(\lambda_1), \dots, f(\lambda_N), f(\lambda_N)), & n \text{ odd.} \end{cases} \end{aligned} \quad (3)$$

As explicitly shown in the supplementary material, this yields the famous Whittle likelihood in the time domain via the transformation theorem

$$p_{\mathbf{Z}_n = F_n^T \tilde{\mathbf{Z}}_n}^W(\mathbf{z}_n | f) \propto \exp\left\{-\frac{1}{2} \sum_{j=0}^{n-1} \left( \log f(\lambda_j) + \frac{I_{n,\lambda_j}(\mathbf{z}_n)}{f(\lambda_j)} \right)\right\}, \quad (4)$$

which provides an approximation of the true likelihood. It is exact only for Gaussian white noise in which case  $f(\lambda_j) = \sigma^2/2\pi$ . It has the advantage that it depends directly on the spectral density in contrast to the true likelihood that depends on  $f$  indirectly via Wiener-Khinchin’s theorem. Sometimes, the summands corresponding to  $j = 0$  as well as  $j = n/2$  (the latter for  $n$  even) are omitted in the likelihood approximation. In fact, the term corresponding to  $j = 0$  contains the sample mean (squared) while the term corresponding to  $j = n/2$  gives the alternating sample mean (squared). Both have somewhat different statistical properties and usually need to be considered separately. Furthermore, the first term is exactly zero if the methods are applied to time series that have been centered first, while the last one is approximately zero and asymptotically negligible (refer also Remark 2).

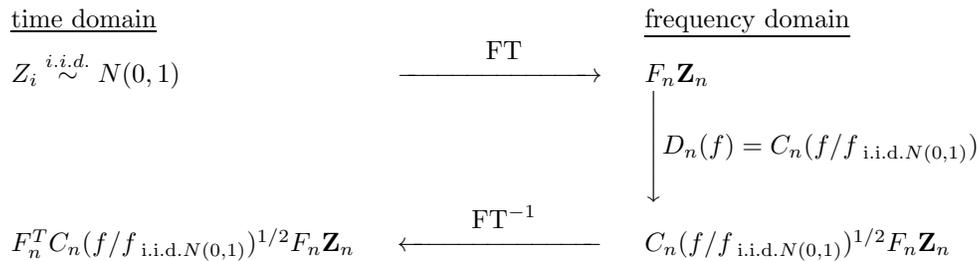
### 2.2 Nonparametric likelihood correction

The central idea in this work is to first conceive the Whittle likelihood as the likelihood of an i.i.d. Gaussian white noise working model that is nonparametrically corrected in the frequency domain. We then generalize the Whittle likelihood by employing more realistic parametric working models that more adequately describe the dependence structure of the time series.

Let us assume that we start with an i.i.d.  $N(0, 1)$  working model which has constant spectral density  $1/2\pi$ . This will in general not give a good spectral density estimate of the time series. Therefore we correct this by defining a correction matrix

$$C_n(f/f_{i.i.d.N(0,1)}) = \begin{cases} \text{diag} \left( \frac{f(\lambda_0)}{1/2\pi}, \frac{f(\lambda_1)}{1/2\pi}, \frac{f(\lambda_1)}{1/2\pi}, \dots, \frac{f(\lambda_N)}{1/2\pi}, \frac{f(\lambda_N)}{1/2\pi}, \frac{f(\lambda_{n/2})}{1/2\pi} \right), & n \text{ even,} \\ \text{diag} \left( \frac{f(\lambda_0)}{1/2\pi}, \frac{f(\lambda_1)}{1/2\pi}, \frac{f(\lambda_1)}{1/2\pi}, \dots, \frac{f(\lambda_N)}{1/2\pi}, \frac{f(\lambda_N)}{1/2\pi} \right), & n \text{ odd.} \end{cases}$$

and correct the Fourier-transformed time series in the frequency domain by multiplying by the square root of the correction matrix and inverse-Fourier-transform back into the time domain. This can be illustrated in a schematic diagram:



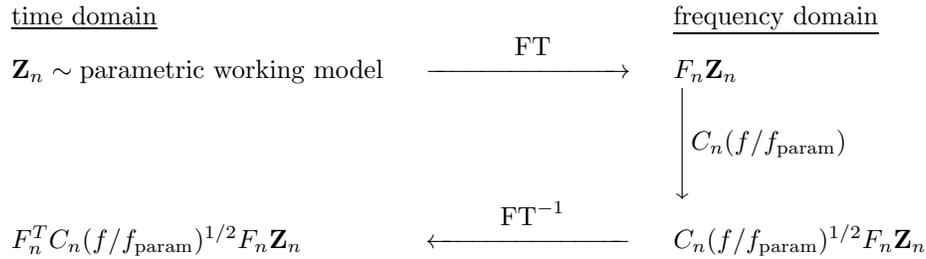
Then the density induced by this linear transformation is the Whittle likelihood as  $D_n(f) = C_n(f/f_{i.i.d.N(0,1)})$  in (3).

Adopting this viewpoint, it is straightforward to generalize the Whittle likelihood. We start with some parametric likelihood in the time domain, such as e.g. obtained from an ARMA-model, that is believed to be a reasonable approximation to the true

time series. We denote the spectral density that corresponds to this parametric working model by  $f_{\text{param}}(\cdot)$ . If the model is misspecified, then this spectral density is also wrong (biased) and needs to be corrected for the corresponding likelihood to have the correct second-order dependence structure. To this end, we define a correction matrix

$$C_n = C_n(f/f_{\text{param}}) = \begin{cases} \text{diag}\left(\frac{f(\lambda_0)}{f_{\text{param}}(\lambda_0)}, \frac{f(\lambda_1)}{f_{\text{param}}(\lambda_1)}, \frac{f(\lambda_1)}{f_{\text{param}}(\lambda_1)}, \dots, \right. \\ \left. \frac{f(\lambda_N)}{f_{\text{param}}(\lambda_N)}, \frac{f(\lambda_N)}{f_{\text{param}}(\lambda_N)}, \frac{f(\lambda_{n/2})}{f_{\text{param}}(\lambda_{n/2})}\right), & n \text{ even,} \\ \text{diag}\left(\frac{f(\lambda_0)}{f_{\text{param}}(\lambda_0)}, \frac{f(\lambda_1)}{f_{\text{param}}(\lambda_1)}, \frac{f(\lambda_1)}{f_{\text{param}}(\lambda_1)}, \dots, \right. \\ \left. \frac{f(\lambda_N)}{f_{\text{param}}(\lambda_N)}, \frac{f(\lambda_N)}{f_{\text{param}}(\lambda_N)}\right), & n \text{ odd.} \end{cases}$$

Our proposed concept of a nonparametric correction of a parametric time domain likelihood is illustrated in the schematic diagram:



As a result we obtain the following nonparametrically corrected likelihood function under the parametric working model

$$p_{\text{param}}^C(\mathbf{Z}_n | f) \propto \det(C_n)^{-1/2} p_{\text{param}}(F_n^T C_n^{-1/2} F_n \mathbf{Z}_n), \quad (5)$$

where  $p_{\text{param}}$  denotes the parametric likelihood.

The diagonal correction via  $C_n(f/f_{\text{param}})$  is a bias correction from the parametric spectral density to the true spectral density, while the covariance structure of the periodogram ordinates is inherited from the parametric model (see Proposition 1 c) below). In particular, for non-i.i.d. parametric models, the covariance of the periodogram ordinates is not diagonal (which can also be seen by the form of the corrected likelihood as given in the proof of Theorem 3.1 in the supplementary material on page 2). This is in contrast to Whittle's likelihood under which the periodogram ordinates are indeed independent.

Such an approach in a bootstrap context has been proposed and successfully applied by Kreiss and Paparoditis (2003) using an AR( $p$ ) approximation. In the context of nonparametric density estimation of i.i.d. data a similar semi-parametric correction approach has been proposed by Hjort and Glad (1995).

**Remark 1.** *Parametric models with a multiplicative scale parameter  $\sigma \neq 1$  yield the same corrected likelihood as the one with  $\sigma = 1$ , i.e. if  $\sigma \mathbf{Z}_n$  is used as working model this leads to the same corrected likelihood for all  $\sigma > 0$ . For instance, if the parametric model*

is given by i.i.d.  $N(0, \sigma^2)$  random variables with  $\sigma^2 > 0$  arbitrary, then the correction also results in the Whittle likelihood (for a proof we refer to the supplementary material). Analogously, for linear models  $Z_t = \sum_{l=-\infty}^{\infty} d_l e_{t-l}$ ,  $e_t \sim (0, \sigma^2)$ , which includes the class of ARMA-models, the corrected likelihood is independent of  $\sigma^2$ .

We can now prove the following proposition which shows two important things: First, the corrected likelihood is the exact likelihood in case the parametric model is correct. Second, the periodograms associated with this likelihood are asymptotically unbiased for the true spectral density regardless of whether the parametric model is true.

**Proposition 1.** *Let  $\{Z_t\}$  be a real zero mean stationary time series with absolutely summable autocovariance function  $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$  and let  $f_{\text{param}}(\lambda) \geq \beta > 0$  for  $0 \leq \lambda \leq \pi$  be the spectral density associated with the (mean zero) parametric model used for the correction.*

- a) *If  $f = f_{\text{param}}$ , then  $p_{\text{param}}^C = p_{\text{param}}$ .*  
 b) *The periodogram associated with the corrected likelihood is asymptotically unbiased for the true spectral density, i.e.*

$$\begin{aligned} \mathbb{E}_{p_{\text{param}}^C} I_{n, \lambda_j}(\mathbf{Z}_n) &= \int I_{n, \lambda_j}(z_1, \dots, z_n) dp_{\text{param}}^C(z_n, \dots, z_n) \\ &= \frac{f(\lambda_j)}{f_{\text{param}}(\lambda_j)} \mathbb{E}_{p_{\text{param}}} I_{n, \lambda_j}(\mathbf{Z}_n) = f(\lambda_j) + o(1), \end{aligned}$$

where the convergence is uniform in  $j = 0, \dots, \lfloor (n-1)/2 \rfloor$ .

- c) *Furthermore, for all  $k, l = 0, \dots, \lfloor (n-1)/2 \rfloor$ ,*

$$\begin{aligned} &\text{Cov}_{p_{\text{param}}^C}(I_{n, \lambda_l}(\mathbf{Z}_n), I_{n, \lambda_k}(\mathbf{Z}_n)) \\ &= \frac{f(\lambda_l) f(\lambda_k)}{f_{\text{param}}(\lambda_l) f_{\text{param}}(\lambda_k)} \text{Cov}_{p_{\text{param}}}(I_{n, \lambda_l}(\mathbf{Z}_n), I_{n, \lambda_k}(\mathbf{Z}_n)). \end{aligned}$$

The proof (which can be found in the supplementary material), shows that the vector of periodograms under the corrected likelihood has exactly the same distributional properties as the vector of the periodograms under the parametric likelihood multiplied with  $f(\cdot)/f_{\text{param}}(\cdot)$ . Hence, asymptotic properties as the ones derived in Theorem 10.3.2 in Brockwell and Davis (2009) carry over with the appropriate multiplicative correction.

In the remainder of the paper we describe how to make use of this nonparametric correction in a Bayesian set-up.

### 3 Bayesian semiparametric approach to time series analysis

To illustrate the Bayesian semiparametric approach and how to sample from the pseudo-posterior distribution, in the following we restrict our attention to an AR( $p$ ) model as

our parametric working model for the time series, i.e.  $Z_i = \sum_{l=1}^p a_l Z_{i-l} + \epsilon_i$ , where  $\{\epsilon_i\}$  are i.i.d.  $N(0, 1)$  random variables with density denoted by  $\varphi(\cdot)$ . Note that without loss of generality,  $\sigma^2 = 1$ , cf. Remark 1. This yields the parametric likelihood of our working model, depending on the order  $p \geq 0$  and on the coefficients  $\mathbf{a} = (a_1, \dots, a_p)$ :

$$p_{\text{param}}(\mathbf{Z}_n | \mathbf{a}) = p_{\text{AR}}(\mathbf{Z}_n | \mathbf{a}) \propto p_{\text{AR}}(\mathbf{Z}_p | \mathbf{a}) \prod_{i=p+1}^n \varphi \left( Z_i - \sum_{l=1}^p a_l Z_{i-l} \right) \quad (6)$$

with spectral density

$$f_{\text{param}}(\lambda; \mathbf{a}) = \frac{1}{2\pi} \left| 1 - \sum_{l=1}^p a_l e^{-il\lambda} \right|^{-2}. \quad (7)$$

The second factor in (6) gives the likelihood of  $Z_{p+1}, \dots, Z_n$  given  $Z_1, \dots, Z_p$  and is usually referred to as partial likelihood. In the paper, we will use the full likelihood with the exception of the analysis of the LIGO data in Section 4.5 below, where the partial likelihood was used for computational reasons. We assume the time series to be stationary and causal. Thus,  $\mathbf{a}$  is restricted such that  $\phi(z) := 1 - a_1 z - \dots - a_p z^p$  has no zeros inside the closed unit disc, cf. Theorem 3.1.1. in Brockwell and Davis (2009). Because the corresponding spectral density is strictly positive and continuous on  $[0, \pi]$ , it is bounded away from 0. Furthermore, each  $X_t$  depends only on the previous  $p$  observations whose influence on future observations decreases exponentially fast. As a consequence, this property is inherited by the autocovariance function of causal AR(p) (in fact causal ARMA(p, q)) time series. More precisely, it holds by (3.3.10) in Brockwell and Davis (2009):  $|\gamma_{\text{AR}}(h)| \leq c|h|^p r^{-|h|} \leq CR^{-|h|}$  for some  $c, C > 0$  and  $r, R > 1$ . In particular, the assumptions on the parametric model as given in Proposition 1 and Assumption A.1 below are fulfilled.

For now, we assume that the parameters  $(p, \mathbf{a})$  of the parametric working model are fixed (and in practice set to Bayesian point estimates obtained from a preceding parametric estimation step). An extension to combine the estimation of the parametric model with the nonparametric correction will be detailed later in Section 3.4.

### 3.1 Nonparametric prior for spectral density inference

For a Bayesian analysis using either the Whittle or nonparametrically corrected likelihood, we need to specify a nonparametric prior distribution for the spectral density. Here we employ the approach by Choudhuri et al. (2004) which is essentially based on the Bernstein polynomial prior of Petrone (1999) as a nonparametric prior for a probability density on the unit interval. We briefly describe the prior specification and refer to Choudhuri et al. (2004) for further details.

In contrast to the approach in Choudhuri et al. (2004), we do not specify a nonparametric prior distribution for the spectral density  $f(\cdot)$ , but for a pre-whitened version thereof, incorporating the spectral density of the parametric working model into the estimation. To elaborate, for  $0 \leq \eta \leq 1$ , consider the *eta-damped correction function*

$$c_\eta(\lambda) = c_\eta(\lambda; \mathbf{a}) := f(\lambda) / f_{\text{param}}(\lambda; \mathbf{a})^\eta. \quad (8)$$

This corresponds to a reparametrization of the likelihood (5) by replacing  $C_n = C_n(f/f_{\text{param}})$  with  $C_n = C_n(c_\eta(\lambda; \mathbf{a})f_{\text{param}}(\lambda; \mathbf{a})^{\eta-1})$ .

We now put a prior on  $c_\eta(\cdot)$ , which – for fixed  $\eta$  – includes the two extreme cases: (i) prior on the spectral density  $f$  (if  $\eta = 0$ ) and (ii) the correction  $f/f_{\text{param}}$  (if  $\eta = 1$ ). Because nonparametric estimation becomes generally easier the smoother the function to be estimated is, a prior on  $f$  is preferable if the density is smoother than the correction while a prior on the correction  $f/f_{\text{param}}$  is preferable otherwise. If the parametric working model is already close to the truth or captures the more prominent peaks it is thus preferable to set  $\eta$  close to one, otherwise it is preferable to set  $\eta$  close to 0. In this sense, the parameter  $\eta$  can be seen as a **model confidence parameter**.

The choice of  $\eta$  will be detailed in Section 3.4, but for now,  $\eta$  is assumed fixed.

We reparametrise  $c_\eta(\cdot)$  to a density function  $q(\cdot)$  on  $[0, 1]$  via  $c_\eta(\pi\omega) = \tau q(\omega)$ ,  $0 \leq \omega \leq 1$  with normalization constant  $\tau = \int_0^1 c_\eta(\pi\omega) d\omega$ . Thus, a prior for  $c_\eta(\cdot)$  may be specified by putting a Bernstein polynomial prior on  $q(\cdot)$  and then an independent Inverse-Gamma( $\alpha_\tau, \beta_\tau$ ) prior on  $\tau$ , its density denoted by  $p_\tau$ . The Bernstein polynomial prior of  $q$  is specified in a hierarchical way as follows:

1.  $q(\omega|k, G) = \sum_{j=1}^k G(\frac{j-1}{k}, \frac{j}{k})\beta(\omega|j, k-j+1)$  where  $G(u, v] = G(v) - G(u)$  for a distribution function  $G$  and  $\beta(\omega|l, m) = \frac{\Gamma(l+m)}{\Gamma(l)\Gamma(m)} \omega^{l-1}(1-\omega)^{m-1}$  is the beta density with parameters  $l$  and  $m$ .
2.  $G$  has a Dirichlet process distribution with base measure  $\alpha = MG_0$ , where  $M > 0$  is a constant and  $G_0$  a distribution function with Lebesgue density  $g_0$ .
3.  $k$  has a discrete distribution on the integers  $k = 1, 2, \dots$ , independent of  $G$ , with probability function  $p(k) \propto \exp(-\theta k \log(k))$ . Note that smaller values of  $k$  yield smoother densities.

Furthermore, we achieve an approximate finite-dimensional characterization of this non-parametric prior in terms of  $2L+3$  parameters  $(V_1, \dots, V_L, W_0, W_1, \dots, W_L, k, \tau)$  by employing the truncated i.e. stick-breaking representation of the Dirichlet process (Sethuraman, 1994)

$$G = \sum_{l=1}^L p_l \delta_{W_l} + (1 - p_1 - \dots - p_L) \delta_{W_0}$$

with  $p_1 = V_1$ ,  $p_l = \left(\prod_{j=1}^{l-1} (1 - V_j)\right) V_l$  for  $l \geq 2$ ,  $V_l \sim \text{beta}(1, M)$ , and  $W_l \sim G_0$ , all independent. This gives a prior finite mixture representation of the eta-damped correction

$$c_\eta(\pi\omega) = \tau \sum_{j=1}^k \tilde{w}_{j,k} \beta(\omega|j, k-j+1), \tag{9}$$

where  $\tilde{w}_{j,k} = \sum_{l=0}^L p_l I\{\frac{j-1}{k} < W_l \leq \frac{j}{k}\}$  and  $p_0 = 1 - \sum_{l=1}^L p_l$ .

The joint prior density of  $c_\eta$  by means of this finite-dimensional approximation can be written as

$$p(V_1, \dots, V_L, W_0, W_1, \dots, W_L, k, \tau) \propto \left( \prod_{l=1}^L M(1 - V_l)^{M-1} \right) \left( \prod_{l=0}^L g_0(W_l) \right) p_k(k) p_\tau(\tau). \quad (10)$$

Here, we specify a diffuse prior by choosing the uniform distribution for  $G_0$  and  $M = 1$ . We set  $\theta = 0.01$ ,  $\alpha_\tau = \beta_\tau = 0.001$  and follow the recommendation by Choudhuri et al. (2004) for the truncation point  $L = \max\{20, n^{1/3}\}$ .

### 3.2 Posterior computation

The prior (9) on  $c_\eta(\cdot)$  induces a prior on  $f(\cdot)$  by multiplication with  $f_{\text{param}}(\cdot; \mathbf{a})^\eta$ , see (8). Accordingly, the pseudo-posterior distribution of  $f(\cdot)$  can be computed as prior times the corrected parametric likelihood:

$$p_{\text{post}}^C(V_1, \dots, V_L, W_0, W_1, \dots, W_L, k, \tau | \mathbf{Z}_n, \mathbf{a}, \eta) \propto p(V_1, \dots, V_L, W_0, W_1, \dots, W_L, k, \tau) \det(C_n)^{-1/2} p_{\text{param}}(F_n^T C_n^{-1/2} F_n \mathbf{Z}_n | \mathbf{a}),$$

where  $C_n = C_n(c_\eta(\lambda; \mathbf{a}) f_{\text{param}}(\lambda; \mathbf{a})^{\eta-1})$  and  $f_{\text{param}}(\lambda; \mathbf{a})$  as in (7). Samples from the pseudo-posterior distribution can be obtained via Gibbs sampling following the steps outlined in Choudhuri et al. (2004). The full conditional for  $k$  is discrete and readily sampled, as is the conjugate full conditional of  $\tau$ . We use the Metropolis algorithm to sample from each of the full conditionals of  $V_l$  and  $W_l$  using the uniform proposal density of Choudhuri et al. (2004).

**Remark 2.** *As in Choudhuri et al. (2004), we omit the first and last terms in the corrected likelihood that correspond to  $\lambda = 0$  and  $\lambda = n/2$  (and setting  $c_\eta(0) = c_\eta(n/2) = 0$  as well as  $F_n \mathbf{Z}_n(0) = F_n \mathbf{Z}_n(n) = 0$ ). This is due to the role that the corresponding Fourier coefficients play (being equal to the sample mean respectively alternating sample mean), which typically requires a special treatment (see Proposition 10.3.1 and equation (10.4.7) in Brockwell and Davis (2009)). For the application to spectral density estimation in this paper this leads to more stable statistical procedures irrespective of the true mean of the time series. However, in situations, where the time series is merely used to describe the error terms such as regression models, change point or unit-root testing, these coefficients should be included and the likelihood used for the time series  $Z_t - \mu$ , where  $\mu$  is the mean (not the sample mean) of the time series.*

### 3.3 Posterior consistency

In this section, we will show consistency of the pseudo-posterior distribution based on the Bernstein polynomial prior and the corrected likelihood for a given working model under the same assumptions as Choudhuri et al. (2004). Bayesian parametric modelling has the advantage of Bernstein-von-Mises theorems that guarantee that frequentist and

Bayesian inference procedures are asymptotically equal to the first order of approximation. Because this is no longer true for Bayesian nonparametric models, we need to explicitly show posterior consistency, i.e. that the posterior distribution eventually concentrates in the neighbourhood of the true parameter value as the sample size gets large. To this end, contiguity is a powerful concept, introduced by Le Cam and Yang (1960), for comparing the asymptotic behaviour of two sequences of probability measures, defined on spaces depending on the sample size  $n$ . The main usefulness of mutual contiguity of the true joint distribution of the Fourier Coefficients  $F_n Z_n$ , the Whittle measure  $N(0, D_n)$  and as shown below the corrected likelihood, is that consistency results (or rates of convergence) obtained under the Whittle measure or corrected likelihood immediately carry over to the true original Gaussian measure. This is important to be able to use techniques of Bayesian nonparametric asymptotics that only work if the posterior is based on the true likelihood, which is not the case when the Whittle measure or the corrected likelihood is used. This is also one of the main challenges in obtaining consistency results in the non-Gaussian case.

The paper by Ghosal et al. (1999) gives sufficient conditions for consistency of the posterior distribution when using Bernstein polynomial priors in terms of the existence of exponentially powerful tests for testing  $H_0 : \theta = \theta_0$  and prior positivity of a Kullback-Leibler neighbourhood. But these require i.i.d. observations. To prove posterior consistency under the Whittle (and hence by contiguity the true) likelihood, Choudhuri et al. (2004) extend this result to independent but not identically distributed observations and apply this to the periodogram ordinates which are independent exponential random variables under the Whittle likelihood. However, because periodogram ordinates under the corrected likelihood are no longer independent, that theorem is no longer applicable. Therefore, we derive an extension to non-independent Gaussian random variables allowing us to prove consistency under the corrected and hence by contiguity the true Gaussian likelihood.

Throughout the section, we will make the following assumption, where from now on we denote the true spectral density by  $f_0$  to avoid confusion.

**Assumption A.1.** a) For the parametric working model with autocovariance function  $\gamma_{param}(\cdot)$  and spectral density  $f_{param}(\cdot)$  it holds that

$$\sum_{h \in \mathbb{Z}} h^\alpha |\gamma_{param}(h)| < \infty \text{ for some } \alpha > 1,$$

$$f_{param}(\lambda) \geq \beta > 0 \text{ for some } \beta > 0 \quad \text{and all } 0 \leq \lambda \leq \pi.$$

b)  $\{Z_t\}$  is a stationary mean zero Gaussian time series with autocovariance function  $\gamma_0(\cdot)$  and spectral density  $f_0(\cdot)$  with

$$\sum_{h \in \mathbb{Z}} h^\alpha |\gamma_0(h)| < \infty \text{ for some } \alpha > 1,$$

$$f_0(\lambda) \geq \beta > 0 \text{ for some } \beta > 0 \quad \text{and for all } 0 \leq \lambda \leq \pi.$$

The density and the distribution of  $\mathbf{Z}_n = (Z_1, \dots, Z_n)$  is denoted by  $p_{n, f_0}(\cdot)$  and  $P_{n, f_0}$ .

An important first observation is, that in the Gaussian case the corrected likelihood, the Whittle likelihood as well as the true likelihood are all mutually contiguous (for a definition we refer to Van der Vaart (1998), Definition 6.3). This fact may also be of independent interest:

**Theorem 1.** *Under Assumptions A.1 the true density  $p_{n,f_0}(\cdot)$ , the Whittle likelihood  $p^W(\cdot|f)$  given in (4) as well as the corrected (Gaussian) parametric likelihood  $p_{param}^C(\cdot|f)$  given in (5) are all mutually contiguous.*

With the help of this theorem we are now able to prove posterior consistency under certain assumptions on the time series and prior (a detailed proof of Theorems 1 and 2 can be found in the supplementary material).

**Theorem 2.** *Let  $0 \leq \eta \leq 1$  be fixed. Let Assumptions A.1 hold in addition to the following assumptions on the prior for  $c_\eta(\cdot)$ :*

(P1) *for all  $k$ ,  $0 < p_k(k) \leq B \exp^{-bk \log k}$  for some constants  $B, b > 0$ ,*

(P2)  *$g_0$  is bounded, continuous and bounded away from 0,*

(P3) *the parameter  $\tau$  is assumed fixed and known.*

*Let  $c_{0,\eta}(\lambda) = f_0(\lambda)/f_{param}^\eta(\lambda)$ . Then the posterior distribution is consistent, i.e. for any  $\epsilon > 0$ ,*

$$\Pi_n(c : \|c - c_{0,\eta}\|_1 > \epsilon | \mathbf{Z}_n) \rightarrow 0$$

*in  $P_{n,f_0}$ -probability, where  $\Pi_n(\cdot | \mathbf{Z}_n)$  denotes the pseudoposterior distribution computed using the corrected likelihood.*

While we prove consistency only for the Gaussian case, we conjecture that consistency of the spectral density holds also for non-Gaussian time series (with certain moment assumptions) as the asymptotic distribution of the spectral density only depends on the first and second-order structure of the time series (see the simulation results for non-Gaussian data in the supplementary material). This is supported by the fact that the asymptotic distribution of the spectral density only depends on the first and second order structure of the time series in combination with theoretical findings on the applicability as well as limitations of the corresponding likelihood approximations in a bootstrap context (see Franke and Härdle (1992), Dahlhaus and Janas (1996), Kreiss and Paparoditis (2003), Kreiss et al. (2011)).

The assumption of a known, fixed  $\tau$  is needed to prove posterior consistency but can be relaxed to a sequence of prior distributions that asymptotically bracket the true value as in Remark 1, Section 5 of Choudhuri et al. (2004). In simulation and case studies, we assumed  $\tau$  to be unknown with an Inverse-Gamma(0.001,0.001) prior distribution.

### 3.4 Prior for the parameters of the working model

In the following we will consider a stationary autoregressive time series of order  $p$  in its causal representation as a working model. This fulfills the assumptions on the autocovariance function (which decreases exponentially with increasing lag  $h$ ) and on the spectral density, see Brockwell and Davis (2009).

In the previous sections, the parameters of the working model were assumed to be fixed, as e.g. obtained in an initial pre-estimation step. From a Bayesian perspective, it is desirable to couple the inference about the parametric working model with the nonparametric correction, allowing for the inclusion of prior knowledge about the model and for uncertainty quantification about the interaction of model and correction. Thus, for a fixed order  $p$ , we include both the autoregressive parameters  $\mathbf{a} = (a_1, \dots, a_p)$  and the spectral shape confidence  $\eta$  from (8) into the Bayesian inference. The introduction of the parameter  $\eta$  effectively robustifies the procedure in the sense that it guarantees our method will not be worse than a corresponding fully nonparametric one. Putting a prior on these additional parameters, may make these parameters (but not the spectral density) unidentifiable, but will lead to superior performance as demonstrated by the following simulations. In the context of priors for variance components, additional non identifiable parameters have also been successfully introduced by Gelman (2006).

Because the domain of the parameter space for stationary and causal AR( $p$ )-processes has a rather complicated shape (namely the roots of the characteristic polynomial need to lie outside the unit circle), we use a reparametrization suggested by Nam et al. (2014) based on the partial autocorrelation  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_p)$ . Then the only condition needed to ensure stationarity and causality (and hence identifiability) of the parametric model is the condition that all partial autocorrelations  $\rho_j$ ,  $j = 1, \dots, p$ , take values between  $-1$  and  $1$ , which is ensured by our i.i.d.  $U(-1, 1)$ -prior. The autoregressive parameters  $\mathbf{a} = \mathbf{a}(\boldsymbol{\rho})$  can be readily obtained from this parametrisation (further details can be found in the supplementary material).

We consider the following prior specification for the spectral density:

$$f(\lambda) = c_\eta(\lambda) f_{\text{param}}(\lambda; \boldsymbol{\rho})^\eta,$$

with a Bernstein-Dirichlet prior on  $c_\eta(\cdot)$  as in Section 3.1, a uniform prior on  $\eta$  and uniform priors on the  $\rho_l$ 's, all a priori independent. Of course, it is possible to employ different prior models (see Liseo and Macaro (2013); Sørbye and Rue (2016)). In conjunction with the corrected parametric likelihood, we obtain samples from the joint pseudo-posterior distribution

$$p_{\text{post}}^C(v_1, \dots, v_L, w_0, w_1, \dots, w_L, k, \tau, \rho_1, \dots, \rho_p, \eta | \mathbf{Z}_n)$$

analogously to Section 3.2 via Gibbs sampling. Note that, since the corrected parametric likelihood is the Lebesgue density of a probability measure, it is sufficient that the prior distributions are proper for the posterior distribution to be proper. We use random walk Metropolis-within-Gibbs steps with normal proposal densities to sample from the full conditionals of  $\eta$  and  $\rho_1, \dots, \rho_p$  respectively. The proposal variance for  $\eta$  is set to 0.01,

where proposals larger than 1 (smaller than 0) are truncated at 1 (at 0). To achieve proper mixing of the parametric model parameters, the proposal variances  $\sigma_t^2$  for  $\rho_l$  are determined adaptively as described in Roberts and Rosenthal (2009) during the burn-in period, aiming for an acceptance rate of 0.44, where proposals with absolute value larger or equal to one are discarded.

**Remark 3.** *While we assume the autoregressive order  $p$  to be fixed, it is also possible to include the autoregressive order in the Bayesian inference by using a Reversible-jump Markov Chain Monte Carlo scheme (Green, 1995) or stochastic search variables (Barnett et al., 1996). However, the simulations in Section 4.1 indicate and the discussion in Section 4.2 points out, that this may only be done in combination with a heavily penalized prior on  $p$ . In our approach, we suggest a preliminary model selection step instead (as outlined in Section 4.2).*

### 3.5 Missing values

The Gibbs sampler can be easily extended to handle missing values in the time series. Missing values are treated as unknown parameters and sampled from the full conditional density. For Gibbs sampling, it is sufficient to consider the case of a single missing value, say  $Z_m$  for  $m \in \{p+1, \dots, n\}$ , cf. the detailed illustration for Gibbs sampling of autoregressive time series in McCulloch and Tsay (1994).

Let  $\boldsymbol{\theta} = (V_1, \dots, V_L, W_0, W_1, \dots, W_L, k, \tau, \mathbf{a}, \eta)$  denote the vector of parameters, then the full conditional posterior distribution of  $Z_m$  given  $\mathbf{Z}_{(-m)}$  and  $\boldsymbol{\theta}$  is proportional to the parametric likelihood  $p_{\text{param}}(F_n^T C_n^{-1/2} F_n \mathbf{Z}_n | \mathbf{a})$  and thus again a normal distribution which is given in the supplementary material.

## 4 Numerical evaluation

In this section, we evaluate the finite sample behavior of our *nonparametrically corrected (NPC)* approach to Bayesian spectral density estimation numerically. To demonstrate the trade-off between the parametric working model and the nonparametric spectral correction, we compare our approach to both fully parametric and fully nonparametric approaches. We first present the results of a simulation study with ARMA data in Section 4.1 before considering sunspot data in Section 4.3, light intensities of the *S. Carinae* variable star in Section 4.4 as well as gravitational wave data in Section 4.5. An implementation of all procedures presented below is provided in the R package `beyondWhittle`, which is available on CRAN, see Meier et al. (2017).

### 4.1 Simulated ARMA data

We consider data generated from the ARMA model

$$Z_t = aZ_{t-1} + be_{t-1} + e_t, \quad 1 \leq t \leq n \quad (11)$$

with standard Gaussian white noise  $e_t$  and different values of  $a, b$  and  $n$  including pure AR (with  $b = 0$ ) and pure MA (with  $a = 0$ ) models. The following competing approaches are compared with NPC:

*Nonparametric estimation (NP).* The procedure from Choudhuri et al. (2004), which is based on the Whittle likelihood and a Bernstein-Dirichlet prior on the spectral density. Note that this coincides with the NPC approach with a white noise parametric working model ( $p = 0$ ), cf. Remark 1.

*Autoregressive estimation (AR).* For  $p = 0, 1, \dots, p_{\max}$ , an autoregressive model of order  $p$  is fitted to the data using a Bayesian approach based on the full parametric likelihood as given in (6) with the same partial autocorrelation parametrization and the same prior assumptions as for the parametric working model within the nonparametrically corrected likelihood procedure, see Section 3.4 (for details on the sampling scheme we refer to the supplementary material). The order  $p^*$  minimizing the Deviance Information Criterion (DIC) is then chosen for model comparison. For computational reasons, the partial likelihood was used to calculate the DIC.

The working model in the NPC approach is chosen to be the  $\text{AR}(p^*)$  model from the AR procedure. The prior for the working model parameters is as described in Section 3.4 and the prior for the nonparametric correction is as described in Section 3.1. For the NP approach, the same Bernstein-Dirichlet prior for  $f(\cdot)$  is used as for  $c_\eta(\cdot)$  in the NPC approach.

We compare the average Integrated Absolute Error (aIAE) of the posterior median spectral density estimate, the empirical coverage probability of a Uniform Credible Interval (cUCI) as well as the median of the median lengths of these uniform credible intervals (mLEN). It is well known, that the frequentist coverage of Bayesian credibility sets can fail badly in theory and practice (see Szabó et al. (2015) for a recent treatment on the matter), but it appears that our methodology yields reasonable results at least in the simple examples considered.

Pointwise posterior credible intervals are not suited for investigating coverage, since they do not take the multiple testing problem into account that arises at different frequencies. Following Häfner and Kirch (2016) (see also Neumann and Polzehl (1998)), a Uniform Credible Interval for the spectral density can be constructed as follows: Denote by  $f_1^*(\cdot), \dots, f_N^*(\cdot)$  the posterior spectral density samples obtained by one of the procedures. Then for  $0 < \alpha < 1$  the Uniform  $\alpha$ -Credible Interval is given by

$$[\exp(\log f^*(\lambda) - C_\alpha^* \sigma^*(\lambda)), \exp(\log f^*(\lambda) + C_\alpha^* \sigma^*(\lambda))],$$

where  $f^*(\lambda)$  denotes the sample median at frequency  $0 \leq \lambda \leq \pi$ ,  $\sigma^*(\lambda)$  the median absolute deviation of  $\log f_1^*(\lambda), \dots, \log f_N^*(\lambda)$  and  $C_\alpha^*$  is chosen such that

$$\frac{1}{N} \sum_{j=1}^N 1 \left\{ \max_{0 \leq \lambda \leq \pi} \frac{\log f^*(\lambda) - \log f_j^*(\lambda)}{\sigma^*(\lambda)} \leq C_\alpha^* \right\} \geq 1 - \alpha.$$

The intervals are constructed on a logarithmic scale to ensure that their covered range contains only positive values. Because small values of  $f^*$  lead to very large absolute

(a) Order selection via DIC.

	AR(1): $a = 0.95$			MA(1): $b = 0.8$			ARMA(1, 1): $a = 0.75, b = 0.8$		
	$n = 64$	$n = 128$	$n = 256$	$n = 64$	$n = 128$	$n = 256$	$n = 64$	$n = 128$	$n = 256$
aIAE									
AR	2.590	2.012	1.533	0.292	0.241	0.190	1.203	1.016	0.851
NP	3.543	2.946	2.370	0.197	0.153	0.119	1.022	0.806	0.625
NPC	2.992	2.240	1.612	0.206	0.157	0.121	1.083	0.907	0.727
cUCI									
AR	0.959	0.970	0.990	0.857	0.869	0.871	0.844	0.830	0.886
NP	0.863	0.771	0.801	0.998	0.999	0.995	0.953	0.919	0.906
NPC	0.952	0.973	0.996	0.998	1.000	1.000	0.999	1.000	0.998
mLEN									
AR	0.184	0.108	0.071	2.402	1.572	0.909	3.304	1.402	0.668
NP	1.024	0.441	0.248	1.303	0.660	0.399	1.441	0.630	0.344
NPC	0.632	0.290	0.150	3.273	1.689	0.985	4.862	2.110	0.974
$\hat{\eta}$	0.697	0.818	0.896	0.285	0.191	0.121	0.483	0.384	0.272

(b) Fixed order  $p = 1$ .

	ARMA(1, 1): $a = 0.75, b = 0.8$								
	aIAE			cUCI			mLEN		
	$n = 64$	$n = 128$	$n = 256$	$n = 64$	$n = 128$	$n = 256$	$n = 64$	$n = 128$	$n = 256$
$AR_{(p=1)}$	1.171	1.234	1.289	0	0	0	0.295	0.177	0.116
$NPC_{(p=1)}$	0.942	0.741	0.586	0.986	0.969	0.954	0.847	0.443	0.264
$\hat{\eta}$	0.601	0.635	0.670	-	-	-	-	-	-

Table 1: Average Integrated Absolute Error (aIAE), Uniform 0.9-Credible Interval coverage (cUCI), median length of credible intervals and average posterior model confidence  $\hat{\eta}$  for different realizations of model (11).

values on a log scale, we do not employ the usual logarithm, but the Fuller-logarithm as described in Fuller (1996), page 496, i.e.

$$\log_{\text{Fuller}}(x) = \log(x + \xi) - \xi/(x + \xi),$$

for some small  $\xi > 0$ . We use  $\alpha = 0.9$  and  $\xi = 0.001$  in our simulations. The chains were run for 12,000 iterations for AR (after a burn-in period of 8,000 iterations) and for 20,000 iterations for NP and NPC (after a burn-in period of 30,000 iterations), where a thinning of 4 was employed for NP and NPC. We choose  $p_{\max} = 15$  and consider lengths  $n = 64, 128, 256$  from model (11) with  $N = 1024$  replicates ( $N$  a power of 2 to use the computational resources efficiently) respectively.

The results are shown in Table 1a. For AR(1) data, the AR procedure yields the best results (in terms of aIAE, cUCI and mLEN), whereas the NP performs worst. It can be seen that AR and NPC benefit from the well-specified parametric model. For MA(1) data, however, the AR approach yields the worst results, whereas NPC benefits from the nonparametric correction, yielding only slightly worse integrated errors than NP, at the

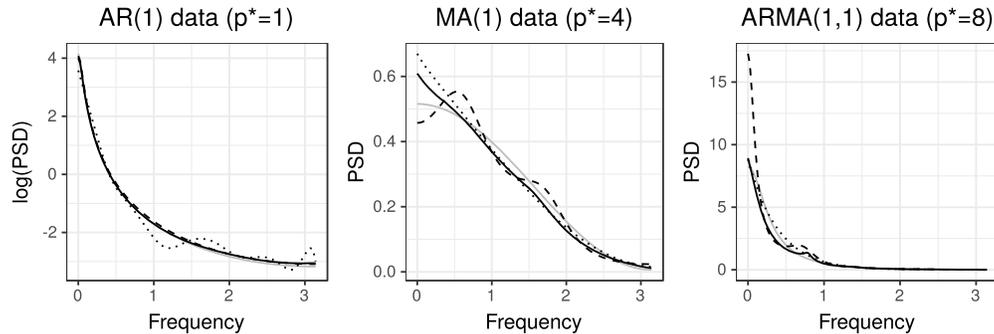


Figure 1: Posterior median spectral density estimates of AR (dashed), NP (dotted) and NPC (solid black) for a single instance of the AR(1), MA(1), and ARMA(1,1) time series compared to the true spectral density (gray).

expense of longer credible intervals. In case of ARMA(1,1) data, the estimation does not benefit from the autoregressive fit, i.e. the moving average misspecification dominates the estimation. Thus the results are similar to the MA(1) case. For the correctly specified AR(1) model, the estimate of the confidence parameter  $\hat{\eta}$  is close to 1, in particular it increases with increasing sample size. In the misspecified case of the MA(1) model,  $\hat{\eta}$  is close to zero and decreases with increasing sample size, while in the intermediate ARMA(1,1) case, its estimate is close to 0.5 but also decreases with increasing sample size.

The fact, that the nonparametric correction does not benefit from the parametric fit in the case of the ARMA(1,1) model is somewhat surprising. However, this is due to the use of the DIC for order selection which is not optimal for the nonparametric correction as explained in Section 4.2 below. Table 1b shows the simulation results for the ARMA(1,1) model and a fixed order of  $p = 1$ . While a parametric AR(1) model is clearly not able to explain the data (see e.g. the zero coverage of the uniform credibility intervals), this choice of the order significantly improves the results of the NPC procedure for the ARMA(1,1) data. In fact, the latter is now better than both the AR procedure as well as the NP procedure in terms of aIAE. At the same time the confidence in the model as indicated by  $\hat{\eta}$  increases and the median length of the credible sets is now smaller than for all other procedures while the coverage is comparable or even larger.

Figure 1 shows the posterior median spectral density estimates for one exemplary replication of each of the three models considered in Table 1a. The true spectral density is shown in gray. The other lines are the (pointwise) posterior median spectral densities of AR (dashed), NP (dotted) and NPC (solid black).

We also explored cases where the time series is almost non-stationary. Table 2 shows average integrated absolute errors and coverage of 90% credible intervals for AR(1) models with autoregressive parameters  $a = 0.98, 0.99, 0.995, \text{ and } 0.999$ . Even for AR(1) models that are close to a unit root, the average integrated absolute error of NPC is generally smaller than that of NP and very close that of the correctly specified parametric

	AR(1): $a = 0.98$			AR(1): $a = 0.99$			AR(1): $a = 0.995$			AR(1): $a = 0.999$		
	$n = 64$	$n = 128$	$n = 256$	$n = 64$	$n = 128$	$n = 256$	$n = 64$	$n = 128$	$n = 256$	$n = 64$	$n = 128$	$n = 256$
aIAE												
AR	7.93	6.39	4.82	16.11	13.40	10.37	26.80	23.21	18.32	37.41	33.26	27.14
NP	10.03	9.13	8.10	18.93	18.02	16.69	29.86	28.99	27.71	40.72	39.94	38.75
NPC	9.15	7.35	5.32	17.87	15.32	11.87	28.63	25.71	20.87	39.25	36.30	29.98
cUCI												
AR	0.85	0.91	0.96	0.73	0.84	0.91	0.65	0.77	0.89	0.61	0.73	0.86
NP	0.51	0.29	0.26	0.22	0.05	0.04	0.11	0.01	0.00	0.06	0.00	0.00
NPC	0.89	0.94	0.99	0.78	0.90	0.97	0.72	0.86	0.96	0.66	0.82	0.93
mLEN												
AR	0.171	0.104	0.067	0.174	0.102	0.066	0.168	0.099	0.065	0.166	0.099	0.065
NP	1.162	0.504	0.281	1.238	0.551	0.307	1.251	0.575	0.342	1.247	0.604	0.372
NPC	0.580	0.262	0.140	0.550	0.242	0.136	0.530	0.232	0.127	0.500	0.220	0.126
$\hat{\eta}$	0.742	0.859	0.932	0.755	0.881	0.942	0.761	0.886	0.950	0.765	0.890	0.948

Table 2: Average Integrated Absolute Error (aIAE), Uniform 0.9-Credible Interval coverage (cUCI) and median length of credibility intervals for different realizations of model (11) close to unit root.

model. Furthermore, the coverage of the 90% credible interval of NPC reach the nominal value for increasing sample sizes whereas the coverage for the NP method is close to 0 even though they are considerably longer than the credibility intervals obtained from NPC with a much better coverage. The values of the model confidence parameter  $\eta$  are increasing with sample size and for  $n = 256$  very close to 1.

Further simulation results can be found in Section 3 in the supplementary material. In particular, we get qualitatively the same results even for non-Gaussian data, indicating that the properties of the NPC approach that we illustrated in this section carry over to the non-Gaussian case.

## 4.2 Elbow Criterion

Under relatively weak conditions (see e.g. Kreiss et al. (2011)) a linear process can be written as an  $AR(\infty)$ -process with white noise errors (similarly to the famous Wold representation). Consequently, an  $AR(p)$ -model with sufficiently large order captures the structure of a (Gaussian) linear process to any degree of accuracy. In this sense, the use of an AR-model with a sufficiently large order can be viewed as a nonparametric procedure, a fact, that has been exploited by AR-sieve-bootstrap methods for quite some time. For a recent mathematical analysis of the validity and limitations of this approach we refer to Kreiss et al. (2011).

Consequently, an AR-model can still be used for spectral density estimation under misspecification as long as the order is sufficiently large. This is why standard order selection techniques such as DIC-minimization tend to choose large orders in misspecified situations. In combination with the nonparametric correction such a large order leads to suboptimal results as has been seen in the above simulations probably due to the use of two competing nonparametric procedures. The simulation study suggests that it is indeed much better for NPC to use a parametric model that improves the spectral density estimate over that of Gaussian white noise but also leaves some room for the

nonparametric correction to work. Consequently, we aim at choosing a relatively small order that is still capturing the main features. In other words, we are not interested in an elaborate  $AR(\infty)$  approximation but rather in a proxy model that captures the main parametric features of the data.

As a rule-of-thumb we suggest to look at scree-like plots of the negative maximum log likelihood: For increasing orders one can often see a clear bend (elbow) in the curve (with a slow decay from that point on that is not slow enough to be captured by standard penalization techniques). Similar to the use of scree plots in the context of Principal Components Analysis (PCA), that point can be seen as a reasonable truncation point (‘elbow criterion’) where those features best explained by the parametric model have been captured. While this small model does not yet fully explain the data, adding more parameters is not helping the nonparametrically corrected procedure that we propose.

In the context of an autoregressive working model for the generation of the elbow plots only, we approximate the negative maximum log-likelihood by the negative log-likelihood evaluated at the Yule-Walker estimate. This is to ensure numerical stability and computational speed, especially for large orders. The approximation is motivated by the asymptotic equivalence of both estimates, see e.g. Chapter 8 in Brockwell and Davis (2009). The estimate is referred to as *negative maximum log-likelihood* in the text.

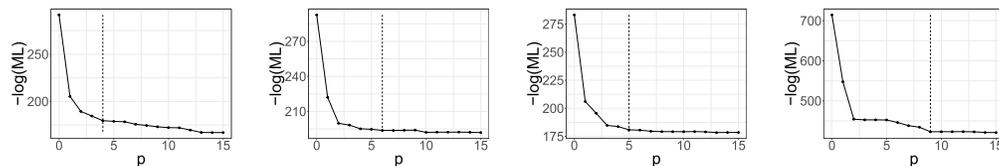


Figure 2: Negative maximum log likelihood for different  $AR(p)$  models applied to three realizations of the  $ARMA(1,1)$  model with  $a = 0.75, b = 0.8$  of length  $n = 128$  (first 3 plots from left) and the sunspot data (rightmost plot). The respective DIC-minimizing order is visualized by a black dashed line.

Figure 2 shows the scree-like plots for three exemplary  $ARMA(1,1)$  realizations from the above model as well as the sunspot data set. In all three realizations the elbow is at  $p = 1$  or  $p = 2$  while the DIC-criterion chooses much larger orders between 4 and 6. Table 1b clearly shows the superiority of the choice of fixed order  $p = 1$  over a DIC-based order selection for the nonparametric correction.

For the sunspot data that effect can also be seen clearly as the above procedure proposes to use  $p = 2$  in the nonparametric procedure while the DIC-criterion suggests  $p = 9$ . For a detailed discussion of this data analysis we refer to Section 4.3, similar effects for the LIGO data are discussed in Section 4.5.

### 4.3 Sunspot data

In this section, we analyse the yearly sunspot data from 1700 until 1987. We take the mean-centered version of the square root of the 288 observations as input data. We

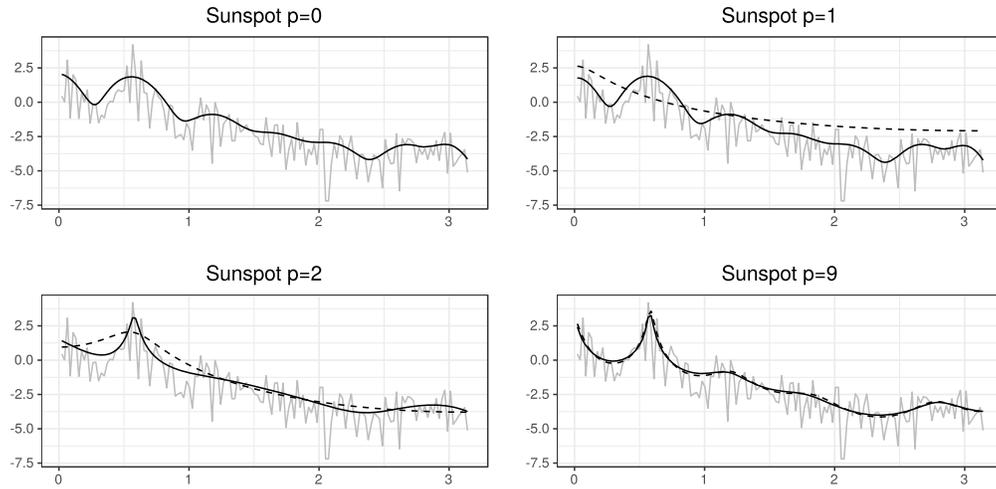


Figure 3: Posterior median spectral density estimates NPC (solid black) and AR (dashed black) for the transformed sunspot data on a logarithmic scale, for different autoregressive orders  $p$ . The log-periodogram is visualised in grey. For  $p = 0$ , NPC corresponds to NP.

compare the AR and the NPC procedure for fixed values  $p = 0, 1, 2, 9$ . Note that for  $p = 0$  the NPC estimate is the same as the pure nonparametric estimate (NP) based on the Whittle likelihood. While  $p = 9$  minimises the DIC,  $p = 2$  captures the main AR-features of the data as indicated by the elbow criterion (see Section 4.2 and Figure 2 (d)). The results are shown in Figure 3. While for  $p = 1$  the Bernstein polynomials of the nonparametric correction cannot yet capture the peaks sufficiently well, this is clearly the case for  $p = 2$  (the choice obtained from the elbow criterion): While the parametric model itself can clearly not yet explain the data well, enough features are captured to improve the nonparametric correction. For larger order choices, the estimate from the NPC method does not change much anymore, so that the correction does indeed not profit from additional parameters in the AR-model. In fact for  $p = 9$  (as indicated by DIC), the Bayes estimator of AR and NPC are very similar.

The posterior means of the parameter  $\eta$  were 0.42, 0.71, 0.72, and 0.75 for  $p = 0, 1, 2, 9$  respectively. As already noted in the simulation study, this demonstrates again that the parameter  $\eta$  can be interpreted as a model confidence parameter and adds robustness to the corrected likelihood procedure. With increasing autoregressive order  $p$  and thus improved fit of the parametric model,  $\eta$  increases. We also see that for  $p = 1$  the NPC estimate based on the AR(1) working model is identical to the NP estimate. The AR(1) spectral density clearly does not give an adequate fit which is reflected in the low value of  $\eta$ , thus giving only little weight to the parametric AR(1) working model in the NPC procedure. For  $p = 9$ , the AR(9) model fits well, the estimate of  $\eta$  is large and the NPC estimate dominated by the parametric AR(9) model.

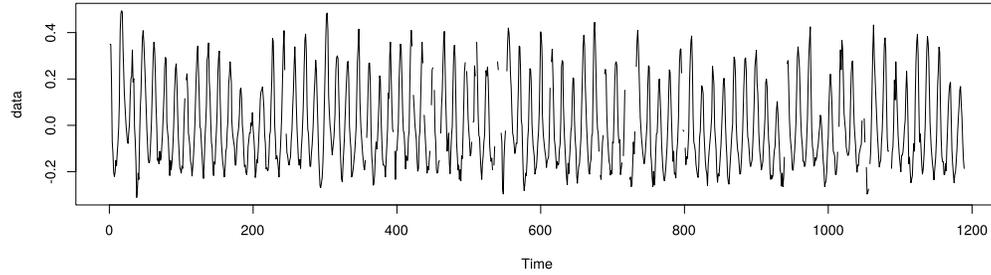


Figure 4: 1189 observations of light intensity from the *S. Carinae* variable star data after mean-centering and a square-root transformation.

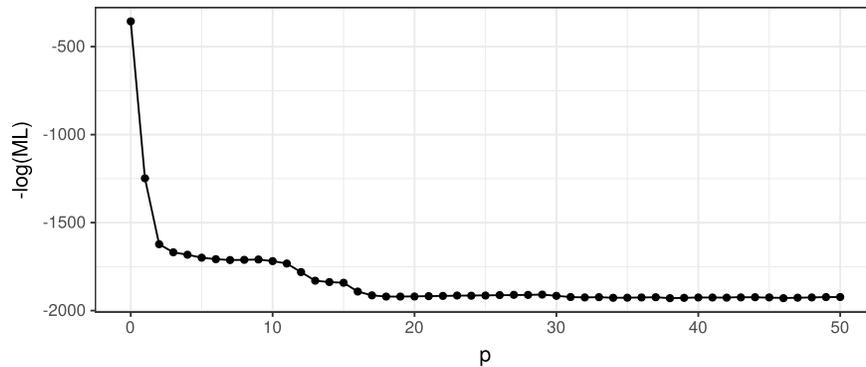


Figure 5: Negative maximum log likelihood for different  $AR(p)$  models applied to the *S. Carinae* data.

#### 4.4 Variable star *S. Carinae* data with missing values

As a further example, we re-analyze the time series from Baldwin and Thomson (1978) of light intensities of the *S. Carinae* variable star, that were previously analyzed by Carter and Kohn (1997) and Huerta and West (1999). The data set consists of 1189 observations of 10-day averages of light intensities over several years. The mean-centered and square-root-transformed observations are given in Figure 4 from which a periodicity of around 150 days can be discerned by visual inspection (fifteen ten-day means, corresponding to a frequency of 0.42). The data set contains 40 missing values.

The scree-type plot in Figure 5 shows elbows at orders  $p = 2$  and 17. Note that the Bayesian analysis by Huerta and West (1999) that kept the autoregressive model order variable gave high posterior probabilities to AR orders between 19 and 25. We compare results of the AR, NP, and NPC procedure for  $p = 2$  and  $p = 17$  in Figure 6.

The Bernstein polynomial prior results in very smooth estimates of the spectral density which do not approximate the major peaks visible in the periodogram. Again,

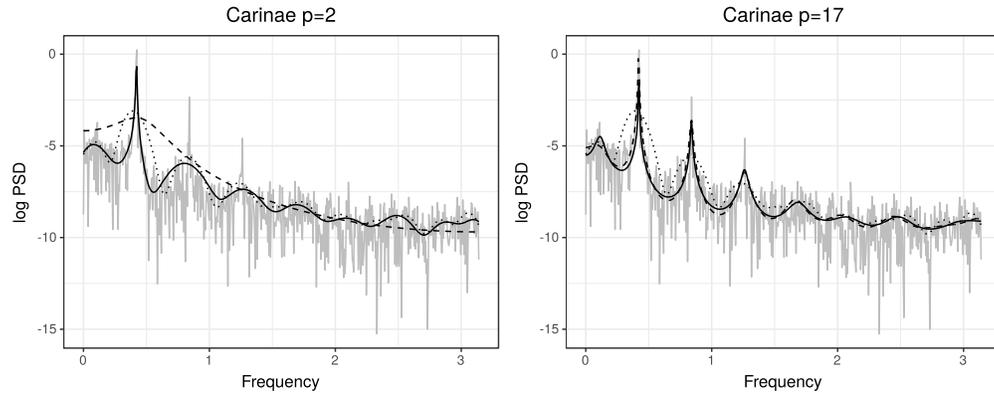


Figure 6: Posterior median spectral density estimates of NPC (solid black,  $p = 2$  (left) and  $p = 17$  (right)), AR(2) (dashed black), NP (dotted black) and periodogram (grey) for the *S. Carinae* data.

we see how the NPC estimate benefits from the fit of the parametric AR( $p$ ) model. An AR(2) model can only pick up the first major peak around 0.42 in the spectral density corresponding to a period of 149.8 days. The NPC estimate benefits from this and is able to considerably sharpen the first peak compared to the NP method. The AR(17) model provides a much better fit and picks up the subsequent peaks at multiples of 0.42, corresponding to harmonics of the fundamental frequency at 0.42. Furthermore, it shows a low amplitude peak near frequency 0.11 corresponding to a period of 568.4 days which was also found in Baldwin and Thomson (1978) and which mirrors the findings in Huerta and West (1999). Furthermore, Huerta and West (1999) estimated the posterior probability of a unitary root to be 0.87.

#### 4.5 Gravitational wave data

Gravitational waves, ripples in the fabric of spacetime caused by accelerating massive objects, were predicted by Albert Einstein in 1916 as a consequence of his general theory of relativity, see Einstein (1916). Gravitational waves originate from non-spherical acceleration of mass-energy distributions, such as binary inspiraling black holes, pulsars, and core collapse supernovae, propagating outwards from the source at the speed of light. However, they are very small (a thousand times smaller than the diameter of a proton) so that their measurement has provided decades of enormous engineering challenges.

On Sept. 14, 2015, the Advanced Laser Interferometric Gravitational Wave Observatory (Advanced LIGO), see Aasi et al. (2015b), made the first direct detection of a gravitational wave signal, GW150914, originating from a binary black hole merger (Abbott et al., 2016b). As of May 2018, five additional gravitational wave signals have been observed by the network of ground-based interferometers, which now includes Advanced Virgo in Cascina, Italy (Acernese et al., 2015). Four of these signals came from binary black hole mergers (Abbott et al., 2016a, 2017a,b,c), and one from a binary neutron

star merger with an electromagnetic counterpart (Abbott et al., 2017d), heralding a new era in astronomy. The two L-shaped LIGO instruments (in Hanford, Washington and Livingston, Louisiana) each consist of two perpendicular arms, each 4 kilometers long. A passing gravitational wave will alternately stretch one arm and squeeze the other, generating an interference pattern which is measured by photo-detectors. The detector output is a time series that consists of the time-varying dimensionless strain  $h(t)$ , the relative change in spacing between two test masses. The strain can be modelled as a deterministic gravitational wave signal  $h(t; \boldsymbol{\theta})$  depending on a vector  $\boldsymbol{\theta}$  of unknown waveform parameters plus additive noise  $n(t)$ , such that

$$y(t) = h(t; \boldsymbol{\theta}) + n(t), \quad t = 1, \dots, T.$$

There are a variety of noise sources at the LIGO detectors. This includes *seismic* noise, due to the motion of the mirrors from ground vibrations, earthquakes, wind, ocean waves, and vehicle traffic, *thermal* noise, from the microscopic fluctuations of the individual atoms in the mirrors and their suspensions, and *shot* noise, due to the discrete nature of light and the statistical uncertainty from the “photon counting” that is performed by the photo-detectors. In particular, Advanced LIGO noise includes high power, narrow band, spectral lines, visible as sharp peaks in the log-periodogram. As the Advanced LIGO spectrum is time-varying and subject to short-duration large-amplitude transient noise events, so-called “glitches”, a precise and realistic modelling and estimation of the noise component jointly with the signal is important for an accurate inference of the signal parameters  $\boldsymbol{\theta}$ . The current approach, which was also used for estimating the parameters of GW150914 in Abbott et al. (2016c), is to first use the Welch method (Welch, 1967) to estimate the spectral density from a separate stretch of data, close to but not including the signal and then to assume stationary Gaussian noise with this known spectral density in order to estimate the signal parameters.

Several approaches have been suggested in the recent gravitational wave literature to simultaneously estimate the noise spectral density and signal parameters. These include generalising the Whittle likelihood to a Student-t likelihood as in Röver et al. (2011), similarly modifying the likelihood to include additional scale parameters and then marginalising over the uncertainty in the PSD as in Littenberg et al. (2013), using cubic splines for smoothly varying broad-band noise and Lorentzians for narrow-band line features as in Littenberg and Cornish (2015), a Morlet-Gabor continuous wavelet basis for both gravitational wave burst signals and glitches as in Cornish and Littenberg (2015), the nonparametric approach of Choudhuri et al. (2004) using a Dirichlet-Bernstein prior (Edwards et al., 2015) and a generalisation of this using a B-spline prior, see Edwards et al. (2018).

We consider 1 s of real Advanced LIGO data collected during the sixth science run (S6)<sup>1</sup>, recoloured to match the target noise sensitivity of Advanced LIGO (Christensen, 2010). The data is differenced and then multiplied by a Hann window to mitigate spectral leakage. A low-pass Butterworth filter (of order 20 and attenuation 0.25) is then applied before downsampling from a LIGO sampling rate of 16384 Hz to 4096 Hz, reducing the volume of data. For reasons of computational complexity, the partial likelihood rather than the full likelihood of the working model is being used.

---

<sup>1</sup>The data is publicly available: <https://losc.ligo.org/S6/>

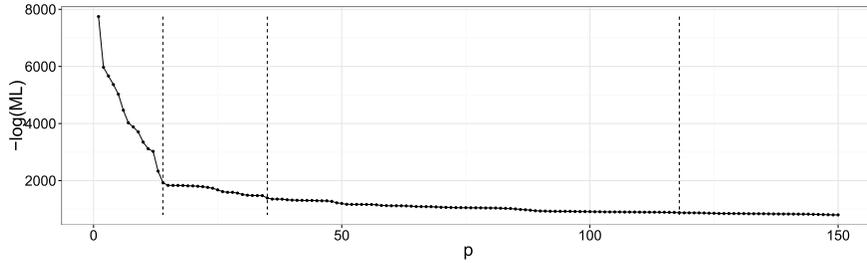


Figure 7: Negative maximum log likelihood for different  $AR(p)$  models applied to Advanced LIGO S6 data.

We first run a pure nonparametric model, corresponding to a nonparametrically corrected likelihood with an  $AR(0)$  working model (i.e. the Whittle likelihood) to estimate the spectral density. We then compare this to a nonparametrically corrected model with an order of  $p = 14$ , where a clear elbow can be seen in the negative log-likelihood plot (see Figure 7 and Remark 4.2). We run these simulations for 100,000 Markov Chain Monte Carlo (MCMC) iterations, with a burn-in of 50,000, and thinning factor of 5. Results are illustrated in Figure 8 (a).

Even though  $k$  converged to  $k \approx 900$  mixture components, it is clear that the Bernstein-Dirichlet prior together with the Whittle likelihood is not flexible enough to estimate the sharp peaks of the Advanced LIGO spectral density. The parametric  $AR(14)$  model (estimated using the Bayesian autoregressive sampler described in the supplementary material), captures the four main peaks but not their sharpness. Additionally, it does not capture the structure well in the frequency bands 0 to 450 Hz as well as larger than 1100 Hz. When compared to the  $AR(0)$  model, the nonparametrically corrected model based on  $p = 14$  estimates the sharp peaks much better. Furthermore, it sharpens all four peaks of the  $AR(14)$ -model (with a slight exception around 400 Hz, where seemingly two very sharp peaks overlap, a feature that is not captured by the  $AR(14)$  model at all). In the frequency bands 0 to 450 Hz as well as larger than 1100 Hz, where the parametric model fails altogether, the correction yields similar results to the nonparametric Whittle procedure. Similarly to the nonparametric Whittle procedure  $k$  tends towards  $k_{\max} = 800$  indicating that the Bernstein-Dirichlet prior together with an  $AR(14)$ -model is not yet flexible enough for this data set.

On the other hand, Bayesian Information Criterion (BIC) based on the Yule-Walker estimators as in Figure 7 – DIC ran into convergence problems for this data set – chooses an order of  $p = 118$  and the results can be seen in Figure 8 (c). For this order, the results of the purely autoregressive method and the nonparametric correction (based on the same order) are very similar which is in line with the discussion in Section 4.2 stating that both methods can be seen as a nonparametric estimation procedure. Both fit the peaks nicely at the expense of a possible overfitting in the high frequencies as well as between peaks. Combined with the observation from the simulation study that a smaller value of  $p$  typically results in a better fit and uncertainty quantification for the nonparametric correction, this leads us to take another look at Figure 7. In fact,

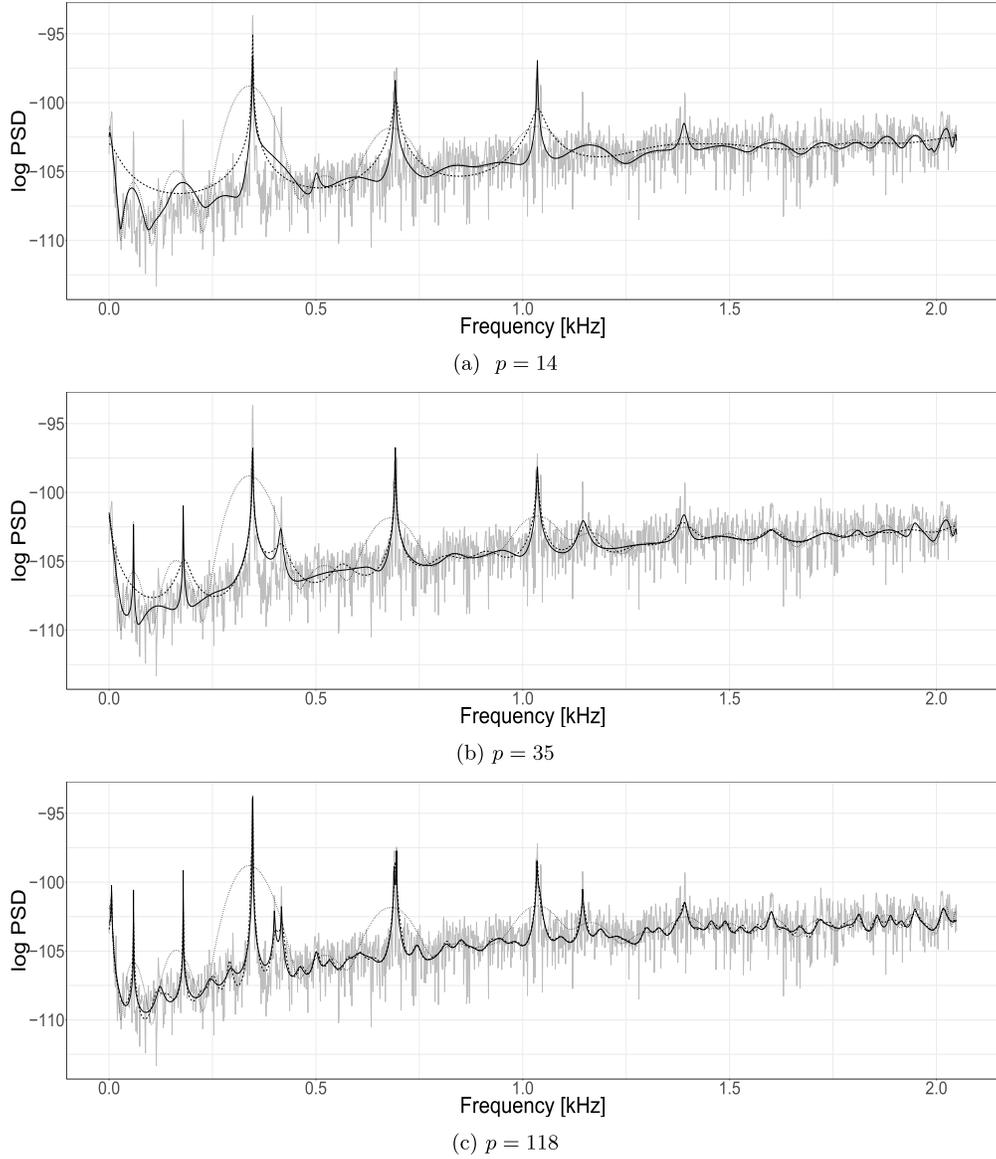


Figure 8: Estimated log spectral density for a 1 s segment of Advanced LIGO S6 data. The posterior median log spectral density estimate of NPC under an  $AR(p)$  working model (solid black),  $AR(p)$  (dashed black), and NPC under an  $AR(0)$  working model (dotted black) are overlaid with the log periodogram (grey), where (a) depicts  $p = 14$ , (b) depicts  $p = 35$  and (c) depicts  $p = 118$ .

the negative log-likelihood between the models AR(14) and AR(35) decreases by 533, which is not as sharp as the elbow at  $p = 14$ , but still significant – keeping in mind that the LIGO data is a very complex data set – much more so than the ARMA(1,1) or the sunspot data. There is a moderately sized jump before  $p = 35$ , while afterwards the descend slows down significantly. In fact, for the order chosen by BIC ( $p = 118$ ) the log-likelihood reaches the level of 879, showing that the difference between  $p = 14$  and  $p = 35$  (of 533) is comparable to the one between  $p = 35$  and  $p = 118$  (of 512). This indicates that there is indeed another change of gradient around  $p = 35$ . When looking at penalized likelihoods, this is also the point, where different penalizations start to obviously diverge. The results for  $p = 35$  can be found in Figure 8 (b).

The parametric AR(35) model already provides a reasonable fit to the periodogram, picking up the major peaks (with the exception around 100 Hz), but potentially under- and overestimates some of the peaks. In particular there are still major problems in the frequency bands 0 to 300 Hz and 400 to 700 Hz. The NPC procedure with  $p = 35$  keeps the peaks that have been captured well by the parametric model but corrects problems most prominently in the above mentioned frequency bands. Compared to the fit for  $p = 118$  the high-frequency spectrum as well as the spectrum between peaks is now much smoother, the only setback being around 400 Hz. It is worth mentioning that the correction works in several ways: Sharpening existing peaks (e.g. at 0 Hz), adding new peaks (e.g. at 100 Hz) as well as smoothing out some conceivably spurious peaks (e.g. at 600 Hz). Overall, the resulting estimate seems to capture the structure quite well in all frequency bands. This impression is complemented by the results of the NPC method with AR(35) working model together with the pointwise and uniform credible bounds obtained from the procedure in Figure 9.

It is important for astrophysicists in their searches of gravitational waves to understand whether the peaks in the noise spectrum are instrumental in origin, or are gravitational waves from astrophysical sources. So, effort is put into understanding the instrumental origin of these spectral lines and peaks, and engineering the detectors to keep the line bandwidth as narrow as possible. For instance, some of the peaks that are most noticeable to the eye are all believed to be due to known instrumental effects:

- 60 Hz power line harmonics, including ‘wings’ extending  $\pm 2$  Hz around multiples of 60 Hz, due to upconversion of low-frequency seismic noise. These arise, for example, due to imperfect electronic shielding, and magnetic coupling to the mirror suspensions.
- Thermally excited mirror suspension ‘violin modes’ at 329.5 Hz and in the 335–350 Hz range, and harmonics due to the numerous mirror suspension wires in the LIGO interferometers. These are an unfortunate but inevitable part of the LIGO detector design.
- calibration lines inserted by moving the end mirrors (using photons from lasers) or injecting phase modulations through particular electronic channels. Calibration lines during S6 were at 46.7, 393.1, and 1144.3 Hz.

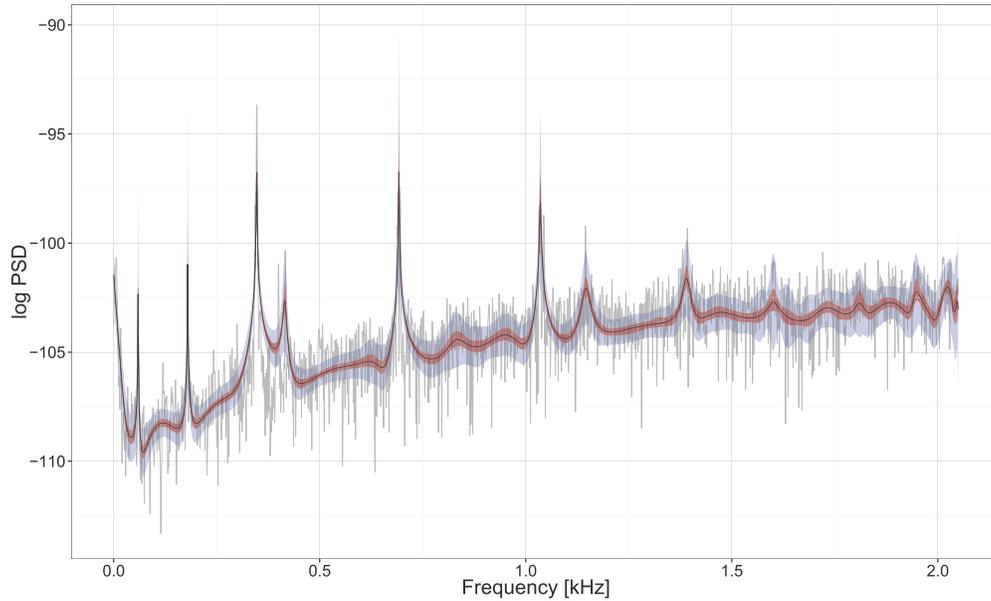


Figure 9: Estimated log spectral density for a 1 s segment of Advanced LIGO S6 data. The posterior median log spectral density estimate of NPC under an AR(35) working model (solid black), pointwise 0.9-credible region (shaded red), and uniform 0.9-credible region (shaded violet) are overlaid with the log periodogram (grey).

Note that continuous waves from non-axisymmetric spinning neutron stars also appear as extremely narrow spectral peaks. Their frequency is modulated by the Doppler shift due to the motion of the detectors with respect to the source. Because this is a small effect, we expect the pulsar produced signals to be buried in the noise, and not visible in the noise spectral densities of the detectors; matched filtering over the full observation time is required to identify potential signals in the data. In addition to spectral lines whose instrumental origin is identified, there are other peaks in the spectrum whose origin remains unknown. All of these features were rejected by the continuous wave searches conducted by the LIGO Scientific Collaboration, because they do not exhibit the expected Doppler modulation, may exhibit very different amplitude and frequency modulation, and/or are not seen with consistent amplitudes in other detectors, see also Coughlin et al. (2010) and Aasi et al. (2015a).

## 5 Conclusion

In this work we propose a nonparametric correction of a parametric likelihood to obtain an approximation of the true likelihood of a stationary time series that benefits from the main parametric features of the data. This approach extends the famous approximation by Whittle. For Gaussian data, the Whittle likelihood, the nonparametric correction as well as the true likelihood are asymptotically equivalent. Secondly, we propose a

Bayesian procedure for spectral density estimation which is based on this nonparametrically corrected likelihood. We show consistency of the resulting pseudo-posterior distribution for a fixed parametric likelihood. Furthermore, we present a Bayesian semi-parametric procedure that combines inference about the parametric working model with the nonparametric correction. The extent of the contribution of the parametric spectral density to the spectral density estimate is controlled by a shape confidence parameter. Simulation results have shown that this procedure inherits the benefits from the parametric working model if the latter is well-specified or describe a part of the features of the data well, while in the misspecified case the results are comparable to the usage of the Whittle likelihood.

Even though we have illustrated the nonparametrically corrected likelihood approach in this paper using *zero mean* stationary time series and the focus has been on spectral density estimation, this approach can be readily extended to more complex models where this zero mean time series constitutes the error term, such as linear or nonlinear dynamic regression models and these types of applications will be addressed in future research. Furthermore, regarding future work, it is interesting to investigate whether in the non-Gaussian case the class of time series for which asymptotically consistent inference holds can be enlarged by choosing an appropriate model. It is important to understand the distributional influence in the non-Gaussian case both in finite samples and asymptotically. As an example, in a bootstrap context, it suffices to capture the fourth order structure of a linear model to obtain asymptotically valid second-order frequentist properties of the autocovariance structure (Dahlhaus and Janas (1996); Kreiss and Paparoditis (2003)). As suggested by Kleijn et al. (2012) in a parametric setting, this property does not simply carry over to a Bayesian context. Preliminary results for non-Gaussian autoregressive time series however have shown considerable benefits (with respect to first and second order frequentist properties) when the innovation distribution is well-specified in comparison to a Gaussian model. Since any parametric likelihood is susceptible to misspecification, the ultimate goal is to consider a Bayesian nonparametric model for the innovation distribution, such as Dirichlet mixtures of normals. Further directions for future work are automation of the elbow criterion discussed in Remark 4.2. Instead of choosing a fixed order in advance, an automation might as well serve as a guideline for specifying a prior on the AR order parameter, which can be included in the inference by means of Reversible Jump Markov Chain Monte Carlo (RJMCMC) (cf. Remark 3).

## Supplementary Material

Supplementary material to ‘Beyond Whittle: Nonparametric Correction of a Parametric Likelihood with a Focus on Bayesian Time Series Analysis’ (DOI: [10.1214/18-BA1126SUPP](https://doi.org/10.1214/18-BA1126SUPP); .pdf). The electronic supplement contains the proofs for this paper, some comments on the Bayesian autoregressive sampler, the full conditional distribution of missing values as well as some additional simulation results.

## References

- Aasi, J., the LIGO Scientific Collaboration, and the Virgo Collaboration (2015a). “Characterization of the LIGO detectors during their sixth science run.” *Classical and Quantum Gravity*, 32: 115012. [1065](#)
- Aasi, J. et al. (2015b). “Advanced LIGO.” *Classical and Quantum Gravity*, 32: 074001. [1060](#)
- Abbott, B. P. et al. (2016a). “GW151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence.” *Physical Review Letters*, 116: 241103. [MR3707758](#). doi: <https://doi.org/10.1103/PhysRevLett.116.061102>. [1060](#)
- Abbott, B. P. et al. (2016b). “Observation of gravitational waves from a binary black hole merger.” *Physical Review Letters*, 116: 061102. [1060](#)
- Abbott, B. P. et al. (2016c). “Properties of the binary black hole merger GW150914.” *Physical Review Letters*, 116: 241102. [MR3707756](#). doi: <https://doi.org/10.1103/PhysRevLett.116.131102>. [1061](#)
- Abbott, B. P. et al. (2017a). “GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2.” *Physical Review Letters*, 118: 221101. [1060](#)
- Abbott, B. P. et al. (2017b). “GW170608: Observation of a 19 Solar-mass Binary Black Hole Coalescence.” *The Astrophysical Journal Letters*, 851: L35. [1060](#)
- Abbott, B. P. et al. (2017c). “GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence.” *Physical Review Letters*, 119: 141101. [1060](#)
- Abbott, B. P. et al. (2017d). “Multi-messenger Observations of a Binary Neutron Star Merger.” *The Astrophysical Journal Letters*, 848: L12. [1061](#)
- Acernese, F. et al. (2015). “Advanced Virgo: A second-generation interferometric gravitational wave detector.” *Classical and Quantum Gravity*, 32(2): 024001. [1060](#)
- Andrews, D. W. K. and Monahan, J. C. (1992). “An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator.” *Econometrica*, 60(4): 953–966. [MR1168742](#). doi: <https://doi.org/10.2307/2951574>. [1039](#)
- Baldwin, A. and Thomson, P. (1978). “Periodogram analysis of S. Carinae.” *Royal Astronomical Society of New Zealand*, 6: 31–35. [1059](#), [1060](#)
- Barnett, G., Kohn, R., and Sheather, S. (1996). “Bayesian estimation of an autoregressive model using Markov chain Monte Carlo.” *Journal of Econometrics*, 74(2): 237–254. [MR1411599](#). doi: [https://doi.org/10.1016/0304-4076\(95\)01744-5](https://doi.org/10.1016/0304-4076(95)01744-5). [1052](#)
- Bauwens, L., Lubrano, M., and Richard, J.-F. (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press. [MR1858176](#). [1038](#)
- Bollerslev, T. (1986). “Generalized autoregressive conditional heteroskedasticity.” *Journal of Econometrics*, 31(3): 307–327. [MR0853051](#). doi: [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1). [1038](#)

- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2013). *Time series analysis: Forecasting and control*. John Wiley & Sons. MR2419724. doi: <https://doi.org/10.1002/9781118619193>. 1038
- Brockwell, P. J. and Davis, R. A. (2009). *Time series: Theory and methods*. Springer. MR1093459. doi: <https://doi.org/10.1007/978-1-4419-0320-4>. 1041, 1042, 1045, 1046, 1048, 1051, 1057
- Cadonna, A., Kottas, A., and Prado, R. (2017). “Bayesian mixture modeling for spectral density estimation.” *Statistics & Probability Letters*, 125: 189–195. MR3626085. doi: <https://doi.org/10.1016/j.spl.2017.02.008>. 1039
- Carter, C. 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. MR1436567. doi: <https://doi.org/10.1111/1467-9868.00067>. 1038, 1059
- 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. MR3173730. doi: <https://doi.org/10.1080/10618600.2013.785293>. 1038, 1041
- 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. MR2109494. doi: <https://doi.org/10.1198/016214504000000557>. 1038, 1040, 1046, 1048, 1049, 1050, 1053, 1061
- Christensen, N. (2010). “LIGO S6 detector characterization studies.” *Classical and Quantum Gravity*, 27: 194010. 1061
- Contreras-Cristán, A., Gutiérrez-Peña, E., and Walker, S. G. (2006). “A note on Whittle’s likelihood.” *Communications in Statistics – Simulation and Computation*, 35(4): 857–875. MR2291357. doi: <https://doi.org/10.1080/03610910600880203>. 1038
- Cornish, N. J. and Littenberg, T. B. (2015). “Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches.” *Classical and Quantum Gravity*, 32: 135012. 1061
- Costa, M. J., Finkenstädt, B., Roche, V., Ladvi, F., Gould, P. D., Foreman, J., Halliday, K., Hall, A., and Rand, D. A. (2013). “Inference on periodicity of circadian time series.” *Biostatistics*, 14(4): 792–806. 1038
- Coughlin, M., the LIGO Scientific Collaboration, and the Virgo Collaboration (2010). “Noise Line Identification in LIGO S6 and Virgo VSR2.” *Journal of Physics: Conference Series*, 243: 012010. 1065
- Dahlhaus, R. and Janas, D. (1996). “A frequency domain bootstrap for ratio statistics in time series analysis.” *The Annals of Statistics*, 24(5): 1934–1963. MR1421155. doi: <https://doi.org/10.1214/aos/1069362304>. 1050, 1066

- Durbin, J. and Koopman, S. (2012). *Time series analysis by state space methods*. 38. Oxford University Press. MR3014996. doi: <https://doi.org/10.1093/acprof:oso/9780199641178.001.0001>. 1038
- Edwards, M. C., Meyer, R., and Christensen, N. (2015). “Bayesian semiparametric power spectral density estimation with applications in gravitational wave data analysis.” *Physical Review D*, 92: 064011. 1061
- Edwards, M. C., Meyer, R., and Christensen, N. (2018). “Bayesian nonparametric spectral density estimation using B-spline priors.” *Statistics and Computing*. DOI: 10.1007/s11222-017-9796-9. 1061
- Einstein, A. (1916). “Approximative integration of the field equations of gravitation.” *Sitzungsberichte Preußischen Akademie der Wissenschaften*, 1916 (Part 1): 688–696. 1060
- Emmanoulopoulos, D., McHardy, I., and Papadakis, I. (2013). “Generating artificial light curves: Revisited and updated.” *Monthly Notices of the Royal Astronomical Society*, 433(2): 907–927. 1038
- Engle, R. (1982). “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.” *Econometrica: Journal of the Econometric Society*, 987–1007. MR0666121. doi: <https://doi.org/10.2307/1912773>. 1038
- Fan, J. and Yao, Q. (2002). *Nonlinear time series*, volume 2. Springer. MR1964455. doi: <https://doi.org/10.1007/b97702>. 1038
- Franke, J. and Härdle, W. (1992). “On bootstrapping kernel spectral estimates.” *The Annals of Statistics*, 121–145. MR1150337. doi: <https://doi.org/10.1214/aos/1176348515>. 1038, 1050
- Fuller, W. (1996). *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. Wiley. MR1365746. 1054
- Gangopadhyay, A., Mallick, B., and Denison, D. (1999). “Estimation of spectral density of a stationary time series via an asymptotic representation of the periodogram.” *Journal of Statistical Planning and Inference*, 75(2): 281–290. MR1678977. doi: [https://doi.org/10.1016/S0378-3758\(98\)00148-7](https://doi.org/10.1016/S0378-3758(98)00148-7). 1038
- Gelman, A. (2006). “Prior distributions for variance parameters in hierarchical models.” *Bayesian Analysis*, 1(3): 515–533. MR2221284. doi: <https://doi.org/10.1214/06-BA117A>. 1051
- Ghosal, S., Ghosh, J., and Ramamoorthi, R. (1999). “Posterior consistency of Dirichlet mixtures in density estimation.” *Annals of Statistics*, 27: 143–158. MR1701105. doi: <https://doi.org/10.1214/aos/1018031105>. 1049
- Green, P. J. (1995). “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.” *Biometrika*, 82(4): 711–732. MR1380810. doi: <https://doi.org/10.1093/biomet/82.4.711>. 1052

- Häfner, F. and Kirch, C. (2016). “Moving Fourier analysis for locally stationary processes with the bootstrap in view.” Preprint. MR3714115. doi: <https://doi.org/10.1111/jtsa.12241>. 1053
- Härdle, W., Horowitz, J., and Kreiss, J.-P. (2003). “Bootstrap methods for time series.” *International Statistical Review*, 71(2): 435–459. MR2906626. doi: <https://doi.org/10.1016/j.jkss.2011.07.002>. 1038
- Härdle, W., Lütkepohl, H., and Chen, R. (1997). “A review of nonparametric time series analysis.” *International Statistical Review*, 65(1): 49–72. MR1433592. 1038
- Hermansen, G. H. (2008). “Bayesian nonparametric modelling of covariance functions, with application to time series and spatial statistics.” Ph.D. thesis, Universitetet i Oslo. 1038
- Hidalgo, J. (2008). “Specification testing for regression models with dependent data.” *Journal of Econometrics*, 143(1): 143–165. MR2384437. doi: <https://doi.org/10.1016/j.jeconom.2007.08.013>. 1038
- Hjort, N. L. and Glad, I. K. (1995). “Nonparametric density estimation with a parametric start.” *The Annals of Statistics*, 882–904. MR1345205. doi: <https://doi.org/10.1214/aos/1176324627>. 1039, 1044
- Hjort, N. L., Holmes, C. C., Müller, P., and Walker, S. G. (2010). “Bayesian nonparametrics.” *AMC*, 10: 12. MR2722987. doi: <https://doi.org/10.1017/CB09780511802478>. 1038
- Huerta, G. and West, M. (1999). “Bayesian inference on periodicities and component spectral structure in time series.” *Journal of Time Series Analysis*, 20: 401–416. MR1711369. doi: <https://doi.org/10.1111/1467-9892.00145>. 1059, 1060
- Hurvich, C. M. and Zeger, S. (1987). *Frequency domain bootstrap methods for time series*. New York University, Graduate School of Business Administration. 1038
- Jentsch, C. and Kreiss, J.-P. (2010). “The multiple hybrid bootstrap - Resampling multivariate linear processes.” *Journal of Multivariate Analysis*, 101(10): 2320–2345. MR2719865. doi: <https://doi.org/10.1016/j.jmva.2010.06.005>. 1039
- Jentsch, C., Kreiss, J.-P., Mantalos, P., and Paparoditis, E. (2012). “Hybrid bootstrap aided unit root testing.” *Computational Statistics*, 27(4): 779–797. MR3041857. doi: <https://doi.org/10.1007/s00180-011-0290-0>. 1039
- Kim, Y. M. and Nordman, D. J. (2013). “A frequency domain bootstrap for Whittle estimation under long-range dependence.” *Journal of Multivariate Analysis*, 115: 405–420. MR3004567. doi: <https://doi.org/10.1016/j.jmva.2012.10.018>. 1038
- Kirch, C. (2007). “Resampling in the frequency domain of time series to determine critical values for change-point tests.” *Statistics & Decisions*, 25(3/2007): 237–261. MR2412072. doi: <https://doi.org/10.1524/stnd.2007.0902>. 1038
- Kirch, C. and Politis, D. N. (2011). “TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain.” *Annals of Statistics*, 39(3): 1427–1470. MR2850208. doi: <https://doi.org/10.1214/10-AOS868>. 1038

- Kirch, C., Edwards, M. C., Meier, A., and Meyer, R. (2018). “Supplementary material to ‘Beyond Whittle: Nonparametric Correction of a Parametric Likelihood with a Focus on Bayesian Time Series Analysis’.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/18-BA1126SUPP>. 1040
- Kleijn, B. J. K., van der Vaart, A. W., et al. (2012). “The Bernstein-von-Mises theorem under misspecification.” *Electronic Journal of Statistics*, 6: 354–381. MR2988412. doi: <https://doi.org/10.1214/12-EJS675>. 1066
- Kreiss, J.-P. and Lahiri, S. N. (2012). “Bootstrap Methods for Time Series.” *Handbook of Statistics*, 30: 3–26. 1038
- Kreiss, J.-P. and Paparoditis, E. (2003). “Autoregressive-aided periodogram bootstrap for time series.” *Annals of Statistics*, 31(6): 1923–1955. MR2036395. doi: <https://doi.org/10.1214/aos/1074290332>. 1039, 1044, 1050, 1066
- Kreiss, J.-P. and Paparoditis, E. (2011). “Bootstrap methods for dependent data: A review.” *Journal of the Korean Statistical Society*, 40(4): 357–378. MR2906623. doi: <https://doi.org/10.1016/j.jkss.2011.08.009>. 1038
- Kreiss, J.-P. and Paparoditis, E. (2012). “The hybrid wild bootstrap for time series.” *Journal of the American Statistical Association*, 107(499): 1073–1084. MR3010895. doi: <https://doi.org/10.1080/01621459.2012.695664>. 1039
- Kreiss, J.-P., Paparoditis, E., and Politis, D. N. (2011). “On the range of validity of the autoregressive sieve bootstrap.” *The Annals of Statistics*, 2103–2130. MR2893863. doi: <https://doi.org/10.1214/11-AOS900>. 1041, 1050, 1056
- Le Cam, L. and Yang, G. . (1960). *Asymptotics in Statistics*. Springer. MR1784901. doi: <https://doi.org/10.1007/978-1-4612-1166-2>. 1049
- Liseo, B. and Macaro, C. (2013). “Objective priors for causal AR (p) with partial autocorrelations.” *Journal of Statistical Computation and Simulation*, 83(9): 1613–1628. MR3169260. doi: <https://doi.org/10.1080/00949655.2012.667409>. 1051
- Liseo, B., Marinucci, D., and Petrella, L. (2001). “Bayesian semiparametric inference on long-range dependence.” *Biometrika*, 88(4): 1089–1104. MR1872221. doi: <https://doi.org/10.1093/biomet/88.4.1089>. 1038
- Littenberg, T. B. and Cornish, N. J. (2015). “Bayesian inference for spectral estimation of gravitational wave detector noise.” *Physical Review D*, 91: 084034. 1061
- Littenberg, T. B., Coughlin, M., Farr, B., and M., F. W. (2013). “Fortifying the characterization of binary mergers in LIGO data.” *Physical Review D*, 88: 084044. 1061
- McCulloch, R. E. and Tsay, R. S. (1994). “Bayesian analysis of autoregressive time series via the Gibbs sampler.” *Journal of Time Series Analysis*, 15(2): 253–250. MR1263893. doi: <https://doi.org/10.1111/j.1467-9892.1994.tb00188.x>. 1052
- Meier, A., Kirch, C., Edwards, M. C., and Meyer, R. (2017). *beyondWhittle: Bayesian Spectral Inference for Stationary Time Series*. R package. 1039, 1052

- Müller, P. and Mitra, R. (2013). “Bayesian nonparametric inference—why and how.” *Bayesian Analysis (Online)*, 8(2). MR3066939. doi: <https://doi.org/10.1214/13-BA811>. 1038
- Nam, C., Aston, J., and Johansen, A. (2014). “Parallel sequential Monte Carlo samplers and estimation of the number of states in a Hidden Markov Model.” *Annals of the Institute of Statistical Mathematics*, 66: 553–575. MR3211874. doi: <https://doi.org/10.1007/s10463-014-0450-4>. 1051
- Neumann, M. H. and Polzehl, J. (1998). “Simultaneous bootstrap confidence bands in nonparametric regression.” *Journal of Nonparametric Statistics*, 9(4): 307–333. MR1646905. doi: <https://doi.org/10.1080/10485259808832748>. 1053
- Petrone, S. (1999). “Random Bernstein polynomials.” *Scandinavian Journal of Statistics*, 26(3): 373–393. MR1712051. doi: <https://doi.org/10.1111/1467-9469.00155>. 1040, 1046
- Roberts, G. O. and Rosenthal, J. S. (2009). “Examples of adaptive MCMC.” *Journal of Computational and Graphical Statistics*, 18(2): 349–367. MR2749836. doi: <https://doi.org/10.1198/jcgs.2009.06134>. 1052
- 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. MR2985940. doi: <https://doi.org/10.1214/11-AOS955>. 1038
- Röver, C., Meyer, R., and Christensen, N. (2011). “Modelling coloured residual noise in gravitational-wave signal processing.” *Classical and Quantum Gravity*, 28: 015010. MR2763397. doi: <https://doi.org/10.1088/0264-9381/28/1/015010>. 1061
- Schuster, A. (1898). “On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena.” *Terrestrial Magnetism*, 3(1): 13–41. 1038
- Sethuraman, J. (1994). “A constructive definition of Dirichlet priors.” *Statistica Sinica*, 639–650. 1047
- Shao, X. and Wu, B. W. (2007). “Asymptotic spectral theory for nonlinear time series.” *Annals of Statistics*, 35(4): 1773–1801. MR2351105. doi: <https://doi.org/10.1214/009053606000001479>. 1042
- Sørbye, S. H. and Rue, H. (2016). “Penalised complexity priors for stationary autoregressive processes.” *arXiv preprint arXiv:1608.08941*. MR3714116. doi: <https://doi.org/10.1111/jtsa.12242>. 1051
- Steel, M. F. J. (2008). “The New Palgrave Dictionary of Economics, chapter Bayesian time series analysis.” 1038
- Sykulski, A., Olhede, S., Lilly, J., Guillaumin, A., and Early, J. (2017). “The de-biased Whittle likelihood.” *arXiv preprint arXiv:1605.06718v2*. 1039

- Szabó, B., van der Vaart, A., van Zanten, J., et al. (2015). “Frequentist coverage of adaptive nonparametric Bayesian credible sets.” *The Annals of Statistics*, 43(4): 1391–1428. MR3357861. doi: <https://doi.org/10.1214/14-AOS1270>. 1053
- Van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3. Cambridge University Press. MR1652247. doi: <https://doi.org/10.1017/CB09780511802256>. 1050
- Walker, S. (2013). *Bayesian Theory and Applications*, chapter Bayesian nonparametrics, 1–34. Oxford University Press. MR3221167. doi: <https://doi.org/10.1093/acprof:oso/9780199695607.003.0013>. 1038
- Wasserman, L. (2006). *All of nonparametric statistics*. Springer. MR2172729. 1038
- Welch, P. D. (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: 70–73. 1061
- Whittle, P. (1957). “Curve and periodogram smoothing.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 38–63. MR0092331. 1038, 1041

#### Acknowledgments

This work was supported by DFG grant KI 1443/3-1. Some of the preliminary research was conducted while Claudia Kirch was at KIT, where her position was financed by the Stifterverband für die deutsche Wissenschaft by funds of the Claussen-Simon-trust. The research was initiated during a visit of Renate Meyer at Karlsruhe Institute of Technology (KIT), which was financed by the German academic exchange service (DAAD). The authors would like to thank the New Zealand eScience Infrastructure (NeSI) and the Universitätsrechenzentrum (URZ) Magdeburg for their high performance computing facilities, the Centre for eResearch at the University of Auckland and Jörg Schulenburg for their technical support, and Nelson Christensen for sharing his insights into the LIGO data. Renate Meyer was also supported by the James Cook Research Fellowships from Government funding, administered by the Royal Society Te Apārangi.