Bayesian spectral density estimation using P-splines

Patricio Maturana-Russel · Renate Meyer

Abstract This article a Bayesian approach to estimating the spectral density of a stationary time series using a prior based on a mixture of P-spline distributions. Our proposal is motivated by the B-spline Dirichlet process prior of Edwards et al. (2019) in combination with Whittle's likelihood and aims at reducing the high computational complexity of its posterior computations. The strength of the B-spline Dirichlet process prior over the Bernstein-Dirichlet process prior of Choudhuri et al. (2004) lies in its ability to estimate spectral densities with sharp peaks and abrupt changes due to the flexibility of B-splines with variable number and location of knots. Here, we suggest to use P-splines of Eilers and Marx (1996) that combine a B-spline basis with a discrete penalty on the basis coefficients. We demonstrate in a simulation study and two real case studies that this approach retains the flexibility of the B-splines, achieves the same ability to accurately estimate peaks due to a new data-driven knot allocation scheme but significantly reduces the computational costs.

Patricio Maturana-Russel
Department of Statistics, University of Auckland, Auckland, New Zealand
E-mail: p.russel@auckland.ac.nz

Renate Meyer
Department of Statistics, University of Auckland, Auckland, New Zealand

1 Introduction

The power spectral density (psd), or simply spectral density, describes the distribution of the power or variance over the individual frequency components of a time series. It embodies useful information for the study of stationary time series. For a zero-mean weakly stationary time series, an absolutely summable autocovariance function, i.e. $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, guarantees the existence of a continuous and bounded spectral density function given by

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp\left(-ih\lambda\right),$$

where $-\pi < \lambda \leq \pi$ is the angular frequency.

Methods to estimate this function range from parametric to nonparametric approaches. The former are mainly based on fitting autoregressive moving average (ARMA) models. However, the corresponding spectral density will be biased if the ARMA model does not adequately describe the dependence structure of the time series. This is a downside of all parametric approaches. They are efficient if the model fits well but because they make restrictive assumptions about the data-generating mechanism, their inference is sensitive

to model misspecifications and may be severely biased. Bayesian nonparametric approaches, on the other hand, put a prior distribution on the set of all possible densities that could have generated the data not just a small parametric subset thereof. They thus avoid the problem of choosing between different parametric models by an intrinsic data-driven determination of the model complexity.

As the periodogram fluctuates around the true spectral density, most nonparametric methods for spectral density estimation hinge on smoothing the periodogram

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} Y_t \exp\left(-it\lambda\right) \right|^2,$$

in one way or another, where $-\pi < \lambda \leq \pi$ is the angular frequency and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ is a stationary time series of length n. The periodogram ordinates evaluated at the Fourier frequencies λ_i are asymptotically independent and follow an exponential distribution with mean $f(\lambda_i)$. Thus, they are unbiased estimates of $f(\lambda_i)$ but not consistent (Brockwell and Davis, 1991).

Most Bayesian approaches to psd estimation use the so-called Whittle likelihood, introduced by (Whittle, 1957) as an approximation to the likelihood of a Gaussian stationary time series. Based on the limiting independent exponential distribution of the periodogram values, the Whittle likelihood for a mean centred weakly stationary time series $\boldsymbol{Y}$ of length n is defined as

$$L(\boldsymbol{Y}|f) = \prod_{l=1}^{\nu} \frac{1}{f(\lambda_l)} e^{-I_n(\lambda_l)/f(\lambda_l)}$$
$$\propto \exp\left(-\sum_{l=1}^{\nu} \log f(\lambda_l) + \frac{I_n(\lambda_l)}{f(\lambda_l)} \right), \quad (1)$$

where $\lambda_l = 2\pi l/n$ are the Fourier frequencies, $f(\cdot)$ spectral density of $\boldsymbol{Y}$, and $\nu = \lfloor (n-1)/2 \rfloor$ is the greatest integer value less than or equal to $(n-1)/2$. Advantages of this approximation are that it depends directly on the spectral density, unlike the true Gaussian likelihood function, and gives a good approximation for non-Gaussian time series under certain conditions as shown by Shao and Wu (2007).

Bayesian nonparametric prior specifications for the psd include Carter and Kohn (1997) who used a smoothing prior on the log psd. Rosen et al.

(2012) decompose the log psd into a linear and nonlinear component and put a Gaussian process prior on the nonlinear part, Pensky et al. (2007) propose Bayesian wavelet-smoothing of the log psd and recently, and Cadonna et al. (2017) modelled the log psd as a mixture of Gaussian distributions with frequency-dependent weights and mean functions. Gangopadhyay et al. (1999) modelled the log-spectral density function by piecewise polynomials of low order between strategically placed knots on the support of the psd, allowing the data to select the number and location of the knots. A reversible jump Markov chain Monte Carlo (RJM-CMC; Green, 1995) algorithm was implemented to sample from the posterior distribution. Similarly, frequentist approaches such as Rodrguez-lvarez et al. and Wood and Fasiolo (2017) model the log spectral density by a linear combination of splines with variable knots and combine these with adaptive smoothing.

Penalized splines have been used for a time-domain analysis of time series data e.g. to model the mean as a smooth function of time by Krivobokova et al. (2006) or for fitting and forecasting univariate nonlinear time series by Wegener and Kauermann (2017), but not in the frequency domain for estimating the spectral density. The P-spline prior on the spectral density function that we are going to propose is related to the Bernstein-Dirichlet process (BDP) prior introduced by Choudhuri et al. (2004) and combined with the Whittle likelihood. The BDP prior is a mixture of beta densities with weights induced by a Dirichlet process. The number of components, which is a smoothing parameter, is given a discrete prior. Edwards et al. (2019) extended the method by replacing the beta densities with B-spline densities. Similar to Gangopadhyay et al. (1999), the specification of the number and location of the B-spline densities is data-driven, however it avoids a RJMCMC algorithm by putting a second Dirichlet process prior on the cdf that induces the inter-knot distances. Unlike the beta densities which have global support on the interval [0,1], B-spline densities have local support, thereby increasing their flexibility and allowing better modelling of spectral densities with sharp peaks and abrupt changes. As demonstrated in Edwards et al. (2019), the B-spline–Dirichlet process (BspDP) prior outperforms the BDP prior in estimating complex psd's.

However, this flexibility comes at a high computational price.

With the aim of reducing the computational cost of the BspDP prior approach, we suggest to use the idea of P-splines of Eilers and Marx (1996) that combines a large but fixed number of B-splines with a simple difference penalty that controls the degree of smoothness of the spectral density. P-splines, i.e. equally-spaced B-splines, have been successfully implemented for frequentist non-parametric regression but to the best of our knowledge not yet for Bayesian spectral density estimation. We suggest two novel Bayesian approaches in this paper, both based on the Whittle likelihood and on a mixture of B-spline densities. Both approaches fix the number of B-spline densities. The first approach uses equally spaced knots with a P-spline prior which we define in Section III. We demonstrate that this approach yields significant savings in computational time without sacrificing accuracy over the BspDP prior approach for simple spectral structures. The second approach has a particular focus on spectral densities with sharp peaks. Instead of an equidistant spacing, the knots are spread based on the distribution of peaks in the periodogram. Thus the knot allocation is also data-driven as with the BspDP prior but the knots remain fixed once allocated. This novel knot allocation scheme avoids the computational complexity of the BspDP prior. We show that with this knot allocation scheme, the estimates of spectral densities with spikes and abrupt changes improves significantly without an increase in computing time. ======= With the aim of reducing the computational cost of the BspDP prior approach, we suggest to use the idea of P-splines Eilers and Marx (1996) that combines a large but fixed number of B-splines with a simple difference penalty that controls the degree of smoothness of the spectral density. P-splines, i.e. equally-spaced B-splines, have been successfully implemented for frequentist non-parametric regression but to the best of our knowledge not yet for Bayesian spectral density estimation. We suggest two novel Bayesian approaches in this paper, both based on the Whittle likelihood and on a mixture of B-spline densities. Both approaches fix the number of B-spline densities. The first approach uses equally spaced knots with a P-spline prior which we define in Section III. We demonstrate that this approach yields significant

savings in computational time without sacrificing accuracy over the BspDP prior approach for simple spectral structures. The second approach has a particular focus on spectral densities with sharp peaks. Instead of an equidistant spacing, the knots are spread based on the distribution of peaks in the periodogram. Thus the knot allocation is also data-driven as with the BspDP prior but the knots remain fixed once allocated. This novel knot allocation scheme avoids the computational complexity of the BspDP prior. We show that with this knot allocation scheme, the estimates of spectral densities with spikes and abrupt changes improves significantly without an increase in computing time.

The paper is organized as follows: In order to set notation and allow for a direct comparison, we start by reviewing the B-spline densities and the BspDP prior in Section II. Section III specifies the P-spline prior for the psd, the novel knot allocations scheme and the Gibbs sampler used to sample from the posterior distribution based on the P-spline prior and Whittle likelihood. In Section IV, we test our P-spline-based approaches to spectral density estimation and compare results to those based on the BspDP prior in a simulation study and the application to two real time series. The paper concludes with a discussion of the relative merits of the new approaches.

2 Notation and review of the B-spline–Dirichlet process prior

By an application of the Weierstrass theorem it can be seen that a mixture of beta densities with only integer parameters can uniformly approximate any continuous density on the interval $[0, 1]$ (Choudhuri et al., 2004). Let G be a cumulative distribution function (cdf), with continuous density g, then

$$\widehat{g}(\omega) = \sum_{k=1}^{K} G\left(\frac{k-1}{K}, \frac{k}{K}\right] \beta(\omega; k, K-k+1) \quad (2)$$

$$= \sum_{k=1}^{K} w_k \beta(\omega; k, K-k+1) \quad (3)$$

converges uniformly to $g(\omega)$, where $G(u, v] = G(v) - G(u)$, $\beta(\omega; a, b)$ is the beta density with parameters a and b, and $w_k = G\left((k-1)/K, k/K\right]$

are the weights of the mixture. The Bernstein polynomial prior, which is used to describe a nonparametric prior for probability densities in the unit interval, is based on this approximation (Petrone, 1999a,b). Choudhuri et al. (2004) proposed it as a prior on the spectral density function.

As shown by Perron and Mengersen (2001) for distributions on the interval [0,1], the mixture of B-spline distributions has better approximation properties than the mixture of beta distributions for a fixed number of components. In particular, the approximation error of the mixture of B-spline distributions can be made arbitrarily small by finding suitable knot locations and increasing the number of knots. This work motivated Edwards et al. (2019) to propose the BspDP prior on the spectral density function which is based on the representation (2) but with the beta densities replaced by B-spline densities as defined in the following.

B-splines and BspDP Prior

A spline of degree $r+1$ is a function defined piecewise by polynomials of degree $\leq r$, which meet at points called *knots* where the function is continuous. Any of these spline functions can be described by basis functions, known as *B-splines*, in other words, all spline functions can be represented as a unique combination of B-splines with the same order and over the same partition. Without loss of generality, the global domain of interest is assumed to be the unit interval $[0, 1]$.

The B-spline basic functions of any order can be recursively defined as

$$B_{k,0}(\omega, \boldsymbol{\xi}) = \begin{cases} 1, \omega \in [\xi_{k-1}, \xi_k] \\ 0, \text{otherwise} \end{cases}$$

$$B_{k,r}(\omega, \boldsymbol{\xi}) = v_{k,r} B_{k,r-1}(\omega; \boldsymbol{\xi}) + \\ (1 - v_{k+1,r}) B_{k+1,r-1}(\omega; \boldsymbol{\xi}),$$

where

$$v_{k,r} = \begin{cases} \dfrac{\omega - \xi_{k-1}}{\xi_{k+r-1} - \xi_{k-1}}, \xi_{k-1} \neq \xi_{k+r-1} \\ 0, \qquad\qquad \text{otherwise} \end{cases}$$

and

$$\boldsymbol{\xi} = \{0 = \xi_0 = \xi_1 = \cdots = \xi_r \leq \xi_{r+1} \leq \cdots \leq \xi_K \\ = \xi_{K+1} = \cdots = \xi_{K+r} = 1\}$$

is the knot vector. This vector is a non-decreasing sequence that contains $K + r + 1$ knots, which can be divided in $2r$ external and $K - r + 1$ internal knots. The latter must be $\geq r$.

The *B-spline densities*, i.e. normalized B-spline functions, are defined by

$$b_{k,r}(\omega; \boldsymbol{\xi}) = (r + 1) \frac{B_{k,r}(\omega; \boldsymbol{\xi})}{\xi_{k+r} - \xi_{k-1}}.$$

The BspDP prior proposed by Edwards et al. (2019) involves two Dirichlet processes, one placed on G inducing the weights of the B-spline densities and the other on the knot spacings. In practice, these are implemented using the stick-breaking representation (Sethuraman, 1994), for details see Edwards et al. (2019). For computational reasons, this infinite series representation of G and H is truncated at a large but finite positive integer, L_G and L_H, respectively. These truncations control the quality of the approximations. Longer series improve the approximation, but increase the number of calculations and consequently the computation time.

As shown by Edwards et al. (2019), the approach based on the BspDP prior outperforms the BDP prior in estimating spiky spectral densities. In the case of smooth spectral densities, there is no a significant difference in the performance of these two approaches. However, in practice, the true psd function is unknown, which makes it necessary to use methods such as the BspDP that will work well in all conceivable cases.

To implement the BDP prior, one only has to calculate the beta densities once and they can be stored and re-utilized across MCMC steps. On the other hand, the BspDP prior requires the calculation of the B-spline densities at each iteration, since these vary due to varying knot locations. Moreover, $2 \times (L_G + L_H)$ calculations are needed in the truncated stick-breaking representation of the two Dirichlet processes at each iteration resulting in a significantly increased complexity of the BspDP algorithm.

This has prompted us to explore the performance of two new algorithms, both based on the P-spline prior algorithm. Both use a fixed number of B-spline densities and knot locations, avoiding the Dirichlet processes and the recalculation of the B-spline densities. Whereas the first approach uses equidistant knots, the second algorithm proposes

a new fast knot allocation strategy. Both methods preserve the flexibility of the B-spline densities and their potential ability to estimate spectral densities with sharp peaks and abrupt changes.

3 P-spline prior

As the spectral density function $f(\cdot)$ is defined on the interval $[0, \pi]$, it is reparametrized and defined as

$$f(\pi\omega) = \tau \times s_r(\omega; \boldsymbol{w}, \boldsymbol{\xi}), \quad \omega \in [0, 1],$$

where $\tau = \int_0^1 f(\tau\omega)\mathrm{d}\omega$ is the normalizing constant and

$$s_r(\omega; \boldsymbol{w}, \boldsymbol{\xi}) = \sum_{k=1}^{K} w_k b_{k,r}(\omega; \boldsymbol{\xi}), \qquad (4)$$

with *fixed* number K of B-spline densities $b_{k,r}(\cdot)$ of fixed degree r, weight vector $\boldsymbol{w} = (w_1, \cdots, w_K)$, and knot sequence $\boldsymbol{\xi}$.

$\boldsymbol{\xi}$ contains fixed equidistant internal knots on $[0, 1]$ in our basic P-spline prior. A different allotment of knots is described in section "Adaptive knots".

Smoothing splines in the frequentist context are based on approximating a given function by a linear combination of B-splines and putting a penalty on the integrated squared second derivative as a measure of roughness. P-splines avoid derivatives by expressing the penalty as the sum of squares of differences of the B-spline coefficients. In a Bayesian context, this penalty can be transformed into a prior distribution for the dth order difference of successive coefficients, see e.g. Lang and Brezger (2004).

In our approach, the B-spline coefficients w_k are weights, so positive and sum to one. We therefore reparametrize to the vector $\boldsymbol{v}$ with

$$v_k = \log\left(\frac{w_k}{1 - \sum_{k=1}^{K-1} w_k}\right).$$

After some calculations, it can be shown that the weights are given by

$$w_k = \frac{e^{v_k}}{1 + \sum_{k=1}^{K-1} e^{v_k}}.$$

The last weight is defined as $w_K = 1 - \sum_{k=1}^{K-1} w_k$, thus they all sum to 1. Then, an indirect prior is placed on the weights via

$$\boldsymbol{v}|\phi \sim \mathrm{N}_{K-1}(\boldsymbol{0}, (\phi\boldsymbol{P})^{-1})$$
$$\phi|\delta \sim \mathrm{Gamma}(\alpha_\phi, \delta\beta_\phi)$$
$$\delta \sim \mathrm{Gamma}(\alpha_\delta, \beta_\delta)$$

where $\boldsymbol{v} = (v_1, \ldots, v_{K-1})^\top$ is a $K - 1$ dimensional parameter vector, ϕ is the smoothing or penalty parameter, $\mathbf{P} = \mathbf{D}^\top\mathbf{D} + \epsilon\mathbf{I}_{K-1}$ is the penalty matrix (which is a full matrix rank matrix for any small quantity ϵ, for instance, 10^{-6}) with $\mathbf{D}$ the dth order difference matrix $(d = 1, 2, \ldots, K-2)$, α_ϕ and α_δ are shape parameters, $\delta\beta_\phi$ and β_δ are rate parameters, and $\mathrm{Gamma}(a, b)$ denotes a gamma distribution with mean a/b and variance a/b^2. We used first and second order difference penalties in the simulation study. The 1st order difference matrix $\mathbf{D} \in \mathbb{R}^{(K-2)\times(K-1)}$ is defined as

$$\begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & & 0 \\ 0 & -1 & 1 & 0 & \cdots & & 0 \\ \vdots & & \ddots & \ddots & & & \vdots \\ 0 & & & & \cdots & 0 & -1 & 1 \end{bmatrix} \qquad (5)$$

and the 2nd order difference matrix $\in \mathbb{R}^{(K-3)\times(K-1)}$ as

$$\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ 0 & & & & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}. \qquad (6)$$

The rationale for using discrete second order roughness penalties as an approximation to the usual continuous roughness measure based on the integral over the second derivatives is given in Eilers and Marx (1996).

The parameter vector for the P-spline prior is $\boldsymbol{\theta} = (\boldsymbol{v}^\top, \phi, \delta, \tau)^\top$. We use a robust specification for the prior distribution of the penalty parameters as suggested by Jullion and Lambert (2007) by choosing small values for α_δ and β_δ, for instance, 10^{-4}. The choice of α_ϕ and β_ϕ do not affect the spectral density estimate. Here we set α_ϕ and β_ϕ equal to 1 as used by Bremhorst and Lambert (2016). The scale parameter τ is given by an inverse $\mathrm{Gamma}(\alpha_\tau, \beta_\tau)$ prior. These are also the prior specifications used in the simulations and examples in Section 4.

Posterior Computation

The joint posterior distribution of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{v}^\top, \phi, \delta, \tau)^\top$ is given by

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto L(\boldsymbol{Y}|f) \times p(\boldsymbol{\theta})$$
$$= L(\boldsymbol{Y}|f) \times p(\boldsymbol{v}|\phi, \delta) \times p(\phi|\delta) \times p(\delta) \times p(\tau)$$

where $L(\boldsymbol{Y}|f)$ denotes the Whittle likelihood defined in (1). This posterior is proper since all the prior distributions are proper. We use the Gibbs sampler to sample from the joint posterior distribution.

Sampling from the full conditional posterior distributions of ϕ, δ, and τ can be performed directly. These are given by

$$\phi|\boldsymbol{Y}, \delta, \boldsymbol{v} \sim \text{Gamma}\left(\frac{K-1}{2} + \alpha_\phi, \frac{1}{2}\boldsymbol{v}^\top \mathbf{P} \boldsymbol{v} + \delta\beta_\phi\right),$$
$$\delta|\boldsymbol{Y}, \phi \sim \text{Gamma}\left(\alpha_\phi + \alpha_\delta, \beta_\phi\phi + \beta_\delta\right), \text{ and}$$
$$\tau|\boldsymbol{Y}, \delta, \phi, \boldsymbol{v} \sim \text{Inv-Gamma}\left(\alpha_\tau + \nu, \sum_{l=1}^{\nu} \frac{I_n(\lambda_l)}{f(\lambda_l)} + \beta_\tau\right).$$

However, to sample from the full conditional distribution of $\boldsymbol{v}$, we use the Metropolis algorithm. In order to improve the mixing of the chains, we apply the Metropolis algorithm on a reparametrized posterior distribution (Lambert, 2007). For this, we need a pilot posterior sample for $\boldsymbol{v}$ from which we calculate its mean vector $\overline{\boldsymbol{v}}$ and covariance matrix $\mathbf{S}$. Then we define the following re-parametrization

$$\boldsymbol{v} = \mathbf{S}^{1/2}\boldsymbol{\beta} + \overline{\boldsymbol{v}},$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{K-1})^\top$.

We can update $\boldsymbol{v}$ by modifying $\boldsymbol{\beta}$ according to a proposal distribution. In this work, we propose a univariate proposal value β_k^* to update $\boldsymbol{v}$ given by

$$\beta_k^* = \beta_k + \sigma z,$$
$$\boldsymbol{v}^* = \mathbf{S}^{1/2}\boldsymbol{\beta}^* + \overline{\boldsymbol{v}},$$

where $\boldsymbol{\beta}^* = (\beta_1, \ldots, \beta_k^*, \ldots, \beta_{K-1})^\top$, $z \sim$ Normal$(0, 1)$, and σ controls the length of the proposals. We vary σ across iterations in order to get an acceptance rate between 0.3 and 0.5. Note that even though the proposal is univariate on $\boldsymbol{\beta}$, it is multivariate on $\boldsymbol{v}^*$. Alternatively to the Metropolis step used for $\boldsymbol{v}$, a data augmentation step based on Pólya-Gamma auxiliary variables can be used which would allow for a full Gibbs sampler (Polson et al., 2013).

The mth cycle of the Gibbs sampler is given by

– Draw $\boldsymbol{v}^m$ from $p(\boldsymbol{v}|\boldsymbol{\beta}^{m-1}, \phi^{m-1}, \delta^{m-1}, \tau^{m-1})$ using $K-1$ univariate Metropolis steps for β_k, with $k = 1, \ldots, K-1$, according to the reparametrization described above;
– Draw ϕ^m from

Gamma$\left(\frac{K-1}{2} + \alpha_\phi, \frac{1}{2}\boldsymbol{v}^{m\top}\mathbf{P}\boldsymbol{v}^m + \delta^{m-1}\beta_\phi\right)$;

– Draw δ^m from Gamma $\left(\alpha_\phi + \alpha_\delta, \beta_\phi\phi^m + \beta_\delta\right)$;
– Draw τ^m from

Inv-Gamma$\left(\alpha_\tau + \nu, \sum_{l=1}^{\nu} \frac{I_n(\lambda_l)}{f(\lambda_l)} + \beta_\tau\right)$.

The influence of the penalty is determined by ϕ. The larger this value is, the smoother is the resulting estimate. It is interesting to note the role of the penalty matrix $\mathbf{P}$ in the rate parameter on the full conditional posterior distribution of ϕ. When the quadratic form $\boldsymbol{v}^\top \mathbf{P} \boldsymbol{v}$ tends to infinity, this distribution becomes concentrated at 0^+. This limiting behaviour yields rougher results. Note that in this case the prior rate $\delta\beta_\phi$ becomes irrelevant in the full conditional posterior density. On the other hand, when $\boldsymbol{v}^\top \mathbf{P} \boldsymbol{v}$ tends to 0^+, the distribution favours large ϕ values, yielding smoother estimates. The magnitude of this quadratic form is controlled by the penalty order and the number of B-spline densities. In general, large order penalties produce lower values for this quantity, causing smoother results. In this way the penalty matrix penalizes the inclusion of B-spline densities.

Ruppert (2002) explained that the degree of smoothness implied by a certain choice of prior parameters also depends on the scale of the responses $\boldsymbol{Y}$. To avoid the problem, the author suggested to standardize $\boldsymbol{Y}$ before the sampling process and apply the inverse operation afterwards. This also allows to avoid numerical problems in the MCMC process. We follow his proposal.

Choosing the number of B-splines

The number of B-spline densities plays a critical role in the model fit and may result in under- or oversmoothing. Even though the penalty parameter controls the smoothness of the fit, a large number of B-splines yields rougher and a low number smoother spectral density estimates. There is a smallest sufficient number which results in a psd estimate that fits the features of the data whereas

choosing a larger number will have an insignificant effect on the fit. Unexpectedly, Ruppert (2002) found some cases in which too many B-splines degrades the fitting in terms of mean square error. Unfortunately, there is no consensus in how to find the optimal value of the number of B-splines.

Eilers et al. (2015) consider the use of $K = 100$ B-splines a wise choice, unless computational constraints are evident. Some algorithms that explore different number of knots K^* in a trial sequence and find an optimal value with respect to a certain goodness-of-fit criterion have been proposed. Ruppert (2002) suggested a full-search and myopic algorithms which are based on the generalized cross-validation statistic. However, their results are not conclusive and this criterion only applies if the penalty parameter is treated as a tuning constant (Kauermann and Opsomer, 2011). Likhachev (2017) also considered a knot sequence and proposed the selection of the number of knots based on Akaike's, corrected Akaike's and the Bayesian Information Criterion.

Ruppert (2002) discussed a heuristic rule of thumb $K^* = \min\{n/4, 40\}$, which is simpler and, in our experience, seems to work very well in general. This criterion allocates a knot between $\max\{4, \lfloor n/40 \rfloor\}$ observations, where $\lfloor \cdot \rfloor$ stands for the floor function. We employ this approach in the application Section for selecting the number of B-spline densities.

Adaptive knots

The knots are often equally spaced in P-spline methods. This scheme works quite well for smooth psds without any abrupt changes. However, for spectral densities with spikes, the algorithm will require a large number of knots to adequately estimate the peaks. The B-spline prior method is able to handle abrupt changes because the knot location is variable and driven by the data. We propose an adaptive method that allocates the knots according to the periodogram but once allocated, the knots stay fixed. The idea is to concentrate more effort in those regions that have potential peaks as detected by the periodogram.

Our adaptive knot location proposal works as follow. We calculate the periodogram and apply a square root transformation in order to have more regular magnitudes and eliminate a potential trend. Second, we standardize these values and apply the absolute value function. We treat this transformed periodogram as a density function $\widetilde{f}$ and calculate its distribution function F. Finally, the knots are allocated according to the quantiles of the inverse distribution function of $\widetilde{f}$. Thus, in those areas in which the periodogram has sharp and abrupt changes, our procedure will assign more knots in proportion to their magnitudes. Hereafter, we will refer to this scheme as *adaptive knots*.

The procedure can be summarized as follows:

1. Let $\boldsymbol{x} = \left(\sqrt{I_n(\lambda_1)}, \ldots, \sqrt{I_n(\lambda_\nu)}\right)$ be the square root of the periodogram values for each angular frequency.

2. Define $z_i = \left|\dfrac{x_i - \bar{x}}{S_x}\right|$ where $\bar{x}$ denotes the average and S_x the standard deviation of $\boldsymbol{x}$ and calculate its distribution function $F_z = P(z \leq z_i)$.

3. Let $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_{K^*}^*)$ be a vector of K^* equally spaced values in $[0, 1]$.

4. The adaptive knot vector is given by $\boldsymbol{\xi} = \left(F_z^{-1}(\xi_1^*), \ldots, F_z^{-1}(\xi_{K^*}^*)\right)$.

For non-equidistant knots, it does not make sense any longer to base the definition of the covariance matrix of the Gaussian prior on the difference matrices as defined in (5) and (6) because this is an adequate approximation of the derivative-based roughness measures – defined as the integral of the first and second order derivatives of the P-splines – for equally spaced knots only. The definition of the difference matrices could be adjusted using divided differences, but here we use the derivative-based penalties as described by Wood (2017). We make use of the implementation of the derivative-based penalty matrix by the function `bsplinepen` in the R-package `fda`. This penalty matrix contains the inner products of the respective first and second order derivatives of the basis functions. In addition, we normalise this matrix by dividing it by the maximum absolute column sum.

The adaptive knot allocation can be regarded as an empirical Bayes prior since the data are used to specify the prior distribution. However, only the locations of the B-splines are determined by the data, not the parameters of the prior distribution which in this case are the weights w_k associated

with each B-spline $b_{k,r}$. The main criticism of an empirical Bayes approach is that the data are being used twice, once to specify the prior and then again to update the prior to the posterior. Here, the data are used only once to specify the knot locations but not again to update these through the likelihood.

4 Application

We first assess the properties of our P-spline spectral density estimates in a simulation study and compare them to estimates based on the BspDP prior. We furthermore apply all approaches to the analysis of the classical sunspot dataset and the *S. Carinae* variable star light intensities, previously analysed by Carter and Kohn (1997); Huerta and West (1999); Kirch et al. (2018). As the psd of the *S. Carinae* time series contains sharp peaks, its analysis allows to assess the impact of the periodogram-spaced knots in contrast to the equidistant knots. For all the analyses, we use cubic B-splines ($r = 3$).

4.1 Simulation study

The setup of our simulation study mirrors that in Edwards et al. (2019) who compared spectral density estimates based on the BspDP and the BDP priors. We generated 300 autoregressive time series of order 1 and 4 with unit variance Gaussian innovations of length $n = \{128, 256, 512\}$. For the AR(1) model, the first order autocorrelation $\rho_1 = 0.9$ was chosen, which has a relatively simple spectral density (see Figure 1). The AR(4) time series with $\rho_1 = 0.9$, $\rho_2 = -0.9$, $\rho_3 = 0.9$ and $\rho_4 = -0.9$ have a spectral density with two large peaks (see Figure 2).

We estimated the spectral density functions using the BspDP and P-spline priors. The BspDP-based analysis was performed via the `gibbs_bspline` function in the R package `bsplinePsd` Edwards et al. (2018). The algorithm was executed for 100,000 iterations with a burn-in period of 25,000 and thinning factor of 10, resulting in 7,500 samples used for posterior inference. Pilot analyses showed that these specifications are suitable in these examples in terms of convergence (results not shown).

The P-spline analysis was performed using the `gibbs_pspline` function implemented in the R package `psplinePsd` Russel and Meyer (2020). A total of 100,000 MCMC samples were generated using a pilot run of 20,000 iterations with a burn-in period of 5,000 and thinning factor of 10 to calibrate the proposals. The final sample consisted of 80,000 samples with a burn-in period of 5,000 and thinning factor of 10, resulting in 7,500 samples used for posterior inferences. This procedure was replicated for both equidistant and adaptive knots considering first and second order penalties.

The theoretical spectral density for an AR(p) model is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \sum_{j=1}^{p} \rho_j \exp(-i\lambda)|^2},$$

where σ^2 is the variance of the white noise innovations and $(\rho_1, \ldots, \rho_p)$ are the model parameters. In order to compare the psd estimates to the true function, we use the integrated absolute error (IAE) or L_1-error which is defined as

$$\text{IAE} = ||\widehat{f} - f||_1 = \int_0^\pi |\widehat{f}(\omega) - f(\omega)| d\omega,$$

where $\widehat{f}(\cdot)$ is the pointwise posterior median of the spectral density $f(\cdot)$. The IAE is computed for each of the 300 replications and its median is calculated. The results are displayed in Table 1.

For the AR(1) time series, the P-spline prior with equidistant knots and penalty order $d = 1$ yields better psd estimates than the BspDP prior. A larger penalty order does not necessarily result in more accurate psd estimates. In this case of a relatively smooth spectral density, the adaptive knots do not yield an improvement over the equidistant knots and the BspDP prior.

For the AR(4) time series, the fit based on the P-spline prior with adaptive knots is better than the fit based on the B-spline Dirichlet process prior and the one based on the P-spline prior with equidistant knots. In general, the BspDP algorithm achieves higher accuracy than the P-spline method with equidistant knots. As shown by Edwards et al. (2019), the posterior distribution based on the BspDP prior yields better results than posterior based on the BDP in terms of median IAE in this example. Consequently, our P-spline algorithm with adaptive knots outperforms the BDP method as well.

	$n = 128$	$n = 256$	$n = 512$
AR(1)			
B-spline	0.825	0.769	0.609
Equidistant knots			
P-spline $d = 1$	0.772	0.644	0.527
P-spline $d = 2$	0.708	0.604	0.629
Adaptive knots			
P-spline $d = 1$	0.874	0.783	0.610
P-spline $d = 2$	0.874	0.815	0.640
AR(4)			
B-spline	3.019	2.250	1.832
Equidistant knots			
P-spline $d = 1$	2.914	2.389	2.307
P-spline $d = 2$	3.150	2.569	2.388
Adaptive knots			
P-spline $d = 1$	2.495	1.830	1.489
P-spline $d = 2$	2.696	2.003	1.677

Table 1 Median IAE for the psd estimates using the BspDP and the P-spline priors for simulated AR(1) and AR(4) time series with different penalty orders (d) and knot location schemes for the P-splines.
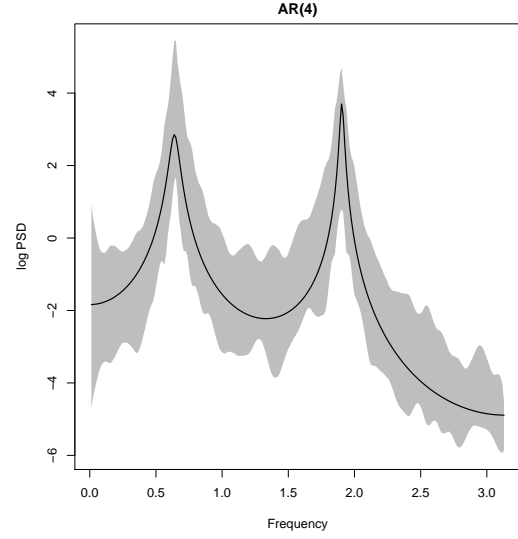

Fig. 2 Estimated log-spectral density for an AR(4) time series via the P-spline prior algorithm with a first order penalty. The solid line stands for the true psd, whereas grey area is the uniform 90% credible band.

Fig. 1 Estimated log-spectral density for an AR(1) time series via the P-spline prior algorithm with a first order penalty. The solid line stands for the true psd, whereas grey area is the uniform 90% credible band.

The IAE measures the accuracy of the point estimates of the spectral density. In the following, we will also compare the coverage of posterior uniform credible bands. We define the $100(1 - \alpha)\%$ uniform credible band as

$$\widehat{f}(\lambda) \pm \zeta_\alpha \times \mathrm{mad}\left(\widehat{f}_i(\lambda)\right), \ \ \lambda \in [0, \pi],$$

where $\widehat{f}(\lambda)$ is the pointwise posterior median spectral density, $\mathrm{mad}(\widehat{f}_i(\lambda))$ is the median absolute deviation of the posterior samples $\widehat{f}_i(\lambda)$ and ζ_α is such that

$$\mathrm{P}\left(\max\left\{\frac{|\widehat{f}_i(\lambda) - \widehat{f}(\lambda)|}{\mathrm{mad}(\widehat{f}_i(\lambda))}\right\} \leq \zeta_\alpha\right) = 1 - \alpha.$$

The idea is to measure the proportion of times that the true spectral density is fully contained by the uniform credible band in the 300 replications. The results are displayed in Table 2. In addition, we calculate the proportion of times that the uniform credible band contains the value of the true psd at each of the Fourier frequencies. The median of these pointwise coverage proportions are displayed in Table 3.

For the AR(1) model, the BspDP and first order penalty P-spline 90% credible bands cover the whole true spectral density function, i.e. for all frequencies λ, in almost all the 300 independent analyses. In general, the latter method shows diminishing performance as the penalty degree increases. For the AR(4) model, the P-spline method with equidistant knots performs poorly. On the other hand, the adaptive knot scheme yields evidently much better results, which are comparable (in the case of the first order penalty) to those obtained via the BspDP algorithm. For the P-spline priors,

	$n = 128$	$n = 256$	$n = 512$
AR(1)			
B-spline	1.000	1.000	1.000
Equidistant			
P-spline $d = 1$	1.000	1.000	0.997
P-spline $d = 2$	0.993	1.000	0.993
Adaptive knots			
P-spline $d = 1$	0.997	1.000	0.997
P-spline $d = 2$	0.973	0.740	0.770
AR(4)			
B-spline	0.933	0.967	0.897
Equidistant knots			
P-spline $d = 1$	0.527	0.353	0.007
P-spline $d = 2$	0.323	0.167	0.007
Adaptive knots			
P-spline $d = 1$	0.790	0.847	0.813
P-spline $d = 2$	0.337	0.387	0.270

Table 2 Proportion of times that the true psd is entirely encapsulated within the 90% uniform credible band of the 300 estimates.

	$n = 128$	$n = 256$	$n = 512$
AR(1)			
B-spline	1.000	1.000	1.000
Equidistant &			
Adaptive knots			
P-spline $d = 1, 2$	1.000	1.000	1.000
AR(4)			
B-spline	1.000	1.000	1.000
Equidistant knots			
P-spline $d = 1$	1.000	0.992	0.984
P-spline $d = 2$	0.984	0.984	0.980
Adaptive knots			
P-spline $d = 1$	1.000	1.000	1.000
P-spline $d = 2$	0.984	0.992	0.992

Table 3 Median of the pointwise proportions that the true psd at each Fourier frequency is contained within the 90% uniform credible bands.

	$n = 128$	$n = 256$	$n = 512$
AR(1)			
B-spline	29.497	32.423	37.896
Adaptive knots			
P-spline $d = 1$	4.859	7.450	9.606
P-spline $d = 2$	4.538	6.863	8.783
B-spline/P-spline	6.278	4.531	4.122
AR(4)			
B-spline	30.111	34.126	40.875
Adaptive knots			
P-spline $d = 1$	4.883	7.453	9.665
P-spline $d = 2$	4.543	6.887	8.827
B-spline/P-spline	6.389	4.760	4.421

Table 4 Median run-times in minutes and relative run-times with respect to an average of the P-spline run-times.

the coverage probabilities decrease as the penalty degree order increases, which can be explained by the forced smoothness imposed by the higher penalty order. As shown by Edwards et al. (2019), the BspDP algorithm produces better results than the BDP prior in the AR(4) case, which coverage probability is zero across $n = \{128, 256, 512\}$. Consequently, our P-spline algorithm with adaptive knots outperforms the BDP method in this case.

Even though the P-spline method with equidistant knots for the AR(4) case performs poorly in terms of uniform coverage, i.e. the proportion of times that the credible band covers the true psd *completely* in the 300 runs, it is important to note that this is only because the uniform band does not cover the psd for just very few frequencies. As is evident from Table 3, only for a small fraction of the frequencies the psd is not covered by the uniform credible band. In this context, BspDP and adaptive knot P-spline algorithms perform equally well.

As Edwards et al. (2019) noted, one of the drawbacks of the BspDP prior is its computational complexity relative to the BDP. This is directly reflected in the run-time, which is approximately 2-3 times higher in this example (see Table 3 in Edwards et al. (2019)). The median run-time in our analyses for the BspDP prior and the P-spline with adaptive knots are displayed in Table 4. Note that

the P-spline method based on equidistant knots has similar run-times to the adapative knots. The P-spline method is approximately 4-7 times faster than BspDP approach in these examples. This is due to a reduced number of calculations per iteration because the B-spline densities are calculated only once at the start of the algorithm and there are no Dirichlet process approximations via the stick-breaking method.

In general terms, the P-spline algorithm with our novel knot allocation scheme, the periodogram-spaced knots, offers the best trade-off between accuracy and computation cost. The rule-of-thumb criterion, $K = \min\{n/4, 40\}$, for the number of B-spline densities in the P-spline prior algorithm worked perfectly well in this example.

4.2 Sunspot data analysis

Sunspots are regions of reduced surface temperature that are visible as darker spots on the sun's photosphere. We analyse the average annual mean sunspot numbers for the year 1700-1987 consisting of 288 observations. This is a classic dataset used to assess spectral density estimation methods. A square root transformation is applied in order to make the observations more symmetrical and stationary.

As in Edwards et al. (2019), the MCMC algorithm based on the BspDP prior is run for 100,000 iterations, with a burn-in period of 50,000 and thinning factor of 10. Thus, 5,000 samples are used for the estimation. This took about 32 minutes. The other specifications are the same as those used in the simulation study.

The MCMC algorithm based on the P-spline prior is run in two stages. In the first stage, the proposal is calibrated for the final run. It consists of 25,000 iterations with a burn-in period of 10,000 and thinning factor of 10, resulting in 1,500 samples. Then the algorithm is run for 75,000 iterations with a burn-in period of 25,000 and thinning factor of 10, resulting in 5,000 samples. The analysis is performed using the first order penalty for equidistant and adaptive knots. All other specifications are the same as those used in the simulation study. A single analysis takes about 6 minutes.

The results for the BspDP and P-spline prior with first order penalty and adaptive knots are displayed in Figure 3. Both methods reveal a large peak at about 0.09. In other words, there is a periodic solar cycle of length $1/0.09 \approx 11$ years. These results are consistent with those obtained via the BDP prior (Choudhuri et al., 2004). However, the BspDP estimation takes approximately 32 minutes compared to only 6 minutes for the the P-spline estimation. Thus the P-spline estimation is approximately five times more efficient.

4.3 Variable star S. Carinae data analysis

This dataset contains 1189 visual observations of the *S. Carinae*, a variable star in the southern hemisphere sky. These are daily observations collected by the Royal Astronomical Society of New Zealand, which correspond to 10 day averages of light intensities over several years. This data set
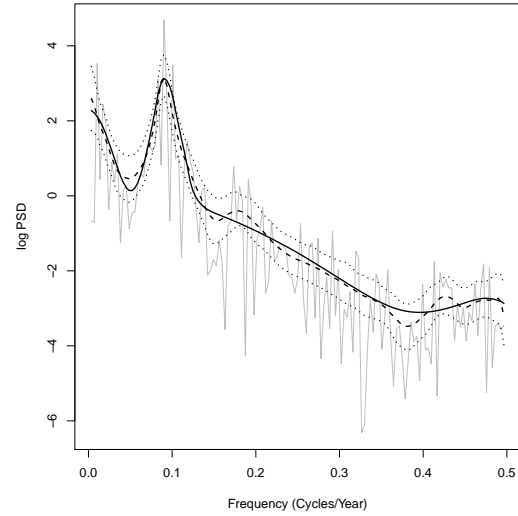

Fig. 3 Log-spectral density estimate for the sunspot data. The continuous grey line represents the periodogram. The continuous and dashed lines stand for the posterior median log psd obtained using the B-spline and P-spline priors, respectively. The dot lines represents the 90% pointwise credible bands for the latter prior, which consider a first order penalty and adaptive knots.

has been analysed in the literature previously (see for instance Carter and Kohn (1997); Huerta and West (1999); Kirch et al. (2018)). It contains 40 missing observations that in this work have been replaced by the mean, thus not affecting the general features of the data. In addition, the data has been squared root transformed and mean centred. The first 150 observations are displayed in Figure 4, where the missing values are replaced by the mean.

Kirch et al. (2018) provided a psd estimate via a nonparametrically corrected (NPC) approach, which is able to detect several peaks in the function. These features make this data set suitable to assess the impact of the knot location on the P-spline estimates.

Again, the P-spline algorithm is run in two stages. First, a Markov chain of 5,000 samples with a burn-in period of 1,000 and thinning factor of 10, resulting in 400 points, is used to calibrate the proposals of the final run. Then, a Markov chain of 10,000 samples with a burn-in period of 2,000 and thinning factor of 10, that is 800 points, is used for the psd estimation. We use 40 B-spline densities, based on the rule-of-thumb criterion discussed above, and the first order penalty. The whole pro-
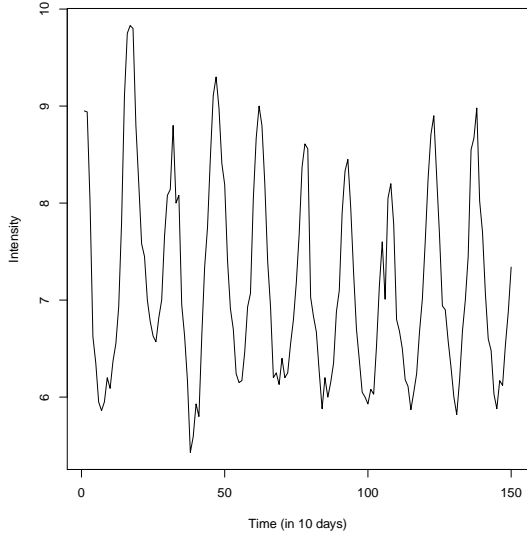
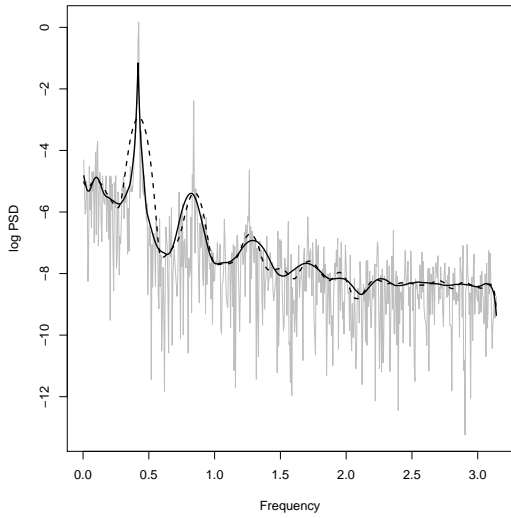**Fig. 4** First 150 observations of the *S. Carinae* data.

Fig. 5 Log-spectral density estimate for the variable star *S. Carinae* data based on the P-spline prior. The continuous grey line represents the periodogram whereas the continuous and dashed black lines stand for the posterior median log psd obtained using adaptive and equidistant knots, respectively.

cess takes less than 2 minutes to run and is replicated for equidistant and adaptive knots. The results are displayed in Figure 5.

The periodogram shows several peaks with the main one located at around 0.42. When the knots are equally spaced, the algorithm is unable to capture the sharpness of this peak, in contrast to the adaptive knots. This is because the periodogram-based knot allocation puts more knots in those re-

gions which contain significant peaks as indicated by the periodogram. The rest of peaks, which are smaller in comparison to the main one, are slightly better captured by the equidistant knots, since a large proportion of the adaptive knots have been allocated around the main peak. However, this can easily be corrected by increasing the number of knots. The psd estimate based on adaptive knots is comparable to the one obtained via the NPC approach (Kirch et al., 2018).

5 Discussion

As shown by Edwards et al. (2019), the BspDP prior outperforms the BDP prior in terms of IAE and uniform coverage probabilities for spiky psds. This is explained by the local support of the B-splines and its better approximation properties (Edwards et al., 2019). However, its performance is only achieved at a much higher computational cost. These characteristics have motivated our P-spline approach.

Although the P-spline prior is very similar to the B-spline prior as both are based on mixtures of B-spline distributions, there is quite a significant difference. The latter works with a variable number and location of knots, which are driven by the data. The P-spline prior assumes a fixed number of knots with fixed locations, thus almost looks like a special case. However, the multivariate prior based on the difference penalty of the basis coefficients controls the smoothness of the spectral density, a feature that can't be achieved with a univariate Dirichlet process prior. Most importantly, fixing the number and location of knots reduces the computational effort significantly, without sacrificing accuracy as shown in the application section. To deal with abrupt peaks in the psd, we propose to locate the knots according to the peaks detected via the periodogram. This approach does not affect the computational time and improves the estimates dramatically.

In our simulation study, we showed that the equidistant knot scheme for the P-spline method works quite well in the case of simple spectral structures (AR(1) model). However, its performance is surpassed by the adaptive knot scheme for complex structures (AR(4) model). In this scenario particularly, we can highlight the good performance of the adaptive knot scheme. This

method outperforms the BspDP prior in terms of IAE and run-times, particularly for the complex psds. Regarding uniform coverage, a *first order* penalty P-spline coupled with adaptive knots works best. For this combination, the uniform coverage almost reaches the nominal level.

We also assessed our proposal in the classic sunspot dataset, in which the posterior distribution based on the BspDP prior is able to estimate correctly the solar cycle that occurs every 11.07 years. The posterior distribution based on the P-spline prior yields almost identical results, but in significantly less computational time.

Finally, we assess the fit of the posterior distribution based on the P-spline prior a under two different knot location schemes: equally spaced and periodogram-spaced knots. For this, we estimated the psd for the *S. Carinae* time series which has a spectral density with several sharp peaks. Equidistant knots failed to capture the peaks of this spectral density whereas adpative knots significantly improved the results without affecting the computational time.

The number of B-spline densities was selected according to the criterion proposed by Ruppert (2002). It seems to work well for simple spectral densities. However, more complex functions could require larger number of B-spline densities. This can be evaluated studying the peaks of the periodogram. Our knot location proposal is very useful in these situations. Future research will explore whether the choice of the number of knots can be related to the entropy of the periodogram.

As already reviewed in the introduction, many Bayesian nonparametric approaches to spectral density estimation propose to put a prior on the log-spectral density. This could be easily implemented by modifying (4) to

$$\log f(\pi\omega) = \sum_{k=1}^{K} v_k b_{k,r}(\omega, \boldsymbol{\xi})$$

and it could be interesting to examine for what class of spectral densities this prior specification would yield better posterior inference.

Another interesting question is whether the P-spline prior can be extended to enable a Bayesian analysis of multivariate time series. In this case, the spectral density is a Hermitian positive definite matrix-valued function. A novel Bayesain

nonparametric approach to multivariate spectral density estimation was proposed by Meier et al. (2020) based on a generalization of the Bernstein-Dirchlet process prior where Bernstein polynomials are used to smooth a matrix-Gamma process. An avenue for future research will be to explore whether a computational speed-up might be possible by making use of P-splines in smoothing the elements of the Cholesky decomposition of the spectral density matrix.

In conclusion, the P-spline prior with the adaptive knot scheme offers a more computationally viable alternative to the BspDP prior for spectral density estimation. As shown in this work, it can handle spectral densities with several peaks in a reasonable run-time.

The R package `psplinePsd` can be downloaded from GitHub.

References

Vincent Bremhorst and Philippe Lambert. Flexible estimation in cure survival models using bayesian p-splines. *Computational Statistics & Data Analysis*, 93: 270 – 284, 2016. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2014.05.009. URL http://www.sciencedirect.com/science/article/pii/S0167947314001492.

Peter J Brockwell and Richard A Davis. *Time Series: Theory and Methods*. Springer-Verlag New York, New York, USA, 2 edition, 1991. ISBN 978-0-387-97429-3. doi: 10.1007/978-1-4419-0320-4. URL https://doi.org/10.1007/978-1-4419-0320-4.

A. Cadonna, A. Kottas, and R. Prado. Bayesian mixture modeling for spectral density estimation. *Statistics & Probability Letters*, 125:189–195, 2017. ISSN 0167-7152. doi:

https://doi.org/10.1016/j.spl.2017.02.008. URL http://www.sciencedirect.com/science/article/pii/S0167715217300573.

C. K. Carter and R. Kohn. Semiparametric bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society, Series B*, 59(1):255–268, 1997. doi: 10.1111/1467-9868.00067. URL https://doi.org/10.1111/1467-9868.00067.

Nidhan Choudhuri, Subhashis Ghosal, and Anindya Roy. Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050–1059, 2004. doi: 10.1198/016214504000000557. URL https://doi.org/10.1198/016214504000000557.

Matthew Edwards, Renate Meyer, and Nelson Christensen. Bayesian nonparametric spectral density estimation using b-spline priors. *Statistics and Computing*, 29(1):67–78, 2019. ISSN 0960-3174.

Matthew C. Edwards, Renate Meyer, and Nelson Christensen. *bsplinePsd: Bayesian Nonparametric Spectral Density Estimation Using B-Spline Priors*, 2018. URL https://CRAN.R-project.org/package=bsplinePsd. R package version 0.6.0.

Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with b -splines and penalties. *Statist. Sci.*, 11(2):89–121, 05 1996. doi: 10.1214/ss/1038425655. URL https://doi.org/10.1214/ss/1038425655.

Paul H. C. Eilers, Brian D. Marx, and María Durbán. Twenty years of p-splines. *SORT: statistics and operations research transactions*, 39(2), 2015. ISSN 1696-2281. URL http://hdl.handle.net/2117/88526.

A.K. Gangopadhyay, B.K. Mallick, and D.G.T. Denison. 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, 1999. ISSN 0378-3758. doi: https://doi.org/10.1016/S0378-3758(98)00148-7. URL http://www.sciencedirect.com/science/article/pii/S0378375898001487.

Peter J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 12 1995. ISSN 0006-3444. doi: 10.1093/biomet/82.4.711. URL https://dx.doi.org/10.1093/biomet/82.4.711.

G. Huerta and M. West. Bayesian inference on periodicities and component spectral structure in time series. *Journal of Time Series Analysis*, 20(4):401–416, 1999. doi: 10.1111/1467-9892.00145. URL https://doi.org/10.1111/1467-9892.00145.

Astrid Jullion and Philippe Lambert. Robust specification of the roughness penalty prior distribution in spatially adaptive bayesian p-splines models. *Computational Statistics & Data Analysis*, 51(5):2542 – 2558, 2007. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2006.09.027. URL http://www.sciencedirect.com/science/article/pii/S0167947306003549.

Göran Kauermann and Jean D. Opsomer. Data-driven selection of the spline dimension in penalized spline regression. *Biometrika*, 98(1):225–230, 2011. ISSN 00063444. URL http://www.jstor.org/stable/29777177.

Claudia Kirch, Matthew C. Edwards, Alexander Meier, and Renate Meyer. Beyond whittle: Nonparametric correction of a parametric likelihood with a focus on bayesian time series analysis. *Bayesian Anal.*, 2018. doi: 10.1214/18-BA1126. URL https://doi.org/10.1214/18-BA1126. Advance publication.

Tatyana Krivobokova, Gran Kauermann, and Theofanis Archontakis. Estimating the term structure of interest rates using penalized splines. *Statistical Papers*, 47(3):443–459, 2006. ISSN 0932-5026.

Philippe Lambert. Archimedean copula estimation using bayesian splines smoothing techniques. *Computational Statistics & Data Analysis*, 51(12):6307 – 6320, 2007. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2007.01.018. URL http://www.sciencedirect.com/science/article/pii/S0167947307000217.

Stefan Lang and Andreas Brezger. Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212, 2004. doi: 10.1198/1061860043010. URL https://doi.org/10.1198/1061860043010.

D.V. Likhachev. Selecting the right number of knots for b-spline parameterization of the dielectric functions in spectroscopic ellipsometry data analysis. *Thin Solid Films*, 636:519 – 526, 2017. ISSN 0040-

6090. doi: https://doi.org/10.1016/j.tsf.2017.06.056. URL http://www.sciencedirect.com/science/article/pii/S0040609017304911.

Alexander Meier, Claudia Kirch, and Renate Meyer. Bayesian nonparametric analysis of multivariate time series: A matrix gamma process approach. *Journal of Multivariate Analysis*, 175, 2020. ISSN 0047-259X.

Marianna Pensky, Brani Vidakovic, , and Daniela De Canditiis. Bayesian decision theoretic scale-adaptive estimation of a log-spectral density. *Statistica Sinica*, 17:635–666, 2007. URL http://www3.stat.sinica.edu.tw/statistica/j17n2/j17n210/j17n210.html.

F. Perron and K. Mengersen. Bayesian nonparametric modeling using mixtures of triangular distributions. *Biometrics*, 57(2):518–528, 2001. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/3068361.

Sonia Petrone. Random Bernstein Polynomials. *Scandinavian Journal of Statistics*, 26(3):373–393, 1999a. doi: 10.1111/1467-9469.00155. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9469.00155.

Sonia Petrone. Bayesian density estimation using bernstein polynomials. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 27(1):105–126, 1999b. ISSN 03195724. URL http://www.jstor.org/stable/3315494.

Nicholas G. Polson, James G. Scott, and Jesse Windle. Bayesian inference for logistic models using plyagamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013. doi: 10.1080/01621459.2013.829001. URL https://doi.org/10.1080/01621459.2013.829001.

Mara Rodrguez-lvarez, Maria Durban, Dae-Jin Lee, and Paul Eilers. On the estimation of variance parameters in non-standard generalised linear mixed models: application to penalised smoothing. *Statistics and Computing*, 29(3):483–500. ISSN 0960-3174.

Ori Rosen, Sally Wood, and David S. Stoffer. Adaptspec: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association*, 107(500):1575–1589, 2012. ISSN 0162-1459. doi: 10.1080/01621459.2012.716340. URL https://doi.org/10.1080/01621459.2012.716340.

David Ruppert. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11(4):735–757, 2002. doi: 10.1198/106186002853. URL https://doi.org/10.1198/106186002853.

Patiricio Maturana Russel and Renate Meyer. *psplinePsd: P-splines for spectral density estimation*, 2020. URL https://github.com/pmat747/psplinePsd.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994. URL http://www3.stat.sinica.edu.tw/statistica/j4n2/j4n216/j4n216.htm.

Xiaofeng Shao and Wei Biao Wu. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35(4):1773–1801, 2007. ISSN 00905364. URL http://www.jstor.org/stable/25464559.

Michael Wegener and Goran Kauermann. Forecasting in nonlinear univariate time series using penalized splines.(report). *Statistical Papers*, 58(3):557, 2017. ISSN 0932-5026.

P. Whittle. Curve and periodogram smoothing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 19(1):38–63, 1957. ISSN 00359246. URL http://www.jstor.org/stable/2983994.

Simon Wood. P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing*, 27(4):985–989, 2017. ISSN 0960-3174.

Simon N. Wood and Matteo Fasiolo. A generalized fellnerschall method for smoothing parameter optimization with application to tweedie location, scale and shape models. *Biometrics*, 73(4):1071–1081, 2017. ISSN 0006-341X.