

Spectral density estimation using P-spline priors

Patricio Maturana-Russel · Renate Meyer

the date of receipt and acceptance should be inserted later

Abstract This article proposes a Bayesian approach to estimating the spectral density of a stationary time series using a P-spline prior. Our proposal is motivated by the B-spline Dirichlet process prior of Edwards et al. (2018a) combined with Whittle’s likelihood with the aim of reducing the computational complexity of 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 fix the number and location of the knots and specify a P-spline prior. 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.

Keywords P-spline prior · B-spline prior · spectral density estimation · Whittle likelihood

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

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

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

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

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

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

in one way or another, where $-\pi < \lambda \leq \pi$ is the angular frequency and $\mathbf{Y} = (Y_1, \dots, Y_n)$ is a stationary time series of length n . The periodogram has interesting features, for instance, it is asymptotically unbiased and follows an exponential distribution with mean $f(\lambda)$. However, it is 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, the Whittle likelihood for a mean centered weakly stationary time series \mathbf{Y} of length n is defined as

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

where $\lambda_l = 2\pi l/n$ are the Fourier frequencies, $f(\cdot)$ spectral density of \mathbf{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 is 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 and Stoffer (2007) 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) modeled 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 (RJMCMC) (RJMCMC; Green, 1995) algorithm was implemented to sample from the posterior distribution.

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. (2018a) 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., 2018a), 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 Eilers and Marx (1996) that combine a large but fixed number of B-splines with a simple difference penalty that controls the degree smoothness of the spectral density. P-splines, i.e. equally-spaced B-splines, have been successfully implemented for frequentist nonparametric 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 B-spline densities 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. 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 remains 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 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

$$\begin{aligned}\hat{G}(\omega) &= \sum_{k=1}^K G\left(\frac{k-1}{K}, \frac{k}{K}\right) \beta(\omega; k, K-k+1) \\ &= \sum_{k=1}^K w_k \beta(\omega; k, K-k+1)\end{aligned}$$

converges uniformly to $G(\omega)$, where $G(u, v] = G(v) - G(u)$ and $\beta(\omega; a, b)$ is the beta density with parameters a and b . 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. (2018a) to propose the BspDP prior on the spectral density function.

B-splines

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

$$\begin{aligned}B_{k,0}(\omega, \boldsymbol{\xi}) &= \begin{cases} 1, & \omega \in [\xi_{k-1}, \xi_k] \\ 0, & \text{otherwise} \end{cases} \\ B_{k,r}(\omega, \boldsymbol{\xi}) &= v_{k,r} B_{k,r-1}(\omega; \boldsymbol{\xi}) + \\ &\quad (1 - v_{k+1,r}) B_{k+1,r-1}(\omega; \boldsymbol{\xi}),\end{aligned}$$

where

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

and

$$\begin{aligned}\boldsymbol{\xi} &= \{0 = \xi_0 = \xi_1 = \dots = \xi_r \leq \xi_{r+1} \leq \dots \leq \xi_K \\ &\quad = \xi_{K+1} = \dots = \xi_{K+r} = 1\}\end{aligned}$$

is the knot vector. This vector is a non-decreasing sequence that contains $K + r + 1$ knots, which can be divided in $2(r+1)$ 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

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

Prior description

The BspDP prior is a mixture of B-spline densities defined as

$$s_r(\omega; K, \mathbf{w}_K, \boldsymbol{\xi}) = \sum_{k=1}^K w_{k,K} b_{k,r}(\omega; \boldsymbol{\xi}), \quad (2)$$

where K is the number of B-spline densities $b_{k,r}(\cdot)$ of fixed degree r , $\mathbf{w}_K = (w_{1,K}, \dots, w_{K,K})$ the weight vector, and $\boldsymbol{\xi}$ is the knot sequence.

If the number of B-spline densities K is considered as another parameter, the \mathbf{w}_K dimension changes according to it. A natural way to deal with it is the RJMCMC (Green, 1995) algorithm, which was implemented by Gangopadhyay et al. (1999) in the similar context of low-order polynomials. However, the method often has implementation difficulties which vary from problem to problem, such as trans-dimensional jumps.

Assuming that the weights are induced by a cdf G on $[0, 1]$, Choudhuri et al. (2004) proposed a Dirichlet process as a prior for G . This approach avoids the potential difficulties caused by jumps between the parameter spaces of different dimensions in RJMCMC. Edwards et al. (2018a) followed this approach, but in addition they assumed that the $K-r$ internal knot differences are induced by a cdf H on $[0, 1]$ with further Dirichlet process prior, that is

$$\Delta_j = \xi_{j+r} - \xi_{j+r-1} = H\left(\frac{j-1}{K-r}, \frac{j}{K-r}\right),$$

for $j = 1, \dots, K-r$.

Therefore, the BspDP prior on the spectral density function parametrized in terms of K , G , and H is given by

$$s_r(\omega; K, G, H) = \sum_{k=1}^K G\left(\frac{k-1}{K}, \frac{k}{K}\right) b_{k,r}(\omega; H).$$

The BspDP prior is a mixture of B-spline densities with local support, unlike the BDP prior which uses beta densities with support on the

whole unit interval. This feature allows the B-spline method more flexibility for spectral density estimation (Edwards et al., 2018a).

When there are no internal knots, the B-spline basis coincides with the Bernstein polynomial basis. Therefore, the BspDP prior can be considered a generalization of the BDP prior.

The BspDP prior discussed above has been defined for the unit interval $[0, 1]$. However, the spectral density function $f(\cdot)$ is defined on the interval $[0, \pi]$, therefore its prior is reparametrized and defined as

$$f(\pi\omega) = \tau \times s_r(\omega; K, G, H), \quad \omega \in [0, 1],$$

where $\tau = \int_0^1 f(\tau\omega) d\omega$ is the normalizing constant.

To sum up, the BspDP prior has the following components:

- $G \sim \text{DP}(M_G, G_0)$, where $M_G > 0$ is the concentration parameter and G_0 is the base distribution. It produces the weights for the K B-spline densities.
- $H \sim \text{DP}(M_H, H_0)$, where $M_H > 0$ is the concentration parameter and H_0 is the base distribution. It produces the knot locations determining the shape and location of the B-spline densities.
- $p(K) \propto \exp(-\theta_K K^2)$ for $K = 1, \dots, K_{\max}$. K is the number of B-spline densities used in Eq. (2). It plays the role of smoothing parameter, that is smaller values implies smoother spectral density estimates. K_{\max} must be limited for computational reasons. In practice, pilot runs are used to determine it.
- $\tau \sim \text{Inverse-Gamma}(\alpha_\tau, \beta_\tau)$ is the normalizing constant. α_τ and β_τ are the shape and scale parameters, respectively.

The BspDP prior proposed by Edwards et al. (2018a) involves two Dirichlet processes, one placed on G inducing the weights of the B-spline densities and the other on H for the knot locations. In practice, these are implemented using the stick-breaking representation (Sethuraman, 1994), for details see Edwards et al. (2018a). For computational reasons, the number of components of this infinite series representation of G and H is limited to a large but finite positive integer, that is L_G and L_H , respectively. These parameters control the quality of the approximations. Larger values improve the approximation, but increase the

number of calculations and consequently the computation time.

As shown by Edwards et al. (2018a), 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

Fixing the number of B-spline densities K in Eq. (2), the prior for the unit interval $[0, 1]$ is defined as

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

where $\mathbf{w} = (w_1, \dots, w_K)$.

$\boldsymbol{\xi}$ contains fixed equidistant internal knots on $[0, 1]$ in our basic P-spline prior. A different allotment of knots is described in section “knot allocation strategy”. Therefore, the P-spline prior on the spectral density function is given by

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

where $\tau = \int_0^1 f(\tau\omega) d\omega$ is the normalizing constant.

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 translated into a prior distribution for the d^{th} 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 \mathbf{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

$$\begin{aligned} \mathbf{v} &\sim N_{K-1}(\mathbf{0}, (\phi \mathbf{P})^{-1}) \\ \phi | \delta &\sim \text{Gamma}(\alpha_\phi, \delta \beta_\phi) \\ \delta &\sim \text{Gamma}(\alpha_\delta, \beta_\delta) \end{aligned}$$

where $\mathbf{v} = (v_1, \dots, v_{K-1})^\top$ 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 d^{th} order difference matrix ($d = 1, 2, 3, \dots$), α_ϕ and α_δ are shape parameters, $\delta \beta_\phi$ and β_δ are rate parameters, and $\text{Gamma}(a, b)$ denotes a gamma distribution with mean a/b and variance a/b^2 . We used first, second and third order difference penalties in the simulation and case study. The 1st order difference matrix \mathbf{D} is defined as

$$\begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & & \cdots & 0 & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(K-2) \times (K-1)},$$

for the 2nd order is

$$\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & & & \cdots & 0 & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(K-3) \times (K-1)},$$

and for the 3rd order is

$$\begin{bmatrix} 1 & -3 & 3 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -3 & 3 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \ddots & & \vdots \\ 0 & & & \cdots & 0 & 1 & -3 & 3 & -1 \end{bmatrix}$$

$$\in \mathbb{R}^{(K-4) \times (K-1)}.$$

We take a robust specification for the penalty prior distribution as suggested by Jullion and Lambert (2007). 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). These are the prior specifications used in Section 4.

The parameter vector for the P-spline prior is $\boldsymbol{\theta} = (\mathbf{v}^\top, \phi, \delta, \tau)^\top$.

Posterior Computation

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

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

where $L(\mathbf{Y}|f)$ 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

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

However, to sample from the full conditional distribution of \mathbf{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 \mathbf{v} from which we calculate its mean vector $\bar{\mathbf{v}}$ and covariance matrix \mathbf{S} . Then we define the following re-parametrization

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

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

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

$$\beta_k^* = \beta_k + \sigma z,$$

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

where $\boldsymbol{\beta}^* = (\beta_1, \dots, \beta_k^*, \dots, \beta_{K-1})^\top$, $z \sim \text{Normal}(0, 1)$, 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 \mathbf{v}^* .

The m^{th} cycle of the Gibbs sampler is given by

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

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

- Draw δ^m from $\text{Gamma}(\alpha_\phi + \alpha_\delta, \beta_\phi\phi^m + \beta_\delta)$;
- Draw τ^m from

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

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 $\mathbf{v}^\top \mathbf{P} \mathbf{v}$ 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 $\mathbf{v}^\top \mathbf{P} \mathbf{v}$ 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 \mathbf{Y} . To avoid the problem, the author suggested to standardize \mathbf{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 produces rougher estimates and a low number produces smoother spectral densities. 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.

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 a relative optimal value according to certain goodness-of-fit criteria 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.

Knot allocation strategy

The knots are often equally spaced in P-spline methods. This scheme works quite well in the P-spline prior algorithm for psds in which there are no 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 a 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 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 \tilde{f} and calculate its distribution function F . Finally, the knots are allocated according to the quantiles of the inverse distribution function of \tilde{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 *periodogram-spaced knots* or short *P-spaced knots*.

The procedure can be summarized as follows:

1. Let $\mathbf{x} = (\sqrt{I_n(\lambda_1)}, \dots, \sqrt{I_n(\lambda_\nu)})$ be the square root of the periodogram values for each angular frequency.
2. Define $z_i = \left| \frac{x_i - \bar{x}}{S_x} \right|$ and calculate its distribution function $F_z = P(z \leq z_i)$.
3. Let $\boldsymbol{\xi}^* = (\xi_1^*, \dots, \xi_{K^*}^*)$ be a vector of K^* equally spaced values in $[0, 1]$.
4. The P-spaced knot vector is given by $\boldsymbol{\xi} = (F_z^{-1}(\xi_1^*), \dots, F_z^{-1}(\xi_{K^*}^*))$.

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 analyzed 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. (2018a) who compared spectral density estimates based on the BspDP and the BDP priors. We generated 300 time series 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. (2018b). 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`. A total of 100,000 MCMC samples were generated using a pilot run of 25,000 iterations with a burn-in period of 5,000 to calibrate the proposals. The final sample consisted of 75,000 samples with a burn-in period of 5,000 and thinning factor of 10, resulting in 7,000 samples used for posterior inferences. This procedure

was replicated for both equidistant and P-spaced knots.

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 (ρ_1, \dots, ρ_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

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

where $\hat{f}(\cdot)$ 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 P-spaced knots do not yield an improvement over the equidistant knots but slightly better or similar accuracy than the BspDP prior.

For the AR(4) time series, the P-spline prior with P-spaced knots categorically outperforms the BspDP algorithm. It also outperforms the equidistant knot scheme. The latter yields more accurate results than the BspDP prior for $n = \{128, 256\}$. For $n = 512$, the third order penalty P-splines has a slightly higher median IAE, but overall, this approach achieves higher accuracy. As shown by Edwards et al. (2018a), the BspDP prior produces better results than the BDP in terms of median IAE in this example. Consequently, our P-spline algorithm outperforms the BDP method as well.

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

$$\hat{f}(\lambda) \pm \zeta_\alpha \times \text{mad}(\hat{f}_i(\lambda)), \quad \lambda \in [0, \pi],$$

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

	$n = 128$	$n = 256$	$n = 512$
AR(1)			
B-spline	1.257	0.738	0.605
<i>Equidistant knots</i>			
P-spline $d = 1$	0.729	0.632	0.517
P-spline $d = 2$	0.844	0.612	0.638
P-spline $d = 3$	0.727	0.728	0.769
<i>P-spaced knots</i>			
P-spline $d = 1$	0.910	0.769	0.595
P-spline $d = 2$	0.994	0.867	0.731
P-spline $d = 3$	1.059	0.829	0.610
AR(4)			
B-spline	4.069	3.016	2.401
<i>Equidistant knots</i>			
P-spline $d = 1$	3.014	2.462	2.335
P-spline $d = 2$	3.247	2.619	2.399
P-spline $d = 3$	3.438	2.786	2.510
<i>P-spaced knots</i>			
P-spline $d = 1$	2.659	1.884	1.465
P-spline $d = 2$	2.627	1.849	1.411
P-spline $d = 3$	2.625	1.835	1.410

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.

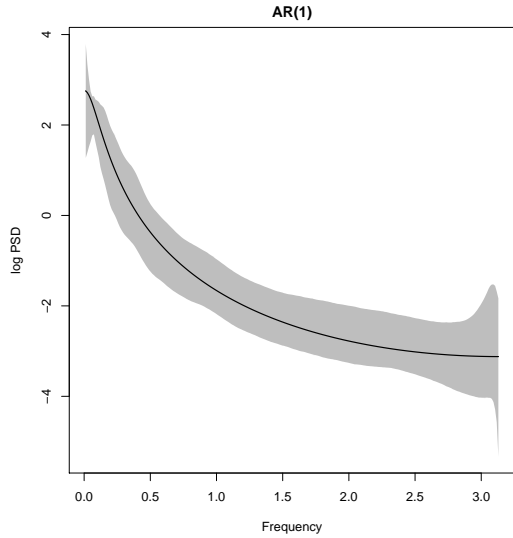


Fig. 1 Estimated log-spectral density for an AR(1) time series via the P-spline prior algorithm. The solid line stands for the true psd, whereas gray area is the uniform 90% credible band.

such that

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

The idea is to measure the proportion of times that the true spectral density is contained by the

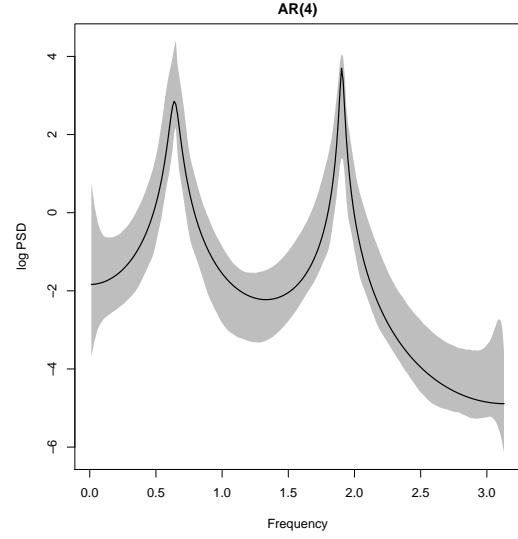


Fig. 2 Estimated log-spectral density for an AR(4) time series via the P-spline prior algorithm. The solid line stands for the true psd, whereas gray area is the uniform 90% credible band.

uniform credible band. This calculation is made for all 300 replications. The results are displayed in Table 2.

For the AR(1) model, all 90% credible bands cover the whole true spectral density function, i.e. for all frequencies λ . For the AR(4) model, the P-spline prior with P-spaced knots has the best coverage accross the different specifications. For the equidistant knots the coverage is always greater than 90%. For the first penalty order P-splines, the coverage probabilities are all higher than the ones produced by the BspDP prior algorithm. For the second order penalty, it has a better performance, except for $n = 512$, where the values are equal. For the third order penalty, we see a mix of results, that is equal, higher and lower values for the increasing sequence of n , respectively. For the P-spline priors, in particular for $n = 512$, the coverage probabilities decrease as the penalty degree order increases, which can be explained by the forced smoothness imposed by the higher penalty order. In general, P-spline priors have a better coverage probability than the BspDP method which as shown by Edwards et al. (2018a) have better coverage probabilities than the BDP prior in the AR(4) case.

As Edwards et al. (2018a) noted, one of the drawbacks of the BspDP prior is its computational complexity relative to the BDP. This is directly re-

	$n = 128$	$n = 256$	$n = 512$
AR(1)			
B-spline	1.000	1.000	1.000
<i>Equidistant \mathcal{E}</i>			
<i>P-spaced knots</i>			
P-spline $d = 1, 2, 3$	1.000	1.000	1.000
AR(4)			
B-spline	0.968	0.945	0.980
<i>Equidistant knots</i>			
P-spline $d = 1$	0.984	0.992	0.984
P-spline $d = 2$	0.984	0.984	0.980
P-spline $d = 3$	0.968	0.976	0.969
<i>P-spaced knots</i>			
P-spline $d = 1$	1.000	1.000	1.000
P-spline $d = 2$	1.000	1.000	1.000
P-spline $d = 3$	0.984	1.000	1.000

Table 2 Coverage probabilities of 90% uniform credible bands.

	$n = 128$	$n = 256$	$n = 512$
AR(1)			
B-spline	30.872	34.873	42.738
<i>P-spaced knots</i>			
P-spline $d = 1$	5.629	8.291	10.507
P-spline $d = 2$	5.650	8.374	10.512
P-spline $d = 3$	5.587	8.301	10.427
B-spline/P-spline	5.491	4.190	4.077
AR(4)			
B-spline	31.663	37.066	47.604
<i>P-spaced knots</i>			
P-spline $d = 1$	5.621	8.240	10.462
P-spline $d = 2$	5.641	8.372	10.549
P-spline $d = 3$	5.611	8.327	10.686
B-spline/P-spline	5.630	4.459	4.506

Table 3 Median run-times in minutes and relative run-times respect to an average of the P-spline times.

flected in the run-time, which is approximately 2-3 times higher in this example (see Table 3 in Edwards et al. (2018a)). The median run-time in our analyses for the BspDP prior and the P-spline with P-spaced knots are displayed in Table 3. Note that the P-spline method based on equidistant knots has similar run-times to the P-spaced knots. The P-spline method is approximately 4-6 times faster than BspDP approach in these examples. This is basically because it does not involve as many calculations as the latter 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.

Overall, the P-spline prior outperforms the B-spline prior in this example in terms of IAE, cov-

erage probability and run-time. Particularly, our novel knot allocation scheme, the periodogram-spaced knots, yield the best results in the case of peaked spectral densities. 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. (2018a), 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 around 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 replicated for the first, second and third order penalties and for equidistant and P-spaced knots. All other specifications are the same as those used in the simulation study. A single analysis takes around 6 minutes.

The results for the BspDP and P-spline prior with first order penalty and P-spaced 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.

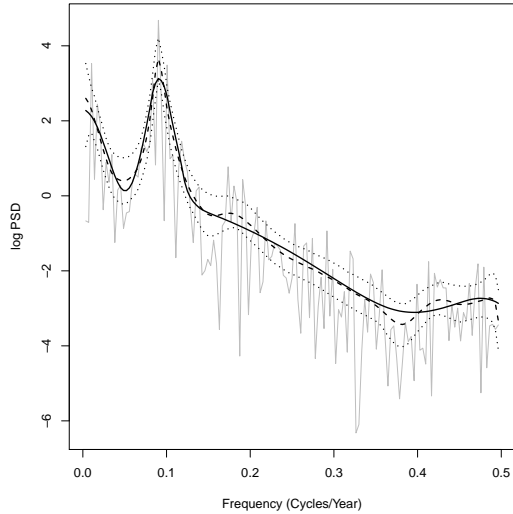


Fig. 3 Log-spectral density estimate for the sunspot data. The continuous gray line represents the periodogram. The continuous and dashed lines stand for the posterior median log psd obtained using the B-spline and P-spline priors, respectively. The dot lines represents the 90% pointwise credible bands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

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 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.

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

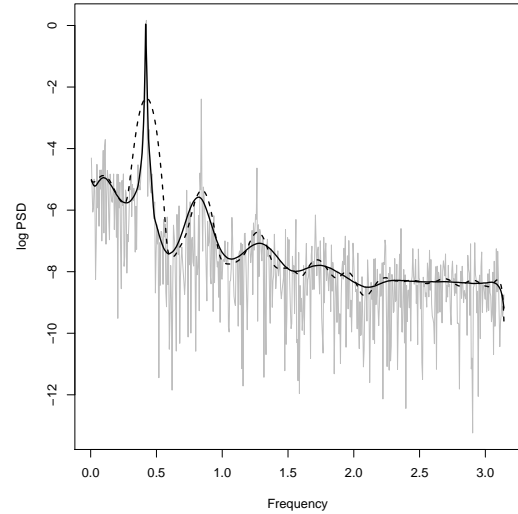


Fig. 4 Log-spectral density estimate for the variable star *S. Carinae* data based on the P-spline prior. The continuous gray line represents the periodogram whereas the continuous and dashed black lines stand for the posterior median log psd obtained using P-spaced and equidistant knots, respectively.

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 second order penalty. The whole process takes less than 2 minutes to run and is replicated for equidistant and P-spaced knots. The results are displayed in Figure 4.

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 P-spaced knots. This is because the periodogram-based knot allocation puts more knots in those regions which contain significant peaks detected 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 P-spaced 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 P-spaced knots is comparable to the one obtained via the NPC approach (Kirch et al., 2018).

5 Discussion

As shown by Edwards et al. (2018a), 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., 2018a). However, its performance is only achieved at a much higher computational cost. These characteristics have motivated our P-spline approach.

The P-spline prior can be regarded as a special case of the B-spline prior. 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. These simplifications reduce 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 improve the estimates dramatically. This also indicates that the BspDP computation time could be reduced by the choice of P-spaced knots as initial values.

In our simulation study, we showed that, in general, the P-spline prior outperforms the BspDP prior in all our assessments, that is in terms of IAE, coverage probabilities, and run-times. From the simulation study, it seems that a second order penalty P-spline works reasonably well.

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

Finally, we assess the P-spline prior algorithm 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 P-spaced 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, at least for the spec-

tral densities that we experimented with. 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.

In conclusion, the P-spline prior represents 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.

Acknowledgements We thank the New Zealand eScience Infrastructure (NeSI) for their high performance computing facilities, and the Centre for eResearch at the University of Auckland for their technical support. PM's and RM's work is supported by Grant 3714568 from the University of Auckland Faculty Research Development Fund and the DFG Grant KI 1443/3-1. All analysis was conducted in R, an open-source statistical software available on CRAN (cran.r-project.org).

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 C. Edwards, Renate Meyer, and Nelson Christensen. Bayesian nonparametric spectral density estimation using b-spline priors. *Statistics and Computing*, Jan 2018a. ISSN 1573-1375. doi: 10.1007/s11222-017-9796-9. URL <https://doi.org/10.1007/s11222-017-9796-9>.
- Matthew C. Edwards, Renate Meyer, and Nelson Christensen. *bsplinePsd: Bayesian Nonparametric Spectral Density Estimation Using B-Spline Priors*, 2018b. 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](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.
- 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>.
- 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>.
- Ori Rosen and David S. Stoffer. Automatic estimation of multivariate spectra via smoothing splines. *Biometrika*, 94(2):335–345, 02 2007. ISSN 0006-3444. doi: 10.1093/biomet/asm022. URL <https://doi.org/10.1093/biomet/asm022>.
- 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>.
- 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>.
- 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>.