Bayesian semiparametric power spectral density estimation with applications in gravitational wave data analysis

Matthew C. Edwards,[1,2,*] Renate Meyer,[1,†] and Nelson Christensen[2,‡]

[1]*Department of Statistics, University of Auckland, Auckland 1142, New Zealand*
[2]*Physics and Astronomy, Carleton College, Northfield, Minnesota 55057, USA*

(Received 30 May 2015; published 9 September 2015)

The standard noise model in gravitational wave (GW) data analysis assumes detector noise is stationary and Gaussian distributed, with a known power spectral density (PSD) that is usually estimated using clean off-source data. Real GW data often depart from these assumptions, and misspecified parametric models of the PSD could result in misleading inferences. We propose a Bayesian semiparametric approach to improve this. We use a nonparametric Bernstein polynomial prior on the PSD, with weights attained via a Dirichlet process distribution, and update this using the Whittle likelihood. Posterior samples are obtained using a blocked Metropolis-within-Gibbs sampler. We simultaneously estimate the reconstruction parameters of a rotating core collapse supernova GW burst that has been embedded in simulated Advanced LIGO noise. We also discuss an approach to deal with nonstationary data by breaking longer data streams into smaller and locally stationary components.

DOI: 10.1103/PhysRevD.92.064011 PACS numbers: 04.30.-w, 02.50.-r, 05.45.Tp, 97.60.Bw

I. INTRODUCTION

Astronomy is entering a new and exciting era, with the second generation of ground-based gravitational wave (GW) interferometers (Advanced LIGO [1], Advanced Virgo [2], and KAGRA [3]) expected to reach design sensitivity in the next few years. Throughout history, developments in astronomy have led to a deeper understanding of the Universe. Each time we probe the Universe with new sensors, we discover exciting and unexpected phenomena that challenge our current beliefs in astrophysics and cosmology. GW astronomy promises to do the same, providing a new set of ears to listen to (potentially unanticipated) cataclysmic events in the cosmos.

Apart from the first direct observation of GWs, extracting astrophysical information encoded in GW signals is one of the primary goals in GW data analysis. Since observations are subject to noise, accurate astrophysical predictions rely on an honest characterization of these noise sources. At its design sensitivity, Advanced LIGO will be sensitive to GWs in the frequency band from 10 Hz to 8 kHz. The main noise sources for ground-based interferometers include seismic noise, thermal noise, and photon shot (quantum) noise [1]. Seismic noise limits the low frequency sensitivity of the detectors. Thermal noise is the predominate noise source in the most sensitive frequency band of Advanced LIGO (around 100 Hz), and it arises from the test mass mirror suspensions and the Brownian motion of the mirror coatings. Photon shot noise is due to quantum uncertainties in the detected photon arrival rate, and it dominates the high frequency sensitivity of the detectors.

Standard assumptions about the noise model in the GW data analysis community rely on detector noise being stationary and Gaussian distributed, with a known power spectral density (PSD) that is usually estimated using off-source data (not on a candidate signal) [4]. Real GW data often depart from these assumptions [5]. It was demonstrated in [6] that fluctuations in the PSD can moderately bias parameter estimates of compact binary coalescence GW signals embedded in LIGO data from the sixth science run (S6).

High amplitude non-Gaussian transients (or "glitches") in real detector data invalidate the Gaussian noise assumption, and misspecifications of the parametric noise model could result in misleading inferences and predictions. A more sophisticated approach would be to make no assumptions about the underlying noise distribution by using nonparametric techniques. Unlike parametric statistical models, which have a fixed and finite set of parameters (e.g., the Gaussian distribution has two parameters: μ and σ^2 representing the mean and variance, respectively), nonparametric models have a potentially infinite set of parameters, allowing for much greater flexibility.

The theory of spectral density estimation requires a time series to be a stationary process. If data are not stationary (which is often the case for real LIGO data), it is important to adjust for this by introducing a time-varying PSD. It was demonstrated in [7] that the noise PSD in real S6 LIGO data is in fact time varying. Variation in detector sensitivity was also shown in [8]. Other GW literature that discusses nonstationary noise include [9,10]. It would be an oversimplification to assume the Advanced LIGO PSD is constant over time, and to use off-source data

*matt.edwards@auckland.ac.nz
†renate.meyer@auckland.ac.nz
‡nchriste@carleton.edu

in characterizing this. On-source estimation of the PSD would therefore be preferable to mitigate the time-varying nature of the PSD.

There have been attempts reported in the literature to improve the modeling of noise present in GW data, primarily concentrating on noise with embedded signals from well-modeled GW sources, such as binary inspirals [4,11–14], and more recently from GW bursts (unmodeled and typically short duration events) [7,15].

Under the Bayesian framework, Röver *et al.* [11] used a Student-t likelihood as a generalization to the commonly used Whittle (approximate Gaussian) likelihood [16]. The benefit of the Student-t setup is twofold: uncertainty in the noise spectrum can be accounted for via marginalization of nuisance parameters, and outliers can be accommodated due to the heavy-tail nature of the Student-t probability density. A drawback of this method is that the choice of hyperparameters can unduly influence posterior inferences.

Using the maximum likelihood approach, Röver [12] later demonstrated that the Student-t likelihood could be used as a generalization to the matched-filtering detection method commonly used in the analysis of GW signals from well-modeled sources. This approach would not be appropriate for GW bursts, since matched-filtering requires accurate signal models with well-defined parameter spaces.

Littenberg and Cornish [13] used Bayesian model selection to determine the best noise likelihood function in non-Gaussian noise. They considered Gaussian noise with a time-varying mean, noise from a weighted sum of two Gaussian distributions (non-Gaussian tails), and a combination of Gaussian noise and glitches (modeled as a linear combination of wavelets).

Littenberg *et al.* [4] demonstrated how one can incorporate additional scale parameters in the Gaussian likelihood and marginalize over the uncertainty in the PSD to reduce systematic biases in parameter estimates of compact binary mergers in S5 LIGO data. This method requires an initial estimate of the PSD. On a related note, Vitale *et al.* [14] highlighted a Bayesian method, similar to iteratively reweighted least squares, that analytically marginalizes out background noise and requires no *a priori* knowledge of the PSD. They applied this to simulated data from LISA Pathfinder.

More recently, Littenberg and Cornish [7] introduced the BayesLine algorithm in conjunction with BayesWave [15] to estimate the underlying PSD of GW detector noise. BayesLine is used to model the Gaussian noise component. They use a cubic spline to model the smooth changing broadband noise and Lorentzians (Cauchy densities) to model wandering spectral lines (due to AC supply, violin modes, etc.). BayesWave, on the other hand, models the non-Gaussian instrument "glitches" and burst sources with a continuous wavelet basis. Both methods make use of the transdimensional reversible jump Markov chain Monte Carlo (RJMCMC) algorithm of Green [17]. BayesLine is very pragmatic and works remarkably well on real Advanced LIGO data. However, the authors did not consider statistically important notions such as the posterior consistency of the PSD [18].

Our approach to improving the GW noise model relies on developments over the past decade in the area of Bayesian nonparametrics. Since parametric modeling can lead to biased estimates when the underlying parametric assumptions are invalid, we prefer nonparametric techniques to estimate the PSD of a stationary noise time series.

A common nonparametric estimate of the spectral density of a stationary time series is the periodogram, calculated using the (normalized) squared modulus of Fourier coefficients. That is,

$$I_n(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t \exp(-\mathrm{i}t\lambda)\right|^2, \qquad \lambda \in (-\pi, \pi], \quad (1)$$

where λ is the frequency, and $\{X_t\}$ is a stationary time series, where $t = 1, 2, \ldots, n$ represents discretized time. The periodogram randomly fluctuates about the true spectral density of a time series but is not a consistent estimator, motivating methods such as periodogram smoothing and averaging [19]. Averaging of off-source periodograms from Tukey windowed simulated Advanced LIGO noise has been used in GW literature relating to reconstructing rotating core collapse GWs [20] and predicting the important astrophysical parameters from these events [21].

In this paper, we implement the nonparametric Bayesian spectral density estimation method and Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) sampler presented by Choudhuri *et al.* [22], which updates a nonparametric Bernstein polynomial prior [23,24] on the spectral density using the Whittle likelihood to make posterior inferences. A Bernstein polynomial prior is essentially a finite mixture of Beta probability densities (see Sec. II C and Appendix A). It was proved that this method yields a consistent estimator for the true spectral density of a (short-term memory) stationary time series [22]—an attractive feature, absent in the periodogram. Posterior consistency in this context essentially means that the posterior probability of an arbitrary neighborhood around the true PSD goes to 1 as the length of the time series increases to infinity. Thus, as the sample size increases, the posterior distribution of the PSD will eventually concentrate in a neighborhood of the true PSD [18]. This is an important asymptotic robustness quality of the posterior distribution in that the choice of prior parameters should not influence the posterior distribution too much. Especially in Bayesian nonparametrics, because of the high dimension of the parameter space, many posterior distributions do not automatically possess this quality [18]. We refer the reader to Appendix C for a visual demonstration of posterior consistency.

Unlike Refs. [4,11,14], we do not treat noise as a nuisance parameter to be analytically integrated out. Although the signal parameters are our primary concern, we are also interested in quantifying our uncertainty in the underlying PSD of the noise in terms of posterior probabilities and credible intervals. Knowledge of this uncertainty will allow us to make honest astrophysical statements.

In this study, we assume that data are the sum of a GW signal embedded in noise (from all noise sources), such that

$$\mathbf{y} = \mathbf{s}(\boldsymbol{\beta}) + \boldsymbol{\epsilon}(\boldsymbol{\theta}), \tag{2}$$

where $\mathbf{y}$ is the (coincident) time-domain GW data vector, $\mathbf{s}$ is a GW signal parametrized by $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}$ is noise parametrized by $\boldsymbol{\theta}$. Notation with a tilde on top, such as $\tilde{\mathbf{y}}$, refers to the frequency-domain equivalent of the same quantity, obtained by the discrete Fourier transform (DFT). Note that we are treating noise in this setup as the conglomeration of detector noise (such as thermal noise and photon shot noise), background noise (such as seismic noise), and residual errors due to parametric statistical modeling of GW signals. An important caveat is to ensure the magnitude of the errors in the statistical model of the signal is minimized, so as to not artificially dominate the noise. Estimation of spectral lines (as done by the BayesLine algorithm [7]) is out of the scope of this paper.

The GW signal could essentially come from any source, but in this paper we will restrict our concentration to those from rotating core collapse supernovae to simplify the problem and demonstrate the power of the method. Using the recent waveform catalogue of Abdikamalov *et al.* [25], we conduct principal component analysis (PCA) and fit a principal component regression (PCR) model of the most important principal components (PCs) on an arbitrary rotating core collapse GW signal [20,21,26]. The (parametric) signal component is easily embedded as an additional Gibbs step in the Metropolis-within-Gibbs MCMC sampler of Choudhuri *et al.* [22]. That is, we utilize a *blocked* Gibbs approach to sequentially sample the signal parameters $\boldsymbol{\beta}$ given the noise parameters $\boldsymbol{\theta}$, and vice versa. As the model now contains a parametric signal component as well as a nonparametric noise component, it is "semiparametric."

To accommodate for nonstationary noise, we adapt an idea presented by Rosen *et al.* [27] and assume that a nonstationary time series can be broken down into smaller locally stationary segments. For each segment, we separately estimate the PSD using the method of Choudhuri *et al.* [22] and look at the time-varying spectrum.

We see this work as being a complement to existing methods, with the following benefits:

(i) A Bayesian framework, allowing us to update prior knowledge based on observed data, as well as quantify uncertainty in terms of probabilistic statements.
(ii) Posterior consistency of the PSD; i.e., the posterior distribution will concentrate around the true PSD as the sample size increases.
(iii) No parametric assumptions about the underlying noise distribution (parametric models are very sensitive to misspecifications), and high amplitude non-Gaussian transients in the noise can be handled.
(iv) Nonstationarities can be taken into account by splitting the data into smaller locally stationary segments.
(v) Estimation of noise and signal parameters are done simultaneously using Gibbs sampling.
(vi) Uncertainty in astrophysically meaningful parameter estimates are honest, with less systematic bias present.
(vii) Noninformative priors can be chosen, and the PSD does not need to be known *a priori*.
(viii) Useful for any signal with a parametric statistical model (including rotating core collapse supernova GWs).

The paper is structured as follows: Section II outlines the methods and models used to simultaneously estimate signal and noise parameters in GW data; results for toy models and simulated Advanced LIGO data are presented in Sec. III; and in Sec. IV, we discuss the consequences of this work, as well as future initiatives. Supplementary material can be found in the three appendixes.

II. METHODS AND MODELS

A. Parametric, nonparametric, and semiparametric models

Statistical models can be classified into two groups—parametric and nonparametric. Parametric models have a fixed and finite set of parameters, are relatively easy to analyze, and are powerful when their underlying assumptions are correctly specified. However, if the model is misspecified, inferences will be unreliable. Nonparametric models have far fewer restrictions but are less efficient and powerful than their parametric counterparts. No assumption about the underlying distribution of the data is made in nonparametric modeling, and the number of parameters is not fixed (and potentially infinite dimensional). Instead, the effective number of parameters increases with more data, providing the model structure.

For example, parametric regression (including linear models, nonlinear models, and generalized linear models) uses the following equation:

$$\mathbf{y} = g(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}|\boldsymbol{\beta}) + \boldsymbol{\epsilon}, \tag{3}$$

where $\mathbf{y}$ is the response variable, $g(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}|\boldsymbol{\beta})$ is a function of k explanatory variables (that aim to explain the variability in $\mathbf{y}$) given some model parameters $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}$ is the statistical error, usually assumed to be independent and identically distributed (iid) Gaussian random variables,

with 0 mean and constant variance σ^2. Here, the functional form of $g(.)$ is known in advance, such as in linear regression, where we have

$$g(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}|\boldsymbol{\beta}) = \beta_0 + \beta_1 \mathbf{x_1} + \cdots + \beta_k \mathbf{x_k}. \quad (4)$$

Nonparametric regression has a similar setup but assumes that the functional form of $g(.)$ is unknown and to be learned from the data. Function $g(.)$ could be thought of as an uncountably infinite-dimensional parameter in a nonparametric setting.

Semiparametric models contain both parametric and nonparametric components. The parametric regression model presented in Eqs. (3) and (4) is essentially the same parametric model used in this paper for GW signal reconstruction, where $(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k})$ are principal component (PC) basis functions. However, we model the noise $\boldsymbol{\epsilon}$ nonparametrically, rather than assuming iid Gaussian noise. Since we have parametric and nonparametric components, our model is semiparametric in nature.

B. Bayesian nonparametrics

Bayesian nonparametrics contains the set of models on the interface between the Bayesian framework and nonparametric statistics, and is characterized by large parameter spaces and probability measures over these spaces [18]. The Bayesian statistical framework is useful for incorporating prior knowledge and is particularly powerful when these priors accurately represent our beliefs. As mentioned in the previous section, nonparametric methods are useful for constructing flexible and robust alternatives to parametric models. A benefit of Bayesian nonparametric models is that they automatically infer model complexity from the data, without explicitly conducting model comparison.

Bayesian nonparametrics is a relatively nascent field in statistics and faces many challenges. The most obvious one is the mathematical difficulty in specifying well-defined probability distributions on infinite-dimensional function spaces. Constructing a prior on these spaces can be arduous, and in the case of noninformative priors, one should ensure large topological support so as not to put too much mass on a small region. Further, creating computationally convenient algorithms to sample from complicated posterior distributions presents its own set of challenges. It is also important to ensure that a Bayesian nonparametric model is statistically consistent (the truth is uncovered asymptotically), as some procedures do not automatically possess this quality [18].

Bayesian nonparametric priors (and posteriors) are stochastic processes rather than parametric distributions. Ferguson [28] provided the seminal paper for the field of Bayesian nonparametrics, introducing the Dirichlet process, an infinite-dimensional generalization of the Dirichlet distribution, now commonly used as a prior in infinite mixture models. This is a popular model (often called the Chinese Restaurant Process) for classification problems where the number of classes is unknown and to be inferred from the data. A formal definition of the Dirichlet distribution and Dirichlet process can be found in Appendix B.

Another popular prior in Bayesian nonparametrics is the Gaussian process prior, which is often used in nonlinear regression contexts. In fact, one could extend the regression example in the previous section into the realm of Bayesian nonparametrics by putting a Gaussian process prior on the function g. Compare this to the Bayesian parametric counterpart, which puts a prior on the model parameters $\boldsymbol{\beta}$.

For further discussion on Bayesian nonparametrics, we refer the reader to [18].

C. Spectral density estimation

A weakly (or second-order) stationary time series $\{X_t\}$ is a stochastic process that has constant and finite mean and variance over time, and an autocovariance function $\gamma(h)$ that depends only on the time lag h. That is, for a zero-mean weakly stationary process, the autocovariance function has the form

$$\gamma(h) = \mathrm{E}[X_t X_{t+h}], \quad \forall\ t, \quad (5)$$

where E[.] is the expected value operator, and t represents time.

Assuming an absolutely summable autocovariance function ($\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$), the (real-valued) spectral density function $f(\lambda)$ of a zero-mean weakly stationary time series is defined as

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-\mathrm{i}h\lambda), \qquad \lambda \in (-\pi, \pi], \quad (6)$$

where λ is the angular frequency. Note that the spectral density function and autocovariance function are a Fourier transform pair. In this paper, we will also call this the *power* spectral density (PSD) function, although this term is sometimes reserved for the empirical spectrum (periodogram) in the GW literature.

For a mean-centered weakly stationary time series $\{X_t\}$ of length n, with spectral density $f(\lambda)$, the Whittle approximation to the Gaussian likelihood, or simply the Whittle likelihood [16], is defined as

$$L_n(\mathbf{x}|f) \propto \exp\left(-\sum_{l=1}^{\lfloor u \rfloor} \left(\log f(\lambda_l) + \frac{I_n(\lambda_l)}{f(\lambda_l)}\right)\right), \quad (7)$$

where $\lambda_l = 2\pi l/n$ are the positive Fourier frequencies, $u = (n-1)/2$, $\lfloor u \rfloor$ is the greatest integer value less than or equal to u, and $I_n(.)$ is the periodogram defined in Eq. (1). If the PSD is known, the $\log f$ term in Eq. (7) is a constant and can be ignored. The Whittle likelihood has an advantage over the true Gaussian likelihood as it has a direct

dependence on the PSD rather than the autocovariance function. The Whittle likelihood is only exact for Gaussian white noise but works well under certain conditions, even when the data are not Gaussian [29]. More information about these concepts can be found in any good time series analysis textbook, such as the one by Brockwell and Davis [30].

We now need to specify a nonparametric prior for the PSD. We will briefly introduce the spectral density estimation technique of Choudhuri *et al.* [22], which is based on the Bernstein polynomial prior of Petrone [23,24]. The Bernstein polynomial prior is a nonparametric prior for a probability density on [0, 1] and is based on the Weierstrass approximation theorem that states that any continuous function on [0, 1] can be uniformly approximated to any desired degree by a Bernstein polynomial. If this function is a density on [0, 1], this Bernstein polynomial is essentially a finite mixture of Beta densities. We refer the reader to Appendix A for a definition of the Bernstein polynomial and Beta density. Instead of putting a Dirichlet prior on the mixture weight vector, the weights are defined via a probability distribution G on [0, 1], and a Dirichlet process prior is put on the space of probability distributions on [0, 1]. Appendix B contains supplementary material on the Dirichlet process.

Since the spectral density is not defined on the unit interval, we reparametrize $f(\lambda)$, such that

$$f(\pi\omega) = \tau q(\omega), \qquad \omega \in [0,1], \tag{8}$$

where $\tau = \int_0^1 f(\pi\omega)\mathrm{d}\omega$ is the normalization constant. To specify a prior on spectral density $f(\pi\omega)$, we put a Bernstein polynomial prior on $q(\omega)$, using the following hierarchical scheme:

(i) $q(\omega) = \sum_{j=1}^{k} G(\frac{j-1}{k}, \frac{j}{k}]\beta(\omega|j, k-j+1)$, where G is a cumulative distribution function, and $\beta(\omega|a,b)$ is a Beta probability density with parameters a and b.
(ii) G is a Dirichlet process with base measure G_0 and precision parameter M.
(iii) k has a discrete probability mass function such that $p(k) \propto \exp(-\theta_k k^2), k = 1, 2, \ldots$.
(iv) τ has an inverse-Gamma$(\alpha_\tau, \beta_\tau)$ distribution.
(v) G, k, and τ are *a priori* independent.

We use the stick-breaking construction of the Dirichlet process by Sethuraman [31], which is an infinite-dimensional mixture model (defined in Appendix B). For computational purposes, we need to truncate the number of mixture distributions to a large but finite number L. The choice of a large L will provide a more accurate approximation but also increase the computation time. Here, we choose $L = \max\{20, n^{1/3}\}$. We therefore reparametrize G to $(Z_0, Z_1, \ldots, Z_L, V_1, \ldots, V_L)$ such that

$$G = \left(\sum_{l=1}^{L} p_l \delta_{Z_l}\right) + \left(1 - \sum_{l=1}^{L} p_l\right)\delta_{Z_0}, \tag{9}$$

where $p_1 = V_1$, $p_l = (\prod_{j=1}^{l-1}(1-V_j))V_l$ for $l \geq 2$, $V_l \sim \mathrm{Beta}(1, M)$ for $l = 1, \ldots, L$, and $Z_l \sim G_0$ for $l = 0, 1, \ldots, L$. Note that δ_a is a probability density, degenerate at a. That is, $\delta_a = 1$ at a and 0 otherwise. This yields the prior mixture of the PSD,

$$f(\pi\omega) = \tau \sum_{j=1}^{k} w_{j,k}\beta(\omega|j, k-j+1), \tag{10}$$

with weights $w_{j,k} = \sum_{l=0}^{L} p_l I\{\frac{j-1}{k} < Z_l \leq \frac{j}{k}\}$ and $p_0 = 1 - \sum_{l=1}^{L} p_l$.

Abbreviating the vector of noise parameters as $\boldsymbol{\theta} = (\mathbf{v}, \mathbf{z}, k, \tau)$, the joint prior is therefore

$$p(\boldsymbol{\theta}) \propto \left(\prod_{l=1}^{L} M(1-v_l)^{M-1}\right)\left(\prod_{l=0}^{L} g_0(z_l)\right)p(k)p(\tau), \tag{11}$$

and is updated using the Whittle likelihood to produce the unnormalized joint posterior.

This method is implemented as a Metropolis-within-Gibbs MCMC sampler. In Choudhuri *et al.* [22], parameters k and τ are readily sampled from their full conditional posteriors, while $\mathbf{V}$ and $\mathbf{Z}$ require the Metropolis algorithm with Uniform proposals. Our only variation on this implementation is our sampling of the smoothness parameter k. We found that a Metropolis step is faster than sampling from the full conditional. The original implementation contains a for() loop that evaluates the log posterior $k_{\max}$ number of times, where $k_{\max}$ is chosen (during pilot runs) to be large enough to cater to the roughness of the PSD. For most well-behaved cases, $k_{\max} = 50$ will suffice, but the Advanced LIGO PSD requires many more mixture distributions (by 1 to 2 orders of magnitude) due to its steepness at low frequencies. This is a significant computational burden, and a well-tuned Metropolis step can therefore outperform the original implementation.

A discussion of the Dirichlet process and stick-breaking representation can be found in Appendix B.

D. Signal reconstruction

To reconstruct a rotating core collapse GW signal that is embedded in noise, we use the (parametric) PCR method described in [20,21,26]. That is, let

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}, \tag{12}$$

where $\tilde{\mathbf{y}}$ is the frequency-domain GW data vector of length n frequency-domain GW data vector, $\tilde{\mathbf{X}}$ is the $n \times d$ matrix of the d frequency-domain principal component basis vectors, $\boldsymbol{\beta}$ is the vector of signal reconstruction parameters (PC coefficients), and $\tilde{\boldsymbol{\epsilon}}$ is the frequency-domain noise vector with a *known* PSD. We assume flat priors on $\boldsymbol{\beta}$. It is

important to highlight that useful astrophysical information (such as the ratio of kinetic to gravitational potential energy of the inner core at bounce, and precollapse differential rotation) can be extracted by regressing the posterior means of the PC coefficients $\boldsymbol{\beta}$ on the known astrophysical parameters from the waveform catalogue, and sampling from the posterior predictive distribution [21].

We include an additional Gibbs step in the MCMC sampler described in the previous section to simultaneously reconstruct a rotating core collapse GW signal, while also estimating the noise power spectrum. Omitting the conditioning on the data for clarity, we sequentially sample the full set of conditional posterior densities $p(\boldsymbol{\theta}|\boldsymbol{\beta})$ and $p(\boldsymbol{\beta}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\mathbf{v}, \mathbf{z}, k, \tau)$ are the noise parameters defined in the previous section and $\boldsymbol{\beta}$ are the signal reconstruction parameters. That is, we sample in a cycle from the full conditional posterior distribution of the signal parameters, given the PSD parameters, and the full conditionals of the PSD parameters, given the signal parameters. This setup is called a *blocked* Gibbs sampler.

To sample the signal parameters, we fix the most recent MCMC sample of the PSD parameters. The conditional posterior of $\boldsymbol{\beta}$ is

$$\mathrm{P}(\boldsymbol{\beta}|\boldsymbol{\theta}) = \mathrm{N}(\boldsymbol{\mu}, \Sigma) \tag{13}$$

where $\Sigma = (\tilde{\mathbf{X}}'\mathbf{D}^{-1}\tilde{\mathbf{X}})^{-1}$ and $\boldsymbol{\mu} = \Sigma\tilde{\mathbf{X}}'\mathbf{D}^{-1}\tilde{\mathbf{y}}$. Here $\mathbf{D} = 2\pi \times \mathrm{diag}(f(\boldsymbol{\lambda}))$ is the noise covariance matrix, and $f(\boldsymbol{\lambda})$ is the most recent estimate of the PSD. More explicitly, at iteration $i+1$ in the blocked Gibbs sampling algorithm, we perform the following steps:

(1) Create a time-domain noise vector: $\boldsymbol{\epsilon}^{(i+1)} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(i)}$. Due to the linearity of the Fourier transform, $\boldsymbol{\beta}$ will be the same whether we are in the time domain or frequency domain.
(2) Sample the PSD parameters $\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\beta}^{(i)}$ using the method of Sec. II C.
(3) Sample the signal parameters $\boldsymbol{\beta}^{(i+1)}|\boldsymbol{\theta}^{(i+1)}$ using Eq. (13) (since the PSD in iteration $i+1$ is now known).

E. Nonstationary noise

As mentioned in Sec. II C, stationary noise has a constant and finite mean and variance over time, and an autocovariance function that depends only on the time lag. Nonstationary noise does not meet these requirements and has a time-varying spectrum. Stationarity of a time series can be tested using classical hypothesis tests such as the Augmented Dickey-Fuller test [32], the Phillips-Perron unit root test [33], and the Kwiatwoski-Phillips-Schmidt-Shin (KPSS) test [34].

To accommodate nonstationary noise, we adapt an idea presented by Rosen *et al.* [27], which assumes a time series can be broken down into locally stationary segments. In their paper, they treat the number of stationary components of a nonstationary time series as unknown and use RJMCMC [17] to estimate the segment breaks.

In a similar fashion, we break a nonstationary time series (or GW data stream) into J equal segments. We have two requirements for the length of these segments: the segment length is large enough for the Whittle approximation to be valid, and the segments are locally stationary according to heuristics or formal stationarity hypothesis tests. This approach fits nicely into our current MCMC framework. For each segment, we estimate the PSD using the nonparametric method introduced in Sec. II C. A benefit of this approach is that change-points in the PSD can be detected without using RJMCMC.

The conditional posterior density for all noise model parameters $\boldsymbol{\theta}$ is the following product:

$$\pi(\boldsymbol{\theta}|\boldsymbol{\beta}, \tilde{\mathbf{y}}) = \prod_{j=1}^{J} \pi_j(\boldsymbol{\theta}_j|\boldsymbol{\beta}, \tilde{\mathbf{y}}_j), \tag{14}$$

where $\pi_j(\boldsymbol{\theta}_j|\boldsymbol{\beta}, \tilde{\mathbf{y}}_j)$ is the conditional posterior density of the model parameters $\boldsymbol{\theta}_j$ in the jth segment given the signal parameters $\boldsymbol{\beta}$ and the jth segment of data $\tilde{\mathbf{y}}_j$.

Note that under this setup, the PC coefficients $\boldsymbol{\beta}$ do not depend on segments $j = 1, 2, \ldots, J$, since we require one set of PC coefficients (not J sets) to reconstruct a rotating core collapse GW signal.

To sample $\boldsymbol{\beta}|\boldsymbol{\theta}$, we use the same approach presented in Sec. II D. The only difference is in the construction of the noise covariance matrix. This is constructed as $\mathbf{D} = 2\pi \times \mathrm{diag}(f_1(\boldsymbol{\lambda}), f_2(\boldsymbol{\lambda}), \ldots, f_J(\boldsymbol{\lambda}))$, where $f_j(\boldsymbol{\lambda})$ is the PSD of the jth noise segment.

III. RESULTS

For the following examples, we set $L = \max\{20, n^{1/3}\}$ and use the noninformative prior setup of Choudhuri *et al.* [22]. That is, let $G_0 \sim \mathrm{Uniform}[0, 1], M = 1, \alpha_\tau = \beta_\tau = 0.001$, and $\theta_k = 0.01$. We use $k_{\max} = 50$ for most examples, and $k_{\max} = 400$ for the example with simulated Advanced LIGO noise to cater to the steep drop in the PSD at low frequencies.

For the examples with a signal embedded in noise, we use a $\mathrm{Uniform}(-\infty, \infty)$ prior on the signal reconstruction parameters $\boldsymbol{\beta}$, and let $d = 25$ PCs. For a discussion on the optimal choice of PCs, we refer the reader to [21]. We also scale the signals to a signal-to-noise ratio (SNR) of $\varrho = 50$. Here SNR (for n even) is defined as

$$\varrho = \sqrt{2\sum_{j=0}^{n/2+1} \frac{|\tilde{s}(\lambda_j)|^2}{|\tilde{\epsilon}(\lambda_j)|^2}}, \tag{15}$$

where λ_j are the positive Fourier frequencies, $\tilde{s}(.)$ is the Fourier transformed signal, and $\tilde{\epsilon}(.)$ is the Fourier transformed noise series. Note that for the zero and Nyquist

frequencies, the factor of 2 in Eq. (15) becomes a factor of 4.

The value of $\varrho = 50$ is physically motivated, as we would expect to see a SNR of approximately 50 to 170 for rotating core collapse supernova GWs at a distance of 10 kpc. We therefore demonstrate how the method works for the lower end of this range.

The units for frequency in most examples are radians per second (rad/s). In the example using simulated Advanced LIGO noise, we rescale to kilohertz (kHz). PSD units are the inverse of the frequency units, and the PSD figures are scaled logarithmically. GW strain amplitude is unitless.

For all examples, we run the MCMC sampler for 150,000 iterations, with a burn-in period of 50,000 and a thinning factor of 10. This results in 10,000 samples retained.

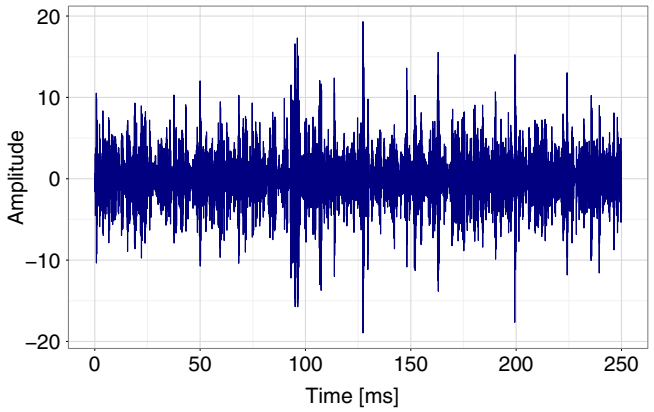

FIG. 1 (color online). Simulated stationary AR(1) process with first-order autocorrelation $\rho = -0.9$ and Student-t innovations ($\nu = 3$ degrees of freedom).

A. Estimating the PSD of non-Gaussian colored noise

To demonstrate how our model is capable of dealing with non-Gaussian transients in the data (or *glitches* as they are sometimes called in GW data analysis), we provide an illustrative toy example, using colored noise generated from a first-order autoregressive process, abbreviated as AR(1).

A mean-centered AR(1) process $\{X_t\}$ is defined as

$$X_t = \rho X_{t-1} + \epsilon_t, \qquad t = 1, 2, \ldots, n, \tag{16}$$

where ρ is the first-order autocorrelation, and ϵ_t is a white noise process (not necessarily Gaussian), with zero mean and constant variance σ_ϵ^2. With this formulation, we see how the current observation at time t depends on the previous observation at time $t-1$ through ρ, as well as some white noise ϵ_t, often referred to as innovations or the innovation process in time series literature.

The AR(1) model is a useful example here since it has a well-defined theoretical spectral density that we can compare our results against. Assuming $|\rho| < 1$, the AR(1) process is stationary and has spectral density

$$f(\lambda) = \frac{\sigma_\epsilon^2}{1 + \rho^2 - 2\rho\cos 2\pi\lambda}, \qquad \lambda \in (-\pi, \pi]. \tag{17}$$

As seen in Eq. (17), the AR(1) process has a PSD that is not flat, and the noise in our toy example is colored (nonwhite), with correlations between frequencies—typical of what we would expect with real Advanced LIGO noise. As the AR(1) process has a colored spectrum, and white noise has a flat spectrum, we will use the term innovations to refer to the white noise component of the model to avoid confusion.

For our example, rather than using Gaussian innovations, which is the most common innovation process used in autoregressive models, we use Student-t innovations with $\nu = 3$ degrees of freedom. The choice of $\nu = 3$ degrees of freedom is the smallest integer that results in a Student-t model with finite variance [a requirement for the innovation process $\{\epsilon_t\}$ of an AR(1) model]. This model has wider tails than that of the Gaussian model (and in fact the widest tails possible while maintaining the finite variance requirement), meaning we can expect extreme values in the tails of the distribution to occur more often. This will be our proxy for glitches.

We refer the reader to a relevant time series analysis textbook such as the one by Brockwell and Davis [30] for further information on AR(1) processes.

For this example, we generate a length $n = 2^{12}$ AR(1) process with $\rho = -0.9$ and Student-t innovations with $\nu = 3$ degrees of freedom. Let this (stationary) time series have a sampling interval $\Delta_t = 1/2^{14}$ (the same as Advanced LIGO). The data setup can be seen in Fig. 1.

We can see the effect of using $\nu = 3$ degrees of freedom in Fig. 1. Notice how there are transient high amplitude non-Gaussian events. These are a result of the wide-tailed nature of the Student-t density. It would be very unlikely to see these high amplitude events if the innovation process was Gaussian.

We now run the noise-only algorithm of Sec. II C to demonstrate that we can accurately characterize a non-Gaussian noise PSD.

The estimated pointwise posterior median log PSD in Fig. 2 is very close to the true log PSD, and the 90% credible region generally contains the true log PSD. This demonstrates that even if there are non-Gaussian transients in the data (which is certainly the case for real LIGO data), this PSD estimation method performs well. This is, however, not surprising as the Whittle likelihood gives a good approximation to Gaussian and some non-Gaussian likelihoods [29].

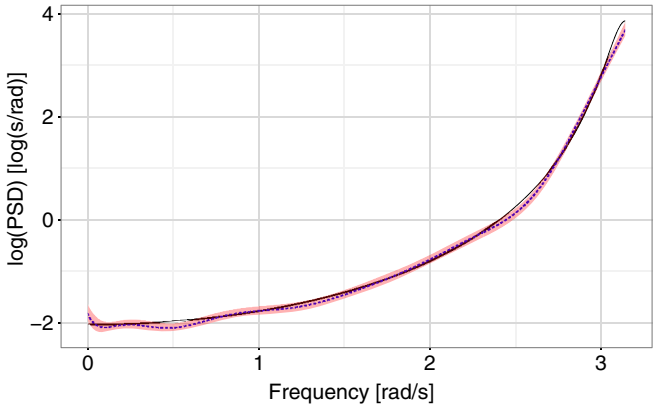

FIG. 2 (color online). Estimated log PSD of the AR(1) time series in Fig. 1. The 90% credible region (shaded pink) and posterior median log PSD (dashed blue) are superimposed with the true log PSD (solid black).

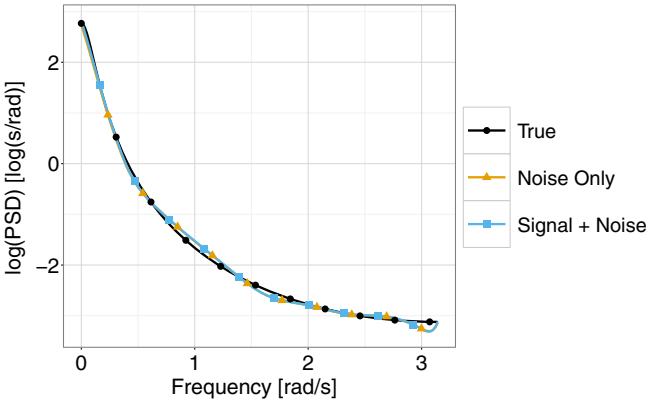

FIG. 4 (color online). Comparison of the noise PSD estimates for the noise-only and signal-plus-noise models. Plotted are the pointwise posterior median log noise PSDs with and without a GW signal. The true log PSD of the AR(1) noise series is overlaid.

B. Extracting a rotating core collapse signal in stationary colored noise

In this example, we aim to extract a rotating GW signal from noisy data using the blocked Gibbs sampler described in Sec. II D. We embed the $A1O12.25$ rotating core collapse GW signal from the Abdikamalov *et al.* [25] test catalogue (i.e., a signal not part of the base catalogue used to create the PC basis functions) in AR(1) noise with $\rho = 0.9$. For clarity, let this process have a Gaussian white noise innovation process with $\sigma_\epsilon^2 = 1$. Let the time series be length $n = 2^{12}$, which corresponds to $1/4$ s of data at the Advanced LIGO sampling rate. The signal is scaled to have a SNR of $\varrho = 50$. The reconstructed signal can be seen in Fig. 3.

The rotating core collapse GW signal in Fig. 3 is reconstructed particularly well during the collapse and bounce phases (the first few peaks or troughs). The post-bounce ringdown oscillations are usually poorly estimated due to stochastic dynamics [21,25], but are acceptable for this particular example.

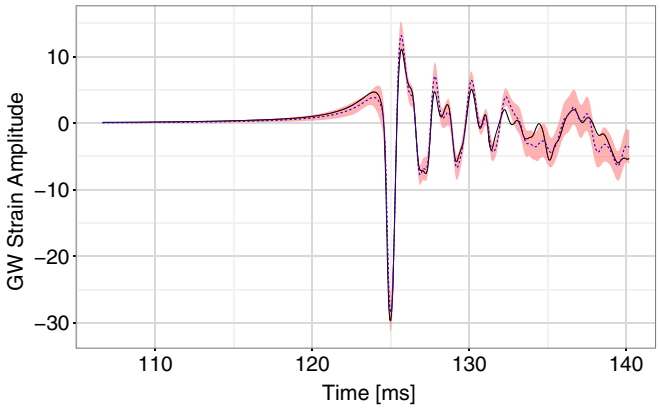

FIG. 3 (color online). Reconstructed rotating core collapse GW signal. The 90% credible region (shaded pink) and posterior median signal (dashed blue) are superimposed with true $A1O12.25$ GW signal from the Abdikamalov *et al.* [25] test catalogue (solid black).

In this example, the signal parameters were simultaneously estimated with the noise PSD using the blocked Gibbs sampler described in Sec. II D. We now compare the performance of the estimated noise PSD with and without a signal present. That is, we compare the noise PSD estimates between the algorithms presented in Sec. II C (noise-only model) and Sec. II D (signal-plus-noise model), using the same noise series for both models.

We can see in Fig. 4 that both models (noise-only and signal-plus-noise) perform similarly when estimating the PSD of colored Gaussian noise. The posterior median log PSDs are approximately equal and are very close to the true log PSD of an AR(1) process with $\rho = 0.9$. This is a useful robustness check, and it demonstrates that we are successfully decoupling the signal from the noise.

C. Comparing input and reconstruction parameters

As there is no analytic form linking the astrophysical parameters of a rotating core collapse stellar event to its GW signal, we can only approximate the GW signal using statistical methods. We do this using PCR, but this means that there are no true input parameters that we can compare with the estimated signal reconstruction parameters. However, if one were to create a fictitious signal as a known linear combination of PCs, we could demonstrate the algorithm's performance in estimating the signal reconstruction parameters.

Consider the following fictitious rotating core collapse GW signal:

$$\mathbf{y} = \sum_{i=1}^{d} \alpha_i \mathbf{x_i}, \tag{18}$$

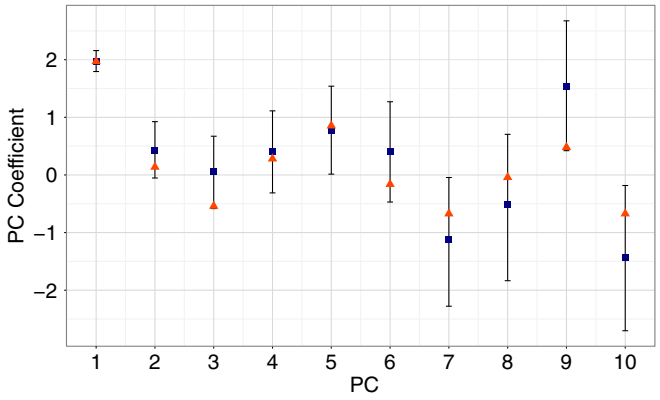

FIG. 5 (color online). Posterior median PC coefficients (blue square) and "true" PC coefficients (orange triangle) for the 10 PCs of a fictitious GW signal embedded in AR(1) noise. The error bands are the 95% credible intervals.

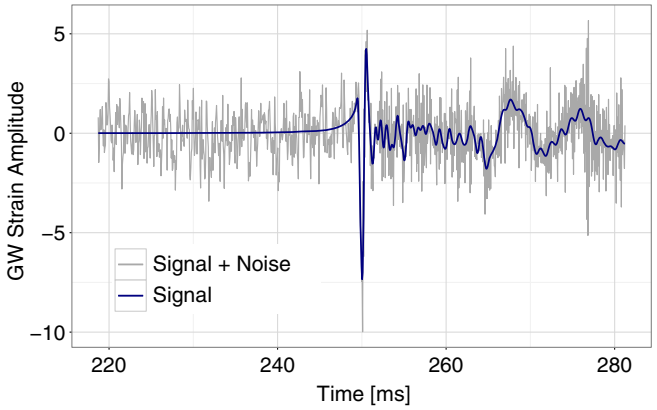

FIG. 6 (color online). Snapshot of the signal superimposed on the signal-plus-noise model. The noise series has length $n = n_1 + n_2 = 2^{13}$ and is segmented into two equal parts. The first half of the noise is generated from an AR(1) with $\rho = 0.5$, and the second half is generated from an AR(1) with $\rho = -0.75$. Both segments use a Gaussian innovation process with variance $\sigma_\epsilon^2 = 1$. The $A1O8.25$ rotating core collapse GW signal from the Abdikamalov *et al.* test catalogue [25] is embedded in this noise with a SNR of $\varrho = 50$.

where $\mathbf{y}$ is the signal of length n, $(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_d})$ are the d PC basis vectors of length n, and $(\alpha_1, \alpha_2, \ldots, \alpha_d)$ are the "true" weights, or PC coefficients. To randomize the weights, we randomly sample each from the standard normal distribution.

In this example, we embed the fictitious length $n = 2^{12}$ GW signal in AR(1) noise with $\rho = 0.9$ and Gaussian innovations with $\sigma_\epsilon = 1$. We set $d = 10$. We rescale the signal to have SNR $\varrho = 50$, and after the algorithm has run, we rescale our estimated PC coefficients back to the original level for comparison.

It can be seen in Fig. 5 that the true PC coefficients are generally contained within the 95% credible intervals, demonstrating that the algorithm can estimate a signal's input parameters well in the presence of stationary colored noise. Notice also that the credible intervals widen as the principal component number increases. This is due to the fact that higher numbered PCs explain lower amounts of variation in the waveform catalogue, resulting in lower amplitude waves. We would therefore be more uncertain about these PCs embedded in noise.

D. Extracting a rotating core collapse signal in time-varying colored noise

Nonstationary noise has a time-varying spectrum. To illustrate how our method can handle nonstationarities (or change-points in the spectral structure), we simulate a noise series with $J = 2$ locally stationary components of equal length $n_1 = n_2 = 2^{12}$. The first segment of the noise series is generated from an AR(1) process with $\rho = 0.5$. The second noise segment comes from an AR(1) process with $\rho = -0.75$. Both segments use a Gaussian innovation process with variance $\sigma_\epsilon^2 = 1$ for clarity. We embed part of the $A1O8.25$ waveform from the Abdikamalov *et al.* catalogue [25]. This waveform is in the test set, not included in the construction of PC basis functions. The data setup can be seen in Fig. 6.

The aim here is to simultaneously estimate both noise PSDs, as well as reconstruct the embedded GW signal using the method described in Sec. II E. Here we are assuming the change-point between the two noise series is known, though we will demonstrate in the next section that our method can locate unknown change-points.

Notice the difference between the first half of the noise series compared with the second half. Each segment has a different dependence structure and is therefore colored differently in the frequency domain. This results in a different time-domain morphology. Estimates of the noise PSDs can be seen in Figs. 7 and 8.

Figures 7 and 8 show the estimated log PSDs for the two noise segments. The pointwise posterior median log PSDs

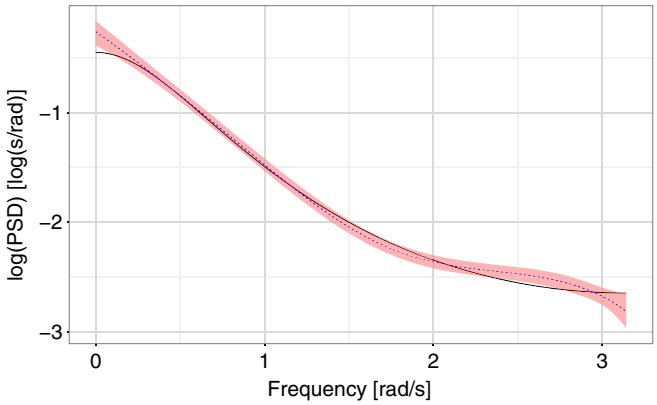

FIG. 7 (color online). Spectral density estimate of the first noise segment ($\rho = 0.5$) from Fig. 6. The 90% credible region (shaded pink), posterior median log PSD (dashed blue), and theoretical log PSD (solid black) are shown.

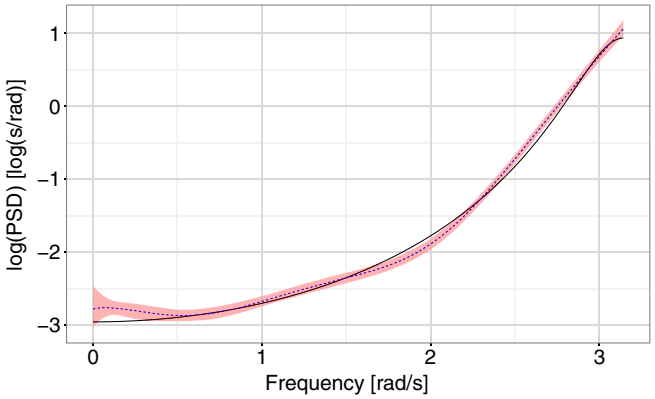

FIG. 8 (color online). Spectral density estimate of the second noise segment ($\rho = -0.75$) from Fig. 6. The 90% credible region (shaded pink), posterior median log PSD (dashed blue), and theoretical log PSD (solid black) are shown.

are close to the true log PSDs, and the 90% credible regions for both segments mostly contain the true log PSDs but veer slightly off towards the low frequencies. Due to posterior consistency of the PSD, these estimates will only get better as the sample size increases. Slight imperfections in the PSD estimates may not be such a problem if the embedded GW signal is extracted well, which happens to be the case in this example. The extracted signal can be seen in Fig. 9.

The 90% credible region for the reconstructed GW signal in Fig. 9 generally contains the true signal and has performed particularly well during collapse and bounce. Again, the post-bounce ringdown oscillations usually have the poorest reconstruction through the time series but have performed remarkably well in this example, regardless of the slight imperfections of the PSD estimates.

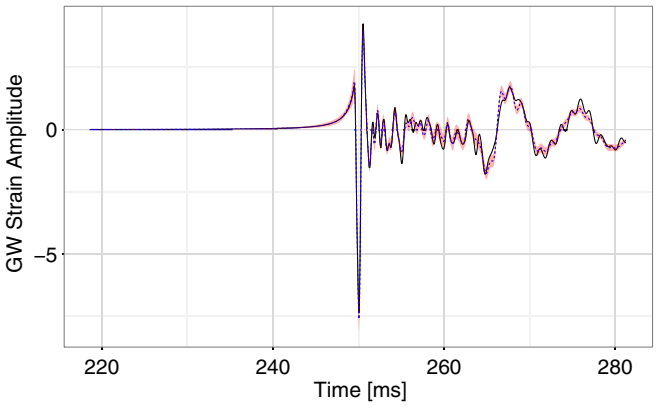

FIG. 9 (color online). Reconstructed rotating core collapse GW. The 90% credible region (shaded pink) and posterior median signal (dashed blue) are shown, superimposed with the true $A1O8.25$ GW signal from the Abdikamalov *et al.* [25] test catalogue (solid black). The first half of the signal was embedded in AR(1) noise with $\rho = 0.5$, and the second half had AR(1) noise with $\rho = -0.75$. Both noise segments had Gaussian white noise with $\sigma_\epsilon^2 = 1$.

E. Detecting a spectral change-point

Consider a change-point problem similar to that of the previous section, where a time series exhibits a change in its spectral structure somewhere in the series. A valuable consequence of the algorithm presented in Sec. II E is its ability to detect change-points regardless of whether the change-point occurs within a segment or on the boundary. For the following examples, let $n = 2^{12}$ and break this into $J = 32$ equal length segments. For clarity, assume the time series does not contain an embedded GW signal.

First consider the case where the change-point occurs on the boundary of two noise series. Let $n_1 = n_2 = 2^{11}$ be the lengths of each noise series, and let the first half of the time series be generated from an AR(1) with $\rho = 0.5$, and the second half from an AR(1) with $\rho = -0.75$. Both AR(1) processes have additive Gaussian innovations with $\sigma_\epsilon^2 = 1$. In this example, the change-point occurs exactly halfway through the series. Figure 10 shows a time-frequency map of the estimated log PSDs for each segment.

It is obvious that a change-point occurs halfway through Fig. 10, as there is a sheer change in the spectral structure at this point between segments 16 and 17. The first half of the time-frequency map exhibits stronger low-frequency behavior, whereas the second half has more power in the higher frequencies.

Now consider the case where the change-point occurs during a segment rather than on the boundary. Here, let each segment have the same setup as before, but instead set $n_1 = 2^{11} - 2^6$ and $n_2 = 2^{11} + 2^6$ such that a change-point occurs halfway through segment 16. A time-frequency map of the estimated log PSDs can be seen in Fig. 11.

Figure 11 demonstrates that there is a noticeable change-point roughly halfway through the series. There is a smoother transition from one PSD structure to the other than in the previous example since the true change-point occurs in the middle of a segment rather than on the boundary.

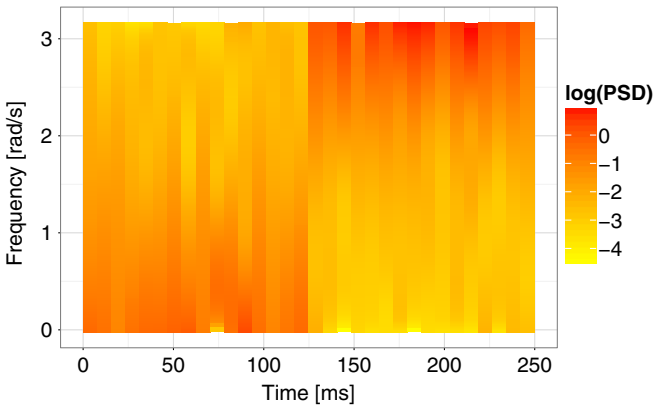

FIG. 10 (color online). Time-frequency map showing the estimated posterior median log PSDs for 32 segments of 1/4 s of AR(1) noise. The change-point in spectral structure occurs exactly halfway through the series.

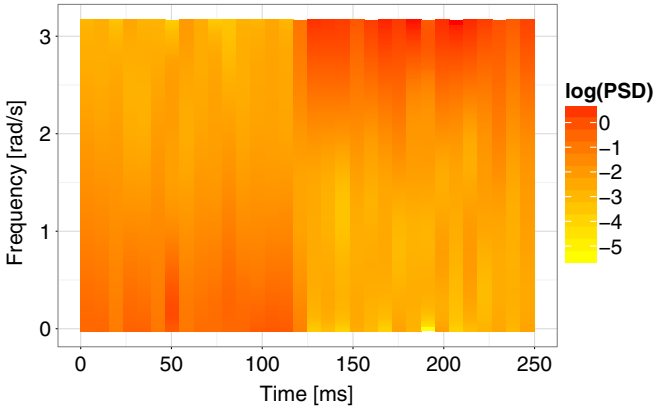

FIG. 11 (color online). Time-frequency map showing the estimated posterior median log PSDs for 32 segments of 1/4 s of AR(1) noise. The change-point in spectral structure occurs in the middle of segment 16 just before the halfway point.

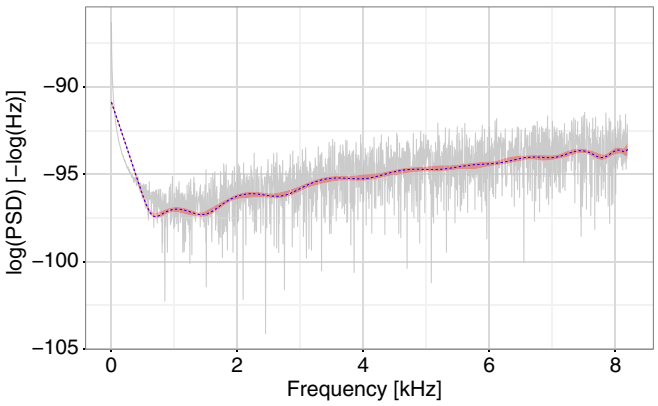

FIG. 12 (color online). Estimated log PSD for simulated Advanced LIGO noise. The 90% credible region (shaded pink) and posterior median (dashed blue) are shown, overlaid with log periodogram (solid grey).

These examples demonstrate that we can detect potentially unknown change-points in a time series. It is important to note that if more segments are used, the time duration within each segment becomes smaller, and our accuracy in detecting the change-point increases. That is, the time at which the change-point occurs becomes more resolved if the segment durations are smaller. However, one must also ensure that the segment durations are long enough for the Whittle approximation to be valid.

F. Simulated Advanced LIGO noise

In this example, we simulate Advanced LIGO noise and embed the $A1O10.25$ rotating core collapse GW signal from the Abdikamalov *et al.* [25] catalogue in it, scaled to a SNR of $\varrho = 50$. We assume a one-detector setup, with a linearly polarized GW signal (zero cross polarization). The Advanced LIGO sampling rate is $r_s = 2^{14}$ Hz, with a Nyquist frequency of $r_* = 2^{13}$ Hz. Let $n = 2^{12}$, which corresponds to quarter of a second of data.

The simulated noise is Gaussian and colored by the Advanced LIGO design sensitivity PSD. Generating this noise blindly results in a perfect matching of the end points and their derivatives, due to the simplified frequency-domain model. This is not realistic, since real data will often not have matching end points. In order to make the noise generation more realistic, we internally generated a longer frequency-domain series (10 times longer), inverse discrete Fourier transformed it, and returned a fraction of it with a random starting point. This is referred to as "padding" the data.

Figure 12 shows the estimated log PSD and the 90% credible region, overlaid with the log periodogram. The method performs remarkably well, particularly at higher frequencies. Even though we will not be able to resolve frequencies below ∼10–20 Hz at the Advanced LIGO design sensitivity, it is still interesting to see how this method performs at lower frequencies. Here, the low frequency estimates are slightly off, but not by much. We believe this to be due to two factors: 1/4 s of simulated Advanced LIGO noise is actually a nonstationary series, and we did not adjust for nonstationarities (simulated Advanced LIGO data are not stationary for more than 1/16 s based on the Augmented Dickey-Fuller test, Phillips-Perron unit root test, and KPSS test); and the Bernstein polynomial basis functions are notoriously slow to converge to a true function [35,36]. These factors considered, the method still provides a reasonable approximation.

The resultant reconstructed GW signal can be seen in Fig. 13. The estimated signal here is very close to the true signal during the collapse and bounce phases, as well as during the ringdown oscillations. The 90% credible region contains most of the true GW signal.

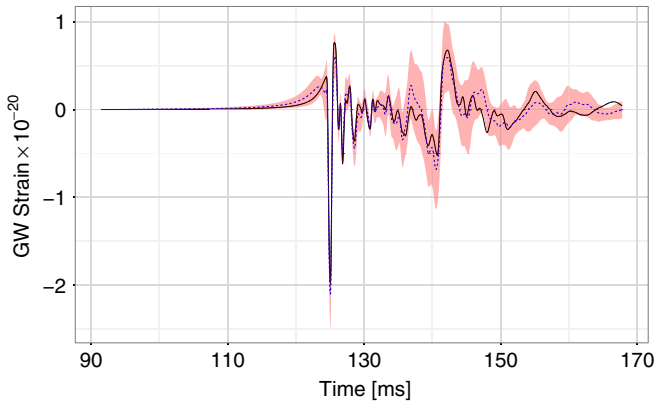

FIG. 13 (color online). Reconstructed rotating core collapse GW signal. The 90% credible interval (shaded pink) and posterior mean (dashed blue) are shown, overlaid with the true $A1O10.25$ signal (solid black) from Abdikamalov *et al.* [25] test catalogue. The signal is scaled to a SNR of $\varrho = 50$.

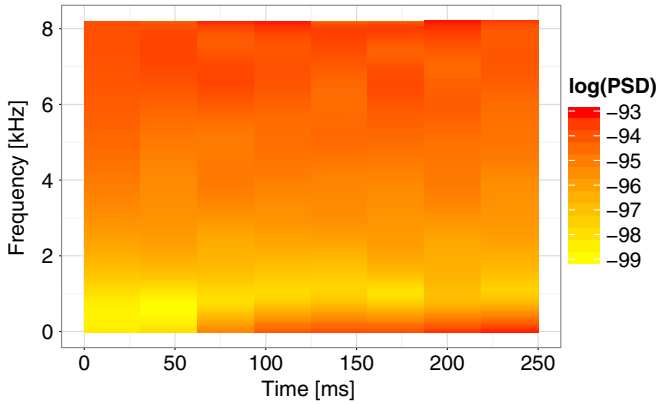

FIG. 14 (color online). Time-frequency map of the estimated time-varying noise spectrum for 8 segments of $1/4$ s simulated Advanced LIGO noise. The posterior median log PSDs for each noise segment are used.

We chose $d = 25$ PCs to reconstruct a rotating core collapse GW signal, but this could be too many or too few basis functions. Model selection methods similar to [21] were not investigated in the current study, and even though Figs. 3, 9, and 13 demonstrated good estimates during all phases (including ringdown), there is a demand for improved reconstruction methods.

We then accommodated for nonstationarities in detector noise by breaking the series into smaller and locally stationary components, and looked at the resulting time-varying spectrum. This can be seen in Fig. 14. Rather than choosing $J = 32$ as in Sec. III E, nonstationarities in the Advanced LIGO noise become more apparent if we slice the noise series into fewer segments, each with longer duration. Instead, consider splitting the data into $J = 8$ equal length segments ($n_j = 2^9$). Here, the Whittle approximation is valid, and the segments look locally stationary.

Figure 14 illustrates that the Advanced LIGO PSD is changing over time. Notice that lower frequencies are gaining more power over time. Assuming that each segment is locally stationary (which should be the case since the duration of each segment is less than $1/16$ s), it is important to accommodate for the changing nature of the PSD since the Choudhuri *et al.* [22] PSD estimation technique is based on the theory of stationary processes. If we did not adjust for nonstationarities, estimates of astrophysically meaningful parameters could become biased.

IV. DISCUSSION AND OUTLOOK

This study was motivated by the need for an improved model for PSD estimation in GW data analysis. The assumptions of the standard GW noise model are too restrictive for Advanced LIGO data. GW data are subject to high amplitude non-Gaussian transients, meaning that the Gaussian assumption is not valid. If the noise model is incorrectly specified, we could make misleading inferences. The stationarity assumption is also not valid, as simulated Advanced LIGO noise is not stationary for much longer than $1/16$ s according to classical statistical hypothesis tests. Using off-source data to estimate the PSD is problematic since the PSD will naturally drift over time, and is not necessarily the same as on the GW source.

The primary goal of this study was to develop a statistical model that allows for on-source estimation of the PSD, while making no assumptions about the underlying noise distribution. We also wanted a method capable of accounting for nonstationary noise. Although we restricted our attention to GWs from rotating core collapse stellar events in this paper, our approach is perfectly valid for any GW signal embedded in noise.

A secondary goal of this paper was to highlight to the GW data analysis community the rich and active area of Bayesian nonparametrics (and semiparametrics). We believe this framework will be a very powerful toolbox going forward, particularly in the analysis of GW bursts, since accurate parametric models for these types of signals are limited. Further, our future research efforts regarding rotating core collapse events involves Bayesian nonparametric regression models to construct GWs from their initial conditions. Regularization methods, such as the Bayesian LASSO [37], are also being considered.

In this paper, we have assumed linearly polarized GWs to be detected by one interferometer. A relatively simple extension of this work is to include a network of detectors, as well as GWs with nonzero cross polarization. Another extension would be to assume an unknown signal arrival time, as done in [20,21]. These extensions can be expected in the second generation of the algorithm.

The noise in our model was assumed to come from all sources, including detector noise, environmental noise, and statistical noise from parametric modeling of the signal. The statistical noise is the residual difference between the true and fitted signals. An important factor to consider was whether statistical noise artificially dominated the noise. We do not believe this to be a dominating contributor to the overall noise.

Since the "theoretical" PSD of Advanced LIGO at its design sensitivity has a very steep decrease at low frequencies until it reaches a minimum at roughly 230 Hz, it is difficult for our algorithm to perfectly characterize the shape at low frequencies without increasing computation significantly. This is due to the well-known slow convergence of Bernstein basis functions to a true curve. That is, many Bernstein polynomials (on order $k = 1000$) are required to accurately characterize the PSD of Advanced LIGO. Compare this to more well-behaved noise sources, such as those from autoregressive processes, which require $k < 50$. We are currently developing a second generation of this algorithm, using a mixture of B-spline densities

(normalized to the unit interval), rather than Beta densities. B-splines have much faster convergence rates than Bernstein polynomials [35,36]. An additional benefit of changing the basis functions to B-splines is that, like BayesLine [7], we will be able to account for spectral lines by peak-loading knots at *a priori* known frequencies at which these occur. The estimation of spectral lines is out of the scope of this paper, but we believe that a change of basis functions from Bernstein polynomials to normalized B-spline densities could work well. Another interesting approach would be to model spectral lines with informative priors using a similar approach to Macaro [38].

We used noninformative priors in this analysis. It may be possible to translate the known shape of the Advanced LIGO design sensitivity PSD into a prior. This may also aid in improving PSD estimates at lower frequencies.

We discussed a simplified method for estimating the time-varying PSD of nonstationary noise. Our approach assumed that a time series is split into equal length segments, and at known times. We demonstrated that it is possible to identify change-points in a time series and its spectrum using this method, and that there is no need to estimate the locations of the segment splits. Thus, a fixed grid of known segment placements suffices, and no RJMCMC is required. RJMCMC would have slowed the algorithm down significantly and created an entire new set of complications.

There is much work to be done on PSD estimation. As the Advanced LIGO and Advanced Virgo interferometers swiftly approach design sensitivity, it is important that we continue to focus not only on parameter estimation techniques, but also on modeling detector noise. PSD estimation is as important as parameter estimation, since we want to make honest statements about our observations based on rigorous statistical theory. It is hoped that in the near future, we can converge on a PSD estimation method that is less strict than the standard noise model, works well on real detector data, and is based on good statistical theory. We believe that the methods presented in this paper are definitely a step in the right direction.

ACKNOWLEDGMENTS

We thank Neil Cornish for a thorough reading of the manuscript and for providing insightful comments, Claudia Kirch and Sally Wood for helpful discussions on stationary and locally stationary time series, and Blake Seers for advice on data visualization. We also 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. The work of R. M. and M. E. is supported by the University of Auckland staff Research Grant 420048358, and the work of N. C. by NSF Grants No. PHY-1204371 and No. PHY-1505373. This paper has been given LIGO Document No. P1500057. All analysis was conducted in R, an open-source statistical software available on CRAN (cran.r-project.org). We used the ggplot2, grid, coda, and bspec packages and would like to acknowledge the authors of these packages.

APPENDIX A: BERNSTEIN POLYNOMIALS AND THE BETA DENSITY

To define the Bernstein polynomial, we first need to discuss the Bernstein *basis* polynomials. There are $k+1$ Bernstein basis polynomials of degree k, having the following form:

$$b_{j,k}(x) = \binom{k}{j} x^j (1-x)^{k-j}, \qquad j = 0, 1, \ldots, k. \tag{A1}$$

A Bernstein polynomial is the following linear combination of Bernstein basis polynomials:

$$B_k(x) = \sum_{j=0}^{k} \beta_j b_{j,k}(x), \tag{A2}$$

where β_j are called the Bernstein coefficients.

As mentioned in Sec. II C, the Bernstein polynomial prior is a finite mixture of Beta probability densities. We use the following parametrization for the Beta probability density function:

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \tag{A3}$$

$$\propto x^{\alpha-1}(1-x)^{\beta-1}, \tag{A4}$$

where $x \in (0,1)$, the shape parameters are positive real numbers (i.e., $\alpha > 0$ and $\beta > 0$), and $\Gamma(.)$ is the gamma function defined as the following improper integral:

$$\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) \mathrm{d}x. \tag{A5}$$

APPENDIX B: THE DIRICHLET DISTRIBUTION, DIRICHLET PROCESS, AND STICK-BREAKING CONSTRUCTION

The Dirichlet distribution is a multivariate generalization of the Beta distribution (defined in Appendix A) with a probability density function defined on the K-dimensional simplex

$$\Delta_K = \left\{ (x_1, \ldots, x_K) \colon x_i > 0, \sum_{i=1}^{K} x_i = 1 \right\}. \tag{B1}$$

The probability density function of the Dirichlet distribution is defined as

$$f(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\prod_{i=1}^{K}x_i^{\alpha_i-1}, \tag{B2}$$

where $\alpha_i > 0, i = 1, \ldots, K$.

The Dirichlet process is an infinite-dimensional generalization of the Dirichlet distribution. It is a probability distribution on the space of probability distributions, and is often used in Bayesian inference as a prior for infinite mixture models. One of the many representations of the Dirichlet process is Sethuraman's stick-breaking construction [18,31]. This is useful for implementing MCMC sampling algorithms.

Let $G \sim \mathrm{DP}(M, G_0)$, where G_0 is the center measure, and M is the precision parameter (larger M implies a more precise prior). The Sethuraman representation is

$$G = \sum_{i=1}^{\infty} p_i \delta_{Z_i}, \tag{B3}$$

$$p_i = \left(\prod_{j=1}^{i-1}(1 - V_j)\right)V_i, \tag{B4}$$

$$Z_i \sim G_0, \tag{B5}$$

$$V_i \sim \mathrm{Beta}(1, M). \tag{B6}$$

Consider a stick of unit length. The weights p_i associated with points Z_i can be thought of as breaking this stick randomly into infinite segments. Break the stick at location $V_1 \sim \mathrm{Beta}(1, M)$, assigning the mass V_1 to the random point $Z_1 \sim G_0$. Break the remaining length of the stick $1 - V_1$ by the proportion $V_2 \sim \mathrm{Beta}(1, M)$, assigning the mass $(1 - V_1)V_2$ to the random point $Z_2 \sim G_0$. At the ith step, break the remaining length of the stick $\prod_{j=1}^{i-1}(1 - V_j)$ by the proportion $V_i \sim \mathrm{Beta}(1, M)$, assigning the mass $(\prod_{j=1}^{i-1}(1 - V_j))V_i$ to the random point $Z_i \sim G_0$. This process is repeated infinitely many times.

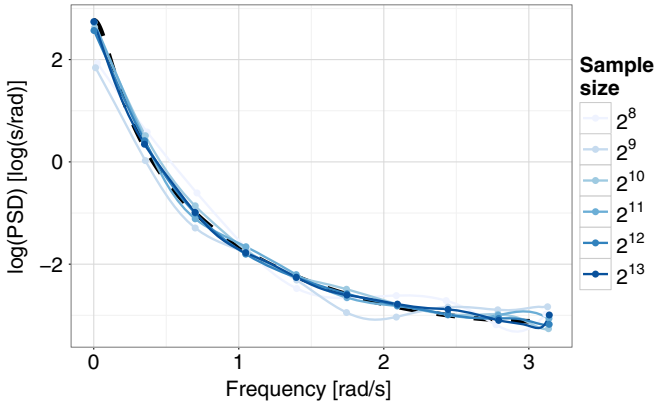

FIG. 15 (color online). Illustration of posterior consistency. The true log PSD (dashed black) is shown, overlaid with pointwise posterior median log PSDs of varying sample sizes.

APPENDIX C: DEMONSTRATION OF POSTERIOR CONSISTENCY

It was proved in [22] that under very general conditions on the prior, the PSD estimation method used in this paper has the property of posterior consistency. We provide an illustrative example of this in Fig. 15.

We generated AR(1) processes (with $\rho = 0.9$ and Gaussian white noise) of varying sample sizes and compared their performance. It can be seen in Fig. 15 that as the sample size of the time series increases, the pointwise posterior median log PSD gets closer to the true log PSD, thus demonstrating posterior consistency.

[1] LIGO Scientific Collaboration, Advanced LIGO, Classical Quantum Gravity **32**, 074001 (2015).
[2] F. Acernese *et al.*, Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2015).
[3] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto, Interferometer design of the KAGRA gravitational wave detector, Phys. Rev. D **88**, 043007 (2013).
[4] T. B. Littenberg, M. Coughlin, B. Farr, and W. M. Farr, Fortifying the characterization of binary mergers in LIGO data, Phys. Rev. D **88**, 084044 (2013).
[5] N. Christensen (for the LIGO Scientific Collaboration and Virgo Collaboration), LIGO S6 detector characterization studies, Classical Quantum Gravity **27**, 194010 (2010).
[6] J. Aasi *et al.*, Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network, Phys. Rev. D **88**, 062001 (2013).
[7] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, Phys. Rev. D **91**, 084034 (2015).
[8] LIGO Scientific Collaboration and Virgo Collaboration, Sensitivity achieved by the LIGO and Virgo gravitational wave detectors during LIGO's sixth and Virgo's second and third science runs, arXiv:1203.2674.
[9] A. M. Sintes and B. F. Schutz, Coherent line removal: filtering out harmonic related line interference from experimental data, with application to gravitational wave detectors, Phys. Rev. D **58**, 122003 (1998).

[10] L. S. Finn and S. Mukherjee, Data conditioning for gravitational wave detectors: A Kalman filter for regressing suspension violin modes, Phys. Rev. D **63**, 062004 (2001).
[11] C. Röver, R. Meyer, and N. Christensen, Modelling coloured residual noise in gravitational-wave signal processing, Classical Quantum Gravity **28**, 015010 (2011).
[12] C. Röver, Student-*t* based filter for robust signal detection, Phys. Rev. D **84**, 122004 (2011).
[13] T. B. Littenberg and N. J. Cornish, Separating gravitational wave signals from instrument artifacts, Phys. Rev. D **82**, 103007 (2010).
[14] S. Vitale *et al.*, Data series subtraction with unknown and unmodeled background noise, Phys. Rev. D **90**, 042003 (2014).
[15] N. J. Cornish and T. B. Littenberg, BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches, Classical Quantum Gravity **32**, 135012 (2015).
[16] P. Whittle, Curve and periodogram smoothing, J. Roy. Stat. Soc., Ser. B **19**, 38 (1957).
[17] P. J. Green, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika **82**, 711 (1995).
[18] S. Ghosal, in *Bayesian Nonparametrics*, edited by N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker (Cambridge University Press, Cambridge, England, 2010), pp. 35–79.
[19] P. D. Welch, The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoustics **15**, 70 (1967).
[20] C. Röver, M.-A. Bizouard, N. Christensen, H. Dimmelmeier, I. S. Heng, and R. Meyer, Bayesian reconstruction of gravitational wave burst signals from simulations of rotating stellar core collapse and bounce, Phys. Rev. D **80**, 102004 (2009).
[21] M. C. Edwards, R. Meyer, and N. Christensen, Bayesian parameter estimation of core collapse supernovae using gravitational wave signals, Inverse Probl. **30**, 114008 (2014).
[22] N. Choudhuri, S. Ghosal, and A. Roy, Bayesian estimation of the spectral density of a time series, J. Am. Stat. Assoc. **99**, 1050 (2004).
[23] S. Petrone, Random Bernstein polynomials, Scand. J. Stat. Theory Appl. **26**, 373 (1999).
[24] S. Petrone, Bayesian density estimation using Bernstein polynomials, Can. J. Stat. **27**, 105 (1999).
[25] E. Abdikamalov, S. Gossan, A. M. DeMaio, and C. D. Ott, Measuring the angular momentum distribution in core-collapse supernova progenitors with gravitational waves, Phys. Rev. D **90**, 044001 (2014).
[26] I. S. Heng, Rotating stellar core-collapse waveform decomposition: A principal component analysis approach, Classical Quantum Gravity **26**, 105005 (2009).
[27] O. Rosen, S. Wood, and D. S. Stoffer, AdaptSPEC: Adaptive spectral estimation for nonstationary time series, J. Am. Stat. Assoc. **107**, 1575 (2012).
[28] T. S. Ferguson, A Bayesian analysis of some nonparametric problems, Ann. Stat. **1**, 209 (1973).
[29] X. Shao and B. W. Wu, Asymptotic spectral theory for nonlinear time series, Ann. Stat. **35**, 1773 (2007).
[30] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. (Springer, New York, 1991).
[31] J. Sethuraman, A constructive definition of Dirichlet priors, Statistica Sinica **4**, 639 (1994).
[32] S. E. Said and D. A. Dickey, Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika **71**, 599 (1984).
[33] P. C. B. Phillips and P. Perron, Testing for a unit root in time series regression, Biometrika **75**, 335 (1988).
[34] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root, Journal of econometrics **54**, 159 (1992).
[35] M. J. D. Powell, *Approximation Theory and Methods* (Cambridge University Press, Cambridge, England, 1981).
[36] W. Shen and S. Ghosal, Adaptive Bayesian procedures using random series priors, arXiv:1403.0625.
[37] T. Park and G. Casella, The Bayesian lasso, J. Am. Stat. Assoc. **103**, 681 (2008).
[38] C. Macaro, Bayesian non-parametric signal extraction for Gaussian time series, Journal of econometrics **157**, 381 (2010).