

Variational inference for correlated gravitational wave detector network noise

Jianan Liu¹,² Avi Vajpeyi¹, Renate Meyer¹, Kamiel Janssens^{1,2,3,4,5}, Jeung Eun Lee¹,
Patricio Maturana-Russel^{1,6}, Nelson Christensen^{1,3}, and Yixuan Liu⁷

¹*Department of Statistics, University of Auckland, 38 Princes Street, Auckland, New Zealand*

²*Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium*

³*Université Côte d'Azur, Observatoire Côte d'Azur, CNRS, Artemis, 06304 Nice, France*

⁴*Department of Physics, The University of Adelaide, Adelaide, South Australia 5005, Australia*

⁵*ARC Centre of Excellence for Dark Matter Particle Physics, Melbourne, Australia*

⁶*Department of Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand*

⁷*MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom*



(Received 23 September 2024; revised 28 November 2024; accepted 7 February 2025; published 10 March 2025)

Gravitational wave detectors like the Einstein Telescope and LISA generate long multivariate time series, which pose significant challenges in spectral density estimation due to the large number of observations as well as the presence of correlated noise. Addressing both issues is crucial for accurately interpreting the signals detected by these instruments. This paper presents an application of a variational inference spectral density estimation method specifically tailored for dealing with correlated noise in the data. It is flexible in that it does not rely on any specific parametric form for the multivariate spectral density. We provide the multivariate Whittle likelihood in a form that is easy to evaluate as it depends on a low-dimensional covariance matrix. To deal with very long time series, the method employs a blocked Whittle likelihood approximation for stationary time series. It utilizes the Cholesky decomposition of the inverse spectral density matrix to ensure a positive definite estimator. A discounted regularized horseshoe prior is applied to the spline coefficients of each Cholesky factor, and the posterior distribution is computed using a stochastic gradient variational Bayes approach. This method is particularly effective in addressing correlated noise, a significant challenge in the analysis of multivariate data from colocated detectors. The method is demonstrated by analyzing 2000 s of simulated Einstein Telescope noise, which shows its ability to produce accurate spectral density estimates and quantify coherence between time series components. This makes it a powerful tool for analyzing correlated noise in gravitational wave data.

DOI: [10.1103/PhysRevD.111.062003](https://doi.org/10.1103/PhysRevD.111.062003)

I. INTRODUCTION

The next-generation gravitational-wave (GW) detectors, e.g., Einstein Telescope (ET) [1], Cosmic Explorer (CE) [2], and Laser Interferometer Space Antenna (LISA) [3], will usher in a transformative era of GW astronomy. The increased sensitivity to GW will lead to a much higher GW detection rate, improving our understanding of astrophysical and cosmic phenomena, ranging from the dynamics of black hole mergers to the nature of the early Universe [2,4–7].

However, multivariate time series analysis of future GW detector data faces challenges in managing correlated multivariate noise [8,9]. ET and LISA, which consist of colocated interferometers, may experience correlated noise in their data streams [10]. For ET, this correlated noise can stem from seismic, Newtonian, and magnetic sources [11–14], while LISA may encounter noise from temperature variations and microthrusters [15,16].

Ignoring correlated noise in analyses of multivariate GW data can lead to biased parameter estimates and incorrect

astrophysical conclusions (e.g., in analyzing the stochastic GW background [17–19] and transient signals from events such as binary black hole mergers [20]). Recent studies have shown the need for methods that can handle correlated noise [10] but did not include a demonstration of Bayesian methods capable of effectively doing so. Other work, [20] did demonstrate the capability to estimate the physical parameters of a compact binary coalescence event in a detector network with correlated noise. Their likelihood included the multivariate spectral density, but they treated the spectral density as fixed with a focus on estimating the parameters of the binary chirp signal. While this is an important step forward in the field, the authors of [20] made a number of simplifying assumptions. First of all, they considered the correlated noise to be in phase. Secondly, they considered the detectors to be (almost) maximally correlated across the entire frequency spectrum. So far, no Bayesian method has been suggested to estimate multivariate spectral densities of gravitational wave network data including correlated noise.

In this work, we demonstrate a method which is able to account for correlated noise as well as uncorrelated noise in a simultaneous and consistent approach. Additionally, we make no assumptions about the phase of the different noise terms and model both real and imaginary cross terms. While we demonstrate our method on a simplified noise example, the method forms the basis for future work with more complex noise scenarios, but this work shows the effectiveness of the chosen approach. Additionally, our proposed method allows for a flexible Bayesian approach for matrix-valued spectral density estimation, in a non-parametric way. Such a method is robust with respect to deviations from certain parametric shapes such as power laws in order to be able to capture all potential small- and large-scale noise characteristics. Furthermore, the approach must exhibit computational efficiency when applied to long time series of GW observations, as in the case of ET and LISA. Meier *et al.* [21] and Liu *et al.* [22] have developed nonparametric Bayesian approaches to multivariate spectral density estimation based on the Whittle likelihood and a Bernstein-Hermitian positive semidefinite matrix-Gamma process prior on the spectral density matrix and used an adaptive Markov chain Monte Carlo (MCMC) algorithm to sample from the posterior. Although theoretically attractive with proven consistency properties and contraction rates, MCMC methods require significant computation time due to the complexity of sampling from high-dimensional posterior distributions, making them impractical for long time series of GW observations, where the computational burden increases dramatically.

In this paper, a stochastic gradient variational Bayes (SGVB) approach, as developed by Hu and Prado [23], is proposed to estimate the multivariate power spectral density (PSD) for correlated ET noise. As a development within the broader framework of variational inference (VI), SGVB optimizes a surrogate posterior distribution by minimizing the Kullback-Leibler (KL) divergence from the true posterior distribution [24–26]. This method transforms complex posterior sampling into an optimization problem, enhancing sampling efficiency and reducing computational demands [27,28]. Recent advances in VI, such as normalizing flows, have demonstrated its effectiveness for pulsar timing-array datasets [29].

In this article, we demonstrate the usefulness and flexibility of a SGVB approach to estimate the spectral density matrix of the correlated residual noise from a network of GW detectors. These residuals could, for instance, be the remaining noise after subtracting individually resolved signals observed by ET or within global fit approaches for LISA [30,31]. Here, the approach is illustrated by using simulated ET noise and is also readily applicable to estimate correlated LISA noise or correlated LIGO–VIRGO–KAGRA (LVK) noise induced by magnetic noise such as from the Schumann resonances [17]. To ease computational efforts of handling large time series, a

blocked Whittle likelihood approximation is introduced and utilized instead of the traditional Whittle likelihood. Advice is provided on choosing tuning parameters such as the learning rate and number of basis functions in the SGVB approach. Finally, as a by-product of this method, the squared coherence is computed as a function of frequency, quantifying the degree of intertemporal correlation between components of the multivariate time series.

The article is structured as follows. Section II introduces the stochastic gradient variational Bayes (SGVB) method and defines the blocked Whittle likelihood. The efficiency and accuracy of the SGVB approach are tested using simulations in Sec. III. Section IV presents results of the method applied to simulated ET datasets consisting of varying levels of correlated noise. Lastly, the article concludes with a summary of findings in Sec. V.

II. METHOD

A. Likelihood

Assume we are given a time series of gravitational wave observations of length n from p channels. To set notation, let $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^\top \in \mathbb{R}^{n \times p}$ be a p -dimensional stationary, mean-zero time series, sampled at time intervals $\Delta_t = 1/(2f_{Ny})$, i.e., $\mathbf{Z}_\ell = \mathbf{Z}(\ell\Delta_t) \in \mathbb{R}^p$ for $\ell = 1, \dots, n$, where f_{Ny} is the Nyquist frequency, for a total observation time of T with a total of $n = T/\Delta_t$ sampled values for each of the p dimensions. The frequency resolution, Δ_f , is given by

$$\Delta_f = \frac{1}{n\Delta_t} = \frac{1}{T}. \quad (1)$$

Let \mathbf{d}_k be the discrete Fourier transform (DFT) of \mathbf{Z} given by

$$\mathbf{d}(f_k) = \Delta_t \sum_{t=1}^n \mathbf{Z}_t \exp\left(-2\pi i \frac{k}{n} t\right), \quad (2)$$

where $f_k = k\Delta_f = k/(n\Delta_t) = k/T$ for $k = 1, \dots, N$, and $N = n/2$ if n is even, $N = (n-1)/2$ if n is odd. For stationary time series with absolutely summable autocovariances, i.e., $\sum_{h=1}^{\infty} \gamma_{lm}(h) < \infty$, the discrete Fourier coefficients $\mathbf{d}(f_k)$ are asymptotically independent and have a complex Gaussian distribution with mean zero and covariance matrix $T\mathbf{S}(f_k)$ with the continuous two-sided spectral density matrix,

$$\mathbf{S}(f) = \frac{1}{2f_{Ny}} \sum_{\ell=-\infty}^{\infty} \Gamma(\ell\Delta_t) \exp(-2\pi i f \ell \Delta_t),$$

and the Fourier transform of the time-invariant autocovariance function $\Gamma(h) = (\gamma_{lm}(h)) = \mathbb{E}(\mathbf{Z}_t \mathbf{Z}_{t+h}^\top)$ that in turn defines the Toeplitz covariance matrix of \mathbf{Z} . This

asymptotic Gaussian distribution is the basis of the multivariate Whittle likelihood approximation in the frequency domain, expressed as

$$\mathcal{L}(\mathbf{d}|\mathbf{S}) \propto \prod_{k=1}^N \det(\mathbf{S}(f_k))^{-1} \times \exp\left(-\frac{1}{T} \mathbf{d}(f_k)^* \mathbf{S}(f_k)^{-1} \mathbf{d}(f_k)\right), \quad (3)$$

where $\mathbf{d}(f_k)^*$ is the conjugate transpose of the $\mathbf{d}(f_k)$ and $\mathbf{S}(f_k)$ is a $p \times p$ Hermitian positive semidefinite spectral density matrix at each f_k . Therefore, the unknown quantity here is \mathbf{S} , a matrix-valued function at each frequency with the additional restriction that its value at each frequency is Hermitian positive semidefinite. Note that when estimating the spectral density matrix, it is important that the estimate is again positive semidefinite at each f_k so that the quadratic form in the exponent of the Whittle likelihood remains positive and thus defines a valid likelihood. In this analysis, we use the two-sided spectral density, which includes both positive and negative frequencies, allowing for a full representation of the frequency spectrum. However, for the purposes of this study, we only retain the positive frequency components.

Given \mathbf{d} , one method to estimate the multivariate spectral density matrix \mathbf{S} is Bayesian inference. In particular, having a flexible Bayesian method instead of simply using a frequentist Welch estimate is important when the ultimate task is to simultaneously estimate the parameters of a GW signal while properly taking the uncertainties of the residual noise estimation into account. The Bayes' theorem is given by

$$p(\mathbf{S}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\mathbf{S})\pi(\mathbf{S})}{\mathcal{Z}(\mathbf{d})} \propto \mathcal{L}(\mathbf{d}|\mathbf{S})\pi(\mathbf{S}), \quad (4)$$

where $\pi(\mathbf{S})$ is the prior distribution and $p(\mathbf{S}|\mathbf{d})$ is the posterior density of unknown \mathbf{S} given \mathbf{d} , and $\mathcal{Z}(\mathbf{d})$ is the Bayesian evidence (see Thrane and Talbot [32] and Christensen and Meyer [33] for discussions on GW Bayesian inference).

In GW parameter estimation with large datasets, such as those from ET and LISA, the computational burden of directly applying the Whittle likelihood can be significant. To mitigate this challenge, a ‘‘blocked’’ Whittle likelihood is adopted. The time series is divided into N_b equal-sized blocks $\mathbf{Z} = (\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(N_b)})^\top$ with each block a p -dimensional time series of each of length n/N_b . The discrete Fourier transform of each block is denoted by $\mathbf{d}^{(i)}$, $i = 1, \dots, N_b$. The stationarity assumption implies that the statistical properties of each block, in particular, their spectral densities, are the same. Making the further simplifying assumption of independence among different

blocks, the likelihood then becomes the product of the individual Whittle likelihoods for each block,

$$\mathcal{L}_b(\mathbf{d}|\mathbf{S}) = \prod_{i=1}^{N_b} \mathcal{L}(\mathbf{d}^{(i)}|\mathbf{S}). \quad (5)$$

Assuming independence of the blocks ignores the correlations that will be present between adjacent blocks, but the correlations will vanish with increasing distance between the blocks. This independence assumption is often made, e.g., in the context of the block bootstrap for time series, see Lahiri [34], or in the recent paper by Rosati and Littenberg [35] on residual noise estimation within the LISA global fit and Criswell *et al.* [36] on anisotropic LISA Galactic foreground estimation. Note that as the length of the blocked data is less than the original dataset, the frequency resolution of the spectral density will become coarser as the number of blocks N_b increases. In practice, the number of blocks should be chosen to achieve a required frequency resolution while also remaining computationally tractable.

B. Parametrization of \mathbf{S}

The prior defined by Hu and Prado [23] and Rosen and Stoffer [37] models the components of the Cholesky factorization of the inverse spectral density matrix via smoothing splines (for details, refer to Hu and Prado [23]). In their approach, the inverse of $\mathbf{S}(f_k)$ can be represented as $\mathbf{S}(f_k)^{-1} = \mathbf{T}_k^* \mathbf{D}_k^{-1} \mathbf{T}_k$, where \mathbf{D}_k is a diagonal matrix with diagonal elements $\delta_{1k}^2, \delta_{2k}^2, \dots, \delta_{pk}^2$, and

$$\mathbf{T}_k = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\theta_{21}^{(k)} & 1 & 0 & \dots & 0 \\ -\theta_{31}^{(k)} & -\theta_{32}^{(k)} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\theta_{p1}^{(k)} & -\theta_{p2}^{(k)} & \dots & -\theta_{p,p-1}^{(k)} & 1 \end{pmatrix} \quad (6)$$

is a $p \times p$ complex unit lower triangular matrix. Thus, the Whittle likelihood can be rewritten as the product of its p components depending on θ_j, δ_j , for $j = 1, \dots, p$,

$$\mathcal{L}(\mathbf{d}|\mathbf{S}) \propto \prod_{j=1}^p \mathcal{L}_j(\mathbf{d}_j|\theta_j, \delta_j, \mathbf{d}_{<j}), \quad (7)$$

where

$$\mathcal{L}_j(\mathbf{d}_j|\theta_j, \delta_j, \mathbf{d}_{<j}) \propto \prod_{k=1}^N \delta_{jk}^{-2} \exp\left(\frac{-|d_j(f_k) - \sum_{l=1}^{j-1} \theta_{jl}^{(k)} d_l(f_k)|^2}{T \delta_{jk}^2}\right), \quad (8)$$

and $\mathbf{d}_{<j}$ denotes the data of channels preceding the j th channel, the parameters $\boldsymbol{\theta}_j, \boldsymbol{\delta}_j$ represent the set of θ_k and δ_k for the j th component of the multivariate time series, $\theta_{jl}^{(k)}$ represents the corresponding element in the matrix T_k for any $j > l$, and $\theta_{jl}^{(k)} = 1$ when $j = l$, $d_j(f_k)$ denotes the Fourier coefficient of the j th component, and $\mathbf{d}_j = (d_j(f_1), \dots, d_j(f_N))^T$. Then $\log(\delta_{jk}^2)$ and the real and imaginary parts of $\theta_{jl}^{(k)}$ are modeled by Demmler-Reinsch basis functions in terms of M truncated smoothing splines, given by

$$\Re(\theta_{jl}^{(k)}) = \alpha_{jl,0} + \alpha_{jl,1}f_k + \sum_{s=1}^{M-1} \psi(f_k)\alpha_{jl,s+1}, \quad (9)$$

$$\Im(\theta_{jl}^{(k)}) = \beta_{jl,0} + \beta_{jl,1}f_k + \sum_{s=1}^{M-1} \psi(f_k)\beta_{jl,s+1}, \quad (10)$$

$$\log \delta_{jk}^2 = \gamma_{j,0} + \gamma_{j,1}f_k + \sum_{s=1}^{M-1} \psi(f_k)\gamma_{j,s+1}, \quad (11)$$

where $\psi(x) = \sqrt{2} \cos(s\pi x)$, with s being an integer that ranges from 1 to $M - 1$, represents the spline basis function. The flexibility of the model can be adjusted by choosing the number of basis splines M . Thus, a flexible model for the spectral densities $\mathbf{S} = \mathbf{S}(\boldsymbol{\nu})$ depending on a parameter vector $\boldsymbol{\nu}$ that comprises all α, β , and γ parameters is constructed. Following Hu and Prado [23], discounted regularized horseshoe priors [38] are used for the α, β , and γ spline coefficients. This prior is adept at handling varying degrees of smoothness in the individual components of the spectral density matrix while avoiding overfitting (for details, refer to [23,38]). Thus, in line with Eq. (7), the subvector of the parameter vector $\boldsymbol{\nu}$ that contains all parameters for component j is denoted as $\boldsymbol{\nu}_j$. This allows us to decompose the posterior into the product of p posteriors for each individual component, given by

$$p(\boldsymbol{\nu}|\mathbf{d}) = \prod_{j=1}^p p_j(\boldsymbol{\nu}_j|\mathbf{d}_j, \mathbf{d}_{<j}). \quad (12)$$

This prescription would enable the application of the SGVB approach to each $p_j(\boldsymbol{\nu}_j|\mathbf{d}_j, \mathbf{d}_{<j})$ in parallel. Here, however, we apply the SGVB algorithm to the whole parameter vector $\boldsymbol{\nu}$ even though Sec. II C details the more general application to each $\boldsymbol{\nu}_j$.

C. Stochastic gradient variational Bayes

This section provides a brief review of SGVB, along with discussions on how to tune the learning rate and the number of basis splines.

1. Variational inference review

The fundamental concept of variational inference (VI) is to approximate the posterior distribution $p(\boldsymbol{\nu}|\mathbf{d})$ by using a surrogate distribution from a family of variational distributions $\mathcal{Q} = \{q_\phi(\boldsymbol{\nu}); \phi \in \Phi\}$, which depends on a parameter vector ϕ within a parameter space Φ . Typically, the variational family is selected for its simplicity and computational tractability. A product of normal distributions with mean μ_{ji} and variance σ_{ji}^2 , $i = 1, \dots, \dim(\boldsymbol{\nu}_j)$, is utilized for each vector $\boldsymbol{\nu}_j$. The goal of the variational approach is to identify the parameters $\phi_j = (\boldsymbol{\mu}_j, \boldsymbol{\xi}_j)$ with $\xi_{ji} = \log \sigma_{ji}^2$ for $i = 1, \dots, \dim(\boldsymbol{\nu}_j)$, that minimize the reverse Kullback-Leibler (KL) divergence between the variational family and the true posterior distribution, denoted as $d_{KL}(q_{\phi_j}||p_j)$, i.e.,

$$\begin{aligned} \phi_j^* &= \arg \min_{\phi_j} d_{KL}(q_{\phi_j}||p_j) \\ &= \arg \min_{\phi_j} \int \log \frac{q_{\phi_j}(\boldsymbol{\nu}_j)}{p_j(\boldsymbol{\nu}_j|\mathbf{d}_j, \mathbf{d}_{<j})} q_{\phi_j}(\boldsymbol{\nu}_j) d\boldsymbol{\nu}_j. \end{aligned} \quad (13)$$

The optimization algorithm employed in this work uses a SGVB approach, as described by Kingma and Welling [28], Xu *et al.* [39], Domke [40].

2. Choice of the number of basis functions

The variation of parameters ($\log \delta_{jk}^2, \Re(\theta_j^{(k)}), \Im(\theta_j^{(k)})$) across frequencies f_k is modeled using M truncated smoothing splines. Choosing an appropriate M is crucial: a low M may cause underfitting, while a high M can lead to overfitting and increased computational complexity due to the higher dimensionality of ϕ_j . To address overfitting concerns, a horseshoe prior on the coefficients acts as a regularization mechanism, effectively reducing the risk of overfitting [38].

The likelihood function [Eq. (8)] measures the consistency of the data with the spectral density, parametrized by ($\log \delta_{jk}^2, \Re(\theta_j^{(k)}), \Im(\theta_j^{(k)})$). We use the likelihood at the maximum likelihood estimate (MLE) to choose M . The likelihood at the MLE tends to increase with M , reflecting an improvement in model fit as the model complexity increases. When M is sufficiently large, the likelihood tends to stabilize or increase steadily. We chose M to prevent underfitting by conservatively selecting a slightly larger M than the point where the likelihood stabilizes.

3. Optimization of the learning rate

The KL minimizer ϕ_j^* [Eq. (13)] is equivalent to the maximizer of the evidence lower bound (ELBO) between q_{ϕ_j} and $p(\boldsymbol{\nu}_j, \mathbf{d}_j, \mathbf{d}_{<j})$, i.e.,

$$\begin{aligned} \text{ELBO}(q_{\phi_j}, p_j(\boldsymbol{\nu}_j|\mathbf{d}_j, \mathbf{d}_{<j})) \\ = \mathbb{E}_{\boldsymbol{\nu}_j \sim q_{\phi_j}(\boldsymbol{\nu}_j)} [\log p_j(\boldsymbol{\nu}_j, \mathbf{d}_j, \mathbf{d}_{<j}) - \log q_{\phi_j}(\boldsymbol{\nu}_j)]. \end{aligned} \quad (14)$$

The SGVB algorithm is used to approximate ϕ_j^* , and it consists of the next two steps:

- (1) Maximize $\log p(\boldsymbol{\nu}_j, \mathbf{d}_j, \mathbf{d}_{<j})$ with respect to $\boldsymbol{\nu}_j$, i.e., $\hat{\boldsymbol{\nu}}_j = \operatorname{argmax}_{\boldsymbol{\nu}_j} \log p(\boldsymbol{\nu}_j, \mathbf{d}_j, \mathbf{d}_{<j})$, and set $\hat{\boldsymbol{\mu}}_j = \hat{\boldsymbol{\nu}}_j$.
- (2) Maximize the ELBO [Eq. (14)] with respect to $\boldsymbol{\xi}_j$ for fixed $\hat{\boldsymbol{\mu}}_j$ from the first step, i.e., $\hat{\boldsymbol{\xi}}_j = \operatorname{argmax}_{\boldsymbol{\xi}_j} \mathbb{E}_{\boldsymbol{\nu}_j \sim q_{\hat{\boldsymbol{\mu}}_j, \boldsymbol{\xi}_j}} [\log p(\boldsymbol{\nu}_j, \mathbf{d}_j, \mathbf{d}_{<j}) - \log q_{\hat{\boldsymbol{\mu}}_j, \boldsymbol{\xi}_j}(\boldsymbol{\nu})]$.¹

The approximation for ϕ_j^* is $\hat{\phi}_j = (\hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\xi}}_j)$. This is equivalent to algorithm 1 in [23] up to and including phase 2. In the three-phase VB approach by [23], $\hat{\phi}$ was used as an initial point for the optimization in the fine-tuning phase 3 to get ϕ^* . The numerical study in [23] demonstrated that $\hat{\phi}$ was already close to the maximum, and the optimization criterion did not markedly improve during the phase 3 stage. Therefore, we omitted phase 3 for the simulation studies and ET data. However, the fine-tuning stage can be added if needed.

The performance of each maximizer depends on the learning rate; if the learning rate is too small or too large, the optimization algorithm may get stuck at a local maximum or take too long to find the global maximum. Thus, choosing the optimal learning rate is crucial. Recently, [41] proposed automated techniques for the blackbox VB in which the learning rate is adaptively decreased. The reliability of the stochastic optimization methods for VB due to substantial hand-tuning parameters to apply effectively were discussed in literature [41,42].

While Hu and Prado [23] selected a specific but arbitrary value for the learning rate, this study proposes a method to select the optimal learning rate and automate the selection procedure.

Let τ_1 and τ_2 be the learning rates associated with gradient ascent methods at the first and second step, respectively. As τ_2 has a broad range of plausible values, it is unnecessary to deviate from the default settings proposed by Hu and Prado [23]. Let $\Phi_{\tau_1} = (\hat{\phi}_1, \dots, \hat{\phi}_p)$ be the result after performing the two-step SGVB algorithm with τ_1 . Given the default setting for τ_2 , an appropriate τ_1 is obtained by maximizing the ELBO,

$$\tau_1^* = \operatorname{argmax}_{\tau_1} \mathbb{E}_{\boldsymbol{\nu} \sim q_{\Phi_{\tau_1}}(\boldsymbol{\nu})} [\log p(\boldsymbol{\nu}, \mathbf{d}) - \log q_{\Phi_{\tau_1}}(\boldsymbol{\nu})]. \quad (15)$$

This optimization is carried out over a continuous parameter space using the Python package HyperOpt [43] and the tree-structured Parzen estimator (TPE) algorithm [44]. From our preliminary study, the optimal learning rates obtained by fixing τ_2 coincide with those obtained when optimizing τ_1 and τ_2 jointly. The maximum number of iterations for the optimization procedure is set to 10 000.

¹ $q_{\hat{\boldsymbol{\mu}}_j, \boldsymbol{\xi}_j}$ is the variational distribution q_{ϕ_j} with $\phi_j = (\hat{\boldsymbol{\mu}}_j, \boldsymbol{\xi}_j)$.

D. Squared coherence

To quantify the frequency-dependent relationship between the channels of the multivariate ET time series, we employ the squared coherence, widely used in many fields [e.g., [45,46]]. Squared coherence, $C_{xy}(f_k)$, is a normalized measure of association between two time series at frequency f_k , ranging from 0 (no correlation) to 1 (perfect correlation). For two components of a multivariate time series, it is defined as

$$C_{xy}(f_k) = \frac{|\mathbf{S}_{xy}(f_k)|^2}{\mathbf{S}_{xx}(f_k)\mathbf{S}_{yy}(f_k)}, \quad (16)$$

where $\mathbf{S}_{xy}(f_k)$ is the cross-spectral density between components x and y , and $\mathbf{S}_{xx}(f_k)$, $\mathbf{S}_{yy}(f_k)$ are the spectral densities of components x and y , respectively, at frequency f_k ($x, y = 1, 2, \dots, p; x \neq y$), for $x, y = 1, 2, \dots, p$, with $x \neq y$.

III. SIMULATION STUDY

In order to test the efficiency and accuracy of SGVB, 500 independent realizations of a bivariate time series are generated from a vector autoregressive model of order 2 [VAR(2)] and a vector moving-average model of order 1 [VMA(1)], using three different sample sizes $n = 256; 512; 1024$ [refer to Liu *et al.* [Sec. IV.2, [22]] for the definitions of the VAR(2) and VMA(1) models]. The spectral densities are estimated using both the SGVB method and Liu *et al.*'s [22] vectorized nonparametrically corrected method (VNPC), an MCMC-based approach that samples from the exact posterior distribution without a variational approximation.²

In order to avoid underfitting in SGVB, an appropriate number of basis functions is determined for each dataset using the method discussed in Sec. II C. Figure 1 shows the log MLE (normalized for comparison between datasets) plotted against the number of basis functions. $M = 30$ is set as the log MLE enters a steady ascent phase and before any perturbations appear, thereby avoiding both underfitting across the three different dataset sizes. When $n = 256$, a rapid increase of the log-likelihood is observed as $M > 50$, indicating overfitting. As shown in Eqs. (9)–(11), the number of parameters is $4(M + 2)$ for a bivariate time series, which exceeds the number of data points.

Given an estimated spectral density matrix $\hat{\mathbf{S}}$ and the true spectral density matrix \mathbf{S}_0 , the accuracy of PSD estimates can be assessed using the L_2 error. The L_2 error provides a quantitative measure of the discrepancy between $\hat{\mathbf{S}}$ and \mathbf{S}_0 , encapsulating the overall difference between the estimated and true spectral density matrices across the frequency

²For the VNPC analyses, we use the software and settings provided by Liu *et al.* [22], specific for the VAR(2) and VMA(1) models.

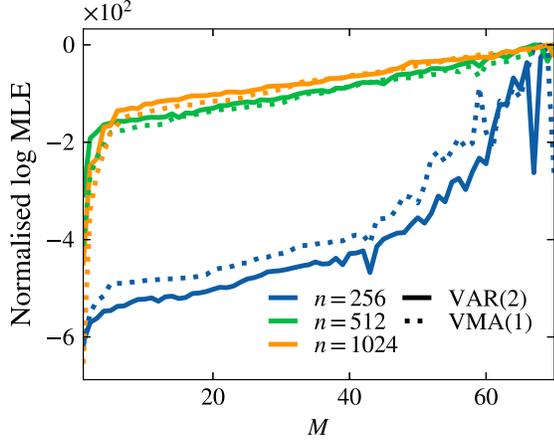


FIG. 1. The relationship between the number of basis functions (M) and the log-normalized likelihood at the MLE for VAR(2) and VMA(1) models (solid and dotted lines respectively) with different data lengths $n = 256$ (blue), $n = 512$ (green), and $n = 1024$ (orange). A fixed learning rate of 0.002 was used across all cases.

range. A smaller L_2 error indicates that $\hat{\mathbf{S}}$ closely approximates \mathbf{S}_0 , while a larger L_2 error reflects greater deviation. The L_2 error can be approximated as

$$\|\hat{\mathbf{S}} - \mathbf{S}_0\|_{L_2} \approx \left(\frac{1}{N} \sum_{k=1}^N \|\hat{\mathbf{S}}_0(f_k) - \mathbf{S}_0(f_k)\|^2 \right)^{1/2}, \quad (17)$$

where $\|\cdot\|$ denotes the Frobenius norm, which for a complex $p \times p$ matrix \mathbf{A} is defined by

$$\|\mathbf{A}\| = \sqrt{\sum_{i,j} |a_{ij}|^2}. \quad (18)$$

Figure 2 displays the L_2 error distributions from both the SGVB and VNPC methods across the different sample sizes. The L_2 errors are calculated by using the estimated spectral densities and the true spectral densities of the VAR(2) simulations (top panel) and the VMA(1) simulations (bottom panel), respectively.

Figure 3 displays the speedup factor of the SGVB compared to the VNPC method. Additionally, Table I presents the simulation study’s median L_2 errors, pointwise coverage, and median width of pointwise 90% credible regions, as well as the median computation time (in seconds).³

Figure 2 reveals that as the sample size increases, both methods achieve lower median L_2 errors, and the median L_2 errors are similar between the two methods. Additionally, Fig. 3 shows the median speedup factor for SGVB becomes

³Pointwise coverage is defined as the proportion of times the true spectral density value at each frequency is captured within the confidence interval across different realizations of the data.

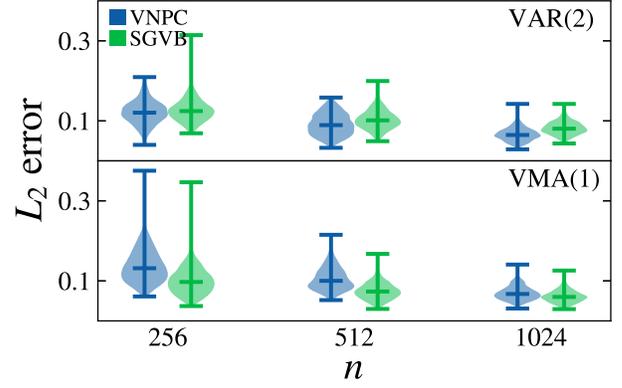


FIG. 2. Violin plots of VNPC and SGVB L_2 errors from 500 realizations for VAR(2) and VMA(1) models, for different data lengths n . The SGVB method uses optimized values for M and τ_1 .

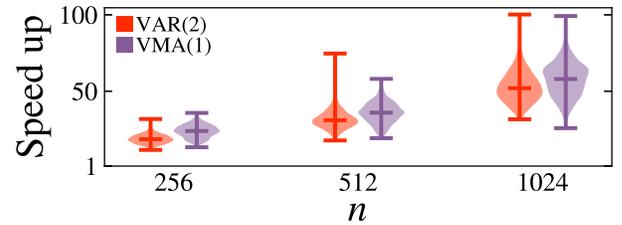


FIG. 3. Violin plots of the speed up gained by using SGVB compared to VNPC from 500 realizations for VAR(2) and VMA(1) models, for different data lengths n . The SGVB method uses optimized values for M and τ_1 .

increasingly pronounced, ultimately surpassing VNPC by more than 50 times, as n is increased. Notably, when $n = 1024$, SGVB’s accuracy is only marginally lower than VNPC’s, yet it reduces median computation time by a factor of approximately 50. The simulation study provides empirical evidence for the consistency of the SGVB approach, demonstrating that as sample size grows, the posterior distributions contracts around the true spectral density, following the same trend as the VNPC approach. However, it does so at a fraction of the computation time.

Finally, Table I illustrates that, even though VNPC method consistently outperforms SGVB across all sample sizes in terms of pointwise coverage, typically achieving 83%–90% coverage compared to SGVB’s 61%–66%, SGVB generally produces narrower credible regions. The limited coverage and narrow credible intervals observed from the SGVB results can be attributed to the mean field approximation [26] which assumes independent components ignoring correlations between parameters, and using the KL divergence as distance measure. The combined effects tend to underestimate the marginal variances of the target distribution [27,47]. To address these limitations, researchers have explored more flexible surrogate distributions and a broader range of divergence measures in variational inference [26,48–51].

TABLE I. Comparison of median L_2 errors, empirical pointwise coverage, median width of pointwise 90% credible intervals, and median computation time (in seconds) with their respective MAD (median absolute deviation) for 500 simulations using VNPC and SGVB methods across different sample sizes ($n = 256, 512,$ and 1024) for both VAR(2) and VMA(1) models.

		$n = 256$		$n = 512$		$n = 1024$	
		VNPC	SGVB	VNPC	SGVB	VNPC	SGVB
VAR(2)	L_2 error	0.12 ± 0.02	0.12 ± 0.02	0.10 ± 0.02	0.09 ± 0.02	0.08 ± 0.01	0.06 ± 0.01
	Pointwise coverage	0.87 ± 0.06	0.62 ± 0.11	0.85 ± 0.06	0.66 ± 0.10	0.84 ± 0.06	0.67 ± 0.07
	\mathbf{S}_{11} CI width	0.09 ± 0.01	0.06 ± 0.01	0.06 ± 0.01	0.04 ± 0.01	0.04 ± 0.00	0.03 ± 0.00
	$\Re\mathbf{S}_{12}$ CI width	0.09 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.04 ± 0.01	0.05 ± 0.00	0.03 ± 0.00
	$\Im\mathbf{S}_{12}$ CI width	0.07 ± 0.01	0.04 ± 0.01	0.06 ± 0.01	0.04 ± 0.01	0.04 ± 0.00	0.03 ± 0.00
	\mathbf{S}_{22} CI width	0.13 ± 0.01	0.07 ± 0.01	0.09 ± 0.01	0.05 ± 0.01	0.07 ± 0.01	0.04 ± 0.00
	Time [s]	$5.4 \text{ K} \pm 0.6 \text{ K}$	280 ± 30	$9.4 \text{ K} \pm 1.3 \text{ K}$	300 ± 30	$18.4 \text{ K} \pm 2.3 \text{ K}$	350 ± 30
VMA(1)	L_2 error	0.10 ± 0.03	0.13 ± 0.03	0.07 ± 0.02	0.10 ± 0.02	0.06 ± 0.01	0.07 ± 0.01
	Pointwise coverage	0.92 ± 0.07	0.64 ± 0.10	0.92 ± 0.08	0.63 ± 0.09	0.87 ± 0.10	0.67 ± 0.10
	\mathbf{S}_{11} CI width	0.12 ± 0.02	0.09 ± 0.02	0.08 ± 0.01	0.07 ± 0.01	0.06 ± 0.01	0.05 ± 0.01
	$\Re\mathbf{S}_{12}$ CI width	0.08 ± 0.01	0.06 ± 0.01	0.05 ± 0.01	0.04 ± 0.01	0.04 ± 0.00	0.03 ± 0.00
	$\Im\mathbf{S}_{12}$ CI width	0.07 ± 0.01	0.04 ± 0.01	0.06 ± 0.01	0.03 ± 0.00	0.03 ± 0.00	0.03 ± 0.00
	\mathbf{S}_{22} CI width	0.17 ± 0.03	0.18 ± 0.04	0.12 ± 0.02	0.13 ± 0.02	0.08 ± 0.01	0.10 ± 0.01
	Time [s]	$4.2 \text{ K} \pm 0.3 \text{ K}$	180 ± 30	$6.9 \text{ K} \pm 0.5 \text{ K}$	190 ± 30	$12.6 \text{ K} \pm 1.0 \text{ K}$	220 ± 40

Additionally, studies have investigated model misspecification and developed assessment methods for approximation inference [52–55]. While more sophisticated frameworks yield improved posterior approximations, they often demand greater computational resources. In this work, we prioritize computational efficiency while maintaining an acceptable level of accuracy.

IV. APPLICATION TO ET NOISE

A. Data generation

Three independent realizations of Gaussian noise were synthesized (one for each of the XYZ channels), each spanning 2000 s, spectrally shaped to match the design sensitivity of the ET xylophone configuration [56,57].⁴ The noise was sampled at a frequency of 2048 Hz, resulting in a

multivariate time series comprising 4096 k data points per channel. In the rest of the paper, we refer to this colored Gaussian noise as *ET noise*.

Following Janssens *et al.* [10], nonidentical correlated noise in the X, Y, and Z channels is simulated by injecting colored Gaussian noise characterized by frequency-domain Gaussian peaks. The power spectral density is given by

$$\mathbf{S}_n^{GP}(f, \mu, A) = \left(\frac{A}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(f - \mu)^2 \right\} \right)^2, \quad (19)$$

where A represents the amplitude, and μ the frequency peak location.

Gaussian noise is injected at specific frequencies in each channel as follows:

$$\begin{aligned} \mathbf{X}: \mathbf{S}_n^{GP}(\mu = 10 \text{ Hz}, A = 4 \times 10^{-24}), & \quad \mathbf{S}_n^{GP}(\mu = 50 \text{ Hz}, A = 2 \times 10^{-24}), \\ \mathbf{Y}: \mathbf{S}_n^{GP}(\mu = 10 \text{ Hz}, A = 4 \times 10^{-24}), & \quad \mathbf{S}_n^{GP}(\mu = 90 \text{ Hz}, A = 1.5 \times 10^{-24}), \\ \mathbf{Z}: \mathbf{S}_n^{GP}(\mu = 50 \text{ Hz}, A = 2 \times 10^{-24}), & \quad \mathbf{S}_n^{GP}(\mu = 90 \text{ Hz}, A = 1.5 \times 10^{-24}). \end{aligned}$$

To introduce correlated noise between channel pairs, we utilize identical Gaussian peaks at matching frequencies across the paired channels. In contrast, for uncorrelated

noise scenarios, independent Gaussian peaks are simulated for each channel, ensuring that there is no cross-channel correlation.

As stated in Janssens *et al.* [10], this dataset is inherently simplified due to the distinct nature of the correlated noise terms, which are highly differentiable. Nevertheless, it serves as a valuable initial demonstration of our approach. Future work aims to investigate more realistic and complex noise scenarios, such as the presence of correlated magnetic Janssens *et al.* [13,14] and/or Newtonian noise [11,12] as

⁴Note that the sensitivity curve used here is the previously called ‘‘ET-D’’ sensitivity curve. We choose this curve rather than the updated version presented by Branchesi *et al.* [58], to allow a direct comparison with previous work [10]. The small difference between PSD curves will have no impact on the key conclusions of the applicability of the work presented in this paper.

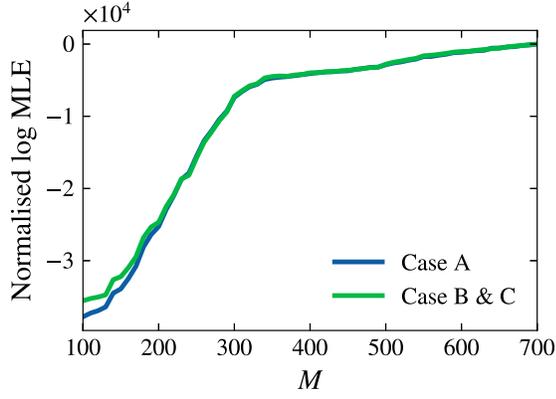


FIG. 4. Relationship between the number of basis functions (M) and the normalized log maximum likelihood estimate (MLE) for different scenarios of Einstein Telescope noise analysis. Case A (blue line) represents correlated noise, while Cases B and C (green line) represent uncorrelated noise scenarios.

well as (multiple) GW signals. However, this is considered beyond the scope of this work where we want to lay the foundation for the estimation framework.

B. Data analysis

Three analyses are performed to estimate the power spectral density (PSD) and the corresponding squared coherences under the following different noise conditions,

- (1) *Case A*: ET noise with correlated Gaussian peaks,
- (2) *Case B*: ET noise with uncorrelated Gaussian peaks,
- (3) *Case C*: ET noise with uncorrelated Gaussian peaks, assuming independence between channels.

The data are divided into 125 equal blocks, each consisting of 16 384 frequency points and a maximum frequency of 1024 Hz for the blocked Whittle likelihood estimation [Eq. (5)]. Each Whittle likelihood block uses identical basis function expressions defined in Sec. II B for the spectral density matrix.

Figure 4 displays the log maximum likelihood estimates (MLE) across a range of M for each case. As the log MLE stabilizes around $M \sim 400$, we set $M = 450$. Using the methodology discussed in Sec. II, the PSDs for each case are estimated and plotted in Figs. 5 and 6. For each case, the diagonal subplots represent the spectral densities for X, Y, and Z channels. Since the off-diagonal elements of the PSD matrix at each frequency are complex numbers, with the elements above the diagonal being the conjugates of those below, we plot the real part of the off-diagonal elements in the upper subplot and the imaginary part in the lower subplot. Additionally, the squared coherences are plotted for the cases in Fig. 7. The plot shows the squared coherences estimation between any pair of channels for case A and case B.

In all three cases, prominent spectral peaks are observed at the specific frequencies in each channel, consistent with the injected Gaussian noise peaks. This demonstrates the

ability of our proposed method to accurately capture and identify these peaks. We note that our PSD estimates exhibit oscillatory features around the Gaussian peaks; this is most likely linked to the relatively loud and sharp Gaussian peak features. This highlights one of the potential weaknesses of the SGVB method: its potential difficulty in fitting very sharp features. For more realistic data sets including correlated noise from magnetic and/or Newtonian noise origin, this will most likely form less of an issue due to the smooth, power-law-like behavior of such noise sources Janssens *et al.* [14], Janssens *et al.* [10–12]. However, this could form an issue with actual detector data which is contaminated with a large amount of instrumental spectral noise lines. Future work will aim to investigate the capabilities of the SGVB method and its efficiency in deal with such complexity using real GW data and whether precleaning steps should be undertaken. However, the dramatic speedup of PSD estimation using the SGVB remains an extremely potent benefit for next generation GW detectors. Without faster novel noise estimates techniques, such as the one proposed in this work, the resulting science will very likely be limited by available computation time.

In case A, where correlated noise is intentionally introduced, substantial coherence between channel pairs is observed at their shared peak frequencies (see the top panel of Fig. 7). Specifically, high coherence is evident between channels X and Y at 10 Hz, X and Z at 50 Hz, and Y and Z at 90 Hz. The median estimates for the squared coherence aligns with the true coherence, demonstrating that our estimate accurately reflects the underlying correlated noise structure. This alignment validates the effectiveness of our spectral analysis in capturing interchannel relationships.

Case B, which considers ET noise data with nominally uncorrelated noise, reveals minor fluctuations in the off-diagonal subplots representing the real and imaginary parts of the estimated spectral densities, despite the absence of deliberately injected correlations. These fluctuations manifest as small coherence values (median $C_{xy} < 0.008$) that fluctuate around the true zero coherence in the corresponding squared coherence plots of Fig. 7 (visible as they are plotted on a log scale). Notably, the median squared coherence fluctuations are lower than those expected from uncorrelated white noise data, which scale as approximately the inverse of the number of data segments used to average over. In this case, i.e., $1/N_b = 0.008$. The squared coherence estimates based on the SGVB method are below this estimate and hence are consistent with zero coherence.

In case C, where the three channels are assumed independent, the model assumes zero cross spectrum (the XY, YZ, XZ components) in the PSD estimates. Consequently, the squared coherence between any pair of channels is uniformly zero across all frequencies, which

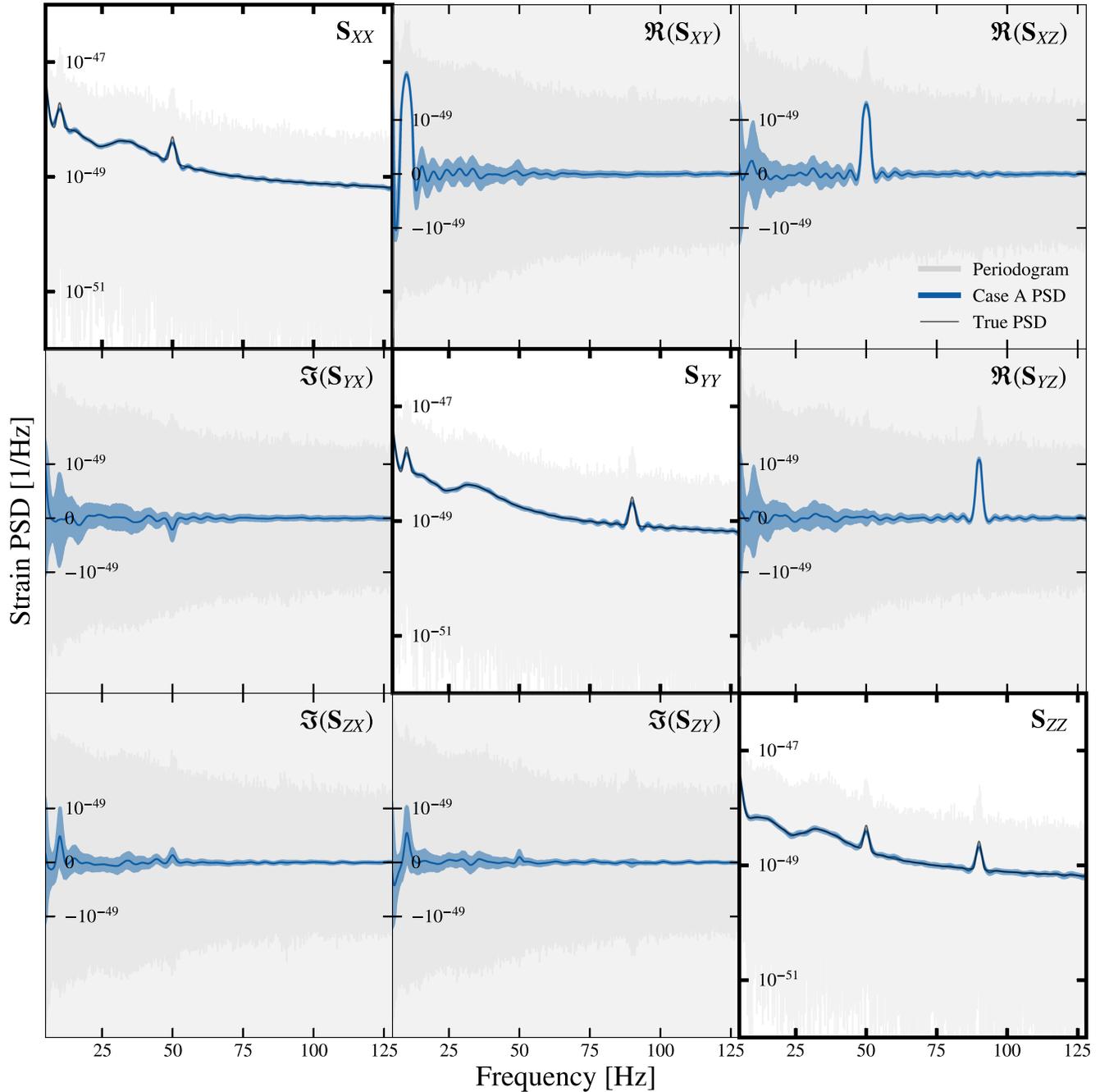


FIG. 5. Case A PSD. The plot shows the periodogram (gray), the PSD estimates (colored, with 90% credible interval shaded), and the true PSD (black curve in the diagonals). The upper triangle subplots show the real parts of the off-diagonal elements of the spectral density matrix, while the lower triangle subplots show the imaginary parts of the off-diagonal elements.

is consistent with theoretical expectations for uncorrelated channels, providing a baseline to assess the degree of correlation in the other cases. Comparing cases B and C highlights an important feature about the effectiveness of our SGVB model. As one would expect, a model with less complexity (case C) typically yields a more accurate estimate compared to a more complex model (case B). More concretely, the average root mean square deviation

(RMSD)⁵ over the frequency band 5 to 128 Hz is 9.773×10^{-50} for case B and 9.725×10^{-50} for case C. The difference between both different scenarios is relatively

⁵The average RMSD is calculated as the mean of the square root of the mean squared differences between the estimated median PSD and the true PSD across the frequency band for X, Y, and Z channels.

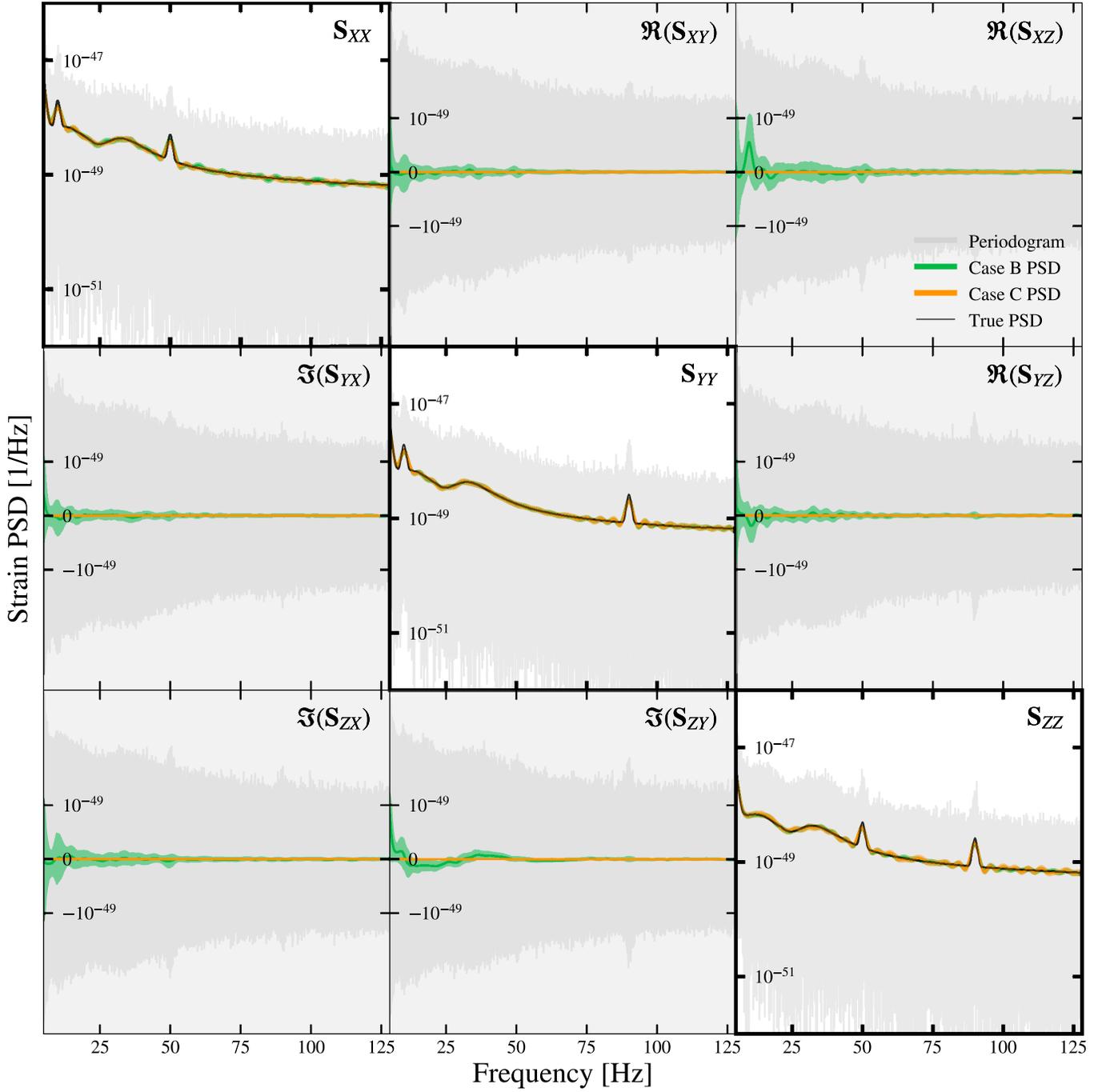


FIG. 6. Cases B and C PSDs. The plot shows the periodogram (gray), the PSD estimates (colored, with 90% credible interval shaded), and the true PSD (black curve in the diagonals). The upper triangle subplots show the real parts of the off-diagonal elements of the spectral density matrix, while the lower triangle subplots show the imaginary parts of the off-diagonal elements.

small, case C's RMSD is about 0.5% lower than that of case B, showcasing the huge potential and flexibility of the SGVB approach in estimating detector noise whether correlated noise is present or not.

To evaluate the accuracy of the estimated spectral density matrices, we whiten the Fourier-transformed data for cases A, B, and C using the estimated PSD and test the real and imaginary whitened data for multivariate normality [59].

The whitening process was performed for each data segment and at every frequency f_k , using the transformation,

$$\mathbf{d}_i^w(f_k) = \mathbf{S}^{-1/2}(f_k)\mathbf{d}_i(f_k), \quad (20)$$

where $\mathbf{d}_i^w(f_k)$ represents the whitened Fourier-transformed data for the i th segment at frequency f_k , and $\mathbf{S}^{-1/2}(f_k)$ is the inverse square root of the estimated PSD matrix at f_k .

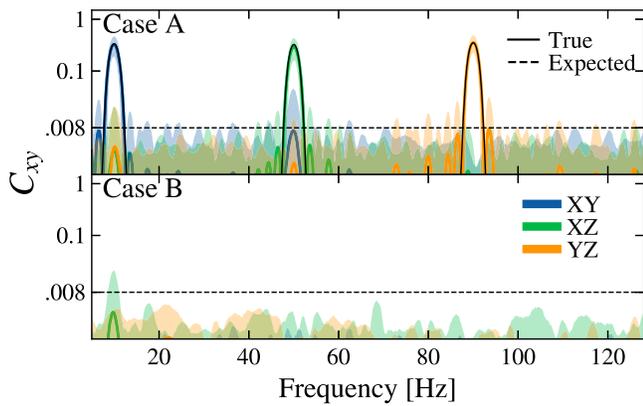


FIG. 7. Squared coherence estimation for any pair of channels in cases A (top) and B (bottom). The solid curves represent the median estimated squared coherences, while the shaded areas indicate the corresponding 90% credible interval. The black solid line is the true squared coherence, and the black dashed line is the expected coherence from uncorrelated white noise data.

Here, $i = 1, 2, \dots, 125$ denotes the index of the data segments.

Table II summarizes the number of segments passing the normality test at a 5% significance level. As expected, with a 5% significance threshold, false positives occur in 5% of tests under the null hypothesis. The results align with these expectations, supporting the validity of the PSD estimation.

V. CONCLUSIONS

Recent work has pointed out the importance of taking into account correlated interferometer network noise when estimating the parameters of gravitational wave signals observed by ground-based detectors such as ET or space-based detectors such as LISA [8–10]. However, so far, no Bayesian statistical methodology has been suggested to estimate the multivariate noise spectral density of a detector network. Such a method will be important, in particular, when estimating residual noise in a joint on-source estimation or global fit approach. It can form a building block for global fit models such as GLASS [30] and Erebor [31] that are cyclically conditioning on resolved signals and noise spectral density estimates. In this study, we propose a computationally efficient variational Bayesian approach for estimating the spectral density of multivariate time series. It is important to note that this technique does not assume any restrictive parametric form of the spectral density, allowing it to adapt to any shape of the PSD. To handle the large length of GW network measurements, we use a blocked Whittle likelihood. We provide a hyperparameter optimization method and guidelines to tune the settings of the method. Simulation studies presented for VAR(2) and VMA(1) demonstrate that our approach achieves accuracy comparable to MCMC while significantly reducing

TABLE II. Number of whitened data segments passing the normality test for real and imaginary parts after whitening, for each case and across all 125 segments.

	$\Re(\mathbf{d}_i^w)$	$\Im(\mathbf{d}_i^w)$
Case A	118	121
Case B	119	119
Case C	119	119

computation time, although with potentially narrower credible intervals.

The application of our method to simulated ET noise data revealed its ability to accurately identify and characterize spectral features across different noise scenarios. In the correlated noise case, our approach successfully detected the injected Gaussian peaks and quantified the coherence between channels at specific frequencies.

These findings have significant implications for GW data analysis. The ability to efficiently and accurately estimate spectral densities and interchannel correlations is crucial for optimizing detection algorithms and understanding noise characteristics in GW detectors. Our method’s computational efficiency makes it particularly suitable for analyzing large-scale datasets expected from next-generation detectors such as ET and LISA. While this paper presented an investigation into the accuracy and efficiency of the SGVB approach to a noise-only scenario, it only lays the foundation for future investigations of on-source modeling in the presence of multiple and potentially overlapping signals. Many aspects remain to be explored: the best way to combine signal parameter estimation methods with the SGVB approach for simultaneous PSD estimation, the impact of data gaps and instrumental glitches, and strategies for their mitigation.

Future work could explore the application of this method to real GW detector data and its extension to simultaneously analyzing a stochastic GW background signal. Jointly analyzing the six channels of two detector networks such as TianQin and LISA could provide a more flexible alternative to cross-correlation detection [60]. Further investigations could examine the implications of these spectral characteristics on GW detection sensitivity and methods to mitigate correlated noise effects in ET or LISA. Extending the analysis to longer time series or different ET configurations could provide insights into the stability and generalizability of these spectral features. Additionally, investigating the method’s performance on GPUs could further validate its scalability, robustness and versatility.

In conclusion, the variational Bayes approach offers a promising tool for spectral density estimation in GW data analysis, combining accuracy with computational efficiency. This method has the potential to enhance our ability to characterize detector noise and ultimately improve GW detection capabilities.

Software used: Python [61], TensorFlow Probability [62,63] NumPy [64], SciPy [65], PANDAS [66], Matplotlib [67], Jupyter Book [68]. Documentation and examples of the software can be found at [69,70].

ACKNOWLEDGMENTS

We thank Zhixiong Hu and Racquel Prado for making their code publicly available and for helpful discussions. A. V., J. E. L., N. C., P. M.-R., and R. M. gratefully acknowledge support by the Marsden Fund Council Grant No. MFP-UOA2131 from New Zealand Government funding, managed by the Royal Society Te

Apārangi. K. J. was supported by FWO-Vlaanderen via Grant No. 11C5720N during part of this work. For this project, N. C. has received financial support from the CNRS through the MITI interdisciplinary programs. We thank the New Zealand eScience Infrastructure (NeSI) [71] for the use of their high performance computing facilities and the Centre for eResearch at the University of Auckland for their technical support.

DATA AVAILABILITY

The data that support the findings of this article are openly available [70,72,73].

-
- [1] M. Punturo, M. Abernathy, F. Acernese, B. Allen, N. Andersson, K. Arun *et al.*, The Einstein Telescope: A third-generation gravitational wave observatory, *Classical Quantum Gravity* **27**, 194002 (2010).
- [2] M. Evans, R. X. Adhikari, C. Afle, S. W. Ballmer, S. Biscoveanu, S. Borhanian *et al.*, A horizon study for cosmic explorer: Science, observatories, and community, *arXiv:2109.09882*.
- [3] P. Amaro-Seoane *et al.*, Laser interferometer space antenna, *arXiv:1702.00786*.
- [4] ET steering committee, Einstein telescope: Science case, design study and feasibility report (2020).
- [5] M. Maggiore, C. V. D. Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard *et al.*, Science case for the Einstein Telescope, *J. Cosmol. Astropart. Phys.* **03** (2020) 050.
- [6] M. Branchesi, M. Maggiore, D. Alonso, C. Badger, B. Banerjee, F. Beirnaert *et al.*, Science with the Einstein Telescope: A comparison of different designs, *J. Cosmol. Astropart. Phys.* **07** (2023) 068.
- [7] P. Amaro-Seoane, J. Andrews, M. Arca Sedda, A. Askar, Q. Baghi, R. Balasov *et al.*, Astrophysics with the laser interferometer space antenna, *Living Rev. Relativity* **26**, 2 (2023).
- [8] ET steering committee, ET Design Report Update 2020 (2020).
- [9] M. Colpi, K. Danzmann, M. Hewitson, K. Holley-Bockelmann, P. Jetzer *et al.*, LISA definition study report, *arXiv:2402.07571*.
- [10] K. Janssens, G. Boileau, M.-A. Bizouard, N. Christensen, T. Regimbau, and N. v. Remortel, Formalism for power spectral density estimation for non-identical and correlated noise using the null channel in Einstein Telescope, *Eur. Phys. J. Plus* **138**, 352 (2023).
- [11] K. Janssens, G. Boileau, N. Christensen, F. Badaracco, and N. van Remortel, Impact of correlated seismic and correlated Newtonian noise on the Einstein Telescope, *Phys. Rev. D* **106**, 042008 (2022).
- [12] K. Janssens *et al.*, Correlated 0.01–40 Hz seismic and Newtonian noise and its impact on future gravitational-wave detectors, *Phys. Rev. D* **109**, 102002 (2024).
- [13] K. Janssens, M. Ball, R. M. S. Schofield, N. Christensen, R. Frey, N. van Remortel *et al.*, Correlated 1–1000 Hz magnetic field fluctuations from lightning over earth-scale distances and their impact on gravitational wave searches, *Phys. Rev. D* **107**, 022004 (2023).
- [14] K. Janssens, K. Martinovic, N. Christensen, P. M. Meyers, and M. Sakellariadou, Impact of Schumann resonances on the Einstein Telescope and projections for the magnetic coupling function, *Phys. Rev. D* **104**, 122006 (2021).
- [15] G. Anderson, J. Anderson, M. Anderson *et al.*, Experimental results from the ST7 mission on LISA Pathfinder, *Phys. Rev. D* **98**, 102005 (2018).
- [16] M. Armano, H. Audley, J. Baird *et al.* (LISA Pathfinder Collaboration), LISA Pathfinder micronewton cold gas thrusters: In-flight characterization, *Phys. Rev. D* **99**, 122003 (2019).
- [17] E. Thrane, N. Christensen, and R. M. S. Schofield, Correlated magnetic noise in global networks of gravitational-wave detectors: Observations and implications, *Phys. Rev. D* **87**, 123009 (2013).
- [18] N. Christensen, Stochastic gravitational wave backgrounds, *Rep. Prog. Phys.* **82**, 016903 (2018).
- [19] G. Boileau, N. Christensen, and R. Meyer, Figures of merit for a stochastic gravitational-wave background measurement by LISA: Implications of LISA Pathfinder noise correlations, *Phys. Rev. D* **106**, 063025 (2022).
- [20] F. Cireddu, M. Wils, I. C. F. Wong, P. T. H. Pang, T. G. F. Li, and W. Del Pozzo, Likelihood for a network of gravitational-wave detectors with correlated noise, *Phys. Rev. D* **110**, 104060 (2024).
- [21] A. Meier, C. Kirch, and R. Meyer, Bayesian nonparametric analysis of multivariate time series: A matrix gamma process approach, *J. Multivariate Anal.* **175**, 104560 (2020).
- [22] Y. Liu, C. Kirch, J. E. Lee, and R. Meyer, A nonparametrically corrected likelihood for Bayesian spectral analysis of multivariate time series, *Comput. Stat. Data Anal.* **199**, 108010 (2024).
- [23] Z. Hu and R. Prado, Fast Bayesian inference on spectral analysis of multivariate stationary time series, *Comput. Stat. Data Anal.* **178**, 107596 (2023).

- [24] M. I. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul, Introduction to variational methods for graphical models, *Mach. Learn.* **37**, 183 (1999).
- [25] M. J. Wainwright and M. I. Jordan, Graphical models, exponential families, and variational inference, *Found. Trends Mach. Learn.* **1**, 1 (2008).
- [26] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, Variational inference: A review for statisticians, *J. Am. Stat. Assoc.* **112**, 859 (2017).
- [27] D. M. Blei and M. I. Jordan, Variational inference for Dirichlet process mixtures, *Bayesian Anal.* **1**, 121 (2006).
- [28] D. P. Kingma and M. Welling, Auto-encoding variational Bayes, [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
- [29] M. Vallisneri, M. Crisostomi, A. D. Johnson, and P. M. Meyers, Rapid parameter estimation for pulsar-timing-array datasets with variational inference and normalizing flows, [arXiv:2405.08857](https://arxiv.org/abs/2405.08857).
- [30] T. B. Littenberg and N. J. Cornish, Prototype global analysis of LISA data with multiple source types, *Phys. Rev. D* **107**, 063004 (2023).
- [31] M. L. Katz, N. Karnesis, N. Korsakova, J. R. Gair, and N. Stergioulas, An efficient GPU-accelerated multi-source global fit pipeline for LISA data analysis, *Phys. Rev. D* **111**, 024060 (2025).
- [32] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, *Pub. Astron. Soc. Aust.* **36**, e010 (2019).
- [33] N. Christensen and R. Meyer, Parameter estimation with gravitational waves, *Rev. Mod. Phys.* **94**, 025001 (2022).
- [34] S. N. Lahiri, *Resampling Methods for Dependent Data*, Springer Series in Statistics (Springer, New York, New York, NY, 2003).
- [35] R. Rosati and T. B. Littenberg, Prototype stochastic gravitational wave background recovery in the LISA global fit residual, [arXiv:2410.17180](https://arxiv.org/abs/2410.17180).
- [36] A. W. Criswell, S. Rieck, and V. Mandic, Templated anisotropic analyses of the LISA galactic foreground, *Phys. Rev. D* **111**, 023025 (2025).
- [37] O. Rosen and D. S. Stoffer, Automatic estimation of multivariate spectra via smoothing splines, *Biometrika* **94**, 335 (2007).
- [38] J. Piironen and A. Vehtari, Sparsity information and regularization in the horseshoe and other shrinkage priors, *Electron. J. Stat.* **11**, 5018 (2017).
- [39] M. Xu, M. Quiroz, R. Kohn, and S. A. Sisson, Variance reduction properties of the reparameterization trick, in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research*, edited by K. Chaudhuri and M. Sugiyama (PMLR, Brookline, MA, 2019), Vol. 89, pp. 2711–2720.
- [40] J. Domke, Provable gradient variance guarantees for black-box variational inference, in *Advances in Neural Information Processing Systems*, Vol. 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (NeurIPS, Red Hook, NY, 2019).
- [41] M. Welandawe, M. R. Andersen, A. Vehtari, and J. H. Huggins, A framework for improving the reliability of black-box variational inference, *J. Mach. Learn. Res.* **25**, 1 (2024), <http://jmlr.org/papers/v25/22-0327.html>.
- [42] A. Agrawal, D. Sheldon, and J. Domke, Advances in black-box VI: Normalizing flows, importance weighting, and optimization, in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)* (Curran Associates, Inc., Red Hook, NY, 2020).
- [43] J. Bergstra, D. Yamins, and D. Cox, Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures, in *Proceedings of the 30th International Conference on Machine Learning, Proceedings of Machine Learning Research*, edited by S. Dasgupta and D. McAllester (PMLR, Atlanta, Georgia, USA, 2013), Vol. 28, pp. 115–123.
- [44] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, Algorithms for hyper-parameter optimization, in *Advances in Neural Information Processing Systems*, edited by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger (NeurIPS, Red Hook, NY, 2011), Vol. 24.
- [45] V. Sakkalis, Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG, *Computers in Biology and Medicine* **41**, 1110 (2011).
- [46] C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* **37**, 424 (1969).
- [47] B. Wang and D. M. Titterton, Inadequacy of interval estimates corresponding to variational Bayesian approximations, in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, edited by R. G. Cowell and Z. Ghahramani (PMLR, Brookline, MA, 2005), pp. 373–380.
- [48] D. P. Kingma and M. Welling, Auto-encoding variational Bayes, [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
- [49] R. Danilo and M. Shakir, Variational inference with normalizing flows, in *Proceedings of the 32nd International Conference on Machine Learning, PMLR* (PMLR, Brookline, MA, 2015), Vol. 37, pp. 1530–1538.
- [50] Y. Yang, D. Pati, and A. Bhattacharya, α -variational inference with statistical guarantees, *Ann. Stat.* **48**, 886 (2020).
- [51] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, beta-VAE: Learning basic visual concepts with a constrained variational framework, in *International Conference on Learning Representations* (Toulon, France, 2017), <https://openreview.net/forum?id=Sy2fzU9gl>.
- [52] Y. Wang and D. Blei, Variational Bayes under model misspecification, in *Advances in Neural Information Processing Systems*, Vol. 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019).
- [53] Y. Yao, A. Vehtari, D. Simpson, and A. Gelman, Yes, but did it work?: Evaluating variational inference, in *Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research*, edited by J. Dy and A. Krause (PMLR, Brookline, MA, 2018), Vol. 80, pp. 5581–5590.
- [54] J. Lee, G. K. Nicholls, and R. J. Ryder, Calibration procedures for approximate Bayesian credible sets, *Bayesian Anal.* **14**, 1245 (2019).

- [55] X. Yu, D. Nott, M.-N. Tranc, and N. Kleind, Assessment and adjustment of approximate inference algorithms using the law of total variance, *J. Comput. Graph. Stat.* **30**, 977 (2021).
- [56] S. Hild, S. Chelkowski, A. Freise, J. Franc, N. Morgado, R. Flaminio, and R. DeSalvo, A xylophone configuration for a third-generation gravitational wave detector, *Classical Quantum Gravity* **27**, 015003 (2009).
- [57] S. Hild *et al.*, Sensitivity studies for third-generation gravitational wave observatories, *Classical Quantum Gravity* **28**, 094013 (2011).
- [58] M. Branchesi, M. Maggiore, D. Alonso, C. Badger, B. Banerjee, F. Beirnaert, E. Belgacem, S. Bhagwat, G. Boileau, S. Borhanian *et al.*, Science with the Einstein Telescope: A comparison of different designs, *J. Cosmol. Astropart. Phys.* **07** (2023) 068.
- [59] N. Henze and B. Zirkler, A class of invariant consistent tests for multivariate normality, *Commun. Stat., Theory Methods* **19**, 3595 (1990).
- [60] Z.-C. Liang, Z.-Y. Li, E.-K. Li, J. dong Zhang, and Y.-M. Hu, Unveiling a multi-component stochastic gravitational-wave background with the Tianqin + LISA network, [arXiv:2409.00778](https://arxiv.org/abs/2409.00778).
- [61] G. van Rossum and F.L. Drake Jr, Python programming language, Python Software Foundation, <https://www.python.org/> (2020).
- [62] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.
- [63] J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, and R. A. Saurous, TensorFlow distributions, [arXiv:1711.10604](https://arxiv.org/abs/1711.10604).
- [64] C. R. Harris *et al.*, Array programming with NumPy, *Nature (London)* **585**, 357 (2020).
- [65] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261 (2020).
- [66] T. pandas development team, PANDAS-dev/PANDAS: PANDAS, [10.5281/zenodo.3509134](https://zenodo.org/record/3509134) (2020).
- [67] J. D. Hunter, Matplotlib: A 2d graphics environment, *Comput. Sci. Eng.* **9**, 90 (2007).
- [68] The Executable Book Project, Jupyter Book, <https://jupyterbook.org/> (2020).
- [69] https://nz-gravity.github.io/sgvb_psd/
- [70] J. Liu, A. Vajpeyi, R. Meyer, K. Janssens, J.E. Lee, P. Maturana-Russel, N. Christensen, and Y. Liu, [sgvb_psd](https://github.com/nz-gravity/sgvb_psd), https://github.com/nz-gravity/sgvb_psd (2024).
- [71] <https://www.nesi.org.nz>
- [72] J. Liu, A. Vajpeyi, R. Meyer, K. Janssens, J.E. Lee, P. Maturana-Russel, N. Christensen, and Y. Liu, Raw dataset for [sgvb_psd](https://github.com/nz-gravity/sgvb_psd), https://github.com/nz-gravity/sgvb_psd_paper/tree/main/data (2024).
- [73] J. Liu, A. Vajpeyi, R. Meyer, K. Janssens, J.E. Lee, P. Maturana-Russel, N. Christensen, and Y. Liu, Dataset accompanying “Variational inference for correlated gravitational wave detector network noise,” [10.5281/zenodo.13789536](https://zenodo.org/record/13789536) (2024).