Gravitational-wave astronomy with an uncertain noise power spectral density

Colm Talbot[1, 2, 3, a] and Eric Thrane[2, 3]

[1]*LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA*
[2]*Monash Centre for Astrophysics, School of Physics and Astronomy, Monash University, VIC 3800, Australia*
[3]*OzGrav: The ARC Centre of Excellence for Gravitational-Wave Discovery, Clayton, VIC 3800, Australia*

In order to extract information about the properties of compact binaries, we must estimate the noise power spectral density of gravitational-wave data, which depends on the properties of the gravitational-wave detector. In practice, it is not possible to know this perfectly, only to estimate it from the data. Multiple estimation methods are commonly used and each has a corresponding statistical uncertainty. However, this uncertainty is widely ignored when measuring the physical parameters describing compact binary coalescences, and the appropriate likelihoods which account for the uncertainty are not well known. In order to perform increasingly precise astrophysical inference and model selection, it will be essential to account for this uncertainty. In this work, we derive the correct likelihood for one of the most widely used estimation methods in gravitational-wave transient analysis, the median average. We demonstrate that simulated Gaussian noise follows the predicted distributions. We then examine real gravitational-wave data at and around the time of GW151012, a relatively low-significance binary black hole merger event. We show that the data are well described by stationary-Gaussian noise and explore the impact of different noise power spectral density estimation methods on the astrophysical inferences we draw about GW151012.

arXiv:2006.05292v2 [astro-ph.IM] 16 Jun 2020

I. INTRODUCTION

The astrophysical parameters of compact binaries are inferred from gravitational-wave data using Bayesian inference. A crucial first step for Bayesian inference is to choose the appropriate likelihood for our data. In gravitational-wave transient data analysis this, typically, hinges on assumptions that the noise is Gaussian and stationary over the period being analysed [1]. If these conditions are met, and if the noise power spectral density (PSD) were known exactly, the appropriate likelihood would be the Whittle likelihood [2]

$$\mathcal{L}(\tilde{d}|\theta, P) = \frac{2}{\pi T P} \exp\left(-\frac{2|\tilde{d} - \tilde{\mu}(\theta)|^2}{T P}\right). \qquad (1)$$

Here P is the PSD, $\tilde{d}$ is the frequency domain interferometer data, T is the duration of the data being analyzed, and $\tilde{\mu}$ is our model for the expected signal.

However, in practice, we do not have access to the true power spectral density of gravitational-wave detectors and so we have to rely on an empirical estimate. There are two commonly used methods to compute these estimates. The simplest method is to average over the power in neighbouring stretches of data to generate an "off-source" estimate. This method assumes that the PSD does not vary over the duration being averaged and that there are no non-Gaussian features in the data. The other commonly used method is to simultaneously fit the signal and power spectral density to obtain an "on-source" estimate, e.g., [3]. While this method does not involve analyzing as much data, and hence if less effected by non-stationarity and non-Gaussianity, it is far more

computationally expensive. In this work, we are going to focus on the former.

To generate an off-source PSD we typically either compute the mean or median average of neighbouring segments. Taking the mean of neighbouring segments is a commonly used method (sometimes referred to as the Welch or Blackwell method [4]) in gravitational-wave data analysis and many other signal processing applications. However, it is not widely used in gravitational-wave transient data analysis due to it's sensitivity to non-Gaussian transients, "glitches," in the detector noise. To mitigate the effect of these glitches a median average is instead used to compute the PSD as the median is more robust to the presence of large outliers. However, there may be effective methods to either remove or exclude these non-Gaussian features [5–11].

The other assumption underlying Equation 1 is that the PSD does not change over time; in other words, the data are stationary. In practice, the PSD of real interferometers varies over the time scale of minutes, and so care must be taken when estimating the PSD using longer stretches of time [1, 3, 12]. Methods for mitigating this non-stationarity have also been considered previously [11]. In this paper, we ignore these possible effects and assume the data we look at are both Gaussian and stationary.

In Chatziioannou *et al.*, the authors compare Advanced LIGO data whitened with a median estimated off-source PSD with on-source PSD estimation methods. They show that the data whitened using the median PSD does not follow a unit normal distribution. They argue that this difference is due to non-Gaussianity and non-stationarity in the data. However, data whitened using an off-source are known to follow a non-Normal distribution *even for Gaussian noise*. For a mean average, the whitened Gaussian-noise data follow a Student's t-distribution, and the correct likelihood to use is the

---

[a] ctalbot@caltech.edu

Student-Rayleigh distribution [14–17]. In this work, we demonstrate that data whitened with a median PSD estimate follows a different distribution, and we show how to marginalise over the uncertainty in this estimated PSD to obtain the correct likelihood for stationary, Gaussian noise.

On-source PSD estimation using the `BayesLine` algorithm [3] marginalizes over a prior which models the PSD as a combination of a slowly varying spline and Lorentzians to fit sharp spectral features. This algorithm is generally combined with the `BayesWave` algorithm [18] to fit astrophysical and terrestrial transients simultaneously with the PSD. However, this does not allow direct inference of physical parameters describing compact binary coalescences, e.g., the masses and spins of merging black holes.

A common approach in compact binary data analysis is to take a median average of the posterior distribution for the PSD obtained using `BayesLine` rather than an off-source averaged PSD, e.g. [19]. Recently, Biscoveanu et al. introduced a method to marginalize over the uncertainty in these on-source PSDs estimates. However, this is done at significant computational cost requiring $\sim 200\times$ the computational resources as a standard analysis. Additionally, under the formalism presented there, it is not possible to compute the Bayesian evidences necessary to perform model comparison.

The remainder of this paper is structured as follows. In Section II, we derive the appropriate distributions for the likelihood and whitened data after marginalising over the uncertainty in a median (and/or mean) PSD estimate. We provide a brief introduction to Bayesian inference in the context of gravitational-wave astronomy in Section III. We then demonstrate the efficacy of our formalism by applying it to simulated Gaussian data in Section IV. Following this, in Section V we consider a case study using real Advanced LIGO data. We analyze the marginal gravitational-wave candidate GW151012 with both mean and median PSD estimates to understand the effect of marginalizing over the statistical uncertainty and of using the different estimation techniques. This event is convenient for our present purposes since the effects we seek to study are most prominent for marginal signals like GW151012. Some closing comments are then provided in Section VI.

II. FORMALISM

A. Gaussian noise

For stationary Gaussian noise $n(t)$, if we do not manipulate the data in any way before performing a Fourier transform, the noise covariance can be written in the frequency domain as

$$C(f, f') = \langle \tilde{n}(f)\tilde{n}^*(f') \rangle = P(f)\delta(f - f'). \quad (2)$$

The angle braces denote an ensemble average over realizations. In practice, we work with discrete Fourier transforms and noise covariance matrices

$$C_{ij} = \langle \tilde{n}_i \tilde{n}_j^* \rangle = \frac{T}{4}P_i\delta_{ij} = \frac{T}{4}A_i\delta_{ij}A_j \quad (3)$$

Here T is the duration over which the discrete Fourier transforms is performed, i, j index frequency bins, and $A_i = P_i^{1/2}$ is the noise amplitude spectral density (ASD).

For real data, a number of manipulations are performed before the data are Fourier transformed, which makes things more complicated. The data are band-passed and windowed in the time domain to prevent aliasing and spectral leakage [1]. As long as the frequency limits of the band-pass filters do not overlap with the frequency range of interest, the band-passing can be ignored. However, the window applied to the data must be considered. Since the window is multiplicative in the time domain, there is a corresponding convolution in the frequency domain,

$$C_{ij}^w = \left\langle (\tilde{n} * \tilde{W})_i^* (\tilde{n} * \tilde{W})_j \right\rangle = \frac{T}{4}A_i\mathcal{T}_{ij}A_j, \quad (4)$$

$$\mathcal{T}_{ij} = \begin{cases} \tilde{W}_{i-j} & i \geq j \\ \tilde{W}_{j-i}^* & i < j \end{cases}. \quad (5)$$

Where $\mathcal{T}$ is a Hermitian Toeplitz matrix and there is no implied summation over i or j. For a rectangular window $\mathcal{T}_{ij} = \delta_{ij}$ and the covariance matrix is diagonal. For generic windows, there is a regular, predictable, off-diagonal power. In reality, the effect of this is much smaller than the effects considered here, and inverting the covariance matrix is a significant computational challenge. We leave a detailed analysis of the effect of non-rectangular windows on parameter estimation and model selection to a future study.

The real and imaginary components of the frequency-domain noise follow a normal distribution with variance matrix P,

$$p(\tilde{n}_i | P_i) = \sqrt{\frac{2}{\pi T P_i}} \exp\left(-\frac{2\tilde{n}_i^2}{T P_i}\right) \quad (6)$$

This is not the likelihood which we use when analyzing gravitational-wave transients as we need to simultaneously consider the real and imaginary components of the noise. The likelihood is given by

$$\mathcal{L}(\tilde{n}_i, \tilde{n}_j | C_{ij}) = \frac{2}{\pi T \det(C_{ij})} \exp\left(-\frac{1}{2}\tilde{n}_i C_{ij}^{-1} \tilde{n}_j^*\right). \quad (7)$$

Since we assume the covariance matrix is diagonal this is often written in the simplified form known as the Whittle likelihood,

$$\mathcal{L}(\tilde{n}_i | C_i) = \frac{2}{\pi T P_i} \exp\left(-\frac{2|\tilde{n}_i|^2}{T P_i}\right). \quad (8)$$

We note that this likelihood is normalized over the complex plane. It is convenient to reduce to one dimension

for visualisation purposes, so we note that the power of the noise $\mathcal{P}_i = |\tilde{n}_i|^2$ follows an exponential distribution,

$$p(\mathcal{P}_i|P_i) = \frac{2}{TP_i} \exp\left(-\frac{2\mathcal{P}_i}{TP_i}\right). \tag{9}$$

All of the expressions above assume that there are no non-Gaussian signals in the data. In order to include signals we simply make the substitution $n = d - \mu$ where d is the data and μ is the signal.

Time-domain windows affect the noise and signal components differently. We assume that the window is always applied such that the window does not cause any loss of signal power in the observing frequency band. In addition to the correlation between different frequency bins induced by the window, there is a net power loss in the Gaussian noise given by the mean square value of the window function. Care must be taken to consistently correct for this power loss to avoid biasing our inference, e.g., [21].

Now that we have established which distributions we want to use when the PSD is known, we can address the distributions we want to use when the PSD is uncertain.

B. Median PSD estimate

The generic expression for the likelihood marginalised over uncertainty in an estimated PSD, $\hat{P}$, is

$$\mathcal{L}_P(\tilde{d}|\theta, \hat{P}) = \int_0^\infty dP\, \mathcal{L}(\tilde{d}|\theta, P)\pi(P|\hat{P}). \tag{10}$$

Where $\mathcal{L}(\tilde{d}|\theta, P)$ is the likelihood of obtaining the data given model parameters θ and the true PSD P, as defined in Eq. 1, and $\pi(P|\hat{P})$ is our prior on the true PSD given the estimated PSD. Similarly, using Equation 6, we can write down an expression for the expected distribution of whitened strain residuals, $\tilde{\nu} = \tilde{n}/\hat{P}^{1/2}$,

$$p_P\left(\tilde{\nu}|\hat{P}\right) = \int_0^\infty dP\, p(\tilde{n}|P)\pi(P|\hat{P}). \tag{11}$$

Here $\tilde{n}$ is the frequency-domain data after removing any signals present.

First we need to define the estimated PSD

$$\hat{P} = \frac{\text{median}(P_\ell)}{\alpha}. \tag{12}$$

Where

$$\alpha = \sum_{\ell=1}^N \frac{(-1)^\ell}{\ell} \tag{13}$$

is a factor to account for the median being a biased estimator of the mean (see, e.g., Appendix B of [22]), and ℓ indexes the segments being averaged over. For simplicity, we assume that we are computing the median of an odd number, N, of non-overlapping stretches ensuring $\alpha > 0$.

It is convenient to work with a regularised version of the PSD,

$$Q = 2\hat{P}/P. \tag{14}$$

Since the data are assumed to follow a zero-mean Gaussian distribution with variance P, the Q_i are drawn from a χ^2 distribution of order 2,

$$p(Q) = \chi_2^2(Q) = \frac{1}{2}\exp\left(-\frac{Q}{2}\right). \tag{15}$$

Additionally, we define the usual cumulative distribution function, Φ, and survival function, S, for this quantity,

$$\Phi(Q) = \int_0^Q dQ' p(Q') = 1 - \exp\left(-\frac{Q}{2}\right) \tag{16}$$

$$S(Q) = \int_Q^\infty dQ' p(Q') = \exp\left(-\frac{Q}{2}\right). \tag{17}$$

The probability of the median of an odd number of segments follows the median order statistic. This is the probability of the getting the median value from the distribution multiplied by the probability of having $m = (N-1)/2$ measurements less than $\hat{Q}$ and m measurements larger than $\hat{Q}$. Symbolically, this is

$$\pi(Q|\hat{P}) = \frac{p(Q)}{2\hat{P}}\frac{\Phi(Q)^m S(Q)^m}{B(m+1, m+1)} \tag{18}$$

$$= \frac{\left(1 - e^{-\frac{1}{2}Q}\right)^m e^{-\frac{(m+1)}{2}Q}}{4\hat{P}B(m+1, m+1)} \tag{19}$$

$$= \sum_{k=0}^m \binom{m}{k}\frac{(-1)^k e^{-\frac{(m+k+1)}{2}Q}}{4\hat{P}B(m+1, m+1)}. \tag{20}$$

Where B is the Beta function and in the last line we perform a binomial expansion. The final piece we need is to relate our prior on Q to our prior on P,

$$\pi(P|\hat{P})dP = \pi(Q|\hat{P})dQ. \tag{21}$$

Substituting this expression into Equation 10, the PSD-marginalized likelihood is

$$\mathcal{L}_P = \int_0^\infty dQ \frac{Q}{8\pi\hat{P}} \frac{\left(1 - e^{-\frac{Q}{2}}\right)^m e^{-\frac{Q}{2}\left(\frac{1}{2}|\tilde{\nu}|^2 + m + 1\right)}}{B(m+1, m+1)} \tag{22}$$

$$= \sum_{k=0}^m \binom{m}{k}\frac{(-1)^k}{2\pi\hat{P}}\frac{\left(m + k + 1 + \frac{|\tilde{\nu}|^2}{2}\right)^{-2}}{B(m+1, m+1)}, \tag{23}$$

and using Equation 11, the distribution of whitened residuals is

$$p(\tilde{n}) = \int_0^\infty dQ \frac{1}{4}\sqrt{\frac{Q}{2\pi\hat{P}}}\frac{\left(1 - e^{-\frac{Q}{2}}\right)^m e^{-\frac{Q}{2}\left(\frac{1}{2}|\tilde{\nu}|^2 + m + 1\right)}}{B(m+1, m+1)} \tag{24}$$

$$= \sum_{k=0}^m \binom{m}{k}\frac{(-1)^k}{\sqrt{2\pi\hat{P}}}\frac{\left(m + k + 1 + \frac{|\tilde{\nu}|^2}{2}\right)^{-3/2}}{B(m+1, m+1)}. \tag{25}$$

While the final expressions Equations 23 and 25 are exact closed form solutions, they are numerically unstable and cannot be safely computed for $m \gtrsim 15$. We therefore simply construct an interpolant over numerically computed values of the integrals in Equations 22 and 24, which can be rapidly evaluated at run time.

C. Mean PSD estimate

The appropriate distribution to use for a mean averaged PSD has been discussed and independently derived multiple times in the literature, e.g., [15, 17]; in this work, we just quote the relevant results. For a mean estimate:

$$\hat{P} = \frac{1}{N} \sum_{i=1}^{N} P_i. \qquad (26)$$

$$\pi(P|\hat{P})dP = \pi(Q|\hat{P})dQ = \chi^2_{2N}(Q)dQ. \qquad (27)$$

$$\mathcal{L}_P = \frac{1}{2\pi\hat{P}} \left(1 + \frac{|\tilde{\nu}|^2}{2N} \right)^{-(N+1)}, \qquad (28)$$

$$p(\tilde{\nu}) = \frac{\Gamma(N+1/2)}{\sqrt{2\pi N \hat{P}}\Gamma(N)} \left(1 + \frac{|\tilde{\nu}|^2}{2N} \right)^{-(N+1/2)}. \qquad (29)$$

Equation 28 is the Student-Rayleigh distribution and Equation 29 is the Student's t-distribution with $2N$ degrees of freedom, two degrees for each segment being averaged over, coming from the real and imaginary components of the frequency domain strain.

D. Limiting cases

When $N = 1$ the mean and median are the same and so Equation 25 should reduce to a Student's t-distribution with two degrees of freedom. As expected, we find

$$p(\tilde{n}) = \frac{1}{\sqrt{2\pi\hat{P}}} \left(1 + \frac{|\tilde{\nu}|^2}{2} \right)^{-3/2}. \qquad (30)$$

Analogously, the PSD uncertainty marginalised likelihoods also match and are both

$$\mathcal{L}_P = \frac{1}{2\pi\hat{P}} \left(1 + \frac{|\tilde{\nu}|^2}{2} \right)^{-2}. \qquad (31)$$

The other important limiting case is when $N \to \infty$. It is a well-known result that the Student's t-distribution converges to a Gaussian in this limit; this follows from Taylor-expanding the distribution. Performing a similar expansion, it is possible to demonstrate that the Student-Rayleigh distribution converges to the Whittle likelihood. We numerically confirm that (25) and (23) also converge to a Gaussian distribution and Whittle likelihood respectively.

III. BAYESIAN INFERENCE FOR GRAVITATIONAL-WAVE TRANSIENTS

In the previous section, we derived likelihood functions, marginalized over the statistical uncertainty in an estimate of the PSD. These likelihood functions $\mathcal{L}(\tilde{d}|\theta, \hat{P}, \mathcal{M})$ are the probability of obtaining data $\tilde{d}$ given a signal model $\mathcal{M}$ described by parameters θ and a PSD estimate $\hat{P}$. However, we are generally interested in measuring the source-model parameters and performing model comparison. Using Bayes' theorem we get

$$p(\theta|\tilde{d}, \hat{P}, \mathcal{M}) = \frac{\mathcal{L}(\tilde{d}|\theta, \hat{P}, \mathcal{M})\pi(\theta, \hat{P}, \mathcal{M})}{\mathcal{Z}(\tilde{d}|\hat{P}, \mathcal{M})}. \qquad (32)$$

The term on the LHS, $p(\theta|\tilde{d}, \hat{P}, \mathcal{M})$, is the posterior probability distribution, the probability of the parameters describing the model given the data. The term $\pi(\theta, \hat{P}, \mathcal{M})$ is our prior distribution which is based on our expectation before analysing the data. The term $\mathcal{Z}(\tilde{d}|\hat{P}, \mathcal{M})$ is the evidence for the data given the model $\mathcal{M}$.

The evidence is used for model comparison by computing Bayes factors for two models

$$BF_0^1 = \frac{\mathcal{Z}(\tilde{d}|\hat{P}, \mathcal{M}_1)}{\mathcal{Z}(\tilde{d}|\hat{P}, \mathcal{M}_0)}. \qquad (33)$$

While the Bayes factor is often used for model selection, strictly speaking, we should compare the probability of the model given the data, rather than the probability of the data given the model. This is given by the odds

$$\mathcal{O}_0^1 = \frac{\mathcal{Z}(\tilde{d}|\hat{P}, \mathcal{M}_1)}{\mathcal{Z}(\tilde{d}|\hat{P}, \mathcal{M}_0)} \frac{\pi(\mathcal{M}_1|\hat{P})}{\pi(\mathcal{M}_0|\hat{P})}, \qquad (34)$$

which is the Bayes factor comparing the two models multiplied by the prior odds. Throughout this work, we will assume all models have equal prior odds and so the odds reduces to the Bayes factor.

Finally, we define the coherent vs incoherent Bayes factor [23], BCI, as a measure of the relative probability that the data contain a coherent signal or incoherent signals in different detectors,

$$\text{BCI} = \frac{\mathcal{Z}(\{\tilde{d}_k\}|\{\hat{P}_k\}, \mathcal{M}))}{\Pi_k \mathcal{Z}(\tilde{d}_k|\hat{P}_k, \mathcal{M}))}. \qquad (35)$$

Here, the k index multiple independent interferometers, e.g., LIGO Hanford and LIGO Livingston. As in [24] we assume that any incoherent signals are described by the same model as the coherent signals, however, this is not necessarily the case [25]. We note that the BCI is not used as the final discriminator between the coherent and incoherent models as it is missing a prior for the relative rates of coherent and incoherent signals. In [24, 25] the priors on rate are empirically calibrated delta functions, however, in [26] the authors fit the rates of coherent and incoherent signals.

In general relativity, non-eccentric binary black hole coalescences are fully described by fifteen parameters. Eight parameters which describe the "intrinsic" properties of the binary (two masses and two three-dimensional angular momentum vectors), and seven "extrinsic" parameters to specify the position, orientation, and coalescence time of the binary relative to Earth. This parameter space is typically explored using stochastic samplers using either Markov-chain Monte Carlo [27] or nested sampling [28].

In order to improve the convergence of the sampling and accelerate our inference, it is possible to use a modification of the Whittle likelihood which is marginalized over the coalescence time, orbital phase, and distance of the source [23, 29]. It is not possible to perform these marginalizations as easily while also marginalizing over uncertainty in the PSD. Therefore, in this work, we perform our inference in two stages following [30].

1. First, we analyze the data using the Whittle likelihood marginalized over coalescence time, binary orbital phase, and distance to obtain samples from the posterior distribution and an estimate of the signal evidence and Gaussian noise evidence. Posterior distributions for these marginalized parameters are then recomputed in post-processing. We use `dynesty` [31], an implementation of the nested sampling algorithm, as implemented in `Bilby` [32] to sample the space.

2. After this, we importance resample the posterior obtained in the previous step by the ratio of the PSD-marginalised likelihood to the Whittle likelihood to obtain posterior samples and an evidence which include the marginalisation over the statistical uncertainty.

We note that the importance sampling in step 2 only works when resampling to a distribution which is similar to the original posterior distribution. We quantify this by evaluating the efficiency of the resampling and the number of effective samples from the PSD marginalised posterior. Since the marginalized likelihoods converge to the non-marginalized likelihood when averaging many segments, we expect the resampling to be efficient. This method also generically gives a much smaller uncertainty on the Bayes factor comparing the two models than would be obtained by performing two independent sampling runs [33]. A similar method has previously been employed for cosmological inference in [34] to marginalize over uncertainty in an estimated covariance matrix.

IV.   DEMONSTRATION WITH GAUSSIAN NOISE

To demonstrate the accuracy of the methods described in Section II we analyze simulated Gaussian noise colored by the Advanced LIGO design sensitivity PSD [35].
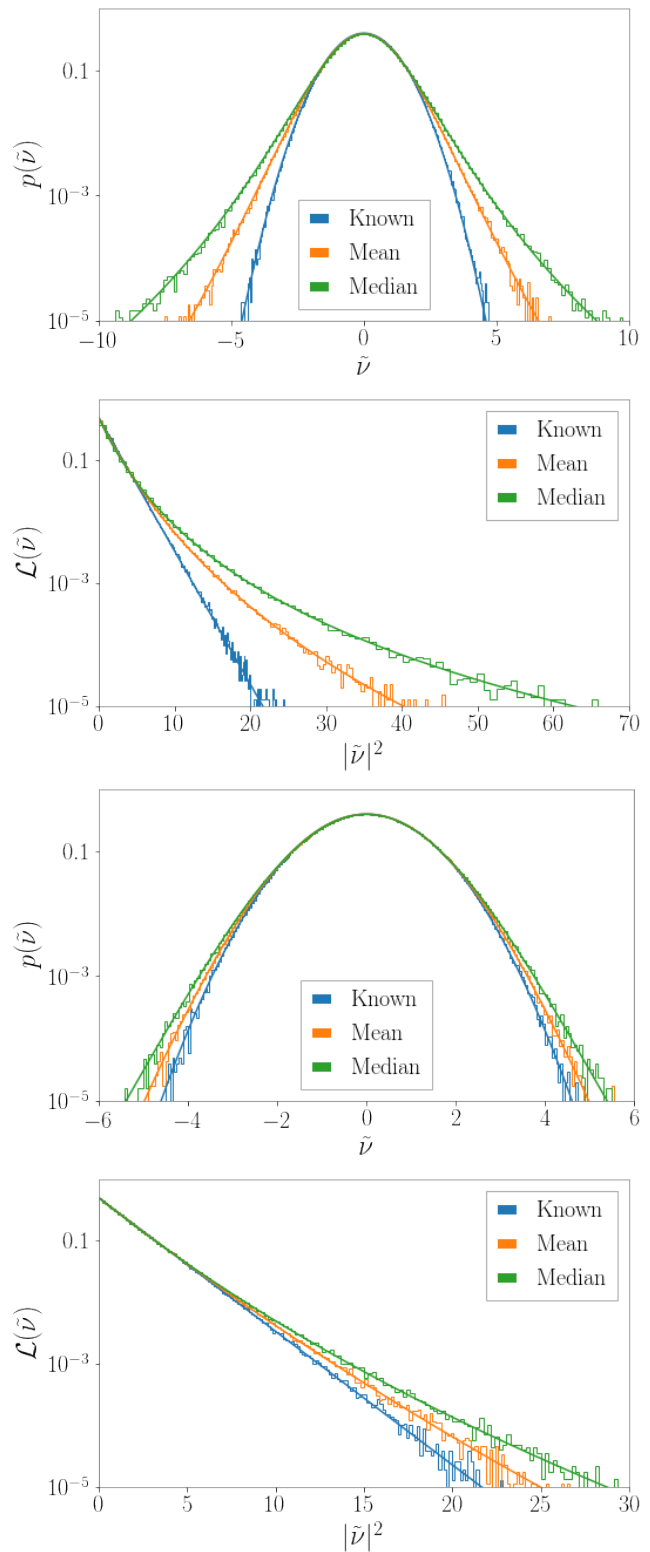

FIG. 1: Comparison of the expected and empirical distribution of whitened frequency-domain strain (first and third panels) and whitened noise power (second and fourth panels) when using three different PSDs for whitening and simulated Gaussian noise colored to the known PSD. The three PSDs used are the true "known" PSD (blue), a mean estimate (orange), and a median estimate (green). The number of averages used to generate the mean and median estimates are 7 (top pair), and 31 (bottom pair). In all cases the data follow the predicted distributions.
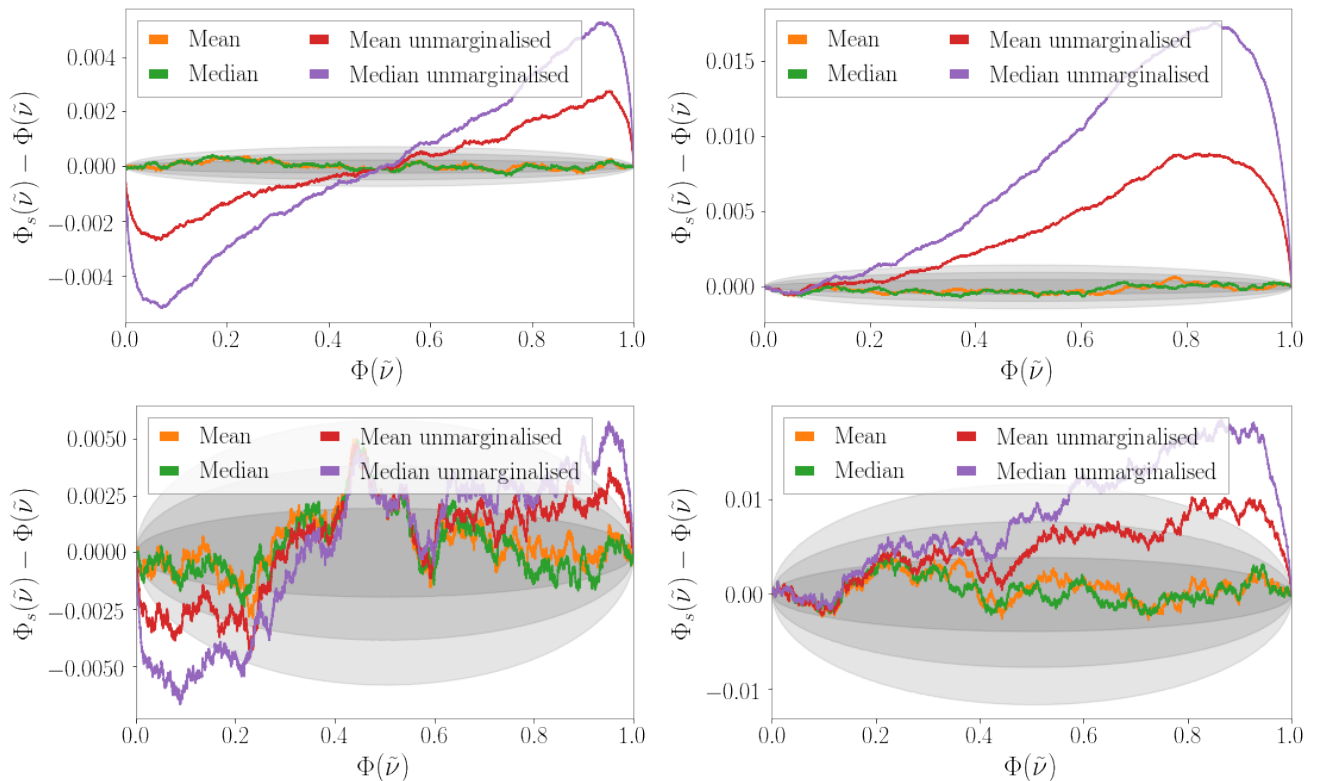
FIG. 2: Difference between the empirical and expected cumulative distributions of whitened frequency-domain strain (left) and whitened noise power (right) when using two different PSDs for whitening and simulated Gaussian noise colored to a known PSD. In the top panels we consider 512 seconds of simulated data and in the bottom panels we consider 8 seconds of simulated data. The two PSDs used are a mean estimate (orange/red), and a median estimate (green/purple). The gray shaded regions show the expected 1σ, 2σ, and 3σ uncertainties. We average over 31 realisations to generate the PSDs. There are significant deviations when not marginalizing over the uncertainty in the PSD once enough data are considered.

Following [13], we perform three tests on data whitened using median and mean PSD estimates for verification. As an extension to the analysis presented in [13], we consider the whitened power, $|\tilde{\nu}|^2$ in addition to the real and imaginary components of the whitened strain $\tilde{\nu}$. For all estimated PSDs, we average non-overlapping segments with the same duration as the analysis segment.

First, we perform a visual test of the whitened data. In Figure 1, we show the distribution of the real and imaginary components of the whitened strain (first and third panels) and whitened power (second and fourth panels) along with the theoretical expectations. In the top (bottom) pair of panels, we average over 7 (31) independent noise realisations. We see that the data whitened using the off-source estimates follow the expected distributions in each case.

In Figure 2 we show the difference between the empirical and expected cumulative distribution functions, $\Phi_s - \Phi$, plotted against the expected cumulative distribution function for the same data as in Figure 1. In orange and red we compare the data whitened with the mean PSD estimate with the expected distributions with and

without marginalizing over the uncertainty in the PSD respectively. In green and purple we compare the data whitened with the median PSD estimate with the expected distributions with and without marginalizing over the uncertainty in the PSD respectively. The grey regions indicate the expected 1σ, 2σ, and 3σ fluctuations. For both PSD estimation methods, we see that the data agree better with the distributions which marginalize over the uncertainty in the PSD.

When comparing the marginalized distributions to the non-marginalized distributions we see two clear deviations from the expected behaviour. The whitened strain uncertainty-marginalized distribution has wider, symmetric, tails than a normal distribution leading to the negative $\Phi_s - \Phi$ for small $\tilde{\nu}$ and positive $\Phi_s - \Phi$ for large $\tilde{\nu}$. The distribution of the whitened power, however, only has a wide tail out to large $|\tilde{\nu}|$, leading to the positive $\Phi_s - \Phi$ for large $\tilde{\nu}$.

To quantify the similarity of the data to the expected distributions we compute the Anderson-Darling statistic

$$A^2 = N \int_{-\infty}^{\infty} d\Phi \, \frac{(\Phi_s - \Phi)^2}{\Phi(1 - \Phi)}. \tag{36}$$
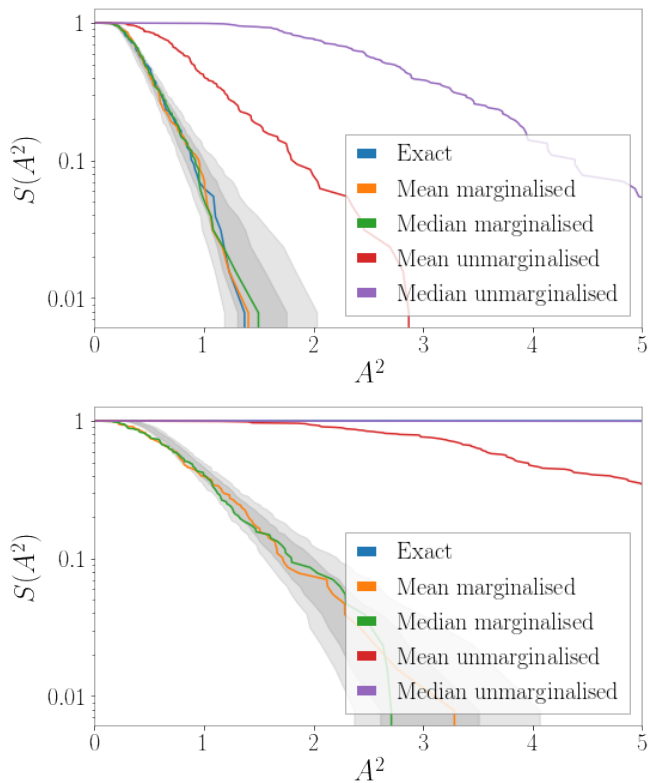
FIG. 3: The survival function of the Anderson-Darling statistic comparing Gaussian noise whitened with three PSDs with the distributions which do/do not marginalize over the statistical uncertainty in the PSD estimate. In the top panel, we analyze the real and imaginary components of the whitened strain. In the bottom panel, we analyze whitened power. The blue curve is generated using data whitened by the exact known PSD. The orange/green curves are generated by comparing data whitened by the mean/median PSD estimates with the distributions which marginalize over the statistical uncertainty. The red/purple curves are generated by comparing data whitened by the mean/median PSD estimates with the distributions which do not marginalize over the statistical uncertainty. The grey shaded regions show the expected 1σ and 2σ uncertainties. We average over 31 realisations to generate the PSDs. The Anderson-Darling statistic does not follow the expected distribution when not marginalizing over the uncertainty in the PSD.

Here N is the number of samples, in this case, the number of frequency bins. The numerator is the square of the quantity on the vertical axis of Figure 2 and the integral is over the horizontal axis.

In Figure 3 we show the survival function of the distribution of the Anderson-Darling statistic for four cases: for both the mean and median PSD estimation methods we compare the distribution of the whitened strain to a unit normal distribution and the expected distribution as

described in the previous section for the whitened strain (left) and whitened power (right). We also show the expected distribution if the two distributions are the same.

We note that the gradient of the expected distribution of the Anderson-Darling statistic is steeper for the whitened strain than for the whitened power. When applying a window the data before performing the discrete Fourier transforms, the real and imaginary components of the frequency domain strain are no longer independent, reducing the appropriate value of N by a factor of two (see, Appendix A of [21]). We, therefore, avoid the case identified in [13] where the distribution of ν appeared to match the correct distribution better than possible.

V. A CASE STUDY - GW151012

To examine the effect of non-Gaussianity and non-stationarity on the noise properties of real gravitational-wave detectors, we analyze the data in the two Advanced LIGO interferometers [36] at and around the time of GW151012 [37], the lowest significance binary black hole merger included in the first gravitational-wave transient catalog [19]. We analyze 256 s of data ending 2 s after the merger time. We subdivide the data into 32 8 s chunks, the first 31 chunks are used to compute the PSD and the final 8 s are the on-source data.

We apply a Tukey window with a roll off of 0.2 s to each of the chunks to suppress spectral leakage. We then fast-Fourier transform the windowed time-domain strain before averaging the PSD chunks. We do not apply the conventional window amplitude correction factor to either the PSD or the data. After applying the fast-Fourier transform, we remove all data below 20 Hz and above 1024 Hz.

The resulting PSDs and the power in the on-source data are shown in Figures 4 and 5. Figure 4 shows data from the LIGO Hanford interferometer and Figure 5 data from the LIGO Livingston interferometer. The orange curves show the mean estimated PSDs and the green show the median estimated PSDs. All the PSD estimates are at the centre of the scatter in the on-source data, as expected. We note that the width of the scatter on the mean PSDs is slightly smaller than for the median due to the slower convergence of the median estimate.

A. Data quality tests

The main reason for using a median estimate over a mean estimate is to mitigate the effect of large non-Gaussian transients. However, the formalism derived above is invalid if there is a large outlier in the data being averaged over. Therefore, we try to identify if any of the segments are clear outliers. We compute the power per segment divided by the mean power in all the other segments. This is essentially testing how well the data in each of the segments is whitened by the data in the other
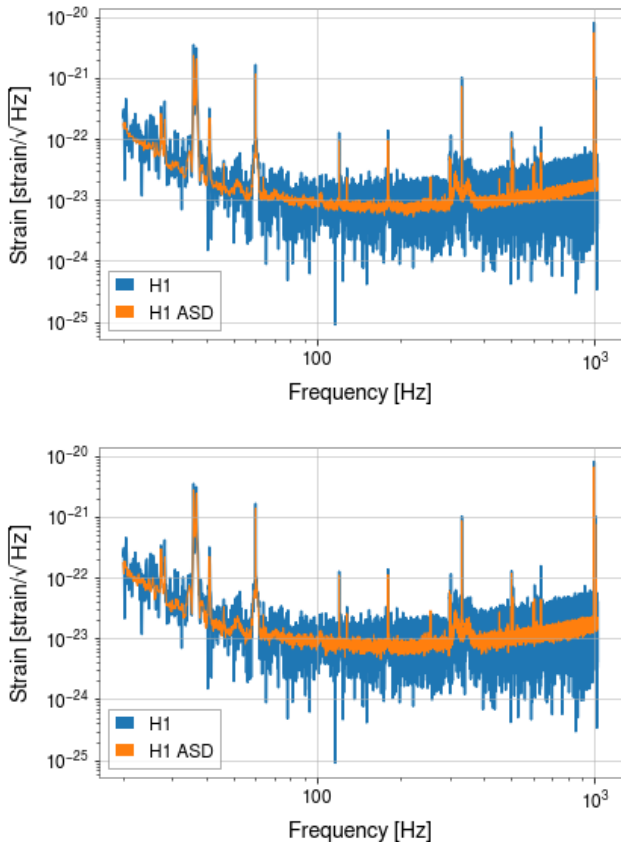
FIG. 4: Frequency domain strain power (blue) and mean (top) and median (bottom) estimated power spectral densities (orange) for LIGO Hanford at the time of GW151012.
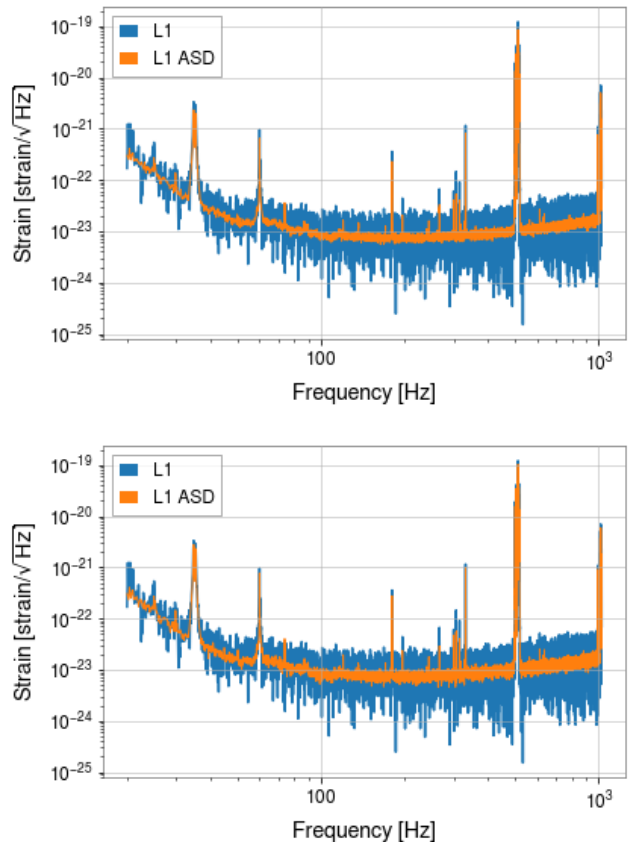
FIG. 5: Frequency domain strain power (blue) and mean (top) and median (bottom) estimated power spectral densities (orange) for LIGO Livingston at the time of GW151012.

segments. We apply an empirically tuned threshold of 1.5 for the mean whitened power per segment. Any segment with a mean power above this value we discard and repeat the test. We identify that one segment of the Hanford data which fails this test with a mean whitened power of 2.66. Visual inspection reveals that this segment has a larger amplitude than all the others below ∼ 100Hz. No significant outliers are present in the selected Livingston data.

Additional tests of the quality are possible and performed routinely during gravitational-wave data analysis. For example, researchers often remove specific frequency bins if the noise at that frequency is known to be non-Gaussian, e.g., around the frequency (and higher harmonics) of mains electricity [1]. A possible extension would be to use the normalised average power used above to track non-stationarity in the data, a similar method is used in in [11]. Implementing further data quality cuts and vetoes will improve the quality of our off-source PSD estimates and is an interesting avenue for further study.

B. Data whitening

We repeat the tests performed in Section IV on the data. In Figure 6, we show the deviations from the expected cumulative distribution functions for the data from the Hanford (top) and Livingston (bottom) interferometers. On the left, we show the real and imaginary components of the whitened strain and on the right the whitened power. In orange we show the difference between the empirical mean-estimated PSD whitened data and expected mean-marginalised distributions (29, 28), in green the difference between the empirical median-estimated PSD whitened data and expected median-marginalised distributions (25, 23). In red (purple) we compare the data whitened using the mean- (median-) estimated PSDs with the distributions which do not account for the uncertainty. In gray we show the 1σ, 2σ, and 3σ expected deviations.

In Table I, we quote the corresponding values of the Anderson-Darling statistic for each of these lines. We see that the largest deviations are observed when using data whitened with a median estimated PSD and compared
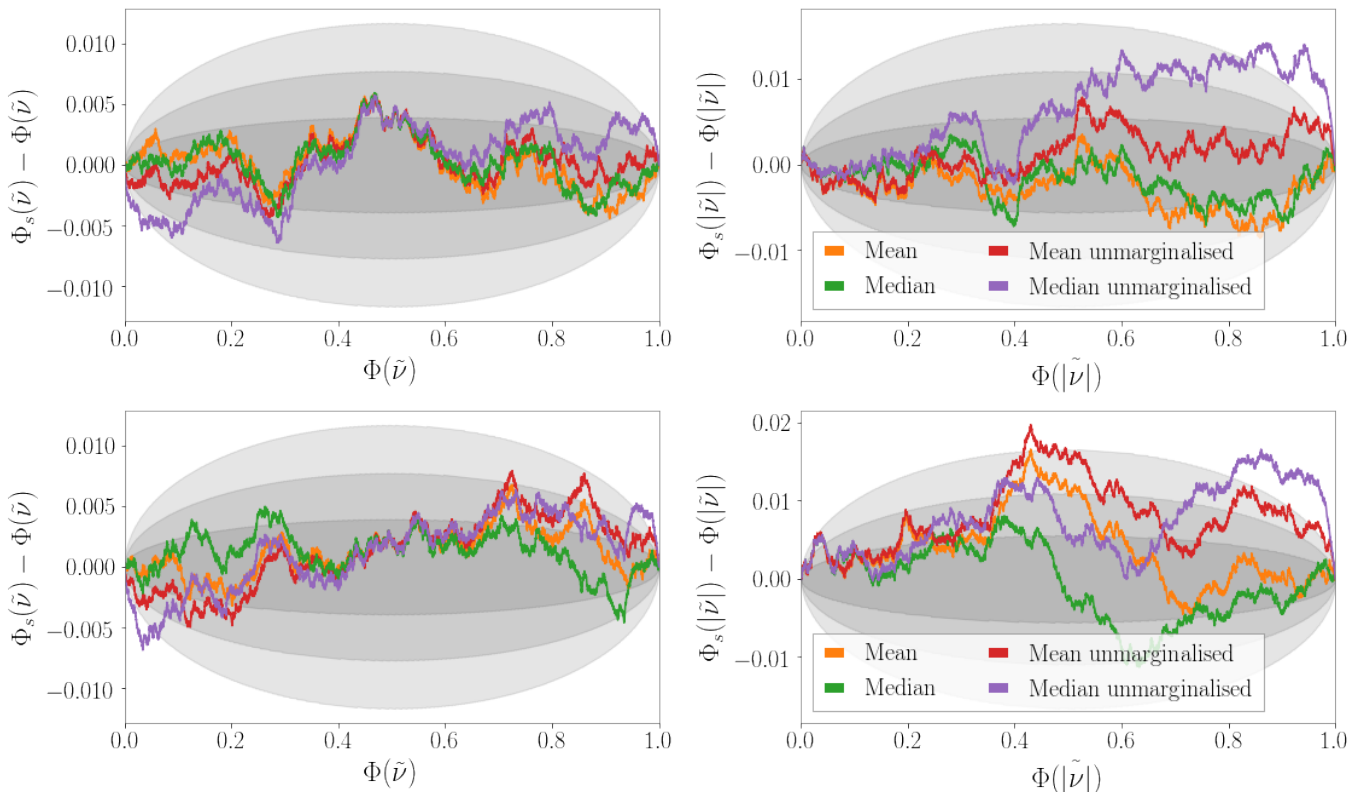
FIG. 6: The difference between the empirical and analytic cumulative distribution functions of whitened strain $\tilde{\nu}$ and whitened power $|\tilde{\nu}|^2$ in the LIGO Hanford (top) and Livingston (bottom) interferometers with two different power spectral density estimation methods at the time of GW151012. The gray shaded regions show the expected 1σ, 2σ, and 3σ fluctuations. For the orange and green curves, the data are compared with the distributions which marginalise over uncertainty in the power spectral density estimate. For the red and purple curves, the data are compared with the distributions which do not marginalize over uncertainty in the power spectral density estimate. We note that the latter pair of curves deviate from the 3σ region, while the former does not. The data are well described by a stationary Gaussian process when marginalising over uncertainty in the power spectral density.

	Mean vs marginalised	Median vs marginalised	Mean vs not marginalised	Median vs not marginalised
Livingston strain	0.53	0.46	1.62	2.66
Hanford strain	0.56	0.42	0.48	2.21
Livingston power	1.48	1.03	4.22	6.71
Hanford power	0.79	0.48	0.84	5.55

TABLE I: Values of the Anderson-Darling statistic for the whitened strain for the mean and median marginalised likelihood at the time of GW151012. Larger values of the Anderson-Darling statistic indicate comparatively worse agreement. The marginalized distributions match the data better.

to the non-marginalised distributions. We also see that, with the exception of the strain components in Hanford, the Anderson-Darling statistic is always smaller when using the appropriate marginalised distributions.

C. Impact on inference

We analyze the data using Bayesian inference as described in Section III twice, once each with the mean-

PSD	No marg	Marg
Mean	9.98	9.89
Median	10.39	10.16

TABLE II: Values of the (natural) log coherent vs incoherent Bayes factor (ln BCI) for different PSD estimates and likelihoods. In both cases the difference between the ln BCI with and without marginalising over the PSD uncertainty is within typical uncertainties due to finite sampling.
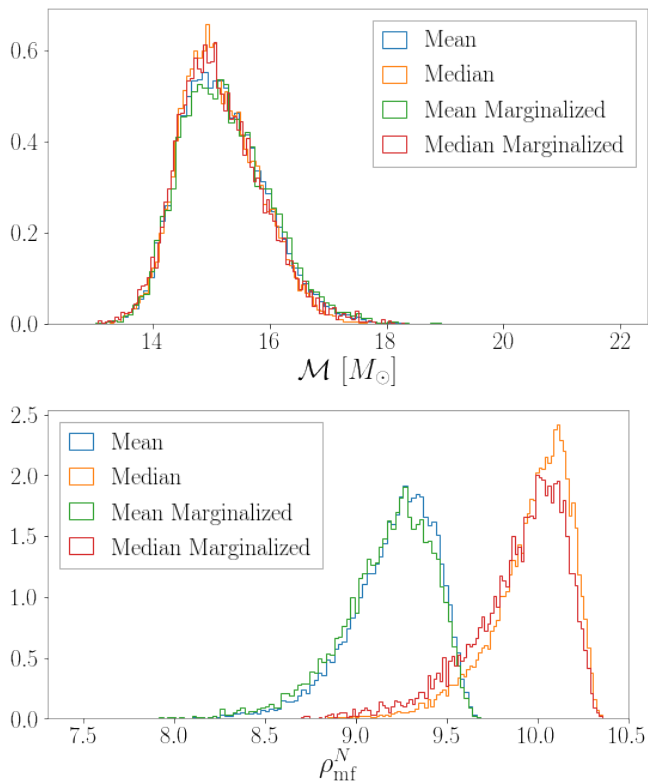
FIG. 7: The posterior distribution for chirp mass (top) and matched filter SNR (bottom) for GW151012 for four different models. In blue and green we use the mean estimated PSD while in yellow and red we use the median estimate. In blue and yellow we neglect the uncertainty in the PSD estimate and in green and red we marginalise over the appropriate statistical uncertainty. We note that in both cases, marginalising over the uncertainty increases the width of the chirp mass posterior and decreases the average SNR. The mean estimated PSD gives a wider chirp mass posterior than the median PSD estimate and we see a corresponding decrease in the recovered matched filter SNR.
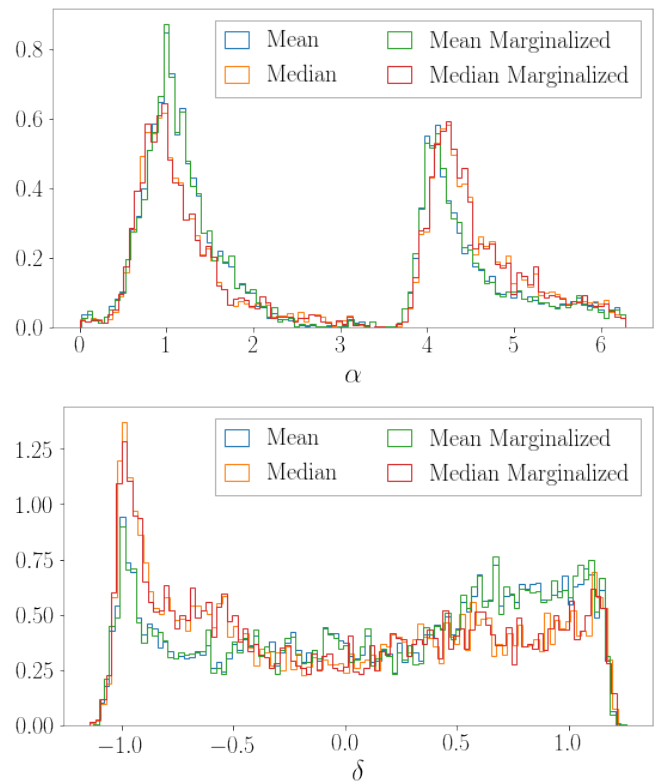

FIG. 8: The posterior distribution for right ascension (top) and declination (bottom) for GW151012 for four different models. In blue and green we use the mean estimated PSD while in yellow and red we use the median estimate. In blue and yellow we neglect the uncertainty in the PSD estimate and in green and red we marginalise over the appropriate statistical uncertainty. In this case, marginalising over the uncertainty does not make a large difference to the inferred posterior distributions. However, the different PSD estimation techniques while giving consistent posterior distributions, give different posterior weights to different parts of the sky.

PSD	Marg vs no marg
Mean	19.26
Median	91.67

TABLE III: Values of the (natural) log Bayes factor comparing the marginalised and unmarginalised likelihood hypotheses. In both cases there is a strong preference for the marginalised likelihood better describing the data than the unmarginalised likelihood.

averaged and median-averaged PSDs to obtain samples from the posterior distribution and Bayesian evidences under four sets of assumptions.

1. The data are well described by the mean-estimated PSD and the Whittle likelihood.

2. The data are well described by the mean-estimated PSD and the Student-Rayleigh likelihood.

3. The data are well described by the median-estimated PSD and the Whittle likelihood.

4. The data are well described by the median-estimated PSD and the median marginalized likelihood, Equation 23.

In Table II we show the natural logarithm of the BCI under these four set of assumptions. We find that both PSD estimation methods have $\ln \mathrm{BCI} \approx 10$ which is a moderately strong preference for the coherent hypothesis, although we note that a full treatment requires careful consideration of prior odds. For both PSD estimates

the BCI decreases slightly when marginalizing over the uncertainty. The increase in the BCI when using the median estimated PSD is mirrored in the increased signal-to-noise ratio ρ in the lower panel of Figure 7. This is likely due to the different handling of non-Gaussian features in the mean and median PSD estimation methods.

In Table III we show the natural log Bayes factors comparing the marginalized to unmarginalized likelihoods for both PSD estimation methods. In both cases, we see a strong preference for the model which marginalizes over the uncertainty. This preference is much larger for the median PSD estimate. This can be understood by the fact that the large $|\nu|$ tail of the median marginalized likelihood is broader than the mean marginalized likelihood, c.f., Figure 1, lower-right panel.

In Figures 7 and 8 we show selected posterior distribution under our four sets of assumptions. In the top panel of Figure 7 we show the posterior distribution for the best measured combination of the component masses, the chirp mass

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_1)^{1/5}}. \tag{37}$$

The parameters m_i are the masses of the two component black holes. In the bottom panel we show the network matched filter signal-to-noise ratio (SNR). We find that the recovered SNR is larger when using the median PSD estimate and the marginalizing over the uncertainty in the PSD increases the posterior support at SNR less than the maximum found SNR but does not decrease the maximum SNR. Correspondingly, we see that the posterior for chirp mass is slightly less strongly peaked when marginalizing over uncertainty in the PSD, and when using the median PSD estimate.

In Figure 8 we show the posterior distribution for the parameters describing the position on the sky, right ascension α and declination δ. The impact of marginalizing over the uncertainty in the PSD does not significantly affect the inferred sky localisation of the binary. However, the two PSD estimation methods recover different posterior distributions within the same region on the sky.

The fact that the differences in the posterior distributions and Bayes factors when using the different PSD estimation methods are larger than the corrections due to marginalizing over the statistical uncertainty mean that either one or both of the estimation methods are producing a biased estimate of the true PSDs. The source of this bias is presumably non-stationarity and/or non-Gaussianity in the data used to estimate the PSD. It is not possible to determine which estimate is less biased and so this must be considered as an extra source of systematic uncertainty in off-source PSD estimates.

VI. DISCUSSION

Performing astrophysical inference on gravitational-wave data requires an estimate of the noise power spectral density (PSD). In practice it is not possible to know this perfectly, only to estimate it from the data. Multiple methods of estimating the PSD are used, and each carries with it a different class of statistical uncertainty. In this work, we derived the relevant statistical uncertainty for an estimation method, which is widely used when analyzing gravitational-wave transients, the median average. We obtained a closed-form expression for the likelihood, which marginalizes over the statistical uncertainty, and demonstrated that simulated Gaussian data matches this distribution.

We then applied our new results to the lowest significance transient in the first LIGO/Virgo gravitational-wave transient catalog, GW151012. We analyzed this event using two different PSDs with likelihoods, which did and did not marginalize over the appropriate statistical uncertainty, one using a median average, and one using a mean average. We showed the PSD estimation method has a clear effect on the inferred posterior probability distribution and Bayesian evidence. The changes in the posterior distributions and Bayesian evidence when marginalizing over the statistical uncertainty is more subtle. However, for applications which require precise estimates of the evidence such as [24–26], these small differences will be crucial.

There are many interesting extensions to the work presented here, which are left to future work. These include implementing data quality tests when analyzing real gravitational-wave data, which are known to be non-Gaussian and non-stationary over timescales of minutes to hours. Examples of this can be found in other areas of gravitational-wave data analysis. For example, searches for gravitational-wave transient signals implement methods to track and mitigate non-stationarity [11] and remove large non-Gaussian transients [10, 11]. Searches for continuous gravitational-wave sources and the stochastic gravitational-wave background include algorithms to detect and remove stretches of data where the PSD is rapidly fluctuating [5] or frequencies where the data are known to be non-Gaussian, e.g., around the frequency of AC mains electricity [1]. By combining these methods and the statistical models presented here, we can enable precision astrophysical inference for gravitational-wave transients without large computational overheads.

[1] B. P. Abbott *et al.*, arXiv:1908.11170.

[2] J. D. Romano and N. J. Cornish, Living Rev. Relativ. **20**, 2 (2017), arXiv:1608.06889.

[3] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **91**, 084034 (2015), arXiv:1410.3852.

[4] P. Welch, IEEE Trans. Audio Electroacoust. **15**, 70 (1967).

[5] B. P. Abbott *et al.*, Phys. Rev. Lett. **118**, 121101 (2017), arXiv:1612.02029 [gr-qc].

[6] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. Dal Canton, T. Dent, S. Fairhurst, M. S. Kehl, D. Keppel, B. Krishnan, A. Lenon, A. Lundgren, A. B. Nielsen, L. P. Pekowsky, H. P. Pfeiffer, P. R. Saulson, M. West, and J. L. Willis, Classical and Quantum Gravity **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].

[7] C. Pankow, K. Chatziioannou, E. A. Chase, T. B. Littenberg, M. Evans, J. McIver, N. J. Cornish, C.-J. Haster, J. Kanner, V. Raymond, S. Vitale, and A. Zimmerman, Phys. Rev. D **98**, 084016 (2018), arXiv:1808.03619 [gr-qc].

[8] J. C. Driggers, S. Vitale, A. P. Lundgren, M. Evans, K. Kawabe, S. E. Dwyer, K. Izumi, R. M. S. Schofield, A. Effler, D. Sigg, *et al.*, Phys. Rev. D **99**, 042001 (2019), arXiv:1806.00532 [astro-ph.IM].

[9] D. Davis, T. Massinger, A. Lundgren, J. C. Driggers, A. L. Urban, and L. Nuttall, Classical and Quantum Gravity **36**, 055011 (2019), arXiv:1809.05348 [astro-ph.IM].

[10] S. Sachdev, S. Caudill, H. Fong, R. K. L. Lo, C. Messick, D. Mukherjee, R. Magee, L. Tsukada, K. Blackburn, P. Brady, P. Brockill, *et al.*, (2019), arXiv:1901.08580 [gr-qc].

[11] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, Phys. Rev. D **100**, 023011 (2019), arXiv:1902.10341 [astro-ph.IM].

[12] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale, Phys. Rev. D **101**, 042003 (2020), arXiv:1911.09083 [gr-qc].

[13] K. Chatziioannou, C.-J. Haster, T. B. Littenberg, W. M. Farr, S. Ghonge, M. Millhouse, J. A. Clark, and N. Cornish, (2019), arXiv:1907.06540.

[14] C. Röver, Phys. Rev. D **84**, 122004 (2011), arXiv:1109.0442.

[15] C. Röver, R. Meyer, and N. Christensen, Classical and Quantum Gravity **28**, 015010 (2011), arXiv:0804.3853 [stat.ME].

[16] T. Yamamoto, K. Hayama, S. Mano, Y. Itoh, and N. Kanda, Phys. Rev. D **93**, 082005 (2016).

[17] S. Banagiri, M. W. Coughlin, J. Clark, P. D. Lasky, M. A. Bizouard, C. Talbot, E. Thrane, and V. Mandic, (2019), arXiv:1909.01934.

[18] N. J. Cornish and T. B. Littenberg, Class. Quantum Gravity **32**, 135012 (2015), arXiv:1410.3835.

[19] B. P. Abbott *et al.*, Physical Review X **9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[20] S. Biscoveanu, C.-J. Haster, S. Vitale, and J. Davies, arXiv e-prints (2020), arXiv:2004.05149 [astro-ph.HE].

[21] C. Talbot, *Astrophysics of Binary Black Holes at the Dawn of Gravitational-Wave Astronomy*, Ph.D. thesis, Monash University (2020).

[22] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, Phys. Rev. D **85**, 122006 (2012), arXiv:0509116 [gr-qc].

[23] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade, Phys. Rev. D **91**, 42003 (2015).

[24] M. Isi, R. Smith, S. Vitale, T. J. Massinger, J. Kanner, and A. Vajpeyi, Phys. Rev. D **98**, 042007 (2018), arXiv:1803.09783 [gr-qc].

[25] G. Ashton, E. Thrane, and R. J. E. Smith, Phys. Rev. D **100**, 123018 (2019), arXiv:1909.11872 [gr-qc].

[26] R. Smith and E. Thrane, Physical Review X **8**, 021019 (2018), arXiv:1712.00688 [gr-qc].

[27] W. K. Hastings, Biometrika **57**, 97 (1970), https://academic.oup.com/biomet/article-pdf/57/1/97/23940249/57-1-97.pdf.

[28] J. Skilling, Bayesian Anal. **1**, 833 (2006).

[29] E. Thrane and C. Talbot, Publ. Astron. Soc. Aust. **36**, e010 (2019), arXiv:1809.02293.

[30] E. Payne, C. Talbot, and E. Thrane, arXiv:1905.05477.

[31] J. S. Speagle, arXiv:1904.02180.

[32] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, *et al.*, Astrophys. J. Suppl. Ser. **241**, 27 (2019), arXiv:1811.02042.

[33] M. Hübner, C. Talbot, P. D. Lasky, and E. Thrane, Phys. Rev. D **101**, 023011 (2020), arXiv:1911.12496 [astro-ph.HE].

[34] E. Sellentin and A. F. Heavens, Mon. Not. R. Astron. Soc. Lett. **456**, L132 (2016), arXiv:1511.05969.

[35] B. P. Abbott *et al.*, Living Reviews in Relativity **21**, 3 (2018), arXiv:1304.0670 [gr-qc].

[36] J. Aasi *et al.*, Classical and Quantum Gravity **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].

[37] B. P. Abbott *et al.*, Physical Review X **6**, 041015 (2016), arXiv:1606.04856 [gr-qc].

[38] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, J. Phys. Conf. Ser. **610**, 012021 (2015), arXiv:1410.4839.

[39] The LIGO Scientific Collaboration, R. the Virgo Collaboration and Abbott, *et al.*, (2019), arXiv:1912.11716 [gr-qc].