Issues of mismodeling gravitational-wave data for parameter estimation

Oliver Edy, Andrew Lundgren, and Laura K. Nuttall

The Issues of Mismodelling Gravitational-Wave Data for Parameter Estimation

Oliver Edy,* Andrew Lundgren,† and Laura K. Nuttall‡

University of Portsmouth, Institute of Cosmology and Gravitation, Portsmouth PO1 3FX, United Kingdom

Bayesian inference is used to extract unknown parameters from gravitational wave signals. Detector noise is typically modelled as stationary, although data from the LIGO and Virgo detectors is not stationary. We demonstrate that the posterior of estimated waveform parameters is no longer valid under the assumption of stationarity. We show that while the posterior is unbiased, the errors will be under- or overestimated compared to the true posterior. A formalism was developed to measure the effect of the mismodelling, and found the effect of any form of non-stationarity has an effect on the results, but are not significant in certain circumstances. We demonstrate the effect of short-duration Gaussian noise bursts and persistent oscillatory modulation of the noise on binary-black-hole-like signals. In the case of short signals, non-stationarity in the data does not have a large effect on the parameter estimation, but the errors from non-stationary data containing signals lasting tens of seconds or longer will be several times worse than if the noise was stationary. Accounting for this limiting factor in parameter sensitivity could be very important for achieving accurate astronomical results. This methodology for handling the non-stationarity will also be invaluable for analysis of waveforms that last minutes or longer, such as those we expect to see with the Einstein Telescope.

I. INTRODUCTION

The Advanced LIGO detectors have been regularly detecting gravitational wave signals since 2015 [1], joined by Advanced Virgo in 2017 [2], and the later addition of KAGRA in 2020 [3]. The signals observed so far have been produced by the coalescence of either a binary system of black holes (BBH) or neutron stars (BNS). The gravitational waves emitted contain information we can infer about the system, such as the mass of each component of the binary or their sky position. The signal is identified through techniques such as matched filtering [4], which determines how closely the observed data matches with a theoretical waveform model known as a template. In practice, the output of the matched filter is calculated for a large bank of templates spanning the parameter space of expected signals [5]. However, the exact physical properties and physics of the system must be extracted using Bayesian analysis [6]; these models evaluate a posterior probability density function of the unknown signal parameters that describe a specific model of the data. While Bayesian analysis is not designed to identify the template that fits the signal best, when a suitable template has been found with matched filtering, Bayesian methods can obtain the parameter uncertainties more directly than the template search.

A gravitational-wave signal is fully described by 15 parameters, assuming negligible orbital eccentricity [6, 7]; 8 of these are intrinsic to the observed system, including the mass of each body before coalescence, or their spins, while the remaining 7 depend on the observer (such as sky position, or the time t_c in the observing frame at which the coalescence occurred). This means that even

inference calculations of signals only a few seconds long are computationally expensive. Although techniques are being developed to improve processing time (for example, [8–13]), they are not yet mature. The most widely used inference codes for gravitational waves are LALInference [6], Bilby [14, 15], and PyCBC Inference [16]. These, as well as all other current methods, assume that the instrumental noise is both stationary and Gaussian (see, for example, [6, 16–19]). Note that we define stationarity to mean that the statistical properties of the noise do not change with time. Signal processing often uses wide-sense stationarity, where only the mean and covariance are assumed to be time-independent [20]. The fact that the statistical parameters of the noise (such as the spectrum shape or the intensity) do not vary with time means the data can be characterised [21]. Consequently, modelling and analysis of the noise in the frequency-domain can be simplified, which makes evaluation computationally efficient [22].

In reality, the gravitational-wave strain data from the LIGO and Virgo detectors is not stationary or Gaussian [23–27]. Hence, inference using the appropriate non-stationary model could not make use of the simpler form. There have been some methodologies proposed for handling the non-stationarity. For example, References [28, 29] propose modelling the noise in a wavelet domain, where non-stationarity is more simply computed. This method has proven particularly useful in analysis of data containing glitches [30]. Other methods include estimating the variation in detector noise and using it to re-rank events to reduce the number of false events being reported [31]), or incorporating the uncertainty in the measurement of the power spectral density into parameter estimation [32].

Typically, though, the assumption of stationarity and Gaussianity is taken instead [33], simply as matter of practicality because it enables a much simpler form for the likelihood to be used, where the 'likelihood' refers to the model we choose for the data. Given that parame-

---

* oliver.edy@port.ac.uk
† andrew.lundgren@port.ac.uk
‡ laura.nuttall@port.ac.uk

ter estimation is known for its computational cost [6], a simpler model means that analysis is inexpensive. However, mismodelling the data in this way will affect the certainty with which we can estimate parameters from a gravitational-wave signal, a probability function known as the posterior. This could mean that uncertainties for parameters inferred from the system could be much smaller or larger than they really are. This compromise on the effectiveness of parameter estimation has already been explored in Reference [27].

In this paper, we focus on the problems with non-stationary data and parameter estimation. While previous work such as those cited above have focused on handling the effects of non-stationarity, we concern ourselves with investigating how gravitational-wave analysis is affected if we do not account for the non-stationarity of the data. Specifically, our goal is to determine the effect that non-stationary noise will have on the estimated parameters, given that the assumed Bayesian model does not fit the data. This will establish the amount of non-stationarity which is allowable before the estimated parameters are significantly affected and it becomes necessary to explicitly account for the non-stationarity. In Section II, we first give a brief overview of how parameter estimation works, then consider whether approximating the data as stationary is a valid assumption to make. In Section III, we suggest a model of Gaussian data with which we can determine deviations from stationarity by considering the covariance matrix. A few examples of how this might work are given in Section IV.

II. DESCRIBING NON-STATIONARITY

A. Extracting Signals from Noisy Data

A signal of known shape $h(t)$ is optimally extracted from noisy strain data $n(t)$ through a matched filter comparison with a number of templates [34]. The mass and spin of a signal must then be approximately the same as those that generated the matching template. Once a signal has been identified with techniques such as matched-filtering, analysis shifts to estimating the parameters of the compact-object binaries that generated the gravitational waves.

Methods for inferring details about a system are rooted in Bayes Theorem, described in Ref [35] as a generalised technique for understanding a system when working with incomplete information. Formally, for a set of parameters θ, we represent this as

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}, \qquad (1)$$

where we refer to $p(\theta|d)$ as the posterior of the data, $p(d|\theta)$ the likelihood, $p(d)$ the evidence, and $p(\theta)$ the prior. The posterior is the probability of obtaining the parameters θ given the data $d(t)$. The prior $p(\theta)$ de-

scribes an understanding of the system and the parameters θ before knowing anything about the data $d(t)$; the choice of prior is very important, as a poorly chosen prior will strongly bias the entire parameter estimation and likely lead to flat posteriors or wildly incorrect parameter estimations [36, 37].

The denominator $p(d)$ is called the evidence, and is a normalisation factor which can be used to directly compare one model of the data with another [38]. Because of the way that Bayesian inference is implemented in determining the gravitational wave parameters, it is normalisation-insensitive, and so it is rare to actually need to calculate the evidence [39, 40]. Note, though, that this isn't always true, as the evidence is necessary in nested sampling, wherein new samples are drawn from a normalised prior probability (for example, Reference [41] describes how this is implemented for the dynesty code).

The likelihood $p(d|\theta)$ is the probability of obtaining the data $d(t)$ observed, given the parameters θ being considered. In essence, it is the model we choose for the data, and the part we have most control over.

As stated, the likelihood form $\mathcal{L}$ is assumed to be Gaussian of the form

$$\mathcal{L} = e^{-\frac{1}{2}\langle n(f)^\dagger | n(f) \rangle}, \qquad (2)$$

where the square brackets denote the inner product averaged over all noise realisations. Note that the full derivation for this likelihood is described in Reference [34].

With the assumption that the data is a multivariate Gaussian in the frequency domain, the noise covariance is diagonal, which we interpret to mean the frequencies are independent. In the case where the noise is one-dimensional, the Gaussian likelihood simplifies to

$$\mathcal{L} = e^{-\frac{1}{2}n^\dagger(f)S^{-1}n(f)}, \qquad (3)$$

where S is the noise covariance.

B. Validity of Stationarity Assumption

We must consider whether it is indeed valid to model the data as stationary. First, we look at a description of stationary noise. According to Wold's Theorem [42], any process X can be represented as the sum of a deterministic part D (which we shall ignore here) and a random part. The random part can then be further decomposed into a purely random component R convolved with a purely deterministic process C. For a discrete dataset sampled at an evenly spaced set of times n, we have

$$X_n = \sum_n C_{n-i}R_i + D_n = C * R + D, \qquad (4)$$

where R is completely uncorrelated, so that

$$\langle R_i R_j \rangle = 0 \text{ if } i \neq j. \qquad (5)$$

This form suggests that stationary noise is simply white noise (R) being acted upon by a filter (C).

Although there is no simple classification of non-stationary noise [43], there are multiple different models (see, for example, [42, 44–48]). We concern ourselves with models that decompose non-stationary noise in a similar manner to stationary noise in Wold's Theorem, this time into a family of non-deterministic stationary processes, or a family of deterministic non-stationary processes. We can compare this to the stationary case by now interpreting a time-varying filter to be acting on the white noise. The motivating example for this and followup work is non-stationary noise generated as the sum of some stationary Gaussian noise n_1 and a second Gaussian sample n_2 with oscillating amplitude $B(t)$:

$$n(t) = An_1(t) + B(t)n_2(t). \qquad (6)$$

This example has been chosen for its simplicity, while also approximating the behaviour assumed by current LIGO calibration models: a stationary background noise n_1, and an unpredictable part $n_2(t)$ [49]. Similar models have already been incorporated into parameter estimation [50, 51], with strong agreement to the results of the LIGO data releases [52, 53]. This indicates that as well as being a simple representation, the model is also a good fit for the real interferometer noise.

The assumption of stationarity is often fair to make providing that the amplitude $B(t)$ evolves slowly in time or is otherwise small. In the case of ground-based interferometers, the noise is measured to be nearly stationary at least over a period of several seconds around the same duration as the signals that LIGO is able to detect [52, 54].

However, Reference [21] have already shown these assumptions of stationarity break down for periods of ~ 64 seconds, the timescales needed for the analysis of binary neutron star mergers. This was further explored in Reference [27], which showed that parameter estimation will be noticeably compromised for segments of duration as small as 128s.

Since we do not have a quantitative knowledge of the effect of non-stationarity, we cannot know how uncertain the posteriors might be even for short signals. Therefore, to understand any impact on parameter estimation, we present a formalism to describe how non-stationary data might differ from stationary data.

C. Characterising the Noise

Gaussian noise can be completely characterised by its mean and covariance [27], where the noise covariance represents random noise characteristics of the data. We define $n(f)$ as the discrete Fourier transform of a stretch of non-stationary noise and we represent it as a complex column vector n. Note that the noise is Gaussian [21], so

$$\langle n \rangle = 0, \qquad (7)$$

where the square brackets denote the ensemble average over all realisations of the noise.

We define the noise covariance Σ as

$$\Sigma = \langle nn^\dagger \rangle, \qquad (8)$$

and in the case of stationary noise, we instead express the same quantity as

$$\Sigma_{\text{stationary}} = S. \qquad (9)$$

The covariance matrix reduces to a simpler form for stationary noise. In the frequency domain, the noise covariance S becomes a positive diagonal matrix of the form

$$S = \begin{pmatrix} S_1 & 0 & \dots & 0 \\ 0 & S_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S_N \end{pmatrix}, \qquad (10)$$

where N is the number of data points. Since the likelihood is of the form $n^\dagger S^{-1} n$, we only care about values along the diagonal. This leaves a very simple likelihood form, as explored in Section III A.

When the noise is non-stationary, we have no robust method for estimating Σ from the noise power spectrum, as we would for stationary noise. This is due to the near impossibility of characterising the data, due to the time variance of the statistical parameters [21]. As outlined in Reference [55], if we cannot assume stationarity, the method commonly used to estimate the power spectral density is to calculate an *off-source* estimate, taking a mean of neighbouring segments, although this method will be more restrictive when considering the longer data segments needed to analyse a BNS.

Additional methods for calculating an estimate for the power spectral density have been proposed. For example, References [56–58] discuss the concept of non-stationarities in detector data evolving the PSD over time, referred to as PSD-drift. This causes a loss in sensitivity of the detectors, so correction methodology is offered by computing a running estimate of all matched filter overlaps. As Reference [59] notes, by correcting for PSD-drift, we see a notable difference in estimated parameters compared to the configurations assuming stationary data.

Even when the non-stationary covariance Σ can be estimated, problems arise because it is not diagonal. This means the same simplification of the likelihood is not possible, and so the computation becomes far more complicated.

Using the likelihood given by Equation (3), and knowing the form of the noise covariance S, we will now derive a probe to investigate how greatly parameter estimation is affected by the non-stationarity of the data.

III. MEASURING THE EFFECT OF NON-STATIONARITY

A. Deriving the Covariance Matrix of Waveform Parameters

We want to determine how greatly the estimated parameters of a particular waveform will be affected by the incorrect assumption of stationary noise. The best way to investigate this would be to derive the covariance matrix of the posterior distribution for the estimated waveform parameters, which describes how confidently we can measure each parameter.

The noise covariance is assumed to be a positive diagonal matrix S. The data, in the frequency domain, is written as a complex column matrix d. We use complex variables throughout the derivation, because the noise vector n is complex in Fourier space. We then convert to a real matrix when we reach the final result (the covariance matrix of the waveform parameters).

We assume that the template is not a perfect fit for the signal. As in Reference [60], we approximate the template waveform h as a linear series around the true signal h_0,

$$h = h_0 + \omega\vec{\theta}, \tag{11}$$

where the ω matrix is comprised of the derivatives of h by each waveform parameter, and $\vec{\theta}$ is the set of parameters (we will also encounter its Hermitian conjugate $\vec{\theta}$ in the derivation). Note that for simplicity we have only taken Equation (11) to first-order.

We assume that the signal has been found in Gaussian data. The corresponding log-likelihood (up to a constant normalisation) is

$$\Lambda \propto -\frac{1}{2}\left(d - h | d - h\right) + \frac{1}{2}\left(d | d\right). \tag{12}$$

Note that the matrix notation of the noise-weighted inner product is given by

$$(a|b) = 4a^\dagger S^{-1}b, \tag{13}$$

so the log-likelihood expression reduces to

$$\Lambda \propto 2\left(d^\dagger S^{-1}h + h^\dagger S^{-1}d - h^\dagger S^{-1}h\right). \tag{14}$$

We decompose the strain data d into the sum of the signal h_0 and the instrumental noise n [27]:

$$d = n + h_0. \tag{15}$$

Using this substitution for d in Equation (14), as well as the expansion (11), the log-likelihood becomes

$$\begin{aligned}\Lambda \propto 2\Big(&n^\dagger S^{-1}h_0 + h_0^\dagger S^{-1}n + h_0^\dagger S^{-1}h_0 + n^\dagger S^{-1}\omega\vec{\theta} \\ &+ \vec{\theta}^\dagger\omega^\dagger S^{-1}n - \vec{\theta}^\dagger\omega^\dagger S^{-1}\omega\vec{\theta}\Big).\end{aligned} \tag{16}$$

Maximising Λ with respect to $\vec{\theta}$ and its Hermitian conjugate $\vec{\theta}^\dagger$, we find the following expressions for the parameters at the maximum likelihood:

$$\begin{aligned}\vec{\theta}_{\mathrm{ml}} &= \left(\omega^\dagger S^{-1}\omega\right)^{-1}\left(\omega^\dagger S^{-1}n\right), \\ \vec{\theta}_{\mathrm{ml}}^\dagger &= \left(n^\dagger S^{-1}\omega\right)\left(\omega^\dagger S^{-1}\omega\right)^{-1}.\end{aligned} \tag{17}$$

Note that we interpret $\vec{\theta}_{\mathrm{ml}}$ as complex to simplify calculations, hence the presence of $\vec{\theta}_{\mathrm{ml}}^\dagger$. We convert $\vec{\theta}$ into a fully real matrix in Section III C.

From References [61, 62], we note that

$$\langle\vec{\theta}_{\mathrm{ml}}\rangle = 0, \tag{18}$$

so long as the template matches the true signal, which shows that the posterior is still centered around the same point. Therefore, to linear order at least, mismodelling the data will only deform the shape of the posterior (thereby over- or under-representing the confidence contours), but does not cause any bias to the estimated parameters.

For ease of reading, the ml subscript will be dropped from $\vec{\theta}$ for the rest of the text unless otherwise stated.

By multiplying the two quantities in Equation (17) together, we create the covariance matrix for $\vec{\theta}$:

$$\langle\vec{\theta}\vec{\theta}^\dagger\rangle = \left(\omega^\dagger S^{-1}\omega\right)^{-1}\left(\omega^\dagger S^{-1}\langle nn^\dagger\rangle S^{-1}\omega\right)\left(\omega^\dagger S^{-1}\omega\right)^{-1}. \tag{19}$$

As the noise covariance is defined in Equation (8), we see that the covariance matrix becomes

$$\langle\vec{\theta}\vec{\theta}^\dagger\rangle = \left(\omega^\dagger S^{-1}\omega\right)^{-1}\left(\omega^\dagger S^{-1}\Sigma S^{-1}\omega\right)\left(\omega^\dagger S^{-1}\omega\right)^{-1} \tag{20}$$

for non-stationary noise. However, in the case when S matches the true covariance Σ, then this greatly simplifies. This is the case for stationary noise, because the quantity $\langle nn^\dagger\rangle = S$ when averaged over infinite noise realisations, as stated in Equation (9), resulting in the familiar

$$\langle\vec{\theta}\vec{\theta}^\dagger\rangle = \left(\omega^\dagger S^{-1}\omega\right)^{-1}. \tag{21}$$

This result is comparable to the inverse of the Fisher matrix [63]. We intend to use deviations from the stationary covariance matrix (Equation (21)) to show how mismodelling non-stationary data can affect parameter estimation.

B. Distinction Between $\vec{\theta}$ and $\vec{\theta}_{\mathrm{ml}}$

To fully understand the difference between $\vec{\theta}$ and $\vec{\theta}_{\mathrm{ml}}$, we must consider the basis of parameter estimation in Bayes' theorem [35], as given by Equation (1).

$\langle\vec{\theta}_{\mathrm{ml}}\vec{\theta}_{\mathrm{ml}}^\dagger\rangle$ is the covariance matrix of the parameter set $\vec{\theta}$ at the maximum likelihood, averaged over all noise realisations. This makes $\vec{\theta}_{\mathrm{ml}}^\dagger$ a point estimate, not a random

variable; we have sought point values for $\vec{\theta}$ which maximise the posterior $p(\vec{\theta}|d)$, and treat the term $\frac{p(\vec{\theta})}{p(d)}$ as a constant. In this way, we do not inject any prior beliefs about the parameters into the calculation for likely values of $\vec{\theta}$, and so effectively ignore the prior.

A full Bayesian treatment would estimate the entire distribution for $p(\vec{\theta}|d)$, treating $\vec{\theta}$ as an array of random variables. Although the inclusion of the prior would more tightly constrain the covariance matrix, we consider only the covariance of the maximum likelihood here both because its inverse is the proper definition of the Fisher matrix [61], and because the prior doesn't change considerably in the regime of high signal-to-noise ratio (SNR) [62]. This last point in particular is important as we are concerning ourselves with whether analysis of a real signal is affected by the non-stationarity of the data.

C. Dealing with Nuisance Parameters

To minimise dimensionality during calculations, the covariance matrix has been treated as complex, but the parameters of $\vec{\theta}$ are real. It is at this point that we decompose the complex matrix from Equation (21) into a fully real matrix, with each column or row representing a parameter. Since the matrix has been treated as complex, however, there are several rows and columns that now contain nuisance parameters, which are the imaginary counterparts to parameters which do exist in the waveform. For example, we show in Section IV A by taking the derivative of the waveform by the phase ϕ, that it is simply the complex counterpart to the derivative by the amplitude $\mathcal{A}$. Similarly, the other parameters of $h(f)$ have complex counterparts, although these are non-physical, and not represented in the waveform $h(f)$.

Given that Equation (20) is a covariance matrix, it is possible to *fix* these nuisance parameters. This is done by inverting it into a Fisher matrix, removing the rows and columns corresponding to these nuisance parameters, and inverting back to find the new covariance matrix. Therefore, we are left with a covariance matrix for only the physical parameters.

For now, let's only consider parameters $\mathcal{A}$, ϕ, and t_c. Given we have decomposed the matrix into to be fully real, we would also need to consider the complex counterpart to the parameters. For ϕ, that is simply $\mathcal{A}$, as mentioned above, but for t_c, we would need to introduce an additional nuisance parameter, ξ, which will shortly be fixed from the covariance matrix.

The log-likelihood would be of a form similar to

$$\begin{aligned}
\Lambda &= \vec{\theta}^T C^{-1} \vec{\theta} \\
&= \begin{pmatrix} \Delta\mathcal{A} & \Delta\phi & \Delta\xi & \Delta t_c \end{pmatrix} \begin{pmatrix} \sigma_{\mathcal{A}\mathcal{A}} & \sigma_{\mathcal{A}\phi} & \sigma_{\mathcal{A}\xi} & \sigma_{\mathcal{A}t_c} \\ \sigma_{\phi\mathcal{A}} & \sigma_{\phi\phi} & \sigma_{\phi\xi} & \sigma_{\phi t_c} \\ \sigma_{\xi\mathcal{A}} & \sigma_{\xi\phi} & \sigma_{\xi\xi} & \sigma_{\xi t_c} \\ \sigma_{t_c\mathcal{A}} & \sigma_{t_c\phi} & \sigma_{t_c\xi} & \sigma_{t_c t_c} \end{pmatrix} \begin{pmatrix} \Delta\mathcal{A} \\ \Delta\phi \\ \Delta\xi \\ \Delta t_c \end{pmatrix},
\end{aligned} \tag{22}$$

where C is our covariance matrix.

We're not actually interested in the nuisance parameter ξ, as it is not a term of the waveform $h(f)$. The means we know that $\xi = 0$ in the waveform, and can be removed from the Fisher matrix by fixing it. This involves stripping out any row or column that would be multiplied by ξ. Therefore, Equation (22) reduces to

$$\Lambda = \begin{pmatrix} \Delta\mathcal{A} & \Delta\phi & \Delta t_c \end{pmatrix} \begin{pmatrix} \sigma_{\mathcal{A}\mathcal{A}} & \sigma_{\mathcal{A}\phi} & \sigma_{\mathcal{A}t_c} \\ \sigma_{\phi\mathcal{A}} & \sigma_{\phi\phi} & \sigma_{\phi t_c} \\ \sigma_{t_c\mathcal{A}} & \sigma_{t_c\phi} & \sigma_{t_c t_c} \end{pmatrix} \begin{pmatrix} \Delta\mathcal{A} \\ \Delta\phi \\ \Delta t_c \end{pmatrix}. \tag{23}$$

The Fisher matrix used here can then be inverted back to get the new covariance matrix.

There are other parameters that we want to remove because they are not useful in the analysis, such as the phase ϕ, which tells us no extra information than the amplitude $\mathcal{A}$. These cannot be treated in the same way because they are real parameters of $h(f)$ and we do not know their value to be able to fix them. Instead, we must marginalise over these parameters, integrating them out of the probability function. An example of such is given in Section IV A.

IV. APPLICATIONS

A. 2D Toy Model

We propose a toy model to demonstrate the calculation and subsequent handling of the covariance matrix of a waveform's parameters. Let us now consider the simplest case, for which we assume that we know all the parameters of the waveform $h(f)$, except for the amplitude $\mathcal{A}$, and the phase ϕ. The waveform can now be described as follows:

$$h(f) = h_0(f) e^{\mathcal{A} + i\phi}, \tag{24}$$

where h_0 is the waveform of known parameters.

The corresponding derivatives of $h(f)$ by these two parameters are

$$\partial_{\mathcal{A}} h = h, \qquad \partial_{\phi} h = ih. \tag{25}$$

Taking the linear waveform approximation as in Equation (11), and recalling that ω is the derivative of the waveform with the parameters, we find an $N \times 1$ row matrix of

$$\omega = (\partial_{\vec{\theta}} h) \tag{26}$$
$$= (\partial_{\mathcal{A}} h + \partial_{\phi} h) \tag{27}$$
$$= (h + ih), \tag{28}$$

where N is the length of h.

Using this ω in Equation (21), we re-express the covariance matrix as

$$\langle \vec{\theta}\vec{\theta}^{\dagger} \rangle = \left((h^{\dagger} - ih^{\dagger}) S^{-1} (h + ih) \right)^{-1}$$
$$= \left(2h^{\dagger} S^{-1} h + i \left(h^{\dagger} S^{-1} h - h^{\dagger} S^{-1} h \right) \right)^{-1}$$
$$= (\sigma_1 + i\sigma_2), \tag{29}$$

where $\sigma_1 = 2h^{\dagger} S^{-1} h$ and $\sigma_2 = 0$.

The covariance matrix can be decomposed into a fully real 2×2 matrix by separating out the real and imaginary values as follows:

$$\langle \vec{\theta}\vec{\theta}^{\dagger} \rangle = \begin{pmatrix} \sigma_1 & \sigma_2 \\ -\sigma_2 & \sigma_1 \end{pmatrix}. \tag{30}$$

With the matrix decomposition demonstrated, from here on σ_2 will be explicitly referred to as 0.

The corresponding log-likelihood is

$$\Lambda = \begin{pmatrix} \Delta\mathcal{A} & \Delta\phi \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_1 \end{pmatrix} \begin{pmatrix} \Delta\mathcal{A} \\ \Delta\phi \end{pmatrix}. \tag{31}$$

Note the form of the covariance matrix:

$$\begin{array}{c|cc} & \Delta\mathcal{A} & \Delta\phi \\ \hline \Delta\mathcal{A} & \sigma_1 & 0 \\ \Delta\phi & 0 & \sigma_1 \end{array} \tag{32}$$

The covariance between $\Delta\mathcal{A}$ and its complex counterpart $\Delta\mathcal{A}$ is 0, while both $\Delta\mathcal{A}$ and $\Delta\phi$ have the same covariance with themselves this is true for all parameters and their complex counterparts.

We can marginalise over any parameter we are not interested in to remove it from the likelihood, and further simplify the likelihood form for computational efficiency.

We receive the same information from both the amplitude and the phase, so we do not need both. For demonstration purposes, we are not interested in the phase of the gravitational wave, and so marginalise over the parameter to remove it from the likelihood in Equation (31). For a Gaussian probability, marginalising over a parameter produces an identical result to maximising by that parameter (up to normalisation), so for ease, we maximise by ϕ instead,

$$\frac{\partial\Lambda}{\partial\Delta\phi} = \frac{\partial}{\partial\Delta\phi} \left(\sigma_1 \Delta\mathcal{A}^2 + \sigma_1 \Delta\phi^2 \right) = 0. \tag{33}$$

We can see that $\Delta\phi = 0$, and so the simplified form of the log-likelihood is

$$\Lambda = (\Delta\mathcal{A}) (\sigma_1) (\Delta\mathcal{A}). \tag{34}$$

Interestingly, in this simple case, this is the same result as if we had fixed ϕ, but typically fixing a parameter and marginalising over it would produce very different matrices. Otherwise, this toy model provides a simple demonstration of how we can find and then simplify the covariance matrix for a system. In the next section, we then show how the covariance matrix might be affected by non-stationarity in a real-life scenario.

B. Visualising the Effect of Non-Stationarity

Although we have described the effect of non-stationarity in a theoretical capacity, we have yet to visualise it. To demonstrate, we shall compare the idealised stationary Gaussian model against two different non-stationary models which we name Model A and B. Model A localises the non-stationary noise over a short period of time, whereas the non-stationarity in Model B is present over the entire dataset and hence models an extended period of non-stationarity.

For Model A, we have generated Gaussian data as described by Equation (6), with the time-varying amplitude $B(t)$ set to 0 at all times except for a period of 8 seconds centred around a time t_0:

$$B(t) = \begin{cases} \text{Tukey}(\alpha = 0.5) & t_0 - 4 < t_0 < t_0 + 4 \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

In the latter instance, the amplitude will describe a Tukey window with the shape parameter $\alpha = 0.5$. The window was specifically chosen to last 8 seconds because it exceeds the length of the majority of signals measured by LIGO to date. Therefore, we will be able to determine whether the non-stationarity would have an effect on signals of comparable length.

In the case of Model B, the amplitude $B(t)$ described in Equation (6) is sinusoidal, oscillating at 3.2 Hz.

These models were chosen as simplified approximations to real phenomena seen in the LIGO interferometers. A short period of stationarity as seen in Model A is comparable to a short burst of non-stationarity that is created by environmental events such as thunderstorms. Model B is instead an approximation for extended periods of non-stationarity, analogous to microseismic ground-motion [64]. The realistic counterparts are never expected to become as extreme as the models described here, which were chosen to make the effect of non-stationarity on parameter estimation suitably explicit. Additionally, due to the very different forms of noise that we have created, it is not appropriate to compare the results of Model A or B with the other, but only against the stationary Gaussian model.
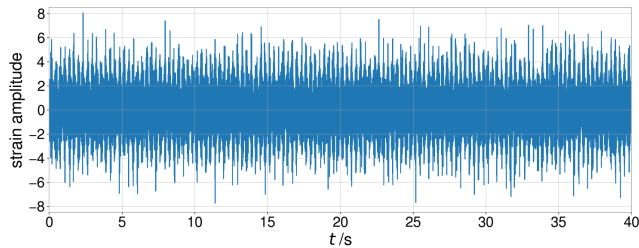
These three data samples (the stationary Gaussian noise, and Models A and B) will be 40 seconds long.
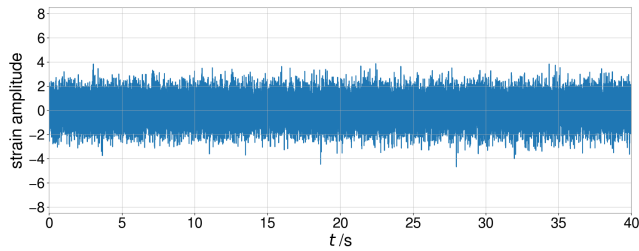
(a) Stationary Gaussian noise.
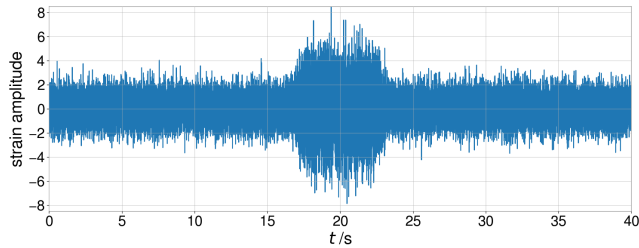
(b) Gaussian noise with an 8-second Tukey window applied at 16 to 24 seconds.

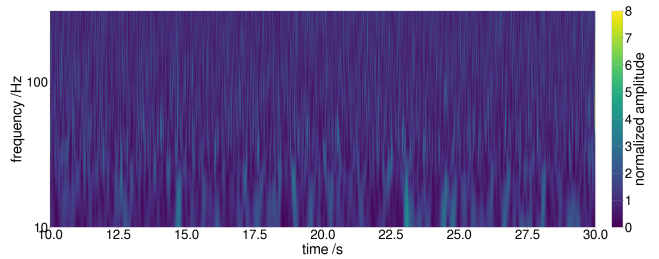(c) Gaussian noise with an oscillating amplitude of 3.2 Hz.

FIG. 1: Strain plots for three artificially generated timeseries over entire 40-second sample.

(a) Time-frequency spectrogram of stationary Gaussian noise.

(b) Time-frequency spectrogram of Gaussian noise with an 8-second Tukey window applied at 16 to 24 seconds.
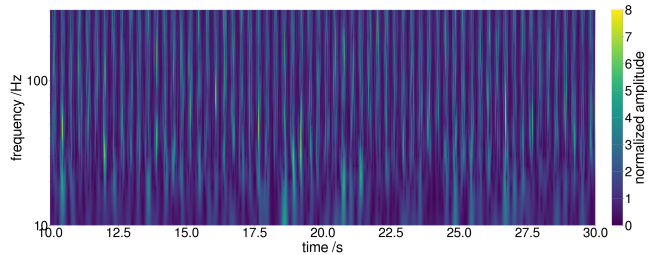
(c) Time-frequency spectrogram of Gaussian noise with an oscillating amplitude of 3.2 Hz.

FIG. 2: Time-frequency representation for three artificially generated timeseries using the Q-transform [65]. Note that all plots have been restricted to the period between 10 and 30 seconds to better highlight the oscillating peaks in plot c.

This is considered a suitable duration to produce reliable results, and not too long a duration to not be relevant to the shorter data segments analysed by inference codes.
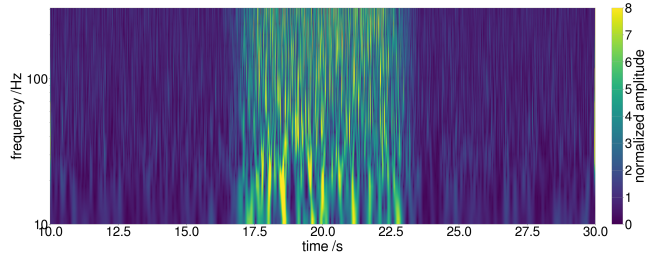
The data and time-frequency spectrograms for one sample of each type of noise have been plotted in Figure 1 and Figure 2 respectively.

Model A is almost identical to the stationary noise in both of these plots, except for the spike in power created by the Tukey window. This is exactly as we would expect given the noise is the sum of two separate Gaussian timeseries for this period ($n_1(t)$ and $n_2(t)$ in Equation (6), where the contribution of $n_2(t)$ is 0 for the rest of the time due to the amplitude $B(t)$).

The power spectra for each of these models was taken using the Welch method. Consequently, this means that the spectra we have calculated correspond to the stationary part of the spectrum. The standard procedure in searches is to estimate the power spectrum from longer stretches of data using the median [17]. This difference shouldn't change our conclusions, as they are comparable spectra for when the size of the non-stationarity is small, as is the case here.

When comparing the power spectra to each other (see Figure 3), the curve for Model A is very similar to the stationary Gaussian case. This is expected, as the power of the 8 seconds of non-stationarity would be washed out by the remaining 32 seconds of stationary Gaussian data.

While the spectrum for Model A has on average 2.8 times the power of the stationary Gaussian model, the spectrum for Model B has on average 13.3 times more power, an entire order of magnitude higher than the stationary Gaussian model. This is because the amplitude $B(t)$ extends over the entire 40-second sample, rather than a fraction of the entire duration. This means that the excess power will not be washed out by the stationary part.

So far, this has been a very qualitative review of the differences between these models. To get a better understanding of the deviations from stationarity, we then
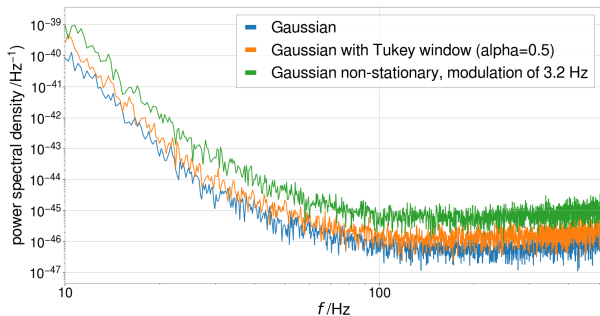
FIG. 3: Power spectrum comparing 40-second samples of stationary Gaussian noise (blue), Gaussian noise made non-stationary by a Tukey window (orange), and Gaussian noise made non-stationary by an sinusoidal modulation of 3.2 Hz (green).

calculated the noise covariance for each of these models using the formalism described by Section III.

According to Equation (8), the noise covariance is calculated as the inner product of the noise n with its Hermitian conjugate $n^\dagger$ averaged over infinite noise realisations. Here we are realistically limited to 10,000 realisations. The resulting covariance matrices have been plotted in Figure 4, limited to the 50 Hz around 256 Hz for both x- and y-axis to better showcase the diagonal and non-diagonal terms.

The stationary Gaussian model is a diagonal line of constant value, with every off-diagonal term on average 100 times smaller than the diagonal terms. This is as we predicted in Equation (9), and we would expect the off-diagonal terms to be exactly zero if we were able to average over infinite noise realisations, rather than the 10,000 we were realistically limited to. The difference between Model A and the stationary Gaussian noise model is evident by the thicker width of the leading diagonal.

Model B also has a diagonal line of constant value, in addition to off-diagonal lines. It is these lines that represent the deviation from stationarity, with their distance from the central diagonal line related to the amplitude $B(t)$. The rest of the values are approximately zero.

While it isn't immediately apparent in the plots shown, there is not just one set of off-diagonal lines created by the non-stationarity. The values of the first pair of lines off the central diagonal line are about a third of those on the central diagonal. The second set is just visible in Figure 4c, with the values close to a tenth of those on the leading diagonal. Values in the other sets of lines are so small as to not be discernible from the zero values. The magnitude of the values and the distance of the lines in relation to the leading diagonal is related to strength of the non-stationarity. In Figure 4c, that refers to the frequency at which $B(t)$ oscillates.

In Section II C, we stated that Gaussian noise can be completely characterised with its noise covariance, and it should now be apparent that the off-diagonal terms are how the non-stationarity manifests in the noise covari-
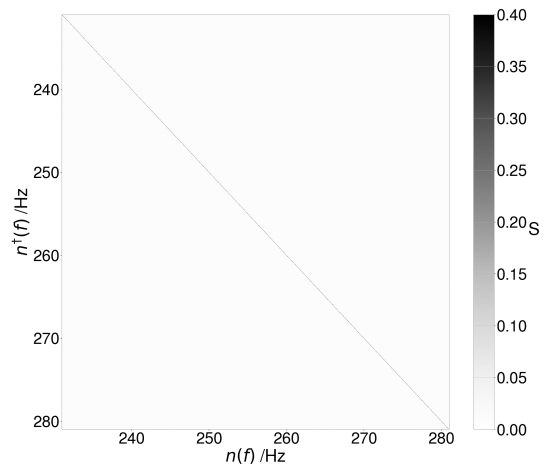
(a) Noise covariance for Gaussian data.
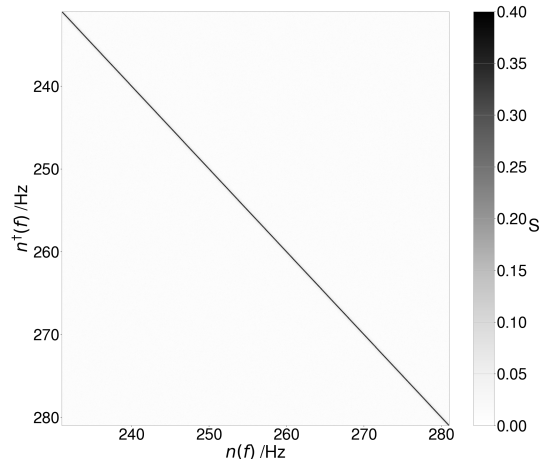
(b) Noise covariance for Gaussian data with an 8-second Tukey window applied at 16 to 24 seconds. Features to be noticed are several sets of off-diagonal lines clustered very close to the leading diagonal.
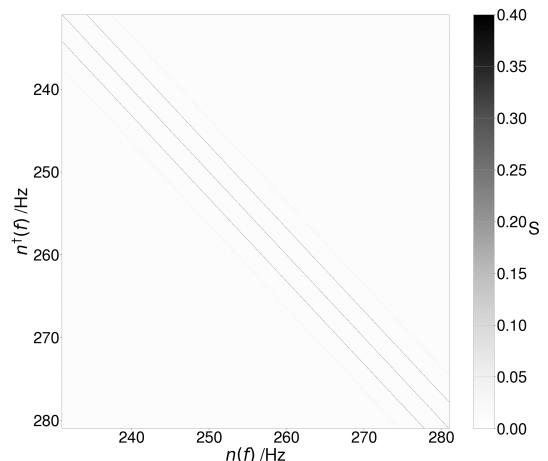
(c) Noise covariance for Gaussian data with an oscillating amplitude of 3.2 Hz. Features to be noticed are the two sets of off-diagonal lines, where the second set is only just visible.

FIG. 4: Noise covariance for three artificially generated timeseries. Note that data has been whitened and rescaled so that all values are between 0 and 1, for better comparison. We further restricted the colour axis to only values between 0 and 0.4, to better highlight the off-diagonal terms in plot c.

ance matrix. It is these non-diagonal terms that prevent us from assuming a simpler likelihood form. However, by knowing the noise covariance, we can calculate the extent that we will mismodel the data. As such, just by knowing Σ, we have a great description of the data. Unfortunately, estimating Σ accurately is very difficult; we have tried to minimise variance in the data by taking the average over 10,000 realisations of the noise, but this is obviously not possible for real noise.

We now look at the discrepancies in the covariance matrix of gravitational-wave waveform parameters that we might expect to see for mergers when we incorrectly assume a stationary noise covariance matrix.

C. The Effect of Non-Stationarity on the Covariance Matrix of a Short Gravitational-Wave Signal

We model a waveform with parameters purposefully chosen to be similar to GW150914 [66], with SNR set to 15. This is because the waveform is short in duration, and representative of the signals that LIGO can currently detect. That means we can show how uncertain the parameters of signals we currently observe will be given the non-stationarity is not accounted for.

The covariance matrix for the waveform parameters shows how confidently we can measure each parameter. We will compare the covariance matrix of the waveform parameters for Model A and Model B with the stationary Gaussian model. This was chosen as the reference because stationary Gaussian noise perfectly matches the likelihood form used, and so the posteriors will take on their true form.

The waveform model we have chosen to use is the TaylorF2, as defined by Reference [19], in which the signal $h(f)$ is of the form

$$h(f) = f^{-\frac{7}{6}} e^{\mathcal{A}+i\psi(f)}, \qquad (36)$$

where to 1.5PN,

$$\psi(f) = 2\pi f t_c + \phi_c - \frac{\pi}{4}$$
$$+ \frac{3}{128} (\pi\mathcal{M}f)^{-\frac{5}{3}} \left(1 + \frac{20}{9} \left(\frac{743}{336} + \frac{11}{4}\eta \right) (\pi M f)^{\frac{2}{3}} \right.$$
$$- 4 (4\pi - \beta) (\pi M f) + 10\varepsilon \left(\frac{3058673}{1016064} + \frac{5429}{1008}\eta \right.$$
$$\left. \left. + \frac{617}{144}\eta^2 - \sigma \right) (\pi M f)^{\frac{4}{3}} \right). \qquad (37)$$

While we are using parameters for a binary black hole here, we believe that the TaylorF2 can be used to sufficiently show the impact that non-stationarity can have on parameter estimation. We are not comparing different waveform models, but instead are comparing the effects of non-stationary noise models with a stationary noise model, using the same waveform throughout.

For simplicity, we reduce the TaylorF2 waveform to a form that only uses parameters $\mathcal{A}$, ϕ, t_c, and $\mathcal{M}$. With the exception of the chirp mass $\mathcal{M}$, each of these parameters are linearised, and so their corresponding values in the covariance matrix can be interpreted easily. The mass component can also be linearised by defining

$$\lambda_m = \frac{3}{128} (\pi\mathcal{M}f_0)^{-\frac{5}{3}}. \qquad (38)$$

This is the chirp time, the time it takes the waveform to go from f_0 to coalescence. Hence, we can re-express $\psi(f)$ as

$$\psi(f) = 2\pi f t_c + \phi_c - \frac{\pi}{4} + \lambda_m \left(\frac{f}{f_0} \right)^{-\frac{5}{3}} (1 + \dots). \qquad (39)$$

Using this waveform and the noise covariance matrices just found (and plotted in Figure 4, the covariance matrix of the parameters can be calculated using Equation (20). The fractional difference between the covariance matrices of these two models and the stationary Gaussian covariance matrix are shown in Tables I and II. Note that the cross-terms of $\Delta\mathcal{A}$ and $\Delta\phi$ will always be 0 as the parameters are complex counterparts. We also reiterate that it doesn't make sense to compare the results of Model A directly with the results of Model B because they are very different models of noise to each other.

The fractional differences in each value of the covariance matrix in Table II are fairly consistent, and relatively small. As the non-stationarity extends over the entire sample, that means the extra power is also spread over the sample, and so does not greatly compromise parameter estimation results for a short signal.

Conversely, Model A was chosen to have the non-stationarity focused into a single section of the data, centered around the coalescence of the waveform. Bearing in mind the short duration of the waveform being considered, this effectively means we have twice as much stationary noise around the waveform than for the stationary Gaussian model. This is reflected by the fact the values of the covariance matrix in Table I are in general four times as large as for the stationary Gaussian model. This would mean that modelling this data as stationary Gaussian would produce much less confident posteriors than if the correct likelihood form was used.

One final point to note with Tables I and II is the presence of several values of the covariance for which the

TABLE I: Fractional difference between the covariance matrix of Gaussian noise with an 8-second Tukey window, and stationary Gaussian noise.

	$\Delta\mathcal{A}$	$\Delta\phi$	$\Delta\lambda_m$	Δt_c
$\Delta\mathcal{A}$	4.008	0.0	4.008	4.007
$\Delta\phi$	0.0	4.008	-0.5674	-4.401
$\Delta\lambda_m$	4.008	-0.5674	4.007	4.007
Δt_c	4.007	-4.401	4.007	4.001

fractional difference is negative. This does not mean that the non-stationary noise produces a more accurate parameter in these cases. Recall that the covariance matrix predicts how well we can measure a parameter, and that the posterior is only correctly found when the likelihood form matches the data. Equation (18) tells us that the posterior will still be centered around the same point, so any mismodelling will only come from the width of the posterior. That means that the best estimate we will get of the parameters will come from stationary Gaussian data, and the non-stationary models will mean the uncertainties of the parameters could be overestimated or underestimated. The negative values correspond to an underestimate of the uncertainties.

We have shown that modelling non-stationary data as stationary will always affect the confidence with which parameters are estimated. However, in cases where the non-stationary extends over the entire noise sample but is relatively small in amplitude, the extent to which short BBH-like signals are under- or overestimated is minimal. The same cannot be said for a loud spike of non-stationarity as in Model A. The limit at which a model can no longer be approximated as stationary is a topic of future research interest. We also reiterate that these results were generated as the average of 10,000 realisations of noise; when considering real interferometer data, we would only be able to use one sample, and so would expect the non-stationary results to be even worse, as the off-diagonal terms of the noise covariance will not average out.

D. The Effect of Non-Stationarity on the Covariance Matrix of a Long Gravitational-Wave Signal

The majority of signals detectors currently observe are only a few seconds long. This means that extended periods of non-stationarity should not cause a significant problem. However, it will become increasingly important to handle the effect of non-stationarity when detectors such as the Einstein Telescope begin observing. These are expected to detect much longer signals [67], which put stringent requirements on stationarity. The analysis problems inherent to the Einstein Telescope are well documented (for example, [67–69]).

We look now at how a longer signal might be affected

by the non-stationarity by comparing the covariance matrices seen in Tables I and II, but now considering a signal modelled after GW190814 [70]. This model was chosen due to the large mass difference and relatively small masses resulting in a merger of duration of tens of seconds. The effect of the mismodelling would be even more significant in an even longer signal, like we would see from a BNS-like object. However, we cannot inject a BNS-like signal into 40 seconds of data, as the signal is of a similar duration, and we would experience wraparound effects in the frequency-domain that would in turn distort the results. On the other hand, analysing data longer than 40 seconds would be both a challenge computationally, and make the results more difficult to compare to the shorter signal.

Just as we saw in Section IV C, both the short period of non-stationarity in Model A and the extended period of non-stationarity in Model B have a detrimental effect on the estimated parameters of the waveform. An interesting comparison would be between Tables I and III, and II and IV. This would allow us to compare how much more greatly affected a long signal would be affected by non-stationarity than a shorter signal.

For Model B, the fractional difference in covariances for GW190814 are approximately the same as for the GW150914-like signal, with the greatest difference being the increase in the magnitude of the cross-terms (such as for $\Delta\mathcal{A}\Delta\lambda_m$). For Model A, the majority of the covariances are slightly smaller, but still the same order of magnitude. The greatest discrepancy comes from the underestimate of the $\Delta\phi\Delta t_c$ term. This arises from the length increase of the signal, which now coalesces in the centre of the Tukey window, but the inspiral starts before the window, transitioning between two radically different forms of the noise. We expect to see even greater discrepancies with the stationary model when we consistently observe longer signals, a problem that has already been identified for LISA [58], for which antenna repointing will create data gaps during the observation of signals. The assumption of stationarity in these periods will lead to parameter estimation biases.

TABLE II: Fractional difference between the covariance matrix of Gaussian noise with an oscillating amplitude of 3.2 Hz, and stationary Gaussian noise.

	$\Delta\mathcal{A}$	$\Delta\phi$	$\Delta\lambda_m$	Δt_c
$\Delta\mathcal{A}$	1.32	0.0	1.299	1.35
$\Delta\phi$	0.0	1.32	1.147	1.018
$\Delta\lambda_m$	1.299	1.147	1.27	1.342
Δt_c	1.35	1.018	1.342	1.332

TABLE III: Fractional difference between the covariance matrix of Gaussian noise with an 8-second Tukey window, and stationary Gaussian noise for a GW190814-like object.

	$\Delta\mathcal{A}$	$\Delta\phi$	$\Delta\lambda_m$	Δt_c
$\Delta\mathcal{A}$	3.362	0.0	0.7657	4.073
$\Delta\phi$	0.0	3.362	-5.171	-21.54
$\Delta\lambda_m$	0.7657	-5.171	1.956	3.907
Δt_c	4.073	-21.54	3.907	4.001

E. Variation of the Fisher Matrix Over Time

Perhaps one of the most surprising things seen when comparing the results in Section IV C and IV D is the decrease in magnitude of the covariance matrix values as the waveform duration increased. This is because we calculate the quantity $\omega^\dagger S^{-1}\omega$, a Fisher matrix, and must invert it to find the covariance matrix. In the process of inverting the matrix, terms of the matrix are affected by the mass components, which are larger for larger values of $\mathcal{M}$. If we were to instead look at the Fisher matrix, we would not see the influence of these mass components on the other values. As a corollary, by comparing Fisher matrices, while we could evaluate the difference in measurability of the different parameters for the non-stationary noise models compared to the stationary noise model, it would not help us understand how greatly under- or overestimated the parameters might be, and an understanding of both is useful.

Because of the non-stationarity of the data, we do expect to see the exact values of the Fisher matrix vary with t_c. To explore what extent the values vary with time, we have calculated the Fisher matrix at different values of t_c for three different waveforms of decreasing total mass. The corresponding values of one parameter of these Fisher matrices, the amplitude-amplitude term $\mathcal{F}_{\mathcal{A}\mathcal{A}}$, has been plotted in Figure 5.

In the case where the noise is stationary Gaussian, the value of $\mathcal{F}_{\mathcal{A}\mathcal{A}}$ is roughly constant regardless of the waveform, as we would expect, with the small fluctuations we do see arising from the fact we could only average over 10,000 realisations of noise, rather than infinite.

At times before and after the Tukey window is applied to Model A ($16 < t_c < 24s$), the values of $\mathcal{F}_{\mathcal{A}\mathcal{A}}$ are comparable to the stationary Gaussian noise model, because it simply is stationary Gaussian noise at these times. In the period when the Tukey window is applied, the values of $\mathcal{F}_{\mathcal{A}\mathcal{A}}$ are still roughly constant for all waveforms, although the values are much smaller. This is consistent with the assertion that the increased level of noise in the Tukey window means that the parameters are less measurable. We also note that the line corresponding to Merger 3 (between masses $m_1 = 9.25 M_\odot$ and $m_2 = 8.0 M_\odot$) has a greater offset than for the other two mergers. The longer waveform of Merger 3 averages

TABLE IV: Fractional difference between the covariance matrix of Gaussian noise with an oscillating amplitude of 3.2 Hz, and stationary Gaussian noise for a GW190814-like object.
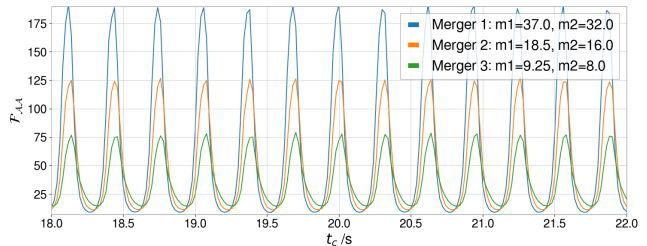
	$\Delta\mathcal{A}$	$\Delta\phi$	$\Delta\lambda_m$	Δt_c
$\Delta\mathcal{A}$	1.234	0.0	1.965	0.8001
$\Delta\phi$	0.0	1.234	-1.838	1.885
$\Delta\lambda_m$	1.965	-1.838	1.219	-0.1868
Δt_c	0.8001	1.885	-0.1868	0.3767
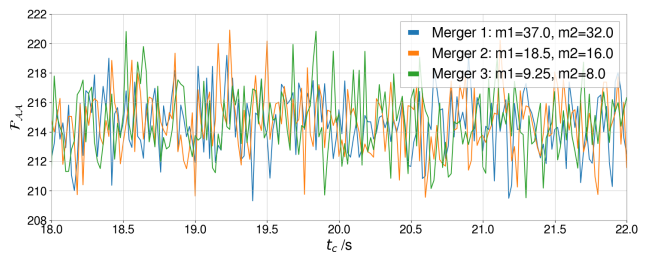
(a) Values of $\mathcal{F}_{\mathcal{A}\mathcal{A}}$ at times t_c for three waveforms injected into Gaussian data.
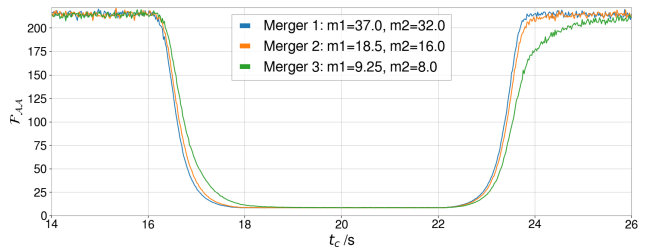
(b) Values of $\mathcal{F}_{\mathcal{A}\mathcal{A}}$ at times t_c for three waveforms injected into Gaussian data with an 8-second Tukey window applied at 16 to 24 seconds.

(c) Values of $\mathcal{F}_{\mathcal{A}\mathcal{A}}$ at times t_c for three waveforms injected into Gaussian data with an oscillating amplitude of 3.2 Hz.

FIG. 5: Values of $\mathcal{F}_{\mathcal{A}\mathcal{A}}$ for different values of coalescence time t_c for three waveforms where only their total masses are varied. Note that masses are in terms of solar mass. The t_c range for c is displayed from 4 to 16s, to better showcase the effect of the entire Tukey window.

over more time, and so the Fisher matrix values shift in time.

The sinusoidal modulation is visible in the results for Model B. We can quite clearly see here that the parameters become more measurable (the values for $\mathcal{F}_{\mathcal{A}\mathcal{A}}$ spikes) whenever the oscillating amplitude is at a minimum. Again, we can see a difference caused by the duration of the signals; in this case, the very short duration of Merger 1 means that the signal is either entirely in or outside the spikes of non-stationarity, but the longer signals will extend outside of the spikes, and so the values of the Fisher matrix will average out in time. While not as tightly spaced as for the stationary Gaussian model, all the values are in the same order of magnitude, and so are still comparable to each other.

The non-stationarity described here is far more extreme than we would expect to see in real experiments, chosen instead to demonstrate the effect on the waveform parameters. Even so, the comparable measurability of the waveforms shows that the estimation of the parameters depends very strongly on when the merger takes place relative to the noise, more than the properties of the waveform itself.

V. CONCLUSION AND FUTURE WORK

We have undertaken one of the first analytical studies of the effect of mismodelling gravitational-wave strain data as stationary in parameter estimation. We established that mismodelling the data in this way will not bias the parameters, but will affect the width of the posterior. This means that there will be an under- or overestimate of the parameter credible intervals.

We demonstrated a methodology for investigating the effect of non-stationarity through calculation of noise covariance Σ, and the covariance of the waveform parameters. Using this methodology, we showed any form of non-stationarity will affect the results of parameter estimation, but the credible intervals are more greatly misrepresented when the non-stationarity increases the amplitude of the noise, rather than the form of the non-stationarity itself; Reference [71] described similar results for given SNR. The effect non-stationarity has on parameter estimation is more pronounced in the case when the signal itself is longer and extends over several forms of the noise, as parameters become harder to measure, as signified by the decrease in magnitude of the Fisher matrix values. This indicates that non-stationarity is a particular obstacle in reliably estimating the parameters of a BNS. Additionally, if we continue to assume the noise is Gaussian, all we need to know about the noise is the matrix Σ, since this square matrix completely characterises the data, with any non-stationarity manifesting in off-diagonal terms.

Now that a method for determining the covariance of stationary and non-stationary Gaussian noise has been determined, we believe there is merit in creating a measure of the effect of non-stationarity by comparing covariance matrices. We propose such a metric would concern deviations from the expected stationary Gaussian covariance we have obtained from the method above, with the main aim of establishing the point at which non-stationarity becomes a problem.

We believe it would be possible to predict the extent to which the posterior will be over or underestimated, and so a formalism to account for this without significantly increasing computational cost should be possible to include in parameter estimation codes. Until a method for handling the non-stationarity is developed, the incorrect credible intervals could also be a limiting factor of astronomical observations using gravitational waves, not least of which being an estimation of the Hubble parameter. For example, Reference [72] proposes that gravitational-wave events can be used to obtain a measurement of H_0 with an uncertainty of only a few percent; the methodology outlined in our paper could be applied to ensure that the confidence intervals are not being underestimated. We would also be concerned that incorporating the uncertainty of the PSD-measurement as suggested by [32] would only serve to compromise the H_0 measurement. Our methodology could also be particularly important when detectors such as the Einstein Telescope begin observing, as these are expected to detect much longer signals [67], which put stringent requirements on stationarity.

One of the major issues we foresee with this work is the calculation of Σ. Therefore, it would be useful to delve into methods for estimating $\langle nn^\dagger \rangle$ on average to accurately find Σ in realistic data.

Another avenue to consider is that real noise is not only non-stationary, it is also non-Gaussian, but modelled as stationary Gaussian for the same computational efficiency reasons as mentioned in III A. Although the effect that the non-stationarity has on the posterior is being examined through the method in this paper and any followup work, the same method cannot be used for non-Gaussian data owing to the different form the likelihood takes. Instead, we see the need for future work to develop suitable techniques to measure the effect on parameter estimation of mismodelling non-Gaussian data.

From this work, it would be possible to determine how non-ideal a dataset can be without compromising parameter estimation, create a robust measure for the extent to which these datasets are affected, and investigate whether current inference codes can be modified to account for the effect without greatly increasing computational cost.

[1] J. Aasi *et al.*, Advanced ligo, Classical and Quantum Gravity **32**, 074001 (2015), 1411.4547 [gr-qc].

[2] B. P. Abbott *et al.* (The LIGO Scientific Col-

laboration and Virgo Collaboration), Gw170814: A three-detector observation of gravitational waves from a binary black hole coalescence, Physical Review Letters **119**, 10.1103/physrevlett.119.141101 (2017), arXiv:1709.09660 [gr-qc].

[3] T. Akutsu *et al.*, Overview of kagra: Calibration, detector characterization, physical environmental monitors, and the geophysics interferometer (2020), arXiv:2009.09305 [gr-qc].

[4] S. Babak *et al.*, Searching for gravitational waves from binary coalescence, Physical Review D **87**, 024033 (2013), arXiv:1208.3491 [gr-qc].

[5] J. B. Camp and N. J. Cornish, Gravitational wave astronomy, Annual Review of Nuclear and Particle Science **54**, 525 (2004).

[6] J. Veitch *et al.*, Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library, Physical Review D **91**, 10.1103/physrevd.91.042003 (2014), arXiv:1409.7215 [gr-qc].

[7] B. P. Abbott *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), Properties of the binary black hole merger gw150914, Physical Review Letters **116**, 10.1103/physrevlett.116.241102 (2016), arXiv:1602.03840 [gr-qc].

[8] M. van der Sluys *et al.*, Parameter estimation of spinning binary inspirals using markov chain monte carlo, Classical and Quantum Gravity **25**, 184011 (2008), arXiv:0805.1689 [gr-qc].

[9] B. Zackay, L. Dai, and T. Venumadhav, Relative binning and fast likelihood evaluation for gravitational wave parameter estimation (2018), arXiv:1806.08792 [astro-ph.IM].

[10] R. Smith *et al.*, Fast and accurate inference on gravitational waves from precessing compact binaries, Physical Review D **94**, 10.1103/physrevd.94.044031 (2016), arXiv:1604.08253 [gr-qc].

[11] H. Qi and V. Raymond, Pyroq: a python-based reduced order quadrature building code for fast gravitational wave inference (2020), arXiv:2009.13812 [gr-qc].

[12] S. Morisaki and V. Raymond, Rapid parameter estimation of gravitational waves from binary neutron star coalescence using focused reduced order quadrature, Physical Review D **102**, 10.1103/physrevd.102.104020 (2020), arXiv:2007.09108 [gr-qc].

[13] N. J. Cornish, Rapid and robust parameter inference for binary mergers (2021), arXiv:2101.01188 [gr-qc].

[14] G. Ashton *et al.*, Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy, The Astrophysical Journal Supplement Series **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[15] I. M. Romero-Shaw *et al.*, Bayesian inference for compact binary coalescences with bilby: Validation and application to the first ligo–virgo gravitational-wave transient catalogue, Monthly Notices of the Royal Astronomical Society **499**, 3295 (2020), arXiv:2006.00714 [astro-ph.IM].

[16] C. M. Biwer *et al.*, Pycbc inference: A python-based parameter estimation toolkit for compact binary coalescence signals, Publications of the Astronomical Society of the Pacific **131**, 024503 (2019), arXiv:1807.10312 [astro-ph.IM].

[17] J. Aasi *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network, Physical Review D **88**, 10.1103/physrevd.88.062001 (2013), arXiv:1304.1775 [gr-qc].

[18] P. Jaranowski and A. Królak, Gravitational-wave data analysis. formalism and sample applications: The gaussian case, Living Reviews in Relativity **8** (2005), arXiv:0711.1115 [gr-qc].

[19] E. Poisson and C. M. Will, Gravitational waves from inspiralling compact binaries: Parameter estimation using second-post-newtonian waveforms, Physical Review D **52**, 848 (1995), arXiv:gr-qc/9502040 [gr-qc].

[20] N. Perraudin and P. Vandergheynst, Stationary signal processing on graphs, IEEE Transactions on Signal Processing **65**, 3462 (2017).

[21] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, Physical Review D **91**, 10.1103/physrevd.91.084034 (2015), arXiv:1410.3852 [gr-qc].

[22] M. van der Sluys *et al.*, Gravitational-wave astronomy with inspiral signals of spinning compact-object binaries, The Astrophysical Journal **688**, L61 (2007), arXiv:0710.1897 [astro-ph].

[23] L. Blackburn *et al.*, The lsc glitch group: Monitoring noise transients during the fifth ligo science run, Classical and Quantum Gravity **25**, 184004 (2008), arXiv:0804.0800 [gr-qc].

[24] B. P. Abbott *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), Characterization of transient noise in advanced ligo relevant to gravitational wave signal gw150914, Classical and Quantum Gravity **33**, 134001 (2016), arXiv:1602.03844 [gr-qc].

[25] R. Abbott *et al.* (The LIGO Scientific Collaboration and the Virgo Collaboration), Open data from the first and second observing runs of advanced ligo and advanced virgo (2019), arXiv:1912.11716 [gr-qc].

[26] B. P. Abbott *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), A guide to ligo-virgo detector noise and extraction of transient gravitational-wave signals, Classical and Quantum Gravity **37**, 055002 (2020), arXiv:1908.11170 [gr-qc].

[27] K. Chatziioannou *et al.*, Noise spectral estimation methods and their impact on gravitational wave measurement of compact binary mergers, Physical Review D **100**, 10.1103/physrevd.100.104004 (2019), arXiv:1907.06540 [gr-qc].

[28] N. J. Cornish, Time-frequency analysis of gravitational wave data, Physical Review D **102**, 124038 (2020), arXiv:2009.00043 [gr-qc].

[29] N. J. Cornish *et al.*, Bayeswave analysis pipeline in the era of gravitational wave observations, Physical Review D **103**, 10.1103/physrevd.103.044006 (2021), arXiv:2011.09494 [gr-qc].

[30] K. Chatziioannou *et al.*, Modeling compact binary signals and instrumental glitches in gravitational wave data, Physical Review D **103**, 10.1103/physrevd.103.044013 (2021), arXiv:2101.01200 [gr-qc].

[31] S. Mozzon *et al.*, Dynamic normalization for compact binary coalescence searches in non-stationary noise, Classical and Quantum Gravity **37**, 215014 (2020), arXiv:2002.09407 [astro-ph.IM].

[32] S. Biscoveanu *et al.*, Quantifying the effect of power spectral density uncertainty on gravitational-wave parameter estimation for compact binary sources, Physi-

cal Review D **102**, 10.1103/physrevd.102.023008 (2020), arXiv:2004.05149 [astro-ph.HE].

[33] C. Röver, Student-t based filter for robust signal detection, Physical Review D **84**, 10.1103/physrevd.84.122004 (2011), arXiv:1109.0442 [physics.data-an].

[34] L. A. Wainstein and V. D. Zubakov, *Extraction of Signals from Noise* (Prentice-Hall International, 1962).

[35] R. Trotta, Bayes in the sky: Bayesian inference and model selection in cosmology, Contemporary Physics **49**, 71 (2008), arXiv:0803.4089 [astro-ph].

[36] C. P. Robert and J. Rousseau, On bayesian data analysis (2010), arXiv:1001.4656 [stat.ME].

[37] S. Vitale *et al.*, Impact of bayesian priors on the characterization of binary black hole coalescences, Physical Review Letters **119**, 10.1103/physrevlett.119.251103 (2017), arXiv:1707.04637 [gr-qc].

[38] A. Gelman and X.-L. Meng, Simulating normalizing constants: From importance sampling to bridge sampling to path sampling, Statistical Science **13**, 163 (1998).

[39] D. W. Hogg and D. Foreman-Mackey, Data analysis recipes: Using markov chain monte carlo, The Astrophysical Journal Supplement Series **236**, 11 (2018), arXiv:1710.06068 [astro-ph.IM].

[40] N. Christensen and R. Meyer, Using markov chain monte carlo methods for estimating parameters with gravitational radiation data, Physical Review D **64**, 10.1103/physrevd.64.022001 (2001), arXiv:gr-qc/0102018 [gr-qc].

[41] J. S. Speagle, dynesty: a dynamic nested sampling package for estimating bayesian posteriors and evidences, Monthly Notices of the Royal Astronomical Society **493**, 3132 (2020), arXiv:1904.02180 [astro-ph.IM].

[42] J. D. Scargle, Degeneracy of arma time series models and the arrow of time (2016).

[43] G. Vajente *et al.*, Machine-learning nonstationary noise out of gravitational-wave detectors, Physical Review D **101**, 10.1103/physrevd.101.042003 (2020), arXiv:1911.09083 [gr-qc].

[44] H. Cramér, On some classes of nonstationary stochastic processes, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory* (University of California Press, Berkeley, California, 1961) pp. 57–78.

[45] A. Hebbal *et al.*, Bayesian optimization using deep gaussian processes (2019), arXiv:1905.03350 [stat.ML].

[46] E. W. Kamen and J. A. Sills, The frequency response function of a linear time-varying system, IFAC Proceedings Volumes **26**, 315 (1993), 12th Triennal Wold Congress of the International Federation of Automatic control. Volume 1 Theory, Sydney, Australia, 18-23 July.

[47] R. M. Loynes, On the concept of the spectrum for nonstationary processes, Journal of the Royal Statistical Society. Series B (Methodological) **30**, 1 (1968).

[48] N. Huang *et al.*, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **454**, 903 (1998).

[49] L. Sun *et al.*, Characterization of systematic error in advanced ligo calibration, Classical and Quantum Gravity **37**, 225008 (2020), arXiv:2005.02531 [astro-ph.IM].

[50] S. Vitale *et al.*, physical: A physical approach to the marginalization of ligo calibration uncertainties (2020), arXiv:2009.10192 [gr-qc].

[51] E. Payne *et al.*, Gravitational-wave astronomy with a physical calibration model, Physical Review D **102**, 10.1103/physrevd.102.122004 (2020), arXiv:2009.10193 [astro-ph.IM].

[52] B. P. Abbott *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), Gwtc-1: A gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs, Physical Review X **9**, 10.1103/physrevx.9.031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[53] R. Abbott *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), The o2 data release (2019).

[54] R. Abbott *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), Gwtc-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run, arXiv:2010.14527 [gr-qc] (2020).

[55] C. Talbot and E. Thrane, Gravitational-wave astronomy with an uncertain noise power spectral density, Physical Review Research **2**, 043298 (2020), arXiv:2006.05292 [astro-ph.IM].

[56] B. Zackay *et al.*, Detecting gravitational waves in data with non-gaussian noise (2019), arXiv:1908.05644 [astro-ph.IM].

[57] T. Venumadhav *et al.*, New search pipeline for compact binary mergers: Results for binary black holes in the first observing run of advanced ligo, Physical Review D **100**, 10.1103/physrevd.100.023011 (2019), arXiv:1902.10341 [astro-ph.IM].

[58] M. C. Edwards *et al.*, Identifying and addressing nonstationary lisa noise, Physical Review D **102**, 10.1103/physrevd.102.084062 (2020), arXiv:2004.07515 [gr-qc].

[59] Y. Huang *et al.*, Source properties of the lowest signal-to-noise-ratio binary black hole detections, Physical Review D **102**, 10.1103/physrevd.102.103024 (2020), arXiv:2003.04513 [gr-qc].

[60] M. Vallisneri, Use and abuse of the fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects, Physical Review D **77**, 042001 (2008).

[61] M. Vallisneri, Beyond the fisher-matrix formalism: Exact sampling distributions of the maximum-likelihood estimator in gravitational-wave parameter estimation, Physical Review Letters **107**, 10.1103/physrevlett.107.191104 (2011), arXiv:1108.1158 [gr-qc].

[62] B. S. Sathyaprakash and B. F. Schutz, Physics, astrophysics and cosmology with gravitational waves, Living Reviews in Relativity **12**, 10.12942/lrr-2009-2 (2009), arXiv:0903.0338 [gr-qc].

[63] D. Coe, Fisher matrices and confidence ellipses: A quick-start guide and software (2009), arXiv:0906.4123 [astro-ph.IM].

[64] A. Biswas, J. McIver, and A. Mahabal, New methods to assess and improve ligo detector duty cycle, Classical and Quantum Gravity **37**, 175008 (2020), arXiv:1910.12143 [astro-ph.IM].

[65] S. Chatterji *et al.*, Multiresolution techniques for the detection of gravitational-wave bursts, Classical and Quantum Gravity **21**, S1809 (2004), arXiv:gr-qc/0412119 [gr-qc].

[66] B. P. Abbott *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), Observation of gravitational waves from a binary black hole merger, Physical Review Letters **116**, 10.1103/physrevlett.116.061102 (2016), arXiv:1602.03837 [gr-qc].

[67] T. Regimbau *et al.*, Mock data challenge for the einstein gravitational-wave telescope, Physical Review D **86**, 10.1103/physrevd.86.122001 (2012), arXiv:1201.3563 [gr-qc].

[68] L. Bosi and E. K. Porter, Data analysis challenges for the einstein telescope, General Relativity and Gravitation **43**, 519 (2009), arXiv:0910.0380 [gr-qc].

[69] M. Punturo *et al.*, The einstein telescope: A third-generation gravitational wave observatory, Classical and Quantum Gravity **27**, 194002 (2010).

[70] R. Abbott *et al.* (The LIGO Scientific Collaboration and Virgo Collaboration), Gw190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object, The Astrophysical Journal **896**, L44 (2020), arXiv:2006.12611 [astro-ph.HE].

[71] C. P. L. Berry *et al.*, Parameter estimation for binary neutron-star coalescences with realistic noise during the advanced ligo era, The Astrophysical Journal **804**, 114 (2015), arXiv:1411.6934 [astro-ph.HE].

[72] W. D. Pozzo, Measuring the hubble constant using gravitational waves, Journal of Physics: Conference Series **484**, 012030 (2014).