

# Bayesian framework to follow-up continuous gravitational wave candidates from deep surveys

J. Martins<sup>1,2,\*</sup>, M. A. Papa<sup>1,2</sup>, B. Steltner<sup>1,2</sup>, R. Prix<sup>1,2</sup>, and P. B. Covas<sup>1,2,3</sup>

<sup>1</sup>Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-30167 Hannover, Germany

<sup>2</sup>Leibniz Universität Hannover, D-30167 Hannover, Germany

<sup>3</sup>Departament de Física, Universitat de les Illes Balears, IAC3—IEEC, Carretera Valldemossa km 7.5, E-07122 Palma, Spain



(Received 7 September 2025; accepted 10 December 2025; published 22 January 2026)

Broad all-sky searches for continuous gravitational waves have high computational costs and require hierarchical pipelines. The sensitivity of these approaches is set by the initial search and by the number of candidates from that stage that can be followed up. The current follow-up schemes for the deepest surveys require careful tuning and setup, and have a significant human-labor cost and this impacts the number of follow-ups that can be afforded. Here we present and demonstrate a new follow-up framework based on Bayesian parameter estimation for the rapid, highly automated follow-up of candidates produced by the early stages of deep, wide-parameter space searches for continuous waves.

DOI: [10.1103/rbyp-157m](https://doi.org/10.1103/rbyp-157m)

## I. INTRODUCTION

Gravitational waves emitted during the merger of two compact objects are now routinely detected in data gathered by the LIGO-Virgo-KAGRA Collaborations [1,2].

Among the gravitational wave signals that remain elusive are continuous gravitational waves emitted by fast-rotating neutron stars with a sustained quadrupole moment [3,4], as well as other more exotic scenarios [5,6].

The long duration of continuous waves renders signals detectable that are at any instant of time below the detector noise, and makes it possible to resolve different waveforms with great accuracy. For this reason, all-sky searches over a broad frequency range using coherent detection methods are computationally not viable. This has long been known [7–12] and hierarchical semicoherent searches have proven to be the most sensitive solution to the problem [13–17].

The initial search in the hierarchy examines the full parameter space with an incoherent combination of the results of coherent searches over short data segments, spanning less than one hour to several days, depending on the search. The shorter coherence times reduce the required number of template waveforms and, hence, the computing cost, at the cost of a reduced sensitivity. Low

detection thresholds can be chosen to ease this effect, but result in a high number of false alarms. These are weeded out with several follow-up stages that employ increasingly higher coherence times, finer resolutions, and more stringent detection thresholds and vetoes. The goal is to reject noise while ensuring that detectable gravitational-wave signals survive.

Each follow-up stage of a hierarchical search requires (i) the identification of signal candidates, typically as local maxima of a detection statistic or as the result of some clustering procedure; (ii) the definition of thresholds and vetoes that reject candidates; (iii) the determination of per-candidate parameter space regions around each candidate that are expected to contain an underlying signal; and (iv) the definition of a setup of the next search stage.

The first hierarchical search with a significant follow-up effort was carried out in [18] on the results of an Einstein@Home search. The latest generation Einstein@Home searches are characterized by long coherence-times—possibly the longest among broad searches—and the results are investigated with the largest follow-up campaigns involving millions of candidates [13,19]. This is about an order of magnitude more candidates than examined with other approaches: for example [14] followed up between  $10^3$  and  $10^5$  candidates, depending on the pipeline. The long coherence time and the low follow-up thresholds pose unique challenges to the follow-up schemes.

All the initial Einstein@Home search stages [13,19], as well as all initial stages in [14], and all but one of the [14] follow-ups are completely deterministic: The signal-parameter uncertainty region around each candidate at stage  $n$  is completely covered with a template bank and

\*Contact author: [jasper.martins@aei.mpg.de](mailto:jasper.martins@aei.mpg.de)

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

systematically investigated at stage  $n + 1$ . This approach is warranted by two factors: (1) It is easiest to simply redo on a smaller parameter space what one has already done on a larger one; and (2) in the first follow-up stages the putative signal of the long coherence-time large-scale follow-ups is still weak enough and/or the uncertainty in signal parameters is large enough that stochastic methods are thought to not perform sufficiently well.

On the other hand stochastic search methods have the potential to be more efficient than deterministic template-bank based methods, as they concentrate compute cycles on the interesting regions of the parameter space. This becomes crucial when inspecting high dimensional spaces, like those pertaining to signals from unknown neutron stars with a companion. The more efficient allocation of compute cycles also allows for a more precise parameter estimation of putative signals [20].

Another advantage of stochastic search methods is that there already exist tools to integrate Monte Carlo sampling techniques in Bayesian parameter estimation frameworks, which would provide a natural way to estimate the posterior probability density function for any parameter after a search. Posterior parameter distributions would solve a short-coming of the current methods that evaluate the uncertainty regions conservatively from the distributions of distances between recovered signal parameters and the actual parameters of fake signals from the target population. This approach results in search-region extents that are the same for every candidate and are overestimated for the majority of them. Conversely, the posterior can be calculated specifically for every candidate and inform the search-sampling in the next stage.

Nondeterministic follow-ups of continuous wave candidates have been demonstrated for short coherence times (less than a day) [14,21–23] and for long coherence times (several days) for signals louder than those targeted by the most sensitive all-sky searches [20]. The regime of the weakest signals searched with long coherence times [13,15] remains unexplored. Here, we present a new Bayesian framework for hierarchical searches that establishes non-deterministic follow-up capabilities in that regime, in a highly automated manner. We demonstrate our framework on real data, following up the Einstein@Home search candidates of [13], and find results consistent with the findings of the previous deterministic search.

Sections I–III motivate the development of the new framework, present the signal model, discuss the background of Bayesian inference, and provide the specific context of Einstein@Home searches.

In Sec. IV, we present the results of applying the framework to the Einstein@Home [13]. We assess the stage of the follow-up at which our framework becomes applicable. We follow up all candidates that remained at that stage in the original search, compare the required computational and human effort, and compare the final results.

Section V concludes the paper with a discussion of our findings, and an outlook on future research.

## II. SIGNAL MODEL AND $\mathcal{F}$ -STATISTIC

We consider a continuous wave signal that at the detector takes the form [24],

$$h(t) = F_+(t)h_+(t) + F_\times(t)h_\times(t). \quad (1)$$

$F_+(t)$  and  $F_\times(t)$  are the detector beam-pattern functions for the two gravitational wave polarizations “+” and “ $\times$ .” They depend on the sky position of the source (defined by the ecliptic longitude  $\lambda$  and ecliptic latitude  $\beta$ ), and the orientation  $\psi$  of the wave-frame with respect to the detector-frame. The waveforms for the two polarizations,  $h_+(t)$  and  $h_\times(t)$ , are

$$\begin{aligned} h_+(t) &= A_+ \cos \phi(t), \\ h_\times(t) &= A_\times \sin \phi(t), \end{aligned} \quad (2)$$

with  $\phi(t)$  being the phase of the gravitational-wave signal at the time  $t$  and  $A_{+,\times}$  the polarizations’ amplitudes:

$$\begin{aligned} A_+ &= \frac{1}{2} h_0 (1 + \cos^2 \iota), \\ A_\times &= h_0 \cos \iota. \end{aligned} \quad (3)$$

Here  $h_0$  is the intrinsic gravitational wave amplitude and  $\iota$  is the angle between the total angular momentum of the star and the line of sight.

The phase  $\Phi(\tau)$  of the signal for an observer at rest with respect to the source (for an isolated source this is typically the solar system barycenter) is,

$$\Phi(\tau) = \Phi_0 + 2\pi \left[ f(\tau - \tau_0) + \frac{1}{2} f(\tau - \tau_0)^2 + \dots \right], \quad (4)$$

where  $f$  is the signal frequency and  $\tau$  is the arrival time of the gravitational wave front at the solar system barycenter and  $\tau_0$  is a reference time. The detector-frame phase  $\phi(t)$  incorporates the effects due to the relative motion between the source and the detector, as well as relativistic effects.  $\phi(t) = \Phi(\tau(t))$ , with the transformation  $\tau(t)$  depending on the position of the source.

We will consider an isolated source with a signal model referred to as the IT1-model [25], i.e., a signal emitted by an isolated neutron with only one spindown. This model is described by two sets of parameters  $\omega := (\vartheta, A)$  that define the

$$\begin{aligned} \text{phase-evolution: } \vartheta &= (\lambda, \beta, f, \dot{f}), \text{ and} \\ \text{amplitude: } A &= (h_0, \cos \iota, \psi, \Phi_0) \end{aligned} \quad (5)$$

in the detector frame.

The optimal Neyman-Pearson detection statistic for a known signal is the likelihood ratio  $\Lambda$ ,

$$\Lambda(d, w) = \frac{f_{D|\text{IT1}}(D = d|\omega)}{f_{D|\text{IT1}}(D = d|0)}, \quad (6)$$

where  $f_{D|\text{IT1}}(D = d|\cdot)$  is the probability density of the data  $d$  under a Gaussian noise hypothesis and the IT1 signal model.

By a judicious reparametrization of the signal amplitudes, the signal waveform can be written as a linear combination of four basis functions that depend on the phase-evolution parameters  $\vartheta$  and time, with coefficients  $\mathcal{A}^\mu$  that depend only on the physical amplitude parameters  $A$  [24],

$$h(t; \omega) = \sum_{\mu=1}^4 \mathcal{A}^\mu(A) h_\mu(t; \vartheta). \quad (7)$$

In this form, an analytic maximization of the likelihood ratio can be carried out over the  $\mathcal{A}^\mu$ . The resulting detection statistic is the  $\mathcal{F}$ -statistic [11,24],

$$2\mathcal{F}(\vartheta) := 2 \log \Lambda(d, \mathcal{A}_{\text{MLE}}, \vartheta), \quad (8)$$

where  $\mathcal{A}_{\text{MLE}} = \arg \max_{\mathcal{A}} \Lambda(d, \mathcal{A}, \vartheta)$ , which implicitly depends on the phase-evolution parameters.

In Gaussian noise, the  $2\mathcal{F}$ -statistic is a sum of four Gaussian random variables and, as such, follows a  $\chi^2$  distribution with four degrees of freedom and noncentrality parameter  $\rho^2(\omega_S, \vartheta_t)$  that depends on the signal parameters  $\omega_S$  and on the template parameters  $\vartheta_t$ . In the perfectly matched case ( $\vartheta_t = \vartheta_S$ ), the noncentrality parameter is [26]

$$\rho^2(\omega_S, \vartheta_S) = \frac{4}{25} \frac{T_{\text{data}}}{S_n} h_0^2 R(\lambda, \beta, \psi, \cos \iota), \quad (9)$$

where  $T_{\text{data}}$  is the total data duration,  $S_n$  is the single-sided noise power spectral density at the signal frequency, and  $R(\cdot)$  describes the sky-position and polarization dependent detector response. Given a template with parameters  $\vartheta_t = \vartheta_S + \Delta\vartheta$ , the noncentrality parameter is lower than for a perfectly matched template,

$$\rho^2(\omega_S, \vartheta_S + \Delta\vartheta) \leq \rho^2(\omega_S, \vartheta_S). \quad (10)$$

When the distance between the signal and the template is small, the relative loss is well approximated by a quadratic function of the distance,

$$\frac{\rho^2(\omega_S, \vartheta_S) - \rho^2(\omega_S, \vartheta_S + \Delta\vartheta)}{\rho^2(\omega_S, \vartheta_S)} \simeq g_{ij}(\omega_S) \Delta\vartheta^i \Delta\vartheta^j, \quad (11)$$

with

$$g_{ij}(\omega_S) := \frac{-1}{2\rho^2(\omega_S, \vartheta_S)} \left. \frac{\partial^2 \rho^2(\omega_S, \vartheta_t)}{\partial \vartheta_t^i \partial \vartheta_t^j} \right|_{\vartheta_t = \vartheta_S}. \quad (12)$$

Equation (12) gives the so-called  $\mathcal{F}$ -statistic metric [27]. Typically, approximations that neglect the dependency on the amplitude parameters  $\mathcal{A}_S$  are used [27,28]. We use such an approximation, the phase metric, defined as

$$\tilde{g}_{ij}(\vartheta_S) := \left[ \left\langle \frac{\partial \phi(t, \vartheta)}{\partial \vartheta_i} \frac{\partial \phi(t, \vartheta)}{\partial \vartheta_j} \right\rangle_T - \left\langle \frac{\partial \phi(t, \vartheta)}{\partial \vartheta_i} \right\rangle_T \left\langle \frac{\partial \phi(t, \vartheta)}{\partial \vartheta_j} \right\rangle_T \right]_{\vartheta = \vartheta_S}, \quad (13)$$

where  $\langle \cdot \rangle_T$  represents an average over the observing time.

Semicoherent searches split the data into  $N_{\text{seg}}$  segments  $\{d_\ell\}$ . The semicoherent  $\mathcal{F}$ -statistic is the (averaged) joint log-likelihood ratio of signals with the same phase-evolution parameters  $\vartheta$  in all  $N_{\text{seg}}$  data segments, but independently maximized over the amplitude parameters for each segment [11],

$$2\bar{\mathcal{F}}(\vartheta) := \frac{1}{N_{\text{seg}}} \sum_{\ell=1}^{N_{\text{seg}}} 2 \log \Lambda(d_\ell, \mathcal{A}_{\text{MLE}, \ell}, \vartheta). \quad (14)$$

The  $\mathcal{A}_{\text{MLE}, \ell}$  are the maximum likelihood estimators for each segment. We will only use segmentations where all segments have the same length,  $T_{\text{coh}}$ .

In Gaussian noise,  $N_{\text{seg}} 2\bar{\mathcal{F}}$  follows a  $\chi^2$  distribution with  $4N_{\text{seg}}$  degrees of freedom and a noncentrality parameter that is the sum of all  $\rho_\ell^2$ . The semicoherent  $\mathcal{F}$ -statistic phase metric is the weighted average of the metrics for each data segment [29]. We neglect the weighting operation. We use the semicoherent metric in Sec. III D when modeling the posterior probability density function of signal candidates.

### III. BAYESIAN INFERENCE SCHEME FOR CONTINUOUS WAVE CANDIDATE FOLLOW-UP

#### A. Notation

In this paper, we adopt the following notation for random variables. If not specified otherwise, all random variables discussed are continuous. For any continuous random variable  $X$  with probability density function  $f_X$ , we define

$$p(x) := f_X(X = x). \quad (15)$$

Similarly, if  $X$  conditionally depends on the value taken by another continuous random variable  $Y$  and some other information  $I$ , we define

$$p(x|y, I) := f_{X|Y, I}(X = x|Y = y, I). \quad (16)$$

We also refer to  $p(x)$  just as the probability of  $x$ .

#### B. Background

Bayesian statistic allows one to make inferences on the values of the parameters  $\omega$  affecting the outcome of a

measurement  $d$  in the presence of noise. The measuring device and the noise are described in terms of  $p(d|\omega)$ , the probability to measure  $d$  in the presence of a signal with parameter values  $\omega$ —this is the probability of  $d$  *conditioned* upon  $\omega$ . The inference is made by deriving the probability distribution function for  $\omega$  given the measurement  $d$ :  $p(\omega|d)$ , the so-called posterior distribution. Bayes’s theorem tells us how to invert the order of conditioning and derive the posterior  $p(\omega|d)$ , given the probability of  $p(d|\omega)$  and our prior beliefs of the probability of the parameters,  $p(\omega)$ . Here, where both the parameters and the data take continuous values, Bayes’s theorem takes the form [30],

$$p(\omega|d) = \frac{p(d|\omega)p(\omega)}{\int_{\Omega} p(d|\omega)p(\omega)d\omega}, \quad (17)$$

where  $\Omega$  indicates the sample space of the parameters:  $\omega \in \Omega$ .

The denominator in Eq. (17),

$$p(d|\Omega) = \int_{\Omega} p(d|\omega)p(\omega)d\omega, \quad (18)$$

is called the *evidence* and measures the support for the hypothesis of a signal with parameter values  $\omega \in \Omega$ . The ratio of the evidences for the signal to the null-hypothesis of Gaussian noise gives the Bayes’s factor

$$\mathcal{B}_{S/G} = \frac{p(d|\Omega)}{p(d|0)}. \quad (19)$$

In Bayesian statistics different models (or hypotheses) are compared via Bayes’s factors. A measurement that yields a Bayes’s factor with  $\mathcal{B} > 1$  favors the signal hypothesis [30].

Equations (17) and (19) have historically been used to search for continuous waves from pulsars with well-known phase-evolution parameters [31]. In those searches the final product is the posterior distribution for  $h_0$ , marginalized over all other unknown signal parameters. A highly peaked posterior, far from zero, would be the first indication of the detection of a signal.

Broad surveys have typically taken a frequentist approach based on the  $\mathcal{F}$ -statistic. The main advantage of this method is that it maximizes the likelihood ratio over the amplitude parameters  $\mathcal{A}$ , hence does not require an explicit search over them and this improves the computational efficiency. The parameters that remain for a search are the phase-evolution parameters  $\vartheta$ .

In recent years, Bayesian frameworks have also utilized the  $\mathcal{F}$ -statistic likelihood [32], for candidate follow-ups [20,21] and for parameter estimation [33]. A significant advantage of this approach is that the  $\mathcal{F}$ -statistic calculations have been highly optimized and are hence very efficient, with a number of tools publicly available [34,35]. The work presented here is in this vein.

The first stage—often referred to as Stage 0—of a large survey is the most challenging because the whole parameter space is searched. Only the parameter space regions associated with “interesting” results are ear-marked for further inspection. These are the so-called candidates that are followed up with the hierarchy of semicoherent searches. Typically, the candidates are identified by some clustering procedure and detection statistic threshold, and represented by a point in parameter space—the *seed*—and a parameter uncertainty region around it (see, e.g., Refs. [13,36]). On a large scale production run, no stochastic method has yet been demonstrated competitive with this sort of approach [37,38].

We shall hence assume that a follow-up using the new framework begins with a set of given candidates, for which we know what the parameter uncertainties are. We will also assume that the coherence time hierarchy of the various semicoherent searches is given.

For every candidate we run a succession of follow-ups. At every follow-up search we derive the posteriors on the signal parameters and the evidence for the presence of a signal. We use the former to define the search at the next stage. At the end of the follow-ups, we use the evidences to decide whether to discard the candidate or not [39]. The remainder of this section describes these steps in more detail.

### C. Using the $\mathcal{F}$ -statistic for Bayesian inference

Our starting point is a list of candidates and their equal-size phase-evolution parameters uncertainty. For each candidate the follow-up infers the phase-evolution parameters  $\vartheta$  of a putative signal in that range. For every candidate each follow-up stage produces (i) the posterior distribution for the  $\vartheta$  that will be used to define the search prior of the next stage and (ii) the evidence for a signal in the searched region. We now show how we use the  $\mathcal{F}$ -statistic to compute these quantities.

For a candidate, the posterior probability for the phase-evolution parameters  $\vartheta$  is obtained by marginalizing the full posterior over all possible amplitude parameters  $\{\mathcal{A}\}$ ,

$$p(\vartheta|d) = \frac{p(d|\vartheta)p(\vartheta)}{\int_{\mathcal{R}_c} p(d|\vartheta)p(\vartheta)d\vartheta} \quad (20)$$

$$= \frac{\int_{\{\mathcal{A}\}} p(d|\mathcal{A}, \vartheta)p(\mathcal{A}, \vartheta)d\mathcal{A}}{\int_{\mathcal{R}_c} \int_{\{\mathcal{A}\}} p(d|\mathcal{A}, \vartheta)p(\mathcal{A}, \vartheta)d\mathcal{A}d\vartheta} \quad (21)$$

where  $\mathcal{R}_c$  is the phase-evolution uncertainty region for that candidate. For detectable signals we will assume that the only non-negligible contributions of  $p(d|\mathcal{A}, \vartheta)$  to the integral above come from  $\mathcal{A} = \mathcal{A}_{\text{MLE}}$  so we simply set the prior to a delta distribution,

$$p(\vartheta, \mathcal{A}) = p(\mathcal{A}|\vartheta)p(\vartheta) \quad (22)$$

$$= \delta(\mathcal{A} - \mathcal{A}_{\text{MLE}}(\vartheta))p(\vartheta). \quad (23)$$

This corresponds to a profile-likelihood approach where the amplitude parameters are treated as nuisance parameters. Using Eqs. (6), (8), and (21) with this prior we find

$$p(\vartheta|d) \approx \frac{e^{\mathcal{F}(\vartheta)}p(\vartheta)}{\int_{\mathcal{R}_c} e^{\mathcal{F}(\vartheta)}p(\vartheta)d\vartheta}. \quad (24)$$

We will come back to the validity of this assumption in Secs. IV B and IV D.

The posterior for semicoherent searches is obtained as follows. Given data  $d$ , first consider a search for  $N_{\text{seg}}$  signals with parameters  $\omega_\ell$ , each constrained to nonoverlapping coherent data segments  $d_\ell$  that span some  $T_{\text{coh}}$ . Because the noise realizations in the different segments  $d_\ell$  are independent random variables, the probability to obtain the set  $\{d_\ell\}$  for signal parameters  $\omega_0, \dots, \omega_\ell \in \Omega$ , is given by

$$p(\{d_\ell\}|\{\omega_\ell\}) = \prod_{\ell=1}^{N_{\text{seg}}} p(d_\ell|\omega_\ell). \quad (25)$$

Following Bayes's theorem, the joint posterior distribution for the signal parameters  $\{\omega_\ell\}$  given the  $\{d_\ell\}$  then reads

$$p(\{\omega_\ell\}|\{d_\ell\}) = \frac{p(\{\omega_\ell\})\prod_\ell p(d_\ell|\omega_\ell)}{p(\{d_\ell\}|\Omega \times \dots \times \Omega)}. \quad (26)$$

Like before, we treat the  $\mathcal{A}_\ell$  as nuisance parameters and assume that the only non-negligible contribution to the evidence integral for each segment comes from the respective  $\mathcal{A}_{\text{MLE},\ell}(\vartheta_\ell)$ . Thus, we apply Eqs. (21)–(24) for each segment to Eq. (26) and obtain the marginal posterior

$$p(\{\vartheta_\ell\}|\{d_\ell\}) = \frac{p(\{\vartheta_\ell\})\prod_\ell p(d_\ell|\vartheta_\ell)}{p(d|\mathcal{R}_c \times \dots \times \mathcal{R}_c)}. \quad (27)$$

Following Eq. (6),

$$\prod_\ell p(d_\ell|\vartheta_\ell) = \prod_\ell p(d_\ell|0)e^{\mathcal{F}_\ell(\vartheta_\ell)}. \quad (28)$$

$\mathcal{F}_\ell$  is the coherent  $\mathcal{F}$ -statistic for segment  $\ell$ . Note that  $\prod_\ell p(d_\ell|0) = p(d|0)$ .

However, following Eq. (14), semicoherent approaches target signals with the same phase-evolution parameters in all segments. We can enforce this condition by choosing reference phase-evolution parameters  $\vartheta$  and setting priors

$$p(\{\vartheta_\ell\}) \rightarrow p(\{\vartheta_\ell\}|\vartheta) = \prod_\ell \delta(\vartheta_\ell - \vartheta). \quad (29)$$

We then marginalize Eq. (28) over all  $\vartheta_\ell$  with this prior to yield

$$p(\{d_\ell\}|\vartheta) = \int_{\mathcal{R}_c} \dots \int_{\mathcal{R}_c} \prod_\ell p(d_\ell|\vartheta_\ell)p(\vartheta_\ell|\vartheta)d\vartheta_\ell \quad (30)$$

$$= p(d|0)\prod_\ell \int_{\mathcal{R}_c} e^{\mathcal{F}_\ell(\vartheta_\ell)}\delta(\vartheta_\ell - \vartheta)d\vartheta_\ell \quad (31)$$

$$= p(d|0)\prod_\ell e^{\mathcal{F}_\ell(\vartheta)} = p(d|0)e^{N_{\text{seg}}\bar{\mathcal{F}}(\vartheta)}, \quad (32)$$

where  $\bar{\mathcal{F}}$  is the semicoherent  $\mathcal{F}$ -statistic defined in Eq. (14). Again, we find the semicoherent posterior for the reference phase-evolution parameters  $\vartheta$  by applying Bayes's theorem to Eq. (32),

$$p(\vartheta|\{d_\ell\}) = \frac{p(\{d_\ell\}|\vartheta)p(\vartheta)}{\int_{\mathcal{R}_c} p(\{d_\ell\}|\vartheta)p(\vartheta)d\vartheta} \quad (33)$$

$$= \frac{e^{N_{\text{seg}}\bar{\mathcal{F}}(\vartheta)}p(\vartheta)}{\int_{\mathcal{R}_c} e^{N_{\text{seg}}\bar{\mathcal{F}}(\vartheta)}p(\vartheta)d\vartheta}, \quad (34)$$

and rediscover the right side of Eq. (24) with the substitution  $\mathcal{F} \rightarrow N_{\text{seg}}\bar{\mathcal{F}}$ . The phase-evolution parameter independent noise term  $p(d|0)$  factors out.

The semicoherent Bayes's factor for a signal with phase-evolution parameters  $\vartheta_S \in \mathcal{R}_c$  is

$$\mathcal{B}_{S/G} = \frac{p(\{d_\ell\}|\mathcal{R}_c)}{p(\{d_\ell\}|0)} \quad (35)$$

$$= \frac{p(d|0)\int_{\mathcal{R}_c} e^{N_{\text{seg}}\bar{\mathcal{F}}(\vartheta)}p(\vartheta)d\vartheta}{p(d|0)}, \quad (36)$$

which is equal to the semicoherent evidence for a signal with parameters  $\vartheta \in \mathcal{R}_c$  [40],

$$Z(\mathcal{R}_c) = \int_{\mathcal{R}_c} e^{N_{\text{seg}}\bar{\mathcal{F}}(\vartheta)}p(\vartheta)d\vartheta. \quad (37)$$

We note that, in practice, we substitute the semicoherent  $\mathcal{F}$ -statistic with (half of) the estimated signal power

$$\hat{\rho}^2(\vartheta) := N_{\text{seg}}[2\bar{\mathcal{F}}(\vartheta) - \mathbb{E}_n(2\bar{\mathcal{F}})] \quad (38)$$

where  $\mathbb{E}_n(2\bar{\mathcal{F}}) = 4$ , the expectation value of the semicoherent  $2\bar{\mathcal{F}}$ -statistic for noise-only data. The substitution factors out in equation Eq. (34) and has no effect on the parameter estimation, but removes an  $N_{\text{seg}}$  dependent scaling of the evidence [see Eq. (37)], which simplifies the comparison of results with different  $N_{\text{seg}}$ .

## D. Estimating the posteriors and evidences

A direct evaluation of Eq. (34) is not possible because the evidence [Eq. (37)] cannot be expressed in closed form

as a function of the data,  $d$ . A numerical approach is needed. One could evaluate

$$p(\vartheta|\{d_\ell\}) \propto e^{N_{\text{seg}}\bar{\mathcal{F}}(\vartheta)}p(\vartheta) \quad (39)$$

over a  $\vartheta$ -template bank as done for grid-based searches, estimate the evidence  $Z(\mathcal{R}_c)$  by numerical quadrature [41], and approximate the posterior density, e.g., by interpolation. However, template banks that resolve the fine structure of the posterior distribution are computationally inefficient, because the majority of statistic evaluations are allocated to regions that contribute little to the evidence integral.

Instead, we use the nested sampling algorithm [42]. We give a detailed description of nested sampling in Appendix A and refer to [43] for a recent review. In short, the algorithm performs a stochastic evaluation of the evidence integral [Eq. (37)] using an ensemble of samples randomly drawn from the prior  $p(\vartheta)$ —called live points—that are iteratively updated toward regions of higher likelihood.

Nested sampling has several advantages for our purpose. On one hand, the algorithm performs a thorough exploration of the parameter space before converging to any found maxima, aiding in the discovery of “interesting,” signal-like features. On the other hand, we apply nested sampling to estimate the evidences and posterior distributions for to up to millions of candidates. Yet, almost all candidates examined are expected to be noise whereas signal-like features are rare. Thus, it is crucial to use a sampling algorithm that will converge even if no interesting modes are found. This is not a trivial property satisfied by nested sampling by virtue of an evidence-based converge criterion [see Eq. (A11)].

As its main downside, nested sampling may terminate without sampling interesting modes if none of the live points finds them by chance. The chance for live points to find interesting modes decreases with a higher number of modes, but increases with a higher number of live points. However, a higher number of live points also linearly increases the runtime of the algorithm [see Eq. (A7)]. If significant modes are discovered, all live points are eventually replaced by samples from the most significant mode in later iterations, and the algorithm is not prone to get “stuck” in secondary modes.

The number of possible modes in a search can be estimated from the number of independent templates needed to cover the searched space  $\mathcal{R}_c$ . A rough estimate is to use the metric of the employed detection statistic and compute the number of unit-mismatch hyperellipsoids [Eq. (12)] necessary to cover  $\mathcal{R}_c$ . This quantity is referred to as  $N_\star$  [7,32,44], and is often used as a measure of the “proper” volume of the search space.

Once converged, nested sampling yields an estimate of the evidence  $Z(\mathcal{R}_c)$  and a set of weighted samples  $\{w_i, \vartheta_i\}$

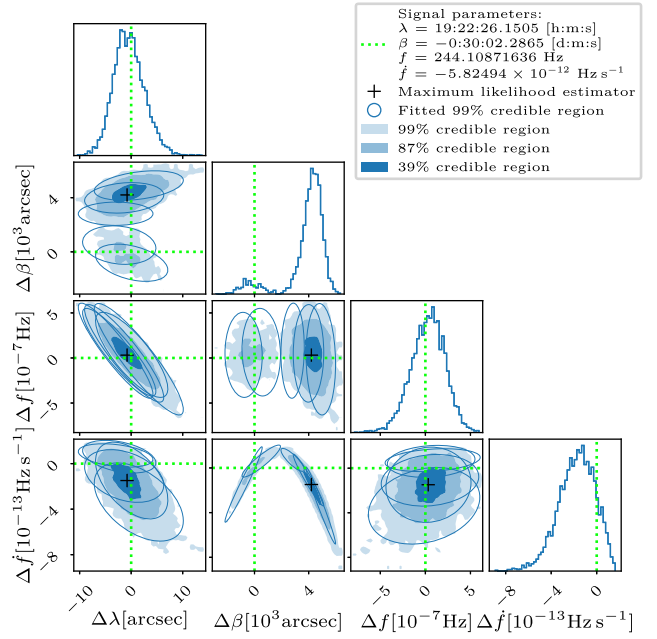


FIG. 1. Posterior distribution for a fake signal near the ecliptic equator ( $\beta \approx 0$ ) added to O3a data and using the semicoherent  $\mathcal{F}$ -statistic with  $T_{\text{coh}} = 120$  h. The posterior shows a bimodal distribution, split between the hemispheres. The plot indicates the  $\Delta(C = 0.99)$  region of each mode of the fitted Gaussian mixture model following Eq. (47).

that can be used to approximate the posterior density [see Eqs. (A12) and (A13)]. However, the follow-up method we describe in the next section requires a functional form of the posterior density. We obtain such a form by modeling the posterior probability density with a Gaussian mixture, accounting for the full covariance and multimodality,

$$p(\vartheta|\{d_\ell\}) \approx \sum_k m_k \mathcal{G}(\vartheta|\mu_k, \Sigma_k) \quad \text{with} \quad \sum_k m_k = 1. \quad (40)$$

This is a weighted sum of multivariate Gaussian distributions  $\mathcal{G}$  with means  $\mu_k$  and covariance matrices  $\Sigma_k$ .

We adopt a two-step process to infer suitable model parameters  $\{m_k, \mu_k, \Sigma_k\}$  from the weighted samples:

- (1) We use the DBSCAN algorithm [45] to identify clusters of samples based on their parameter-space distances. We use the semicoherent phase metric [Eq. (12) and the end of Sec. II] to define a minimum distance between neighboring samples to be part of the same cluster. We also set a minimum occupancy per cluster of at least  $\dim(\mathcal{R}) + 1$  (5, in our case) to ensure nondegenerate covariance matrices. Samples marked as noise by DBSCAN are dropped.
- (2) The weighted samples in each cluster are fitted with a Gaussian mixture model [46]. We do not know *a priori* the number of modes, so we estimate them from the set of samples. We have found that a recursive algorithm that splits the cloud of samples

using a model with two Gaussian components, until a one-component model yields a better description, works well. We use the Bayesian information criterion [30] to quantify the quality of the fit and place thresholds to stop the recursion. The thresholds are calibrated with posteriors obtained from simulated signals. We define a minimum occupancy per mode. If after a split, a set of points is below this limit, this set is also dropped.

The process prioritizes the modelling of signal-like posteriors. For any search, the process is validated with a reference population of simulated signals.

Figure 1 shows the Gaussian mixture model for the posterior of a simulated signal. The signal was placed near the ecliptic equator to produce a characteristic multimodal shape. The Gaussian mixture model captures the overall shape well with only small discrepancies we consider acceptable.

### E. The candidate follow-up

We use the inference from one search stage to inform the next one. Given a posterior distribution  $p^i(\vartheta|\{d_\ell\})$  for a Stage  $i$ , we let it serve as the prior for stage  $i + 1$ ,

$$p^{i+1}(\vartheta) = p^i(\vartheta|\{d_\ell\}). \quad (41)$$

We then use nested sampling and Eq. (34) to estimate  $p^{i+1}(\vartheta|\{d'_\ell\})$ . This conditions the inference at higher coherence times to only consider the most promising parameter-space regions. The efficient sampling of the prior in Eq. (41) is done using the inverse-transform method on our Gaussian mixture models [Eq. (40)] as described in Appendix C.

The process is repeated until the last stage, when the coherence time is equal to the data span.

We note that higher coherence times do not reveal entirely new information. Repeated application of Bayes's theorem and Eq. (41) may, thus, reuse some information, which incurs compounding biases in the estimated signal parameters. However, the goal of a search hierarchy is not to produce the most accurate parameter estimation, but to conduct computationally efficient follow-ups. Any candidates left after the final coherent stage concludes will be subject to more detailed examinations in any case. Further, if longer coherence times yield sufficient new information, the biases will be negligible. The extent of this effect on the follow-up presented in Sec. IV is investigated in Sec. IV D.

Conversely, the initial stochastic stage is initialized with a list of signal candidates marked by seeds  $\vartheta_s$  and an uncertainty range  $\Delta\vartheta$  given by the previous, deterministic stage. For each candidate, we choose uninformative priors based on Jeffreys' invariance principle [30] that span the respective uncertainty region,  $\mathcal{R}_c = \vartheta_s \pm \Delta\vartheta$ .

Jeffreys' invariance principle demands that results from Bayesian inference are invariant under reparametrization of

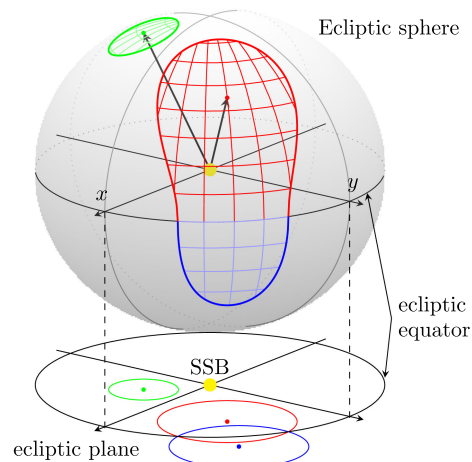


FIG. 2. Circles defining two uncertainty regions on the ecliptic plane and their projection to the ecliptic sphere. The size of the uncertainty regions is exaggerated. Red: uncertainty region for a candidate that is intersecting the ecliptic equator. Blue: the extension of the uncertainty region onto the opposing hemisphere. Green: uncertainty region of a candidate at higher frequency, which corresponds to a smaller circle on the ecliptic plane.

the model. We obtain such priors by first choosing a parametrization for which the response of the  $\mathcal{F}$ -statistic to a signal does not depend on the signal parameters. This corresponds to parametrizations for which the  $\mathcal{F}$ -statistic metric is flat. Using this parametrization, we set uniform priors, and transform them back to our physical parameters. A detailed derivation can be found in Appendix B.

For a search over sky position, frequency and frequency derivative, the  $\mathcal{F}$  metric is already flat for frequency and frequency derivative. For the sky position, with a fixed frequency and long data spans, the metric is approximately flat when projected to the ecliptic plane.<sup>1</sup> The sky-position prior for a candidate at a position  $\beta, \lambda$  and with frequency  $f$  translated from a uniform distribution on the ecliptic plane is accordingly

$$p(\lambda, \beta|f) = \frac{2\pi}{A_{\text{ecl}}(\mathcal{R}_{\text{sky}})} \sin(|\beta|) \cos(|\beta|), \quad (42)$$

where  $A_{\text{ecl}}(\mathcal{R}_{\text{sky}}) = \pi r^2$  is the area of the sky-uncertainty region on the ecliptic plane (see Appendix B for a derivation). This is a circle of radius  $r$  on the ecliptic plane, which is inversely proportional to the signal frequency. This circle projects on the celestial sphere to  $\mathcal{R}_{\text{sky}}$ , whose position and shape depend on where the circle lies on the ecliptic plane.

Figure 2 shows the sky-parameter uncertainty regions and their projection on the celestial sphere for two

<sup>1</sup>This is why the deterministic searches employ sky position grids that are a uniform (hexagonal) area tiling of the orthogonal projection of each celestial hemisphere on the ecliptic plane.

examples. One of these examples depicts a special case that arises when the uncertainty circle on the ecliptic plane intersects the ecliptic equator. The case corresponds to the  $\mathcal{F}$ -statistic response to signals near the ecliptic equator, which is degenerate in the ecliptic hemispheres. To account for such cases, we place a second circle with the same radius that intersects the equator in the same locations as the original circle. Sky points from the opposing hemisphere of the candidate seed are included in the search if their projected ecliptic coordinates lie inside this second circle.

### F. Uncertainty on candidates' parameters

For each candidate, the posterior on the signal parameters defines the *credible region*

$$\Delta(C) := \arg \min_{\Theta \subset \mathcal{R}_c} \left[ \text{Vol}(\Theta) : \int_{\Theta} p(\vartheta|d) d\vartheta = C \right]. \quad (43)$$

$\Delta(C)$  is the smallest possible subregion of  $\mathcal{R}_c$  where the parameters of a putative signal lie with probability  $C$ . The definition does not demand  $\Delta$  to be connected, thus naturally accounting for multimodal posteriors.

The  $C$ -credible regions of an unimodal Gaussian with covariance matrix  $\Sigma$  are nested ellipsoids with surfaces defined by all  $\vartheta$  with constant Mahalanobis distances from the mean-vector  $\mu$ ,

$$d_{\text{Mah}}^2(\vartheta) := (\vartheta - \mu) \Sigma^{-1} (\vartheta - \mu). \quad (44)$$

Each distance  $d_{\text{Mah}}^2$  is associated with the cumulative probability  $C$  of the signal parameters laying within that ellipsoid, given by the CDF of a  $\chi^2$  distribution with  $\dim(\mathcal{R})$  degrees of freedom. Inverting this relation, for any  $C$ , there is some distance

$$\xi^2(C) = F_{\chi^2(\text{df}=\dim(\mathcal{R}))}^{-1}(C), \quad (45)$$

and, thus,

$$\Delta(C) = \{\vartheta | d_{\text{Mah}}^2(\vartheta) < \xi^2(C)\}. \quad (46)$$

For a multimodal Gaussian, no closed form expression exists. Instead, we approximate  $\Delta(C)$  from the union of the per-mode credible regions  $\Delta_k(C)$ ,

$$\Delta(C) \approx \cup_k \Delta_k(C), \quad (47)$$

$$\Delta_k(C) = \{\vartheta | d_{\text{Mah}}^2(\vartheta, \Sigma_k, \mu_k) < \xi^2(C)\}. \quad (48)$$

This approximation overestimates  $\Delta(C)$ : First, while each  $\Delta_k(C)$  contributes  $m_k \times C$  to the cumulative probability according to its weight  $m_k$ , each mode also contributes some residual probability from those points within the credible regions of other modes but not within  $\Delta_k(C)$ . Thus, the

cumulative probability of the union of all  $\Delta_k(C)$  is larger than  $C$ , and so is its volume. Second, the union may include points with lower posterior probability density than some points outside the union. However, the approximation is sufficient for an order-of-magnitude characterization of credible-region volumes.

Out of all regions  $\Theta \subset \mathcal{R}_c$  with  $\text{Vol}(\Theta) = \text{Vol}(\Delta(C))$ ,  $\Delta(C)$  *maximizes* the integrated posterior probability and thus must contain the  $\{\vartheta\}$  with the highest posterior probability density. Defining

$$X'(\mathcal{P}) := \int_{\{\vartheta | p(\vartheta|\{d_\ell\}) \geq \mathcal{P}\}} p(\vartheta|\{d_\ell\}) d\vartheta, \quad (49)$$

there is a  $\mathcal{P}_C$  such that  $X'(\mathcal{P}_C) = C$ , and we obtain

$$\Delta(C) = \{\vartheta | p(\vartheta|\{d_\ell\}) \geq \mathcal{P}_C\}. \quad (50)$$

Equations (49) and (50) allow us to associate any parameter point  $\vartheta$  with a credible level  $C_\vartheta = X'(\mathcal{P}_\vartheta)$ , where  $\mathcal{P}_\vartheta = p(\vartheta|\{d_\ell\})$ .  $C_\vartheta$  is the cumulative probability of all parameter points with higher posterior probability density than  $\vartheta$ .

## IV. APPLICATION ON REAL DATA

We demonstrate the presented framework in a realistic and demanding setting by applying it to the results of the Einstein@Home search by Steltner *et al.* [13], which we refer to henceforth as “the original search.”

The original search was conducted on public Advanced LIGO data from the third observation run [47]. The observation run was split into two subperiods, O3a and O3b. The first stages of the original search only considered O3a, leaving the O3b data for validation.

The original search targets a 90% detection efficiency for signals with sensitivity depth [48] of  $\mathcal{D} = 56[1/\sqrt{\text{Hz}}]$ , frequencies  $20 \text{ Hz} < f_s < 800 \text{ Hz}$ , first-order spindowns  $-2.6 \times 10^{-9} \text{ Hz s}^{-1} < \dot{f}_s < 2.6 \times 10^{-10} \text{ Hz s}^{-1}$ , and sky position parameters uniform all over the sky.

### A. Setup

The target detectable signal amplitudes are such that at the output of Stage 0, the signal candidates' detection statistic should be at least at the level of the loudest Gaussian noise candidates. A stochastic framework like ours could miss such a signal either because the search region is too large—the number of live points required for the sampler to converge to the signal is computationally not viable—or because even if there are enough live points for the search region, this is inhabited by very loud disturbances, which the sampler converges to.

The second problem is mitigated by data-preparation and cleaning procedures before any processing is carried out on the data [49], and by using line-robust detection statistics to

rank the initial search results [50]. Some disturbances do, however, persist and if they resemble an actual astrophysical signal, then one’s ability to detect a signal within the same region will be significantly lowered. This is an issue that is common to stochastic and deterministic searches—in the latter case typically through top-list saturation effects. This effect is however rare, affecting less than a few percent of the results.

Conversely, the first problem of too large search regions is specific to stochastic methods and hence is very relevant to the characterization of our framework. In particular, since the uncertainty regions around candidates at each stage in the deterministic follow-up shrink as the stages proceed, this problem determines at which stage in the follow-up hierarchy we can begin applying our framework. In principle, we would like to begin the stochastic follow-up directly on the Stage 0 candidates, i.e., with Stage 1.

To determine at which stage we can begin applying our framework we study the recovery of test signals at various stages of the original search. Our detection criterion tests that the maximum likelihood estimator  $\vartheta_{\text{MLE}}$  that the sampler has found is related to the test signal with parameters  $\vartheta_s$ . This is enforced by requiring that the power  $\hat{\rho}^2(\vartheta_{\text{MLE}})$  estimated at the maximum likelihood point is close to the *estimated* test-signal power  $\hat{\rho}^2(\vartheta_s)$  [using Eq. (38) for both estimates]. Since the power  $\rho^2$  is a monotonic function of the likelihood, if the likelihood sampler has converged to the signal, then it must be that  $\hat{\rho}^2(\vartheta_{\text{MLE}}) \geq \hat{\rho}^2(\vartheta_s)$ . On the other hand the  $\hat{\rho}^2(\vartheta_{\text{MLE}})$  cannot be too much larger than  $\hat{\rho}^2(\vartheta_s)$ , as that would also indicate convergence to some other local maximum. So in all the recovery criterion is,

$$\hat{\mu}_{\min} \leq \hat{\mu}_S := \frac{\hat{\rho}^2(\vartheta_s) - \hat{\rho}^2(\vartheta_{\text{MLE}})}{\hat{\rho}^2(\vartheta_s)} \leq 0 + \epsilon. \quad (51)$$

We use a tolerance of  $\epsilon = 10^{-4}$  for the upper bound to allow for a small offset of  $\vartheta_{\text{MLE}}$  from the exact maximum likelihood template due to sampling noise. We set  $\hat{\mu}_{\min} = -0.125$ . In Gaussian noise the distance between the  $\vartheta_{\text{MLE}}$  and  $\vartheta_s$  is a multivariate Gaussian random variable with covariance matrix  $[\rho^2 g_{ij}(\vartheta_s)]^{-1}$  [27]. Following Eqs. (11), (44), and (45) the expected distance for the  $C$  quantile is

$$d_{\text{Mah}}^2 = \rho^2 g_{ij}(\vartheta_s) \Delta \vartheta_i \Delta \vartheta_j = \xi^2(C) = \rho_S^2 \mu. \quad (52)$$

The weakest signal from our target population has  $\rho^2 \gtrsim 200$ . At this amplitude, for  $C = 99.995\%$  then  $d^2 \approx 25$  and  $\mu \approx 0.125$ .

We note that the detection criterion can produce false positives in the early stages of the follow-up because the semicoherent  $\mathcal{F}$ -statistic response to a signal is generally multimodal (see Refs. [28,51] and the example posterior shown in Fig. 1). In this case, unlucky noise fluctuations can elevate secondary modes just enough for the sampler to converge on them, missing the “correct” signal parameters.

TABLE I. Per-stage coherence times  $T_{\text{coh}}$ , number of coherent data segments  $N_{\text{seg}}$ , and live-point counts of the search stages. The search regions in frequency, spindown, and sky covered by the deterministic follow-up of [13] are given in the fourth through sixth columns. These are relevant for Stage 1, where the search regions of the seeds were too large to follow up with our scheme, and for Stage 2, which is the first stage conducted with our framework, and that hence uses those uncertainty regions as the prior support. Stages 3–7 use the posterior of the preceding stage as their prior with the full coherence time hierarchy. We also investigate if Stages 3–7 can be replaced with a single, fully coherent stage, Stage 3\*, using the Stage 2 posterior as the prior. Stages 7 and 3\* use a single coherent segment spanning the full O3a observation period.

	$T_{\text{coh}}$ h	$N_{\text{seg}}$	$\Delta f$ μHz	$\Delta \dot{f}$ $10^{-14}$ Hz s $^{-1}$	$\frac{r_{\text{sky}}}{d_{\text{sky}}(f)}$	$n_{\text{live}}$
Stage 1	120	37	1000	11 250	10.0	1500
Stage 2	120	37	50	1200	2.0 <sup>a</sup>	1500
Stage 3	240	19				750
Stage 4	490	9				500
Stage 5	1100	4				500
Stage 6	2200	2				500
Stage 7	coh.	1				500
Stage 3*	coh.	1				750

<sup>a</sup>Due to an error the Stage 2 sky uncertainty used for the follow-up is *larger* than necessary. This makes the follow-up more challenging and hence does not invalidate our findings.

Another cause of false positives is the presence of disturbances also leading to convergence far from the “correct” signal parameters. The impact of both these factors is reduced substantially at higher coherence times, so that as the coherence time increases at each stage, this empirical detection criterion becomes a more and more accurate proxy.

We investigate whether we can use our framework for the entire hierarchy of follow-up stages, i.e., starting with Stage 1. So we apply our framework to the test-signals seeds as they are produced after Stage 0 in the original search, and to the test-signals candidates surviving Stage 1 of the original search. We take as support for the priors the uncertainty regions shown in the first two rows of Table I. We use the DYNESTY sampler [52] for the nested sampling. The specific sampler configuration is given in Appendix D. The number of live points for each configuration is given in Table I or specified explicitly. The recovery results, based on the criterion of Eq. (51) are shown in Fig. 3.

At Stage 1, a significant fraction (10% ~ 20%) of our test signals is lost. A higher number of live points increases the recovery rate, but shows the expected linear scaling in the number of likelihood evaluations [Eq. (A7) and Ref. [43]]. The number of live points necessary to recover all simulated signals incurs computing costs that are not markedly competitive with the deterministic follow-up.

At Stage 2, with 1500 live points, we recover all test signals. We take this as our configuration and use our framework for Stages 2–7.

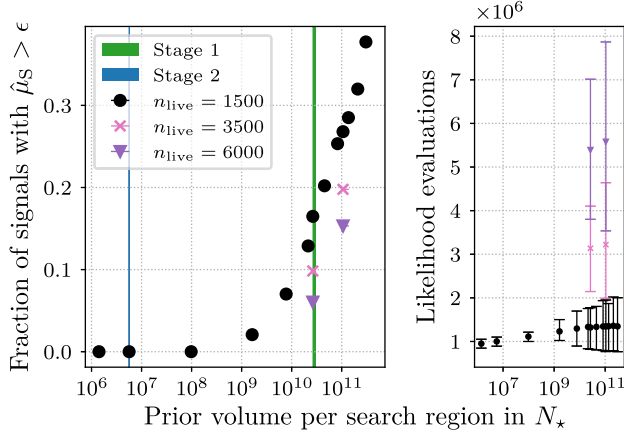


FIG. 3. Left: the fraction of missed test signals for search configurations with different prior volumes. Right: the average number of per-signal likelihood evaluations for the recovered signals. Vertical bars show the 99th quantile. In both plots, vertical colored areas mark configurations listed in Table I. The volumes are given in terms of the number of unit-mismatch hyperellipsoids contained in the search region,  $N_*$ .

After Stage 7 all test signals are still recovered (see Fig. 10 for the distribution of the  $\hat{\mu}$ ). As a precaution against disturbances, we perform a visual inspection of corner plots of the posterior distributions for each signal. We do not observe cases where the sampler converged on disturbances.

Through Stages 3–7, we expect that the parameter space volumes to be examined have decreased substantially. We thus reduce the number of live points as listed in Table I to improve the computational efficiency. However, the next stages are computationally relatively trivial compared to Stage 2, while recovering all test signals anyway. Thus, we did not optimize the number of live points for later stages. The used numbers of live points here are chosen *ad hoc*.

### B. Follow-up of candidates

We follow up the  $\approx 350000$  candidates produced by Stage 1 of the original follow-up.

In Fig. 4 we show the distribution of the evidence from our Stage 2 follow-up. We can see that the test signals from our target population at this stage do not yet separate from the candidates.

We expect the evidence for a real signal to stay approximately constant as the coherence time of the search increases. In contrast, disturbances are unlikely to accumulate significance just as a signal would when scrutinized over longer and longer coherence lengths, so we expect that the results of our noise-dominated data will separate from the results of the test signals as the stages progress.

We define the quantity

$$R^i := \frac{\log Z^i - \log Z^2}{\log Z^2} \quad \text{with } i = 3, \dots, 7, \quad (53)$$

which characterizes the evolution of the search candidates across the different stages. Figure 5 shows the distributions

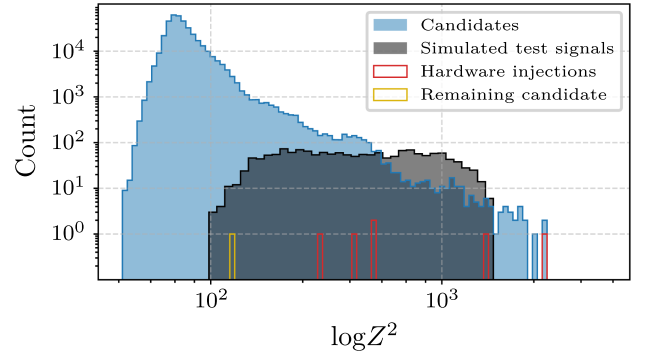


FIG. 4. Log-evidences  $\log(Z^2)$  recorded for each candidate and simulated test signal. The candidates corresponding to hardware injections and the candidate remaining compatible with the test signal population after Stage 7 are emphasized.

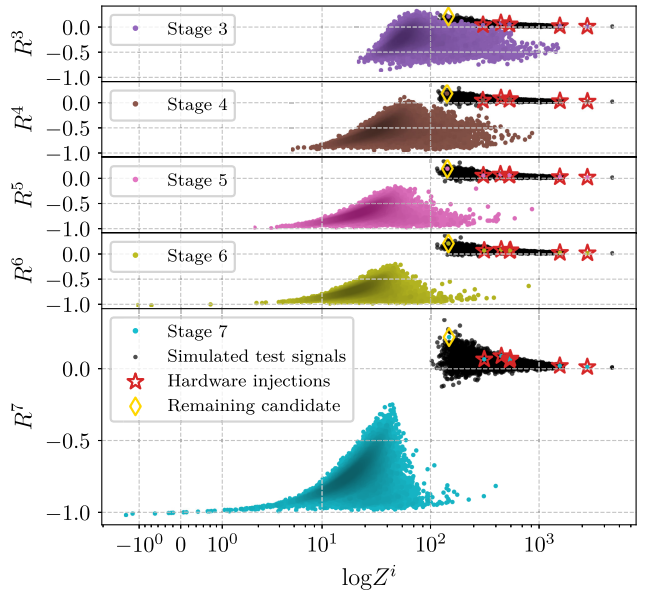


FIG. 5.  $\log Z^i$  recorded across all follow-ups at stages  $i = 3, \dots, 7$  and the relative change  $R^i$ . The candidates corresponding to hardware injections and the remaining significant candidate are emphasized. Two hardware injections closely match in  $\log Z^i$  and  $R^i$  and coincide in the plots at each stage.

of the evidences  $R^i$  versus  $\log Z^i$  for all candidates and test signals of each Stage  $i = 3, \dots, 7$ .

As the stages advance the  $(\log Z^i, R^i)$  points of the search candidates progressively separate from those of the target signals, and at Stage 7 the bulk of the candidate distribution has markedly separated from the target signal distribution. At the tail of the bulk (high  $\log Z^7$ , small  $R^7$ ), a few candidates show log evidences compatible with our test signals. However, the decrease in evidence relative to the initial stochastic stage (Stage 2) of these candidates is not consistent with what is expected from a signal, and hence we do not consider these defensible signal candidates.

TABLE II. The frequency, first spindown, right ascension, and declination of the hardware injections, and the distances between these parameters and the maximum likelihood estimators obtained from our follow-up. The sky coordinates are in equatorial coordinates. The reference time for these values is 1253764756 (GPS time). We note the lower offsets reported in the original search using the full O3 data set, i.e., about twice as much data (see Table 3 of [13]).

ID <sub>inj</sub>	$f_{\text{inj}}$ (Hz)	$\dot{f}_{\text{inj}}$ (Hz s <sup>-1</sup> )	$\alpha_{\text{inj}}$ (h:m:s)	$\delta_{\text{inj}}$ (deg:m:s)	$\Delta f$ (Hz)	$\Delta \dot{f}$ (Hz s <sup>-1</sup> )	$\Delta_{\text{sky}}$ (deg:m:s)
0	265.57502155	$-4.15 \times 10^{-12}$	4:46:12.4628	-56:13:02.9490	$6.5 \times 10^{-8}$	$9.0 \times 10^{-15}$	0:00:00.4012
2	575.16350421	$-1.37 \times 10^{-13}$	14:21:01.4800	3:26:38.3626	$3.6 \times 10^{-8}$	$4.3 \times 10^{-15}$	0:00:00.2198
3	108.8571594	$-1.46 \times 10^{-17}$	11:53:29.4178	-33:26:11.7687	$1.8 \times 10^{-7}$	$2.4 \times 10^{-14}$	0:00:02.9579
5	52.80832436	$-4.03 \times 10^{-18}$	20:10:30.3939	-83:50:20.9036	$4.8 \times 10^{-8}$	$6.0 \times 10^{-15}$	0:00:01.6488
9	763.84731649	$-1.45 \times 10^{-17}$	13:15:32.5397	75:41:22.5205	$-7.9 \times 10^{-8}$	$-9.5 \times 10^{-15}$	0:00:00.1649
10	26.33144237	$-8.5 \times 10^{-11}$	14:46:13.3549	42:52:38.2953	$5.7 \times 10^{-8}$	$8.3 \times 10^{-15}$	0:00:03.9060

Only seven candidates remain compatible with the target distribution of simulated signals. Six of these candidates are due to hardware injections. These are fake signals added to the data at the hardware level to validate search pipelines [47]. The signal parameters of the hardware injections, and their offsets from the respective maximum likelihood estimators are listed in Table II. The framework recovers the parameters with excellent agreement.

The parameter space examined by the original search contains a seventh hardware injection (ID 11 [47]), which was not recovered by Stage 0 the original search, thus was not in the list of the  $\approx 350\,000$  candidates that we follow up and hence our framework could never recover it. Hardware injection 11 has an amplitude lower than the upper limit level of the original search, and we refer the reader to Steltner *et al.* [13] for more information.

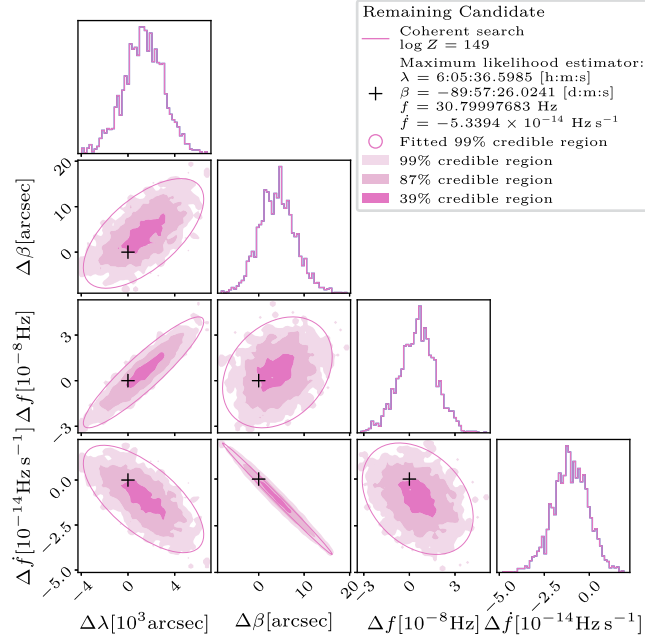


FIG. 6. Posterior distribution for the remaining candidate at Stage 7. The plot indicates the  $\Delta(C = 0.99)$  region of the fitted Gaussian mixture model following Eq. (47).

After closer inspection of the seventh remaining candidate, we cannot defend it as a significant signal candidate. Figure 6 shows the posterior distribution and lists the parameters of the maximum likelihood estimator at the final, coherent stage. We see that the first derivative of the frequency is close to zero, and that the sky position is at the southern pole of the ecliptic sphere ( $\beta \approx -\pi/2$ ) where the Doppler-modulation

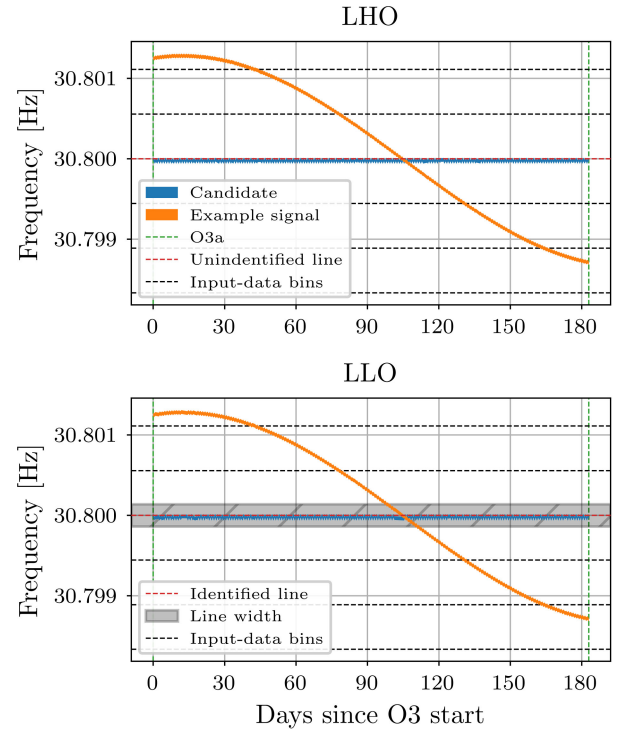


FIG. 7. Instantaneous frequency evolution of the remaining candidate in the detector-frame of the Hanford (LHO) and Livingston (LLO) observatories and known lines. The signal parameters are that of the maximum likelihood estimator in Fig. 6. The same signal but offset in the sky coordinates to ( $\lambda = 0$ ,  $\beta = -1.25$ ) is shown as an example of a signal modulated by Earth's motion around the sun. The nature of the line in LHO is unknown, whereas the LLO line is confirmed as nonastrophysical. The official lines lists can be found in [53]. We mark the input-data bin widths and the linewidths provided in [53] by LIGO.

due to Earth’s motion around the Sun is minimal—our signal has a meek modulation depth of  $\leq 3 \mu\text{Hz}$ . Astrophysical signals with these parameters are very similar to constant-frequency lines due to coherent, long-lasting disturbances in the data. In the case of the remaining candidate, both detectors feature lines close to/overlapping with the candidates waveform, as shown in Fig. 7. LIGO provides detailed information on lines of known and unknown origin [53]. The line in the Hanford detector is of unknown origin. The line in the Livingston detector was identified as of nonastrophysical origin, however, only after the Einstein@Home was deployed. Thus, this line was not removed in the data preparation phase (see Ref. [49]), resulting in the spurious candidate. This stresses the importance of data cleaning and noise identification [54–57].

The fully coherent stage of the original search on O3a data recorded 12 candidates, including the same six hardware injections, but not the candidate that remains in this follow-up. The remaining six candidates from the original search were finally rejected by incorporating the O3b data set. On O3a data alone, the new framework recorded fewer false alarms.

The clean separation of noise candidates and test signals shows that our approximation in deriving the (semicoherent)  $\mathcal{F}$ -statistic—Eqs. (22), (24), and (27)—did not negatively affect the effectiveness of the follow-up relative to our population of test signals and the candidates.

### C. Search volume compression

A further quantity that can be used to characterize the results of the search are the per-candidate prior and posterior volumes at each stage, and the achieved reduction of the volume. These distributions for each stage are shown in Fig. 8. Follow-ups of candidates yield significantly larger posterior volumes than those on simulated signals, which translates to larger prior volumes in the next stage. This difference is explained by the lack of pronounced peaks of the likelihood in follow-ups of candidates.

### D. Parameter estimation

If the inferred Gaussian mixture model posterior density  $p(\vartheta|\{d_\ell\})$  estimated from data containing a signal with parameters  $\vartheta_S$  accurately models the parameter uncertainties, the posterior must be consistent with  $\vartheta_S$ ,

$$P(\vartheta_S \in \Delta(C)) = C. \quad (54)$$

We can evaluate the accuracy of our posteriors by estimating  $P(\vartheta_S \in \Delta(C))$  using the results from the reference signal population. This procedure leads to the “pp-plots” (probability-probability plots) that are often used to validate results from Bayesian parameter estimation.

Let us fix a stage. For each of the  $N_{\text{signal}}$  test signals with parameters  $\vartheta_S$ , we take the Gaussian mixture model posterior  $p(\vartheta|\{d_\ell\})$  and estimate  $X'(\mathcal{P}_S)$  of Eq. (49) as

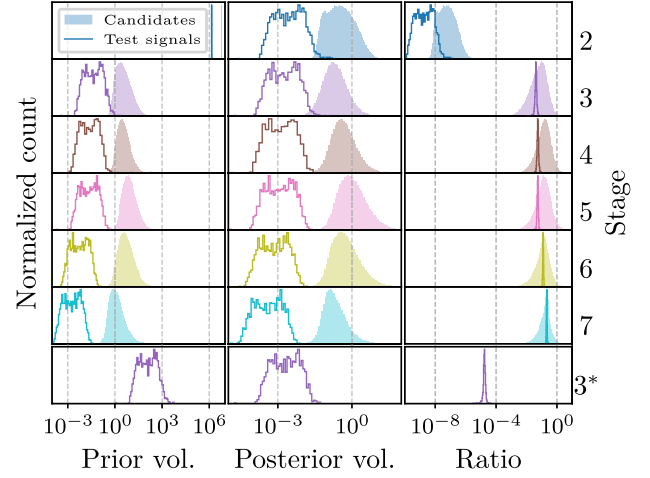


FIG. 8. Distributions of the parameter space volumes covered by the prior distributions, the posterior distribution, and their ratio for each stage. The search results for the candidates of the original search are the shaded regions. The solid lines show the results for the target signal population. All volumes are given in  $N_*$ . All volumes except the prior volumes at Stage 2 correspond to  $\text{Vol}(\hat{\Delta}(99.9\%))$  of the fitted Gaussian mixture models [see Eq. (47)].

$$X'(\mathcal{P}_S) \approx \frac{\sum_{j=1}^N \mathbb{1}_{\{\vartheta|p(\vartheta|\{d_\ell\}) > \mathcal{P}_S\}}(\vartheta_j)}{N}, \quad (55)$$

with  $\mathcal{P}_S = p(\vartheta_S|\{d_\ell\})$ ,  $N = 100\,000$  being the total number of  $\vartheta_j$  samples drawn from the posterior and  $\mathbb{1}_X(x)$  the indicator function

$$\mathbb{1}_X(x) \equiv \begin{cases} 1 & \text{if } x \in X \\ 0 & \text{otherwise.} \end{cases} \quad (56)$$

$X'(\mathcal{P}_S)$  of Eq. (55) is the probability of parameter-space volumes with posterior density higher than that of the signal. We then count the fraction of test signals for which  $X'(\mathcal{P}_S) < C$  and take it as an estimate of  $P(\vartheta \in \Delta(C))$ ,

$$P(\vartheta_S \in \Delta(C)) \approx \frac{\sum_S \mathbb{1}_{\{X'(\mathcal{P}_S) < C\}}(\mathcal{P}_S)}{N_{\text{signal}}}. \quad (57)$$

Figure 9 shows  $P(\vartheta \in \Delta(C))$  as a function of  $C$  for all stochastic search stages. Systematic biases in the parameter estimation would lead to deviations from Eq. (54).

Already at Stage 2 more test signals than expected are found in low posterior probability regions, corresponding to  $P(\vartheta \in \Delta(C)) < C$ . This shows that the posterior is slightly biased toward being narrower than it actually is, and indicates that for this stage the underlying parameter uncertainties are not captured perfectly. Two approximations made by our framework are the likely causes. The first approximation was made in deriving the  $\mathcal{F}$ -statistic based version of Bayes’s theorem in Eqs. (22), (24), and (27),

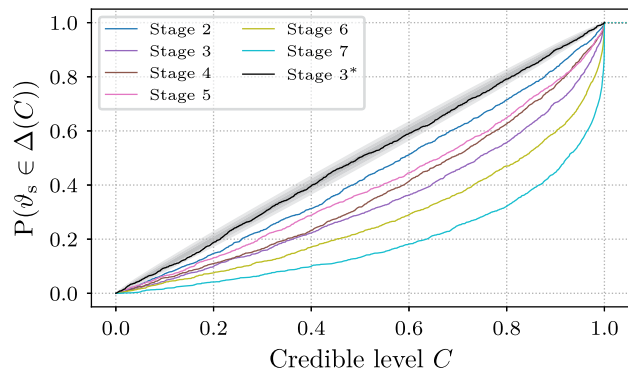


FIG. 9. Probability-probability plot showing the test signal recovery for all stochastic follow-up stages. The  $x$  axis plots the credible level interval  $[0, 1)$ . The  $y$  axis plots the (frequentist) probability of recovering test signals within a credible level  $C$ ,  $P(\vartheta_s \in \Delta(C))$ . Plotted in gray are the 1, 2, and  $3\sigma$  confidence regions expected for unbiased parameter estimation given the number of test signals.

where we assumed that only the most likely amplitude parameters  $\mathcal{A}_{\text{MLE}}$  contribute to the evidence integral. Instead, if small offsets in the phase-evolution parameters can be absorbed by small offsets in the amplitude parameters, the posterior distributions would be wider than our results. The second approximation is the fitted Gaussian mixture model, which may not be able to fit the posterior distributions perfectly.

For Stages 3–7, the bias increases further. This is the expected behavior for repeated application of Bayes’s theorem using the same information, as anticipated in Sec. III E.

We compare these results to a modified coherence time hierarchy of the follow-up after Stage 2: Instead of applying Stages 3–7, we immediately transition to full coherence (Stage 3\* in Table I). Because of the steep increase in coherence time, more information is used for the inference at Stage 3\*, and the effect of the compounding bias is expected to lessen.

On the other hand, the steeper increase of the coherence time also means that the sampler has to explore a prior volume resolving more independent templates (Fig. 8), which could result in the loss of signals similar to our examination of Stage 1. We evaluate Eq. (51) and find that no signals are lost.

Figure 9 shows the  $(C, P(\vartheta_s \in \Delta(C)))$ -graph for Stage 3\* alongside the other stages. The graph agrees well with the expected shape for unbiased parameter estimation. This result also indicates that at full coherence, the approximation made in deriving Bayes’s theorem for the  $\mathcal{F}$ -statistic does not appear to negatively affect the parameter estimation.

Thus, considering only biases in the parameter estimation, a two-stage nondeterministic setup performs better than the full stage hierarchy.

As a further check, we also investigate the frequentist uncertainties of the maximum likelihood estimators by

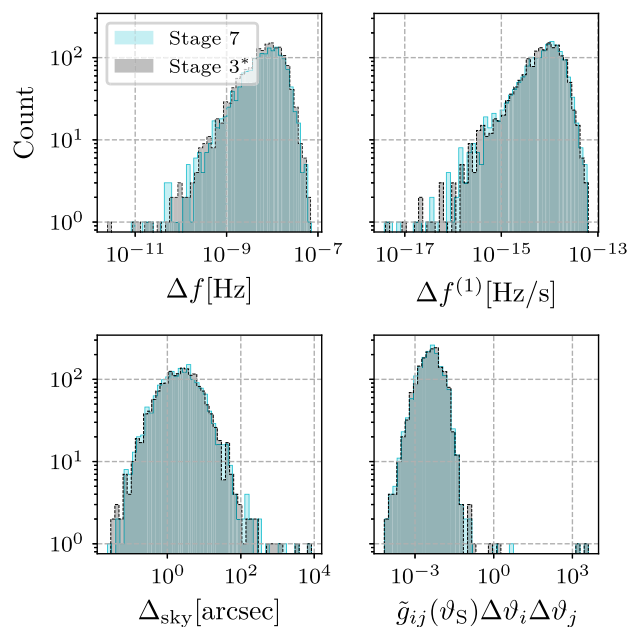


FIG. 10. Parameter-space distances between the signal parameters  $\vartheta_s$  and the maximum likelihood estimators  $\vartheta_{\text{MLE}}$  at the coherent search stages (Stage 7 and 3\*) of each test signal, and the calculated metric-based distance following Eqs. (11)–(13).

considering the distribution of the parameter-space distances between the maximum likelihood estimators and the test signal parameters. The distributions are shown in Fig. 10. For  $f$  and  $\dot{f}$ , the uncertainties are  $\Delta f \lesssim 10^{-7}$  Hz and  $\Delta \dot{f} \lesssim 10^{-13}$  Hz s $^{-1}$ , respectively. For the sky position uncertainty, we find  $\Delta_{\text{sky}} \lesssim 100$  arcseconds for almost all signals, and it is centered at a few arcseconds, consistent with the investigated frequency range and an effective “aperture” of 1 AU.

We note five outliers. The outliers correspond to signals with sky-position parameters very close to the ecliptic equator, and due to noise, the maximum likelihood estimators lie on the “wrong” ecliptic hemisphere. These same signals also produce outliers when considering the metric-based distance measure (lower right plot in Fig. 10): The metric does not account for multi-modality, and, thus, significantly overestimates the mismatch when the  $\mathcal{F}$ -statistic has multiple, nearby modes. Wette [28] provides a different parametrization for the metric that accounts for the specific multimodality encountered. In that parametrization, the calculated metric mismatch of these outliers would be significantly reduced.

### E. Computational cost

Figure 11 shows the number of likelihood evaluations per stage for candidates and test signals. The cost of the initial of Stage 2 on the candidates is comparable with that on simulated test signals. The slight difference—with the searches on noise taking slightly less time than those on

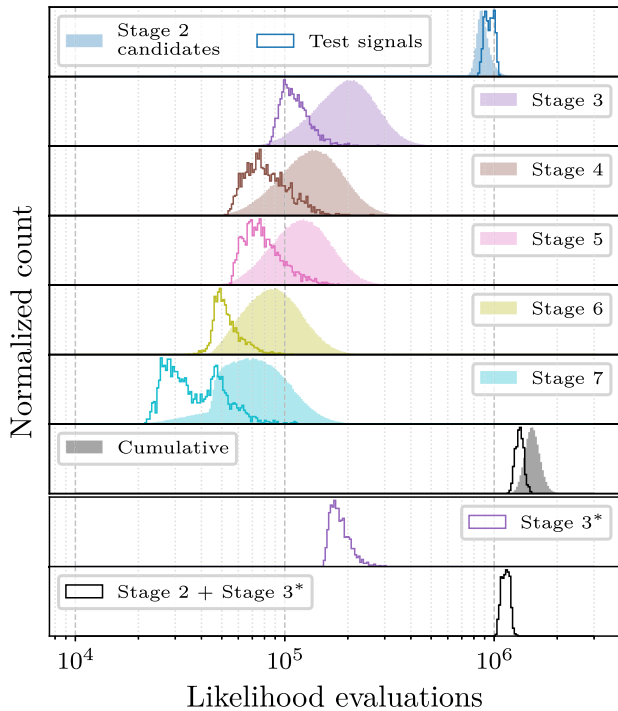


FIG. 11. Distribution of the number of likelihood evaluations performed per simulated signal or candidate at each stage, and cumulative for each simulated signal or candidate across all stages.

simulated signals—is due to the presence of pronounced peaks in the signal likelihoods, which result in larger evidences that take longer to integrate [see Eq. (A11)].

Conversely, in the subsequent stages, the noise candidates take longer to follow up than the signals. The reason is that the noise candidates typically have wider posteriors compared to signal candidates, thereby larger prior volumes to sample in the next stage (see Fig. 8).

Stages 3–7 require significantly fewer likelihood evaluations per candidate compared to Stage 2, for both the candidates and for the test signals. The reduction in computational cost that ensues makes it viable to apply Stages 3–7 to *all* candidates without any intermediate candidate vetoing, and this incurs for the full hierarchy only 1.5 times the computational cost of Stage 2 alone. In contrast, deterministic follow-up schemes require vetoing candidates at all intermediate stages, which requires non-trivial safety studies.

For the simulated test signals, the immediate coherent follow up in Stage 3\* requires approximately twice as many likelihood evaluations as the semicoherent Stage 3, caused by the higher number of waveforms resolved compared to Stage 3. However, because of the smaller number of stages, the total computational cost is smaller than the full hierarchy. We refrain from reapplying Stage 3\* to the search candidates because of the computational cost of a repeated analysis.

To provide context for these results, we give an order-of-magnitude comparison to the signal template grids used by

the original search. These numbers are estimates from the description in [13].

In the original search, the deterministic Stages 2–7 examine  $\approx 10^9$  templates per candidate and stage. Cumulatively, and accounting for candidate rejections after each stage, a total of  $4 \times 10^{14}$  templates are examined.

Assuming  $10^{-2}$  s per likelihood evaluation and neglecting overhead [20], on a computer cluster with  $10^4$  CPUs, this yields a runtime of  $\mathcal{O}(\text{days})$  for our follow-up, which closely matches the actual runtime on the ATLAS computing cluster. For a deterministic search and with the same cost per likelihood evaluation, this would yield  $> \mathcal{O}(\text{years})$  for the original search with candidate rejection. However, we emphasize that searches with grid-based template banks, such as the original search, use optimized algorithms to compute the  $\mathcal{F}$ -statistic that cache certain intermediate results and (significantly) decrease their computational cost. The exact speedup has complex dependencies on the exact setup and available hardware. With these optimizations and accounting for stage setup and accounting for postprocessing, Stages 2–7 of the original follow-up in practice required  $\mathcal{O}(\text{week})$  in combined human and computing time *per stage*.

## V. CONCLUSION AND DISCUSSION

We presented a novel Bayesian framework for large-scale follow-ups of candidates of continuous wave searches. The framework implements a novel stage transition scheme that uses the posterior distribution of one stage as the prior distribution of a new search stage with a higher coherence time.

We present results of a real-world application of this framework to results of the O3 data Einstein@Home all-sky search for continuous waves emitted by isolated neutron stars [13].

We are able to apply the framework to the candidates produced by the Stage 1 follow-up of that search, covering search regions with per-candidate volumes of  $N_{\star} \lesssim 10^8$ . This is consistent with recent results [20] and exceeds the volumes that were previously thought to be the limit of Monte Carlo based frameworks by several orders of magnitude.

We show a significant reduction in the computational cost of subsequent stages and achieve the coherent follow-up of *all*  $\approx 350\,000$  candidates. To our knowledge, no follow-up *without vetoing* between stages on this scale has been reported. Our final search results agree with the original search.

We emphasize the significant reduction of the manual and computational overhead achieved by this approach compared to traditional deterministic follow-up frameworks. Once a viable sampler configuration for an initial target search region volume is found, the full follow-up hierarchy is performed on every candidate to the end. Reference to a target signal population is necessary only for the initial configuration and the final thresholding.

The framework provides a significant simplification of the original follow-up procedure. Still, the necessity of one traditional, grid-based follow-up stage remains for the first stage, due to the large uncertainties in signal parameters in the candidates from that stage. Pushing toward even larger initial volumes is necessary to replace existing follow-up frameworks entirely.

In order to compare with the deterministic follow-ups [13], we use the coherence time hierarchy of that search. However, we show that our framework performs better with a two-stage setup compared to the full coherence time hierarchy when applied to the test signals: The computational cost is reduced and the systematic biases of the parameter estimation are eliminated. This is relevant looking ahead at a follow-up that produces interesting candidates: one of the next steps would then be to search for an electromagnetic counterpart signal informed by our parameter estimates.

Our follow-up only considers O3a data, approximately half of the available time baseline of the third LIGO observation period [47], and about one fifth of the planned time baseline of the fourth LIGO-Virgo-KAGRA observation period [58]. We consider it likely that a two-stage setup is not viable for such a long time baseline. If intermediate stages remain necessary, the PYFSTAT framework and its “offsprings” offer optimization schemes for the construction of the coherence time hierarchy of stochastic searches [21,59,60]. The optimization is based on the *expected* prior volumes of the *population* of candidates at each stage. In our framework, the search regions defined by Eq. (47) would enable a per-candidate optimization based on the actually *observed* volumes—once we implement these schemes.

All-sky searches for continuous waves from neutron stars in binary systems are plagued by the computational cost steeply rising with increasing search coherence length. The development of frameworks like this opens prospects for carrying out extensive follow-ups in these challenging higher dimensionality parameter spaces [1,20].

This project made use of LALSuite [34], SWIGLAL [35], and the SCIKIT-LEARN Python package [61]. The work utilized the ATLAS computing cluster at the MPI for Gravitational Physics Hannover.

## ACKNOWLEDGMENTS

This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 101029058.

## DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

## APPENDIX A: NESTED SAMPLING

Nested sampling is a Monte Carlo algorithm designed to estimate evidences [42]. The method is based on a reformulation of the evidence integral. The prior probability over parameter space points  $\{\vartheta\}$  such that the likelihood  $p(\{d_\ell\}|\vartheta)$  exceeds a certain value  $\mathcal{L}$

$$X(\mathcal{L}) := \int_{\{\vartheta \in \mathcal{R}_c | p(\{d_\ell\}|\vartheta) > \mathcal{L}\}} p(\vartheta) d\vartheta, \quad (\text{A1})$$

allows us to express Eq. (37) as

$$Z(\mathcal{R}_c) = \int_0^\infty X(\mathcal{L}) d\mathcal{L} = \int_0^1 \mathcal{L}(X) dX. \quad (\text{A2})$$

Nested sampling stochastically estimates the  $X_i$  and associated  $\mathcal{L}(X_i)$  through an iterative process. A set of  $n_{\text{live}}$  parameter-tuples, called live points, is sampled from the prior distribution  $p(\vartheta)$ . At iteration  $i$  the live point with the smallest likelihood defines the threshold  $\mathcal{L}_i^t$ . New samples are proposed until one exceeds  $\mathcal{L}_i^t$ . The point is accepted as a live point and the lowest likelihood point is removed. This generates a new set of  $n_{\text{live}}$  live points for iteration  $i + 1$ . A new threshold  $\mathcal{L}_{i+1}^t > \mathcal{L}_i^t$  is defined, and the process repeats itself. The prior probability  $X(\mathcal{L}_i^t)$  is expected to decrease exponentially [see Eq. (A5)]. The  $X_i$  is the prior probability of *all* parameter space points with likelihood exceeding  $\mathcal{L}_i^t$ . The parameter space volumes defined by these points are nested, in the sense that all points in the volume at iteration  $i + 1$  are also in the volume of iteration  $i$ .

The “shrinkage” of  $X_i$  relative to  $X_{i-1}$  is independent of  $i$  and follows a beta distribution,

$$Y := \frac{X_i}{X_{i-1}} \sim \text{Beta}(1, n_{\text{live}}), \quad X_0 = 1. \quad (\text{A3})$$

The expected value of  $\log Y$  is [42]

$$\mathbb{E}[\log Y] = -1/n_{\text{live}} \quad (\text{A4})$$

and hence

$$\mathbb{E}[\log X_k] = \mathbb{E}\left[\log \frac{X_k}{X_0}\right] \approx -k/n_{\text{live}}. \quad (\text{A5})$$

This last equation allows one to estimate the  $X_i$  values in

$$Z_k \approx \sum_i^k \frac{1}{2} (X_{i-1} - X_{i+1}) \mathcal{L}_i^t. \quad (\text{A6})$$

The expected number of iterations necessary such that the remaining probability  $X_k$  reaches some reference  $X^*$  scales as

$$\mathcal{O}(n_{\text{iter}}) \approx n_{\text{live}} \log X^*, \quad (\text{A7})$$

i.e., linear in the number of live points but logarithmically in  $X^*$ .

At each iteration, we want to find a new sample with likelihood  $> \mathcal{L}_i^t$ . The efficiency of this process depends on the number of trials necessary until a sample is accepted. The simplest way to implement this is to sample the initial prior directly. However, the smaller the prior probability of the parameter space that meets this condition ( $X_i$ ), the greater the number of trials necessary for a randomly sampled point to have likelihood  $> \mathcal{L}_i^t$ . Since  $X_i$  shrinks exponentially with each iteration, drawing samples from the initial prior volume becomes exponentially less efficient. In short, the method still allocates most likelihood evaluations to uninteresting parameter space regions.

Implementations of nested sampling offer different algorithms to solve this problem. An in-depth discussion of these algorithms is beyond the scope of this paper. We refer the interested reader to [43,52].

We use a random walk, following the well-established Metropolis-Hastings algorithm [62]. Starting at a randomly chosen live point, a series of random steps (each with a randomly chosen direction) is taken. Steps are accepted only if the stepped-upon point exceeds  $\mathcal{L}_i^t$ . After some number of steps  $n_{\text{steps}}$  the current position can be taken as a new, independently sampled live point. The computational cost per iteration is constant, and, following Eq. (A7), the overall cost to reach some reference  $\log X^*$  remains linear.

Whatever method is chosen for finding the new live point, they all involve sampling the prior. The most efficient way to do this is the inverse transform method. The method takes a function  $g$  for which

$$p(g(U) = \vartheta) = p(\vartheta), \quad (\text{A8})$$

where  $U$  is a uniform random variable. Then,  $\vartheta = g(u)$  is a sample from the prior. The function  $g$  is the inverse cumulative distribution function of the prior—if it exists in closed form.

After  $k$  iterations, the remaining evidence is [Eqs. (A2) and (A6)]

$$\Delta Z_k := Z(\mathcal{R}_c) - Z_k = \int_0^{X_k} \mathcal{L}(X) dX. \quad (\text{A9})$$

An upper bound on  $\Delta Z_k$  can be set assuming that all points within the remaining volume have the same likelihood,

$$\Delta Z_k \leq \max_{\vartheta} p(\{d_{\ell}\}|\vartheta) X_k. \quad (\text{A10})$$

We can estimate  $\max_{\vartheta} p(\{d_{\ell}\}|\vartheta)$  with the highest likelihood of the current live points,  $\mathcal{L}_{\text{max},k}$ , and use Eq. (A10) as a stopping criterion for the evidence integration,

$$\frac{\Delta Z_k}{Z_k} \lesssim \frac{\mathcal{L}_{\text{max},k} X_k}{Z_k} < \delta^t, \quad (\text{A11})$$

where  $\delta^t$  is some chosen threshold. For low  $k$ , the estimate  $\mathcal{L}_{\text{max},k}$  has a large uncertainty. However, for low  $k$ , Eq. (A11) is dominated by the still large  $X_k$  and small  $Z_k$ . Once  $\mathcal{L}_{\text{max},k}$  and  $Z_k$  start to converge, Eq. (A11) is dominated by  $X_k$ , which continues to decrease exponentially.

Once converged, nested sampling yields the integrated evidence,  $Z \approx Z_{n_{\text{iter}}}$ , and a set  $\{\vartheta_i\}$  with  $i = 1, \dots, n_{\text{iter}}$ , where  $\{\vartheta_i\}$  are the live points that were *replaced* by a new sample at each iteration  $i$ . The local density of the points  $\vartheta_i$ , weighted by their relative contribution to the evidence integral

$$w_i = \frac{1}{Z} \left[ \frac{1}{2} (X_{i-1} - X_{i+1}) \mathcal{L}_i \right], \quad (\text{A12})$$

approximates the (local) posterior probability,

$$\sum_i w_i \mathbb{1}_{\Delta\vartheta}(\vartheta_i) \approx \int_{\Delta\vartheta} p(\{\vartheta_{\ell}\}|d) d\vartheta, \quad (\text{A13})$$

with  $\mathbb{1}_{\mathcal{Y}}(y) = 1$  if  $y \in \mathcal{Y}$ , and 0 otherwise. In Sec. III E, we discuss how we construct a functional approximation of the posterior density function from the weighted  $\vartheta_i$ .

## APPENDIX B: PRIORS AT THE INITIAL STAGE

The initial search stage of the stochastic framework uses uninformative priors based on Jeffreys's invariance principle. The principle demands Bayesian inference to be invariant under reparametrization of the likelihood, which can be achieved by setting [30]

$$p(\vartheta) \propto \sqrt{\det \Gamma_{ij}(\vartheta)}. \quad (\text{B1})$$

$\Gamma_{ij}(\vartheta)$  is the Fisher information matrix. If  $\vartheta$  is a parametrization for which  $\Gamma_{ij}(\vartheta)$  is constant,  $p(\vartheta)$  is a uniform distribution.

We want to find such priors for the IT1-signal model we consider, typically parametrized by frequency  $f$ , spindown  $\dot{f}$ , and the sky position in equatorial coordinates  $\alpha$  and  $\delta$  or ecliptic coordinates  $\lambda$  and  $\beta$ .

The Fisher information matrix is proportional to the  $\mathcal{F}$ -statistic metric [27], which is constant in the frequency and spin down components, but generally not in the sky. The  $f$  and  $\dot{f}$  metric components are constant: they do not depend on the parameter space point. The metric  $\alpha$  and  $\delta$  components on the other hand do depend on the parameter-space point, in particular they always depend on  $f$ . They also depend on sky position, but for observation times much longer than days, the metric is approximately constant in the sky when projected to the ecliptic plane [28]. We neglect the correlations between  $\dot{f}$  and sky.

Since the candidates' uncertainty in frequency is very small for each candidate we can consider the frequency as fixed. We will hence take uniform priors for  $f$ ,  $\dot{f}$  and  $(x, y)$ ,

which are the orthogonally projected sky points on the ecliptic plane:

We start with a uniform distribution on the unit circle ( $r \leq 1$ ) describing the ecliptic plane, parametrized in polar coordinates

$$(x, y) = (r \cos(\varphi), r \sin(\varphi)), \quad 0 \leq \varphi < 2\pi. \quad (\text{B2})$$

In this parametrization, a uniform distribution is given by

$$p(\varphi) = \frac{1}{2\pi}, \quad p(r) = 2r. \quad (\text{B3})$$

The validity is easily checked by integrating over circles with different radii.

The projection of a point  $(r \cos(\varphi), r \sin(\varphi))$  on the ecliptic plane to the surface of the unit sphere (in spherical coordinates) is achieved by

$$\lambda = \varphi \quad |\beta| = \arcsin\left(\sqrt{1 - r^2}\right), \quad (\text{B4})$$

where there are two solutions for  $\beta$ , one for each hemisphere. The coordinates  $(\lambda, \beta)$  are the ecliptic longitude and latitude, respectively.

Evidently,  $P(\lambda) = P(\varphi)$ . The distribution for  $\beta$  is given by the Jacobian of the transformation [30]. Considering only positive values of  $\beta$ ,

$$\begin{aligned} p(\beta_+) &= p(r(\beta_+)) \cdot \left| \frac{\partial r(\beta_+)}{\partial \beta_+} \right| \\ &= p(r(\beta_+)) \cdot \left| \frac{\partial}{\partial \beta_+} \left( \sqrt{1 - \sin^2(\beta_+)} \right) \right| \\ &= 2r(\beta_+) \frac{\sin(\beta_+) \cos(\beta_+)}{r(\beta_+)} \\ &= 2 \sin(\beta_+) \cos(\beta_+). \end{aligned} \quad (\text{B5})$$

Negative  $\beta$  values yield the same distribution. The symmetry implies

$$p(\beta) = p(\text{sign}(\beta)) p(\text{sign}(\beta)\beta) \quad (\text{B6})$$

$$= \frac{1}{2} \cdot 2 \sin(|\beta|) \cos(|\beta|), \quad (\text{B7})$$

and the joint probability reads

$$p(\lambda, \beta) = p(\lambda) \cdot p(\beta) = \frac{1}{2\pi} \sin(|\beta|) \cos(|\beta|). \quad (\text{B8})$$

If the distribution is constrained to some subregion  $\mathcal{R}_{\text{sky}}$ , the corresponding probability density is

$$p(\lambda, \beta | \mathcal{R}_{\text{sky}}) = \begin{cases} \frac{2\pi}{A_{\text{ecl}}(\mathcal{R}_{\text{sky}})} p(\lambda, \beta) & \text{if } (\lambda, \beta) \in \mathcal{R}_{\text{sky}}, \\ 0 & \text{else,} \end{cases} \quad (\text{B9})$$

where  $A_{\text{ecl}}(\mathcal{R}_{\text{sky}})$  is the projected area spanned by  $\mathcal{R}_{\text{sky}}$  on the ecliptic plane, summed over both hemispheres. This is simply a consequence of

$$p(x_{\text{ecl}}(\lambda, \beta), y_{\text{ecl}}(\lambda, \beta) | \mathcal{R}_{\text{sky}}) = \frac{1}{A_{\text{ecl}}(\mathcal{R}_{\text{sky}})} \quad (\text{B10})$$

being a uniform distribution by construction.

The CDF for  $\beta$  is

$$u(\beta) := \text{CDF}(\beta) = \int_{-\pi/2}^{\beta} \sin(|\beta'|) \cos(|\beta'|) d\beta' \quad (\text{B11})$$

$$\begin{aligned} &= \int_{-\pi/2}^{\beta} \text{sign}(\beta') \frac{1}{2} \sin(2\beta') d\beta' \\ &= -\frac{1}{2} \int_{-\pi/2}^0 \sin(2\beta') d\beta' \\ &\quad + \frac{1}{2} \text{sign}(\beta) \int_0^{\beta} \sin(2\beta') d\beta' \\ &= \frac{1}{2} + \frac{1}{2} \text{sign}(\beta) [-\cos^2(\beta')]_0^{\beta} \\ &= \frac{1}{2} + \frac{1}{2} \text{sign}(\beta) \sin^2(\beta), \end{aligned} \quad (\text{B12})$$

which can be inverted in closed form,

$$\beta(u) = \arcsin(\sqrt{|2u - 1|}) \text{sign}(2u - 1). \quad (\text{B13})$$

This enables sampling  $\beta$  from a uniform distribution.

If  $\beta$  is constrained to a subrange,  $\beta \in [\beta_{\min}, \beta_{\max}]$ , similar forms can be obtained by renormalizing the  $u(\beta)$ . However, we note that the constraint applied in Eq. (B9) yields a joint distribution where  $\lambda$  and  $\beta$  are not necessarily independent. In such cases, e.g., when  $\mathcal{R}_{\text{sky}}$  describes a circle on the ecliptic plane, Eq. (B13) can still be used by discarding any samples beyond  $\mathcal{R}_{\text{sky}}$ .

### APPENDIX C: THE INVERSE TRANSFORM METHOD FOR GAUSSIAN MIXTURE MODELS

In our scheme, we approximate the posterior distribution of the phase-evolution parameters  $\vartheta$  with a Gaussian mixture model,

$$p(\vartheta | \{d_e\}) \approx \sum_k m_k \mathcal{G}(\vartheta | \mu_k, \Sigma_k) \quad \text{with} \quad \sum_k m_k = 1, \quad (\text{C1})$$

where  $\mathcal{G}(\mu_k, \Sigma_k)$  are multivariate Gaussian distributions with means  $\mu_k$  and covariance matrices  $\Sigma_k$ , weighted by weights  $m_k$ . Here, we describe how such a prior can be sampled using the inverse transform method [Eq. (A8)].

For a single-mode, multivariate Gaussian, the transformation  $g$  from uniform random variables  $u \sim \mathcal{U}$  to

our-phase evolution  $\vartheta$  variables is

$$\vartheta^j = g^j(u) = [L^{-1}]_l^j F_{\mathcal{N}}^{-1}(u^l) + \mu^j, \quad (\text{C2})$$

where  $F_{\mathcal{N}}^{-1}$  is the inverse cumulative distribution function of a standard Gaussian random variable

$$F_{\mathcal{N}}^{-1}(u^l) = \sqrt{2}(\text{erf}^{-1}(2u^l - 1)), \quad (\text{C3})$$

and  $l = 1, \dots, \dim(\mathcal{N})$ ,  $\text{erf}^{-1}$  is the inverse error function, and  $L$  is the Cholesky decomposition of the inverse covariance matrix ( $\Sigma^{-1} = L^T L$ ). The Cholesky decomposition exists for positive definite matrices. While only semidefinite generally, the covariance matrices found by the Gaussian mixture model are nondegenerate. The formula uses the Einstein summation notation.

For multimodal distributions, we introduce a hyperparameter  $k = 1, \dots, N_{\text{modes}}$  identifying each mode, with prior probability  $P(k) = m_k$ , which is eventually marginalized over.

TABLE III. Sampler configuration used in this work. The configuration is oriented at the recommended settings given by the DYNesty documentation. We use DYNesty version 2.0.3. The reader is referred to its documentation for a detailed description of these options and their effect on the sampling.

Option	Setting
$\Delta \log Z$	0.1
Bounding	Multi-ellipsoids
Sampler	Random walk
Walks per iteration	25
Update interval	$2n_{\text{live}}$
Minimum efficiency	5%
Bootstrap	25
Enlarge	1

## APPENDIX D: SAMPLER CONFIGURATION

In Table III, we list the configuration of the DYNesty sampler used throughout all presented results.

- 
- [1] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. X* **13**, 041039 (2023).
- [2] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Astrophys. J. Lett.* **970**, L34 (2024).
- [3] G. Pagliaro, M. A. Papa, J. Ming, J. Lian, D. Tsuna, C. Maraston, and D. Thomas, *Astrophys. J.* **952**, 123 (2023).
- [4] K. Riles, *Living Rev. Relativity* **26**, 3 (2023).
- [5] S. J. Zhu, M. Baryakhtar, M. Alessandra Papa, D. Tsuna, N. Kawanaka, and H.-B. Eggenstein, *Phys. Rev. D* **102**, 063020 (2020).
- [6] A. Arvanitaki, M. Baryakhtar, and X. Huang, *Phys. Rev. D* **91**, 084011 (2015).
- [7] P. R. Brady, T. Creighton, C. Cutler, and B. F. Schutz, *Phys. Rev. D* **57**, 2101 (1998).
- [8] P. R. Brady and T. Creighton, *Phys. Rev. D* **61**, 082001 (2000).
- [9] B. F. Schutz and M. A. Papa, [arXiv:gr-qc/9905018](https://arxiv.org/abs/gr-qc/9905018).
- [10] B. Krishnan, A. M. Sintes, M. Alessandra Papa, B. F. Schutz, S. Frasca, and C. Palomba, *Phys. Rev. D* **70**, 082001 (2004).
- [11] C. Cutler and B. F. Schutz, *Phys. Rev. D* **72**, 063006 (2005).
- [12] V. Dergachev, *Phys. Rev. D* **85**, 062003 (2012).
- [13] B. Steltner, M. A. Papa, H.-B. Eggenstein, R. Prix, M. Bensch, B. Allen, and B. Machenschalk, *Astrophys. J.* **952**, 55 (2023).
- [14] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. D* **106**, 102008 (2022).
- [15] V. Dergachev and M. A. Papa, [arXiv:2507.12161](https://arxiv.org/abs/2507.12161).
- [16] B. McGloughlin, B. Steltner, J. Martins, M. A. Papa, H.-B. Eggenstein, J. Ming, B. Machenschalk, R. Prix, and M. Bensch, [arXiv:2508.20073](https://arxiv.org/abs/2508.20073).
- [17] B. McGloughlin, J. Martins, B. Steltner, M. A. Papa, H.-B. Eggenstein, B. Machenschalk, R. Prix, and M. Bensch, [arXiv:2508.16423](https://arxiv.org/abs/2508.16423).
- [18] M. A. Papa *et al.*, *Phys. Rev. D* **94**, 122006 (2016).
- [19] J. Ming, M. A. Papa, H.-B. Eggenstein, B. Beheshtipour, B. Machenschalk, R. Prix, B. Allen, and M. Bensch, *Astrophys. J.* **977**, 154 (2024).
- [20] P. B. Covas, R. Prix, and J. Martins, *Phys. Rev. D* **110**, 024053 (2024).
- [21] R. Tenorio, D. Keitel, and A. M. Sintes, *Phys. Rev. D* **104**, 084012 (2021).
- [22] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **103**, 064017 (2021).
- [23] P. B. Covas, M. A. Papa, and R. Prix, *Astrophys. J.* **985**, 192 (2025).
- [24] P. Jaranowski, A. Królak, and B. F. Schutz, *Phys. Rev. D* **58**, 063001 (1998).
- [25] V. Dergachev and M. A. Papa, *Phys. Rev. D* **103**, 063019 (2021).
- [26] C. Dreissigacker, R. Prix, and K. Wette, *Phys. Rev. D* **98**, 084058 (2018).
- [27] R. Prix, *Phys. Rev. D* **75**, 023004 (2007).
- [28] K. Wette, *Phys. Rev. D* **92**, 082003 (2015).
- [29] P. B. Covas and R. Prix, *Phys. Rev. D* **112**, 044041 (2025).
- [30] A. Gelman *et al.*, *Bayesian Data Analysis*, 3rd ed., Texts in Statistical Science Series (CRC Press, Taylor and Francis Group, Boca Raton, London, New York, 2014).
- [31] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Astrophys. J.* **935**, 1 (2022).
- [32] G. Ashton and R. Prix, *Phys. Rev. D* **97**, 103020 (2018).
- [33] A. Ashok, P. B. Covas, R. Prix, and M. A. Papa, *Phys. Rev. D* **109**, 104002 (2024).

- [34] LIGO Scientific Collaboration, LALSuite: LIGO Scientific Collaboration algorithm library suite (2020), [10.7935/GT1W-FZ16](https://arxiv.org/abs/10.7935/GT1W-FZ16).
- [35] K. Wette, *SoftwareX* **12**, 100634 (2020).
- [36] B. Steltner, T. Menne, M. A. Papa, and H.-B. Eggenstein, *Phys. Rev. D* **106**, 104063 (2022).
- [37] B. Beheshtipour and M. A. Papa, *Phys. Rev. D* **101**, 064009 (2020).
- [38] B. Beheshtipour and M. A. Papa, *Phys. Rev. D* **103**, 064027 (2021).
- [39] A. C. Searle, [arXiv:0804.1161](https://arxiv.org/abs/0804.1161).
- [40] R. Prix, S. Giampanis, and C. Messenger, *Phys. Rev. D* **84**, 023007 (2011).
- [41] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, USA, 2007).
- [42] J. Skilling, *AIP Conf. Proc.* **735**, 395 (2004).
- [43] G. Ashton *et al.*, *Nat. Rev. Methods Primers* **2**, 1 (2022).
- [44] C. Messenger, R. Prix, and M. A. Papa, *Phys. Rev. D* **79**, 104017 (2009).
- [45] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96 (AAAI Press, Portland, Oregon, 1996), pp. 226–231.
- [46] Gaussian mixture models, <https://scikit-learn.org/stable/modules/mixture.html>, accessed: 2025-12-08.
- [47] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA, Collaborations), *Astrophys. J. Suppl. Ser.* **267**, 29 (2023).
- [48] B. Behnke, M. A. Papa, and R. Prix, *Phys. Rev. D* **91**, 064007 (2015).
- [49] B. Steltner, M. A. Papa, and H.-B. Eggenstein, *Phys. Rev. D* **105**, 022005 (2022).
- [50] D. Keitel, R. Prix, M. A. Papa, P. Leaci, and M. Siddiqi, *Phys. Rev. D* **89**, 064023 (2014).
- [51] B. Allen, *Phys. Rev. D* **100**, 124004 (2019).
- [52] J. S. Speagle, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [53] E. Goetz *et al.*, LIGO-T2100200-v2: O3 lines and combs in found in self-gated C01 data, <https://dcc.ligo.org/LIGO-T2100200/public> (2021), date accessed: 2025-08-06.
- [54] P. Covas *et al.* (LSC Collaboration), *Phys. Rev. D* **97**, 082002 (2018).
- [55] D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall, *Classical Quantum Gravity* **36**, 055011 (2019).
- [56] D. Davis *et al.* (LIGO Collaboration), *Classical Quantum Gravity* **38**, 135014 (2021).
- [57] E. Capote *et al.*, *Phys. Rev. D* **111**, 062002 (2025).
- [58] D. Shoemaker, N. Arnaud, and T. Sawada, LIGO-G2002127-v33: Observing Scenario timeline graphic, post-O3, <https://dcc.ligo.org/LIGO-G2002127/public> (2025), date accessed: 2025-12-08.
- [59] G. Ashton *et al.*, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [60] L. Mirasola and R. Tenorio, *Phys. Rev. D* **110**, 124049 (2024).
- [61] F. Pedregosa *et al.*, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- [62] W. K. Hastings, *Biometrika* **57**, 97 (1970).