



This is the accepted manuscript made available via CHORUS. The article has been published as:

Application of a hierarchical MCMC follow-up to Advanced LIGO continuous gravitational-wave candidates

Rodrigo Tenorio, David Keitel, and Alicia M. Sintes

Phys. Rev. D **104**, 084012 — Published 1 October 2021

DOI: [10.1103/PhysRevD.104.084012](https://doi.org/10.1103/PhysRevD.104.084012)

Application of a hierarchical MCMC follow-up to Advanced LIGO continuous gravitational-wave candidates

Rodrigo Tenorio,^{*} David Keitel, and Alicia M. Sintes

*Departament de Física, Institut d'Aplicacions Computacionals i de Codi Comunitari (IAC3),
Universitat de les Illes Balears, and Institut d'Estudis Espacials de Catalunya (IEEC),
Carretera de Valldemossa km 7.5, E-07122 Palma, Spain*

We present the first application of a hierarchical Markov Chain Monte Carlo (MCMC) follow-up on continuous gravitational-wave candidates from real-data searches. **The follow-up uses an MCMC sampler to draw parameter-space points from the posterior distribution, constructed using the matched-filter as a log-likelihood.** As outliers are narrowed down, coherence time increases, imposing more restrictive phase-evolution templates. We introduce a novel Bayes factor to compare results from different stages: The signal hypothesis is derived from first principles, while the noise hypothesis uses extreme value theory to derive a background model. The effectiveness of our proposal is evaluated on fake Gaussian data and applied to a set of 30 outliers produced by different continuous wave searches on O2 Advanced LIGO data. The results of our analysis suggest **all but five outliers are inconsistent with an astrophysical origin under the standard continuous wave signal model. We successfully ascribe four of the surviving outliers to instrumental artifacts and a strong hardware injection present in the data. The behavior of the fifth outlier suggests an instrumental origin as well, but we could not relate it to any known instrumental cause.**

I. INTRODUCTION

Continuous gravitational waves (CWs) are persistent forms of gravitational radiation. These yet-to-be detected signals are orders of magnitude weaker than compact binary coalescences [1], requiring long integration times (months to years) to differentiate them from noise. **Potentially detectable sources** using the current generation of ground-based interferometric detectors, Advanced LIGO [2] and Advanced Virgo [3], are neutron stars (NSs) presenting some non-axisymmetry such as crustal deformations, r-mode instabilities or free precession [4], or the annihilation of ultralight boson clouds around spinning black holes [5].

Searching for a CW consists in filtering a data stream against a set of **signal templates**, each of which is related to a certain set of parameters describing the CW model being searched for. The number of templates required to properly cover a certain parameter space region, however, scales as a **large** power of observing time [6]. At a fixed computing cost, the optimal strategy is to split the data stream into segments on which the filtering is performed, and then combine the resulting statistics [7, 8]. Since phase information is only fully preserved within each of these segments, they are usually referred to as *coherent segments* spanning a certain *coherence time*.

The approach taken by current implementations of wide parameter space searches such as [9–14] lies in the middle ground. **Wide parameter space regions are analyzed using a relatively low coherence time, ranging from half an hour to a few weeks.** Surviving outliers are then sieved through a suite of vetoes testing their (in)consistency with a CW signal; this includes studying their persistence over the data stream,

comparing their significance in different detectors or checking whether they cross a frequency band containing known instrumental artifacts [15–21]. Other common strategies are coincidence analyses between detectors or clustering neighbouring outliers in order to relate them to a common cause [22–25]. Finally, if there are any surviving outliers, various follow-up strategies use longer coherence times [21, 26–31], either in a single stage or in a hierarchical scheme where candidates are narrowed down over a “ladder” of coherence times.

Large-scale CW searches would benefit from a simple, general hierarchical setup, as it would allow for the systematic follow-up of CW outliers using longer coherence times, imposing tighter constraints and reducing the presence of outliers due to background noise.

Here we present the first complete framework to conduct hierarchical Markov Chain Monte Carlo (MCMC) follow-ups and its application to a set of outliers obtained by different CW search pipelines on Advanced LIGO O2 data. Our work builds on top of [32], which introduced the MCMC follow-up of CW outliers and studied its performance on simulated signals in pure Gaussian noise. We propose a new hypothesis test for the presence of a signal in the data after the full follow-up procedure. The probability of the signal hypothesis is derived from first principles as proposed in [33]; the probability of the noise hypothesis is derived from the application of extreme value theory. We demonstrate the general applicability of this follow-up strategy by analyzing outliers stemming from different analysis pipelines.

Although we restrict ourselves to outliers from CW searches for unknown *isolated* sources, this framework and the corresponding software [34] can also be applied to outliers from searches for sources in binary systems [21, 30], glitching NSs [35] and long-duration gravitational-wave transients [36, 37].

The paper is organized as follows: Section II describes

^{*} rodrigo.tenorio@ligo.org

the basic tools of CW data analysis and overviews the application of MCMC samplers to the follow-up problem; Sec. III introduces a new statistic in terms of hypothesis testing; Sec. IV introduces the O2 outliers to be analyzed and the follow-up setup. The results are presented in Sec. V, concluding in Sec. VI. We briefly comment on the statistical properties of the maximum \mathcal{F} -statistic over correlated templates in appendix A.

II. CONTINUOUS-WAVE DATA ANALYSIS: SEARCH AND FOLLOW UP

A CW signal can be parametrized in terms of two families of parameters, namely the *phase-evolution parameters* λ and the *amplitude parameters* \mathcal{A} . This separation is motivated by the response of a GW detector to such signals

$$h(t; \lambda, \mathcal{A}) = \sum_{\mu=0}^3 \mathcal{A}^\mu h_\mu(t; \lambda), \quad (1)$$

where the functions \mathcal{A}^μ are independent of time [38].

The search for a CW signal can be stated in a Bayesian framework as a hypothesis test between the noise hypothesis \mathcal{H}_G , under which the data consists of Gaussian noise $n(t)$, and the signal hypothesis $\mathcal{H}_S(\lambda, \mathcal{A})$, supporting the presence of a CW signal with a defined set of parameters within said noise $n(t) + h(t; \lambda, \mathcal{A})$. The support of a stream of data x for either of these hypotheses is quantified by the *Bayes factor* [39]

$$B_{S/G}(x; \lambda, \mathcal{A}) = \frac{P(x|\mathcal{H}_S(\lambda, \mathcal{A}))}{P(x|\mathcal{H}_G)}. \quad (2)$$

Following [40, 41], and motivated by the linear dependency of Eq. (1) on the amplitude functions \mathcal{A}^μ , one can choose an appropriate set of priors $P(\mathcal{A})$ such that Eq. (2) can be analytically marginalized:

$$B_{S/G}(x; \lambda) = \int d\mathcal{A} B_{S/G}(x; \lambda, \mathcal{A}) P(\mathcal{A}) \propto e^{\mathcal{F}(x; \lambda)}. \quad (3)$$

The statistic \mathcal{F} , which depends only on the data and the phase parameters, was originally derived as the maximum-likelihood estimator with respect to \mathcal{A} [38, 42]. This is a general detection statistic which only relies on the waveform decomposition presented in Eq. (1) and hence can be applied also to variations of the CW signal model such as sources in binaries [43] and transients [36]. Furthermore, the methods developed in this work can also be applied to CW outliers from any kind of search using a different detection statistic, as long as they can be associated with a parameter-space point with a certain uncertainty.

The role of Eq. (3) is to update the prior probability on the phase evolution parameters $P(\lambda)$ by means of the information conveyed by the data stream x . This can be stated in terms of Bayes' theorem as

$$P(\lambda|x, \mathcal{H}_S) \propto B_{S/G}(x; \lambda) P(\lambda). \quad (4)$$

We note that $B_{S/G}$ and \mathcal{F} have the same statistical power as they are related by a strictly monotonic function. For the sake of later consistency, we will focus on \mathcal{F} from now on. We refer the reader to [36, 40, 44, 45] for a more in-depth analysis of these statistics.

The detection problem is now stated in terms of a maximization: Given a stream of data x , we are interested in finding the phase-evolution parameters λ (also referred to as *templates*) which maximize Eq. (4) or, equivalently, $\mathcal{F}(x; \lambda)$.

A. Coherent and semicoherent searches

The fully-coherent \mathcal{F} -statistic can be expressed in terms of a linear filter between the data stream and a signal template,

$$\tilde{\mathcal{F}}(\lambda) \propto |\langle x, h(\lambda) \rangle|^2, \quad (5)$$

where $\langle \cdot \rangle$ represents a functional scalar product. Throughout this work, and following the convention of [46], fully-coherent quantities will be represented with a tilde; semicoherent quantities, introduced in Eq. (7), will be represented with a caret. The response of $\tilde{\mathcal{F}}$ to an offset $\Delta\lambda$ in the phase-evolution parameters λ is quantified using the *mismatch* [47], which can be defined in terms of a local quadratic approximation around the true signal parameters λ where the mismatch has a minimum:

$$m(\Delta\lambda; \lambda) = \frac{\tilde{\mathcal{F}}(\lambda) - \tilde{\mathcal{F}}(\lambda + \Delta\lambda)}{\tilde{\mathcal{F}}(\lambda)} \simeq \Delta\lambda^T \cdot \bar{g} \cdot \Delta\lambda + \mathcal{O}(\Delta\lambda^3). \quad (6)$$

The symmetric tensor \bar{g} is referred to as the parameter-space metric, and can be used to set up parameter-space coverings, also known as template banks, at a certain mismatch level. This quadratic approximation is known to be valid up to $m \lesssim 0.3 - 0.5$, although latest developments on this subject suggest to further extend the approximation up to $m \sim 1$ [48–50].

Maximizing Eq. (5) poses a computational challenge, as the number of templates to be considered in the optimization scales with a large power of the total length of the data stream [6, 51], while the sensitivity only scales as the square root of it [52, 53]. As discussed in [13, 54], such a strong scaling stems from the tight restrictions imposed by the \mathcal{F} -statistic on the signal model, requiring phase coherence over the whole duration of the data stream. A looser statistic can be constructed by imposing said coherence in a segment-wise manner. To do so, the data stream, spanning a time of T_{obs} , is divided into N_{seg} segments, each of them with a duration of T_{coh} . The *semicoherent* \mathcal{F} -statistic is then constructed by adding the coherent \mathcal{F} -statistics computed in each segment

$$\hat{\mathcal{F}}(\lambda) = \sum_{n=0}^{N_{\text{seg}}-1} \tilde{\mathcal{F}}_n(\lambda), \quad (7)$$

where $\tilde{\mathcal{F}}_n$ refers to the coherent \mathcal{F} -statistic computed using only data within segment n . This approach

uncorrelates the template's phase-evolution between consecutive coherent segments, loosening the constraints imposed on the data and widening \mathcal{F} -statistic peaks in the parameter space [32, 54]. In other words, given a parameter-space coordinate volume, the number of templates required to cover it at a given mismatch *decreases* with lower T_{coh} . This implies a dependency of the parameter space metric \bar{g} on T_{coh} .

The optimal strategy to sweep a wide parameter-space region under a controlled computational budget is then to use a hierarchical scheme with a varying T_{coh} : The first stage surveys a parameter-space region with $T_{\text{coh}} \ll T_{\text{obs}}$, using an affordable number of templates. \mathcal{F} -statistic outliers are then analyzed with an increased coherence time, further narrowing down the parameter-space region of interest. This process continues either until $T_{\text{coh}} = T_{\text{obs}}$ or the candidate is vetoed by a complementary procedure [8, 28, 32, 46].

B. MCMC-based follow-ups

The follow-up of CW outliers requires to set up a template bank across the parameter-space region of interest. Typical gridded approaches use a parameter-space metric to cover the parameter space at fixed maximum mismatch [27, 28, 55]. This approach usually requires an extensive campaign of software injections to be performed in order to calibrate the optimal set up in terms of sensitivity and computing cost [28].

Alternatively, one could view the problem from the point of view of Bayesian inference. Eq. (4) relates the \mathcal{F} -statistic to a posterior probability distribution. This distribution can be sampled using a MCMC method, effectively constructing an adaptative random template bank in the parameter space in which \mathcal{F} -statistic values will be more densely evaluated around high posterior probability regions. As first discussed in [32], this approach achieves close to the theoretical optimal sensitivity for signals in Gaussian noise as long as the parameter space region is small enough to ensure a good convergence of the MCMC.

For the purpose of estimating the effectiveness of an MCMC, as discussed in [32], the effective size of a parameter-space region can be computed in terms of the *number of templates* \mathcal{N} required to cover it at a mismatch of unity using a lattice with unit normalized thickness [51, 56]

$$\mathcal{N}(T_{\text{coh}}, \Delta\lambda) = \int_{\Delta\lambda} d\lambda \sqrt{g(T_{\text{coh}})}, \quad (8)$$

where $g(T_{\text{coh}})$ is the determinant of the parameter-space metric, which depends on T_{coh} as explained in Sec. II A, and $\Delta\lambda$ represents the region being followed up. The integral in Eq. (8) must be computed along the resolved parameter-space dimensions only; i.e., one should *not* include fractional templates, as doing so would underestimate the actual number of templates [8, 43, 46]. For a follow-up search, the parameter-space

region under analysis is typically smaller than the scale of parameter-space correlations, meaning \sqrt{g} can be taken out of the integral as a constant and Eq. (8) simplifies to

$$\mathcal{N}(T_{\text{coh}}, \Delta\lambda) \simeq \sqrt{g(T_{\text{coh}})} \text{Vol}(\Delta\lambda) \quad (9)$$

where $\text{Vol}(\Delta\lambda)$ is the coordinate volume of the region being followed up. Seminal analyses in [32] and follow-up searches performed in [21, 30] suggest that values up to $\mathcal{N}^* \simeq 10^{3-4}$ are compatible with effective MCMC runs in terms of convergence.

CW outliers are identified as a parameter-space point carrying an uncertainty which depends on the pipeline used to conduct the search. Upon entering the follow-up pipeline, these uncertainties are converted into prior probability distributions to start the MCMC sampling. Ref. [32] proposed the use of bounded uniform priors in order to restrict the surveyed parameter-space region; however, such hard boundaries may prevent the successful follow-up of CW candidates whose parameters are shifted due to the presence of parameter-space correlations. We propose the use of uncorrelated Gaussian priors, which concentrate their probability density around a characteristic region while being unbounded. See Sec. V for details on the choice of Gaussian priors.

An MCMC-based follow-up is implemented in the `PyFstat` package [34] using the parallel-tempered ensemble MCMC sampler `ptemcee` [57, 58] to sample the posterior distribution Eq. (4) using either the coherent (Eq. 5) or semicoherent (Eq. 7) \mathcal{F} -statistic. We refer the reader to [32] for an extended discussion on the characteristics of this particular MCMC implementation. The analyses presented in this work were performed using `PyFstat` version 1.11.3 [59].

C. A coherence-time ladder

Early setups of hierarchical schemes were based on the optimization of computing resources in order to achieve a prescribed level of sensitivity [7, 8]. Alternatively, if no computational cost model was available, software injection campaigns were used to calibrate the number of stages [28]. For the case of an MCMC-based follow-up, one can use the quantity $\mathcal{N}(T_{\text{coh}}, \Delta\lambda)$ to design a hierarchical scheme by imposing the proper convergence of the MCMC run at each stage [32].

Suppose a wide parameter-space semicoherent search produces an interesting outlier in the parameter-space region $\Delta\lambda^{(0)}$, where the exact shape is entirely dependent on the pipeline. **Round-bracketed superindices denote different stages of the follow-up.** To set up a first follow-up stage, we choose a coherence time $T_{\text{coh}}^{(0)}$ such that $\mathcal{N}(T_{\text{coh}}^{(0)}, \Delta\lambda^{(0)}) \lesssim \mathcal{N}^*$, ensuring the effective parameter-space resolution is coarse enough for the MCMC algorithm to properly converge towards the region of interest. If successful, the resulting parameter-space region will be narrower, $\Delta\lambda^{(1)} \leq \Delta\lambda^{(0)}$, and a second

MCMC stage using a new coherence time $T_{\text{coh}}^{(1)}$ will be applied. This procedure is repeated until $T_{\text{coh}} = T_{\text{obs}}$ and a final fully-coherent follow-up is performed.

In [32], a simple method was proposed to find the coherence time for a stage j given the previous stage's results. The idea is to increase the coherence time as much as possible such that the MCMC is able to converge to the target distribution. Since this convergence can be quantified in terms of a maximum number of templates within a region \mathcal{N}^* , the new coherence time $T_{\text{coh}}^{(j)}$ can be obtained by solving

$$\mathcal{N}(T_{\text{coh}}^{(j)}, \Delta\lambda^{(j)}) = \mathcal{N}^*. \quad (10)$$

This choice minimizes the number of stages in the scheme, reducing the overall computing cost, while ensuring the effectiveness of the MCMC approach. The explicit dependency of Eq. (10) on the parameter-space region under analysis $\Delta\lambda^{(j)}$, however, hinders the construction of a complete hierarchical scheme.

This dependency can be removed by noticing the inherent self-similarity of MCMC stages: A successful MCMC follow-up stage ends up with a set of samples around a prominent global maximum, the fine structure of which is *underresolved* because of the chosen coherence time. By progressing to the next stage, this fine structure gets resolved and the MCMC *zooms in* further towards the parameter-space maximum. The setup of a coherence-time ladder is simply a problem of minimizing the number of stages to reduce computing cost while maintaining sufficiently big underresolved regions for the MCMC follow-up to properly sample the region of interest. This condition can be simply expressed as $\mathcal{N}(T_{\text{coh}}^{(j)}, \Delta\lambda^{(j+1)}) \simeq 1$; hence, comparing consecutive stages factors out the problematic dependency and the hierarchical scheme can be constructed by solving the recurrence

$$\mathcal{N}^* \simeq \frac{\mathcal{N}(T_{\text{coh}}^{(j+1)}, \Delta\lambda^{(j+1)})}{\mathcal{N}(T_{\text{coh}}^{(j)}, \Delta\lambda^{(j+1)})} = \frac{\sqrt{g(T_{\text{coh}}^{(j+1)})}}{\sqrt{g(T_{\text{coh}}^{(j)})}} \quad (11)$$

given $T_{\text{coh}}^{(0)}$ and \mathcal{N}^* . A numerical solver for Eq. (11) is included in the PyFstat package [34]. Constructing the coherence-time ladder as proposed by [32] makes use of the so-called SuperSky metric [60, 61] to compute the parameter-space volume element. This metric is numerically well-conditioned, but requires $T_{\text{coh}} \gtrsim 1$ day.

Alternatively, we derive an equivalent coherence-time ladder by considering the parameter-space volume reduction from one stage to the next. Let us define

$$\gamma^{(j+1)} = \frac{\text{Vol}(\Delta\lambda^{(j)})}{\text{Vol}(\Delta\lambda^{(j+1)})} \quad (12)$$

as the parameter-space volume *shrinkage* from stage j to stage $j+1$. In a practical application, this quantity can be computed by comparing the volume containing a certain

amount of posterior probability from two consecutive stages.

Eq. (10) can now be re-expressed as

$$1 = \frac{\mathcal{N}(T_{\text{coh}}^{(j+1)}, \Delta\lambda^{(j+1)})}{\mathcal{N}(T_{\text{coh}}^{(j)}, \Delta\lambda^{(j)})}, \quad (13)$$

and Eq. (11) is generalized by including Eq. (12)

$$\gamma^{(j+1)} = \frac{\sqrt{g(T_{\text{coh}}^{(j+1)})}}{\sqrt{g(T_{\text{coh}}^{(j)})}}, \quad (14)$$

where we can recognize $\gamma^{(j+1)}$ as a generalized version of the refinement factor γ introduced in Eq. (73) of [62] to account for the template bank refinement from a semicoherent stage to a fully-coherent one. To fully recover Eq. (11), we simply set $\gamma^{(j+1)} = \mathcal{N}^*$ in every stage j .

According to this derivation, constructing a coherence ladder is equivalent to imposing a ratio of posterior volume shrinkage. For example, choosing $\mathcal{N}^* \simeq 10^4$ is equivalent to imposing an overall volume shrinkage of $\gamma \simeq 10^4$ (i.e. posterior volume is a ten-thousandth fraction of the prior volume) at each step of the ladder. As a result, the behaviour of an MCMC stage is dependent upon its capability to fulfill the required shrinkage rate.

III. EVALUATING THE HIERARCHICAL FOLLOW-UP WITH A BAYES FACTOR

A multi-stage MCMC follow-up analyzes CW outliers by converging towards parameter-space regions with a high posterior probability. After each stage, coherence time is increased, breaking up underresolved regions into smaller ones and allowing the MCMC to further narrow down the parameters associated to the loudest outlier. We are interested in evaluating the significance of the loudest template resulting from the multi-stage follow-up by comparing its actual fully-coherent \mathcal{F} -statistic to the expected value predicted by a previous stage of the ladder.

We construct a new Bayes factor for this comparison, using the fully-coherent \mathcal{F} -statistic of the loudest candidate of the MCMC, $2\tilde{\mathcal{F}}^*$, in order to quantify the support for the presence or lack of a CW signal in the data. Following the definition in Eq. (2),

$$\ln \mathcal{B}_{\text{S/N}}^* = \ln \frac{P(2\tilde{\mathcal{F}}^*|\mathcal{H}_{\text{S}})}{P(2\tilde{\mathcal{F}}^*|\mathcal{H}_{\text{N}})}, \quad (15)$$

where the hypotheses \mathcal{H}_{S} and \mathcal{H}_{N} correspond to the presence or lack of a signal, respectively. As discussed in Sec. III A, the use of extreme value theory allows us to formulate \mathcal{H}_{N} such that it is not restricted to Gaussian noise, but includes any exponentially-bounded

distributions with unbounded domain. The following subsections are devoted to deriving the probability distributions under each of these hypotheses.

A. Noise hypothesis

The noise hypothesis \mathcal{H}_N ascribes the obtained value of $2\tilde{\mathcal{F}}^*$ to pure noise. Under the presence of Gaussian noise, the coherent \mathcal{F} -statistic follows a chi-squared distribution with 4 degrees of freedom¹, $2\tilde{\mathcal{F}} \sim \chi_4^2$. If we consider the resulting MCMC samples as a template bank $\{\lambda\}$, it is clear that $2\tilde{\mathcal{F}}^* = \max_{\lambda \in \{\lambda\}} 2\tilde{\mathcal{F}}(\lambda)$ and the corresponding probability distribution is that of the maximum over a certain number of templates n [63]:

$$P(\max 2\tilde{\mathcal{F}}) = n \cdot \chi_4^2(\max 2\tilde{\mathcal{F}}) \cdot \left[\int_0^{\max 2\tilde{\mathcal{F}}} d\xi \chi_4^2(\xi) \right]^{n-1}, \quad (16)$$

where χ_4^2 denotes the probability density function. The argument equally holds for the case of the *semicoherent* \mathcal{F} -statistic; in that case, however, the number of degrees of freedom of the chi-squared distribution would be $4N_{\text{seg}}$.

By construction, the *effective* number of templates in a CW template bank is different from the actual number of templates. This is because template banks are set up such that no parameter-space point is further than a certain mismatch m from a template in the bank, implying a certain degree of correlation among neighbouring templates [56]. The problem of estimating the effective number of templates in a template bank has not found a definitive solution in the CW literature.

A common approach, see e.g. [64], is to evaluate the template bank on several realizations of Gaussian noise to numerically sample the probability distribution of the loudest outlier; the effective number of templates is then obtained by fitting n from Eq. (16) to the data. Another approach, firstly proposed in [65], splits the results of a wide parameter-space search into disjoint partitions such that they are equivalent to different realizations of a smaller search. The fraction of effective templates can be fitted using Eq. (16) to the loudest outlier per partition, obtaining n through extrapolation. Further developments on this method proposed a non-parametric ansatz to directly estimate the distribution of the loudest candidate of a search [66].

Here we will use a solution based on extreme value theory, which describes the three possible asymptotic distributions followed by the maximum of n independent trials according to the tail of their individual probability distribution. Short-scale correlated variables, such as the

ones arising in the search for CW signals, are also covered by the theory [67]. The family of three distributions, usually referred to as the *generalized extreme value distribution*, is parametrized by a single parameter $c \in \mathbb{R}$ (aside from the location and scale parameters), and encompasses every possible *max-stable distribution*: the maximum value of a set of random variables following a generalized extreme value distribution follows itself a generalized extreme value distribution of the same class, albeit with different parameters. Each of the three possible distributions is related to c being positive, null or negative, and encloses a different set of probability distributions in its domain of attraction [68–70].

For our CW application, we focus on the case $c = 0$, also known as the *Gumbel distribution*

$$\text{Gumbel}(\xi; \mu, \sigma) = \frac{1}{\sigma} \exp \left[- \left(\frac{\xi - \mu}{\sigma} \right) - e^{-\left(\frac{\xi - \mu}{\sigma} \right)} \right], \quad (17)$$

where μ and σ are its location and scale parameters, respectively. The domain of attraction of this distribution comprises a variety of exponentially bounded distributions, including the chi-squared distribution. A similar procedure could be carried out for the other two families $c \neq 0$, including power-law and finite tails, if the behavior of the background noise required so. This argument is consistent with the empirical proposal of [71].

As noted in Appendix D of [53], the presence of correlated templates renders Eq. (16) unsuitable to describe the background noise distribution of CW searches. This is because the family of Gumbel distributions spanned by Eq. (16) as $n \rightarrow \infty$ has a *fixed* scale parameter $\sigma = 2$. The inclusion of correlated templates makes the underlying distribution deviate from a chi-squared [72], but exponential tails still allow the distribution of the maxima to be described by a Gumbel distribution but with $\sigma \neq 2$. Further discussion on this topic is presented in Appendix A.

As a result, we construct $P(2\tilde{\mathcal{F}}^*|\mathcal{H}_N)$ by fitting *both* the location and scale parameters of a Gumbel distribution to the background distribution associated to $2\tilde{\mathcal{F}}^*$

$$P(2\tilde{\mathcal{F}}^*|\mathcal{H}_N) = \frac{1}{\sigma_N} \exp \left[- \left(\frac{2\tilde{\mathcal{F}}^* - \mu_N}{\sigma_N} \right) - e^{-\left(\frac{2\tilde{\mathcal{F}}^* - \mu_N}{\sigma_N} \right)} \right]. \quad (18)$$

This approach has the advantage of circumventing the computation of an effective number of templates by directly using the asymptotic distribution, the functional form of which is robust as long as the individual distribution tails fall off exponentially. The typical number of templates evaluated in an MCMC follow-up is consistent with a good convergence of the maximum distribution towards a Gumbel [73]. Further discussion on the suitable application of extreme value theory to evaluate the loudest outlier of a gravitational-wave search will be presented elsewhere [74].

To estimate the scale and location parameters of the background distribution μ_N, σ_N , we apply the *off-sourcing* procedure, the effectiveness of which was studied in [75]. *Off-sourcing consists in evaluating the*

¹ We recall for the sake of consistency with the statistics literature that a chi-squared distribution with ν degrees of freedom corresponds to a Gamma distribution with shape parameter $k = \nu/2$ and scale parameter $\theta = 2$.

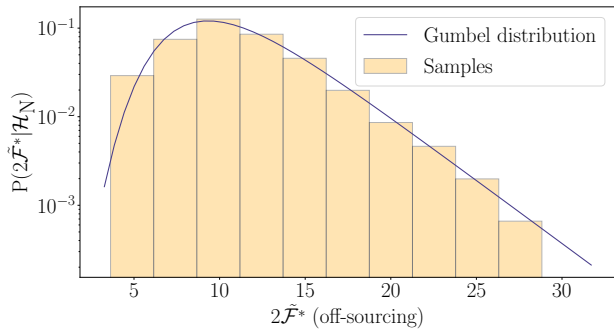


FIG. 1. Distribution of the maximum $2\mathcal{F}$ value of a template bank obtained from its evaluation at $N_o = 600$ different off-sourced right ascensions, excluding 90° around the sky position of the outlier of interest. The template bank corresponds to MCMC samples from the fully-coherent stage follow-up of a simulated signal in Gaussian noise. The solid line represents the fit of a Gumbel distribution.

\mathcal{F} -statistic on a template bank whose sky positions have been purposely *shifted* with respect to that of the outlier of interest. This blinds the detection statistic to the outlier under analysis while still sampling the same background distribution from the dataset. Incidentally, this takes into account template-bank correlations induced by non-Gaussian noise components. These correlations do not arise due to different templates sampling the same spectrogram data (i. e. overlapping frequency-evolution tracks) [25, 65], but due to the presence of correlated spectrogram data spanning different iso-mismatch ellipsoids in the parameter space. The former kind is fundamental in the sense that it is independent of the background; the latter is entirely dependent upon the observed data: the wider the bandwidth of the disturbance, the lower the number of effective independent templates.

In our concrete application, we produce $N_o = 600$ off-sourced template banks by randomly shifting the template's right ascension (azimuthal spherical angle), excluding a 90° region around the sky position of interest. The declination (polar spherical angle) is unchanged in order to maintain a constant level of sensitivity in terms of \mathcal{F} -statistic values. Fig. 1 shows an example of a background noise distribution obtained through this procedure.

The evaluation of off-sourced template banks represents the main contribution to the computing cost of the follow-up. The small number of outliers evaluated in this work allowed us to evaluate a set of off-sourced samples for each of them. For the case of a large-scale follow-up, however, one could benefit from the general properties of the Gumbel distribution to re-use a set of Gumbel parameters for different parameter-space regions, lowering the overall computing cost.

B. Signal hypothesis

The presence of a signal is characterized by its (squared) signal-to-noise ratio (SNR) ρ^2 , which gauges the (squared) amplitude of a signal against that of the background noise [9, 38, 42, 47]. Exact expressions for ρ^2 , which include amplitude-modulation effects due to the antenna pattern of the detectors, are available in [38, 52, 53]. The effect of this parameter on the probability distribution of the \mathcal{F} -statistic is to shift the chi-squared distribution towards a non-central chi-squared distribution, $2\mathcal{F} \sim \chi^2_{4N_{\text{seg}}}(\rho^2)$, where the fully-coherent case corresponds to $N_{\text{seg}} = 1$.

As previously discussed, we are interested in comparing the consistency of $2\tilde{\mathcal{F}}^*$ to the values $2\hat{\mathcal{F}}^*$ obtained in a previous stage of the ladder. Any semicoherent stage of the ladder can be used to construct a signal hypothesis; as discussed in more detail in Sec. V A, we select the second-to-last stage in order to benefit from the tighter constraints imposed by the signal model. For the remainder of this section we simplify our notation by removing the asterisks, assuming every \mathcal{F} -statistic value refers to that of the loudest candidate from the fully-coherent stage.

We construct $P(2\tilde{\mathcal{F}}|\mathcal{H}_S)$ following the developments of [33]. The basic idea goes as follows: Assume a single-template search perfectly matching a signal is performed. The presence of a signal in the data, characterized by ρ^2 , produces an \mathcal{F} -statistic value which depends only on ρ^2 and the number of coherent segments N_{seg} . More specifically, obtaining a value of $2\hat{\mathcal{F}}$ on N_{seg} segments automatically produces an estimate on ρ^2 , which, in turn, yields an estimation of the expected $2\tilde{\mathcal{F}}$ that will be retrieved after performing a fully-coherent search.

The exact flow of information from the semicoherent to the coherent statistic can be readily expressed by marginalizing over the unknown non-centrality parameter ρ^2

$$\begin{aligned} P(2\tilde{\mathcal{F}}|\mathcal{H}_S) &= \int_0^\infty d\rho^2 P(2\tilde{\mathcal{F}}|\rho^2, 2\hat{\mathcal{F}}, N_{\text{seg}}) P(\rho^2|2\hat{\mathcal{F}}, N_{\text{seg}}) \\ &\propto \int_0^\infty d\rho^2 P(2\tilde{\mathcal{F}}|\rho^2) P(2\hat{\mathcal{F}}|\rho^2, N_{\text{seg}}) P(\rho^2), \end{aligned} \quad (19)$$

where constant factors with respect to $2\hat{\mathcal{F}}$ and N_{seg} were omitted and the same data is being used to compute both statistics.² The choice of a prior distribution on ρ^2 depends on the type of search carried out; for a wide parameter-space search such as the ones in which we are interested it is enough to consider an improper uniform prior.

In going to the second line in Eq. (19) we have assumed no dependency between $2\hat{\mathcal{F}}$ and $2\tilde{\mathcal{F}}$ in the sense of

² This corresponds to $\kappa = 1$ in the notation of [33].

$P(2\tilde{\mathcal{F}}|\rho^2, 2\hat{\mathcal{F}}, N_{\text{seg}}) = P(2\tilde{\mathcal{F}}|\rho^2)$. This relation holds exactly if one computes each statistic on a different dataset, corresponding to the *fresh data mode* in [8]. On the other hand, if both statistics are evaluated on the same data, it represents a *conservative* choice in the sense of producing a wider distribution. This is because it neglects any correlations between $2\hat{\mathcal{F}}$ and $2\tilde{\mathcal{F}}$. The lack of a simple way of quantifying correlations amongst said statistics in a general case justifies the safe approach of fresh data mode even though the same data is actually being used [33].

The functional forms of the distributions in Eq. (19) have already been discussed in this subsection:

$$P(2\tilde{\mathcal{F}}|\rho^2) = \chi_4^2(2\tilde{\mathcal{F}}; \rho^2), \quad (20)$$

$$P(2\hat{\mathcal{F}}|\rho^2, N_{\text{seg}}) = \chi_{4N_{\text{seg}}}^2(2\hat{\mathcal{F}}; \rho^2). \quad (21)$$

It is useful to further simplify Eq. (19) to a closed analytical form. A proxy value for ρ^2 can be obtained by simply subtracting the expected noise-only value of a chi-squared distribution with $4N_{\text{seg}}$ degrees of freedom, namely $\rho_0^2 = 2\hat{\mathcal{F}} - 4N_{\text{seg}}$. Assuming $\rho_0^2 \gg 1$, chi-squared distributions can be replaced by Gaussian distributions [76, 77] and Eq. (19) can be further replaced by a Gaussian, the peak of which corresponds to $\mu_S = \rho_0^2$. We refer to [33] for further details on this derivation and simply quote the final result

$$P(2\tilde{\mathcal{F}}|2\hat{\mathcal{F}}, N_{\text{seg}}) = \text{Gauss}(2\tilde{\mathcal{F}}; \mu_S, \sigma_S), \quad (22)$$

where

$$\begin{aligned} \mu_S &= \rho_0^2, \\ \sigma_S^2 &= 8 \cdot (1 + N_{\text{seg}} + \rho_0^2). \end{aligned} \quad (23)$$

These expressions are useful to discuss the qualitative behavior of our newly proposed Bayes factor in different signal regimes. It will also be applicable in the analysis of software-injected signals in Sec. V A. However, due to the regime in which real-data outliers are typically found, we do not apply this Gaussian approximation to their analysis; instead, we numerically evaluate the full version of Eq. (19).

C. Bayes factor

We will now construct an overall Bayes factor to compare the two hypotheses supporting the presence or lack of a signal in a given stream of data. The distribution associated to the noise hypothesis, given in Eq. (18), is constructed by fitting the location and scale parameters of a Gumbel distribution to background data samples obtained through off-sourcing. The noise hypothesis can be defined in terms of said parameters, namely $\mathcal{H}_N = \{\mu_N, \sigma_N\}$, and the resulting distribution is

$$\ln P(2\tilde{\mathcal{F}}^*|\mathcal{H}_N) = - \left(\frac{2\tilde{\mathcal{F}}^* - \mu_N}{\sigma_N} + e^{-\left(\frac{2\tilde{\mathcal{F}}^* - \mu_N}{\sigma_N}\right)} + \ln \sigma_N \right). \quad (24)$$

Parameters in Gaussian noise	
μ_S	$(3 - 10) \times 10^3$
μ_N	10 – 20
σ_S	30 – 80
σ_N	2 – 3

TABLE I. Typical location and scale parameters obtained from an injection campaign on Gaussian noise with an observing time of $T_{\text{obs}} = 9$ months. Signal location and scale parameters were computed using the second-to-last stage of the coherence-time ladder. See Sec. V for further details.

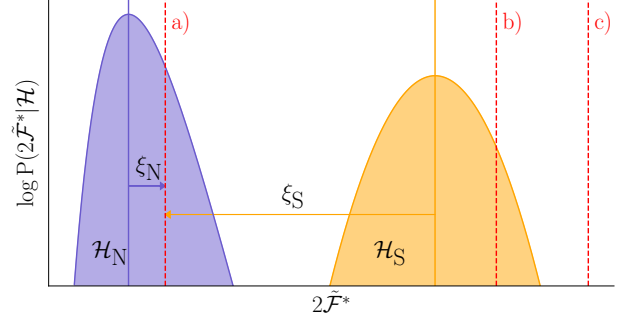


FIG. 2. Illustration of different regimes in which an outlier could be located. Shaded regions represent probability distributions associated to the indicated hypothesis. Dashed vertical lines refer to the enumerated labels in the text.

The signal hypothesis compares the statistical behavior of the loudest candidate across different stages of the coherence-time ladder. We state the signal hypothesis as $\mathcal{H}_S = \{\mu_S, \sigma_S\}$ and, to simplify the following discussion, we write everything in this section using the Gaussian approximation given in Eq. (22):

$$\ln P(2\tilde{\mathcal{F}}^*|\mathcal{H}_S) = -\frac{1}{2} \left[\left(\frac{2\tilde{\mathcal{F}}^* - \mu_S}{\sigma_S} \right)^2 + \ln 2\pi\sigma_S \right]. \quad (25)$$

We note again that this approximated formula will *not* be applied to real-data candidates, as they are not located within the strong signal regime. Instead, we will then numerically evaluate Eq. (19).

It is useful to introduce the following auxiliary variables

$$\xi_S = \frac{2\tilde{\mathcal{F}}^* - \mu_S}{\sigma_S}, \quad \xi_N = \frac{2\tilde{\mathcal{F}}^* - \mu_N}{\sigma_N}, \quad (26)$$

which measure the *discrepancy* of the retrieved $2\tilde{\mathcal{F}}^*$ value with respect to the most probable values under the signal and noise hypothesis, respectively.

Combining Eqs. (24) and (25) we obtain an explicit expression for Eq. (15)

$$\ln \mathcal{B}_{S/N}^* = -\frac{1}{2}\xi_S^2 + \xi_N + e^{-\xi_N} + \ln \frac{\sigma_N}{\sqrt{2\pi}\sigma_S}. \quad (27)$$

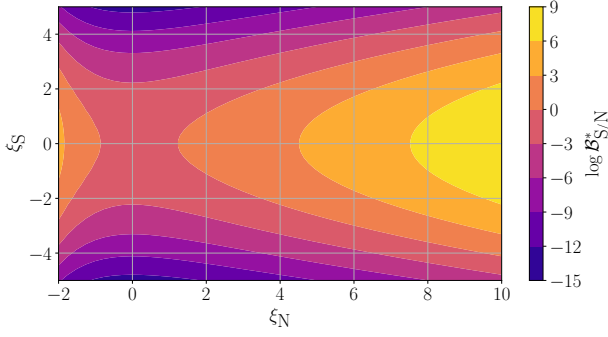


FIG. 3. Bayes factor in terms of the discrepancy of an outlier with respect to the noise and signal hypothesis as described in Eq. (27). Numerical values are computed using $\sigma_N = 3$ and $\sigma_S = 30$, consistent with Table I. This representation will be referred to as the (ξ_N, ξ_S) plane.

Example values of the involved quantities for the use case later in this paper are summarized in Table I. We proceed to analyze the general behavior of this new statistic under different conditions.

The operating point of wide parameter-space searches is generally such that outliers being followed up are significant enough so that $\xi_N > 0$, in the sense that a more sensitive method can be applied once the parameter-space region has been narrowed down. It is also reasonable to expect $\mu_S > \mu_N$, although this assumption may not be valid in case of very deep searches.

We distinguish three interesting regimes of behavior of Eq. (27), labeled in Fig. 2 using dashed vertical lines:

- The candidate is consistent with a noise fluctuation, returning $\xi_N < \xi_S$, hence $\ln \mathcal{B}_{S/N}^* < 0$ and the signal hypothesis is disfavored.
- The candidate is consistent with the signal hypothesis $\xi_S \sim 0$; hence, the dominant contribution to the Bayes factor is given by the discrepancy with respect to the noise hypothesis $\ln \mathcal{B}_{S/N}^* \sim \xi_N$. This is the expected behavior of a detection statistic: the favoring towards the signal hypothesis is directly proportional to the discrepancy with respect to background noise.
- The candidate is beyond the region expected by the signal hypothesis, meaning $\ln \mathcal{B}_{S/N}^* \simeq -\frac{1}{2}\xi_S^2 + \xi_N$. This novel behavior is due to the chosen signal hypothesis: *As opposed to the \mathcal{F} -statistic's signal hypothesis, which results in a monotonic function of SNR, Eq. (19) establishes a particular region of interest centered at $\xi_S = 0$, penalizing deviations towards both sides of it.*

A complementary description of Eq. (27) is shown in Fig. 3, where $\ln \mathcal{B}_{S/N}^*$ is shown on the (ξ_N, ξ_S) plane. These two variables, which represent the discrepancy of

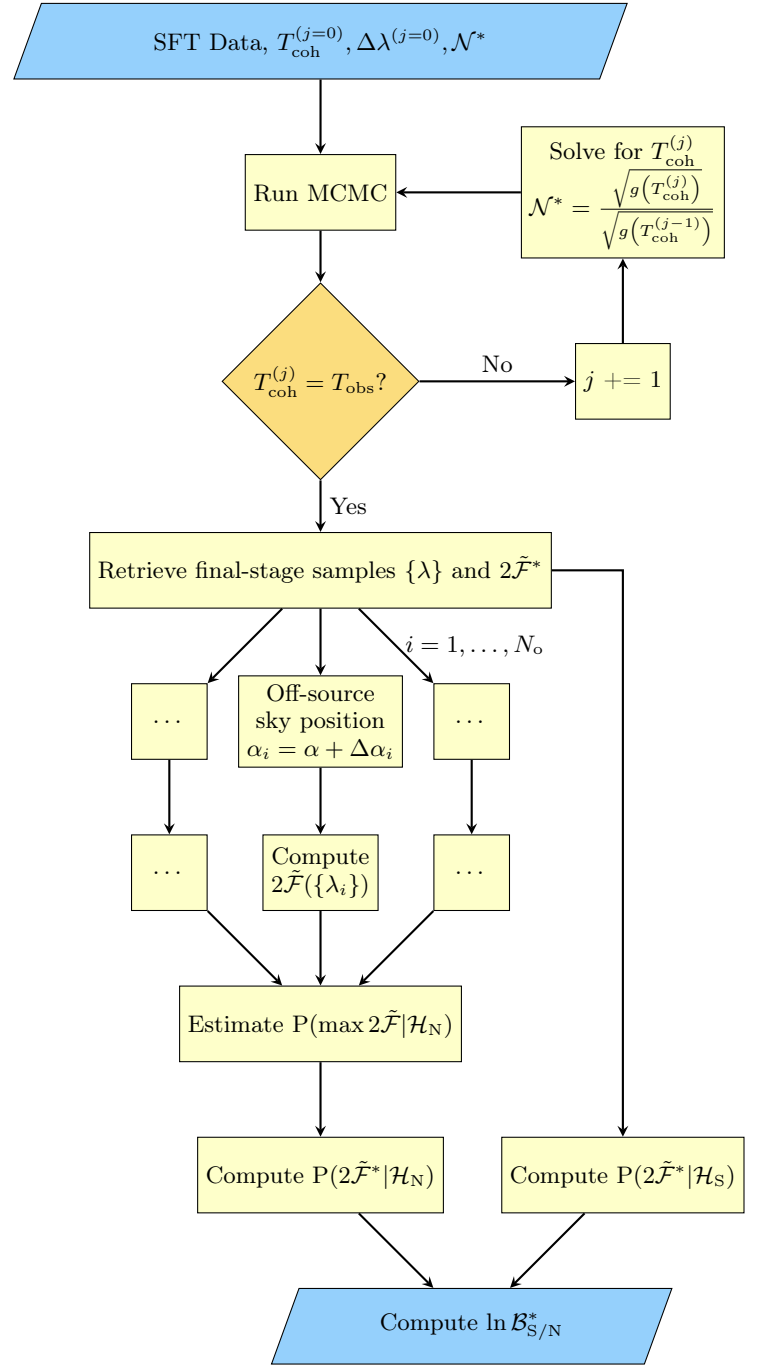


FIG. 4. Flowchart illustrating the computation of $\ln \mathcal{B}_{S/N}^*$ for a CW outlier.

$2\tilde{\mathcal{F}}^*$ with respect to the noise and signal hypothesis, are related by

$$\xi_N = \frac{\sigma_S}{\sigma_N} \xi_S + \frac{\mu_S - \mu_N}{\sigma_N}, \quad (28)$$

meaning that once $\{\mu_S, \sigma_S\}$ and $\{\mu_N, \sigma_N\}$ are determined, the detection statistic is restricted to a straight line in (ξ_N, ξ_S) . This description also clarifies the behavior of $\ln \mathcal{B}_{S/N}^*$ in case b) of Fig. 2. In said case, $\xi_S \sim 0$ and then $\ln \mathcal{B}_{S/N}^* \sim (\mu_S - \mu_N)/\sigma_N$, so that

it is the *combined* action of a high SNR ($\mu_S > \mu_N$) and a low discrepancy with respect to the expected value according to previous stages ($\xi_S \sim 0$) what decides on the consistency of a CW candidate with respect to the signal or noise hypothesis.

A summary of the construction and practical computation of $\ln \mathcal{B}_{S/N}^*$ is shown as a flowchart in Fig. 4.

IV. FOLLOW-UP OF OUTLIERS IN LIGO O2 DATA

We now present the first application of a multi-stage MCMC-based hierarchical follow-up on real data by studying a set of 30 outliers obtained by different CW searches on Advanced LIGO O2 data. These are final-stage outliers resulting from the application of a complete search pipeline, including a set of vetoes depending upon the particularities of each search.

Section IV A briefly describes the main traits of the searches from which outliers are collected. The complete set of outliers to be followed up is reported in Table II. The follow-up setup is described in Sec. IV B.

A. Continuous-wave search outliers from O2 data

1. All-sky *Falcon* search

The *Falcon* pipeline [13] is designed to survey wide parameter-space regions using a so-called *loosely coherent* approach [54, 78, 79], increasing its robustness against small deviations from the standard CW signal model [54].

We are interested in 18 outliers reported in two all-sky searches targeting two different frequency bands of the Advanced LIGO O2 dataset: mid frequency (500 – 1700 Hz) [80] and high frequency (1700 – 2000 Hz) [81]. These searches intended to unveil unknown low-ellipticity sources by analyzing a restricted set of spindown rates ($|f_1| \lesssim 3 \cdot 10^{-12}$ Hz/s). These outliers are the result of a four-stage search using four different coherent times, namely 12, 24, 48 and 144 hours. After each stage, only those templates over a specified threshold were further followed up.

An additional low-frequency *Falcon* search was recently reported in [82]. As will be shown in Sec. V B regarding low-frequency outliers from the other searches discussed below, the greater number and variety of instrumental artifacts in the low-frequency data somewhat hinder the effectiveness of this first incarnation of our follow-up method, as they are not directly addressed by the noise hypothesis. Therefore, we leave a re-analysis of the new low-frequency *Falcon* outliers for future work.

2. Directed *Einstein@Home* search

Einstein@Home is a large-scale computing framework based on the volunteer-computing platform BOINC [83] on which the Global Correlations Transform pipeline [84–86], intended to perform deep (very sensitive) searches across wide parameter-space regions, is deployed. This pipeline is flexible enough so as to be reconfigured into a directed pipeline, using astrophysical information obtained by electromagnetic means to restrict the sky positions to search on.

We are interested in the surviving outlier from a directed search for CWs from central compact objects in three supernova remnants [64]. Said outlier is associated to the central compact object known as 1 WGA J1713.4-949 [87] and located in SNR G347.3-05; for consistency with [64], we will simply refer to it as J1713.

This outlier is a sub-threshold candidate from an earlier *Einstein@Home* search on O1 data directed towards the same supernova remnants [29], which was then re-analyzed using O2 data. The statistical basis of the re-analysis was similar to the techniques explained in Sec. III B, comparing the significance of a candidate on different data streams with respect to the expected significance deduced from the initial analysis. As reported in [64], the outlier under analysis is inconsistent with Gaussian noise, but cannot be associated to the signal hypothesis either.

3. Fomalhaut b *Viterbi* search

The *Viterbi* method spans a family of search pipelines which use a Hidden Markov model (HMM) to describe the frequency evolution of a CW signal [71, 88–90]. Such a signal model is able to incorporate stochastic contributions into the analysis (e.g. timing noise or spin-wandering due to the presence of an accreting companion [91]).

Ref. [92] reports on a *Viterbi* search for CWs directed at Fomalhaut b, an astrophysical object whose exact nature is still surrounded by debate [93–96]. This search complements a previous one performed on Advanced LIGO O1 data using an \mathcal{F} -statistic search assuming the standard deterministic evolution of a CW [97].

The search setup assumes spindown to be the main contribution to the frequency evolution, considering timing noise as a sub-dominant component. This is done by imposing a *biased* random walk as a HMM, in the sense that evolution towards higher frequencies is forbidden³. The search was performed using $T_{\text{coh}} = 5$ days and surviving candidates were sieved

³ This condition drastically reduces the space of possible frequency evolutions contemplated by the HMM model, easing the application of a model-based pipeline to follow up or estimate the exact parameters of any resulting candidates.

though a set of consistency vetoes. In the end, a single outlier was reported for further exploration.

4. *H.E.S.S. Viterbi search*

Another implementation of the *Viterbi* pipeline, similar in scope and assumptions to that mentioned above, was used to perform a search on a set of ten pulsars observed by very high-energy γ -ray surveys in [98].

The search looks for CW emission at once, twice and $4/3$ of the rotational frequency of the targeted pulsars in order to address several emission mechanisms [4]. After assessing the sub-dominant role of spin-wandering on frequency evolution, a biased random walk is implemented in a similar manner to [92], selecting the maximum T_{coh} allowed by the spindown rate of each pulsar so that the frequency evolution is within the range of the HMM.

After applying a set of consistency vetoes, twelve outliers are reported for further exploration; we only considered ten of them as independent follow-up targets since for two pairs of outliers, the corresponding prior parameter-space regions significantly overlap.

B. Follow-up setup

We demonstrate the general application of an MCMC-based multi-stage follow-up to a set of real-data outliers regardless of the pipeline producing them. To do so, outliers will be analyzed ignoring any information gathered from any of the vetoes or follow-up stages reported in their respective searches.

The second Advanced LIGO observing run [99, 100] comprises nine months of data taken by the two Advanced LIGO detectors H1 (Hanford) and L1 (Louisiana) [2]. The employed time segments are those with the “all” tag in [101]. The dataset was divided into segments with a duration of $T_{\text{SFT}} = 1800$ s in which Fourier transforms were computed as explained in [20]. We take the observing time to be $T_{\text{obs}} = 270$ days in order to convert the number of segments of a stage N_{seg} to a coherence time as $T_{\text{coh}} = T_{\text{obs}}/N_{\text{seg}}$.

Our follow-ups are conducted assuming a CW signal model with two spindown components. Since the second spindown component is only reported by the *Einstein@Home* search, we assume it to be compatible with a null value for the other outliers and apply a canonical uncertainty of $\delta f_2 = 2 \cdot (T_{\text{coh}} \cdot T_{\text{obs}}^2)^{-1}$ [9]. As discussed in [102], this increases the robustness of a search method against unmodeled physics, such as neutron star glitches, due to an increase of the available parameter-space correlations.

Table III collects the approximated sensitivity depth achieved by each search according to their reported results. A comparison to the results in Figs. 8 and 9 of [32], which compute the detection efficiency of a four-stage MCMC follow-up starting at $T_{\text{coh}} = 1$ day, places

the outliers within the effective region of the follow-up procedure.

Most wide parameter-space searches currently operate at $T_{\text{coh}} \sim \mathcal{O}(\text{hours})$. As demonstrated in [21, 30], CW candidates with uncertainties at such short coherence times can be successfully recovered by an MCMC follow-up at $T_{\text{coh}} = 0.5$ days.

We construct a hierarchical follow-up by imposing a first stage using $T_{\text{coh}} = 0.5$ days followed by a second stage using $T_{\text{coh}} = 1$ day. Further stages are constructed by means of Eq. (14) using $\mathcal{N}^* = 10^4$. The resulting coherence-time ladder, which is independent of the parameter-space region and the prior specification due to the locality of the analysis, is collected in Table IV. As per the previous discussion on the sensitivity of the considered searches, this ladder can be seamlessly applied to every one of the outliers under analysis.

Table V specifies the hyperparameter setup of every MCMC stage, following the setups employed in [21, 30]. As demonstrated in Sec. VA, this setup suffices to successfully follow up CW candidates within the probed sensitivity range.

The choice of initial priors is directly related to the outlier’s uncertainty returned by each of the analysis pipelines. Pipelines like *Falcon* or *Einstein@Home* return a well-determined parameter-space region in which the outlier was found. The *Viterbi* pipelines, on the other hand, return only the frequency-evolution track of each candidate, which can then be related to a certain parameter-space region if the stochastic contributions are sub-dominant. The scope of a search also affects the prior setup, as searches directed towards a particular sky position (such as the ones performed using *Viterbi*) allow us to place a narrower prior on the sky position of the outlier. It is recommended in [32] to choose a flat prior with fixed bounds containing the outlier’s parameters. Instead, we use a set of Gaussian priors centered at the outlier’s parameters with scale parameters corresponding to the uncertainty in each dimension. After each MCMC step, we re-center the priors on the median value of the resulting posterior distribution, taking half of the (centered) 90% credible interval as the new scale parameter, and re-sample the initial state of the MCMC ensemble. This particular setup ensures a fresh start-up at each stage of the ladder, preventing spurious samples dissociated from the ensemble to pollute the final results. Moreover, the use of unbounded priors prevents the follow-up from missing the true parameters of an outlier due to the presence of parameter-space correlations [47, 103].

The uncertainty associated to *Falcon* outliers is specified in [80] as

$$\begin{aligned} \delta f_0 &= 5 \times 10^{-5} \text{ Hz} , \\ \delta f_1 &= 1 \times 10^{-12} \text{ Hz/s} , \\ \delta \theta &= 0.06 \text{ Hz}/f_0 \text{ rad} , \end{aligned} \tag{29}$$

where $\delta \theta$ refers to the sky position of an outlier projected onto the ecliptic plane. These uncertainties

Outlier ID	Search	f_0 [Hz]	f_1 [Hz/s]	f_2 [Hz/s ²]	α [rad]	δ [rad]	t_{ref} [GPS]	Ref.
Falcon 4	High-frequency Falcon	1891.756740	$-8.22 \cdot 10^{-12}$	—	2.986956	1.005798	1183375935	[81]
Falcon 5	High-frequency Falcon	1892.991060	$-1.08 \cdot 10^{-12}$	—	3.779161	-0.816273	1183375935	[81]
Falcon 15	Mid-frequency Falcon	900.218805	$-2.20 \cdot 10^{-12}$	—	2.084418	-0.102264	1183375935	[80]
Falcon 19	Mid-frequency Falcon	514.148927	$1.60 \cdot 10^{-12}$	—	2.170421	0.092501	1183375935	[80]
Falcon 23	Mid-frequency Falcon	1001.366228	$4.30 \cdot 10^{-12}$	—	1.355837	-0.770266	1183375935	[80]
Falcon 24	Mid-frequency Falcon	676.195421	$2.80 \cdot 10^{-12}$	—	3.847021	-0.101619	1183375935	[80]
Falcon 25	Mid-frequency Falcon	744.219166	$2.40 \cdot 10^{-12}$	—	3.344985	0.612566	1183375935	[80]
Falcon 29	Mid-frequency Falcon	512.490814	$1.20 \cdot 10^{-12}$	—	2.468975	-0.043050	1183375935	[80]
Falcon 31	Mid-frequency Falcon	983.151889	$2.20 \cdot 10^{-12}$	—	3.561119	0.017979	1183375935	[80]
Falcon 34	Mid-frequency Falcon	886.880087	$-1.60 \cdot 10^{-12}$	—	4.912788	-0.703498	1183375935	[80]
Falcon 35	Mid-frequency Falcon	988.373199	$1.20 \cdot 10^{-12}$	—	0.981835	0.778338	1183375935	[80]
Falcon 39	Mid-frequency Falcon	514.291681	$3.20 \cdot 10^{-12}$	—	0.569033	-0.128357	1183375935	[80]
Falcon 40	Mid-frequency Falcon	831.988473	$4.00 \cdot 10^{-13}$	—	4.917347	1.160537	1183375935	[80]
Falcon 41	Mid-frequency Falcon	873.524608	$4.00 \cdot 10^{-13}$	—	0.618991	-0.189450	1183375935	[80]
Falcon 42	Mid-frequency Falcon	895.421949	$3.60 \cdot 10^{-12}$	—	5.105590	0.249163	1183375935	[80]
Falcon 43	Mid-frequency Falcon	1224.745666	$-2.16 \cdot 10^{-12}$	—	1.715268	0.196184	1183375935	[80]
Falcon 45	Mid-frequency Falcon	698.728032	$-2.00 \cdot 10^{-13}$	—	4.557347	-0.724141	1183375935	[80]
Falcon 46	Mid-frequency Falcon	1095.557400	$-1.08 \cdot 10^{-12}$	—	4.354664	-0.260254	1183375935	[80]
J1713	Einstein@Home	368.801379	$-4.37 \cdot 10^{-9}$	$5.9 \cdot 10^{-19}$	4.509371	-0.695189	1131943508	[64]
Fomalhaut b	Fomalhaut b Viterbi	876.503400	$-1.00 \cdot 10^{-12}$	—	6.011130	0.517000	1167545066	[92]
J0534+2200	H.E.S.S. Viterbi	29.813738	$-3.77 \cdot 10^{-10}$	—	1.459675	0.384225	1164556817	[98]
J1420-6048	H.E.S.S. Viterbi	14.511294	$-1.70 \cdot 10^{-11}$	—	3.753057	-1.061240	1164556817	[98]
J1420-6048	H.E.S.S. Viterbi	19.515033	$-2.30 \cdot 10^{-11}$	—	3.753057	-1.061240	1164556817	[98]
J1420-6048	H.E.S.S. Viterbi	29.522611	$-3.50 \cdot 10^{-11}$	—	3.753057	-1.061240	1164556817	[98]
J1718-3825	H.E.S.S. Viterbi	17.503470	$-3.00 \cdot 10^{-12}$	—	4.530116	-0.670585	1164556817	[98]
J1831-0952	H.E.S.S. Viterbi	14.501823	$-1.00 \cdot 10^{-12}$	—	4.850147	-0.172213	1164556817	[98]
J1831-0952	H.E.S.S. Viterbi	15.401223	$-1.00 \cdot 10^{-12}$	—	4.850147	-0.172213	1164556817	[98]
J1831-0952	H.E.S.S. Viterbi	19.999146	$-2.00 \cdot 10^{-12}$	—	4.850147	-0.172213	1164556817	[98]
J1849-0001	H.E.S.S. Viterbi	26.308209	$-9.00 \cdot 10^{-12}$	—	4.850147	-0.000375	1164556817	[98]
J1849-0001	H.E.S.S. Viterbi	26.341209	$-9.00 \cdot 10^{-12}$	—	4.850147	-0.000375	1164556817	[98]

TABLE II. CW search outliers of interest as reported by their original searches. H.E.S.S. Viterbi outliers will be further referred to by including their corresponding frequency.

Search	Estimated $\mathcal{D}^{95\%}$ [Hz ^{-1/2}]
High-frequency Falcon	55 — 65
Mid-frequency Falcon	45 — 55
Directed Einstein@Home*	80 — 90 (75 — 85)
Fomalhaut b Viterbi	45 — 55
H.E.S.S. Viterbi	45 — 55

TABLE III. Estimated ranges of 95% efficiency sensitivity depths achieved by each of the searches according to their reported results. The depth marked with an asterisk corresponds to a 90% efficiency instead. Values in parentheses refer to the sensitivity depth achieved by the original search producing the outlier [29].

Stage	0	1	2	3	4
N_{seg}	500	250	55	5	1
T_{coh} [days]	0.5	1	5	55	270

TABLE IV. Coherence-time ladder constructed using $\mathcal{N}^* = 10^4$ and including an initial stage of $T_{\text{coh}} = 0.5$ days before imposing $T_{\text{coh}} = 1$ days and applying the SuperSky metric. The results are independent of the parameter-space region at which the SuperSky metric was evaluated.

Hyperparameter	Value
Parallel chains	3
Walkers per chain	100
Burn-in & Production steps	250 + 250

TABLE V. MCMC hyperparameter choices for each stage of the follow-up. The number of parallel chains equals the number of temperatures at which the likelihood is being sampled, following the recommendations in [32, 58].

are conservatively lower than the canonical parameter-space resolution defined in [9] for a coherence time of $T_{\text{coh}} = 0.5$ days, meaning their corresponding parameter-space size is within acceptable values to ensure an effective MCMC stage [21, 30]. The Einstein@Home search reports uncertainties corresponding to a coherence time of several months; since we start our follow-up at a lower coherence time, we used the same set of uncertainties as for the Falcon follow-up Eq. (29). Viterbi outliers were not reported as a parameter-space point, but as a frequency band on which a significant frequency-evolution track was found; since both searches were targeted at a particular sky position, we reduced the sky position uncertainty and increased the frequency

uncertainty by the same factor in order to cover all possible frequencies at a similar parameter-space size.

V. RESULTS

Before presenting results on the O2 outliers in Sec. VB, here we first describe an injection campaign in simulated Gaussian noise to demonstrate the efficacy of the follow-up procedure and calibrate a threshold on the newly introduced Bayes factor.

A. Injections in Gaussian noise

We characterize the behavior of $\ln \mathcal{B}_{S/N}^*$ using three sets of 100 artificial signals at different signal strengths. These are injected into Gaussian noise data compatible with the O2 observing run characteristics, i.e. simulating data for both Advanced LIGO detectors and with a duration of $T_{\text{obs}} = 9$ months, using `lalapps_Makefakedata_v5` [104]. The actual O2 data stream covers 60% of the duration of the run T_{obs} due to down time in the detectors (actual fractions are 65.3% and 61.8% for the H1 and L1, respectively) [99]. Since SNR scales as the square root of observing time, this would reduce the actual SNR of a signal to a fraction of 77%. For the simulated Gaussian noise, we set the average amplitude spectral density to a fiducial value of $\sqrt{S_n} = 10^{-23} \text{ Hz}^{-1/2}$. We injected the artificial signals at a fiducial frequency of 100 Hz, uniformly spread across the whole sky and log-uniformly distributed in spindown parameter f_1 within $[-10^{-8}, -10^{-11}] \text{ Hz/s}$. The particular choice of a frequency band does not affect the results of this analysis, since its effects are automatically taken into account by parameter-space resolutions.

The CW amplitude h_0 is fixed in terms of the sensitivity depth [16, 53]

$$\mathcal{D} = \frac{\sqrt{S_n}}{h_0}. \quad (30)$$

Additionally, we define an *effective* sensitivity depth by explicitly including the effects of the cosine of the inclination angle ι [105]:

$$\mathcal{D}_{\text{Eff}} = \frac{\mathcal{D}}{\sqrt{\cos^4 \iota + 6 \cos^2 \iota + 1}}. \quad (31)$$

We selected three depth values, enumerated in Table VI, bracketing the estimated 95% efficiency depth of the analyzed pipelines. The rest of the amplitude parameters were randomly drawn from uniform distributions [106].

We start by estimating the detection efficiency of the follow-up. To do so, we run the full hierarchical follow-up as specified in the previous section and count an injection as “detected” if the injection parameters are within the final-stage posterior probability support. This criterion ensures the CW signals are strong enough to guide the

Depth $[\text{Hz}^{-1/2}]$	Efficiency (%)
40	97 ± 2
60	98 ± 1
80	96 ± 2
Overall	97 ± 1

TABLE VI. Detection efficiencies for each set of 100 injections. An injection was labeled as detected if the final-stage posterior probability contained the injection parameters in its support. Error bars correspond to binomial errors.

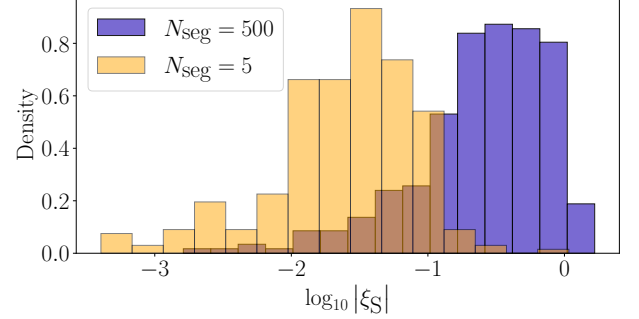


FIG. 5. Distribution of $|\xi_S|$ for the complete set of detected injections using different semicoherent stages, namely $N_{\text{seg}} = 500$ and $N_{\text{seg}} = 5$, as the reference to compute μ_S and σ_S .

MCMC ensemble towards the relevant parameter space region, preventing a signal from being lost. Results are reported in Table VI. As expected from previous analyses in [32], we obtain a detection efficiency above 95% across the sensitivity range, meaning the follow-up is a suitable tool to further analyze the selected set of outliers.

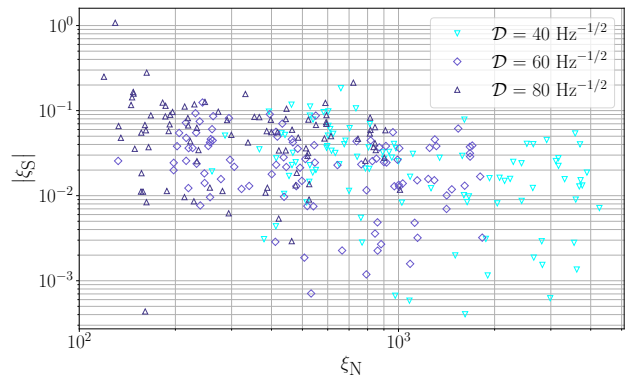


FIG. 6. (ξ_S, ξ_N) plane for the complete set of detected injections using $N_{\text{seg}} = 5$ as the reference stage to compute μ_S and σ_S . The horizontal axis represents the discrepancy with respect to the noise hypothesis, while the vertical axis represents the discrepancy with respect to the signal hypothesis.

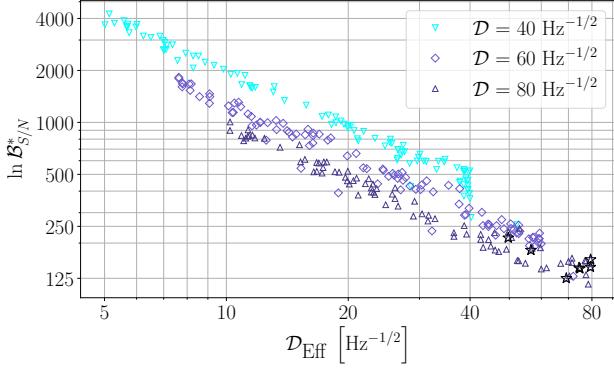


FIG. 7. $\ln \mathcal{B}_{S/N}^*$ computed by applying the multi-stage MCMC follow up on the three sets of software injections. Reference values were computed with respect to the $N_{\text{seg}} = 5$ stage and the Gaussian approximation was used to compute the signal contribution. Outliers marked by a star did not display an ensemble-level volume shrinkage, as explained in the text.

The computation of $\ln \mathcal{B}_{S/N}^*$ requires a particular semicoherent step from the ladder to be selected as the one from which the expected fully-coherent distribution will be propagated. As discussed in Sec. III B, using longer coherence times imposes a more restrictive signal model, reducing the number of outliers due to the presence of detector artifacts and increasing the significance of signal candidates (see e.g. Figure 6 of [32]). Fig. 5 shows the obtained distribution of signal-hypothesis discrepancies $|\xi_S|$ for the complete set of detected injections with respect to two different stages. The use of a lower number of segments (i.e. a longer coherence time) yields a tighter consistency with respect to the expected distribution. We decide to carry out the analysis by taking the second-to-last stage of the ladder ($N_{\text{seg}} = 5$) as the reference from which the expected fully-coherent \mathcal{F} -statistic distribution will be computed.

Figure 6 displays the distribution of injection results on the (ξ_N, ξ_S) plane, showing the discrepancy of an outlier with respect to the noise and signal hypotheses, respectively. The configuration is such that $\xi_S \sim 0$ and $\xi_N \gg \xi_S$, corresponding to case b) in Sec. III. This means that the computation of the signal contribution to $\ln \mathcal{B}_{S/N}^*$ can be assumed to follow a Gaussian distribution and, correspondingly, Eq. (27) applies. Figure 7 shows the Bayes factor $\ln \mathcal{B}_{S/N}^*$ computed by comparing the last two stages of the semicoherent ladder. The observed behavior $\ln \mathcal{B}_{S/N}^* \propto \mathcal{D}_{\text{Eff}}^{-1}$ can be simply explained by noting that \mathcal{D}_{Eff} is inversely proportional to SNR by definition [53].

This injection campaign covers the sensitivity ranges reported in Table III for all searches except one. The **Einstein@Home** search differs in that it was built as a subthreshold search: the reported outlier was thoroughly scrutinized using a variety of tools, including a fully-coherent analysis on O2 data, following a similar scheme as the one presented in this work. in order to assess the follow-up capabilities of our proposed method,

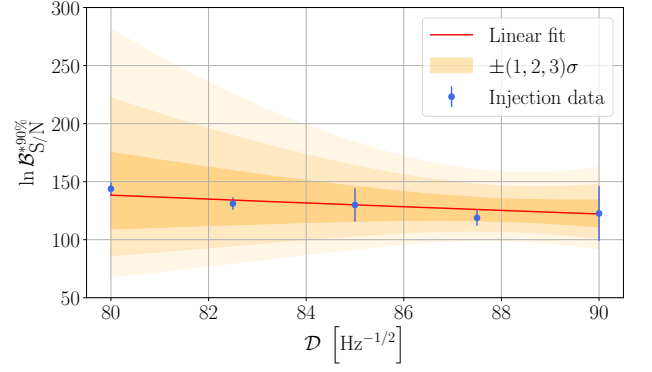


FIG. 8. Estimation of 90% detection-probability threshold on $\ln \mathcal{B}_{S/N}^*$ for different sensitivity depths beyond the injections shown in Fig. 7. Each dot represents an empirical estimate of the 90% detection-probability threshold using 100 simulated signals at a fixed depth value. Error bars correspond to the bootstrap standard deviation using 200 resamples of 50 samples each. The solid line and the associated envelopes represent a linear fit using `scipy.optimize.curve_fit` [107] with 1, 2, and 3 sigma uncertainties.

we perform a second injection campaign akin to the previous one, covering the deepest **Einstein@Home** search sensitivity range. Results are reported as 90% detection-probability thresholds in Fig. 8, following the approach proposed in [53]. Were any of the considered outliers due to a genuine CW signal, the corresponding $\ln \mathcal{B}_{S/N}^*$ should lie within the shaded region or higher. Based on this argument, we set a safe decision threshold at $\ln \mathcal{B}_{S/N}^* = 30$, also accounting for the reduced SNR in the real dataset due to detector downtime.

Lastly, we comment on the behavior of the multi-stage MCMC itself in terms of the volume shrinkage rate introduced in Sec. II C. Figure 9 shows the behavior of the posterior volume of a successfully detected injection. The quantities $V_{\text{prior}}^{(0)}$ and V_{post} represent approximations to the initial prior volume at the first stage of the ladder and the posterior volume after each of the MCMC stages. These quantities are computed by taking the product of parameter-wise central 90% credible intervals, since we are only interested in the overall scaling along the coherence-time ladder. The volume shrinkage shows a power-law behavior, the exponent of which (i.e. the slope in log-log scale) should be approximately given by $\log_{10} \gamma^{(j+1)} \sim 4$ from Eq. (12). The same procedure is performed on the complete set of detected injections, collecting the power-law indices into a histogram in Fig. 10. The rate of volume shrinkage accumulates a prominent peak within the order of magnitude of the expected result.

Figure 10 also displays a small set of injections for which the MCMC ensemble did not produce a clear shrinkage of the (approximated) central 90% credible region, even though the true injection parameters are contained within said region. Their corresponding

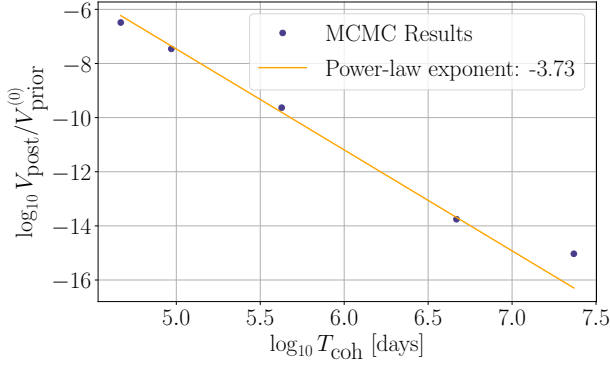


FIG. 9. Posterior volume shrinkage of a successfully detected software injection. Parameter-space volumes are estimated by taking the product of parameter-wise 90% central credible regions as explained in the text. The vertical axis represents the posterior volume as a fraction of the initial prior volume. The slope of the log-log plot is an approximation to the inverse of the volume shrinkage $1/\gamma^{(j+1)}$, where the volume shrinkage $\gamma^{(j+1)}$ was defined in Eq. (12).

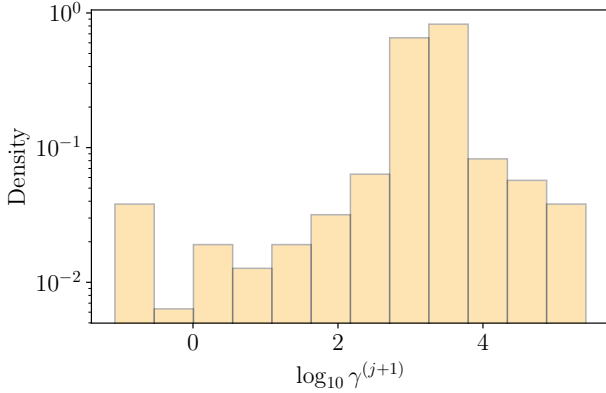


FIG. 10. Distribution of (\log_{10}) posterior volume shrinkage rates of the detected injections. The prominent peak is within the order of magnitude of the expected value according to Eq. (12). The presence of *negative* shrinkage rates for a few injections is discussed in the text.

$\ln \mathcal{B}_{S/N}^*$ values are marked using stars in Fig. 7, belonging to the weakest set of performed injections. This is a consequence of the parameter-space structure in the vicinity of a signal [47]: The effective (squared) SNR recovered by a template falls off as a linear function of the mismatch with respect to the true signal parameters. Strong injections, associated to higher SNR values, are able to sustain an \mathcal{F} -statistic above background throughout a wider parameter-space region than weak injections. Weak injections, as a result, require tighter priors to display a similar behavior to that of stronger injections. The fact that the ensemble is unable to focus into a particular parameter-space region, however, is still compatible with a good recovery of $\ln \mathcal{B}_{S/N}^*$, as for that

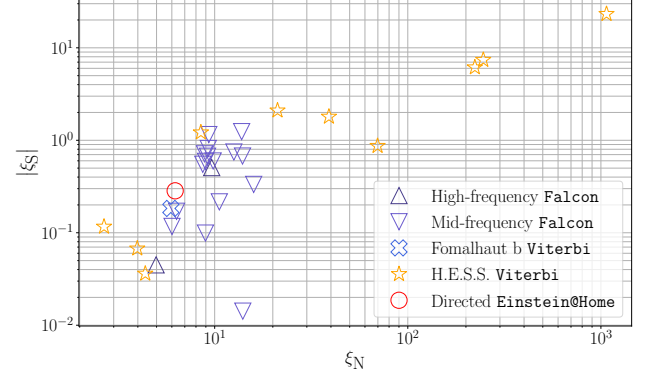


FIG. 11. (ξ_N, ξ_S) plane associated to the outliers found in O2 data by the specified searches.

it is only required to *sample* the region of interest during the production stage. This is in fact the principle upon which the application of a single-stage MCMC follow-up as a simple veto was based in [21, 30], and can be justified by interpreting the MCMC follow-up as being equivalent to a search starting from a *random* template bank at higher mismatches than traditionally suggested in CW searches [108, 109].

B. Follow-up of CW outliers from Advanced LIGO O2 data

We now report the results of the multi-stage MCMC-based follow-up on the set of outliers described in Sec. IV A in terms of the obtained (ξ_N, ξ_S) values and the corresponding $\ln \mathcal{B}_{S/N}^*$.

Figure 11 shows each pipeline's outliers across the (ξ_N, ξ_S) plane, quantifying their discrepancy with respect to the noise and signal hypotheses. The bulk of outliers show discrepancies with respect to the signal hypothesis, quantified by $|\xi_S|$, an order of magnitude larger than those displayed by software injections in Fig. 6. Discrepancies with respect to background noise, quantified by ξ_N , are more than an order of magnitude lower. The retrieved values of μ_N and σ_N are within the brackets obtained in Gaussian noise, suggesting these results are not because of an elevated background noise but rather a low SNR associated to the outliers. We note the presence of three Viterbi outliers at high values of ξ_N , namely J1831-0952@19.9991Hz, J1849-0001@26.3410Hz and J1831-0952@15.4012Hz; and another marginal pair stemming from the same pipeline in the middle ground, namely J1718-3825@17.5034Hz and J1831-0952@14.5018Hz.

We compute $\ln \mathcal{B}_{S/N}^*$ by numerically integrating Eq. (19) due to the regime in which outliers are placed. Results are listed in Table VII and displayed in Fig. 12. Five outliers score over the decision threshold $\ln \mathcal{B}_{S/N}^* = 30$, all of them related to the H.E.S.S. Viterbi pipeline.

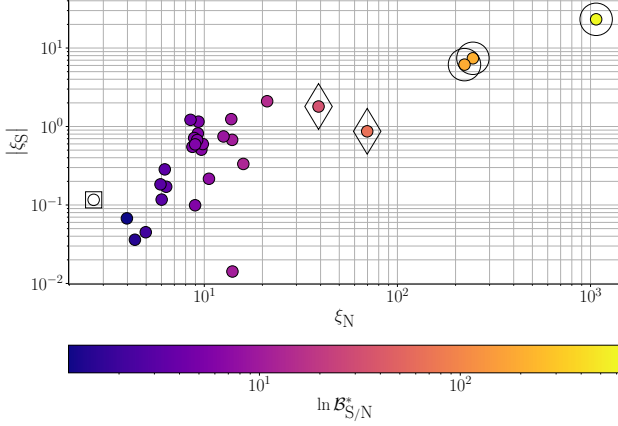


FIG. 12. $\ln \mathcal{B}_{S/N}^*$ values obtained after the hierarchical MCMC follow-up of O2 outliers. Relative outlier positions in this figure are consistent with Fig. 11. Outliers enclosed by circles and diamonds score a $\ln \mathcal{B}_{S/N}^*$ value above 30. The outlier enclosed by a square returns a negative value of $\ln \mathcal{B}_{S/N}^*$ and is displayed as white due to the logarithmic color scale.

The first set of outliers, J1831-0952@15.4012Hz, J1831-0952@19.9991Hz, and J1849-0001@26.3410Hz, is highlighted using circular markers in Fig. 12. The original search [98] ascribed them to instrumental artifacts in the L1 detector. We confirm that to be the case for the outlier J1831-0952@19.9991Hz: the loudest fully-coherent \mathcal{F} -statistic recovered by our follow-up is located at $f_0 \simeq 20.0011$ Hz, crossing a well-known instrumental comb at both LIGO detectors [110]. For outlier J1849-0001@26.3410Hz, we note the presence of a hardware injection (a CW-like signal simulated by direct actuation of the interferometer mirrors, used to test calibration and analysis pipelines) at $f_0 \simeq 26.3396$ Hz with an amplitude corresponding to $\mathcal{D} \sim \mathcal{O}(1 \text{ Hz}^{-1/2})$ [100, 111]. Even though the spindown and sky positions are completely mismatched, such strong artificial signals are known to produce loud candidates across wide parameter space regions [28, 30, 112–114]. We are unable to relate J1831-0952@15.4012Hz to any of the listed narrow spectral artifacts in [100, 110].

A manual check of the segment-wise semicoherent \mathcal{F} -statistic values of J1831-0952@15.4012Hz reveals a rapid accumulation of \mathcal{F} -statistic as the frequency evolution crosses a narrow sub-band. This kind of behavior, shown in Fig. 13, is inconsistent with a CW signal and usually can be related to instrumental artifacts, but said identification becomes more difficult at low frequencies as they are populated by a wider variety of noise sources. As a result, outlier J1831-0952@15.4012Hz is also likely related to an instrumental artifact.

The second group of outliers is enclosed by diamonds in Fig. 12. Outlier J1831-0952@14.5018Hz's loudest candidate is recovered at $f_0 \simeq 14.4953$ Hz. This is

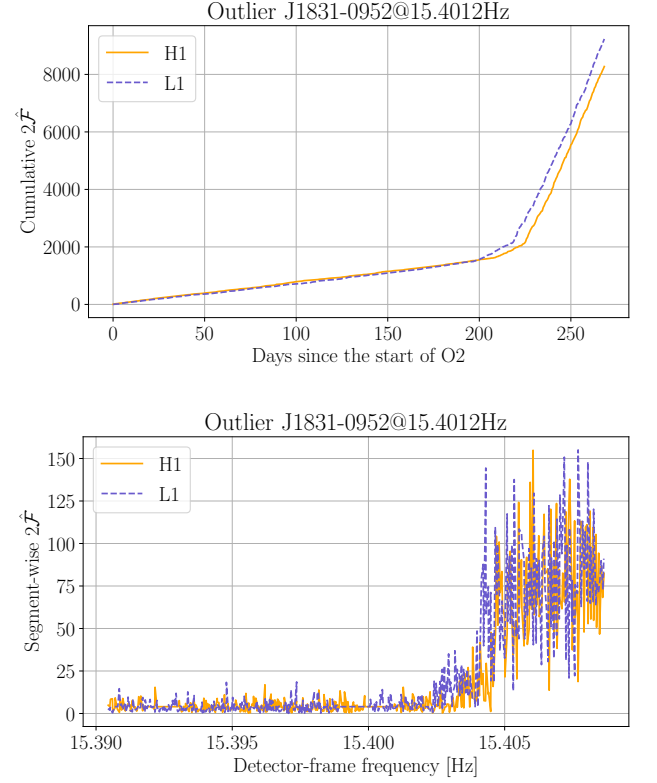


FIG. 13. Segment-wise $2\hat{\mathcal{F}}$ accumulation of the loudest template associated to the outlier J1831-0952@15.4012Hz throughout the observing run using 500 coherent segments. The solid orange line shows the results using the LIGO Hanford detector (H1) only, while the dashed blue line shows the results using the LIGO Livingston detector (L1). The upper panel shows the $2\hat{\mathcal{F}}$ accumulation throughout the duration of the run. The lower panel shows the segment-wise $2\hat{\mathcal{F}}$ values per frequency bin. Frequency values are computed evaluating the frequency-evolution template at the starting time of each segment.

consistent with a 1 Hz comb with an offset of 0.5 Hz.⁴ Another harmonic of the same comb can be related to outlier J1718-3825@17.5034Hz, whose loudest candidate is located at $f_0 \simeq 17.5005$ Hz.

The remaining outliers from all searches return a $\ln \mathcal{B}_{S/N}^*$ value below the decision threshold $\ln \mathcal{B}_{S/N}^* = 30$. For completeness, we list the parameters recovered by the final follow-up stage in Table VII. We highlight outlier J1713, initially found by the Einstein@Home search and scoring below our decision threshold. This result is consistent with the latest Einstein@Home search for J1713 reported in [115], covering up to 400 Hz in O2 data, not finding any significant outliers.

⁴ The spindown value reported by the original search is such that also positive values are covered by the initial prior volume, and indeed our followup recovered the loudest candidate at positive spindown.

Outlier ID	f_0 [Hz]	f_1 [Hz/s]	f_2 [Hz/s ²]	α [rad]	δ [rad]	$\ln \mathcal{B}_{S/N}^*$
Falcon 4	1891.756615	$-6 \cdot 10^{-12}$	$1 \cdot 10^{-20}$	2.987285	1.005941	2.10
Falcon 5	1892.991046	$-2 \cdot 10^{-12}$	$-1.8 \cdot 10^{-19}$	3.778973	-0.816265	6.46
Falcon 15	900.218764	$-1 \cdot 10^{-12}$	$7 \cdot 10^{-20}$	2.084412	-0.102330	10.49
Falcon 19	514.148984	$6 \cdot 10^{-12}$	$3.7 \cdot 10^{-19}$	2.170669	0.092978	9.04
Falcon 23	1001.366278	$2 \cdot 10^{-12}$	$4.9 \cdot 10^{-19}$	1.355553	-0.769952	5.88
Falcon 24	676.195493	$3 \cdot 10^{-12}$	$-2.6 \cdot 10^{-19}$	3.846438	-0.102138	5.80
Falcon 25	744.219196	$2 \cdot 10^{-12}$	$0.4 \cdot 10^{-20}$	3.344781	0.612270	7.42
Falcon 29	512.490782	$-8 \cdot 10^{-12}$	$-2 \cdot 10^{-20}$	2.468688	-0.041880	3.37
Falcon 31	983.151151	$-1 \cdot 10^{-12}$	$-7.0 \cdot 10^{-19}$	3.562362	0.018926	3.06
Falcon 34	886.880063	$-2 \cdot 10^{-12}$	$-2.1 \cdot 10^{-19}$	4.912748	-0.703663	10.85
Falcon 35	988.373241	$2 \cdot 10^{-12}$	$-1.3 \cdot 10^{-19}$	0.982043	0.778393	6.48
Falcon 39	514.291753	$3 \cdot 10^{-12}$	$-3.5 \cdot 10^{-19}$	0.569150	-0.128791	5.47
Falcon 40	831.988457	$-3 \cdot 10^{-12}$	$-1 \cdot 10^{-20}$	4.917884	1.160566	5.48
Falcon 41	873.524663	$3 \cdot 10^{-12}$	$4 \cdot 10^{-20}$	0.619107	-0.189295	5.77
Falcon 42	895.421995	$1 \cdot 10^{-12}$	$-1.9 \cdot 10^{-19}$	5.105728	0.249030	5.37
Falcon 43	1224.745693	$1 \cdot 10^{-12}$	$-1.2 \cdot 10^{-19}$	1.715372	0.196097	5.70
Falcon 45	698.728033	$1 \cdot 10^{-12}$	$1.6 \cdot 10^{-19}$	4.557448	-0.723930	12.69
Falcon 46	1095.557373	$-4 \cdot 10^{-12}$	$-5.7 \cdot 10^{-19}$	4.354405	-0.260292	9.60
J1713	368.801590	$-4.380 \cdot 10^{-9}$	$1.18 \cdot 10^{-18}$	4.511570	-0.694137	3.21
Fomalhaut b	876.517914	$-4.2979 \cdot 10^{-10}$	$-5.67 \cdot 10^{-18}$	6.011153	0.516952	2.96
J0534+2200	29.813469	$-2.3430 \cdot 10^{-10}$	$1.158 \cdot 10^{-17}$	1.461040	0.385286	-0.16
J1420-6048	14.511112	$-2.5 \cdot 10^{-11}$	$1.364 \cdot 10^{-17}$	3.750570	-1.061001	1.46
J1420-6048	19.512364	$-4.4 \cdot 10^{-11}$	$9.80 \cdot 10^{-18}$	3.753242	-1.061227	1.12
J1420-6048	29.526774	$3.9 \cdot 10^{-11}$	$5.30 \cdot 10^{-18}$	3.753011	-1.060915	4.46
J1718-3825	17.500500	$-4.0 \cdot 10^{-11}$	$7.58 \cdot 10^{-18}$	4.528719	-0.670563	<i>33.45</i>
J1831-0952	14.495361	$2.95 \cdot 10^{-10}$	$-3.89 \cdot 10^{-18}$	4.848071	-0.172356	<i>65.64</i>
J1831-0952	15.389002	$8.72 \cdot 10^{-10}$	$-2.980 \cdot 10^{-17}$	4.853052	-0.165697	203.35
J1831-0952	20.0016854	$-7.13 \cdot 10^{-10}$	$7.017 \cdot 10^{-17}$	4.859380	-0.179332	633.204
J1849-0001	26.3062476	$-6.4 \cdot 10^{-11}$	$1.101 \cdot 10^{-17}$	4.850380	-0.000331	14.71
J1849-0001	26.333433	$-7.1 \cdot 10^{-11}$	$2.297 \cdot 10^{-17}$	4.850122	0.003055	192.71

TABLE VII. Loudest template recovered by the multi-stage MCMC follow up for each of the analyzed outliers. Boldface and italic $\ln \mathcal{B}_{S/N}^*$ values correspond to the two sets of outliers highlighted with circles and diamonds in Fig 12, respectively.

VI. CONCLUSION

We have introduced the first complete framework to analyze outliers from arbitrary CW searches using a multi-stage MCMC-based follow-up. After demonstrating its general behavior on Gaussian noise, we applied it to a set of 30 outliers obtained by different CW search pipelines on O2 Advanced LIGO data [64, 80, 81, 92, 98].

The procedure constructs a Bayes factor comparing whether the behavior of the \mathcal{F} -statistic across different stages of the analysis is more consistent with the presence of a signal rather than with pure noise. The expected evolution of this detection statistic as the follow-up progresses can be derived from first principles. The noise contribution is described by applying extreme value theory to samples of background noise data. These samples can be obtained by sampling shifted sky positions with respect to the outliers, blinding the analysis from the presence of a signal.

The application of a multi-stage MCMC follow-up deemed 25 of the analyzed outliers as less consistent with a standard CW signal than with background noise. The remaining five outliers passed the specified threshold

and were manually inspected. Four of them were successfully associated to known instrumental artifacts in the Advanced LIGO detectors. The fifth outlier displays a behavior inconsistent with a CW signal but consistent with an instrumental artifact; the exact instrumental cause, however, could not be identified.

Although the outliers were analyzed assuming a standard signal model corresponding to an isolated CW source, the framework presented here (and the PyFstat software used [32, 34, 59]) can be seamlessly applied to more general models, such as sources in binary systems [21, 30], sources producing glitches [35], and long gravitational-wave transient signals [36, 37].

This represents the first application of a multi-stage MCMC-based follow-up to CW outliers from real data. The scalability of this development is such that it can be taken as a default follow-up strategy to outliers produced by virtually any CW search, as long as they can be related to a well-defined parameter space region. This allows for the general application of long-coherence follow-ups, massively reducing the complexity associated with the setup and calibration of ad hoc vetoes in CW searches.

ACKNOWLEDGMENTS

We gratefully thank Luana M. Modafferi, Luca Rei, and the LIGO-Virgo-KAGRA Continuous Wave working group for many fruitful discussions. We thank an anonymous referee for valuable comments which enhanced the quality of this manuscript. This work was supported by European Union FEDER funds, the Ministry of Science, Innovation and Universities and the Spanish Agencia Estatal de Investigación grants PID2019-106416GB-I00/AEI/10.13039/501100011033, RED2018-102661-T, RED2018-102573-E, Comunitat Autònoma de les Illes Balears through the Direcció General de Política Universitaria i Recerca with funds from the Tourist Stay Tax Law ITS 2017-006 (PRD2018/24), Generalitat Valenciana (PROMETEO/2019/071), EU COST Actions CA18108, CA17137, CA16214, and CA16104. R. T. is supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (ref. FPU 18/00694). D. K. is supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (ref. BEAGAL 18/00148) and cofinanced by the Universitat de les Illes Balears. The authors thankfully acknowledge the computer resources at Picasso and the technical support provided by Barcelona Supercomputing Center - Centro Nacional de Supercomputación through grants No. AECT-2020-3-0022 and No. AECT-2021-1-0029 from the Red Española de Supercomputación (RES). This research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (<https://www.gw-openscience.org/>), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. This paper has been assigned document number LIGO-P2100187.

Appendix A: On the distribution of the maximum \mathcal{F} -statistic and the effective number of templates

The validity of using an effective number of templates to fit Eq. (16) for the expected maximum \mathcal{F} -statistic from a search over a certain actual number of templates, **in the presence** of *non-independent* templates, has been

discussed in the CW literature [52, 53]. We attempt to shed some light on the topic using extreme value theory. Concretely, we analyze the toy model posed in Appendix D of [53].

The basic point in [53] is that the presence of correlated templates not only changes the effective number of templates, but also the “functional form” of the resulting distribution, rendering Eq. (16) inaccurate. As an example, a toy model is constructed by generating a time series of zero-mean unit-variance Gaussian noise and computing the power of its Fourier transform. By choosing a suitable normalization, said power is the squared sum of two identical zero-mean Gaussian variables, following a chi-squared distribution with two degrees of freedom. This distribution can be properly fitted using Eq. (16), and the effective number of templates \mathcal{N}' is consistent with the number of frequency samples $\mathcal{N} = N/2 - 1$, where N is the number of elements from the original time series. Correlated templates are then introduced by over-resolving the Fourier transform applying zero-padding to the time series. The resulting distribution cannot be properly fitted using Eq. (16). The effective number of independent templates \mathcal{N}' is found to *increase* with the length of zero-padding, but it remains bounded by the actual number of power samples \mathcal{N} .

We provide an explanation for the two main issues raised in [53], namely what is the actual “functional form” of the target distribution and why the effective number of templates seems to increase as more correlated templates are included.

Let $x_{n=1,\dots,N}$ be a zero-mean unit-variance Gaussian process. We define its Fourier transform as

$$\tilde{x}_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i n \frac{k}{N}} \quad (\text{A1})$$

where $k = 0, \dots, N-1$. Since $x_n \in \mathbb{R}$, the real and imaginary parts of Eq. (A1) follow a zero-mean Gaussian distribution

$$\begin{aligned} \Re \tilde{x}_k &\sim \text{Gauss}(0, \sqrt{N/2}) \\ \Im \tilde{x}_k &\sim \text{Gauss}(0, \sqrt{N/2}) \end{aligned} \quad (\text{A2})$$

We then define power as

$$\tilde{\rho} = \left(\sqrt{\frac{2}{N}} \Re \tilde{x}_k \right)^2 + \left(\sqrt{\frac{2}{N}} \Im \tilde{x}_k \right)^2 \quad (\text{A3})$$

which, by definition, follows a chi-squared distribution with *two* degrees of freedom $\tilde{\rho} \sim \chi_2^2$. This same quantity is referred to as $2\mathcal{F}_2$ in [53].

The case of a chi-squared distribution with two degrees of freedom is degenerate with an exponential distribution. For the sake of clarity, we re-express it as a gamma distribution with shape parameter $k = 1$ and scale parameter $\theta = 2$, i.e. $\tilde{\rho} \sim \Gamma(1, 2)$. We note that chi-squared distributions correspond to the locus $\theta = 2$ in the parameter space of Gamma distributions, with k equal

to half the degrees of freedom; exponential distributions correspond to the locus $k = 1$, with θ equal to the inverse of the rate parameter.

Let us now define x_n^p as the zero-padded time series containing Np elements, the last $N(p - 1)$ of which are purposely zero. This padding re-scales the variance of the original distribution by a factor $1/p$ and the resulting power can be expressed as

$$\tilde{\rho}_p = \left(\sqrt{\frac{2p}{N}} \Re \tilde{x}_k^p \right)^2 + \left(\sqrt{\frac{2p}{N}} \Im \tilde{x}_k^p \right)^2 = p\tilde{\rho}. \quad (\text{A4})$$

Then, by the properties of the Gamma function, $\tilde{\rho}_p \sim \Gamma(1, 2p)$, which is *not* a chi-squared distribution for $p > 1$, but an exponential distribution with rate parameter $\lambda = (2p)^{-1}$.

Finally, we discuss the asymptotics of the distribution followed by the maximum of a Γ -distributed random variable. As explained in Sec. III A, such light-tailed distributions fall under the domain of attraction of the Gumbel distribution, meaning

$$\max_{\mathcal{N}} \Gamma(k, \theta) \xrightarrow{\mathcal{N} \rightarrow \infty} \text{Gumbel}(\mu, \sigma), \quad (\text{A5})$$

where the location and scale parameters (μ, σ) are given by [70]

$$\mu = \theta [\ln \mathcal{N} + (k - 1) \ln \ln \mathcal{N} - \ln \Gamma(k)], \quad (\text{A6})$$

$$\sigma = \theta. \quad (\text{A7})$$

In particular, the case of $\max_{\mathcal{N}} \tilde{\rho}_p$ results in

$$\mu_p(\mathcal{N}) = 2p \ln \mathcal{N}, \quad \sigma_p = 2p. \quad (\text{A8})$$

It is clear from Eq. (A7) that the asymptotic distribution described by Eq. (16) is a Gumbel distribution with a scale parameter $\sigma = 2$. On the other hand, the asymptotic distribution followed by zero-padded Gaussian noise ($p > 1$) follows a Gumbel distribution with a scale factor $\sigma_p = 2p > 2$. Since the scale parameter is *independent* of \mathcal{N} , Eq. (16) *fails* to describe the asymptotic distribution stemming from correlated templates. In other words, parameter-space correlations shift the distribution followed by the power statistic away from the locus of chi-squared distributions;

since these correlations are generally contained in a certain characteristic length, the resulting light tails are still, however, within the Gumbel distribution's domain of attraction [67].

This result is consistent with the findings reported in Fig. 11 of [53], which we reproduce in Fig. 14. As the zero-padding increases, σ_p increases and the resulting distribution, which is well described by a Gumbel distribution, spreads beyond the fit provided by Eq. (16).

The location parameter μ , on the other hand, does depend on the number of templates. Indeed, if one tries to compute the required effective number of templates \mathcal{N}' so that $\mu_{p=1}(\mathcal{N}')$ coincides with $\mu_p(\mathcal{N})$,

$$\mu_{p=1}(\mathcal{N}') = 2 \ln \mathcal{N}' = 2p \ln \mathcal{N} = \mu_p(\mathcal{N}), \quad (\text{A9})$$

the result is

$$\mathcal{N}' = \mathcal{N}^p, \quad (\text{A10})$$

which is a monotonic function of p . As a result, the effective number of templates *increases* with the zero-padding factor, again in agreement with [53]. We note, however, that this is just a consequence of the chosen Fourier normalization. If the normalized power was constructed using Np as a normalization (the actual number of samples) rather than N (the number of non-zero-padded samples), then Eq. (A4) would be re-written as

$$\hat{\rho}_p = \frac{1}{p} \tilde{\rho}. \quad (\text{A11})$$

Consequently $\hat{\rho}_p \sim \Gamma(1, 2/p)$ and the effects on the standard deviation would be exactly the opposite, as shown in Fig. 15. Indeed, in such a case the effective number of templates would be $\mathcal{N}' = \mathcal{N}^{1/p}$, which decreases as the zero-padding increases.

Our proposed solution to the problem of estimating the effective number of templates is then *not to do so*, as it depends strongly on the specific distribution followed by the noise, which is generally unknown in a real case. Instead, we propose to describe the background noise distribution by fitting an extreme value distribution to a set of samples (see e.g. Section III A). For light-tailed noise, the proper distribution is Gumbel; other distributions are available for noise falling off as a power law or presenting an upper cut-off.

-
- [1] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, arXiv e-print (2020), [arXiv:2010.14527 \[gr-qc\]](#).
 - [2] J. Aasi *et al.*, Advanced LIGO, *Classical and Quantum Gravity* **32**, 074001 (2015), [arXiv:1411.4547 \[gr-qc\]](#).
 - [3] F. Acernese *et al.*, Advanced Virgo: a second-generation interferometric gravitational wave detector, *Classical and Quantum Gravity* **32**, 024001 (2014), [arXiv:1408.3978 \[gr-qc\]](#).
 - [4] M. Sieniawska and M. Bejger, Continuous gravitational waves from neutron stars: current status and prospects, *Universe* **5**, 217 (2019), [arXiv:1909.12600 \[astro-ph.HE\]](#).
 - [5] S. J. Zhu, M. Baryakhtar, M. A. Papa, D. Tsuna, N. Kawanaka, and H.-B. Eggenstein, Characterizing the continuous gravitational-wave signal from boson clouds around galactic isolated black holes, *Physical Review D*

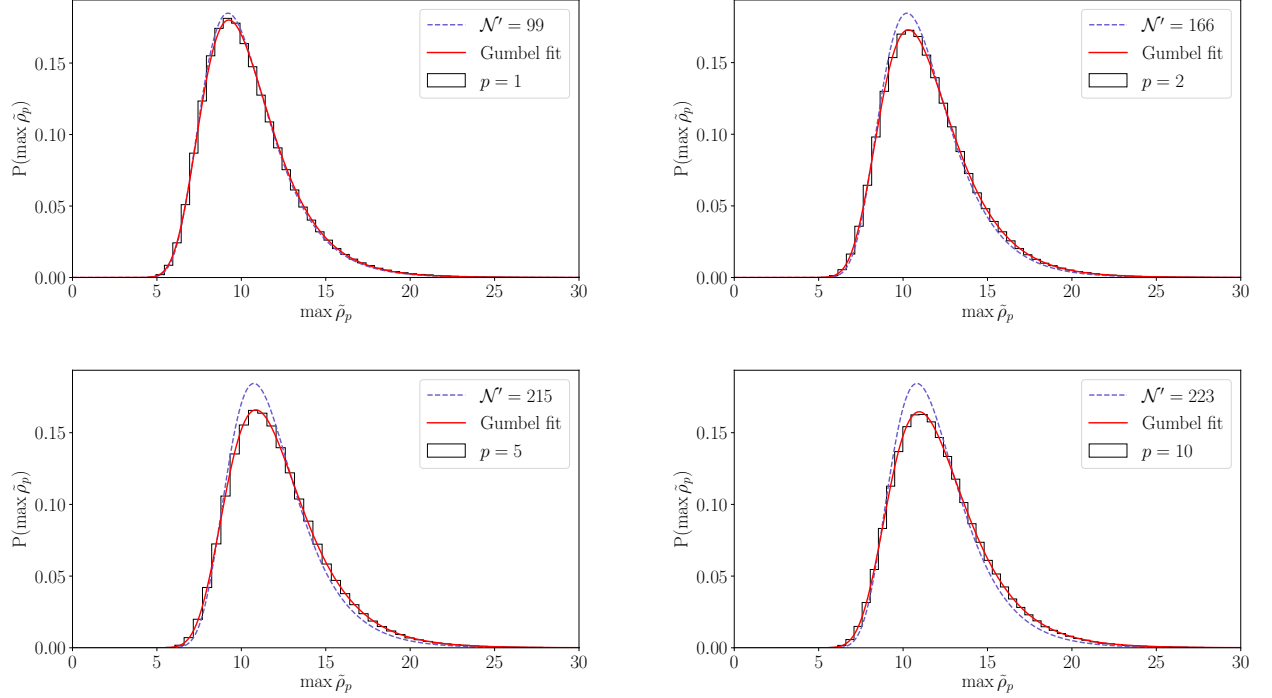


FIG. 14. Maximum Fourier power over $N = 200$ samples of zero-mean unit-variance zero-padded Gaussian noise. In each panel, the stair-case line represents a histogram over 10^6 repeated trials of $\max_N \hat{\rho}_p$. The dashed line is the best fit of Eq. (16) on the effective number of templates \mathcal{N}' , and the solid line is the best fit of a Gumbel distribution on the location and scale parameters. Zero-padding is indicated by p , where $p = 1$ represents no zero-padding, as explained in the text.

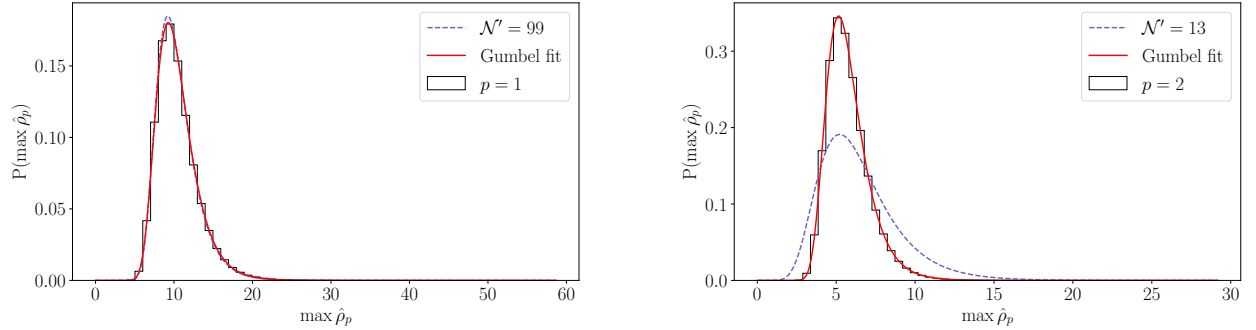


FIG. 15. Equivalent figure to Fig. 14 using the alternative normalization of Fourier power $\hat{\rho}_p$. In this case, increasing the number of correlated templates *narrows* the resulting distribution with respect to Eq. (16).

- 102, 063020 (2020), [arXiv:2003.03359 \[gr-qc\]](#).
- [6] P. R. Brady, T. Creighton, C. Cutler, and B. F. Schutz, Searching for periodic sources with LIGO, [Physical Review D](#) **57**, 2101 (1998), [arXiv:gr-qc/9702050](#).
 - [7] P. R. Brady and T. Creighton, Searching for periodic sources with LIGO II: Hierarchical searches, [Physical Review D](#) **61**, 082001 (2000), [arXiv:gr-qc/9812014](#).
 - [8] C. Cutler, I. Gholami, and B. Krishnan, Improved stack-slide searches for gravitational-wave pulsars, [Physical Review D](#) **72**, 042004 (2005), [arXiv:gr-qc/0505082](#).
 - [9] B. Krishnan, A. M. Sintes, M. A. Papa, B. F. Schutz, S. Frasca, and C. Palomba, Hough transform search for continuous gravitational waves, [Physical Review D](#) **70**, 082001 (2004), [arXiv:gr-qc/0407001](#).
 - [10] P. Astone, A. Colla, S. D'Antonio, S. Frasca, and C. Palomba, Method for all-sky searches of continuous gravitational wave signals using the frequency-Hough transform, [Physical Review D](#) **90**, 042002 (2014), [arXiv:1407.8333 \[astro-ph.IM\]](#).
 - [11] K. Wette, S. Walsh, R. Prix, and M. A. Papa, Implementing a semicoherent search for continuous gravitational waves using optimally-constructed template banks, [Physical Review D](#) **97**, 123016 (2018), [arXiv:1804.03392 \[astro-ph.IM\]](#).
 - [12] M. Sieniawska, M. Bejger, P. Cieciela, and A. Królak, Followup procedure in time-domain \mathcal{F} -statistic searches for continuous gravitational waves, in *XXXVIII Polish Astronomical Society Meeting*, Vol. 7, edited by

- A. Rózańska (2018) pp. 37–40, [arXiv:1810.02224 \[astro-ph.IM\]](#).
- [13] V. Dergachev and M. A. Papa, Sensitivity improvements in the search for periodic gravitational waves using O1 LIGO data, *Physical Review Letters* **123**, 101101 (2019), [arXiv:1902.05530 \[gr-qc\]](#).
 - [14] P. B. Covas and A. M. Sintes, New method to search for continuous gravitational waves from unknown neutron stars in binary systems, *Physical Review D* **99**, 124019 (2019), [arXiv:1904.04873 \[astro-ph.IM\]](#).
 - [15] L. Sancho de la Jordana and A. M. Sintes, A χ^2 veto for continuous wave searches, *Classical and Quantum Gravity* **25**, 184014 (2008), [arXiv:0804.1007 \[gr-qc\]](#).
 - [16] B. Behnke, M. A. Papa, and R. Prix, Postprocessing methods used in the search for continuous gravitational-wave signals from the galactic center, *Physical Review D* **91**, 064007 (2015), [arXiv:1410.5997 \[gr-qc\]](#).
 - [17] P. Leaci, Methods to filter out spurious disturbances in continuous-wave searches from gravitational-wave detectors, *Physica Scripta* **90**, 125001 (2015).
 - [18] S. J. Zhu, M. A. Papa, and S. Walsh, New veto for continuous gravitational wave searches, *Physical Review D* **96**, 124007 (2017), [arXiv:1707.05268 \[gr-qc\]](#).
 - [19] F. Morawski, M. Bejger, and P. Ciecieląg, Convolutional neural network classifier for the output of the time-domain \mathcal{F} -statistic all-sky search for continuous gravitational waves, *arXiv e-prints*, [arXiv:1907.06917](#) (2019), [arXiv:1907.06917 \[astro-ph.IM\]](#).
 - [20] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO O2 data, *Physical Review D* **100**, 024004 (2019), [arXiv:1903.01901 \[astro-ph.HE\]](#).
 - [21] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), All-sky search in early O3 LIGO data for continuous gravitational-wave signals from unknown neutron stars in binary systems, *Physical Review D* **103**, 064017 (2021), [arXiv:2012.12128 \[gr-qc\]](#).
 - [22] A. Singh, M. A. Papa, H.-B. Eggenstein, and S. Walsh, Adaptive clustering procedure for continuous gravitational wave searches, *Physical Review D* **96**, 082003 (2017), [arXiv:1707.02676 \[gr-qc\]](#).
 - [23] B. Beheshtipour and M. A. Papa, Deep learning for clustering of continuous gravitational wave candidates, *Physical Review D* **101**, 064009 (2020), [arXiv:2001.03116 \[gr-qc\]](#).
 - [24] B. Beheshtipour and M. A. Papa, Deep learning for clustering of continuous gravitational wave candidates II: identification of low-SNR candidates, *Physical Review D* **103**, 064027 (2021), [arXiv:2012.04381 \[gr-qc\]](#).
 - [25] R. Tenorio, D. Keitel, and A. M. Sintes, Time-frequency track distance for comparing continuous gravitational wave signals, *Physical Review D* **103**, 064053 (2021), [arXiv:2012.05752 \[gr-qc\]](#).
 - [26] M. Shaltev and R. Prix, Fully coherent follow-up of continuous gravitational-wave candidates, *Physical Review D* **87**, 084057 (2013), [arXiv:1303.2471 \[gr-qc\]](#).
 - [27] M. Shaltev, P. Leaci, M. A. Papa, and R. Prix, Fully coherent follow-up of continuous gravitational-wave candidates: an application to Einstein@Home results, *Physical Review D* **89**, 124030 (2014), [arXiv:1405.1922 \[gr-qc\]](#).
 - [28] M. A. Papa *et al.*, Hierarchical follow-up of subthreshold candidates of an all-sky Einstein@Home search for continuous gravitational waves on LIGO sixth science run data, *Physical Review D* **94**, 122006 (2016), [arXiv:1608.08928 \[astro-ph.IM\]](#).
 - [29] J. Ming *et al.*, Results from an Einstein@Home search for continuous gravitational waves from Cassiopeia A, Vela Jr. and G347.3, *Physical Review D* **100**, 024063 (2019), [arXiv:1903.09119 \[gr-qc\]](#).
 - [30] P. B. Covas and A. M. Sintes, First All-Sky Search for Continuous Gravitational-Wave Signals from Unknown Neutron Stars in Binary Systems Using Advanced LIGO Data, *Physical Review Letters* **124**, 191102 (2020), [arXiv:1904.04873 \[astro-ph.IM\]](#).
 - [31] B. Steltner, M. A. Papa, H. B. Eggenstein, B. Allen, V. Dergachev, R. Prix, B. Machenschalk, S. Walsh, S. J. Zhu, and S. Kwang, Einstein@Home All-sky Search for Continuous Gravitational Waves in LIGO O2 Public Data, *The Astrophysical Journal* **909**, 79 (2021), [arXiv:2009.12260 \[astro-ph.HE\]](#).
 - [32] G. Ashton and R. Prix, Hierarchical multistage MCMC follow-up of continuous gravitational wave candidates, *Physical Review D* **97**, 103020 (2018), [arXiv:1802.05450 \[astro-ph.IM\]](#).
 - [33] Prix, Reinhard, Coherent \mathcal{F} -statistic on semi-coherent candidate, <https://dcc.ligo.org/LIGO-T1700236/public> (2019).
 - [34] D. Keitel, R. Tenorio, G. Ashton, and R. Prix, PyFstat: a Python package for continuous gravitational-wave data analysis, *Journal of Open Source Software* **6**, 3000 (2021).
 - [35] G. Ashton, R. Prix, and D. I. Jones, A semicoherent glitch-robust continuous-gravitational-wave search method, *Physical Review D* **98**, 063011 (2018), [arXiv:1805.03314 \[gr-qc\]](#).
 - [36] R. Prix, S. Giampanis, and C. Messenger, Search method for long-duration gravitational-wave transients from neutron stars, *Physical Review D* **84**, 023007 (2011), [arXiv:1104.1704 \[gr-qc\]](#).
 - [37] D. Keitel and G. Ashton, Faster search for long gravitational-wave transients: GPU implementation of the transient \mathcal{F} -statistic, *Classical and Quantum Gravity* **35**, 205003 (2018), [arXiv:1805.05652 \[astro-ph.IM\]](#).
 - [38] P. Jaranowski, A. Królak, and B. F. Schutz, Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection, *Physical Review D* **58**, 063001 (1998), [arXiv:gr-qc/9804014](#).
 - [39] E. T. Jaynes, *Probability Theory: The Logic of Science*, edited by G. L. Bretthorst (Cambridge University Press, 2003).
 - [40] R. Prix and B. Krishnan, Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics, *Classical and Quantum Gravity* **26**, 204013 (2009), [arXiv:0907.2569 \[gr-qc\]](#).
 - [41] J. T. Whelan, R. Prix, C. J. Cutler, and J. L. Willis, New Coordinates for the Amplitude Parameter Space of Continuous Gravitational Waves, *Classical and Quantum Gravity* **31**, 065002 (2014), [arXiv:1311.0065 \[gr-qc\]](#).
 - [42] C. Cutler and B. F. Schutz, Generalized \mathcal{F} -statistic: Multiple detectors and multiple gravitational wave pulsars, *Physical Review D* **72**, 063006 (2005), [arXiv:gr-qc/0504011](#).
 - [43] P. Leaci and R. Prix, Directed searches for continuous gravitational waves from binary systems: parameter-space metrics and optimal Scorpius X-1 sensitivity, *Physical Review D* **91**, 102003 (2015), [arXiv:1502.00914](#)

- [gr-qc].
- [44] D. Keitel, R. Prix, M. A. Papa, P. Leaci, and M. Siddiqi, Search for continuous gravitational waves: Improving robustness versus instrumental artifacts, *Physical Review D* **89**, 064023 (2014), [arXiv:1311.5738 \[gr-qc\]](#).
- [45] D. Keitel, Robust semicoherent searches for continuous gravitational waves with noise and signal models including hours to days long transients, *Physical Review D* **93**, 084024 (2016), [arXiv:1509.02398 \[gr-qc\]](#).
- [46] R. Prix and M. Shaltev, Search for Continuous Gravitational Waves: Optimal StackSlide method at fixed computing cost, *Physical Review D* **85**, 084010 (2012), [arXiv:1201.4321 \[gr-qc\]](#).
- [47] R. Prix, Search for continuous gravitational waves: Metric of the multi-detector \mathcal{F} -statistic, *Physical Review D* **75**, 023004 (2007), [Erratum: *Physical Review D* **75**, 069901(E) (2007)], [arXiv:gr-qc/0606088](#).
- [48] K. Wette, Empirically extending the range of validity of parameter-space metrics for all-sky searches for gravitational-wave pulsars, *Physical Review D* **94**, 122002 (2016), [arXiv:1607.00241 \[gr-qc\]](#).
- [49] B. Allen, Spherical ansatz for parameter-space metrics, *Physical Review D* **100**, 124004 (2019), [arXiv:1906.01352 \[gr-qc\]](#).
- [50] B. Allen, Optimal Template Banks, *arXiv e-print* (2021), [arXiv:2102.11254 \[astro-ph.IM\]](#).
- [51] K. Wette, Lattice template placement for coherent all-sky searches for gravitational-wave pulsars, *Physical Review D* **90**, 122010 (2014), [arXiv:1410.6882 \[gr-qc\]](#).
- [52] K. Wette, Estimating the sensitivity of wide-parameter-space searches for gravitational-wave pulsars, *Physical Review D* **85**, 042003 (2012), [arXiv:1111.5650 \[gr-qc\]](#).
- [53] C. Dreissigacker, R. Prix, and K. Wette, Fast and Accurate Sensitivity Estimation for Continuous-Gravitational-Wave Searches, *Physical Review D* **98**, 084058 (2018), [arXiv:1808.02459 \[gr-qc\]](#).
- [54] V. Dergachev, On blind searches for noise dominated signals: a loosely coherent approach, *Classical and Quantum Gravity* **27**, 205017 (2010), [arXiv:1003.2178 \[gr-qc\]](#).
- [55] M. Shaltev, Optimizing the StackSlide setup and data selection for continuous-gravitational-wave searches in realistic detector data, *Physical Review D* **93**, 044058 (2016), [arXiv:1510.06427 \[gr-qc\]](#).
- [56] R. Prix, Template-based searches for gravitational waves: Efficient lattice covering of flat parameter spaces, *Classical and Quantum Gravity* **24**, S481 (2007), [arXiv:0707.0428 \[gr-qc\]](#).
- [57] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, emcee: The MCMC Hammer, *Publications of the Astronomical Society of the Pacific* **125**, 306 (2013), [arXiv:1202.3665 \[astro-ph.IM\]](#).
- [58] W. D. Vouden, W. M. Farr, and I. Mandel, Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations, *Monthly Notices of the Royal Astronomical Society* **455**, 1919 (2016), [arXiv:1501.05823 \[astro-ph.IM\]](#).
- [59] G. Ashton, D. Keitel, R. Prix, and R. Tenorio, *Pyfstat/pyfstat: v1.11.3* (2021), <https://doi.org/10.5281/zenodo.4542822>.
- [60] K. Wette and R. Prix, Flat parameter-space metric for all-sky searches for gravitational-wave pulsars, *Physical Review D* **88**, 123005 (2013), [arXiv:1310.5587 \[gr-qc\]](#).
- [61] K. Wette, Parameter-space metric for all-sky semicoherent searches for gravitational-wave pulsars, *Physical Review D* **92**, 082003 (2015), [arXiv:1508.02372 \[gr-qc\]](#).
- [62] H. J. Pletsch, Parameter-space metric of semicoherent searches for continuous gravitational waves, *Physical Review D* **82**, 042002 (2010), [arXiv:1005.0395 \[gr-qc\]](#).
- [63] J. Abadie *et al.* (LIGO Scientific Collaboration), First search for gravitational waves from the youngest known neutron star, *The Astrophysical Journal* **722**, 1504 (2010), [arXiv:1006.2535 \[gr-qc\]](#).
- [64] M. A. Papa, J. Ming, E. V. Gotthelf, B. Allen, R. Prix, V. Dergachev, H.-B. Eggenstein, A. Singh, and S. J. Zhu, Search for Continuous Gravitational Waves from the Central Compact Objects in Supernova Remnants Cassiopeia A, Vela Jr., and G347.3–0.5, *The Astrophysical Journal* **897**, 22 (2020), [arXiv:2005.06544 \[astro-ph.HE\]](#).
- [65] K. W. Wette, *Gravitational waves from accreting neutron stars and Cassiopeia A*, Ph.D. thesis, Australian Natl. U., Canberra (2009).
- [66] K. Wette, L. Dunn, P. Clearwater, and A. Melatos, Deep exploration for continuous gravitational waves at 171–172 Hz in LIGO second observing run data, *arXiv e-print* (2021), [arXiv:2103.12976 \[gr-qc\]](#).
- [67] M. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics (Springer New York, 1983).
- [68] J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, D. De Waal, and C. Ferro, *Statistics of Extremes: Theory and Applications*, Wiley Series in Probability and Statistics (Wiley, 2004).
- [69] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*, Springer Series in Operations Research and Financial Engineering (Springer New York, 2006).
- [70] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events: for Insurance and Finance*, Stochastic Modelling and Applied Probability (Springer Berlin Heidelberg, 2013).
- [71] S. Suvorova, P. Clearwater, A. Melatos, L. Sun, W. Moran, and R. J. Evans, Hidden Markov model tracking of continuous gravitational waves from a binary neutron star with wandering spin. II. Binary orbital phase tracking, *Physical Review D* **96**, 102006 (2017), [arXiv:1710.07092 \[astro-ph.IM\]](#).
- [72] A. Mathai and S. Provost, *Quadratic Forms in Random Variables*, Statistics: A Series of Textbooks and Monographs (Taylor & Francis, 1992).
- [73] A. Gasull, J. López-Salcedo, and F. Utzet, Maxima of Gamma random variables and other Weibull-like distributions and the Lambert W function, *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research* **24**, 714 (2015).
- [74] R. Tenorio, L. M. Modafferi, D. Keitel, and A. M. Sintes, *in preparation*, Tech. Rep. LIGO-P2100277 (LIGO Scientific Collaboration, 2021).
- [75] M. Isi, S. Mastrogiovanni, M. Pitkin, and O. J. Piccinni, Establishing the significance of continuous gravitational-wave detections from known pulsars, *Physical Review D* **102**, 123027 (2020), [arXiv:2010.12612 \[gr-qc\]](#).
- [76] R. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley Series in Probability and Statistics (Wiley, 2005).
- [77] D. Horgan and C. C. Murphy, On the convergence of the chi square and noncentral chi square distributions to the

- normal distribution, *IEEE Communications Letters* **17**, 2233 (2013).
- [78] V. Dergachev, Loosely coherent searches for sets of well-modeled signals, *Physical Review D* **85**, 062003 (2012), [arXiv:1110.3297 \[gr-qc\]](#).
 - [79] V. Dergachev, Loosely coherent searches for medium scale coherence lengths, *arXiv e-print* (2018), [arXiv:1807.02351 \[astro-ph.IM\]](#).
 - [80] V. Dergachev and M. A. Papa, Results from the First All-Sky Search for Continuous Gravitational Waves from Small-Ellipticity Sources, *Physical Review Letters* **125**, 171101 (2020), [arXiv:2004.08334 \[gr-qc\]](#).
 - [81] V. Dergachev and M. A. Papa, Results from high-frequency all-sky search for continuous gravitational waves from small-ellipticity sources, *Physical Review D* **103**, 063019 (2021), [arXiv:2012.04232 \[gr-qc\]](#).
 - [82] V. Dergachev and M. A. Papa, The search for continuous gravitational waves from small-ellipticity sources at low frequencies, *arXiv e-print* (2021), [arXiv:2104.09007 \[gr-qc\]](#).
 - [83] D. P. Anderson, BOINC: A Platform for Volunteer Computing, *arXiv e-prints* (2019), [arXiv:1903.01699 \[cs.DC\]](#).
 - [84] H. J. Pletsch, Parameter-space correlations of the optimal statistic for continuous gravitational-wave detection, *Physical Review D* **78**, 102005 (2008), [arXiv:0807.1324 \[gr-qc\]](#).
 - [85] H. J. Pletsch and B. Allen, Exploiting global correlations to detect continuous gravitational waves, *Physical Review Letters* **103**, 181102 (2009), [arXiv:0906.0023 \[gr-qc\]](#).
 - [86] S. Walsh, K. Wette, M. A. Papa, and R. Prix, Optimizing the choice of analysis method for all-sky searches for continuous gravitational waves with Einstein@Home, *Physical Review D* **99**, 082004 (2019), [arXiv:1901.08998 \[astro-ph.IM\]](#).
 - [87] E. Pfeiffermann and B. Aschenbach, Rosat observation of a new supernova remnant in the constellation scorpius., in *Roentgenstrahlung from the Universe* (1996) pp. 267–268.
 - [88] S. Suvorova, L. Sun, A. Melatos, W. Moran, and R. J. Evans, Hidden Markov model tracking of continuous gravitational waves from a neutron star with wandering spin, *Physical Review D* **93**, 123009 (2016), [arXiv:1606.02412 \[astro-ph.IM\]](#).
 - [89] L. Sun, A. Melatos, S. Suvorova, W. Moran, and R. J. Evans, Hidden Markov model tracking of continuous gravitational waves from young supernova remnants, *Physical Review D* **97**, 043013 (2018), [arXiv:1710.00460 \[astro-ph.IM\]](#).
 - [90] J. Bayley, C. Messenger, and G. Woan, Generalized application of the Viterbi algorithm to searches for continuous gravitational-wave signals, *Physical Review D* **100**, 023006 (2019), [arXiv:1903.12614 \[astro-ph.IM\]](#).
 - [91] A. Mukherjee, C. Messenger, and K. Riles, Accretion-induced spin-wandering effects on the neutron star in Scorpius X-1: Implications for continuous gravitational wave searches, *Physical Review D* **97**, 043016 (2018), [arXiv:1710.06185 \[gr-qc\]](#).
 - [92] D. Jones and L. Sun, Search for continuous gravitational waves from Fomalhaut b in the second Advanced LIGO observing run with a hidden Markov model, *Physical Review D* **103**, 023020 (2021), [arXiv:2007.08732 \[gr-qc\]](#).
 - [93] P. Kalas, J. R. Graham, E. Chiang, M. P. Fitzgerald, M. Clampin, E. S. Kite, K. Stapelfeldt, C. Marois, and J. Krist, Optical Images of an Exosolar Planet 25 Light Years from Earth, *Science* **322**, 1345 (2008), [arXiv:0811.1994 \[astro-ph\]](#).
 - [94] T. Currie, J. Debes, T. J. Rodigas, A. Burrows, Y. Itoh, M. Fukagawa, S. J. Kenyon, M. Kuchner, and S. Matsumura, Direct Imaging Confirmation and Characterization of a Dust-enshrouded Candidate Exoplanet Orbiting Fomalhaut, *The Astrophysical Journal Letters* **760**, L32 (2012), [arXiv:1210.6620 \[astro-ph.EP\]](#).
 - [95] R. Neuhauser, M. M. Hohle, C. Ginski, J. G. Schmidt, V. V. Hambaryan, and T. O. B. Schmidt, The companion candidate near Fomalhaut - a background neutron star?, *Monthly Notices of the Royal Astronomical Society* **448**, 376 (2015), [arXiv:1501.07083 \[astro-ph.SR\]](#).
 - [96] A. Gaspar and G. Rieke, New HST data and modeling reveal a massive planetesimal collision around Fomalhaut, *Proceedings of the National Academy of Science* **117**, 9712 (2020), [arXiv:2004.08736 \[astro-ph.EP\]](#).
 - [97] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Searches for Continuous Gravitational Waves from 15 Supernova Remnants and Fomalhaut b with Advanced LIGO, *The Astrophysical Journal* **875**, 122 (2019), [arXiv:1812.11656 \[astro-ph.HE\]](#).
 - [98] D. Beniwal, P. Clearwater, L. Dunn, A. Melatos, and D. Ottaway, Search for continuous gravitational waves from ten H.E.S.S. sources using a hidden Markov model, *Physical Review D* **103**, 083009 (2021), [arXiv:2102.06334 \[astro-ph.HE\]](#).
 - [99] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Open data from the first and second observing runs of advanced ligo and advanced virgo, *SoftwareX* **13**, 100658 (2021), [arXiv:1912.11716 \[gr-qc\]](#).
 - [100] LIGO Scientific Collaboration and Virgo Collaboration, Gravitational Wave Open Science Center - Advanced LIGO O2 Data Release, <https://www.gw-openscience.org> (2019).
 - [101] Goetz, Evan (LIGO Scientific Collaboration and Virgo Collaboration), Segments used for creating standard SFTs in O2 data, <https://dcc.ligo.org/LIGO-T1900085/public>.
 - [102] G. Ashton, R. Prix, and D. I. Jones, Statistical characterization of pulsar glitches and their potential impact on searches for continuous gravitational waves, *Physical Review D* **96**, 063004 (2017), [arXiv:1704.00742 \[gr-qc\]](#).
 - [103] R. Prix and Y. Itoh, Global parameter-space correlations of coherent searches for continuous gravitational waves, *Classical and Quantum Gravity* **22**, S1003 (2005), [arXiv:gr-qc/0504006](#).
 - [104] LIGO Scientific Collaboration, *LIGO Algorithm Library - LALSuite*, free software (GPL) (2018).
 - [105] J. T. Whelan, S. Sundaresan, Y. Zhang, and P. Peiris, Model-Based Cross-Correlation Search for Gravitational Waves from Scorpius X-1, *Physical Review D* **91**, 102005 (2015), [arXiv:1504.05890 \[gr-qc\]](#).
 - [106] S. Walsh, M. Pitkin, M. Oliver, S. D’Antonio, V. Dergachev, A. Królak, *et al.*, Comparison of methods for the detection of gravitational waves from unknown neutron stars, *Physical Review D* **94**, 124010 (2016), [arXiv:1606.00660 \[gr-qc\]](#).

- [107] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, *et al.*, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, [Nature Methods](#) **17**, 261 (2020).
- [108] C. Messenger, R. Prix, and M. A. Papa, Random template banks and relaxed lattice coverings, [Physical Review D](#) **79**, 104017 (2009), [arXiv:0809.5223 \[gr-qc\]](#).
- [109] B. Allen and A. A. Shoom, Template banks based on \mathbb{Z}^n and A_n^* lattices, arXiv e-print (2021), [arXiv:2102.11631 \[astro-ph.IM\]](#).
- [110] P. B. Covas *et al.*, Identification and mitigation of narrow spectral artifacts that degrade searches for persistent gravitational waves in the first two observing runs of Advanced LIGO, [Physical Review D](#) **97**, 082002 (2018), [arXiv:1801.07204 \[astro-ph.IM\]](#).
- [111] C. Biwer, D. Barker, J. C. Batch, J. Betzwieser, R. P. Fisher, E. Goetz, *et al.*, Validating gravitational-wave detections: The Advanced LIGO hardware injection system, [Physical Review D](#) **95**, 062002 (2017), [arXiv:1612.07864 \[astro-ph.IM\]](#).
- [112] J. Aasi *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Directed search for continuous gravitational waves from the Galactic center, [Physical Review D](#) **88**, 102002 (2013), [arXiv:1309.6221 \[gr-qc\]](#).
- [113] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), First low-frequency Einstein@Home all-sky search for continuous gravitational waves in Advanced LIGO data, [Physical Review D](#) **96**, 122004 (2017), [arXiv:1707.02669 \[gr-qc\]](#).
- [114] O. J. Piccinni, P. Astone, S. D’Antonio, S. Frasca, G. Intini, I. La Rosa, P. Leaci, S. Mastrogiovanni, A. Miller, and C. Palomba, Directed search for continuous gravitational-wave signals from the Galactic Center in the Advanced LIGO second observing run, [Physical Review D](#) **101**, 082004 (2020), [arXiv:1910.05097 \[gr-qc\]](#).
- [115] J. Ming, M. A. Papa, H.-B. Eggenstein, B. Machenschalk, B. Steltner, R. Prix, B. Allen, and O. Behnke, Results from an Einstein@Home search for continuous gravitational waves from G347.3 at low frequencies in LIGO O2 data, arXiv e-print (2021), [arXiv:2108.02808 \[gr-qc\]](#).