Semicoherent method to search for continuous gravitational waves

Sabrina D'Antonio,[1] Cristiano Palomba,[2] Pia Astone,[2] Simone Dall'Osso,[2] Stefano Dal Pra,[3] Sergio Frasca,[4] Paola Leaci,[4,2] Federico Muciaccia,[4] Ornella J. Piccinni,[5] Lorenzo Pierini,[4,2] and Marco Serra[2]

[1]*INFN, Sezione di Roma Tor Vergata, I-00133 Roma, Italy*
[2]*INFN, Sezione di Roma, I-00185 Roma, Italy*
[3]*INFN-CNAF, I-40127 Bologna, Italy*
[4]*Università di Roma "La Sapienza," I-00185 Roma, Italy*
[5]*Institut de Física d'Altes Energies (IFAE), E-08193 Barcelona, Spain*

(Received 20 June 2023; accepted 11 October 2023; published 6 December 2023)

The emission of continuous gravitational waves (CWs), with duration much longer than the typical data taking runs, is expected from several sources, notably spinning neutron stars, asymmetric with respect to their rotation axis and more exotic sources, like ultralight scalar boson clouds formed around Kerr black holes and subsolar mass primordial binary black holes. Unless the signal time evolution is well predicted and its relevant parameters accurately known, the search for CWs is typically based on semicoherent methods, where the full dataset is divided into shorter chunks of given duration, which are properly processed and then incoherently combined. In this paper, we present a semicoherent method, in which the so-called "five-vector" statistics is computed for the various data segments and then summed after the removal of the Earth Doppler modulation and signal intrinsic spin-down. The method can work with segment duration of several days, thanks to a double-stage procedure in which an initial rough correction of the Doppler and spin-down is followed by a refined step in which the residual variations are removed. This method can be efficiently applied for directed searches, where the source position is known to a good level of accuracy, and in the candidate follow-up stage of wide-parameter space searches.

DOI: 10.1103/PhysRevD.108.122001

I. INTRODUCTION

Gravitational wave astronomy started in 2015 with the detection of gravitational waves (GWs) emitted in the last stages of the coalescence of a black hole binary system [1]. To date, a total of 90 binary events have been observed by the LIGO [2] and Virgo detectors [3], all due to the coalescence of compact binaries made, in the vast majority of the cases, by a pair of black holes and, in very few cases, by a pair of neutron stars or by a black hole and a neutron star [4].

However, many more kinds of sources are expected to exist; see, e.g., [5–7] for recent reviews. In particular, we are interested in continuous gravitational wave (CW), sources, which emit signals with duration longer than one day, for which the modulations due to Earth's rotation play a critical role. This kind of emission characterizes, for instance, spinning neutron stars asymmetric with respect to the axis of rotation. Asymmetric spinning neutron stars, isolated or in a binary system, are considered the prototypical source of CWs. Their detection will be a fundamental milestone in GW physics because they can be observed for very long times, becoming true laboratories for fundamental physics and astrophysics. Recently, more exotic sources of CWs have been also proposed which, if detected, would shed light on several important aspects of fundamental physics and cosmology, including dark matter. One example is represented by ultralight boson clouds that may form around Kerr black holes, as a consequence of a superradiance process [8,9]. Once formed, the cloud will dissipate through the emission of a CW signal, with a secular spin-up in frequency. Another interesting example is binary systems made of subsolar mass primordial black holes [10,11]. Such systems, for values of the chirp mass smaller than $\sim 10^{-3}$ solar masses, are characterized by a very long coalescence time and thus emit a nearly periodic signal with a slowly increasing frequency.

The search for CWs can be based on optimal fully coherent methods (using matched filtering), only when the source sky position, frequency, and frequency evolution are accurately known; see, e.g., [12,13]. Otherwise, regardless of the source, the search is computationally very heavy and relies on semicoherent approaches that strongly reduce the required computing power—with respect to matched filtering—at the price of a sensitivity loss (see, e.g., [14–27]). Several such methods have been applied to LIGO-Virgo data from various runs; see, e.g., [23,23,28–32] for recent results concerning all-sky searches and [33–35] for general reviews on CW search methods. Generally speaking, in a semicoherent method, the full dataset is divided into several shorter chunks that are independently processed and then recombined incoherently, i.e., without taking into account the signal phase. The various approaches differ under different aspects: the segment length, the way in which each data segment is processed, the statistics used to measure the significance of the results, the way in which noise artifacts are dealt with, and in several implementation details. In any case, the goal is an analysis method that is as sensitive, robust, and computationally cheap as possible.

In this paper, we introduce a new semicoherent method of analysis that exploits the sidereal amplitude modulation of CW signals, induced by the time-varying response of the detector. The method is built on the so-called five-vector statistics [13], largely used in targeted searches for known pulsars, which is here adapted to a semicoherent scheme. This new pipeline allows one to make sensitive and computationally cheap searches, with coherence time of several sidereal days, toward specific sky directions. As such, it can be used, for instance, to make directed searches toward globular clusters or the Galactic Center and for the follow-up of outliers found in wide-parameter space searches (like all-sky searches).

The paper is organized as follows. In Sec. II, a brief introductory description of the method is given. In Sec. III, the computation of the coarse-frequency grid is described. In Sec. IV, the reader is briefly reminded of five-vector statistics, and its use in the context of a semicoherent method discussed. Section V is devoted to the outline of the removal of the residual Doppler effect and the computation of the semicoherent statistics. Section VI extends the algorithm to the presence of a source intrinsic spin-down. Section VII describes the experimental procedure used to estimate the method sensitivity, including a comparison with the theoretical computation, discusses some implementation details of the analysis procedure, and, finally, briefly comments on the computational cost of the algorithm. In Sec. VIII, the validation tests done with hardware and software-simulated signals are discussed. Finally, Sec. IX contains the conclusions. Details on the theoretical sensitivity computation are given in Appendix A.

II. OVERVIEW OF THE METHOD

In this section, we give an overview of the analysis method, leaving details to following sections. A scheme of the method is shown in Fig. 1, where, as an example, data from two detectors are considered and a fixed-sky direction is assumed. The starting point is represented by band sampled data (BSD) files containing detector calibrated data, each file covering 1 month and a band of 10 Hz, and cleaned from short duration disturbances [36]. For a given target of the search to be performed, a *coarse* grid on the parameter space, consisting of frequency, spin-down, and possibly sky position, is built. The range covered by each parameter depends on the specific target. For instance, for the follow-up of an all-sky outlier, it will consist of a small interval around the outlier parameters, determined by the uncertainty associated with each of them. In the case of a directed search toward, e.g., a globular cluster, we will likely employ a large range of values for frequency and spin-down, which are typically unknown, and just one, or a few, sky position(s) in order to cover the globular cluster extension.

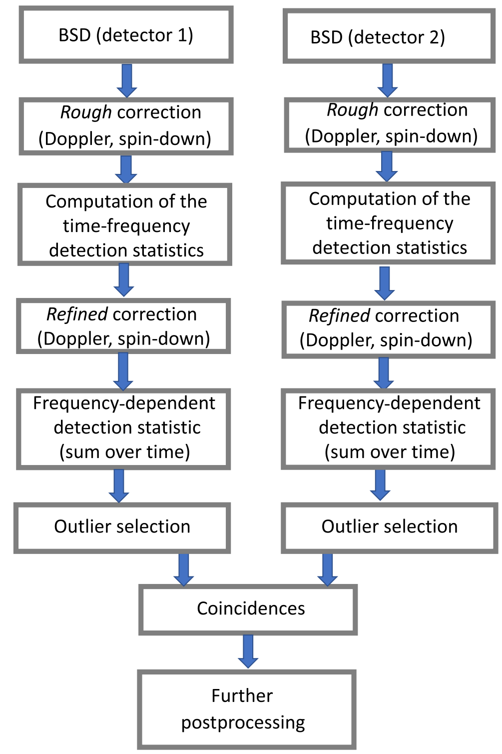

FIG. 1. Scheme of the analysis method, assuming the analysis is done toward a single sky direction and that data from two detectors are used. See text for more details.

For each dataset, the data time series is subject to a heterodyne correction [36] of the Doppler modulation (Sec. III) and intrinsic source spin-down (Sec. VI), done over the coarse grid points. This allows for a partial substraction of those frequency variations. The grid is built in such a way that, over data segments of duration T_{FFT} (where FFT represents the fast Fourier transform), any residual frequency variation is confined within one frequency bin of width $\delta f = 1/T_{\mathrm{FFT}}$. The choice of T_{FFT} and then the corresponding number of grid points is a matter of compromise between sensitivity (longer T_{FFT}) and computational cost (higher number of grid points). When a large-frequency range is considered, for practical reasons, this is split into several smaller bands, say 1 Hz wide, and the analysis steps described in the following are repeated for each of them. For each time segment of duration T_{FFT}, the five-vector statistics is computed (Sec. IV) for each frequency bin (within a 1 Hz subband) and a time-frequency map of the statistics values is built. The residual variation in frequency and spin-down is then removed by building a "refined" grid and applying the needed corrections by properly shifting the frequency bins in the time-frequency plane (Secs. V and VI). At this point, a total statistics is built by summing the statistics computed over each segment. Finally, a fixed number of the most significant outliers are selected in each 1 Hz subband, over the whole sky and spin-down ranges [31]. The outliers are subject to subsequent analyses steps, which have been commonly used in previous searches [31] and for this reason are not discussed in detail in this paper. In brief, coincidences among outliers found in different datasets (see penultimate box in Fig. 1) consist of initially taking only those for which the *distance* in the parameter space is below some predefined threshold, ranking them on the basis of a combination of distance and critical ratio and keeping a given number of the most significant among these; see, e.g., [37] for a detailed discussion. These surviving outliers are subject to additional postprocessing [31,38,38] to discard those that are not compatible with an astrophysical signal. Further iterations of the semicoherent procedure are possibly applied to deeply follow any remaining candidate. This method represents the evolution and improvement of a previous simpler approach used in [39,40].

III. COARSE FREQUENCY GRID

In this section, we describe how the coarse frequency grid is built, deferring the discussion on spin-down correction to Sec. VI and assuming a fixed-sky position. At each point of this grid, a heterodyne correction is applied in order to partially remove the Doppler effect.

A CW signal before reaching the detector can be represented, in complex notation, as $h(t) = h_0 e^{i\phi(t)}$. At the detector, the signal is characterized by an amplitude modulation, not relevant here and discussed in Sec. IV, and a frequency modulation. The corresponding phase evolution for a signal with intrinsic frequency f_0 and assuming, for simplicity, zero spin-down (an assumption which will be relaxed later) can be expressed as

$$\phi(t) = \phi_0 + \omega_0\left(t + \frac{\vec{r}_\mathrm{t}\cdot\hat{n}}{c}\right), \tag{1}$$

where $\omega_0 = 2\pi f_0$, $\vec{n}$ is the unit vector identifying the direction to the source, and $\vec{r}_\mathrm{t}$ is the time-dependent detector position in the reference frame of the solar system barycenter (SSB). The Roemer delay $\frac{\vec{r}_\mathrm{t}\cdot\hat{n}}{c}$ is responsible for the Doppler effect due to the motion of the detector.

If we perform a heterodyne correction over the whole dataset to compensate the Doppler effect, using the right sky position and a wrong angular frequency ω_0', i.e., we multiply the data by a factor $e^{-i\omega_0'\frac{\vec{r}_\mathrm{t}\cdot\hat{n}}{c}}$, the resulting signal phase is

$$\phi_\mathrm{corr}(t) = \omega_0 t + (\omega_0 - \omega_0')\frac{\vec{r}_\mathrm{t}\cdot\hat{n}}{c}. \tag{2}$$

Because of the wrong correction, the resulting signal frequency is affected by a residual Doppler modulation,

$$f(t) = \frac{1}{2\pi}\frac{d\phi_\mathrm{corr}}{dt} = f_0 + (f_0 - f_0')\frac{\vec{v}_\mathrm{t}\cdot\hat{n}}{c}, \tag{3}$$

where $\vec{v}_\mathrm{t}$ is the detector velocity in the SSB reference frame and $f_0' = \omega_0'/2\pi$. At two different times t_1 and t_2, the wrongly corrected and the true signal frequencies differ by

$$f(t_1) - f_0 = (f_0 - f_0')\frac{\vec{v}_{\mathrm{t}_1}\cdot\hat{n}}{c}, \tag{4}$$

$$f(t_2) - f_0 = (f_0 - f_0')\frac{\vec{v}_{\mathrm{t}_2}\cdot\hat{n}}{c}. \tag{5}$$

It follows that

$$f(t_1) - f(t_2) = (f_0 - f_0')\left(\frac{\vec{v}_{\mathrm{t}_2}\cdot\hat{n}}{c} - \frac{\vec{v}_{\mathrm{t}_1}\cdot\hat{n}}{c}\right). \tag{6}$$

We use this equation to define the maximum time interval $T_\mathrm{FFT} = t_2 - t_1$ such that the full signal power is confined within a single frequency bin $\delta f = 1/(t_2 - t_1)$. Specifically, the frequency variation given by Eq. (6) over the time interval T_FFT must meet the condition

$$\left|(f_0 - f_0')\left(\frac{\vec{v}_{\mathrm{t}_2}\cdot\hat{n}}{c} - \frac{\vec{v}_{\mathrm{t}_1}\cdot\hat{n}}{c}\right)\right|_\mathrm{max} \le \frac{1}{T_\mathrm{FFT}}, \tag{7}$$

where the maximum of the expression in parentheses is taken across the whole observing time. For a fixed T_FFT, and for a given detector and sky position, Eq. (7) provides the maximum allowed coarse-frequency step $\Delta_\mathrm{f} = f_0 - f_0'$ when looking for a CW source emitting at an unknown frequency f_0. When searching over a frequency band of width B_0 and starting point f_start, the grid frequencies $f_0' = f_\mathrm{start} + K\cdot\Delta_\mathrm{f}$, $K = 1, \ldots, M$, where $M = \mathrm{round}(B_0/\Delta_\mathrm{f})$, will thus differ from f_0 by at most Δ_f. It is important to note that here we are not taking into account the signal sidereal modulation, which determines an additional spread of the signal power, which will be considered in Sec. IV. Through Eq. (7) we can then set the values for the pair $(T_\mathrm{FFT}, \Delta_\mathrm{f})$ in order to find the best compromise between computational load and search sensitivity. As an example, in Fig. 2 we plot the frequency grid step Δ_f as a function of the FFT duration T_FFT for a source direction ($\lambda = 193.3162°$, $\beta = -30.9956°$). Increasing T_FFT improves the sensitivity, at the same time requiring a smaller step Δ_f and hence a larger number of frequency grid points over B_0 to for which to search. Moreover, in the more general situation in which position and spin-down are not exactly known, the number of grid points in the parameter space increases proportionally to T_FFT^4 [20], further increasing the computational cost of the analysis.

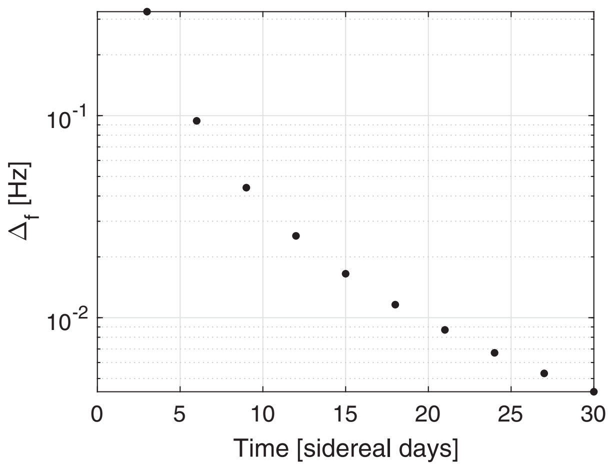

FIG. 2. Frequency coarse step Δ_f as a function of data segment duration T_{FFT}, for a source direction ($\lambda = 193.3162°$, $\beta = -30.9956°$).

For implementation purposes, see Sec. VII A, T_{FFT} is taken as an integer number of sidereal days, $T_{\oplus} = 86164.09$ sec.

IV. FIVE-VECTOR STATISTICS

After the heterodyne coarse correction of the Doppler effect, a CW signal is still not monochromatic in each data segment if T_{FFT} is larger than one sidereal day. This is due to the sidereal modulation induced by the time-varying detector response toward the source direction. In this section, we remind the reader of the definition of the five-vector statistics, which allows one to take this effect into account.

The signal at the detector, after the Doppler correction, is given by [13]

$$h(t) = \mathrm{Re}\left[H_0\left(H_+A_+(t) + H_\times A_\times(t)\right)e^{i(\omega_0 t+\phi_0)}\right], \quad (8)$$

where H_0 is the signal amplitude, and

$$H_+ = \frac{\cos 2\psi - i\eta \sin 2\psi}{\sqrt{1+\eta^2}},$$
$$H_\times = \frac{\sin 2\psi + i\eta \cos 2\psi}{\sqrt{1+\eta^2}}. \quad (9)$$

In Eq. (9), $\eta = \frac{-2\cos\iota}{1+\cos^2\iota}$, with ι being the angle between the source rotation axis and the line of sight, while ψ is the wave polarization angle. The two functions $A_{+/\times}(t)$ are periodic functions of Earth's sidereal angular frequency $\Omega_{\oplus} = 2\pi/T_{\oplus}$. They are linked to the classical radiation pattern functions $F_{+/\times}(\psi; t)$ [12] by $A_{+/\times} = F_{+/\times}(\psi = 0)$. The signal amplitude H_0 in Eq. (8) is related to the classical strain amplitude h_0 by the relation

$$h_0 = \frac{2H_0}{\sqrt{1+6\cos^2\iota + \cos^4\iota}}. \quad (10)$$

The sidereal modulation, which affects both the amplitude and the phase of the signal, produces a splitting of the signal power in five frequencies, $\omega_0 \pm \mathbf{k}\Omega_{\oplus}$, $\mathbf{k} = 0, \pm 1, \pm 2$. This was exploited in [13] to introduce a detection statistics, based on the concept of a five-vector, defined as the complex vector containing the Fourier components of the signal at the five frequencies associated with the sidereal modulation. The five-vector of a given time series of duration T_{FFT}, and at an angular frequency ω_0, is given by (working, for simplicity of notation, in the continuous)

$$\mathbf{X} = \int_{T_{\mathrm{FFT}}} x(t)e^{-i(\mathbf{k}\Omega_{\oplus}+\omega_0)t}dt. \quad (11)$$

In addition to the data five-vector $\mathbf{X}$, the signal template five-vectors $\mathbf{A}_+, \mathbf{A}_\times$ are also computed for each ω_0 by means of Eq. (11), replacing the time series $x(t)$ with the two functions $A_{+/\times}(t)$. Although analytical formulas have been derived for $A_{+/\times}(t)$ (see, e.g., [13]), the corresponding five-vectors are computed numerically in order to take into account features of real data, for instance, data gaps, which would be difficult to deal with otherwise. These three five-vectors are then combined, computing two matched filters of the data with the signal templates,

$$\hat{H}_+ = \frac{\mathbf{X}\cdot\mathbf{A}^+}{|\mathbf{A}^+|^2} \quad \text{and} \quad \hat{H}_\times = \frac{\mathbf{X}\cdot\mathbf{A}^\times}{|\mathbf{A}^\times|^2}. \quad (12)$$

It can be shown [13] that $\hat{H}_{+/\times}$ are estimators of the quantities $H_0 H_{+/\times} e^{i\phi_0}$ in Eq. (8). They are used to define the five-vector statistics as

$$\mathcal{S} = |\mathbf{A}_+|^4|\hat{H}_+|^2 + |\mathbf{A}_\times|^4|\hat{H}_\times|^2, \quad (13)$$

which collects the signal power, spread due to the sidereal modulation, over a time interval T_{FFT} and which also depends on the detector noise through the data five-vector.

V. REMOVAL OF THE RESIDUAL DOPPLER AND COMPUTATION OF THE FINAL STATISTICS

In principle, once we have computed the five-vector statistics for all frequencies of the grid and over all segments of duration T_{FFT}, the final statistics value would be simply obtained by summing all the five-vector statistics values at fixed frequency. In practice, however, we have to take into account the remaining frequency spread due to the coarse Doppler correction described in Sec. III; otherwise, the signal power at a given frequency would not be fully recovered.

The choice of Δ_{f} for a given T_{FFT} on the basis of Eq. (7) guarantees that, in each time interval T_{FFT}, the signal power

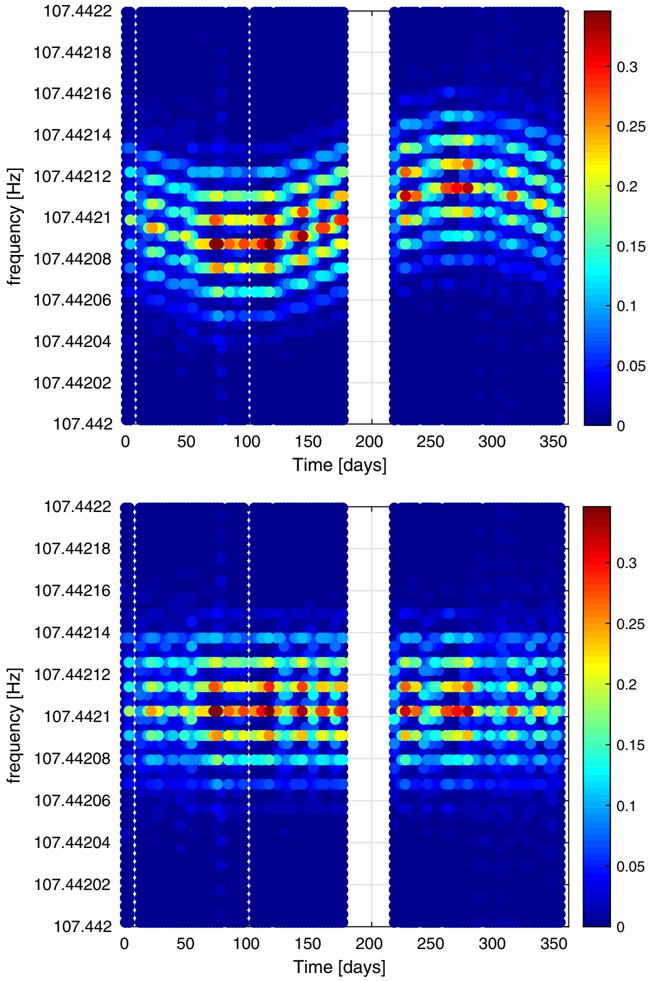

FIG. 3. Time-frequency plot of the statistics computed with $T_{\mathrm{FFT}} = 3T_{\oplus}$ for a simulated source with unitary amplitude, sky position, and polarization parameters corresponding to signal s_1 in Table I, frequency $= 107.4421$ Hz, zero spin-down, and assuming it is observed by the LIGO Livingston detector. Top: refers to the coarse correction done for the grid frequency value nearest to the true signal frequency; bottom: refers to the corresponding refined correction. The color bar gives the value of the five-vector statistics computed through Eq. (13). Even though the plots have been obtained considering only the signal, i.e., without detector noise, data gaps of the O3 run have been taken into account in the simulation.

remains confined into a frequency bin. As a consequence, in each time segment, the five-vector statistics is not affected by the not-optimal Doppler correction. The residual Doppler, however, acts by offsetting the values of the statistics in different segments of the time-frequency plan, as can be clearly seen in the top plot of Fig. 3, obtained considering a simulated signal of unitary amplitude with parameters shown in the first row of Table I ("s_1"), generated by assuming it is observed by the LIGO Livingston detector. The plot shows the time-frequency map of the five-vector statistics of the signal after the coarse Doppler correction for the coarse-frequency value nearest

TABLE I. Position (in ecliptical coordinates) and polarization parameters of the two sets of simulated signals used to test the pipeline performances.

Signal	λ (°)	β (°)	$\cos\iota$	ψ (°)
s_1	193.3162	−30.9956	−0.081	25.4390
s_2	276.8964	−61.1909	0.463	−20.8530

to the true signal frequency, taking data segments of duration $T_{\mathrm{FFT}} = 3T_{\oplus}$. In this case, the residual Doppler amounts to about 8×10^{-5} Hz, which is much smaller than the full uncorrected Doppler shift, $\simeq 0.01$ Hz, but larger than the frequency bin of $\frac{1}{3T_{\oplus}} \simeq 3.9 \times 10^{-6}$ Hz. Therefore, before summing the statistics on the time axis, to obtain the final semicoherent statistics, we need to properly shift the frequencies in order to realign them correctly. Specifically, from Eq. (3) it follows that

$$f_0 = \frac{f(t) + f_0' \frac{\vec{v}_t \cdot \hat{n}}{c}}{\left(1 + \frac{\vec{v}_t \cdot \hat{n}}{c}\right)}, \tag{14}$$

where f_0' denotes the frequency grid values. We have to shift the frequencies $f(t)$ by an amount $D_f(t)$ such that $f(t) - D_f(t) = f_0$. It thus follows, from Eq. (14), that

$$D_f(t) = \frac{(f(t) - f_0') \frac{\vec{v}_t \cdot \hat{n}}{c}}{\left(1 + \frac{\vec{v}_t \cdot \hat{n}}{c}\right)}, \tag{15}$$

and the new corrected frequencies are obtained as

$$f_c(t) = f(t) - D_f(t). \tag{16}$$

By construction, this shift will realign the signal peaks only when f_0' is the grid value nearest to the true signal frequency. The bottom plot of Fig. 3 shows the time-frequency plot of the five-vector statistics after the refined Doppler correction, considering the nearest point to the signal frequency. As expected, after the refined correction, the statistics peaks are aligned. Because of the computation of scalar products between the signal and the templates at different frequency bins [Eq. (12)], the statistics presents nine prominent peaks, although both the signal and the templates are characterized by only five peaks. See Sec. VI for a more detailed discussion. In the simulation, we have taken into account gaps in O3 Livingston data: the empty region in the plots corresponds to a one-month detector commissioning break that occurred during the run. Figure 4 shows the final statistics before and after the residual Doppler correction. Again, the effect of the correction is clearly visible and produces stronger peaks. When the refined step is applied to the other coarse grid points, the resulting correction will be less accurate and produce less significant peaks in the final statistics.

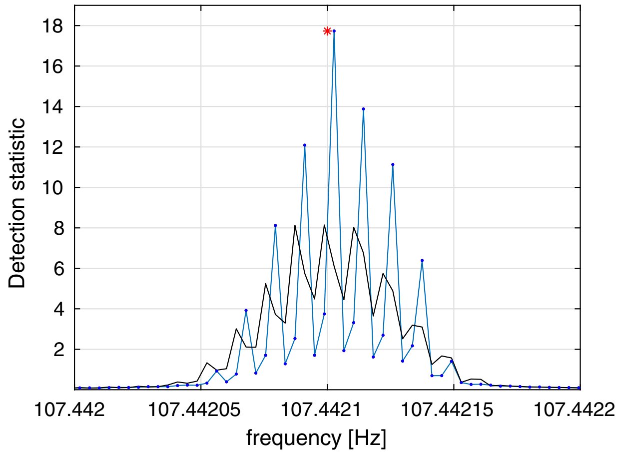

FIG. 4. Global statistics, see Eq. (19), before (black, continuous line) and after (blue, dotted line) the refined correction for a simulated signal with parameters given in the caption of Fig. 3 and assuming it is observed by the LIGO Livingston detector. The asterisk indicates the signal frequency.

VI. SPIN-DOWN CORRECTION

In previous sections we have focused on the Doppler correction, neglecting any intrinsic frequency variation of the signal. In this section, we describe how to remove from the data the frequency variations due to the spin-down. In particular, we consider here only the correction for the first-order spin-down term (i.e., the first time derivative of the frequency). As discussed in Appendix B, for any fixed T_{FFT} longer than one sidereal day, some portions of the potentially explorable parameter space—especially for very large absolute values of the first-order spin-down—would require the second spin-down term to also be taken into account. The correction of the second-order spin-down is deferred to a future work. Suppose we carry out a search for a source located at a given sky position and emitting a CW signal with unknown frequency and spin-down, ranging, respectively, in the interval $[f_{\mathrm{min}}, f_{\mathrm{max}}]$ and $[\dot{f}_{\mathrm{min}}, \dot{f}_{\mathrm{max}}]$. As for the Doppler, also in this case we first apply a coarse spin-down correction followed by a refined correction to subtract the residual spin-down frequency variation and to realign the frequencies of the statistics values due to a CW signal. For a fixed coherence time T_{FFT}, the coarse spin-down step, defined as the maximum mismatch such that the signal power during the time interval T_{FFT}, is confined to a single frequency bin and is

$$\delta\dot{f}_0 = \frac{\delta f}{2T_{\mathrm{FFT}}}. \tag{17}$$

For that given sky position, we then perform a coarse heterodyne data correction (for the Earth Doppler effect), to scan the frequency range of interest at steps Δ_{f} as discussed in Sec. III and, for each frequency, a coarse heterodyne spin-down correction at spin-down values $\dot{f}_{\mathrm{n}} = \dot{f}_{\mathrm{min}} + n \times \delta\dot{f}_0$, where $n = 0, \ldots, \mathrm{round}(|\dot{f}_{\mathrm{max}}|/\delta\dot{f}_0)$. After this stage, the time-frequency plot of the statistics is affected by the residual Doppler and spin-down effects, due to nonoptimal signal corrections, that propagate over the observation time as shown in the top plot of Fig. 5, which refer to the values of the coarse frequency and spin-down nearest to the true signal values. In this example, we show for illustrative purposes the results of the analysis performed on a fake signal with unitary amplitude, emitted by a source with sky position and polarization parameters corresponding to signal s_1 in Table I, $f = 107.4421$ Hz, $\dot{f} = -8.3410^{-11}$ Hz/s, assuming it is observed by the LIGO Livingston detector. We have run the algorithm over the band [107–108] Hz, with step $\Delta_{\mathrm{f}} = 0.3275$ Hz, as given by Eq. (7) for the specific value of $T_{\mathrm{FFT}} = 3T_{\oplus}$. At this point, we remove the residual Doppler due to the approximate frequency correction by properly shifting the frequency of the statistics values, see Eq. (16). The middle plot of Fig. 5 shows that the signal, after the residual Doppler removal, is only affected by the residual spin-down, due to the previous not-optimal spin-down correction. Now we make a loop over the refined spin-down grid, with step $\delta\dot{f} = \frac{\delta f}{2\mathrm{T}_{\mathrm{obs}}}$, which now covers the interval between each pair of successive spin-down values of the coarse grid. The grid step is chosen in such a way that the whole signal power, over the total observation time $\mathrm{T}_{\mathrm{obs}}$, is confined within a single frequency bin. In practice, the correction is performed by shifting the frequency of the statistics values according to the rule

$$f_{\mathrm{c}}(t) = f(t) - K\delta\dot{f} \cdot t, \tag{18}$$

where $K = 1, \ldots, \mathrm{round}(\delta\dot{f}_0/\delta\dot{f}) - 1$, and the time t refers to the central time of each data segment. The bottom plot of Fig. 5 shows the result of the refined corrections for both Doppler and spin-down effects, done using the refined spin-down value nearest to the correct one: the time-frequency values due to the signal are now aligned in frequency. Finally, the statistics are added on the time axis. Figure 6 shows the cumulative corrected detection statistics before and after the spin-down correction. We find that both the frequency and the spin-down of the fake signal have been correctly recovered inside the refined frequency and the refined spin-down bins.

As already noticed in Sec. V, from Figs. 5 and 6 it can be seen that actually there are nine frequency values associated with the injected signal, because the convolution between the five-comb of the data and the five-comb of the theoretical kernel leads to a nine-comb. The nine peaks of the statistics are separated by one sidereal frequency bin $1/T_{\oplus}$ and their relative amplitude depends on the unknown signal polarization. For the specific case shown in Figs. 5 and 6, the central peak is the most significant. In general, the effect of signal polarization, combined with that of noise fluctuations, can result in a most prominent peak

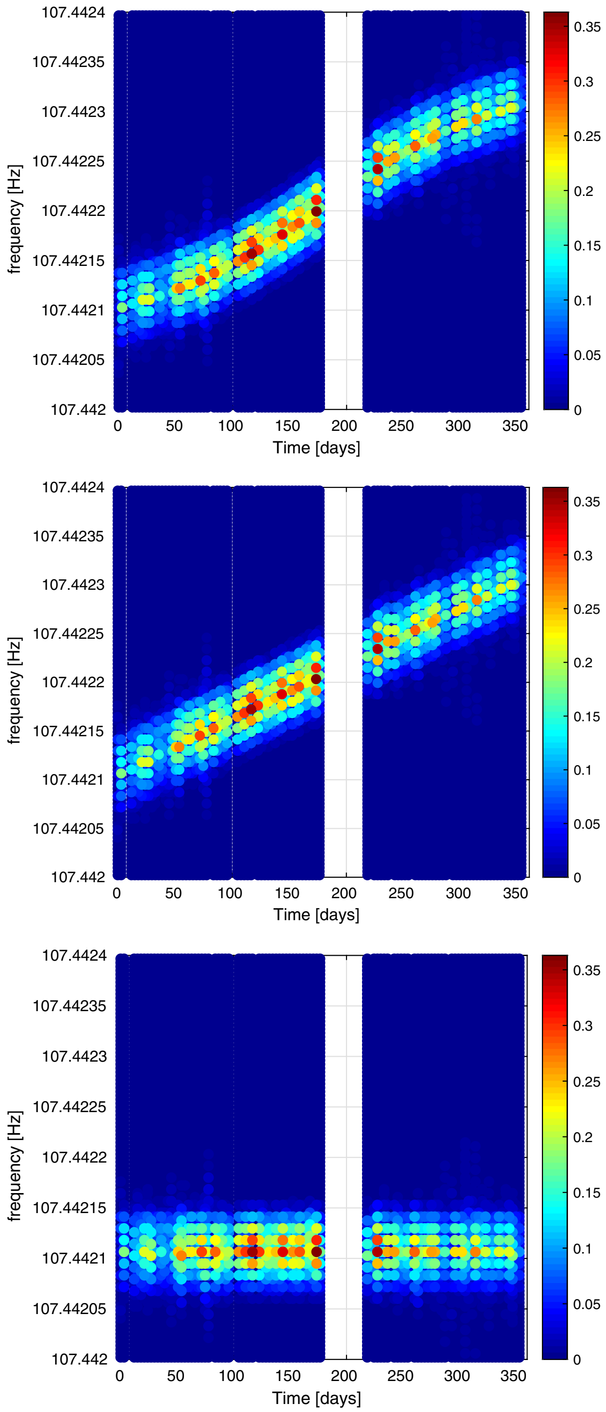

FIG. 5. Time-frequency plot of the detection statistics for a simulated signal with unitary amplitude, parameters given in the text, and assuming it is observed by the LIGO Livingston detector. The analysis has been carried out with $T_{\rm FFT} = 3T_{\oplus}$. Top: statistics after Doppler and spin-down coarse correction, using the frequency and spin-down values nearest to the correct signal values. Middle: statistics after refined Doppler correction. Bottom: statistics after refined spin-down correction, done using the refined spin-down value nearest to the correct one. Even though the plots have been obtained considering only the signal, i.e., without detector noise, data gaps of the O3 run have been taken into account in the simulation.

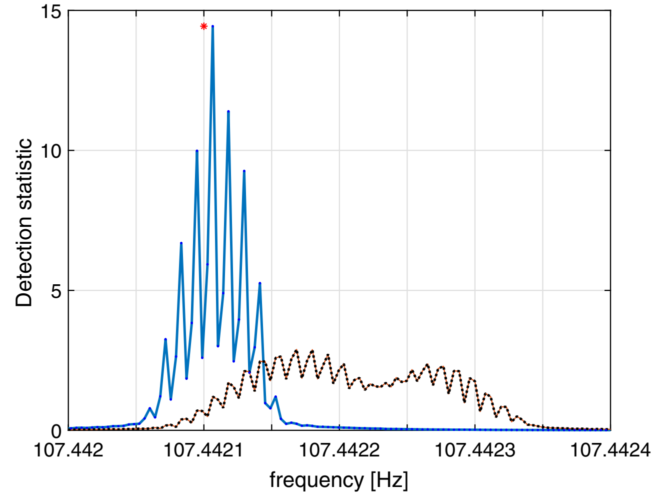

FIG. 6. Final statistics, see Eq. (19), as a function of the frequency, before (red, dotted line) and after (blue, continuous line) spin-down correction for $T_{\rm FFT} = 3T_{\oplus}$ for a simulated signal of unitary amplitude with parameters given in the text. The asterisk indicates the signal frequency.

different from the central one, that is, not corresponding to the real signal frequency. This has implications for the choice of the coincidence window, as we show in Sec. VII, when outliers found in different datasets are compared.

A. Summary of the method

We conclude this section with a brief summary of Secs. V and VI in order to clarify the main steps of the pipeline. Consider a dataset covering a frequency band of width B. For a chosen coherence time $T_{\rm FFT}$ and sky position, a coarse-frequency and spin-down grid are defined. For each point of the coarse-frequency grid, with step Δ_f derived from Eq. (7), and for each point of the coarse spin-down grid, with step $\delta\dot{f}_0$ given by Eq. (17), a coherent data correction over the whole dataset is performed via heterodyne correction. A time-frequency map of the detection statistics is computed over data segments of length $T_{\rm FFT}$, through the definition in Eq. (13). For each coarse-frequency value, the residual Doppler is removed by Eq. (16) to get the time-frequency Doppler-corrected statistics. On the corrected time-frequency map, further shifts are applied for the refined spin-down values between each pair of consecutive coarse spin-down bins, via Eq. (18). Finally, for each refined spin-down value, the sum of the statistics $\mathcal{S}_{\rm i}$ for each time segment of duration $T_{\rm FFT}$ is computed, where the index i identifies the frequency bin, of width $1/T_{\rm FFT}$. The final semicoherent statistics is then a function of the frequency

$$\mathcal{S}(\mathrm{f}) = \sum_{\mathrm{k}=1}^{\mathrm{N}} \mathcal{S}_{\mathrm{k}}, \tag{19}$$

where the sum extends to the $\mathrm{N} = \mathrm{floor}(\frac{T_{\mathrm{obs}}}{T_{\mathrm{FFT}}})$ segments contained in the observation window. Outliers are selected on this final statistics, dividing the searched frequency band in a number of subbands and choosing a given number of the most significant candidates, based on the critical ratio (see Sec. VII for definition). Outliers found in different datasets are then subject to further analysis steps, as briefly discussed in Sec. II. The whole procedure can be repeated considering different sky positions, if needed.

VII. PIPELINE CHARACTERIZATION

This section is dedicated to characterizing the pipeline in terms of sensitivity and computational cost. We start, however, by providing a couple of implementation details that will be relevant also for real searches.

A. Implementation details

As anticipated in Sec. III, the data segment duration T_{FFT} is chosen to be a multiple integer of the sidereal day of Earth $T_{\oplus}$. In this way, the five-vector components correspond to integer frequency bins. Hence, any five-vector can be computed by selecting the proper frequency bins in a FFT of the data. This approach, first introduced in [41], brings a significant speedup (about 3 orders of magnitude) with respect to the computation based on the direct application of Eq. (11).

A second detail concerns the frequency discretization that can lead to losses in the recovered signal power up to 36%, due to the mismatch between the signal frequency and the central frequency of the bins [42]. A cheap method to reduce this effect consists of estimating the FFT values at half bins, using an "interbinning" interpolation [41,42],

$$X_{\mathrm{FFT},k+1/2} \approx \frac{\pi}{4}\left(X_{\mathrm{FFT},k} - X_{\mathrm{FFT},k+1}\right), \tag{20}$$

where $X_{\mathrm{FFT},k}$ denotes the value of the FFT sample at the kth frequency bin. The impact of interbinning in the sensitivity estimation will be discussed in the next section.

B. Sensitivity

The sensitivity is defined as the minimum strain amplitude detectable with a given confidence level (CL), which we choose to be 95%.

We have made an empirical estimation of the sensitivity via software injections of simulated signals in a few frequency bands of real detector data, which has been then extrapolated to the full frequency band 10–2048 Hz, as outlined in the following. We have used O3 LIGO Livingston data in three different 1-Hz frequency bands: [107, 108], [585, 586], and [883, 884] Hz. In each of them, two sets of 80 and 40 signals, denoted as s_1 and s_2, have been generated, respectively, with random frequency, while spin-down, position, and polarization parameters were fixed at the values given in Table I and added to the detector data. Each set of 80 and 40 signals has been injected from 10 to 15 times, each time with different values for the amplitude H_0 [see Eq. (8)], chosen in a range that is expected to contain the minimum detectable value, at the 95% confidence level. Data have been analyzed as they would be in a real search, considering the whole 1-Hz band but only two coarse spin-down bins[1] around the injected values (to save computing time). Both the coarse and the refined corrections have been applied.

For each set of injected signals of amplitude H_0 and for each spin-down value, after running the analysis we select the 300 most significant outliers, across the 1-Hz frequency band. As standard in several wide-parameter searches, the significance of a candidate is represented by its critical ratio (CR) [20] computed on the projection on the frequency axis of the time-frequency map of the detection statistics values and defined as

$$\mathrm{CR(f)} = \frac{\mathcal{S}(\mathrm{f}) - \mu_{\mathrm{n}}}{\sigma_{\mathrm{n}}}, \tag{21}$$

where $\mathcal{S}$ is given by Eq. (19), and μ_n and σ_n are the mean and standard deviation of the noise statistics. The noise statistics is evaluated by replacing the data five-vector $\mathbf{X}$, Eq. (12), with a noise five-vector whose components are randomly chosen over the 1-Hz frequency band so that it cannot represent a physical signal. Adapting the procedure typically used in searches for selecting coincidences among outliers found in different datasets [43], we choose as outliers those points in the search parameter space $(f, \dot{f})$ (the sky position is fixed), for which the a-dimensional distance from any of the injected signals is smaller than the coincidence window D_{max}. In other words, a signal is considered as detected if the dimensionless distance [20]

$$d = \sqrt{\left(\frac{\Delta f}{\delta f \times \Delta}\right)^2 + \left(\frac{\Delta \dot{f}}{\delta \dot{f}}\right)^2} \leq D_{\mathrm{max}}, \tag{22}$$

where $\Delta f, \Delta \dot{f}$ are the dimensional distances of the outlier from the injected signal, and the factor Δ weights in the proper way the frequency distance that can be as large as four sidereal frequency bins ($4/T_{\oplus}$). This is due to the unknown source polarization, as shown in Fig. 7. The choice of D_{max} has been studied in [43] in the context of all-sky searches. In practice, here we take $\Delta = 4$ (in units of sidereal frequency $1/T_{\oplus}$) and, conservatively, $D_{\mathrm{max}} = 2$.[2]

[1] Out of $M = \mathrm{round}(|\dot{f}_{\mathrm{max}}|/\delta\dot{f}_0)$, see discussion in Sec. III.

[2] In the case of coincidences among outliers found in different datasets, the value of D_{max} is chosen depending on the number of follow-ups that can be afforded, given an available amount of time and computing power. A bigger D_{max} allows for a more sensitive search, at the cost of a bigger number of outliers to be followed up.

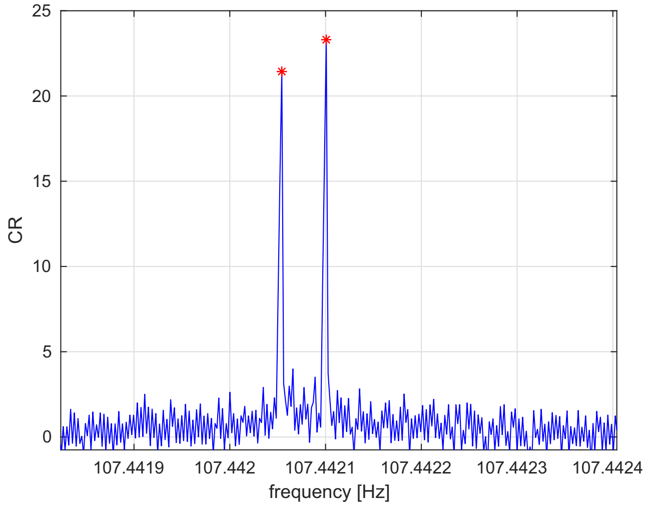

FIG. 7. Critical ratio $\mathcal{CR}$ over a frequency band containing the simulated signal s_2, see Table I, injected in O3 LIGO Livingston data and amplitude $H_0 = 10^{-25}$. The statistics shows two peaks having comparable significance. The two peaks are distant four sidereal frequency bins, with the smaller one at the signal frequency.

The expected number of random outliers, due to noise, that verify the above condition is much smaller than 1. For each signal amplitude, we count the fraction of detected signals and construct detection efficiency curves. Figure 8 shows an example of detection efficiency curves, for two different coherence times $T_{\mathrm{FFT}} = 5T_{\oplus}$ (circles) and $T_{\mathrm{FFT}} = 10T_{\oplus}$

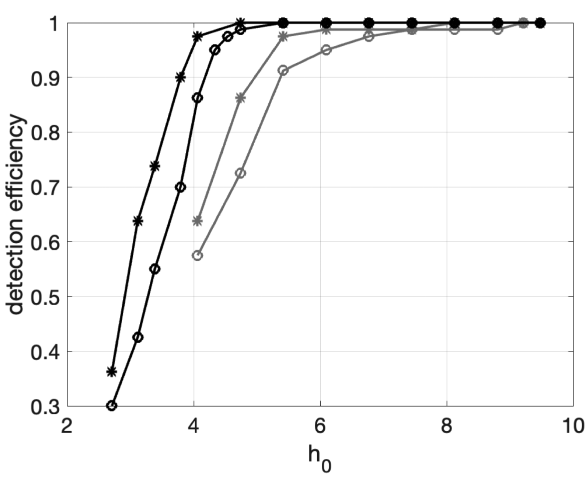

FIG. 8. Detection efficiency curves, done using several sets of 80 software-simulated signals with the parameters of the HI P3, injected in the band [107–108] Hz of LIGO Livingston O3 data with different amplitudes. The pair of gray curves are the detection efficiencies as a function of the signal amplitude, without interbinning and using a coherence time of five (circles) and ten (asterisks) sidereal days. The pair of black curves correspond to the application of the interbinning procedure, which significantly increases the detection efficiency. Signal amplitude is in units of 10^{-26}.

(asterisks), with (black curves) and without (gray curves) interbinning. The sensitivity gain is clearly visible both when we use ten sidereal days, rather than five, and when the interpolation is used. The signal amplitude H_0, such that 95% of the injected signals have a coincident outlier, corresponds to the sensitivity $H_{0,95\%}$ for a specific set of source parameters and for the specific 1-Hz frequency band we are considering. In practice, the 95% level is estimated by linearly interpolating the detection efficiency between the two data points immediately below and above that value.

In order to compute an average sensitivity on the standard strain amplitude h_0, we rescale $H_{0,95\%}$ with two factors, one to average over the sky and polarization from the specific sky position and polarization angle used in the injections, and one to go from $H_{0,95\%}$ to $h_{0,95\%}$, given by the mean value of the coefficient in Eq. (10), which implies the average over the cosine of the star's inclination angle. This is equivalent and computationally much cheaper to generating signals with random parameters. At the end, we have three values of the sensitivity for the three frequency bands we are considering, which correspond to regions with different detector noise. Each of these three sensitivity values has been extrapolated to the full frequency band by applying a further frequency-dependent scaling factor, given by the square root of the ratio $S_n(f)/\langle S_{n,j}\rangle$ of the data power spectrum estimation $S_n(f)$, over the band 10–2048 Hz, to the average power spectrum estimation (with respect to the frequency) in each of the three bands where the injections have been done, i.e., $\langle S_{n,j}\rangle$, with $j = 1, 2, 3$ corresponding, respectively, to [107, 108], [585, 586], and [883, 884] Hz. The final sensitivity estimation is given by the average (computed in each frequency bin) of the three sensitivity curves. Figure 9 shows the final sensitivity $h_{\mathrm{min},95\%}$ for O3 LIGO Livingston data for two different segment durations, $T_{\mathrm{FFT}} = 5T_{\oplus}$ and $T_{\mathrm{FFT}} = 10T_{\oplus}$. The number of combined data segments depends on T_{FFT} and on the presence of gaps in the data and is 60 for $T_{\mathrm{FFT}} = 5T_{\oplus}$ and 30 for $T_{\mathrm{FFT}} = 10T_{\oplus}$. From the sensitivity curves, we have estimated a sensitivity depth [44,45] $\mathcal{D} = \sqrt{S_n(f)}/h_{\mathrm{min},95\%}(f) = 124 \pm 4$ for $T_{\mathrm{FFT}} = 10T_{\oplus}$. The sensitivity values $h_{\mathrm{min},95\%}$ for specific frequency bands and for the above mentioned data segment duration T_{FFT}, together with the corresponding CR, $\mathrm{CR}_{95\%}$, are shown in Tables II and III. The estimated sensitivity has an associated false alarm rate of $O(10^{-5})$, which corresponds to an expectation of $O(10)$ outliers in Gaussian noise for a 1-Hz frequency band, one refined spin-down bin, and one sky location. We have also computed a theoretical sensitivity via a mixed analytical-numerical approach, which assumes Gaussian and stationary noise, as described in Appendix A. The last column of the tables reports such theoretical value. Overall, the theoretical values underestimate the empirical sensitivity, as expected due to non-Gaussian and nonstationary features of real data, by at most ∼10%–15%.

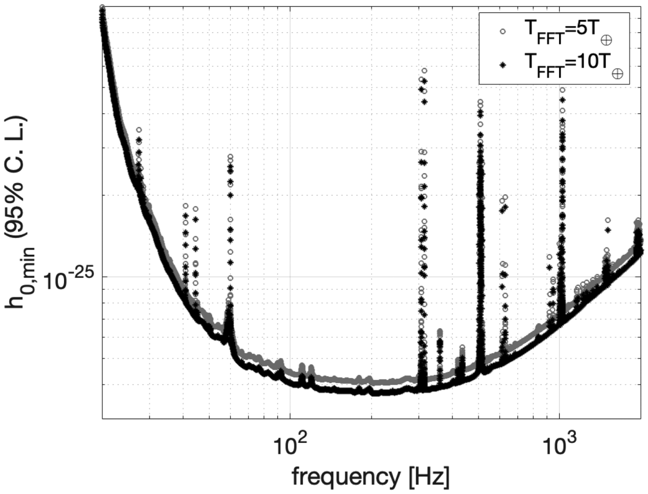

FIG. 9. Search estimated sensitivity for O3 LIGO Livingston (95% CL), obtained through the injection of simulated signals in detector real data. The curves correspond to two different segment duration, $5T_{\oplus}$ (gray, circles) and $10T_{\oplus}$ (black, asterisks). Interbinning has been applied, see text for more details.

C. Computational cost

The semicoherent five-vector method is not suited—at least in its current implementation—for carrying out the follow-up of a large ($\gg 10^4$) number of candidates, as those produced in a typical all-sky search. Rather, it represents an effective method to analyze deeper a relatively small number [$O(10^3$–$10^4)$] of significant candidates. To give an idea of the required computational cost, to analyze one year of data from n_{det} detectors, setting $T_{\mathrm{FFT}} = 3T_{\oplus}$, a frequency band of n_{f} refined bins, a number n_{sky} of sky points, and n_{sd} refined spin-down bins, the code takes less than $2 \times 10^{-7} n_{\mathrm{f}} \cdot n_{\mathrm{sky}} \cdot n_{\mathrm{sd}} \cdot n_{\mathrm{det}}$ core hours, corresponding to about 7×10^{-4} s/template. This would correspond to about 80 core hours for a typical follow-up, covering say $n_{\mathrm{sky}} = 9$, 0.1 Hz frequency band, five coarse spin-down points, and a network of three detectors. It would be then able to follow-up O(5000) candidates previously selected, in about 1% of the time needed to perform the bulk of an all-sky search.

Another reasonable use of the procedure, both in terms of sensitivity and computational cost, concerns directed searches toward, e.g., the Galactic Center or globular clusters. Assuming, for instance, to run a directed search over one year of data of a single detector, looking for a single sky point, a frequency band of 2 kHz, and exploring ten coarse spin-down values, would take about 1.2×10^5 core hours.

The algorithm is characterized by a high level of parallelism, which can be exploited on suitable hardware devices to speed it up. Porting the code on graphics processing units will be the subject of a future work.

VIII. TESTS WITH HARDWARE INJECTIONS

Hardware injections (HIs) are simulated CW signals injected during scientific runs by directly moving detector mirrors. Checking the ability of an analysis pipeline to correctly recover HIs is a standard validation test for CW pipelines. In this section, we present results for two

TABLE II. Sample results from the sensitivity estimation, using the software-simulated signal s_1 (see Table I). The various columns represent the injection band (at random frequency), the data segment duration used in the analysis, the mean of the CR for the detected signals, the sensitivity estimation for H_0 obtained through injections, the corresponding classical strain sensitivity, the corresponding average detector noise amplitude power spectrum (which, for a given frequency band, is independent of the used segment duration), and finally the theoretical estimation. All the sensitivity estimations are in units of 10^{-26}.

Band (Hz)	$T_{\mathrm{FFT}}\ (T_{\oplus})$	$\mathrm{CR}_{95\%}$	$H_{\mathrm{min},95\%}$	$h_{\mathrm{min},95\%}$	$\sqrt{\langle S_{\mathrm{n}}(f)\rangle}(\mathrm{Hz}^{-1/2})$	$h^{\mathrm{th}}_{\mathrm{min},95\%}$
[107, 108]	5	6.1	3.13	4.24	4.97×10^{-24}	3.86
[107, 108]	10	7.0	2.93	3.97		3.57
[585, 586]	5	6.3	3.85	5.21	5.87×10^{-24}	4.37
[585, 586]	10	6.4	3.38	4.58		3.99
[883, 884]	5	7.9	4.94	6.70	7.53×10^{-24}	6.54
[883, 884]	10	7.2	4.62	6.26		5.44

TABLE III. Same as in Table II, but using simulated signal s_2 in Table I. All the sensitivity estimations are in units of 10^{-26}.

Band (Hz)	$T_{\mathrm{FFT}}\ [T_{\oplus}]$	$\mathrm{CR}_{95\%}$	$H_{\mathrm{min},95\%}$	$h_{\mathrm{min},95\%}$	$\sqrt{\langle S_{\mathrm{n}}(f)\rangle}(\mathrm{Hz}^{-1/2})$	$h^{\mathrm{th}}_{\mathrm{min},95\%}$
[107, 108]	5	4.27	3.37	4.37	4.97×10^{-24}	3.25
[107, 108]	10	6.19	3.00	3.88		3.23

TABLE IV. Parameters of HIs used to test the search pipeline with LIGO Livingston O3 data. Frequency and spin-down refer to the GW signal; position is in equatorial coordinates; $\cos(\iota)$ is the cosine of the angle among the source rotation axis and the line of sight; ψ is the wave polarization angle, and H_0 is the signal strain amplitude.

HI	Frequency (Hz)	Spin-down ($\dot{f}$) (Hz/s)	(α, δ) (°)	$\cos\iota$	ψ (°)	H_0
P3	108.857	-1.46×10^{-17}	(178.372, −33.437)	−0.081	25.455	6.615×10^{-26}
P5	52.808	-4.03×10^{-18}	(302.627, −83.839)	0.463	−20.853	3.043×10^{-25}
P11	31.425	-5.07×10^{-13}	(285.097, −58.272)	−0.329	23.589	2.045×10^{-25}

different kinds of tests. The first one consists of running the analysis for specific HIs, assuming the exact sky position and spin-down values and covering a 1-Hz range around the signal frequency, using different data segment durations. Table V shows the results of the analysis for three HIs present in LIGO Livingston O3 data, namely, P3, P5, and P11 for three different choices of T_{FFT}, as indicated in the second column. The third column gives the coarse-frequency bin Δ_{f} used in the analysis, whose value depends on the position of the source as well as the coherence time. The fourth column is the frequency error in detecting the signal, while the last column gives the CR. HI parameters are shown in Table IV. In all cases the signal is well recovered: the error in frequency recovery is always smaller than one bin and the CR increases with T_{FFT}, as expected for sufficiently strong signals.

The second test consists of running a multistage analysis, in which a small parameter space volume around HIs is initially considered and explored with a given segment duration, the most significant candidate selected, and then followed up with longer segment duration. This test closely resembles what would be done in the follow-up of an outlier coming from a wide-parameter search.

In principle, the procedure has not any intrinsic limit on the increasing data segment duration T_{FFT}, which would result in improving the sensitivity, in subsequent steps, to confirm a potential CW candidate. The maximum achievable T_{FFT} is mainly constrained by the available computing power and is related to the size of the search parameter space and to the number of candidates that can be reasonably handled. One further limitation may be present for nearby sources with a high transverse velocity [46] with respect to the line of sight, for which using values of T_{FFT} too large would introduce a sensitivity loss due to a Doppler residual term associated with the variation of source position during the observation time. In the following, we do not take into account this possibility and, for each HI, a double-step analysis has been done. First, a small portion of the parameter space, specified below, around each of sources has been analyzed with a coherence time $T_{\mathrm{FFT}} = 3T_{\oplus}$. A follow-up, with $T_{\mathrm{FFT}} = 10T_{\oplus}$, has been then performed over a smaller region around the most significant candidate found in the previous step. As representative of the results, we discuss here the case of HI P5. The first analysis step focused on the spin-down range $[-2.2449\times10^{-11}, 2.2449\times10^{-11}]$ Hz/s, a sky region covering 0.02° in β and ±1.125° in λ centered at the signal position, corresponding to 15 points in the sky. The point in this three-dimensional grid having the highest CR ($\mathrm{CR}\simeq 93$) has been selected as the signal candidate. It corresponds to the exact source position and to frequency and spin-down values less than one bin off the signal values. We have applied the second analysis stage in a small region around the candidate, increasing the coherence time to $10T_{\oplus}$ and exploring a sky patch of ±0.4° in λ and ±0.015° in β, centered at the candidate position, a frequency range of ±0.03 Hz around the candidate values, and a spin-down range $[-6.7347\times10^{-13}, 6.7347\times10^{-13}]$ Hz/s, corresponding to four coarse spin-down values (and 152 total refined spin-down values) around it. Figure 10 shows the maximum CR for each point of the sky grid (whose position is measured with respect to that of the initial candidate). Figure 11 shows the distribution of the CR as a function of the distance in frequency and spin-down from the starting candidate. All parameters of the most significant candidate, which has $\mathrm{CR} = 243$, are recovered within one bin of the refined grid obtained with $T_{\mathrm{FFT}} = 10T_{\oplus}$, with an error reduced by a factor of ∼3.3 compared to the initial analysis. Furthermore, the increase in CR is compatible with the improvement in sensitivity. Several other high CR outliers appear in Fig. 11, with slightly wrong parameters. This is due to the fact that P5 is a rather strong signal and parameters have some degree of correlation.

TABLE V. Test with HIs. The various columns represent the data segment duration T_{FFT} (in sidereal days), the coarse-frequency step, the error in signal frequency recovery (in bins), and finally, the recovered CR.

HI	T_{FFT} ($T_{\oplus}$)	Δf (Hz)	δf_{err} (bins)	CR
P3	3	0.330	−0.34	17.7
P3	12	0.025	0.11	31.0
P3	24	0.007	0.21	46.7
P5	3	0.500	0.12	94.3
P5	12	0.047	−0.02	291.2
P5	24	0.012	−0.04	465.8
P11	3	0.380	0.05	10.1
P11	12	0.027	0.28	18.5
P11	24	0.007	0.05	36.0

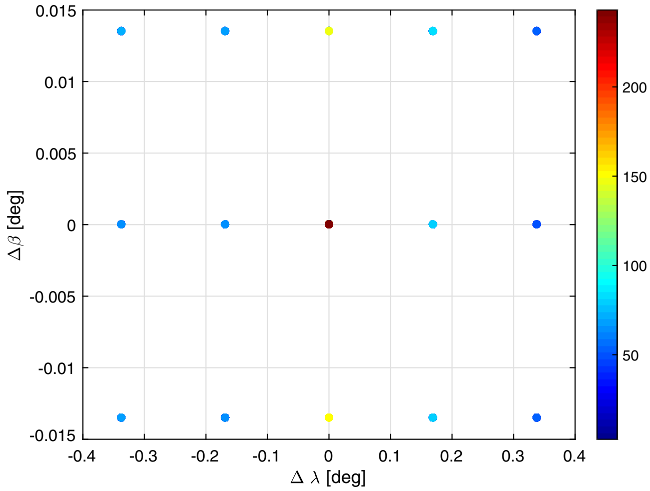

FIG. 10. Analysis of the HI P5, with $T_{\mathrm{FFT}} = 10T_{\oplus}$ (second step). We show the maximum CR over the refined sky grid around the position of the first step candidate. The maximum takes place exactly at the initial candidate position, corresponding to $\Delta\lambda = 0, \Delta\beta = 0$.

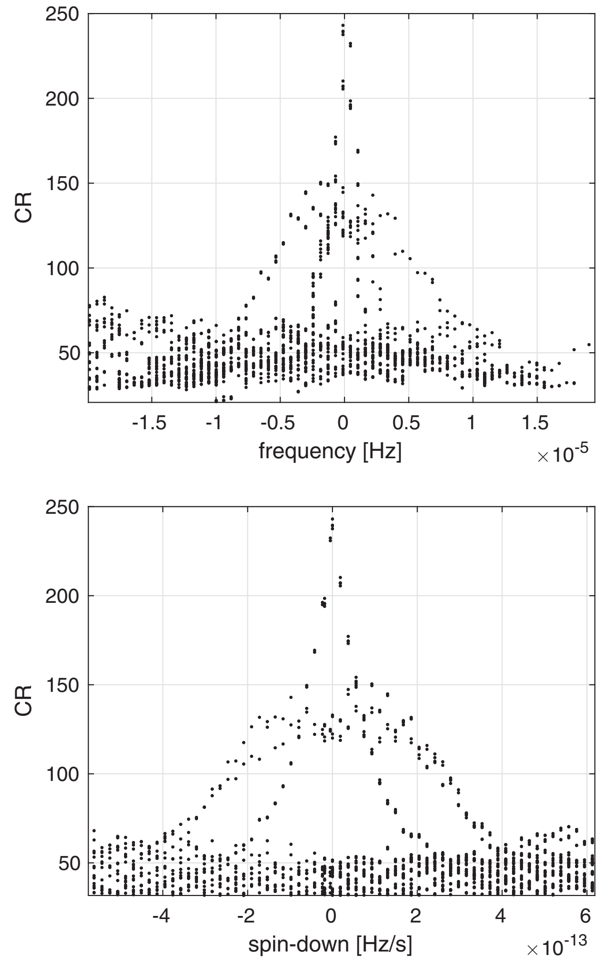

FIG. 11. Analysis of HI P5, second step with $T_{\mathrm{FFT}} = 10T_{\oplus}$: distribution of the CR as a function of the distance in frequency and spin-down from the starting candidate. The maximum takes place at a frequency and spin-down within one bin from the starting candidate values.

IX. CONCLUSIONS

In this paper, we have presented a semicoherent analysis method for the search of CW signals. The method is based on a computationally efficient incoherent combination of five-vector statistics computed over data segments of duration larger than one sidereal day. The concept of five-vectors has been originally introduced in the context of full-coherent targeted searches and exploits the signal sidereal modulation. Here we use it as a coherent step in a semicoherent method in which an initial coarse heterodyne Doppler and spin-down correction is followed by a more refined correction based on the shift of frequencies in the time-frequency plane. On one hand, we demonstrate the heterodyne correction is robust: in order to confine the signal power, in each time window T_{FFT}, within the "natural" bin width, given by the inverse of the data segment duration $\delta f = 1/T_{\mathrm{FFT}}$, it is enough to apply such correction on a coarse grid with step $\Delta_{\mathrm{f}} \gg \delta f$. On the other, we show that such coarse correction leaves a residual frequency variation that can be efficiently removed by shifting the bins of a time-frequency map built computing the five-vector statistics over the single data segments.

The method can be thought as a building block of a multistep procedure in which longer and longer data segments are used, inspecting around a given interesting point (or region) in the search parameter space. Two natural applications are (i) the follow-up of significant candidates found, e.g., in all-sky searches and (ii) directed searches toward specific sky locations, like the Galactic Center or globular clusters, over a large range of frequency and spin-down values.

We have proved, by analyzing both software- and hardware-simulated signals injected in O3 data, that the procedure behaves as expected both in terms of improvement of the candidate significance, when the data segment duration is increased, and in terms of overall sensitivity as compared to a theoretical computation.

The application of the method to wider parameter space, like all-sky searches, is a future milestone for which additional work is needed.

ACKNOWLEDGMENTS

This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. In addition we acknowledge 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 di Fisica Nucleare (INFN), and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, and Spain. S. D. acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (Grant Agreement No. 75449).

APPENDIX A: THEORETICAL SENSITIVITY

In this appendix, we provide some details on the computation of the theoretical sensitivity we refer to in Sec. VII B. An analytical expression for the sensitivity is difficult to derive for the sum of five-vector statistics, while it has been obtained in [37] for the single five-vector statistics, which we briefly summarize. Assuming Gaussian noise with zero mean and variance σ^2 and given the linearity of the Fourier transform, each component of the five-vector, defined by Eq. (11), is also distributed according to a Gaussian with zero mean and variance $\sigma_X^2 = \sigma^2 \cdot T_{\rm FFT}$. The two complex amplitude estimators of Eq. (12), then, have still a Gaussian distribution with zero mean and variance $\sigma^2_{+/\times} = \frac{\sigma_X^2}{|\mathbf{A}_{+/\times}|^2}$. As a consequence, the probability density function of the square modulus of the two estimators is an exponential and then the detection statistics defined in Eq. (13) is distributed according to a linear combination of two exponentials with mean values $\sigma^2_{+/\times}$, see Eq. (34) in [37].

In the presence of a signal of amplitude H_0, each term of the linear combination follows a noncentral χ^2 distribution, with noncentrality parameter

$$\beta_{+/\times} = 2H_0^2 |e^{j\Phi_0} H_{+/\times} \mathbf{A}^{+/\times}|^2/\sigma^2_{\mathbf{X}}. \tag{A1}$$

This applies to each term $\mathcal{S}_i$ in Eq. (19). The distribution of the final statistics $\mathcal{S}$ can be numerically obtained in a straightforward way by generating the two aforementioned distributions (exponentials or noncentral χ^2, respectively, for noise and noise plus signal) and then taking the sum in Eq. (19).

The theoretical sensitivity is computed in the following way. First, we generate the noise-only distribution of the statistics, taking the data average power spectrum $P(f^*)$ at a given frequency, choose a p-value p, e.g., 0.01, and determine the corresponding value of the statistics $\mathcal{S}^*$. Then, for each value of the signal amplitude H_0 in a given range, a population of random source parameters is generated and the corresponding noise plus signal probability distribution is computed. The area of the above distribution on the right of $\mathcal{S}^*$ is evaluated, and the value H_0^* such that the area equals a given value of the detection probability Γ, e.g., 0.95, is determined. The number H_0^* is multiplied by a factor $\sqrt{\pi/2.4308}$, to take into account the average loss due to the uncertainty of the frequency with respect to the frequency bin center [37], and by a factor 1.3258 to convert H_0 to the standard strain h_0 [see Eq. (10)]. This number is the minimum detectable signal amplitude at CL Γ and p-value p for the given value of the data power spectrum. The sensitivity over the whole frequency band is obtained by multiplying that value by $\sqrt{P(f)/P(f^*)}$, i.e., the square root of the ratio of the frequency-dependent data power spectrum to the reference value. Rather than plotting the full theoretical sensitivity, in Table II we reported the theoretical sensitivity computed over O3 LIGO Livingston data for a few frequency bands and different data segment durations $T_{\rm FFT}$, together with the empirical values obtained through the injection of simulated signals, as described in Sec. VII B. The two estimations are in good agreement, with the theoretical one slightly better–by 15% at the most–as expected given they are computed assuming an ideal Gaussian distribution for the noise.

APPENDIX B: SECOND-ORDER SPIN-DOWN

The spin-down correction described in Sec. VI regards only the first-order spin-down term $\dot{f}$. A second-order term $\ddot{f}$ would not be corrected and could determine a sensitivity loss. The condition for an uncorrected second-order spin-down term to not produce a sensitivity loss is that the frequency variation it causes during the observing period $T_{\rm obs}$ is less than half frequency bin $\delta f = 1/2T_{\rm FFT}$. The frequency variation for spinning neutron stars can be expressed through the second-order term of a Taylor expansion as

$$\Delta f = \ddot{f}\frac{T_{\rm obs}^2}{2}. \tag{B1}$$

Hence, the condition for neglecting the second-order spin-down is

$$\ddot{f} \leq \frac{1}{T_{\rm obs}^2 \cdot T_{\rm FFT}}. \tag{B2}$$

This can be translated in a maximum value of the second-order spin-down term. Let us consider a power law to describe the relation between the signal frequency and its first time derivative,

$$\dot{f} \propto f^n, \tag{B3}$$

where n is the "braking index," for which the value depends on the mechanism driving the rotational evolution of the star. By integrating the equation, we find the well-known relation for the time dependency of the frequency,

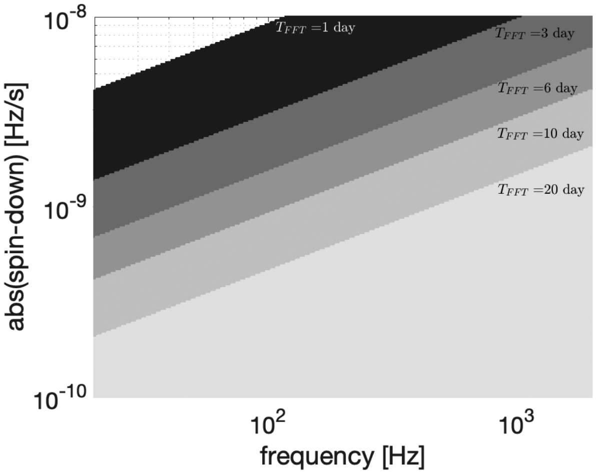

FIG. 12. For various values of $T_{\rm FFT}$, and $T_{\rm obs} = 1$ yr, the plot shows the portion of parameter space for which the second-order spin-down can be neglected.

$$f(t) = f_0\left(1+\frac{t}{\tau}\right)^{\frac{1}{1-n}}, \tag{B4}$$

where $\tau = (1-n)\frac{\dot{f}_0}{f_0}$ is the characteristic spin-down age that, for values of the frequency and spin-down typical for spinning neutron stars, is much bigger than any reasonable observation time of gravitational wave detectors. By deriving Eq. (B4) two times, we obtain the following expression for the second time derivative:

$$\ddot{f} = \frac{f_0}{\tau^2}\frac{n}{(1-n)^2}\left(1+\frac{t}{\tau}\right)^{\frac{2n-1}{1-n}}. \tag{B5}$$

Neglecting the very weak time dependency of $\ddot{f}$ and assuming $n = 5$, which holds for objects in which spin-down is dominated by the emission of gravitational waves, we have the well-known relation

$$\ddot{f}_0 = 5\frac{\dot{f}_0^2}{f_0}. \tag{B6}$$

For each pair $(f_0, \dot{f}_0)$ of the searched parameter space, we can then determine if the corresponding value of $\ddot{f}_0$ satisfies Eq. (B2). Figure 12 shows, for various values of $T_{\rm FFT}$, and $T_{\rm obs} = 1$ yr, the portion of parameter space defined by $f_0 \in [20, 2000]$ Hz and $|\dot{f}_0| \le 10^{-8}$ Hz/s, for which the second-order spin-down can be neglected. The upper left white corner of the plot, corresponding to small signal frequency and very high spin-down, is the region for which the second-order spin-down is never negligible as soon as $T_{\rm FFT} \ge 1$ day. On the other hand, as an example, a source emitting a signal at $f_0 = 100$ Hz and analyzed dividing the data in segments of duration $T_{\rm FFT} = 6$ days could be searched, neglecting the second-order spin-down only if its first-order spin-down was $|\dot{f}_0| < 1.55 \times 10^{-9}$ Hz/s. For each value of $T_{\rm FFT}$, the parameter space cut (the inclined straight line separating regions of different color) corresponds to a specific value of the second-order spin-down, according to Table VI, which is the maximum allowed value not requiring an explicit correction.

TABLE VI. Maximum allowed value of the second-order spin-down that does not require an explicit correction, as a function of $T_{\rm FFT}$ and assuming $T_{\rm obs} = 1$ yr. Such value corresponds to the separation line among different colored regions in Fig. 12.

$T_{\rm FFT}$ (days)	$\ddot{f}_0$ (Hz/s^2)
1	4.25×10^{-18}
3	4.72×10^{-19}
5	1.18×10^{-19}
10	4.25×10^{-20}
20	1.06×10^{-20}

Two comments are in order. First, assuming a different spin-down mechanism, which is a different value for the braking index in Eq. (B3), would affect the position of the cuts in Fig. 12 and the corresponding maximum allowed second-order spin-down values of Table VI. In general, when the spin-down of a spinning neutron star has non-GW contributions, the resulting braking index is smaller than 5 (e.g., it is equal to 3 for pure dipole EM emission). This results in a smaller $\ddot{f}$ for given values of $(f, \dot{f})$. As a consequence, the allowed regions shown in Fig. 12 are conservative.

Second, known pulsars typically have second-order spin-down values smaller than the values shown in the table. Specifically, there is only one known pulsar, J0534 + 2200, with $\ddot{f} \simeq 1.11 \times 10^{-20}$ (Hz/s^2), for which the correction for the second-order spin-down would be needed if $T_{\rm FFT} \ge 20$ days. While properties of the unknown neutron star population could not be directly related to that of known pulsars, we may expect that much larger second-order spin-downs should not be extremely common.

Nevertheless, the extension of the analysis method to include the correction of the second-order spin-down will be an important step that we defer to future work.

[1] B. P. Abbott *et al.*, Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. **116**, 061102 (2016).

[2] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, Classical Quantum Gravity **32**, 074001 (2015).

[3] F. Acernese *et al.*, Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2015).

[4] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, Phys. Rev. X **13**, 041039 (2023).

[5] Paul D. Lasky, Gravitational waves from neutron stars: A review, Pub. Astron. Soc. Aust. **32**, e034 (2015).

[6] Kostas Glampedakis and Leonardo Gualtieri, Gravitational waves from single neutron stars: An advanced detector era survey, Astrophysics and Space Science Library **457**, 673 (2018).

[7] Karl Wette, Liam Dunn, Patrick Clearwater, and Andrew Melatos, Deep exploration for continuous gravitational waves at 171–172 Hz in LIGO second observing run data, Phys. Rev. D **103**, 083020 (2021).

[8] Asimina Arvanitaki and Sergei Dubovsky, Exploring the string axiverse with precision black hole physics, Phys. Rev. D **83**, 044026 (2011).

[9] Chen Yuan, Richard Brito, and Vitor Cardoso, Probing ultralight dark matter with future ground-based gravitational-wave detectors, Phys. Rev. D **104**, 044011 (2021).

[10] Martti Raidal, Christian Spethmann, Ville Vaskonen, and Hardi Veermäe, Formation and evolution of primordial black hole binaries in the early universe, J. Cosmol. Astropart. Phys. 02 (2019) 018.

[11] Sebastien Clesse and Juan García-Bellido, Seven hints for primordial black hole dark matter, Phys. Dark Universe **22**, 137 (2018).

[12] Piotr Jaranowski, Andrzej Królak, and Bernard F. Schutz, Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection, Phys. Rev. D **58**, 063001 (1998).

[13] P. Astone, S. D'Antonio, S. Frasca, and C. Palomba, A method for detection of known sources of continuous gravitational wave signals in non-stationary data, Classical Quantum Gravity **27**, 194016 (2010).

[14] Patrick R. Brady, Teviet Creighton, Curt Cutler, and Bernard F. Schutz, Searching for periodic sources with LIGO, Phys. Rev. D **57**, 2101 (1998).

[15] Patrick R. Brady and Teviet Creighton, Searching for periodic sources with LIGO. II. Hierarchical searches, Phys. Rev. D **61**, 082001 (2000).

[16] Badri Krishnan, Alicia M. Sintes, Maria Alessandra Papa, Bernard F. Schutz, Sergio Frasca, and Cristiano Palomba, The Hough transform search for continuous gravitational waves, Phys. Rev. D **70**, 082001 (2004).

[17] 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, Phys. Rev. D **93**, 123009 (2016).

[18] Paola Leaci, Pia Astone, Sabrina D'Antonio, Sergio Frasca, Cristiano Palomba, Ornella Piccinni, and Simone Mastrogiovanni, Novel directed search strategy to detect continuous gravitational waves from neutron stars in low- and high-eccentricity binary systems, Phys. Rev. D **95**, 122001 (2017).

[19] V. Dergachev, On blind searches for noise dominated signals: A loosely coherent approach, Classical Quantum Gravity **27**, 205017 (2010).

[20] Pia Astone, Alberto Colla, Sabrina D'Antonio, Sergio Frasca, and Cristiano Palomba, Method for all-sky searches of continuous gravitational wave signals using the frequency-Hough transform, Phys. Rev. D **90**, 042002 (2014).

[21] J. Aasi *et al.*, Implementation of an $\{f\}$-statistic all-sky search for continuous gravitational waves in Virgo VSR1 data, Classical Quantum Gravity **31**, 165014 (2014).

[22] S. D'Antonio, C. Palomba, S. Frasca, P. Astone, I. La Rosa, P. Leaci, S. Mastrogiovanni, O. J. Piccinni, L. Pierini, and L. Rei, Sidereal filtering: A novel robust method to search for continuous gravitational waves, Phys. Rev. D **103**, 063030 (2021).

[23] B. Steltner, M. A. Papa, H. B. Eggenstein, R. Prix, M. Bensch, B. Allen, and B. Machenschalk, Deep Einstein@Home all-sky search for continuous gravitational waves in LIGO O3 public data, Astrophys. J. **952**, 55 (2023).

[24] Curt Cutler, Iraj Gholami, and Badri Krishnan, Improved stack-slide searches for gravitational-wave pulsars, Phys. Rev. D **72**, 042004 (2005).

[25] Avneet Singh, Maria Alessandra Papa, Heinz-Bernd Eggenstein, Sylvia Zhu, Holger Pletsch, Bruce Allen, Oliver Bock, Bernd Maschenchalk, Reinhard Prix, and Xavier Siemens, Results of an all-sky high-frequency Einstein@home search for continuous gravitational waves in LIGO's fifth science run, Phys. Rev. D **94**, 064061 (2016).

[26] Karl Wette, Parameter-space metric for all-sky semicoherent searches for gravitational-wave pulsars, Phys. Rev. D **92**, 082003 (2015).

[27] Karl Wette, Empirically extending the range of validity of parameter-space metrics for all-sky searches for gravitational-wave pulsars, Phys. Rev. D **94**, 122002 (2016).

[28] B. P. Abbott *et al.*, Narrow-band search for gravitational waves from known pulsars using the second LIGO observing run, Phys. Rev. D **99**, 122002 (2019).

[29] Vladimir Dergachev and Maria Alessandra Papa, Results from high-frequency all-sky search for continuous gravitational waves from small-ellipticity sources, Phys. Rev. D **103**, 063019 (2021).

[30] R. Abbott *et al.*, All-sky search in early O3 LIGO data for continuous gravitational-wave signals from unknown neutron stars in binary systems, Phys. Rev. D **103**, 064017 (2021).

[31] R. Abbott *et al.*, All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO and Advanced Virgo O3 data, Phys. Rev. D **106**, 102008 (2022).

[32] R. Abbott *et al.*, Narrowband searches for continuous and long-duration transient gravitational waves from known pulsars in the LIGO-Virgo third observing run, Astrophys. J. **932**, 133 (2022).

[33] Rodrigo Tenorio, David Keitel, and Alicia M. Sintes, Search methods for continuous gravitational-wave signals from unknown sources in the advanced-detector era, Universe **7**, 474 (2021).

[34] Ornella Juliana Piccinni, Status and perspectives of continuous gravitational wave searches, Galaxies **10**, 72 (2022).

[35] Keith Riles, Searches for continuous-wave gravitational radiation, Living Rev. Relativity **26**, 3 (2023).

[36] O. J. Piccinni, P. Astone, S. D'Antonio, S. Frasca, G. Intini, P. Leaci, S. Mastrogiovanni, A. Miller, C. Palomba, and A. Singhal, A new data analysis framework for the search of continuous gravitational wave signals, Classical Quantum Gravity **36**, 015008 (2018).

[37] P. Astone, A. Colla, S. D'Antonio, S. Frasca, C. Palomba, and R. Serafinelli, Method for narrow-band search of continuous gravitational wave signals, Phys. Rev. D **89**, 062008 (2014).

[38] Rodrigo Tenorio, David Keitel, and Alicia M. Sintes, Application of a hierarchical MCMC follow-up to Advanced LIGO continuous gravitational-wave candidates, Phys. Rev. D **104**, 084012 (2021).

[39] R. Abbott *et al.*, Searches for continuous gravitational waves from young supernova remnants in the early third observing run of Advanced LIGO and Virgo, Astrophys. J. **921**, 80 (2021).

[40] R. Abbott *et al.*, Search for continuous gravitational wave emission from the Milky Way center in O3 LIGO-Virgo data, Phys. Rev. D **106**, 042003 (2022).

[41] S. Mastrogiovanni, P. Astone, S. D'Antonio, S. Frasca, G. Intini, P. Leaci, A. Miller, C. Palomba, O. J. Piccinni, and A. Singhal, An improved algorithm for narrow-band searches of continuous gravitational waves, Classical Quantum Gravity **34**, 135007 (2017).

[42] Scott M. Ransom, Stephen S. Eikenberry, and John Middleditch, Fourier techniques for very long astrophysical time-series analysis, Astron. J. **124**, 1788 (2002).

[43] B. P. Abbott *et al.*, All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO O2 data, Phys. Rev. D **100**, 024004 (2019).

[44] Berit Behnke, Maria Alessandra Papa, and Reinhard Prix, Postprocessing methods used in the search for continuous gravitational-wave signals from the Galactic Center, Phys. Rev. D **91**, 064007 (2015).

[45] Christoph Dreissigacker, Reinhard Prix, and Karl Wette, Fast and accurate sensitivity estimation for continuous-gravitational-wave searches, Phys. Rev. D **98**, 084058 (2018).

[46] P. B. Covas, Effects of proper motion of neutron stars on continuous gravitational-wave searches, Mon. Not. R. Astron. Soc. **500**, 5167 (2020).