PHYSICAL REVIEW D **87,** 064036 (2013)

Resolving multiple supermassive black hole binaries with pulsar timing arrays. II. Genetic algorithm implementation

Antoine Petiteau,[1,2,*] Stanislav Babak,[2,†] Alberto Sesana,[2,‡] and Mariana de Araújo[2,§]

[1]*APC, FACe, Université Paris Diderot, CNRS/IN2P3, CEA/Irfu, Observatoire de Paris, Sorbonne Paris Cité, 10 rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France*

[2]*Albert Einstein Institute, Am Muhlenberg 1 D-14476 Golm, Germany*

(Received 6 October 2012; published 22 March 2013)

Pulsar timing arrays might detect gravitational waves from massive black hole binaries within this decade. The signal is expected to be an incoherent superposition of several nearly monochromatic waves of different strengths. The brightest sources might be individually resolved, and the overall deconvolved, at least partially, in their individual components. In this paper we extend the maximum-likelihood-based method developed in Babak and Sesana [Phys. Rev. D **85**, 044034 (2012)] to search for individual massive black hole binaries in pulsar timing array data. We model the signal as a collection of circular monochromatic binaries, each characterized by three free parameters: two angles defining the sky location and the frequency. We marginalize over all other source parameters, and we apply an efficient multisearch genetic algorithm to maximize the likelihood function and look for sources in synthetic data sets. On data sets characterized by white Gaussian noise plus few injected sources with signal-to-noise ratio in the range 10–60, our search algorithm performs well, recovering all the injections with no false positives. Individual source signal-to-noise ratios are estimated within a few percent of the injected values, sky locations are recovered within a few degrees, and frequencies are determined with sub-Fourier-bin precision.

DOI: 10.1103/PhysRevD.87.064036 PACS numbers: 04.30.−w, 04.80.Nn, 97.60.Gb, 97.60.Lf

I. INTRODUCTION

Precision timing of millisecond pulsars provides a unique opportunity to get the very first low-frequency gravitational wave (GW) detection. This prospect is attracting the attention of the wider astrophysical community, causing a recent boost of activity in the field. The European Pulsar Timing Array (EPTA) [1], the Parkes Pulsar Timing Array [2] and the North American Nanohertz Observatory for Gravitational Waves [3], joining together in the International Pulsar Timing Array [4], are collecting data and improving their sensitivity in the frequency range of $\sim 10^{-9}$–10^{-6} Hz. In the coming years, the Chinese five-hundred-meter aperture spherical telescope [5] and the planned Square Kilometer Array [6] will provide a major leap in sensitivity. Current surveys are already placing interesting upper limits on the level of a putative GW background [7,8], skimming the range predicted by state-of-the-art models of massive black hole (MBH) evolution [9,10]. Within the next few years, the combined International Pulsar Timing Array data might either result in a first detection or start placing interesting limits on the MBH binary formation efficiency in massive galaxies.

The detection principle is very simple: GWs affect the propagation of radio signals from the pulsar to the receiver on Earth, leaving a characteristic fingerprint in the time of arrival of the radio pulses [e.g., Refs. [11,12]]. Such fingerprint depends on the properties of the underlying cosmological population of inspiralling binaries and will consist of a superposition of quasimonochromatic waves, similar to the white dwarf-white dwarf foreground [e.g., Ref. [13]] in the mHz window relevant to space-based interferometry [14]. This signal has generally been regarded as a stochastic background, and data analysis techniques have been developed accordingly [7,8,15–18]. The actual expected signal, however, is far from being isotropically distributed in the sky, with just a few sources dominating the power at each frequency [9,10,19]. The possibility of resolving an individual source offers appealing astrophysical prospects, and pulsar timing array (PTA) capabilities on this front were also investigated in detail by many authors [20–25].

What is still missing is a detailed study of what kind of information a PTA can extract out of a complex superposition of multiple sources. Is the signal going to be similar to confusion noise? Can we resolve individual sources? How many of them? Can we locate them in the sky, and to what level of accuracy? All these questions are of great interest for the astrophysical community; precise sky localization of individual sources will allow the efficient search for electromagnetic counterparts [26,27], opening the new horizon of multimessenger astronomy.

This is the second in a series of papers devoted to the exploration of the PTA potential of resolving multiple GW sources. In Ref. [28] (hereinafter Paper I), we demonstrated PTA efficiency in disentangling monochromatic sources at the same frequency. The key idea is to estimate

*petiteau@apc.univ-paris7.fr
†Stanislav.Babak@aei.mpg.de
‡Alberto.Sesana@aei.mpg.de
§marianabdaraujo@gmail.com

 © 2013 American Physical Society

the likelihood that a certain number of sources with certain parameters are present in the data. We developed a formalism that allowed us to maximize analytically the likelihood over the extrinsic source parameters, restricting the search to the source sky location only ($2 \times N$ parameters, where N is the number of GW sources in the template). There, we did not implement any proper algorithm to search over the parameter space, and we made a lot of simplifying assumptions, suitable to a first, exploratory investigation.

Our aim is to implement a proper search algorithm, progressively relaxing our limiting assumptions to develop a detection pipeline able to handle the whole complexity of a realistic data set. We start in this paper with the following two major steps: (i) we extend our mathematical formalism to include a frequency scan and (ii) we present an upgraded version of the genetic algorithm employed by Petiteau *et al.* [29] in the LISA mock data challenge [30,31] specifically developed to search for a global maximum on the multimodal likelihood surface embedded in the multidimensional parameter space. We have found (similarly to the mock LISA data challenge) that the genetic algorithm (GA) is very efficient in finding the correct number of sources and their parameters.

The paper is organized as follows. In Sec. II we spell out our main assumptions, and in Sec. III we present the genetic algorithm and its feature. The data sets used to test the algorithm are detailed in Sec. IV, and the algorithm performances and results are presented in Sec. V. In Sec. VI we draw our conclusion and discuss improvements we will present in future work.

This research is the result of the common effort to directly detect gravitational waves using pulsar timing, known as the European Pulsar Timing Array (EPTA) [1,32].

II. DETECTION STRATEGY, EXTENSION TO FREQUENCY SEARCH

The main purpose of this paper is to extend our formalism to include search in frequency and to implement a proper search algorithm to identify maxima in the likelihood. Accordingly, we relax numbers 1, 3, and 8 from the limitations and assumptions described in Sec. II of Paper I, i.e.,

(1) We consider only data sets with noise.
(2) We inject sources at different frequencies.
(3) We implement a proper search algorithm to maximize the likelihood.

A. Choice of the template

In computing the likelihood function, we consider monochromatic GW sources assuming that the orbital frequency does not change appreciably over the observation period [see, e.g., Ref. [21]] which we took to be 10 years. Each GW signal is therefore characterized by seven parameters only: the overall signal amplitude $\{\mathcal{A}\}$, the source frequency and phase $\{f, \Phi_0\}$, and the angles defining its location in the sky $\{\phi, \theta\}$, inclination $\{\iota\}$, and polarization $\{\psi\}$. Contrary to Paper I, we do not fix all the systems at the same frequency but we consider the unknown frequencies of the sources as additional search parameters.

We do not use the full response of the pulsar-earth detector, which is composed of the "earth term" and the "pulsar term" [see, e.g., Ref. [21]] (from now on we drop the quotes and use those notions as jargon), but we construct a signal template to match the earth term only. There are several reasons to drop the pulsar term in the analysis. In all pulsars, the earth terms add up coherently: they all have the same frequency and phase, and the amplitude of the signal in the residuals depends on the relative position of the pulsar and the GW source on the sky. Conversely, the pulsar terms are, in general, incoherent: they usually appear at different frequencies, and the phase and amplitude of the signal depends not only on the position of the source relative to the pulsar but also on the distance to the pulsar, which is usually poorly known (in most of the cases to $\sim$10% precision). Even if we assume we know the pulsar distance exactly, the pulsar term carries the imprint of the binary system as it was emitting at a time $\Delta t = L(1 + \hat{k}.\hat{n})$ in the past as compared to the earth term, where L is the earth-pulsar distance and $\hat{k}$ and $\hat{n}$ are the unit vectors pointing to the sky location of the source and the pulsar, respectively. This means that to connect pulsar and earth terms, we need to know the evolution of the binary system for Δt, which is typically 10^3–10^4 years. Adding the pulsar term in our template could enhance the overall detection signal-to-noise ratio (SNR) and potentially improve the estimation of the distance L. However, in order to extract useful information, we need to model the past evolution of the binary over a rather long time. This requires specific assumptions about the binary system: whether it is precessing due to misaligned spins [33] or eccentricity and whether it is driven completely by gravitational radiation reaction or there is a measurable influence on the dynamic from the gasous [19,26] and stellar environment [34]. Any answer obtained using the pulsar term is therefore strongly conditional: it depends on the assumed model for the binary evolution, and it may not be unique. In addition it will require searching for more parameters, which will inevitably increase false alarm probability and may reduce the efficiency of the search. In this work we are trying to avoid these complexities. However, as mentioned above, if the detection of the pulsar term is possible (and correctly associated to a given GW source), then it will bring valuable information about the binary. Adding of the pulsar term in the search template is also necessary if the frequencies of the pulsar and earth terms are the same and interfere with each other. This happens for very low-frequency binaries, which we do not consider in this paper.

B. Likelihood function and detection statistics

The details of the detection statistics were outlined in Paper I. We briefly summarize here the main points and describe the extension of the formalism to sources with different frequencies. As justified in the previous section, we use the matched filtering technique, assuming the earth term as a template. The mathematical description of the signal template for an individual source as a function of the parameters $\vec{\lambda} = \{\mathcal{A}, f, \Phi_0, \phi, \theta, \iota, \psi\}$ is given by equations (11)–(16) of Paper I. We denote the measured timing residuals as x_α, which includes the measurement noise and the influence of the GW signal r_α, where the subscript α identifies the αth pulsar in the array. The log-likelihood ratio (likelihood that a data set $x_\alpha(t)$ contains a GW signal $r_\alpha(t;\vec{\lambda})$ over the likelihood that it is pure noise) is

$$\log\Lambda = \langle x_\alpha | r_\alpha\rangle - \frac{1}{2}\langle r_\alpha | r_\alpha\rangle, \tag{1}$$

where $r_\alpha = r_\alpha^E$ is the expected earth term in the data. We neglect here all possible stochastic GW signals and look for individual binaries standing above a putative unresolved background only. The inner product appearing in Eq. (1) is defined as

$$\langle x|r\rangle = \frac{2T_o}{NS(f)}\sum_{i=1}^{N} x(t_i)r(t_i), \tag{2}$$

where N is the number of points in the time series, T_o is the observation time, and $S(f)$ is the one-sided noise power spectral density, which we assume to be white Gaussian. Equation (2) is the discrete version of the inner product used in Paper I; it has the advantage of being applicable to unevenly sampled data, which will be the case in reality. It was shown in Paper I that the GW signal imprinted in the data by each individual source can be written as

$$r_\alpha = \sum_{j=1}^{4} a_{(j)} h^\alpha_{(j)}, \tag{3}$$

where $h_{(j)}$ denote the time-dependent parts of the GW signal and includes the parameters for which we want to search, while $a_{(j)}$ are constants over the observation period.

We can then maximize the likelihood ratio over the $a_{(j)}$ constants for each GW source analytically,

$$\frac{\partial\log(\Lambda)}{\partial a_{(j)}} = 0, \qquad \rightarrow a_{(k)} = M^{-1}_{kj} X_j, \tag{4}$$

$$\{\log(\Lambda)\}_{\max\{a_{(j)}\}} \equiv \mathcal{F}_e = \frac{1}{2} X_k M^{-1}_{jk} X_j, \tag{5}$$

where

$$X_j \equiv \sum_{\alpha=1}^{P}\langle x_\alpha | h^\alpha_{(j)}\rangle, \qquad M_{jk} \equiv \sum_{\alpha=1}^{P}\langle h^\alpha_{(j)} | h^\alpha_{(k)}\rangle. \tag{6}$$

The statistical properties of $\mathcal{F}_e$ are investigated in detail in Ref. [25]. In the presence of N_s GW sources in the template, the coefficients $a_{(j)}$ are represented by a $4\times N_s$ array, X_j is also a $4\times N_s$ array, while the M matrix is a $4N_s\times 4N_s$, two-dimensional matrix. The matrix can be decomposed in N_s 4×4 row- and column-matrices, each corresponding to the cross terms between the Ith and Jth GW sources,

$$M^{IJ} = \sum_\alpha \begin{pmatrix} U^{IJ}\mathcal{I}^{IJ}_{ss} & Q^{IJ}\mathcal{I}^{IJ}_{ss} & U^{IJ}\mathcal{I}^{IJ}_{sc} & Q^{IJ}\mathcal{I}^{IJ}_{sc} \\ Q^{JI}\mathcal{I}^{IJ}_{ss} & V^{IJ}\mathcal{I}^{IJ}_{ss} & Q^{JI}\mathcal{I}^{IJ}_{sc} & V^{IJ}\mathcal{I}^{IJ}_{sc} \\ U^{IJ}\mathcal{I}^{IJ}_{cs} & Q^{IJ}\mathcal{I}^{IJ}_{cs} & U^{IJ}\mathcal{I}^{IJ}_{cc} & Q^{IJ}\mathcal{I}^{IJ}_{cc} \\ Q^{JI}\mathcal{I}^{IJ}_{cs} & V^{IJ}\mathcal{I}^{IJ}_{cs} & Q^{JI}\mathcal{I}^{IJ}_{cc} & V^{IJ}\mathcal{I}^{IJ}_{cc} \end{pmatrix}, \tag{7}$$

where

$$U^{IJ} = (F^\alpha_c)^I(F^\alpha_c)^J, \qquad Q^{IJ} = (F^\alpha_c)^I(F^\alpha_s)^J, \qquad V^{IJ} = (F^\alpha_s)^I(F^\alpha_s)^J, \tag{8}$$

and $F^\alpha_{c,s}$ represent the decomposition of the antenna pattern given by Eq. (16) of Paper I. The $\mathcal{I}$ terms come from the inner products of the time-dependent parts of $h^\alpha_{(j),}$ which for each source I are cosine and sine functions of phase $\phi_I = 2\pi f_I t \equiv \omega_I t$. We can evaluate those inner products analytically by using the integral representation adopted in Paper I; for example,

$$\begin{aligned}\langle h^I_{\alpha,(1)} | h^J_{\alpha,(1)}\rangle &= (F^\alpha_c)^I(F^\alpha_c)^J\langle\sin(\phi_I)|\sin(\phi_J)\rangle \\ &\sim (F^\alpha_c)^I(F^\alpha_c)^J\frac{2}{T_o}\int_0^{T_o}\sin(\phi_I)\sin(\phi_J)dt \\ &\equiv (F^\alpha_c)^I(F^\alpha_c)^J\mathcal{I}^{IJ}_{ss} \equiv U^{IJ}\mathcal{I}^{IJ}_{ss}.\end{aligned} \tag{9}$$

The explicit form of the $\mathcal{I}$ integrals for all possible sine and cosine combinations are given by

$$\begin{aligned}\mathcal{I}^{IJ}_{ss} &= \frac{2}{T_o}\int_0^{T_o}\sin(\omega^I t)\sin(\omega^J t)dt \\ &= \mathrm{sinc}(\Delta\phi) - \mathrm{sinc}(\Sigma\phi),\end{aligned} \tag{10}$$

$$\begin{aligned}\mathcal{I}^{IJ}_{cc} &= \frac{2}{T_o}\int_0^{T_o}\cos(\omega^I t)\cos(\omega^J t)dt \\ &= \mathrm{sinc}(\Delta\phi) + \mathrm{sinc}(\Sigma\phi),\end{aligned} \tag{11}$$

$$\begin{aligned}\mathcal{I}^{IJ}_{sc} &= \frac{2}{T_o}\int_0^{T_o}\sin(\omega^I t)\cos(\omega^J t)dt \\ &= \sin\left(\frac{\Sigma\phi}{2}\right)\mathrm{sinc}\left(\frac{\Sigma\phi}{2}\right) + \sin\left(\frac{\Delta\phi}{2}\right)\mathrm{sinc}\left(\frac{\Delta\phi}{2}\right),\end{aligned} \tag{12}$$

$$\begin{aligned}\mathcal{I}^{IJ}_{cs} &= \frac{2}{T_o}\int_0^{T_o}\cos(\omega^I t)\sin(\omega^J t)dt \\ &= \sin\left(\frac{\Sigma\phi}{2}\right)\mathrm{sinc}\left(\frac{\Sigma\phi}{2}\right) - \sin\left(\frac{\Delta\phi}{2}\right)\mathrm{sinc}\left(\frac{\Delta\phi}{2}\right),\end{aligned} \tag{13}$$

where $\Delta\phi \equiv (\omega^I - \omega^J)T_o$ and $\Sigma\phi \equiv (\omega^I + \omega^J)T_o$. Note that the M matrices reduce to the expression given in Eq. (25) of Paper I when $\omega^I = \omega^J$. The $I \neq J$ terms give beatings between two signals at different frequencies, and they are usually smaller than the terms in the $I = J$ matrices. We found that one can consider the sources approximately at the same frequency if $|f_I - f_J| \approx (2/3)\Delta F$, where $\Delta F = 1/T_o$ is the size of the Fourier frequency bin.

We use $\mathcal{F}_e$ as a detection statistic that depends on $3N_s$ search parameters (frequency and sky position of each GW source). Note that the number of GW sources N_s is not known and has to be determined during the search procedure. We can also estimate the relative contribution of each source as $\mathcal{F}_e^J = \frac{1}{2}X_k^J(M_{jk}^{-1})^J X_j^J$. Following [35], the relation between the expectation of the analytically maximized likelihood $\mathcal{F}_e$ and the SNR can be expressed as

$$E(\mathcal{F}_e) = \frac{1}{2}(4N_s + \mathrm{SNR}^2). \tag{14}$$

To search for an individual source, we use the same mathematical framework assuming $N_s = 1$. We refer to Ref. [25] for more details on the statistical properties of $\mathcal{F}_e$ in this latter case.

III. MULTISEARCH GENETIC ALGORITHM: DESCRIPTION AND IMPLEMENTATION

We search for the maximum of $\mathcal{F}_e$ with an improved version of the genetic algorithm (GA) described in Ref. [29], performing multiple searches in parallel. There are other stochastic optimization methods which are currently used in the GW data analysis: Monte Carlo Markov Chain [36], particle swarm optimization [37], etc. All these methods (besides the grid-based method) work well if (i) the initial guess is near global maximum or (ii) if there is only one maximum. In all other cases the optimization methods guarantee finding a maximum (if there is one) but do not guarantee that it is the global one (for reasonable computation time). In our case neither (i) nor (ii) are true. We do not know *a priori* where the global maximum could be, and we have a large number of strong local maxima. What is crucial in this search is not the GA itself, but its specific implementation (multiple searches genetic algorithm as described below), which deals with the multiple maxima of the likelihood surface in a multidimensional parameter space.

A. Genetic algorithm

The GA is a method to perform global searches on large parameter spaces (optimization method) based on the natural selection principle. We apply this method to the search for individual GWs in PTA data using equivalences described in Table I. A template described by a set of parameters $\{\theta_I, \phi_I, f_I\}$ is one organism described by a set of genes; the $\mathcal{F}_e$ of the template is the quality of the organism; parameters are converted to the binary form, and the set of bits represents a gene of the organism.

We start with a group of organisms (templates) chosen randomly (initial search) or constructed from the results of previous searches. We evaluate the quality of each organism ($\mathcal{F}_e$). We select set of pairs (parents) based on qualities: organisms with better quality (templates with higher $\mathcal{F}_e$) are chosen more often than weak organisms. We combine the genotypes of two parents to produce a child (we combine parameters of two chosen templates to produce a new one). We impose the number of produced children to be equal to the number of parents (i.e., we keep the number of evolving organisms constant at each generation). Next, we allow with a certain probability a random mutation in the children's genes (with some probability we randomly change the parameters of the new templates, exploring a larger area of the parameter space). The parents are discarded, and the resulting children form a new generation. We repeat the procedure until we reach a steady state (maximum in the quality) or a maximum number of generations. We keep only one generation active (one group of templates).

The selection and the random mutation processes are controlled by tunable parameters (temperature and probability mutation rate, respectively) that define the breadth of the search. Depending on their value, the colony of organisms can either explore the whole parameter space or sample only the areas around the local maxima, allowing the alternation of global and local exploration phases. The definition and implementation of such parameters and a general discussion of the selection, breeding, and mutation processes can be found in Secs. III and IV of Petiteau *et al.* [29].

TABLE I. Correspondence between GA and GW data analysis notions.

Genetic algorithm		GW search
Organism	$\Longleftrightarrow$	Template: signal from N_s GW sources
Gene (of an organism)	$\Longleftrightarrow$	Parameter (of a template): $3 \times N_s$
Allele (of a gene)	$\Longleftrightarrow$	Bits (of the value of the parameter)
Quality Q	$\Longleftrightarrow$	Maximized likelihood, i.e., F-statistic $\mathcal{F}_e$
Colony of organisms	$\Longleftrightarrow$	Evolving group of templates
nth generation	$\Longleftrightarrow$	The state of colony at nth step of evolution
(selection + breeding) + mutation	$\Longleftrightarrow$	w parameter space exploration strategy

We typically use 50 organisms per generation and 1000 generations. The run of one GA takes a few minutes on a standard laptop (one Intel core at 2 GHz). Since the size of the parameter space increases with the number of sources in the template, the convergence speed decreases accordingly. The algorithm usually converges around the true solution in less than 400 generations for the highest SNR sources. One of the most interesting features of the GA is its efficiency in finding maxima in the $\mathcal{F}_e$ surface first (during the large exploration phase) and then in exploring them deeply to extract the global one (during the local exploration phase). One GA run is usually sufficient to find most of the sources, but sometimes it gets stuck on some local maximum. To overcome this problem, we run several GAs in parallel, as described in the next section.

B. Multiple searches (MultiSearch)

The GA described in the previous section provides the basis for a more general method called the "multiple searches" (MS) algorithm. This method consists in running several GAs in parallel with different properties and initial parameters.

We take an initial population with parameters chosen randomly. We start a GA on this population, tuning the parameters to perform a large exploration. In the resulting population, we select only the best organisms which are well separated. This means that the selected organisms have $\mathrm{SNR} > 97\%\mathrm{SNR}_{\mathrm{Best}}$ and the distance in parameter space between two organisms is higher than a certain threshold chosen empirically after a number of tests: $|\cos(\theta_{I,i}) - \cos(\theta_{I,j})| > \Delta_{c\theta} = 0.1$, $|\phi_{I,i} - \phi_{I,j}| > \Delta_\phi = 20^o$ and $|f_{I,i} - f_{I,j}| > \Delta_f = 0.5$ nHz, where I refers to the source and i and j to the solutions. The selected solutions are called "modes," and this selection process is called "mode separation."

The next step is to start one GA on each mode, tuned for local exploration. The goal is to explore the vicinity of the mode to find the local highest value of $\mathcal{F}_e$. The organisms of each GA are allowed to explore only their mode neighborhood and are forbidden to go on the area of interest of other modes. The area of interest of a mode $\{\cos\theta_i, \phi_i, f_i\}$ is defined within $[\cos\theta_i - \Delta_{c\theta}, \cos\theta_i + \Delta_{c\theta}]$, $[\phi_i - \Delta_\phi, \phi_i + \Delta_\phi]$ and $[f_i - \Delta_f, f_i + \Delta_f]$. In parallel to these "mode GAs," we start another GA tuned for large exploration. We forbid the organisms of this GA to go on the areas of the modes. The aim of this GA is to find new modes (overlooked in previous searches) if there are any left (it can also give a null result).

At the end of this step, all the solutions are grouped together and we apply the "mode separation" to identify "modes." Then we iterate the procedure by restarting several "local" GAs.

In the long run, this method, like other stochastic methods (e.g., Markov Chain Monte Carlo methods), is guaranteed to converge to the global maximum. However, there is no way to exactly know *a priori* how fast it will do so, and one has to decide when to stop it, being somehow confident that the best solution has been found. We usually do two to five iterations of the procedure outlined above before stopping. The number of modes N_{modes} found increases with the number of iterations. Since we are running $N_{\mathrm{modes}} + 1$ GAs at each iteration, the first one takes just a few minutes (one initial exploratory GA run), the second one can take up to an hour (depending on the number of modes found in the first iteration), and the later ones up to a few hours. In total, we run between 50 to 300 GAs for a search. The correct solution is usually found after two iterations (i.e., about one hour). As a pseudotest for convergence, we run several times (typically ten) our MS-GA code, with different initial conditions. If all the runs give almost the same results, we claim convergence.

IV. DESCRIPTION OF THE TEST DATA SETS

The genetic algorithm described in the previous section was used to analyze four blind data sets, which we describe here in detail. Each data set consists of a collection of time series representing the residuals obtained by timing an ensemble of millisecond pulsars (MSPs). To keep our investigation as general as possible, in all data sets MSPs are placed randomly in the celestial sphere, each time series consists of 523 equally sampled data points over a total observing time of ten years (one datapoint every week), and the noise is assumed to be white Gaussian (we will specialize our searches to PTA pulsars in our future work). The injected sources were all equal mass, circular, nonspinning binaries with chirp mass of $10^9 M_\odot$, placed at the same redshift (distance), but with sky location, inclination, polarization, and initial phase drawn randomly, resulting in a range of signal strengths. The number of injected sources and their parameters were chosen to generate suitable signals with the desired SNR and do not reflect the expected distribution of MBH binaries in the Universe. We will test our algorithm on more realistic data sets in future work. Based on the statistics of individually resolvable sources presented in [Ref. [10] see their Fig. 6], the frequency was drawn from a random distribution in the range 10^{-8}–10^{-7} Hz.

Sources were evolved according to an equation of motion accurate to 3.5 post-Newtonian order in phase evolution [38], and gravitational waveforms were generated following [21] (see also Refs. [39,40]). The final residuals injected in the data sets were obtained by time integration of the waveforms (see Eqs. (8) and (9) of Paper I). Note that the injected data are quite different from the adopted circular, nonevolving monochromatic templates we use in the search; we are therefore mimicking the (likely) situation in which the template does not perfectly match the signal. Especially at high frequency, there might be a nonnegligible evolution of the source frequency over ten years, possibly introducing a bias in our source recovery. We will

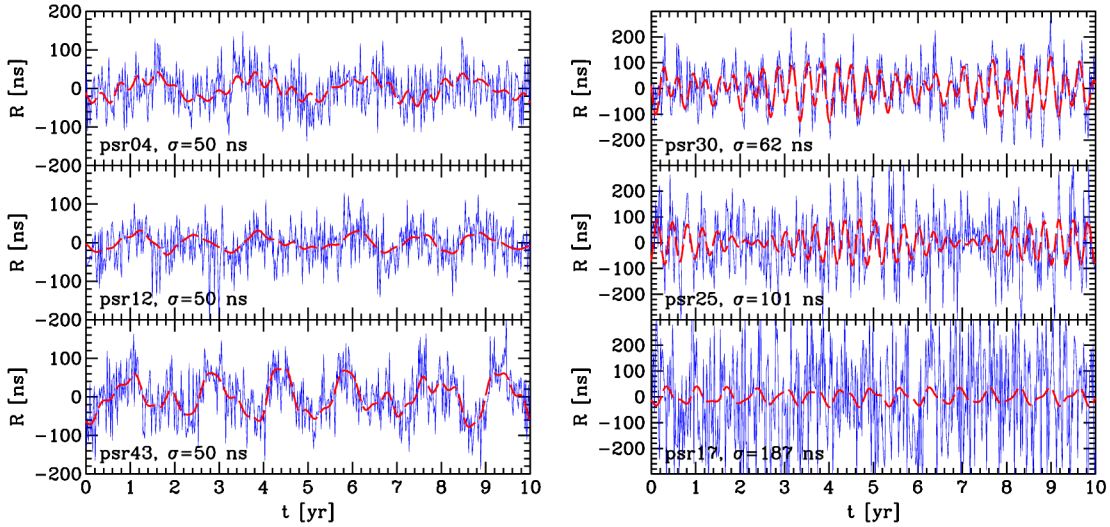

FIG. 1 (color online). Sample of simulated time series. In each panel, the red dashed curve is the injected signal, where the blue jagged line represents the total raw data set, including signal plus white Gaussian noise. Left panel: pulsars extracted from Dataset3; sources are injected in white Gaussian noise with rms $\sigma = 50$ ns. Right panel: pulsars extracted from Dataset4; here each pulsar has a different noise level, as labeled in each panel.

quantify this effect on our results in Sec. V. As in Paper I, we considered the earth term only (issues related to pulsar terms will be explored in the next paper). Here we follow the details of the four data sets:

(i) *Dataset1*: 30 MSP; rms noise 50 ns in each pulsar; 5 binaries at $z = 0.01$, with individual SNR in the range ~30–60;
(ii) *Dataset2*: 30 MSP; rms noise 50 ns in each pulsar; 4 binaries at $z = 0.02$, with individual SNR in the range ~15–55;
(iii) *Dataset3*: 50 MSP; rms noise 50 ns in each pulsar; 8 binaries at $z = 0.03$, with individual SNR in the range ~10–40;
(iv) *Dataset4*: 50 MSP; rms noise of each pulsar randomly drawn in the range 30–200 ns; 3 binaries at $z = 0.01$, with individual SNR in the range ~30–40.

Data sets are in order of increasing complexity (more sources, lower SNR). In the last data set we tested the algorithm performance when combining time series with different noise levels. Sample time series extracted from Dataset3 and Dataset4 are visualized in Fig. 1, where we can appreciate the variety of imprints depending on the pulsar location in the sky relative to each individual source.

V. RESULTS AND DISCUSSION

The data sets were generated separately by A. Sesana and were blindly analyzed by A. Petiteau and S. Babak. The MS-GA was applied to all data sets, adding sources one by one to the template. By doing this, we could test the effectiveness of the code in determining both the number of sources in the data set and their sky location. A summary of the results is given in Table II.

For each data set we evolved several colonies of organisms assuming $N_s = 1, 2, 3, \ldots$ in the template, we computed the SNR of the best organism at the end of each search and tracked its evolution with N_s. Results are shown in the left panel of Fig. 2 for Dataset3. The maximum SNR steadily increases by adding sources up to $N_s = 8$. Adding a ninth source to the template does not significantly improve the match with the data, indicating that the data set is best described by an eight-source model; in fact, there were eight sources in Dataset3. The algorithm identified the correct number of sources in all data sets. We stress here that all the injected sources had SNR > 10, high enough to be dug out of the noise. In the presence of many low SNR sources, we do not expect any search algorithm to recover the correct number of binaries, only to identify the brightest ones. We will address this "confusion problem" in a future paper. A complementary view of this result is given in the right panel of Fig. 2. We ran several GAs using nine-source colonies of organisms, we identified the solution (organism) with highest SNR ($\mathrm{SNR}_{\mathrm{best}}$), and we stored all the organisms having $\mathrm{SNR}^2_{\mathrm{tot}} > 99\%\mathrm{SNR}^2_{\mathrm{best}}$. The figure shows the location in the sky of all sources found in all these "best solutions." The location of sources 1 to 8 does not change much for different solutions, and it is generally consistent with the true (blue crosses) locations of the injected sources. Conversely, the 9th source (green circles) is extremely scattered around the sky. Moreover, the frequency-SNR plot at the extreme right shows that the individual SNR of those 9th sources are almost always <5, compatible with noise fluctuations.

Having tested the code effectiveness in finding the number of sources present in the data, we turn now to the description of the results obtained on the individual data sets. Best solutions (those with $\mathrm{SNR}^2_{\mathrm{tot}} > 99.5\%\mathrm{SNR}^2_{\mathrm{best}}$)

TABLE II. Recovered (and injected) parameter values of all the simulated sources in each data set. The last column represents the sky offset of the recovered sources with respect to the injection (see text for details).

	SNR	f[nHz]	θ[rad]	ϕ[rad]	$\Delta\Theta$[deg]
Dataset1	60.70 (61.11)	56.6 (56.4)	1.249 (1.237)	2.604 (2.601)	0.706
	45.85 (42.48)	38.0 (38.0)	1.750 (1.748)	3.765 (3.764)	0.127
	43.71 (40.43)	36.4 (36.4)	1.555 (1.529)	1.722 (1.712)	1.596
	35.67 (36.14)	53.7 (53.5)	0.537 (0.534)	5.522 (5.451)	2.085
	32.27 (31.33)	48.3 (48.0)	1.286 (1.295)	5.144 (5.123)	1.266
Dataset2	54.64 (54.07)	18.88 (18.9)	1.774 (1.774)	3.839 (3.841)	0.112
	48.01 (47.24)	11.25 (11.3)	1.870 (1.858)	5.720 (5.718)	0.696
	13.64 (13.05)	77.42 (76.5)	0.617 (0.651)	6.158 (6.115)	2.434
	12.23 (12.78)	57.19 (57.0)	1.613 (1.549)	6.050 (6.048)	3.669
Dataset3	44.91 (42.99)	19.33 (19.3)	0.474 (0.468)	1.450 (1.454)	0.359
	37.39 (37.72)	25.42 (25.4)	0.883 (0.878)	2.733 (2.749)	0.763
	26.02 (27.09)	13.21 (13.2)	1.769 (1.764)	5.078 (5.087)	0.581
	20.19 (20.88)	83.42 (82.4)	0.689 (0.668)	4.133 (4.162)	1.593
	19.67 (18.51)	39.79 (39.8)	0.541 (0.509)	0.386 (0.429)	2.211
	17.27 (16.59)	33.16 (33.1)	1.381 (1.397)	3.621 (3.693)	4.160
	13.07 (13.19)	73.83 (73.0)	1.534 (1.536)	5.054 (5.078)	1.379
	10.66 (11.51)	82.75 (81.8)	0.809 (0.864)	6.182 (6.085)	5.192
Dataset4	42.73 (43.92)	98.2 (96.3)	2.028 (2.043)	0.977 (0.961)	1.200
	28.62 (29.28)	91.5 (90.1)	2.655 (2.661)	1.174 (1.121)	1.454
	27.56 (28.28)	48.2 (48.1)	1.231 (1.245)	5.774 (5.769)	0.827

for Dataset1 are shown in the top panel of Fig. 3. All the five injected sources were found at approximately the right sky location, with the right frequency and SNR. Our GA is designed to find the modes corresponding to maxima in the likelihood function, but not to explore the exact shape of the likelihood function around those modes. The lack of parameter space exploration around the maxima prevents us from attaching fully meaningful errors to our best solutions. We plan to include systematic exploration of the maxima in future work; here we just estimate the sky location error as the angular offset between the best solution and the injected signal. This is defined as $\Delta\Theta = \arccos(\vec{n}_t \cdot \vec{n}_r)$, where $\vec{n}_t$ and $\vec{n}_r$ are the unit vectors defining the true sky location of the sources and the recovered value, respectively. This is reported in the last column of Table II. All sources in Dataset1 are offset by less

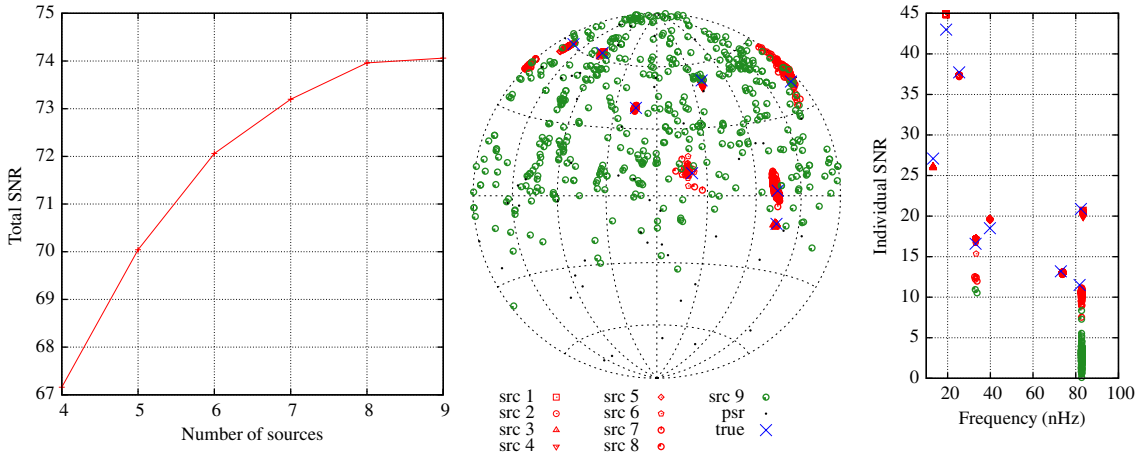

FIG. 2 (color online). Performance of the MS-GA in finding the number of sources. Left panel: signal SNR as a function of the number of sources N_S assumed in the template. Right panel: source localization in the sky (sky map) and in the frequency-SNR space (rightmost panel). Blue crosses (×) correspond to injected values and black dots to the position of the MSPs forming the array. Red marks represent sources 1–8 found by all the organisms with $\mathrm{SNR}^2_{\mathrm{tot}} > 99\%\mathrm{SNR}^2_{\mathrm{best}}$, while green circles represent the 9th source. Note that this latter one does not have a defined position, and typically has SNR < 5.

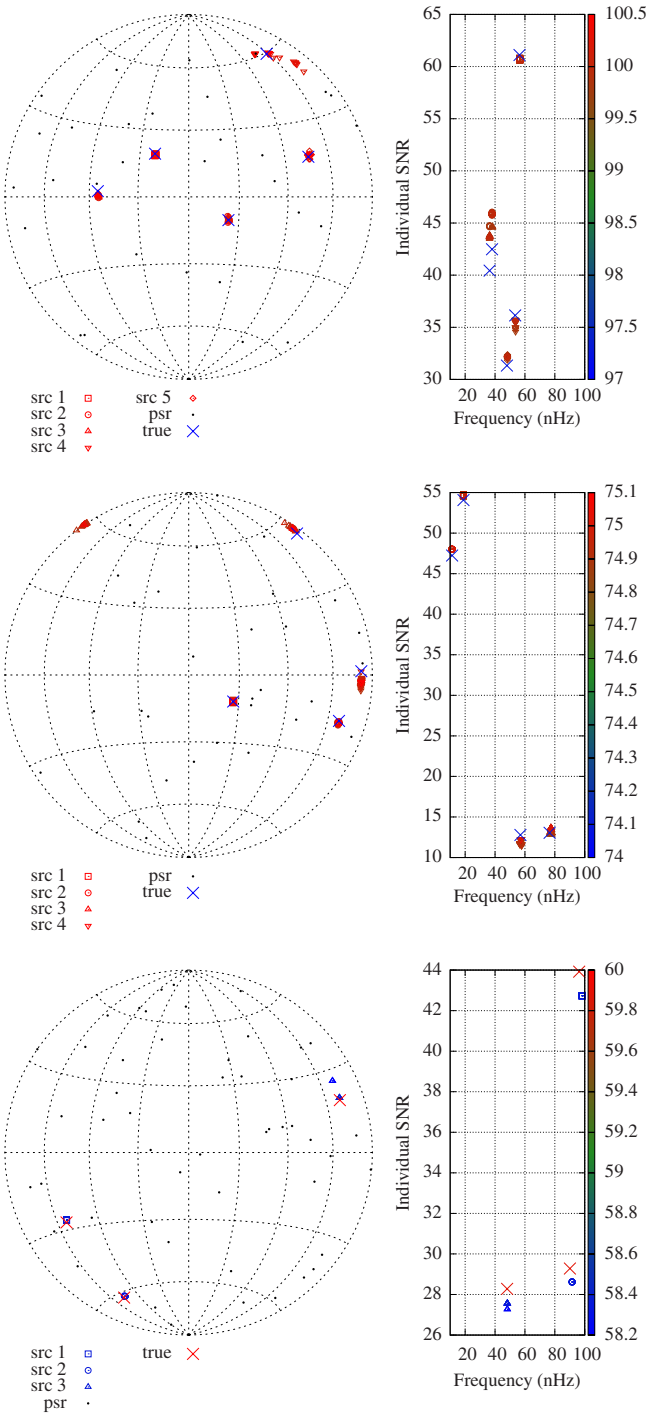

FIG. 3 (color online). Best solutions for Dataset1 (top panel), Dataset2 (central panel) and Dataset4 (bottom panel). In each panel all solutions with $\mathrm{SNR}^2_{\mathrm{tot}} > 99.5\%\mathrm{SNR}^2_{\mathrm{best}}$ are shown. Symbols have the same meaning as in the right panel of Fig. 2. All recovered sources are color-coded according to the rightmost scale, based to the total SNR of the solution they belong to.

than 2 degrees. Results for Dataset2 are shown in the central panel of Fig. 3. Again, we see that all sources are correctly identified, despite two of them having SNR just above 10. Sky location offsets $\Delta\Theta$ are less than 1 degree for the two low-frequency bright sources, but degrade to $\sim$3 degrees for the high-frequency, faint ones. Dataset3 was the richest of all, with eight injected sources. Best solutions are shown in Fig. 4, for different SNR threshold, to give a sense of how fast points cluster toward the maximum of the likelihood. Also in this case, all sources are well located in the sky, with brighter sources located best. Looking at Table II, we notice that we tend to overestimate the frequencies of sources above 60 nHz. This is because at such high f, the $10^9\mathrm{M}_\odot$ chirp mass binaries injected in the data chirp significantly over the ten-year duration of the observations. This is the bias we mentioned in Sec. IV. Since we are matching the signal with non-evolving monochromatic templates, the estimated frequency is higher than the one injected at the beginning of the observation. The mismatch is larger for higher frequencies due to the faster evolution of the sources. We have checked that this bias is indeed due to measurable orbital evolution (over the observation time) by injecting in the data strictly monochromatic GW sources (the same used for the templates). On the other hand, we do not observe any bias in the sky localization caused by the frequency evolution. Since the bias in the frequency is small, a meaningful strategy would be to perform an initial search assuming $\dot{f} = 0$ throughout the whole frequency band, and then use the best solution as an initial guess for a refined local search including frequency derivative for the GW sources with $f > 60$ nHz. We will discuss how likely such sources are (with measurable $\dot{f}$) at the end of this section. The same effect is seen in Dataset4 (bottom panel of Fig. 3 and Table II). Also in this case, we find source offsets within $\sim$1 degree of their true position, but we give a couple of extra nHz to the high-frequency sources. Different noise levels in the pulsars do not affect the performance of our search algorithm.

Overall, our MS-GA performed well on all data sets, recovering all the injected sources without returning any false positives. The parameters of the recovered sources well matched the injections with (i) sky location offsets less than a few degrees, (ii) individual source SNR estimations within few percent of the true ones, and (iii) sub-Fourier-bin frequency accuracy (sometimes within 0.1 nHz for low-frequency sources). Without a complete exploration of the likelihood function around the maxima, it is difficult to asses proper errors on the parameters. We can, however, take sky position offsets as a proxy of the sky localization accuracy. In fact, offsets shown in the last column of Table II scale (with a large scatter) with the inverse of the SNR. This has to be expected: an offset scaling with 1/SNR implies an area of uncertainty scaling with $1/\mathrm{SNR}^2$, in agreement with theoretical expectations. If we approximate the error box in the sky as $\Delta\Omega \approx \pi[\Delta\Theta]^2$, we get values in the range 10–70 deg^2 for sources with SNR in the range 11–13. This is broadly consistent with Ref. [21], which estimated an average

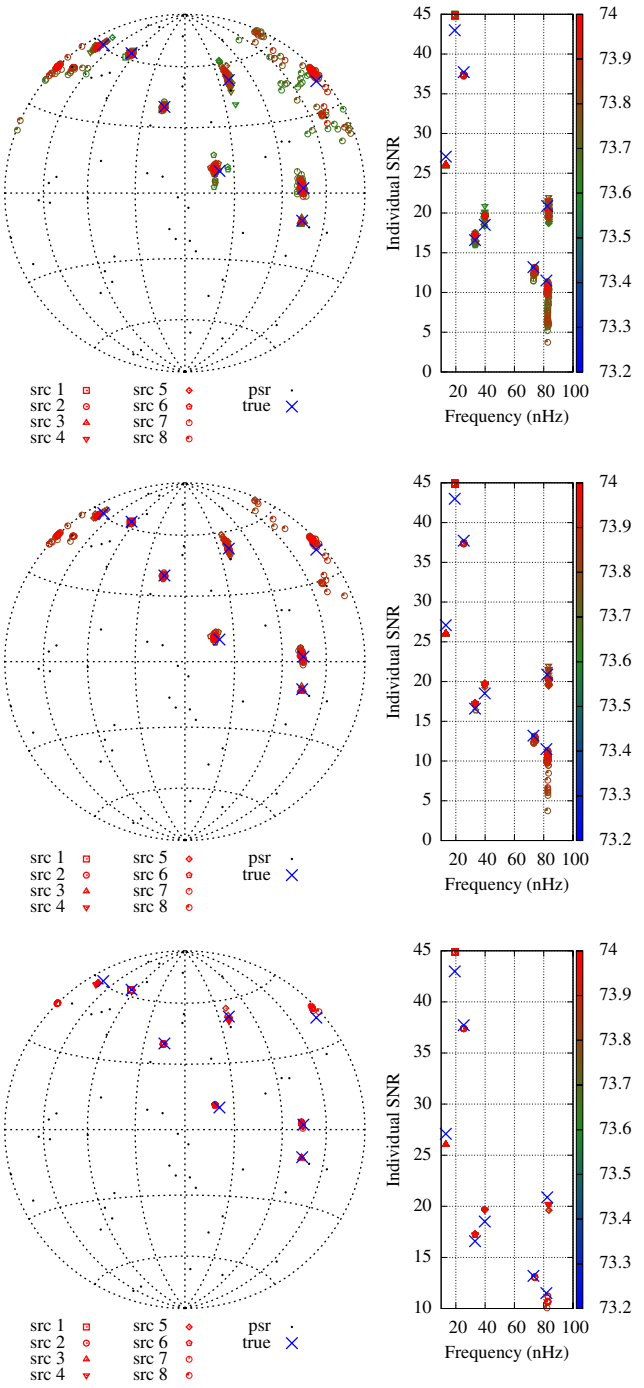

FIG. 4 (color online). Best solutions for Dataset3. The top panel shows all solutions with $\mathrm{SNR}^2_{\mathrm{tot}} > 99\%\mathrm{SNR}^2_{\mathrm{best}}$, the central panel all solutions with $\mathrm{SNR}^2_{\mathrm{tot}} > 99.5\%\mathrm{SNR}^2_{\mathrm{best}}$, and the bottom panel all solutions with $\mathrm{SNR}^2_{\mathrm{tot}} > 99.8\%\mathrm{SNR}^2_{\mathrm{best}}$. Symbols have the same meaning as in the right panel of Fig. 2. All recovered sources are color coded according to the rightmost scale, based to the total SNR of the solution to which they belong.

sky location accuracy of $\Delta\Omega \approx 50\ \mathrm{deg}^2$ for a source observed by an array of 50 pulsars, randomly located in the sky, with total SNR = 10 (in the earth term). As mentioned above, another interesting fact is the frequency mismatch for sources approaching 10^{-7} Hz, caused by their frequency evolution over the observing time (ten years). This means that, in principle, for such sources, we can measure the frequency drift $\dot{f}$, i.e., the chirp rate. The measure of $\dot{f}$ breaks the chirp mass/luminosity distance degeneracy in the source amplitude, allowing for a direct measurement of the source luminosity distance. This, ultimately, will narrow down significantly the number of candidate electromagnetic counterparts in the source sky error box, facilitating a positive identification. However, evolution on such short timescales is detectable only for very massive ($\mathcal{M} \sim 10^9\mathrm{M}_\odot$, like the systems injected in the data) binaries, emitting at frequencies higher than $\sim 7 \times 10^{-8}$ Hz. Intrigued by this possibility, we checked how likely is to find such extreme systems in realistic populations of MBH binaries in the Universe. We took the models investigated by Sesana *et al.* [10] and computed the average number of expected sources with $\mathcal{M} > 10^9\mathrm{M}_\odot$ and $f > 7 \times 10^{-8}$. Depending on the adopted MBH mass-bulge relation and on the accretion implementation [see Ref. [10], for details], we found an average number of sources, ranging from 10^{-3} to 0.04, i.e., there is less than a 5% chance of having such a bright high-frequency source in the sky. If we relax the mass requirement to $\mathcal{M} > 5 \times 10^8\mathrm{M}_\odot$, figures grow to 0.01–0.4. To properly quantify the probability of measuring $\dot{f}$, one should estimate its minimum measurable value for a given array, and then select in the MBH binary population all the sources occupying the portion of the chirp mass-frequency parameter space compatible with such value. The crude figures estimated here indicate that $\dot{f}$ measurements using the earth term only should be unlikely. The presence of up to two to three such sources in our data sets reflects the fact that they were constructed purely as a test bed for our multiple searches genetic algorithm search and not based on realistic MBH binary populations.

VI. CONCLUSIONS

This is the second in a series of papers devoted to the exploration of the PTA potential of resolving multiple GW sources. In Paper I we addressed basic issues, such as the number of sources per frequency bin that can be resolved by an array of N pulsars, demonstrating our findings with primitive searches on several (mostly noiseless) synthetic data sets. Here we pushed our analysis a bit further by (i) extending the mathematical formulation of the likelihood function to include the source frequencies as additional free parameters and by (ii) implementing a multisearch genetic algorithm to efficiently find the maximum of the likelihood function.

We constructed synthetic data sets consisting of collections of time series representing the residuals obtained by timing an ensemble of MSPs. MSPs were placed randomly in the sky, each time series consisted of 523 equally sampled data points over an observing time of ten years (one data

point every week), and the noise in the data was assumed to be white Gaussian. In each data set, we injected an unknown number N_S of sources with random parameters and individual SNR > 10, and we apply our multisearch genetic algorithm to search for their sky location and frequency. Note that we assumed circular monochromatic sources in our template, but we allowed for full post-Newtonian evolution of the injected sources. By doing so, we placed ourselves in the (likely) situation in which the theoretical model of the signal does not perfectly represent its real nature, and we explored the consequences of this mismatch.

Our main results can be summarized as follows:

(i) The MS-GA generally converged to the true maximum of the likelihood function in two to five iterations (a few hours on one core at 2 GHz).

(ii) The MS-GA successfully identified all the injected sources in all data sets. No false positives were found.

(iii) The search on all source parameters was successful: inferred sky locations were offset by less than a few degrees, individual source SNR estimations matched the injections within a few percent, and frequencies were determined with sub-Fourier-bin precision (most of the times to better than 0.1 nHz).

(iv) The sky location offsets roughly scaled with 1/SNR, implying a sky location accuracy scaling as $1/\mathrm{SNR}^2$. Even though we did not compute proper error boxes in the sky, we estimated source localization capabilities broadly consistent with theoretical expectations derived in Ref. [21] under similar assumptions.

(v) We overestimated the frequency of sources approaching $f = 10^{-7}$ Hz. This is because massive systems at such high frequency significantly chirp during the observation time (whereas chirp was not allowed in our template). This means that we can measure $\dot{f}$ and, therefore, estimate the chirp mass and, in turn, the luminosity distance of the source. Although this is a very appealing prospect, we estimated on average less than one source with a measurable $\dot{f}$ in a realistic realization of the MBH binary population in the Universe.

(vi) The MS-GA performances do not seem to be affected by unequal noise levels in different MSPs.

Our results are encouraging; however, they were still obtained under a number of simplifying assumptions that we wish to relax in our future work. First, data sets were still evenly sampled, with no gaps, an ideal situation that is not going to occur in reality. Second, we took noisy data streams and fit for the GW sources only, implicitly assuming perfectly known MSP parameters; any realistic detection pipeline must fit for MSP parameters and GW signals simultaneously. Finally, we still did not include the pulsar terms in our injections; those are likely to blend together with lower-frequency earth terms to bias estimated source parameters and (possibly) to create false positives. Only by relaxing those assumptions will we be able to demonstrate the effectiveness of our MS-GA algorithm in tackling a problem with realistic complexity. We plan to investigate these issues in the next paper of the series. We will then try to apply our search algorithm to raw times of arrival, carrying the imprint of a realistic population of MBH binaries.

ACKNOWLEDGMENTS

Work of A. S. and S. B. was supported in part by DFG Grant No. SFB/TR 7 Gravitational Wave Astronomy and by DLR (Deutsches Zentrum fur Luft- und Raumfahrt). The authors acknowledge the support of their colleagues in the EPTA.

[1] G. H. Janssen *et al.*, in *40 Years of Pulsars: Millisecond Pulsars, Magnetars and More*, edited by C. Bassa *et al.*, American Institute of Physics Conference Series Vol. 983 (AIP, New York, 2008), p. 633.

[2] R. N. Manchester, in *40 Years of Pulsars: Millisecond Pulsars, Magnetars and More*, edited by C. Bassa *et al.*, American Institute of Physics Conference Series Vol. 983 (AIP, New York, 2008), p. 584.

[3] F. Jenet *et al.*, arXiv:0909.1058.

[4] G. Hobbs *et al.*, Classical Quantum Gravity **27**, 084013 (2010).

[5] R. Smits, D. R. Lorimer, M. Kramer, R. Manchester, B. Stappers, C. J. Jin, R. D. Nan, and D. Li, Astron. Astrophys. **505**, 919 (2009).

[6] J. Lazio, in *Proceedings of Panoramic Radio Astronomy: Wide-field 1–2 GHz Research on Galaxy Evolution, Groningen, the Netherlands, 2009*, edited by G. Heald and P. Serra, http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=89.

[7] R. van Haasteren *et al.*, Mon. Not. R. Astron. Soc. **414**, 3117 (2011).

[8] P. B. Demorest *et al.*, arXiv:1201.6641.

[9] A. Sesana, A. Vecchio, and C. N. Colacino, Mon. Not. R. Astron. Soc. **390**, 192 (2008).

[10] A. Sesana, A. Vecchio, and M. Volonteri, Mon. Not. R. Astron. Soc. **394**, 2255 (2009).

[11] M. V. Sazhin, Sov. Astron. **22**, 36 (1978).

[12] R. W. Hellings and G. S. Downs, Astrophys. J. Lett. **265**, L39 (1983).

[13] G. Nelemans, L. R. Yungelson, and S. F. Portegies Zwart, Astron. Astrophys. **375**, 890 (2001).

[14] P. Amaro-Seoane *et al.*, arXiv:1201.3621.

[15] F. A. Jenet, G. B. Hobbs, K. J. Lee, and R. N. Manchester, Astrophys. J. Lett. **625**, L123 (2005).

[16] F. A. Jenet, G. B. Hobbs, W. van Straten, R. N. Manchester, M. Bailes, J. P. W. Verbiest, R. T. Edwards, A. W. Hotan, J. M. Sarkissian, and S. M. Ord, Astrophys. J. **653**, 1571 (2006).
[17] M. Anholm, S. Ballmer, J. D. E. Creighton, L. R. Price, and X. Siemens, Phys. Rev. D **79**, 084030 (2009).
[18] R. van Haasteren, Y. Levin, P. McDonald, and T. Lu, Mon. Not. R. Astron. Soc. **395**, 1005 (2009).
[19] B. Kocsis and A. Sesana, Mon. Not. R. Astron. Soc. **411**, 1467 (2011).
[20] F. A. Jenet, A. Lommen, S. L. Larson, and L. Wen, Astrophys. J. **606**, 799 (2004).
[21] A. Sesana and A. Vecchio, Phys. Rev. D **81**, 104008 (2010).
[22] V. Corbin and N. J. Cornish, arXiv:1008.1782.
[23] D. R. B. Yardley *et al.*, Mon. Not. R. Astron. Soc. **407**, 669 (2010).
[24] K. J. Lee, N. Wex, M. Kramer, B. W. Stappers, C. G. Bassa, G. H. Janssen, R. Karuppusamy, and R. Smits, Mon. Not. R. Astron. Soc. **414**, 3251 (2011).
[25] J. A. Ellis, X. Siemens, and J. D. E. Creighton, Astrophys. J. **756**, 175 (2012).
[26] A. Sesana, C. Roedig, M. T. Reynolds, and M. Dotti, Mon. Not. R. Astron. Soc. **420**, 860 (2012).
[27] T. Tanaka, K. Menou, and Z. Haiman, Mon. Not. R. Astron. Soc. **420**, 705 (2012).
[28] S. Babak and A. Sesana, Phys. Rev. D **85**, 044034 (2012).
[29] A. Petiteau, Y. Shang, S. Babak, and F. Feroz, Phys. Rev. D **81**, 104016 (2010).
[30] S. Babak *et al.*, Classical Quantum Gravity **25**, 184026 (2008).
[31] Mock LISA Data Challenge Task Force, S. Babak *et al.*, Classical Quantum Gravity **27**, 084009 (2010).
[32] European Pulsar Timing Array, http://www.epta.eu.org/.
[33] C. M. F. Mingarelli, K. Grover, T. Sidery, R. J. E. Smith, and A. Vecchio, Phys. Rev. Lett. **109**, 081104 (2012).
[34] A. Sesana, Astrophys. J. **719**, 851 (2010).
[35] P. Jaranowski and A. Królak, Living Rev. Relativity **15**, 4 (2012).
[36] N. J. Cornish and E. K. Porter, Classical Quantum Gravity **24**, 5729 (2007).
[37] Y. Wang and S. D. Mohanty, Phys. Rev. D **81**, 063002 (2010).
[38] W. Junker and G. Schaefer, Mon. Not. R. Astron. Soc. **254**, 146 (1992).
[39] L. Barack and C. Cutler, Phys. Rev. D **69**, 082005 (2004).
[40] B. Willems, A. Vecchio, and V. Kalogera, Phys. Rev. Lett. **100**, 041102 (2008).