Bayesian inference on EMRI signals using low frequency approximations

Bayesian inference on EMRI signals using low frequency approximations

Asad Ali[1,2]**, Nelson Christensen**[3]**, Renate Meyer**[1] **and Christian Röver**[4,5]

[1] Department of Statistics, The University of Auckland, Auckland, New Zealand
[2] Department of Space Science, Institute of Space Technology (IST) Islamabad, Islamabad
44000, Pakistan
[3] Department of Physics and Astronomy, Carleton College, Northfield, MN, USA
[4] Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut) and Leibniz Universität
Hannover, 30167 Hannover, Germany

E-mail: asad.ali@ist.edu.pk, nchriste@carleton.edu, renate.meyer@auckland.ac.nz and
christian.roever@med.uni-goettingen.de

Abstract

Extreme mass ratio inspirals (EMRIs) are thought to be one of the most exciting
gravitational wave sources to be detected with LISA. Due to their complicated
nature and weak amplitudes the detection and parameter estimation of such
sources is a challenging task. In this paper we present a statistical methodology
based on Bayesian inference in which the estimation of parameters is carried out
by advanced Markov chain Monte Carlo (MCMC) algorithms such as parallel
tempering MCMC. We analysed high and medium mass EMRI systems that
fall well inside the low frequency range of LISA. In the context of the Mock
LISA Data Challenges, our investigation and results are also the first instance
in which a fully Markovian algorithm is applied for EMRI searches. Results
show that our algorithm worked well in recovering EMRI signals from different
(simulated) LISA data sets having single and multiple EMRI sources and holds
great promise for posterior computation under more realistic conditions. The
search and estimation methods presented in this paper are general in their
nature, and can be applied in any other scenario such as AdLIGO, AdVIRGO
and Einstein Telescope with their respective response functions.

PACS numbers: 04.80.Nn, 02.70.Uu

1. Introduction

It is likely that most of the galaxies, including our own milky way, host super massive black
holes (SMBHs) with masses of order $10^6 M_\odot$ ($M_\odot$ = solar mass) or larger in their centres.
These SMBHs are surrounded by a large population of stellar mass compact objects (COs) such

[5] Present address: Department of Medical Statistics, University Medical Center Göttingen, 37073 Göttingen,
Germany.

as neutron stars, white dwarfs and small black holes with masses $\sim 10\,M_\odot$. Due to multi-body interactions most of these COs occasionally end up being trapped in an orbit which passes too close to the central mass. Once captured in the strong gravitational field of an SMBH, a CO then starts orbiting around the central mass in an eccentric orbit which decays with time and the CO eventually spirals into the central mass. These inspirals are called extreme mass ratio inspirals (EMRIs) because of the large difference in the masses of the two bodies. EMRIs are considered to be one of the most important potential sources of gravitational waves (GW's), to be detected with the laser interferometer space antenna[6] (LISA) [1]. These inspirals may encode information about the structure of the central black hole, its evolution and other features such as the Lense–Thirring effects [2] and spin–orbit coupling [3].

Bayesian approaches along with MCMC methods have been in use by different groups working on the GW source detection and parameter estimation problems (see e.g. [4–7] and many others). In the context of EMRI signals the reversible jump MCMC algorithm was employed in [8]. Another Monte Carlo, but not rigorously Markovian, approach in which the successive states are chosen from directed proposal distributions, was used in [9–11].

In this work we present results of the applications of the Bayesian approach using the parallel tempering MCMC (PTMCMC) algorithm [12] for the detection and parameter estimation of EMRI signals in LISA data. The noise spectrum is assumed to be unknown and for this reason we used Whittle's approximation to the Gaussian likelihood [13]. The Whittle likelihood uses the approximate properties of discrete Fourier transform (DFT) and assumes that the DFTs are approximately independent normally distributed with mean zero and variance proportional to the power spectral density.

This paper is organized as follows: in section 2 we present the Bayesian statistical model, comprising the waveform model, our assumptions for the error distribution with unknown power spectrum, and prior distributions of the parameters. Section 3 briefly explains the Bayesian approach to detection and parameter estimation and details the MCMC search algorithm. The implementation of the algorithm is described in detail in section 4. Conclusions are given in section 5.

2. Bayesian statistical model

In this section, we outline all components of our statistical model that encompasses the waveform model and detector response, the observation error assumptions that lead to the Whittle likelihood and the prior distribution of the parameters.

2.1. The waveform model and detector response

The EMRI sources given in MLDC [14] LISA data are generated by Barack and Cutler's *analytic kludge* waveform (AKW) approximation [15]. The waveform model is described by 14-dimensional parameter set $\theta = (\nu_0, M, \mu, e_0, \tilde{\gamma}_0, \Phi_0, \alpha_0, \theta_S, \phi_S, \lambda, \chi, \theta_K, \phi_K, D_L)$, where ν_0 is the initial orbital frequency, μ and M are the masses of CO and SMBH, respectively, e_0 is the initial eccentricity, $\tilde{\gamma}_0$, Φ_0 and α_0 are the initial orbital phase angles, θ_S and ϕ_S are the ecliptic latitude and longitude, respectively, λ is the orbital inclination, χ is SMBH's spin with θ_K, ϕ_K its orientation angles and D_L is luminosity distance. In the Barack and Cutler parametrization, the log-transformed values of the parameters ν_0, μ, M and D_L are used. For the MCMC searches a suitable alternative is to use the truncated AKWs [10]. These

---

[6] We realize that the state of LISA is in limbo after NASA's decision to leave the project, but the European Space Agency is investigating carrying on with a scaled down version of the project, and the conclusions we reach in this paper will certainly be applicable to that mission if it moves forward.

waveforms can be computed $\sim$3 times faster than full AKWs and there is a typical overlap of $\sim$90–95% between the two. The multiple harmonics and the complex structure of EMRIs pose a challenge as far as their computation and the statistical estimation of their source parameters is concerned. The likelihood surface for EMRIs contains multiple local peaks corresponding to different harmonics, of which some are as high as 85% of the peak of the dominant harmonic. Furthermore, the orbital evolution (over time) introduces much more uncertainty to the likelihood surface. This means that the orbital parameters keep changing throughout the signal life. To overcome these issues, sophisticated algorithms are needed that can deal with multi-modality due to multiple harmonics and the increased uncertainty due to the time-varying nature of orbital parameters. One such algorithm will be presented in the subsequent sections.

The full descriptions of both the full AKW and truncated AKW models are extensive so we give only the final expressions of the model which is a pair of two polarization signals defined as

$$h_+(t) = A^+(t)\cos 2\psi(t) + A^\times(t)\sin 2\psi(t), \tag{1}$$

$$h_\times(t) = -A^+(t)\sin 2\psi(t) + A^\times(t)\cos 2\psi(t), \tag{2}$$

where $\psi(t)$ is the polarization angle defined as

$$\psi(t) = \arctan\left(\frac{\cos\theta_S \sin\theta_L(t)\cos(\phi_S - \theta_L(t)) - \cos\theta_L(t)\sin\theta_S}{\sin\theta_L(t)\sin(\phi_S - \theta_L(t))}\right), \tag{3}$$

where θ_S and ϕ_S are ecliptic latitude and longitude, respectively, and $\theta_L(t)$ and $\phi_L(t)$ are the time-varying angles specifying the instantaneous direction of the angular momentum.

Unlike ground-based interferometers the LISA arms are at $60°$ angles, and the interferometers are susceptible to '+' and '$\times$' polarizations. Due to its orbital dynamics, the detector response is complicated. The detector output may be derived numerically [16–18], while in the limit of long wavelength/low frequency approximation (LFA) the mapping simplifies [15, 19], which is the regime we will be concerned with here. The three spacecraft's output may be re-combined into two variables with stochastically independent noise components, namely the '*A*' and '*E*' variables [20]; data analysis in the following is going to be based on these two time series.

Although the LFA is simpler and faster, the amplitude of the final LISA response is not exactly the same as that of [17, 18], which we will refer to as the *full LISA response* throughout this paper. While working on MLDC 4 blind data, we found that the amplitudes of the LFA response are by a factor of $\sim$3 larger than those derived by full LISA response. We compared the two responses for several (noise-less) EMRI signals posted on the MLDC webpage [21] including those given in the training data of MLDC 4 and saw the same phenomena. This difference in amplitudes results in an incorrect estimation of luminosity distance (D_L) and sky location angles (θ_S, ϕ_S) as we have found in several test MCMC searches on MLDC 1B EMRI data sets. Particularly, during MCMC searches the chains for sky location angles used to get locked at wrong positions and then nothing could move them from those positions. The same phenomenon (wrong estimation of distance and sky location) was observed by another group too [10, 22]. As an ad hoc solution we divided the overall amplitude in our signal model by 3 and used this option throughout the MLDC 4 blind searches, more investigation regarding the exact origin is required. When the overall amplitude in LFA was divided by 3, the sampler was able to converge to the correct values of the luminosity distance and sky location angles. It is of interest to note that in both cases (i.e. the adjusted and unadjusted amplitudes) the other key parameters, e.g. ν, μ, M, e and χ are almost unaffected and remain the same. Later on, near the submission of MLDC 4 blind entries, we further found that for high mass EMRIs the difference factor is generally rather larger than 3, it is generally $\sim$6.

2.2. Prior information

In order to implement the *a priori* information about the parameters to be inferred, we followed
the specifications given in the MLDC Task Force's documentation [14, 21], and for different
parameters we employed the following prior distributions: $\mu \in$ Uniform[9.5, 10.5]$M_\odot$,
$e_0 \in$ Uniform[0.15, 0.25], $\chi \in$ Uniform[0.5, 0.7]. The range of the prior distribution of e_0 was
changed to [0.10, 0.40] because of its initial values given for different MLDC EMRI sources
being beyond the range [0.15, 0.25]. The prior for SMBH mass are different for different
sources: $M \in$ Uniform[0.95, 1.05] $10^7 M_\odot$ (high mass), $M \in$ Uniform[4.75, 5.25] $10^6 M_\odot$
(medium mass) and $M \in$ Uniform[0.95, 1.05] $10^6 M_\odot$ (low mass). For initial azimuthal orbital
frequency ν_0, the prior distribution was set to be uniform over the range [5.0×10^{-5}, 0.01].
For the polar angles we assumed isotropic priors. For some parameters such as ν_0, μ, M
and D_L, their logarithmic values were used; as in [15], for ease of specification of proposal
distributions, the prior densities for these parameters were transformed accordingly. For the
noise power spectrum we used the conjugate Inv$-\chi^2$ prior distribution, where the prior scale
parameters were estimated based on a disjoint stretch of the same data set [23]. In the following
subsections we present some results obtained with the application of the above search and
estimation algorithm on different MLDC data sets containing single and multiple EMRI
sources.

3. Parameter estimation

In this section, we briefly describe the Bayesian approach to statistical inference and posterior
computation via MCMC. Furthermore, we give an overview of the tempering methods and the
parallel tempering strategy that we employed.

3.1. Posterior inference

We are interested in inferring *a priori* unknown parameters θ from measured data y. To
this end we derive the parameters' posterior probability distribution, which expresses the
information on the actual parameter values after consideration of the observed data by assigning
probabilities to regions of the parameter space. Given the relationship between parameters and
data through the likelihood function $p(y|\theta)$, the posterior density function is given by Bayes'
theorem as

$$p(\theta|y) = \frac{p(y|\theta) \, p(\theta)}{p(y)} \tag{4}$$

[24]. Here $p(\theta)$ is the prior probability density, expressing any information we have about the
parameters before accounting for the measured data. The posterior distribution (4) is essentially
given by the product of prior and likelihood, while the *evidence* $p(y) = \int p(y|\theta) \, p(\theta) \, d\theta$
is commonly of minor concern for parameter estimation purposes, as it only constitutes a
normalizing constant.

3.2. Monte Carlo integration

Bayes theorem supplies the posterior probability distribution in terms of its density function.
In order to extract information relevant for any particular purpose, one may be interested
in marginal posterior distributions, posterior expectations, quantiles, etc; what is commonly
required is the evaluation of *integrals* with respect to the posterior distribution. As these
integrals are rarely analytically tractable, stochastic integration methods are commonly used

to approach posterior inference. Such Monte Carlo methods aim at approximating the desired integrals via random sampling; instead of computing an expectation value analytically, random draws from the distribution of interest are generated, and the expectation is then approximated by an average. Analogous procedures are applied for other figures of interest, like quantiles or marginal densities for example, and an arbitrary accuracy may be achieved by increasing the sample size. A popular variety of such Monte Carlo procedures is the Metropolis(–Hastings) algorithm [25]; what makes this particular algorithm attractive for Bayesian inference is the fact that it only requires the unnormalized probability density function of the distribution of interest as an input, which in this case is supplied by Bayes' theorem (4). Instead of providing stochastically independent samples from the distribution, the Metropolis–Hastings algorithm will generate a Markov chain of subsequently dependent draws whose stationary distribution is the distribution of interest. At each step in the generation of the random sequence, only ratios of probability density values need to be considered, so that an overall normalization constant (like $p(y)$ in (4)) does not need to be known. To approximate a target distribution with density function $p(\theta|y)$, the Metropolis(–Hastings) algorithm employs an acceptance–rejection rule to construct a Markov chain from an auxiliary density $q(\cdot|\cdot)$, which is referred to as proposal density or transition probability function. Being at current state $\theta^{(t)}$ the acceptance probability for moving to a new state θ' is defined as

$$\alpha = \min\left\{1, \frac{p(\theta'|y)q(\theta^{(t)}|\theta')}{p(\theta^{(t)}|y)q(\theta'|\theta^{(t)})}\right\}. \tag{5}$$

Taking $q(\cdot|\cdot)$ to be symmetric, i.e. $q(\theta|\theta') = q(\theta'|\theta)$, leads to the basic Metropolis algorithm [26] with acceptance probability given by

$$\alpha = \min\left\{1, \frac{p(\theta'|y)}{p(\theta^{(t)}|y)}\right\}. \tag{6}$$

3.3. Tempering methods

In most multi-dimensional cases, the density surface of complicated target distributions turns out to have multiple secondaries or local modes that are well separated by deep valleys of low probability regions. The simple Metropolis(–Hastings) algorithm tends to get stuck at some of these local modes for a prohibitively long time before reaching the global mode. Analogous to an annealing process, temperature is used to scale the target density in order to flatten the local modes so that the MCMC sampler moves freely towards the global maximum without being trapped in local modes. At a given temperature $T \geqslant 1$, samples are generated from a tempered version of the target density $p(\theta|y)$ defined as

$$p_T(\theta|y) \propto p(\theta|y)^{\frac{1}{T}}, \tag{7}$$

where $T = 1$ yields the actual target distribution. This heating is equivalent to increasing the standard deviation of the target density by a factor $\sqrt{T}$, therefore as T increases the heated distribution becomes flatter and gets closer to the uniform distribution, which enables the Markov chain to move more freely and hence faster towards higher probability regions.

Some of the most popular tempering methods are simulated annealing [27], simulated tempering [28] and PTMCMC aka Metropolis coupled-MCMC. PTMCMC introduced in [12], is a powerful optimization of the simple Metropolis(–Hastings) algorithm which is very effective in improving the mixing of MCMC chains and in particular in escaping the local modes. The algorithm works by running multiple MCMC chains in parallel, each simulating a separate target density characterized by a different temperature and occasionally attempting swaps of its current states. As stated above, in principle, a high temperature chain sees

the density of the target distribution as more *flattened* relative to a low temperature chain, which means that the high temperature chain can move more freely across the valleys of low probability regions in between modes. In order to make the low temperature chain benefit from the high temperature chain that may be sampling near another mode, an exchange of information about the current states is attempted by proposing a swap of the current states using an additional Metropolis acceptance–rejection step. Detailed descriptions of this algorithm can be found in [7, 29, 30].

3.4. The likelihood function

The data, which is sampled at discrete time steps, can be represented by the following signal plus noise model,

$$y(t) = s(t, \theta) + \epsilon(t) \qquad t = 1, \ldots, N, \tag{8}$$

where the deterministic component $s(t, \theta)$ is the true signal model depending on parameters θ and $\epsilon(t)$ is a zero-mean stationary time series with unknown spectral density [23].

When the data are in (discrete) frequency domain for an unknown noise spectral density the likelihood function is defined as

$$p(\tilde{y}|\theta) = K \times \exp\left[-\sum_{j=0}^{\nu} \left(\log(S(f_j)) + \frac{|\tilde{y}(f_j) - \tilde{s}(f_j, \theta)|^2}{S(f_j)} \right) \right], \tag{9}$$

where $\nu = \lfloor (N-1)/2 \rfloor$ is the greatest integer less than or equal to $(N-1)/2$, $\tilde{y}(f_j)$ and $\tilde{s}(f_j, \theta)$ are Fourier transformed observables and model signal, respectively, $S(f_j)$ is the one-sided power spectral density and K is the normalizing constant.

When the noise spectrum is assumed known then one can omit the constant term, $\log(S_n(f_j))$, from equation (9) to obtain

$$p(\tilde{y}|\theta) = K \times \exp\left[-\sum_{j=0}^{\nu} \frac{|\tilde{y}(f_j) - \tilde{s}(f_j, \theta)|^2}{S(f_j)} \right]. \tag{10}$$

In the literature, equation (9) is known as Whittle's approximation to the Gaussian likelihood or simply Whittle likelihood [31]. The Whittle likelihood assumes that the DFTed residuals are approximately independent complex normally distributed with mean zero and power spectrum $S(f_j)$. The complete description of the Bayesian estimation of the noise spectral density is explained in [23]. The likelihood computation is commonly simplified by restricting the summation to the limited frequency range relevant to the signals of interest. Furthermore, as mentioned in section 2, the two TDI observables A and E are independent, their joint likelihood is just the product of their individual likelihoods.

4. Implementation

Although LFA is much more time efficient compared to full LISA response, EMRI waveforms are still computationally expensive and hence we had to use only shorter data segments ranging in length from one to four weeks. We used heavy-tailed Student-t proposals which are very useful for good mixing. The parallel computation on multiple processors or cores is accomplished by including the message passing interface library [32] with the relevant additional programming scripts in the main analysis code. In all of our searches we used eight or ten chain PTMCMC.
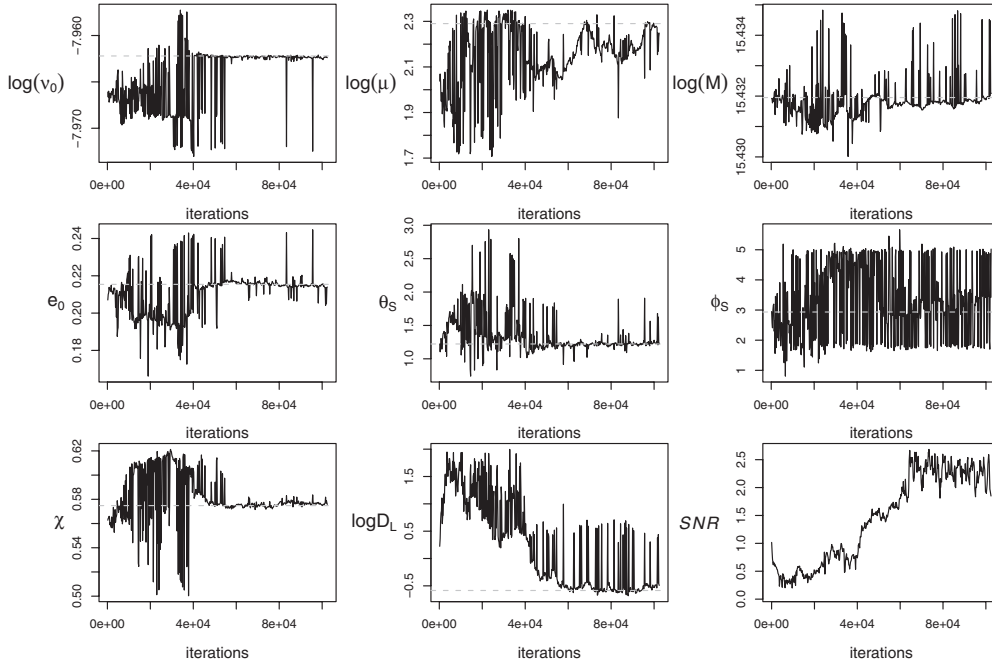
Figure 1. A typical (initial) shorter MCMC run. The trace plots of marginal posterior MCMC samples for some key parameters for the EMRI training source MLDC 1B.3.2. The grey dashed lines indicate the true values.

4.1. Single EMRI sources: example searches

In the simplest scenario the algorithm was applied to two EMRI data sets taken from earlier MLDC rounds; 1B.3.2 and 1C.3.1. 1B.3.2 contains a single medium mass EMRI source and 1C.3.1 contains a high mass EMRI source buried in the LISA instrument noise only. In these attempts the adjusted LFA was used and there was no problem in recovering the signal parameters using the true values as the starting points for the MCMC search, thus we used completely blind searches on two weeks long stretches of the data, to test the performance of our algorithm. In both cases the blind searches were conducted in multiple stages, i.e. first several shorter MCMC chains were run in parallel without a swapping step from random starting points and those chains were chosen that showed stability and for which the SNR and likelihood values were larger than others. The modes of those chains were then set as the starting points in the next MCMC run. These steps were repeated for a few times to narrow down the search range. This approach is somewhat similar to that used in [9]. In the following, summaries of the posterior distribution of the parameters of the EMRI sources given in MLDC 1B.3.2 (medium mass) and MLDC 1C.3.1 (high mass) are given. Figure 1 display a typical shorter MCMC run demonstrating how the chains for different parameters find their true modes. Table 1 and figure 2 presents the posterior summary of the parameters for source 1B.3.2. Similarly, table 2 and figure 3 show the results for source 1C.3.1. From these results it is clear that both signals were recovered with a great accuracy. The widths of different marginal posterior densities show that almost the whole prior range was searched before convergence.
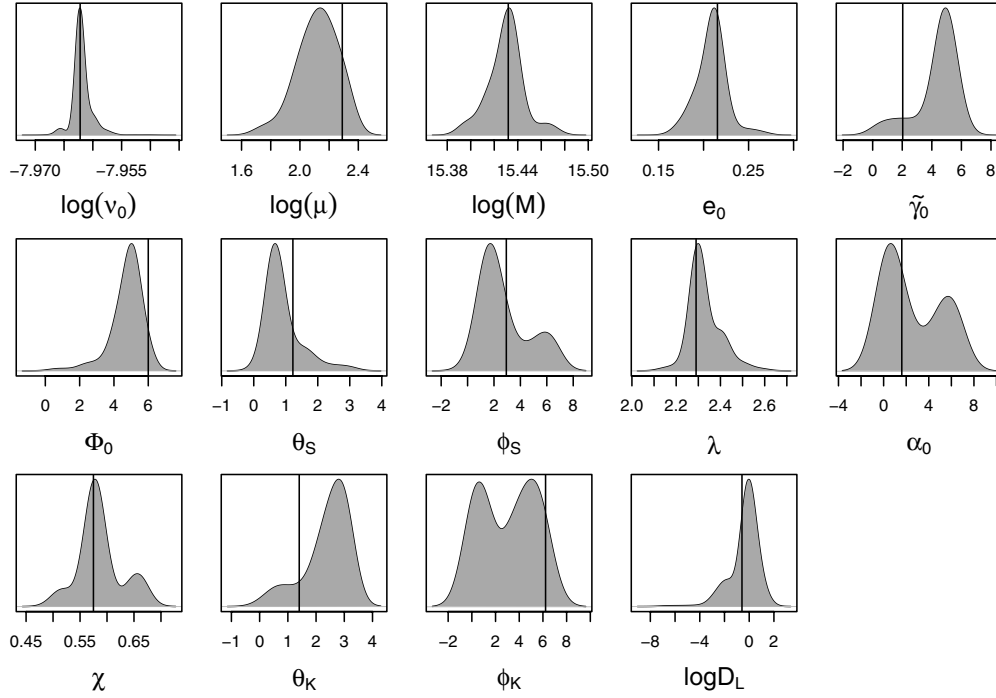
Figure 2. Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI source MLDC 1B.3.2. The solid lines indicate the true values.

Table 1. Posterior results and true parameter values for the recovered medium mass EMRI signal given in actual MLDC 1B.3.2 training data set.

Parameters	Mean	Std. dev.	Mode	95% BCI	True values	Prior range
$\log(\nu_0)$	−7.961 883	0.001 556	−7.962 237	(−7.962 645, −7.959 406)	−7.962 205	$\log[5.0 \times 10^{-5}, 0.01]$Hz
$\log(\mu)$	2.108 143	0.137 913	2.122 627	(1.853 307, 2.323 680)	2.290 258	$\log[9.5, 10.5]M_\odot$
$\log(M)$	15.429 324	0.014 565	15.432 732	(15.401 371, 15.460 317)	15.431 952	$\log[4.75, 5.25]\,10^6 M_\odot$
e_0	0.208 397	0.015 804	0.210 424	(0.183 391, 0.229 404)	0.215 401	[0.10, 0.40] units
$\tilde{\gamma}_0$	4.384 615	1.229 447	4.845 658	(1.232 442, 5.543 105)	2.033 2973	$[0, 2\pi]$ rad
Φ_0	4.682 341	0.695 824	5.011 981	(3.201 36, 5.340 058)	5.999 900	$[0, 2\pi]$ rad
θ_S	0.862 168	0.531 047	0.649 613	(0.492 233, 2.114 493)	1.222 330	$[0, \pi]$ rad
ϕ_S	2.710 713	1.905 070	1.580 461	(1.092 997, 6.237 064)	2.934 625	$[0, 2\pi]$ rad
λ	2.331 862	0.068 092	2.302 875	(2.251 945, 2.463 038)	2.289 951	$[0, \pi]$ rad
α_0	2.884 914	2.493 591	0.649 972	(0.087 803, 6.210 357)	1.609 215	$[0, 2\pi]$ rad
χ	0.589 667	0.039 206	0.578 192	(0.515 752, 0.659 08)	0.574 818	$[0.5, 0.7]\,M^2$
θ_K	2.478 239	0.773 029	2.870 962	(0.597 917, 3.096 335)	1.403 416	$[0, \pi]$ rad
ϕ_K	3.373 622	2.155 152	5.050 124	(0.223 237, 6.042 101)	6.223 129	$[0, 2\pi]$ rad
$\log(D_L)$	−0.538 350	1.128 170	−0.135 892	(−2.721 751, 0.832 533)	−0.584 778	$\log[0, \infty]\,\log$(GPC)

4.2. Multiple EMRIs

4.2.1. MLDC 4 results. The approach was applied to detect signals generated by EMRI sources given in both training and blind data sets issued in the revised MLDC round 4. Looking at the amount of noise in these data, we attempted to recover signals from high mass EMRI systems only. Moreover, there were no medium mass sources in the training data set. Some preliminary results were presented at GWPAW (January 26–29, 2011, Milwaukee, WI,
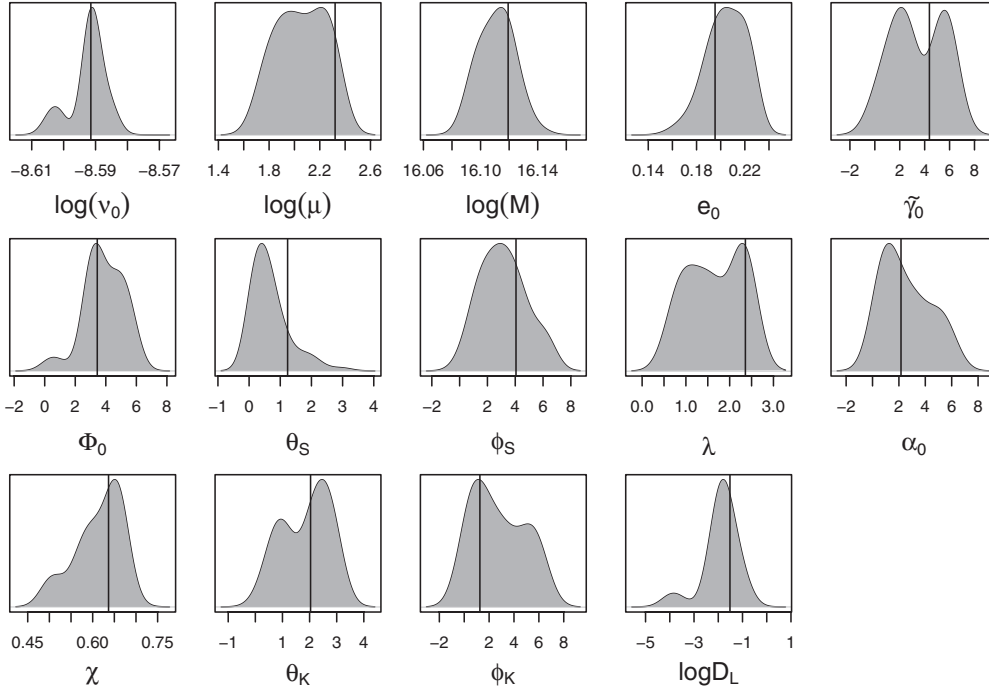
Figure 3. Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI source MLDC 1C.3.1 (high mass source). The solid lines indicate the true values.

Table 2. Posterior results and true parameter values for the recovered high mass EMRI signal given in actual MLDC 1C.3.1 training data set.

Parameters	Mean	Std. dev.	Mode	95% BCI	True values	Prior range
$\log(\nu_0)$	−8.592 307	0.005 407	−8.591 044	(−8.604 332, −8.584 98)	−8.591 472	$\log[5.0 \times 10^{-5}, 0.01]$Hz
$\log(\mu)$	2.026 831	0.185 767	2.209 483	(1.741 593, 2.322 630)	2.320 877	$\log[9.5, 10.5]M_\odot$
$\log(M)$	16.110 918	0.001 331	16.114 347	(16.090 741, 16.132 173)	16.119 304	$\log[0.95, 1.05]\,10^7 M_\odot$
e_0	0.205 405	0.014 601	0.204 826	(0.181 549, 0.226 693)	0.195 337	[0.10, 0.40] units
$\tilde{\gamma}_0$	3.425 854	2.011 777	2.116 807	(0.189 037, 6.143 949)	4.381 526	$[0, 2\pi]$ rad
Φ_0	3.948 827	1.273 087	3.366 273	(0.966 292, 5.721 986)	3.441 184	$[0, 2\pi]$ rad
θ_S	0.616 343	0.537 823	0.406 881	(0.063 561, 1.823 876)	1.235 677	$[0, \pi]$ rad
ϕ_S	2.944 769	1.580 446	2.915 200	(0.817 490, 6.224 024)	4.054 785	$[0, 2\pi]$ rad
λ	1.652 216	0.605 062	2.291 762	(0.719 750, 2.352 924)	2.358 963	$[0, \pi]$ rad
α_0	2.610 697	1.849 872	1.261 552	(0.361 536, 5.797 795)	2.158 356	$[0, 2\pi]$ rad
χ	0.612 676	0.051 827	0.650 991	(0.502 682, 0.662 555)	0.636 644	$[0.5, 0.7]$ M^2
θ_K	1.887 538	0.871 457	2.464 749	(0.472 588, 2.974 532)	2.036 360	$[0, \pi]$ rad
ϕ_K	2.827 086	2.093 720	1.124 890	(0.152 617, 6.081 997)	1.260 128	$[0, 2\pi]$ rad
$\log(D_L)$	−1.943 526	0.763 724	−1.795 853	(−3.843 257, −0.939 327)	−1.518 042	$\log[0, \infty]$ log(GPC)

USA). The MLDC notation for a high mass EMRI sources are EMRI-1-0, EMRI-1-1 and so on, and for a medium mass EMRI sources are EMRI-2-0, EMRI-2-1 and so on, whereas we denote the corresponding estimated sources (in the blind data) simply by High-0, High-1 and Med-0.

4.2.2. Training data. The training data contains three high mass EMRI sources which are somewhat similar to each other, therefore a joint MCMC search was conducted to recover
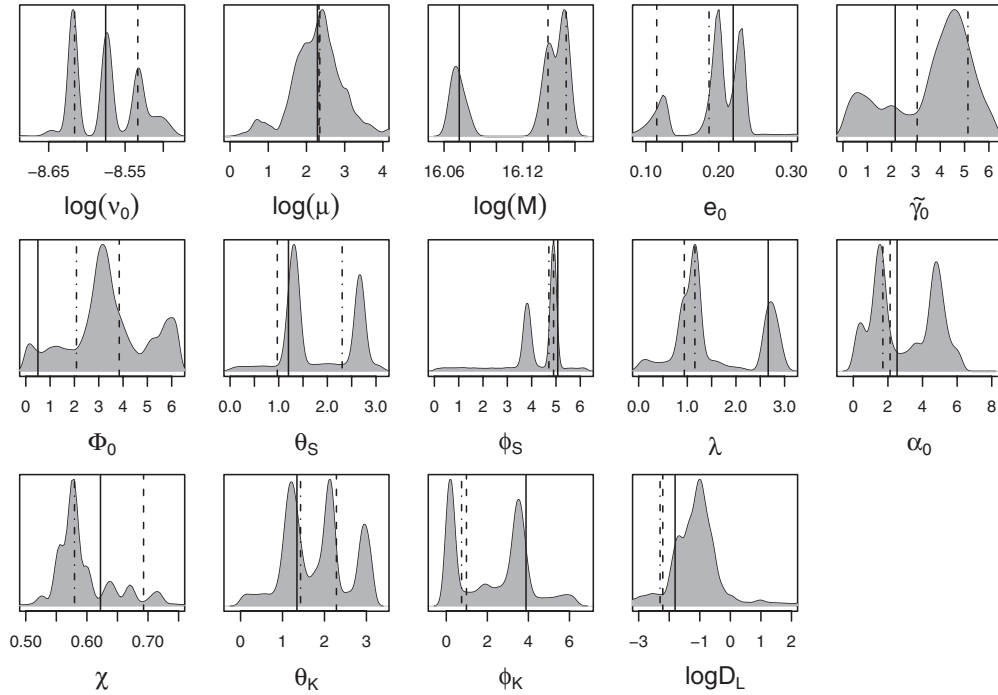
Figure 4. Kernel density estimates of the marginal posterior densities for all 14 parameters of the first three EMRI sources in MLDC 4 training set. In these plots the solid lines indicate the true parameter values of the source EMRI-1-0, whereas the dashed and dashed-dot-dashed lines indicate the true parameter values for sources EMRI-1-1 and EMRI-1-2, respectively, in MLDC 4 training data.

them. In an eight chain MCMC search on two weeks long data segments, three chains were started from the true parameter values of the three signals while the rest of the chains were started from the values in the vicinity of the true parameters. Figure 4 displays the results of this joint search. In the plots of kernel density estimates, different types of vertical lines denote the true values of the parameters of the three different high mass EMRI signals, named EMRI-1-0, EMRI-1-1 and EMRI-1-2 in MLDC 4 training keys, that can be found at the challenge webpage [33]. After running for a sufficiently large number of iterations ($\sim 4 \times 10^6$) it was observed that the third signal (dashed-dot-dashed vertical lines) was dominating the other two as can be seen in figure 4, even though the overall mean swap acceptance rate between chains was $\sim 35\%$. Thus the code was restarted with the starting values of different chains somewhat similar to the true parameter values of the third signal. These results are given in table 3 and figure 5. We can see that all the parameters, except the luminosity distance, D_L, and some of the angles, are estimated with great accuracy and most of the parameters' chains show stability. From the MCMC trace plots the distance parameter seemed to be over estimated and the sky location (θ_S, ϕ_S) seemed to have been locked at a different position. The wrong sky location was attributed to the fact that we were using LFA and are experiencing the same problem as was observed in [10], which also used the LFA [15]. At this stage, the problem of amplitude differences between the full LISA response and LFA was not known to us.

It was also evident that after some 800 000 iterations one of the neighbouring chains of the true ($T = 1$) chain found some other mode; however, overall the true chain was unaffected.
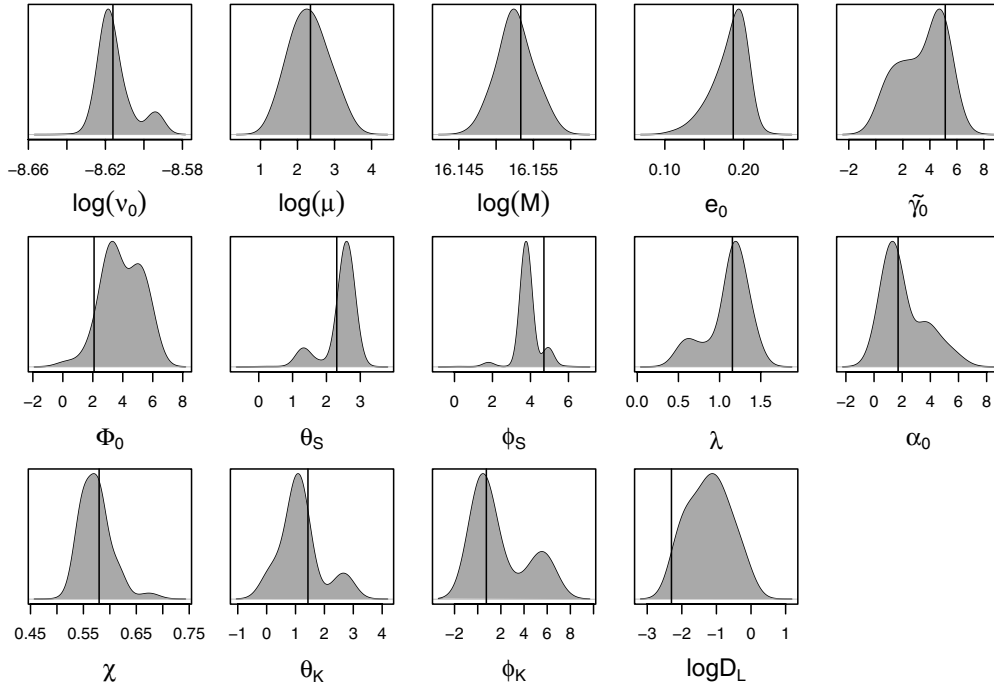
Figure 5. Kernel density estimates of the marginal posterior densities for all 14 parameters of the third EMRI (EMRI-1-2) signal in MLDC 4 training data. The solid lines indicate the true parameter values.

Table 3. Posterior results for the third EMRI (EMRI-1-2) signal given in MLDC 4 training data set.

Parameters	Mean	Std. dev.	Mode	95% BCI	True values	Prior range
$\log(\nu_0)$	−8.615 057	0.008 435	−8.618 036	(−8.624 964, −8.593 739)	−8.616 096	$\log[5.0 \times 10^{-5}, 0.01]$Hz
$\log(\mu)$	2.247 796	0.519 295	2.185 567	(1.433 541, 3.218 01)	2.350 281	$\log[9.5, 10.5]M_\odot$
$\log(M)$	16.152 486	0.002 285	16.152 319	(16.148 756, 16.156 314)	16.153 307	$\log[0.95, 1.05]\ 10^7 M_\odot$
e_0	0.184 634	0.019 464	0.195 090	(0.145 334, 0.203 263)	0.186 712	[0.10, 0.40] units
$\tilde{\gamma}_0$	3.830 427	1.407 267	4.734 815	(1.089 336, 5.408 43)	5.138 873	$[0, 2\pi]$ rad
Φ_0	3.770 239	1.079 385	3.244 351	(2.128 266, 5.591 350)	2.084 779	$[0, 2\pi]$ rad
θ_S	2.455 871	0.434 043	2.611 267	(1.311 937, 2.755 807)	2.305 033	$[0, \pi]$ rad
ϕ_S	3.901 775	0.485 260	3.780 766	(3.583 514, 4.948 959)	4.707 928	$[0, 2\pi]$ rad
λ	1.115 781	0.224 159	1.180 035	(0.572 802, 1.395 177)	1.155 677	$[0, \pi]$ rad
α_0	1.965 982	1.351 007	1.372 157	(0.521 520, 5.008 045)	1.708 861	$[0, 2\pi]$ rad
χ	0.573 573	0.024 460	0.576 902	(0.540 204, 0.616 836)	0.579 723	$[0.5, 0.7]\ M^2$
θ_K	1.235 264	0.685 402	1.127 951	(0.096 748, 2.711 275)	1.429 748	$[0, \pi]$ rad
ϕ_K	1.719 245	2.280 232	0.380 733	(0.088 012, 5.899 353)	0.745 268	$[0, 2\pi]$ rad
$\log(D_L)$	−1.262 741	0.596 673	−1.216 699	(−2.187 479, −0.196 473)	−2.299 291	$\log[0, \infty]\ \log($GPC$)$

This second mode could either correspond to, most probably, a low strength harmonic of this same EMRI signal as such a close harmonic usually corresponds to higher frequencies than the true one, or to a harmonic of another EMRI signal. This sort of overlapping and sharing of characteristics between different signals will be quite common in such complicated cases and will result in confusions among different EMRI signals.
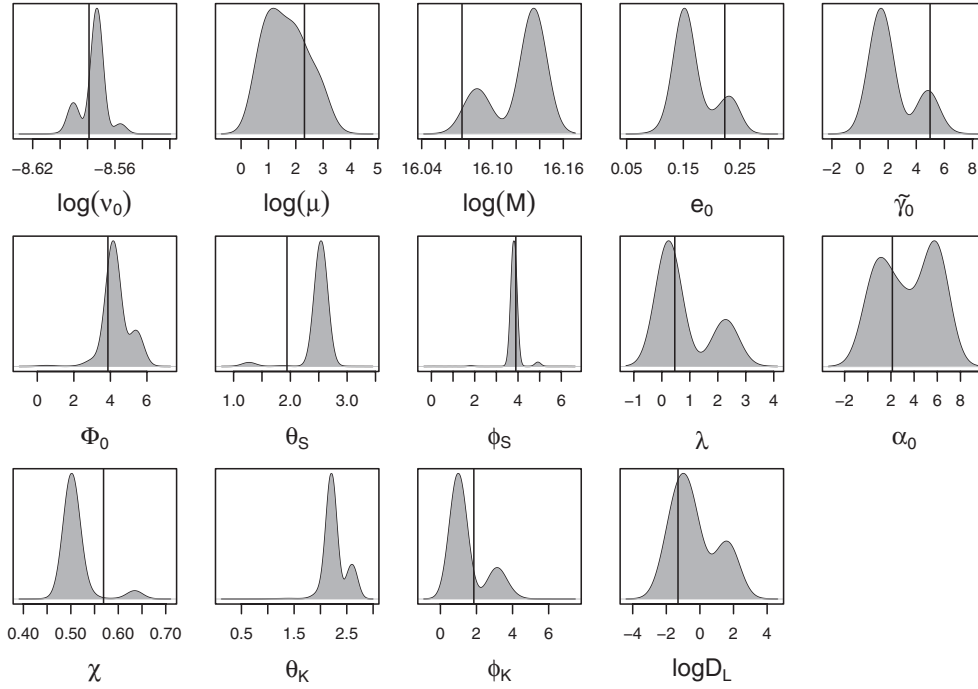
Figure 6. Kernel density estimates of the marginal posterior densities for all 14 parameters for the detected EMRI source in MLDC 4 blind data. The solid lines indicate the true solutions of high mass source EMRI-1-1 in MLDC 4 blind data.

Table 4. Posterior results for the detected high mass EMRI signal given in MLDC 4 blind data set in the first step.

Parameters	Mean	Std. dev.	Mode	95% BCI	Prior range
$\log(\nu_0)$	−8.573 457	0.003 015	−8.573 259	(−8.574 035, −8.572 342)	$\log[5.0 \times 10^{-5}, 0.01]$Hz
$\log(\mu)$	1.409 599	0.623 259	1.179 280	(0.589 129, 2.263 096)	$\log[9.5, 10.5]M_\odot$
$\log(M)$	16.133 081	0.008 141	16.134 940	(16.132 126, 16.136 896)	$\log[0.95, 1.05] \, 10^7 M_\odot$
e_0	0.154 880	0.014 667	0.152 115	(0.139 163, 0.166 299)	[0.10, 0.40] units
$\tilde{\gamma}_0$	1.584 359	0.708 484	1.486 492	(0.859 918, 2.269 012)	$[0, 2\pi]$ rad
Φ_0	4.197 498	0.336 251	4.147 553	(3.834 165, 4.592 633)	$[0, 2\pi]$ rad
θ_S	2.534 345	0.098 177	2.540 198	(2.507 845, 2.579 459)	$[0, \pi]$ rad
ϕ_S	3.822 423	0.097 805	3.809 877	(3.744 967, 3.886 976)	$[0, 2\pi]$ rad
λ	0.298 106	0.365 414	0.249 978	(0.150 734, 0.347 221)	$[0, \pi]$ rad
α_0	3.072 884	2.491 800	5.761 360	(0.150 223, 6.189 003)	$[0, 2\pi]$ rad
χ	0.499 902	0.009 253	0.501 648	(0.488 186, 0.508 558)	$[0.5, 0.7]$ M^2
θ_K	2.218 992	0.075 129	2.207 951	(2.161 368, 2.274 100)	$[0, \pi]$ rad
ϕ_K	1.046 394	0.370 040	0.986 633	(0.870 475, 1.121 522)	$[0, 2\pi]$ rad
$\log(D_L)$	−0.928 824	0.803 5100	−0.986 087	(−1.974 076, −0.018 514)	$\log[0, \infty] \log(\text{GPC})$

4.2.3. Blind data. For the MLDC 4 blind data as a first step an eight chain MCMC search was conducted on the first two weeks in which all the eight chains were started from random values corresponding to a high mass EMRI source. Table 4 shows a summary of the posterior estimates and figure 6 shows the kernel density estimates of the marginal posterior densities. The MCMC chains for all parameters showed great stability except for the CO's mass μ and distance $\log D_L$ the chains were showing somewhat oscillatory and correlated behaviour.
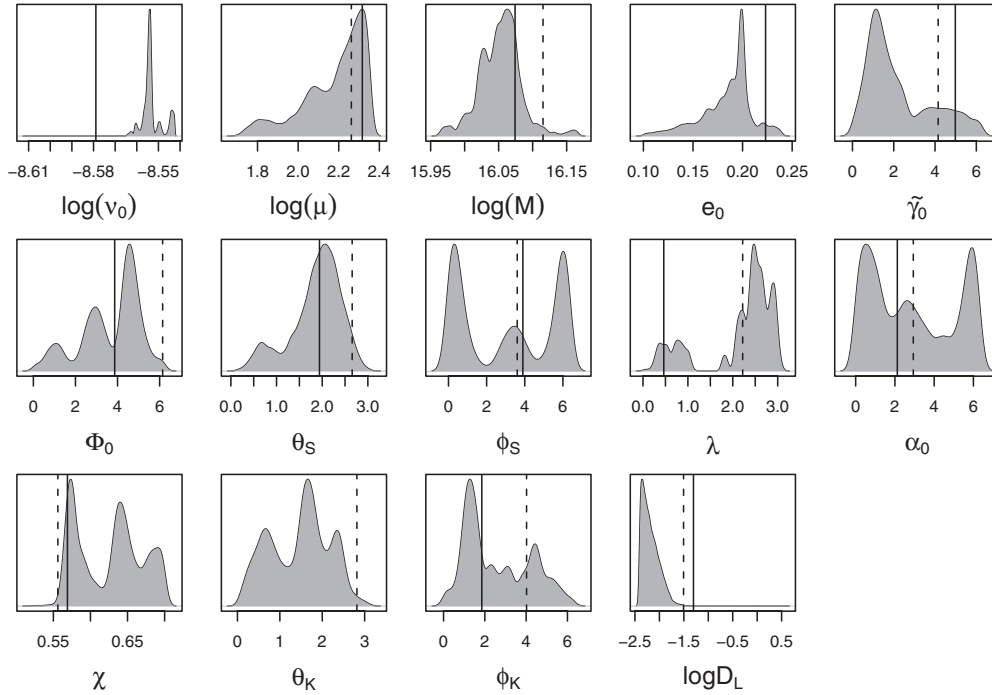
Figure 7. Kernel density estimates of the marginal posterior densities for all 14 parameters for the high mass EMRI source 'High-3' in MLDC 4 blind data. The dashed and solid lines corresponds to the true solutions of two high mass sources EMRI-1-0 and EMRI-1-1, respectively, in MLDC 4 blind data.

The empirical correlation between these two parameters was ∼0.92, which is quite high. It could not be deciphered why this (high correlation) happened as such a phenomenon was found neither in earlier nor in subsequent searches. The chains for all the angles were also stable except for the parameter α_0, which was vibrating between two different modes. An interesting result which was observed in our earlier searches, in which the LFA response was not adjusted (dividing by 3), is that for all time regions the joint plots of the sky location angles indicated a similar behaviour, though in the plots of the kernel density estimates of these two angles obtained for different time regions the (strongest) modes were different. For different time regions the joint plots of the two sky location angles are shown in figure 8. These sky locations remained the same despite using different proposal distributions. At this stage we realized that amplitudes of LFA response are in general by a factor of ∼3 larger than full LISA response amplitudes. As an ad hoc solution to this problem, due to time limitation, the LFA amplitudes were divided by 3, which, in single source searches such as MLDC 1B (as shown in section 4.1), resulted not only in the correct estimation of distance and sky location parameters but also, in MLDC 4 searches, the sky location angles never got locked on wrong positions.

In order to search the entire two years of the blind data several more MCMC searches were conducted on different time regions (data segments). In these searches we used a week long data segments chosen at regular gaps (generally six weeks) covering the entire two years duration of the data or till the anticipated plunge time of a typical EMRI signal (for most of the EMRI sources given in earlier rounds of MLDC the plunge generally occurs during $[1\frac{1}{2} - 1\frac{3}{4}]$ years, although in some cases the plunges are occurred in the last month of the two years
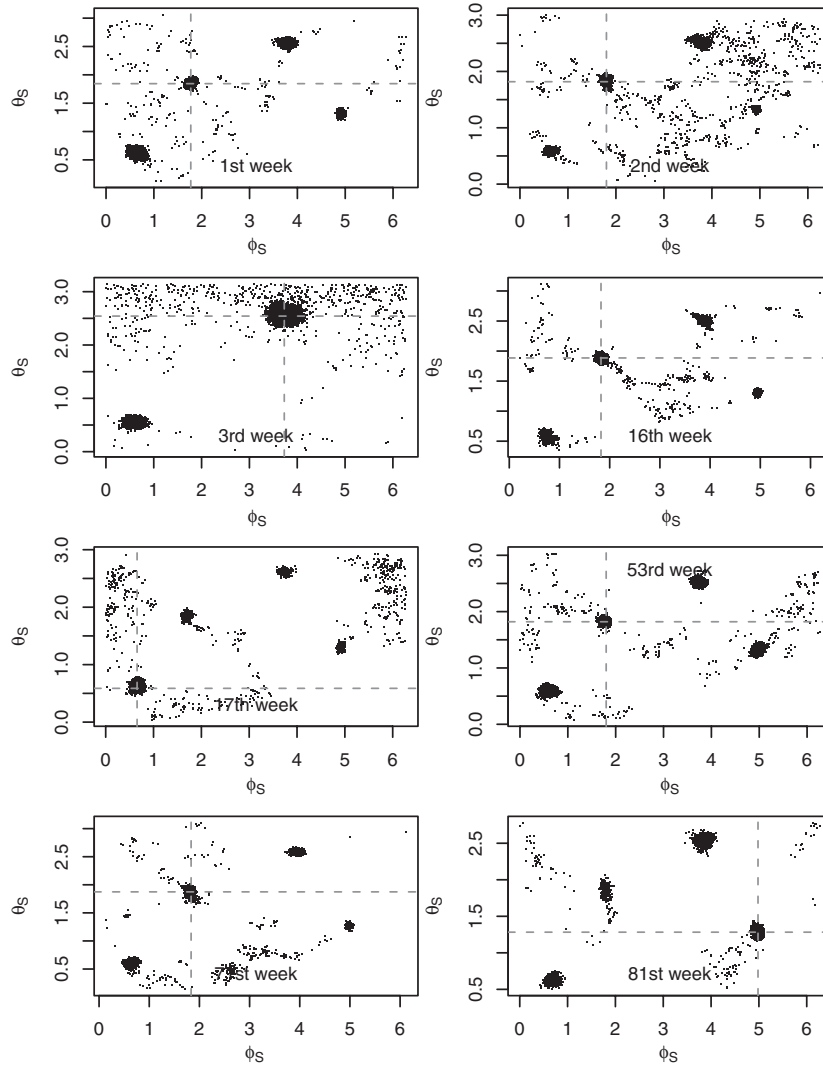
Figure 8. The joint plots of sky location angles for different time regions demonstrate that there are four most probable sky positions either of the same source or there are two or more EMRI sources located in different sky regions. The dashed lines indicate the *maximum a posteriori* values.

duration). Each time region was searched individually by running eight or ten chain MCMCs. The search strategy was the same as explained in section 4.1. Attempts were made to recover both types of EMRI signals (high mass and medium mass) in these searches. In the end, for each type of EMRI source four possible best fits were chosen on the basis of high SNRs, to be submitted to MLDC 4. Among the submitted entries, the posterior results for a medium mass source 'Med-2' and a high mass source 'High-3' are presented in tables 5 and 6 and figure 7. The rest of our entries and solutions can be found on MLDC 4 webpage [33]. The kernel density estimates of the marginal posterior densities for most of the parameters show multi-modality. However, in most cases the strongest modes can be clearly recognized. In all these searches the average rates of the regular Metropolis acceptance were in 20–40% while the average swap acceptance rates were in 16%–30%. The recovered medium EMRI source

Table 5. Posterior results for the detected EMRI signal 'Med-2' given in MLDC 4 blind data set.

Source Parameters	Med-2 Mean	SNR-A = 62.921 Mode	Std. dev.	SNR-E = 70.950 Mean ± 2StdDev	SNR-AE = 103.708 Prior range
ν_0	0.000 341 1537	0.000 340 9980	9.80e−7	(0.000 339 0515, 0.000 342 9558)	[5.0×10^{-5}, 0.01] Hz
μ	13.036 900 58	10.505 164 49	3.21e+0	(6.603943913, 16.710 996 09)	[9.5, 10.5]$M_\odot$
M	4840 967.8525	4838 561.9866	6.93e+4	(4706 125.2326, 4974 725.6905)	[0.95, 1.05] $10^7 M_\odot$
e_0	0.218 266	0.217 780	2.23e−2	(0.173 1910, 0.262 3679)	[0.10, 0.40] units
$\tilde{\gamma}_0$	4.691 678	5.305 615	1.23e+0	(2.840 7054, 7.770 524)	[0, 2π] rad
Φ_0	5.176 204	4.961 252	8.86e−1	(3.189 7754, 6.732 728)	[0, 2π] rad
θ_S	1.382 196	1.408 894	2.88e−1	(0.833 3279, 1.984 459)	[0, 2π] rad
ϕ_S	4.412 142	4.646 086	9.16e−1	(2.814 0819, 6.478 091)	[0, 2π] rad
λ	0.694 621	0.681 557	2.78e−1	(0.126 5363, 1.236 579)	[0, 2π] rad
α_0	2.624 212	0.261 676	2.68e+0	(−5.101 531, 5.624 883)	[0, 2π] rad
χ	0.581 468	0.578 720	2.12e−2	(0.536 2277, 0.621 211)	[0.5, 0.7] M^2
θ_K	1.588 288	1.562 271	3.12e−1	(0.938 9395, 2.185 603)	[0, 2π] rad
ϕ_K	2.572 490	2.615 519	5.01e−1	(1.612 6596, 3.618 377)	[0, 2π] rad
D_L(GPC)	0.079 2323	0.090 8401	3.62e−2	(0.016 161 34, 0.510 5964)	[0, ∞] GPC

Table 6. Posterior results for the detected EMRI signal 'High-3' given in MLDC 4 blind data set.

Source Parameters	High-3 Mean	SNR-A = 19.777 Mode	Std. dev.	SNR-E = 17.406 Mean ± 2SthDev	SNR-AE = 26.346 Prior range
ν_0	0.000 192 9449	0.000 192 7831	8.81e−7	(0.000 191 0318, 0.000 194 5503)	[5.0×10^{-5}, 0.01] Hz
μ	8.816 421 59	10.126 919 09	1.24e+0	(7.479 619 72, 13.711 190 41)	[9.5, 10.5]$M_\odot$
M	9362 002.4476	9462 096.5945	2.86e+5	(8901 852.0130, 10 057 600.5800)	[0.95, 1.05] $10^7 M_\odot$
e_0	0.184 721	0.198 859	2.89e−2	(0.148 464, 0.249 254)	[0.10, 0.40] units
$\tilde{\gamma}_0$	2.247 122	1.133 451	1.59e+0	(−2.049 031, 4.315 933)	[0, 2π] rad
Φ_0	3.681 986	4.570 893	1.38e+0	(1.791 032, 7.350 754)	[0, 2π] rad
θ_S	1.864 284	2.062 0089	5.55e−1	(0.952 001, 3.172 016)	[0, π] rad
ϕ_S	3.088 018	0.309 6088	2.44e+0	(−4.581 664, 5.200 881 77)	[0, 2π] rad
λ	2.147 808	2.472 5720	8.12e−1	(0.848 402, 4.096 742 41)	[0, π] rad
α_0	2.908 103	0.541 1894	2.13e+0	(−3.719 176, 4.801 554 57)	[0, 2π] rad
χ	0.626 321	0.573 2350	4.31e−2	(0.486 933, 0.659 537 07)	[0.5, 0.7]
θ_K	1.466 041	1.656 5712	6.97e−1	(0.262 243, 3.050 899 78)	[0, π] rad
ϕ_K	2.621 246	1.278 5729	1.58e+0	(−1.894 009, 4.451 155)	[0, 2π] rad
D_L (GPC)	0.112 901 98	0.094 923 33	3.15e−2	(0.066 4814, 0.135 5331)	[0, ∞] GPC

'Med-2', had a very high SNR but the estimated parameters are somewhat away from the true ones, demonstrating that these are probably the secondaries of the true densities. The 'High-3' has a lower SNR but the estimated parameters are quite close to the true ones.

4.3. Overlaps and SNRs

4.3.1. True signals: LFA versus full LISA response. When the true solutions for the MLDC 4 blind EMRIs were released, we tried to assess the performance of LFA and, in particular, the performance of the full AKW and TAKW in LFA regime by comparing the overlaps of their $A (\equiv h_I)$, $E (\equiv h_{II})$ channels with those obtained with full LISA response. For this purpose all the true signals (noiseless) were generated using both lisatools package and our model codes. To calculate the overlap the following formula was used

$$O = \frac{\langle a, b \rangle}{\sqrt{\langle a, a \rangle \langle b, b \rangle}}, \tag{11}$$

where $\langle a, b \rangle = \int_{-\infty}^{\infty} \frac{a(f)b^*(f)}{S(f)} \, df$, the noise weighted inner product of two functions a, b. All these quantities are in frequency domain. The lisatools package uses the full AKW model to generate the polarization signals $(h_+, h_\times)$ of EMRIs and then employs the full LISA response

Table 7. True signals: overlaps of the AKW and TAKW in LFA against AKW in full LISA response.

	AKW+LFA versus AKW+LISA			TAKW+LFA versus AKW+LISA		
Signals	EMRI-1-0	EMRI-1-1	EMRI-2-0	EMRI-1-0	EMRI-1-1	EMRI-2-0
Overlap A	0.238	0.750	0.893	0.978	0.886	0.986
Overlap E	0.191	0.844	0.913	0.952	0.874	0.977
Combined overlap	0.214	0.795	0.903	0.965	0.880	0.981

Table 8. Overlaps and SNRs of the detected signals for different sources.

		Med-0*	Med-1	Med-2*	Med-3*
EMRI-2-0*	Overlap A	0.501	0.102	0.219	0.185
	Overlap E	0.664	0.161	0.134	0.115
	Combined overlap	0.577	0.128	0.171	0.146
	SNR A	38.133	1.647	65.307	62.921
	SNR E	42.003	1.190	80.562	70.950
	Combined SNR	56.731	2.032	103.708	94.831
		High-0	High-1$^\blacklozenge$	High-2	High-3$^\clubsuit$
EMRI-1-0$^\clubsuit$	Overlap A	0.021	0.114	0.017	0.055
	Overlap E	0.018	0.084	0.041	0.125
	Combined overlap	0.019	0.098	0.027	0.083
EMRI-1-1$^\blacklozenge$	Overlap A	0.018	0.199	0.018	0.031
	Overlap E	0.007	0.286	0.024	0.057
	Combined overlap	0.012	0.238	0.021	0.042
	SNR A	11.173	11.262	2.688	19.777
	SNR E	15.589	10.227	3.303	17.406
	Combined SNR	19.180	15.213	4.258	26.346

function to generate TDI variables X, Y and Z, that can then be transformed to A, E channels. Similarly, our codes can generate h_+ and $h_\times$ polarizations using AKW or TAKW and then compute h_I, h_{II} using LFA. The overlap statistics are given in tables 7 and 8. These results demonstrate that in the LFA regime the TAKW model gives better results than the full AKW. Therefore, with the LFA, it makes sense to use TAKW model for estimating the EMRI signals buried in LISA data.

4.3.2. True versus detected signals. Overlaps of the estimated signals (TAKW + LFA) with true signals (AKW + full LISA response) and their SNRs are shown in table 8. Different symbols indicate the correspondence between the best matches of estimated signals with true signals. The estimated medium mass signal 'Med-2' is believed to have a best match with the only medium mass source in MLDC 4 blind data. The low degree of overlaps and SNRs of the estimated high mass signals is due to the fact that for these sources the difference factor between the LFA and full LISA response is ∼6 rather than ∼3.

5. Conclusion

In this paper we report on the application of a Bayesian approach to statistical inference on the highly challenging problem of detecting and estimating parameters of GW signals produced by EMRI's in the data from a future space-based interferometric mission such as LISA. The LFA is used to model the LISA response and the Whittle likelihood as a realistic approximation to

the unknown error distribution. Differences in the amplitudes of LFA and the LISA response function were found and temporarily fixed. Further investigations are required to overcome these discrepancies. We assume the noise power spectrum to be unknown and formulate an Inv-χ^2 posterior with a conjugate prior distribution for the unknown noise parameters. We have implemented a parallel tempering MCMC algorithm to sample from the posterior distribution of all 14 unknown parameters, the first instance of a fully Markovian algorithm for EMRI detection and characterization in the context of MLDC LISA data analysis challenges. This algorithm has the potential to sample from multi-modal distributions and thus to avoid getting trapped in local modes. Our simulation results using single as well as multiple EMRI sources in MLDC training and blind data should be seen as a snapshot of an investigation still in progress. However, they show that this strategy holds great promise for posterior computation under more realistic conditions and hints at areas where the algorithm could be refined and tuned.

Acknowledgments

References

[1] Danzmann K and Rüdiger A 2003 LISA technology—concept, status, prospects *Class. Quantum Grav.* **20** S1–9
[2] Everitt C W F *et al* 2011 Gravity probe B: final results of a space experiment to test general relativity *Phys. Rev. Lett.* **106** 221101
[3] Vecchio A 2004 LISA observations of rapidly spinning massive black hole binary systems *Phys. Rev. D* **70** 042001
[4] Christensen N and Meyer R 1998 Markov chain Monte Carlo methods for Bayesian gravitational radiation data analysis *Phys. Rev. D* **58** 082001
[5] Christensen N and Meyer R 2001 Using Markov chain Monte Carlo methods for estimating parameters with gravitational radiation data *Phys. Rev. D* **64** 022001
[6] Röver C *et al* 2007 Inference on inspiral signals using LISA MLDC data *Class. Quantum Grav.* **24** S521
[7] Röver C 2007 Bayesian inference on astrophysical binary inspirals based on gravitational wave measurements *PhD Thesis* The University of Auckland
[8] Stroeer A, Gair J and Vecchio A 2006 Automatic Bayesian inference for LISA data analysis strategies *AIP Conf. Proc.* **873** 444–51
[9] Babak S, Gair J R and Porter E K 2009 An algorithm for the detection of extreme mass ratio inspirals in LISA data *Class. Quantum Grav.* **26** 135004
[10] Gair J R, Porter E K, Babak S and Barack L 2008 A constrained Metropolis–Hastings search for EMRIs in the Mock LISA Data Challenge 1B *Class. Quantum Grav.* **25** 184030
[11] Cornish N J 2008 Detection strategies for extreme mass ratio inspirals arXiv:0804.3323v1
[12] Geyer C J 1991 Markov chain Monte Carlo maximum likelihood *Computing Science and Statistics: Proc. 23rd Symp. Interface* pp 156–63
[13] Whittle P 1953 Estimation and information in stationary time series *Ark. Mat.* **2** 423–34
[14] Arnaud K A *et al* (The Mock LISA Data Challenge Task Force) 2007 An overview of the second round of the Mock LISA Data Challenges *Class. Quantum Grav.* **24** S551
[15] Barack L and Cutler C 2004 LISA capture sources: approximate waveforms, signal-to-noise ratios and parameter estimation accuracy *Phys. Rev. D* **69** 082005
[16] Vallisneri M 2005 Synthetic LISA: simulating time delay interferometry in a model LISA *Phys. Rev. D* **71** 022001
[17] Cornish N J and Rubbo L J 2003 LISA response function *Phys. Rev. D* **67** 022001
[18] Cornish N J, Rubbo L J and Poujade O The LISA simulator www.physics.montana.edu/LISA
[19] Cutler C 1998 Angular resolution of the LISA gravitational wave detector *Phys. Rev. D* **57** 7089–102

[20] Prince T A, Tinto M, Larson S L and Armstrong J W 2002 LISA optimal sensitivity *Phys. Rev.* D **66** 122002
[21] MLDC webpage: http://astrogravs.nasa.gov/docs/mldc
[22] Gair J R and Porter E K 2008 EMRI parameter recovery for training source 1.3.2. unpublished report
[23] Röver C, Meyer R and Christensen N 2011 Modelling coloured residual noise in gravitational-wave signal processing *Class. Quantum Grav.* **28** 015010
[24] Gelman A, Carlin J B, Stern H S and Rubin D B 2003 *Bayesian Data Analysis* 2nd edn (London: Chapman and Hall)
[25] Hastings W K 1970 Monte Carlo sampling methods using Markov chains and their applications *Biometrika* **57** 97–109
[26] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 Equation of state calculations by fast computing machines *J. Chem. Phys.* **21** 1087–92
[27] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671–80
[28] Marinari E and Parisi G 1992 Simulated tempering: a new Monte Carlo scheme *Europhys. Lett.* **19** 451
[29] Gilks W R, Richardson S and Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall)
[30] Liu J S 2001 *Monte Carlo Strategies in Scientific Computing* (New York: Springer)
[31] Whittle P 1957 Curve and periodogram smoothing *J. R.Stat. Soc. Ser. B* **19** 38–63
[32] Gropp W, Huss-Lederman S, Lumsdaine A, Lusk E, Nitzberg B, Saphir W and Snir M 1998 *MPI: The Complete Reference* 2nd edn (Cambridge, MA: MIT Press)
[33] MLDC 4 webpage:. http://www.tapir.caltech.edu/mldc/results4/index.html