Bayesian Inference on EMRI Signals in LISA Data

Asad Ali

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Statistics,
The University of Auckland, 2011.

Dedicated to my family.

ABSTRACT

This research activity is concerned with the applications of Bayesian Monte Carlo methods for the detection and parameter estimation of the gravitational wave signals generated by a special type of gravitational wave sources called extreme mass ratio inspirals (EMRIs). EMRIs are considered to be one of the most important potential sources of gravitational waves in space-time, to be observed with the planned Laser Interferometer Space Antenna (LISA). The data analysis of such sources is a challenging statistical and computational problem because of the large parameter space, weak amplitudes of the signals, complicated nature of the underlying waveform model and that of the LISA data. The posterior density surface of EMRI signals is full of local modes and ordinary Monte Carlo samplers usually fail to explore such multi-modal densities. This thesis tackles the problem of the detection and parameter estimation of such sources by establishing a Bayesian framework in which the posterior distribution is explored with the help of advanced Monte Carlo sampling methods such as parallel tempering Markov chain Monte Carlo. The LISA response to the incoming gravitational wave signals is not simple and requires some further manipulations to adjust the measured signals for different dynamics to which LISA will be exposed during its operation. This response is derived in two different ways, i.e. full LISA response and the low frequency approximation. This framework was applied with a great success in different scenarios ranging from the detection and estimation of parameters of a single EMRI source buried in LISA instrument noise to the detection and estimation of parameters of a particular EMRI source from data in which there are multiple EMRI sources contaminated with the instrument

noise as well as other gravitational sources of noise. The results show that our Bayesian methodology is indeed capable of facing the challenge of the detection and parameter estimation of the signals from EMRI sources in realistic LISA data.

ACKNOWLEDGEMENTS

All thanks and praises to Allah, the most gracious, most merciful. Without His gracious help, I firmly believe, this accomplishment would not have been possible.

My deep appreciation and gratitude go to my supervisors Dr. Renate Meyer and Dr. Nelson Christensen for their incredible and continuous encouragement and kind guidance in a friendly atmosphere with much more considerations for my weaknesses, which enabled me to complete this research work. I feel great satisfaction and sense of gratitude to express my sincere thanks to my senior colleague Dr. Christian Röver for his invaluable help throughout my research. I am grateful to Jonathan Gair for his sincere help in my research problem.

I want to express my special regards to Stephen Cope (compute servers) and Dr. Paul Murrell (R graphics) for their sincere and wonderful help. Special thanks go to Chew Chee for his many fruitful discussions at different times.

I am grateful to Higher Education Commission of Pakistan for providing funds for my studies. My thanks goes to all the people, whom I interacted with, of the Department of Statistics and The University of Auckland for providing me with a wonderful company during my studies here.

I am also thankful to Max Plank Institute for Gravitational Physics (Albert

Einstein Institute), Hannover, Germany for allowing me to use their supercomputer, ATLAS.

I extend my thanks to all my friends, whose names are not mentioned here, with whom I shared great times during the last four years.

Last but not least, I wish to express my deeply felt gratitude towards my parents and other family members for their constant inspiration and unceasing support over the years.

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Introduction

The application of Bayesian methods to make inferences about the characteristics of astrophysical phenomena has received a tremendous attention in recent decades, particularly after the invention of sophisticated Monte Carlo methods to overcome the mathematical challenges that generally arise in the applications of these methods. However, the interaction of the Bayesian approach and astrophysics is not new, rather it is a renewal of their forgotten relationship. After being introduced by Thomas Bayes (1702–1761), who provided the fundamental mathematical formalism in the form of the Bayes' rule, the Bayesian methods were originally applied and further developed by an astrophysicist, Pierre-Simon Laplace (1749–1827). Laplace used Bayesian probability theory to address many astrophysical problems, such as the estimation of masses of the planets from astronomical data, and to quantify the uncertainty in the measurement of the masses due to observational errors [2]. Logically, the Bayesian methods can be thought of as the only reliable tools for making inferences about the parameters associated with an astrophysical phenomenon as there is no way to directly access and observe the underlying objects and conduct experiments repeatedly in a laboratory like the other sciences. In such circumstances the inferences are exposed to *uncertainties* in the estimated quantities, which need to be quantified by a suitable probability

distribution.

In the past, the majority of scientists were reluctant to use Bayesian methods, partly because of the lack of a proper rationale for the assignment of prior probabilities, the famous debate of subjectivity versus objectivity in the definition of prior probability distribution, and partly because of the mathematically complicated structure of the resulting posterior; i.e. difficulties in evaluating the integrals to find the marginal posteriors or the expected values of the parameters and their functions. The first problem was addressed by Edwin Thompson Jaynes (1922–1998) with great clarity in [3]. One of his simple statements is, "the only thing objectivity requires of a scientific approach is that experimenters with the same state of knowledge reach the same conclusion" [4]. Thus the role of the experimenter, whose state of knowledge is to be quantified, has become a decisive factor in the Bayesian approach. Furthermore, he provided a logical way to establish a prior probability using the principle of maximum entropy, for example see [3–5] for more details. For applications of the Bayesian approach to the analysis of signals, the underlying theory has also been established very nicely in the above stated references. A similar Bayesian approach for the analysis of gravitational radiation data was proposed in [6], which was later clarified and extended in [7]. It is this later approach which is now widely used for the data analysis of gravitational waves.

The difficulties arising in the analytical evaluation of posterior distribution were overcome to a great extent by the availability of different numerical integration methods, particularly the Markov chain Monte Carlo (MCMC) methods and fast computing resources.

Earlier studies, e.g. [8–11] and many others, show that Bayesian methods equipped with suitable MCMC sampling algorithms have been successfully applied to estimate the parameters of gravitational wave signals

radiated by different sources. The focus of this thesis is the Bayesian inference on the gravitational wave signals generated by a special kind of sources called extreme mass ratio inspirals (EMRIs).

1.2 About this Work

This work was started in March 2007. The first one and half years were spent in gaining knowledge about the theory of the Bayesian statistics, Monte Carlo sampling methods, gravitational wave physics, signal processing techniques and Bayesian spectrum analysis. Along with these, knowledge about different programming languages and parallel programming techniques was also gained. Some time was spent to gain understanding of the EMRI waveform model and its parameters, and the mechanism of the laser interferometer space antenna (LISA) data simulator for generating the simulated data. At first there were a few problems in computing due to the large memory and high speed required for the efficient computations of the waveform signals as well as the corresponding detector response. In the beginning, because of memory problems a single chain MCMC sampler was used to simulate the posterior density and after memory issues were resolved the sampler was upgraded to a multiple chains sampler. Initially, most of the computations were carried out on a local server having eight processor cores. In September 2009, access to BeSTGRID (Auckland cluster) [12] was gained, which eased up the computational difficulties to a great extent. For the first time some real insight into the waveform model and the nature of its parameters was gained. Some initial results were informally discussed at 14th Gravitational Wave Data Analysis Workshop (GWDAW 14) (January 26-29, 2010, Rome, Italy) with some of the leading people in this field. With the BeSTGRID, in the beginning good results were obtained, however later on it became too inconsistent with regard to its performance. Firstly, there were storage problems. For the whole cluster the total available storage was only 250 GB, which, because of the large input/outputs of several users,

often got quickly filled up and thus the output of the long running simulations would be just destroyed. Secondly, since there were several other users from around the university; sometimes one had to wait for weeks to get enough cores to load his job. Recently, a few fast servers were installed at our own department that solved the computational problems to a great extent. Some of the recent results were presented at Gravitational wave Physics and Astronomy Workshop (GWPAW) (January 26-29, 2011, Milwaukee, Wisconsin, USA).

1.3 Organization of this Thesis

The organization of this thesis is as follows. Chapter 2 presents a brief overview of the general concepts of gravitational waves and their effects on the space-time, different gravitational wave detectors that are used to measure the effect of passing by gravitational waves, different sources that LISA is expected to be able to detect and different data analysis techniques that were used in earlier studies. Chapter 3 introduces the EMRI sources, and two different models that are used to characterize the signals from these sources. In addition, this chapter also discusses the detector response to the incoming gravitational waves and two different methods to approximate this response. Chapter 4 introduces the concepts of the Bayesian approach, different Monte Carlo sampling algorithms and signal processing methods, that are collectively used to conduct Bayesian inference on signals. Chapter 5 presents applications of the methods, that were discussed in Chapter 4, to carry out the Bayesian inference on EMRI signals and the relevant results that were obtained in different scenarios. Finally, some conclusions are presented in Chapter 6. The appendix contains a part of one of the source models given in Chapter 3, which was extremely long and was shifted to appendix for the ease of exposition.

CHAPTER 2

GRAVITATIONAL WAVES, SOURCES AND DETECTORS

2.1 Gravitational Waves

Einstein's theory of general relativity postulates that rapidly moving objects will emit energy. This energy produces periodic ripples or deformations in the fabric of space-time. These ripples are known as gravitational waves (GWs). GWs can be described by oscillations in the so-called fabric of space-time, causing space and everything in it to stretch and squeeze as the waves pass by. Propagating at the speed of light, GWs do not travel through the space-time, rather the fabric of space-time itself is oscillating, i.e. GWs can be thought as "messengers". An important feature of GWs is that all astronomical bodies are transparent to GWs as they pass relatively unaffected through the matter. The GWs carry information about the sources at which they were produced such as information about the movements of stellar bodies, the structures and sizes of merging neutron stars or black-hole binaries and the map of space-time around super massive black holes.

However, the events producing GWs are at great distances from us. Even if caused by the most violent events in a nearby galaxy the GWs become very weak when they reach our planet. This makes the detection of GWs an extremely difficult task.

Until now any attempts to directly detect GWs have not been successful, however, astrophysicists are confident that GWs exist since there is a very strong evidence for their existence based on the indirect measurements of the orbital decay of binary neutron star system PSR1913+16, that are very consistent with the theoretical predictions of energy loss because of gravitational radiation [13, 14].

2.2 The Effect of GWs

The effect of a GW is to stretch and squeeze the space-time fabric in directions perpendicular to the direction of wave propagation. This can be observed by comparing the distance between two (or more) independent freely falling/floating test masses. The GWs are quadrupole, that is they occur in two fundamental states of polarization namely plus- ('$+$') and cross- ('$\times$') polarization. The effect of GWs on a ring of free particles is shown in Figure 2.1. Here the GW is traveling perpendicular to the plane of the ring. The top panel shows the plus polarization while the bottom panel shows the cross polarization. The ring is distorted by the passing GWs and the effect is shown at different phases. Plus polarization changes the distance of free particles by first squeezing the ring along a horizontal direction and then along a vertical direction. Cross polarization has basically the same effect but at 45° inclination. The total area of the plane remains the same and if the GWs are traveling along the plane then there will be no distortion at all.

2.3 GWs' Detectors

As shown above the effect of GWs on free falling particles is that the distance between these particles changes as the wave passes. The change in the distance, called strain amplitude and traditionally denoted by h, is used for measuring the effect of a passing GW directly. However, this

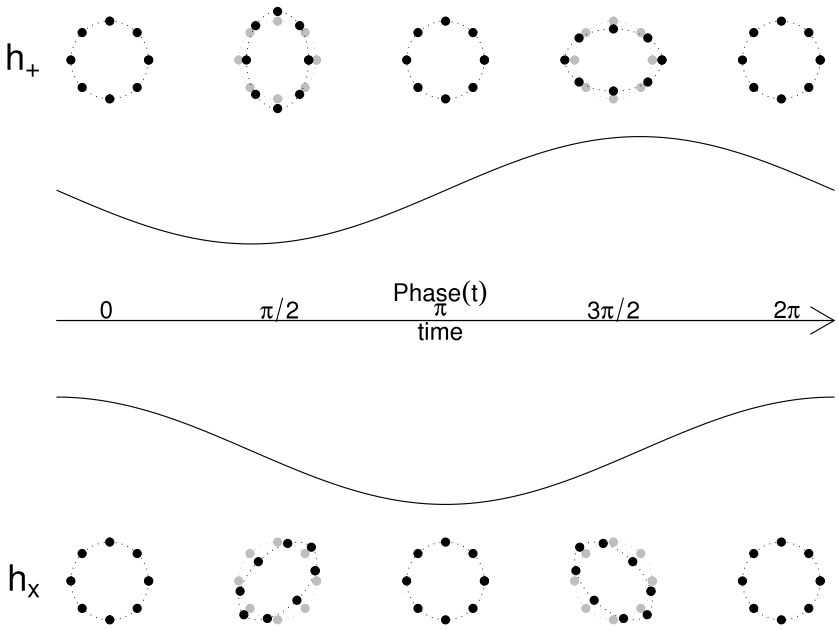

Figure 2.1: Illustration of the effect of "+" and "×" polarizations on a ring of test particles.

change in the distance is extremely small and requires highly sensitive detectors. There are two types of detectors used for the detection of GWs. One is a resonant mass detector in which large masses are used and the deformation caused by GWs in them is measured [15, 16]. The other type is a laser interferometric detector which actually works in a network of more than one detectors located at sufficiently long distance from each other [17, 18]. As the name indicates, laser interferometric detectors use laser interferometry to measure the changes in the strain amplitudes. When a GW passes through the plane of the detector, the distance between the masses changes by an amount ΔL, where L is the distance between the two masses, known as arm-length, resulting in a strain amplitude $h = \Delta L/L$. There are several detectors around the world that use laser interferometry to detect GWs. The ground based detectors are GEO600 (Germany) [19], LIGO (Hanford, USA) and LIGO (Livingston, USA) [20], VIRGO (Italy) [21], and TAMA300 (Japan) [22]. These ground based detectors are large L-shaped instruments with up to $4km$ long arms at 90° to each other. From the central station laser beams

are sent to the ends of the arms where they are reflected by mirrors that are suspended by wires to work as approximately free-falling masses. When a GW passes through the plane of a detector the distance between the mirrors changes by a small amount, which is monitored by photo detectors that measure the phase change of the light. These detectors are at most sensitive to gravitational radiation in the range $1 - 10^4$Hz.

A space-borne detector called Laser Interferometer Space Antenna (LISA) is currently being designed under a joint mission of NASA and ESA. The design of LISA allows to have very long arms and therefore is expected to be sensitive to low frequency gravitational radiation in the range $5 \times 10^{-5} - 10^{-1}$Hz. The basic detector consists of three freely flying spacecraft located at the vertices of an imaginary equilateral triangle configuration with $L = 5 \times 10^9 m$ long sides. Each spacecraft will carry two free-falling test masses and laser instruments that exchange laser beams with other two spacecraft to track the distances between the test masses within them to indicate the passage of a GW. The LISA triangle will move around the sun 20° ($\sim$ $5.2 \times 10^7 km$) behind the Earth with its guiding centre (a point which is equidistant from the three vertices) at the Earth's orbit about the Sun. The LISA constellation is depicted in Figure 2.2.

LISA is not a pointed instrument, and can never be, rather it is an all-sky monitor and at any one time, LISA maps the whole sky. At different positions it will have different sensitivities to GWs from a particular source depending on the location of the source and the polarisation of the waves. LISA will measure simultaneously both polarization components of the incoming GWs. The data will initially consist of at least three time series, recorded along three arms of the detector, from which all physical parameters of the source, including its position, can be extracted. Mapping the whole sky means that there will be GWs from tens of thousands of resolvable as well as unresolvable sources all overlapping in frequencies and phases [23, 24].

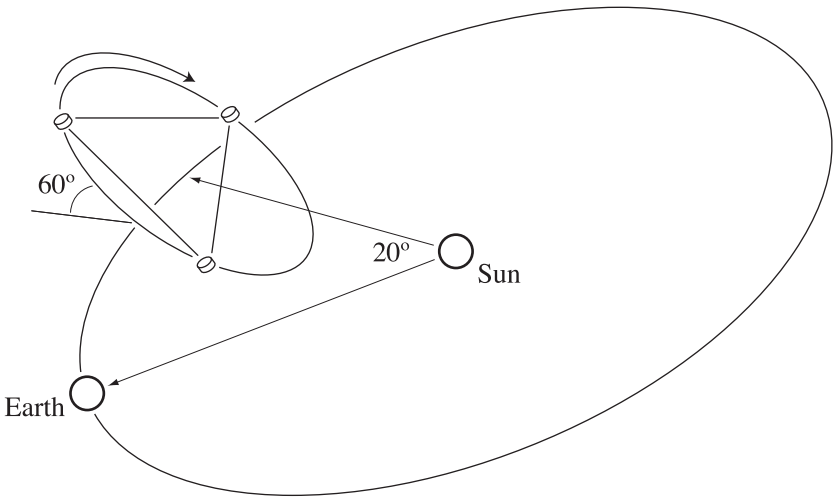

Figure 2.2: LISA constellation orbiting the Sun. The sides of the triangle are $5 \times 10^9 m$ long, and the guiding centre follows the Earth in its orbit, about 20° behind. The triangle is inclined 60° to the ecliptic plane as viewed from the Sun and rotates clockwise.

2.4 LISA Sources

LISA will be observing signals from several different GW sources. Some of the sources which fall in the LISA sensitivity band are briefly mentioned as following.

2.4.1 Galactic Binaries

These types of binaries develop when two objects with very dense masses such as neutron stars (NSs) or white dwarfs (WDs), with roughly equal masses, orbit about each other. The compact objects move initially at an elliptical orbit about the common centre of mass. Over the course of time their individual orbits become increasingly circular as the two objects come closer and closer and at some point in time the two bodies appear to move around the centre of mass at the same circular orbit. This orbit decays with time like a spiral, as the two bodies come closer and closer to each other, which proceeds towards the common centre of the masses. Such spirals are called inspirals (in-spirals). As the two masses gradually

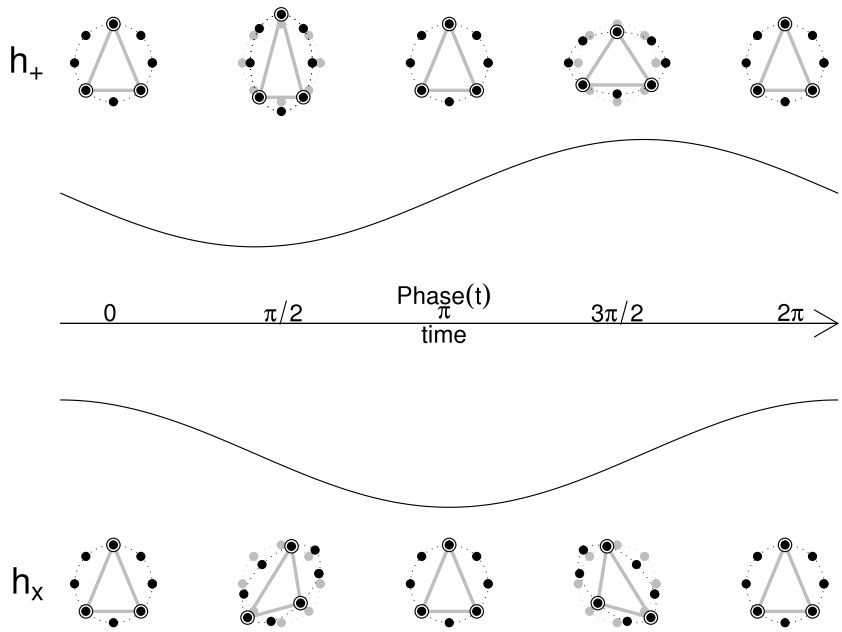

Figure 2.3: Illustration of the effect of "+" and "×" polarizations on LISA constellation. The squeeze and stretch caused by passing GW is changing the distances between the adjacent spacecraft.

inspiral, they emit gravitational radiation with a frequency falling well inside the LISA sensitivity band. These signals encode the luminosity distance to a binary, its sky location, and information about other physical parameters. During its operation the LISA is expected to observe several thousands individual galactic binary systems [25, 26].

2.4.2 Mergers of Massive Black Hole Binaries

Similar to galactic binary systems are massive black hole binary systems in which both members are massive black holes with a total mass in range $10^5 M_\odot - 10^9 M_\odot$. These sources will be detectable by LISA at extremely large distances due to large masses involved. The GWs from these sources encode information about the masses and the spins of the two members. Once such sources are detected and the physical parameters are estimated, this information can be used to find out how these massive black holes are formed and what is the rate of their mergers, that is how often massive

black holes encounter each other, which will put light on the evolution and structure of the large scale galactic dynamics [27, 28].

2.4.3 Extreme Mass Ratio Inspirals

It is predicted that most galaxies, in their centres, host super massive black holes ($\sim \ 10^5 M_\odot - 10^7 M_\odot$) surrounded by a dense population of stellar objects such as NSs, WDs or small black holes and other normal stars. Due to multi-body interactions these stellar objects are often pushed into an orbit which passes too close to the central mass. The captured object then spirals in by orbital decay through the emission of gravitational radiation and eventually plunges into the central mass. Due to large differences between the two masses, such inspirals are called extreme mass ratio inspirals (EMRIs), and are one of the most exciting sources to be observed by LISA. The EMRI studies will help to understand the structure of the space-time around the SMBH using the physical parameters such as spin and mass, and the interactions between SMBH and the cluster of stellar masses around it [29]. The detection of these signals and the estimation of their physical parameters is the focus of this work and will be discussed in details in Chapter 3.

2.4.4 Cosmological Background, Stochastic Sources and Bursts

Stochastic GWs are random signals generated by large number of independent, incoherent and unresolved or diffuse sources. These signals can originate either from the early cosmological events or the astrophysical events happening throughout the history of the universe. The cosmological background could consist of the left over GWs that were produced as the result of the processes that took place very shortly after the big bang. The astrophysical background on the other hand is produced by very recent processes, such as supernova bursts and signals from millions of

unresolved compact or massive black hole inspirals [30, 31]. These GWs might encode information about the early structure of the universe and other high energy astrophysical events. The superposition of GWs from these random and unresolved sources will form a very strong background noise which will be there in LISA data stream and therefore will be of great importance for the precise detection of the other sources.

The relative strengths of individual signals from all sources, mentioned above, along with the LISA instrument noise is shown in Figure 2.4.

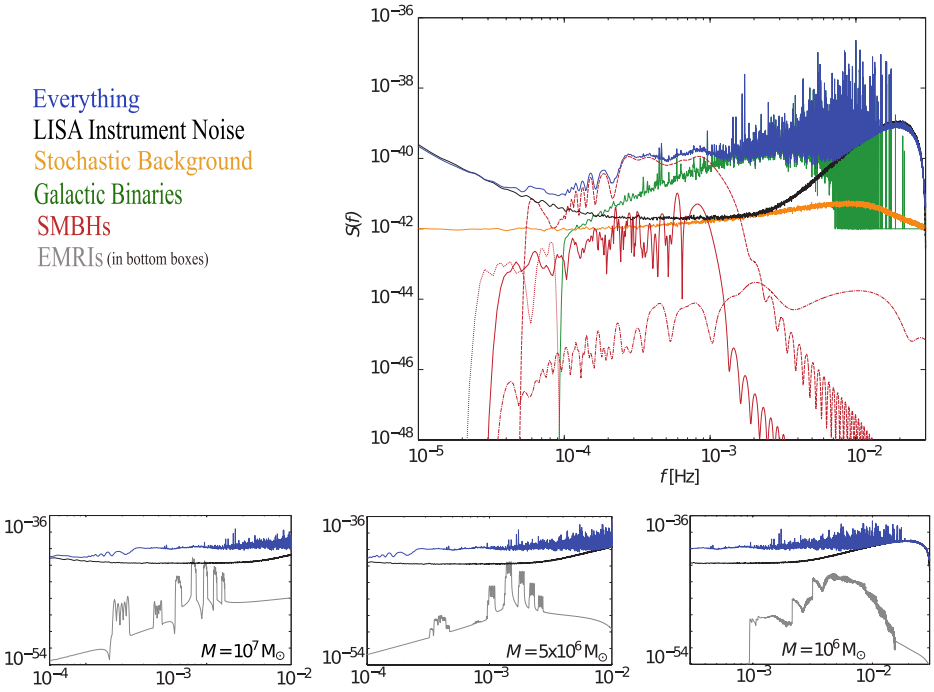

Figure 2.4: (Colour online) The large box shows the one-sided power spectrum (see also Chapter 4) of two years of instrument noise (in black), 60 million Galactic binaries (in green), 4 SMBH binaries (in red), cosmological stochastic background (in yellow) and everything (in blue). The three small boxes in the bottom show the relative strengths of the 3 different EMRIs (in grey) (further details in Chapter 3) to LISA instrument noise and everything. Credit: MLDC 4 website [1].

2.5 Data Analysis

Since the incoming GWs are being measured over time, the data sets consist of multiple time series of strains measured along each arm of the LISA triangle. Because of the LISA motion around the sun the actual incoming signal is Doppler shifted, before being used for actual data analysis, this effect of Doppler shift must be deconvolved in the data. Each spacecraft sends and receives laser beams to and from the other two spacecraft to measure the strain caused by the passage of GWs. In an equal arm detector the lasers beams experience the same delay in each arm and when measurements from all the arms are combined the laser phase errors in all laser beams are common and cancel exactly. There would have been no problem with the received data if LISA was an equal arm detector. However, in reality each spacecraft is moving in its own orbit, therefore LISA's armlengths will not be equal, which gives rise to an overwhelming laser phase noise. To cancel this noise, the phase measurements at each spacecraft are carefully recombined into time delayed observables called time-delay interferometry [32] (more details in Chapter 3). Unlike the ground-based detector data, which is dominated by local noises, the LISA data is expected to be dominated by GW signals from millions of sources. Because of the large number of GW sources there will be huge confusion noise, in which different sources can not be easily isolated because of similarity of their signals [33].

2.5.1 Global Analysis Techniques: Overview

To date, several approaches have been developed for the detection and estimation of parameters of GW signals from different sources. These approaches are very general. To avoid excessive details we present a brief overview of only those methods that were employed and found useful for the search and estimation of parameters of EMRI signals.

2.5.1.1 Matched Filtering Approach

In the matched filtering approach, a bank of templates for the theoretical waveforms is generated over the allowable ranges of parameters. These templates are individually subtracted from the data and the likelihood is computed over the resulting residuals. For stationary noise, the Gaussian likelihood is closely related to the signal-to-noise ratio (SNR) (see also Section 4.5.1 and 4.6). When a chosen template has a sufficient overlap with the signal buried in the data it will maximize the SNR and a detection can be claimed. However, this method is suitable for signals with a small number of parameters because the number of templates increases exponentially with the number of parameters, which can be time consuming if the model signals are computationally expensive and the required waveforms are long. Due to a large number of source parameters (a typical EMRI signal depends on 17 parameters) and longer waveforms (large number of cycles) required for characterizing the EMRI signals, a full scale template based search would require roughly 10^{40} templates which is impossible to do with the capabilities of current computational resources. Instead of searching the whole data at once, which typically is $\sim 1\frac{1}{2} - 2$ years long to cover the entire EMRI event, a semi-coherent search was conducted in [34]; in which shorter data segments of few weeks were used along with a maximization of several source parameters to reduce the dimensionality of parameter space in order to lessen the computational costs.

2.5.1.2 Time-Frequency Approach

A computationally cheaper approach that was used for EMRI detection is based on time-frequency searches in which the LISA data is searched for tracks left by EMRI signals in the spectrograms of LISA data. These methods consist of constructing a spectrogram of the signal. The time domain data is first divided up into shorter segments of a few weeks length, each segment is then Fourier transformed and thus a time-frequency spectrogram is constructed for each segment. These spectrograms are examined

for the identification of possible tracks of an EMRI signal. These tracks correspond to different harmonics and the source parameters can be extracted from the identified tracks [35, 36]. Although this approach has much lower computational costs, it has a lower detection sensitivity. So far, this approach has some success in detecting isolated EMRI sources buried in MLDC LISA noise, it is not yet known how this algorithm will perform when applied to more realistic data containing multiple intersecting tracks of signals from millions of galactic binaries, multiple massive black hole binaries and hundreds of overlapping-in-time EMRIs.

2.5.1.3 Monte Carlo Approach

With the above methods only limited information about the EMRI sources can be extracted from LISA data. Because of the large number of EMRI parameters a suitable alternative is to conduct Monte Carlo searches based on a Bayesian approach in which the posterior distribution is either (if possible) analytically evaluated or is numerically integrated using some Markov chain Monte Carlo (MCMC) algorithm to yield the marginal distributions of the parameters. The theoretical structure of EMRI waveforms yields a very complicated posterior distribution which has multiple local maxima along the global maximum, therefore more advanced MCMC algorithms are required for efficient searches. The first ever MCMC applications for EMRI search and estimation were employed in [37], in which the use of reversible jump MCMC (RJMCMC) was demonstrated to characterize a simplified EMRI signal. A variant of MCMC called Metropolis-Hastings Monte Carlo (MHMC) was used in [38–40], which yields a non-Markovian chain because of the use of several independent and purposely directed proposal distributions that allow a range of steps of different sizes in the parameter space [41]. To enhance the performance of their algorithms, several additional tricks such as conducting several small MCMC runs using multiple chains to identify stronger signals prior to final MCMC search, constrained MCMC jumps scaled by Fisher information matrix (FIM) of parameters, simulated annealing, and island hopping

techniques were also used to ensure the chains move toward the global mode without getting stuck at local modes.

CHAPTER 3

EMRIS, SOURCE MODEL AND LISA RESPONSE

3.1 Introduction

It is predicted that most of the galaxies, including our own milky way, host a super massive black hole (SMBH) with a typical mass of $10^5 M_\odot \lesssim M \lesssim 10^7 M_\odot$ ($M_\odot$ = solar mass) in their centres. These SMBHs are surrounded by a large population of stellar mass compact objects (COs) such as NSs , WDs and small black holes (BHs) and other normal stars with a typical mass $\sim 10 M_\odot$. Due to multi-body interactions most of these COs occasionally wrestle with each other which can result in some of them being pushed into and hence captured in, an orbit which passes too close to the central mass [42]. Once captured in the strong gravitational field of SMBH, the CO then starts orbiting about the central mass in an eccentric orbit which decays over time and the CO gradually spirals in into central mass, followed by a final plunge. Inspirals of such COs into an SMBH 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 GWs in space-time, to be observed by LISA [39]. Such sources are very important in the sense that they encode information about the formation and structure of SMBHs, the characteristics of space-time around them such as Lense-Thirring effects, and spin-orbit coupling [40].

EMRI signals are expected to be very weak as the instantaneous amplitude of a typical EMRI signal is an order of magnitude below even the noise fluctuations in LISA alone and, further, it will be lying behind the strong foreground of confusion noise created by other GWs sources such as massive black hole binaries and NSs, WDs binaries. This makes the detection of such sources in the LISA data stream and parameter estimation for them a challenging technical problem. Due to extreme mass ratio (typically $\mu/M \sim 10^{-5}$) these inspirals proceed very slowly as the captured objects remain in the strong field of the central mass for a significant amount of time before being eventually pulled into it. This means that the EMRI signals will require to be followed for a very long time and hence a large number of waveform cycles (typically 10^5) will be required to accumulate enough signal power for their precise detection [38, 39]. The large parameter space of this type of source (17 dimensional parameter space of signals, see next section) and the complicated likelihood surface makes it a statistical challenge too.

The Mock LISA Data Challenges (MLDC) are a programme to demonstrate and facilitate the development of LISA data analysis capabilities by providing information on common data formats, standard models for LISA response, noises and waveforms and other necessary tools and software [43]. The EMRI waveform model given in the MLDC releases is based on the *Analytic Kludge* waveform (AKW) approximation given in [29, 44]. These waveforms use Pater-Mathews as basis waveforms and Post-Newtonian (PN) approximations for orbital dynamics. While this model is not a very good approximation to the realistic EMRIs, it still captures the main features of EMRI signals and as compared to others (see [45–47] for more details) it is theoretically simplest and computationally cheapest waveform model, though it is still expensive to handle with the currently available computing equipments. Furthermore, although AKW is the simplest among others, this is still very complicated in the sense that the generated waveforms have emission at multiple frequencies cor-

responding to different harmonics of the fundamental orbital frequencies such as the harmonics of the orbital frequency, harmonics of the perihelion precession rate, and harmonics of the orbital-plane precession rate. These frequencies evolve over time during the waveform progression thereby making these signal rather more complicated from a data analysis point of view. The EMRI sources can be divided into three groups depending on the size of the central mass, i.e. $M \sim [0.95, 1.05] \times M_{\odot}10^{7}$ (high mass EMRI sources), $M \sim [4.75, 5.25] \times M_{\odot}10^{6}$ (medium mass EMRI sources) and $M \sim [0.95, 1.05] \times M_{\odot}10^{6}$ (low mass EMRI sources). These system can also be grouped on the basis of the type of GW they generate. For example, the high and medium mass EMRI systems are also called low frequency GW sources and analogously the low mass systems are called high frequency EMRI systems. The one-sided power spectrum (see also Section 4.4.3) of these sources is shown in Figure 3.1 (see also Figure 2.4 in Chapter 2). For all the three sources, the same waveform model, with different parameter values corresponding to different sources, is used; this is described in the following section.

3.2 The AKW Model

A full two-body EMRI system is described by 17 parameters but since the spin of the CO is negligible with respect to the spin of SMBH, the three parameters describing the spin of the CO can be ignored. Thus the system now depends on 14 parameters that are listed in Table 3.1 with their standard symbols and units of measurement as adopted by MLDC [44]. The waveform parameters can be categorized into three groups namely *intrinsic* parameters: which affect the frequency and phase evolution of the different harmonics, *extrinsic* parameters: which affect the waveform projection into a detector response, and *phase* parameters: which define the relative phase of different harmonics at one fiducial time [38].

To describe the model, let us consider an EMRI system in which a CO of mass μ is rotating around an SMBH of M ($\mu/M \ll 1$) on an elliptical

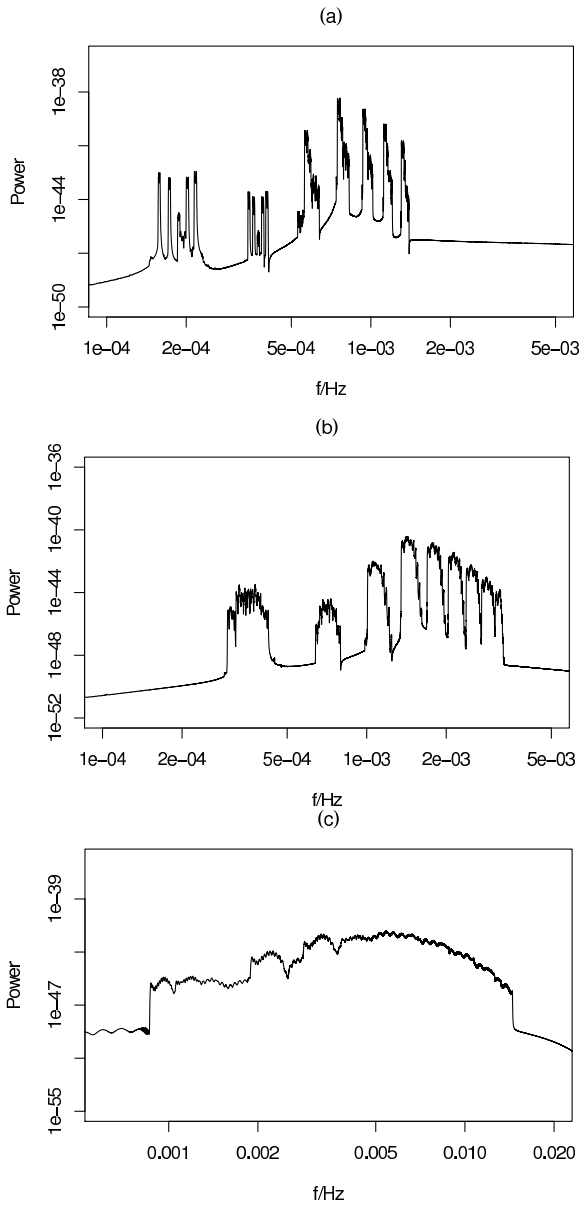

Figure 3.1: The three plots (a), (b) and (c) show the one-sided power spectrum of different EMRI systems with central masses $M \sim [0.95, 1.05] \times M_{\odot}10^7$, $M \sim [4.75, 5.25] \times M_{\odot}10^6$ and $M \sim [0.95, 1.05] \times M_{\odot}10^6$ respectively. For all these sources the CO's mass is $\mu \sim [9.5, 10.5] \times M_{\odot}$ and luminosity distance is $D_L \sim [0.10, 1.0]$ gigaparsec (GPC).

orbit (see Figure 3.2) with semi-major axis a, eccentricity e, and orbital frequency $\nu = (2\pi M)^{-1}(M/a)$. Let also $G = c = 1$, where G is the

Table 3.1: Summary of EMRI parameters and their meaning.

Category	Parameter	Symbol	Standard Units
Intrinsic	CO's mass	μ	$M_\odot$
	SMBH's mass	M	$M_\odot$
	Initial azimuthal orbital... ...frequency	ν_0	Hertz
	Initial eccentricity	e_0	1
	SMBH's spin	χ	M^2
	Angle between spin and... ...angular momentum	λ	Radian
Extrinsic	Ecliptic latitude	θ_S	Radian
	Ecliptic longitude	ϕ_S	Radian
	Polar angle of spin	θ_K	Radian
	Azimuthal angle of spin	ϕ_K	Radian
	Distance to the source	D_L	Parsec
Phase	Initial azimuthal orbital... ...phase	Φ_0	Radian
	Initial azimuthal angle of... ...orbital angular momentum	α_0	Radian
	Initial direction of pericenter	$\tilde{\gamma}_0$	Radian

gravitational constant and c the speed of light. Let n denote the number of all possible harmonics associated with the orbital frequency ν of a given EMRI source. For an n-harmonic waveform the amplitude coefficients of the two polarizations are defined as:

$$h_+ \equiv \sum_n A_n^+, \qquad h_\times \equiv \sum_n A_n^\times \tag{3.2.1}$$

where A_n^+ and $A_n^\times$ are given by,

$$\begin{aligned} A_n^+ &= -[1 + (\hat{L}\cdot\hat{n})^2][a_n \cos(2\gamma) - b_n \sin(2\gamma)] + c_n[1 - (\hat{L}\cdot\hat{n})^2] \\ A_n^\times &= 2(\hat{L}\cdot\hat{n})[b_n \cos(2\gamma) + a_n \sin(2\gamma)] \end{aligned} \tag{3.2.2}$$

For $e < 0.136$ the sums in Equation (3.2.1) are truncated at $n = 4$ otherwise at $n = 30e$. The coefficients a_n, b_n and c_n are given by Peter and Mathews

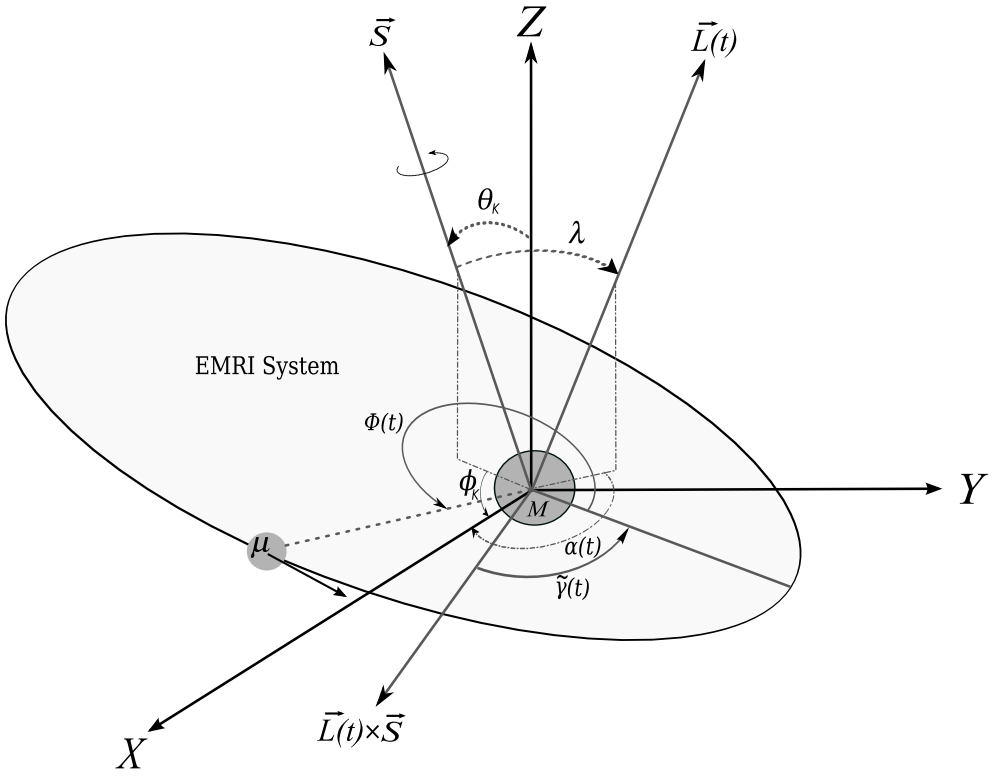

Figure 3.2: An EMRI system, in which a CO μ orbits the central mass M on an elliptical path. $\vec{S}$ is the rotation axis (spin axis) of the SMBH which is parametrized by its magnitude χ, polar angle θ_K and azimuth ϕ_K. Similarly $\vec{L}(t)$ is the instantaneous orbital angular momentum; its direction is parametrized by angle λ, between $\vec{S}$ and $\vec{L}$, and its time varying azimuth $\alpha(t)$. Finally $\tilde{\gamma}(t)$ is the time varying direction of pericenter measured from $\vec{L}(t) \times \vec{S}$ and $\Phi(t)$ is the average orbital phase with respect to pericenter.

expressions,

$$\begin{aligned} a_n &= -n\mathcal{A}\Big\{[J_{n-2}(ne) - 2eJ_{n-1}(ne) + \frac{2}{n}J_n(ne) + 2eJ_{n+1}(ne) \\ &\quad -J_{n+2}(ne)]\cos[n\Phi(t)]\Big\}, \\ b_n &= -n\mathcal{A}\Big\{(1-e^2)^{1/2}[J_{n-2}(ne) - 2J_n(ne) + J_{n+2}(ne)]\sin[n\Phi(t)]\Big\}, \\ c_n &= 2\mathcal{A}J_n(ne)\cos[n\Phi(t)] \end{aligned} \tag{3.2.3}$$

where $\mathcal{A}$ is the overall amplitude defined as,

$$\mathcal{A} \equiv (2\pi\nu M)^{2/3}\frac{\mu}{D_L}, \tag{3.2.4}$$

$J_n(\cdot)$ the Bessel functions of the first kind and γ is an azimuthal angle measuring the direction of pericenter with respect to $\hat{x} \equiv [-\hat{n} + \hat{L}(\hat{L}\cdot\hat{n})]/[1 - (\hat{L}\cdot\hat{n})^2]^{1/2}$ and is related to $\tilde{\gamma}$ by

$$\gamma = \tilde{\gamma} + \beta \tag{3.2.5}$$

where β is the angle from $\hat{x} \propto [-\hat{n} + \hat{L}(\hat{L}\cdot\hat{n})]$ to $\hat{L}\times\hat{S}$, defined as,

$$\beta = \tan^{-1}\left[\frac{(\cos\lambda)\hat{L}\cdot\hat{n} - S\cdot\hat{n}}{\hat{n}\cdot(\hat{S}\times\hat{L})}\right]. \tag{3.2.6}$$

The vector products appearing in the above expressions can be represented in terms of equivalent trigonometric expressions given as following.

$$\begin{aligned}
\hat{S}\cdot\hat{n} &= \cos\theta_S\cos\theta_K + \sin\theta_S\sin\theta_K\cos(\phi_S - \phi_K),\\
\hat{n}\cdot(\hat{S}\times\hat{L}) &= \sin\theta_S\sin(\phi_K - \phi_S)\sin\lambda\cos\alpha(t)\\
&\quad + \frac{(\hat{S}\cdot\hat{n})\cos\theta_K - \cos\theta_S}{\sin\theta_K}\sin\lambda\sin\alpha(t)\\
\hat{L}\cdot\hat{n} &= \cos\theta_S\cos\theta_L - \sin\theta_S\sin\theta_L\cos(\phi_S - \phi_L).
\end{aligned} \tag{3.2.7}$$

The angles $\theta_L(t)$ and $\phi_L(t)$ specify the instantaneous direction of $\hat{L}$ and are given in terms of θ_K, ϕ_K, λ, $\alpha(t)$ as

$$\theta_L(t) = \arccos\left(\cos\theta_K\cos\lambda + \sin\theta_K\sin\lambda\cos\alpha(t)\right) \tag{3.2.8}$$

$$\phi_L(t) = \arctan\left(A/B\right) \tag{3.2.9}$$

where A and B are given as,

$$\begin{aligned} A &= \sin\theta_K \cos\phi_K \cos\lambda + \sin\phi_K \sin\lambda \sin\alpha(t) \\ &\quad - \cos\theta_K \cos\phi_K \sin\lambda \cos\alpha(t) \\ B &= \sin\theta_K \sin\phi_K \cos\lambda - \cos\phi_K \sin\lambda \sin\alpha(t) \\ &\quad - \cos\theta_K \sin\phi_K \sin\lambda \cos\alpha(t). \end{aligned} \tag{3.2.10}$$

The evolution of $\Phi(t)$, $\nu(t)$, $\tilde{\gamma}(t)$, $e(t)$, and $\alpha(t)$ is given by the following PN formulas.
Let us define $Y = 1 - e^2$, $Z = 2\pi M\nu$, and $\chi = S/M^2$,

$$\frac{d\Phi}{dt} = 2\pi\nu, \tag{3.2.11}$$

$$\begin{aligned} \frac{d\nu}{dt} &= \frac{96}{10\pi}(\mu/M^3) Z^{11/3} Y^{-9/2} \Bigg\{ \left[1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right] Y \\ &\quad + Z^{2/3}\left[1273336 - \frac{2561}{224}e^2 - \frac{3885}{128}e^4 - \frac{13147}{5376}e^6\right] \\ &\quad - Z\,\chi\cos\lambda Y^{-1/2}\left[\frac{73}{12} + \frac{1211}{24}e^2 + \frac{3143}{96}e^4 + \frac{65}{64}e^6\right]\Bigg\}, \end{aligned} \tag{3.2.12}$$

$$\begin{aligned} \frac{d\tilde{\gamma}}{dt} &= 6\pi\nu Z^{2/3} Y^{-1}\left[1 + \frac{1}{4}2\pi\nu M^{2/3}(1-e^2)^{-}1(26 - 15e^2)\right] \\ &\quad - 12\pi\nu\,\chi\cos\lambda Z\, Y^{-3/2}, \end{aligned} \tag{3.2.13}$$

$$\begin{aligned} \frac{de}{dt} &= -\frac{e}{15}(\mu/M^2)\, Y^{-7/2} Z^{8/3}\Bigg\{(304 + 121e^2)(1 - e^2(1 + 12\, Z^{2/3})) \\ &\quad - \frac{1}{56}\, Z^{2/3}(133640 + 108984e^2 - 25211e^4)\Bigg\} \\ &\quad + e(\mu/M^2)\,\chi\cos\lambda Z^{11/3}\, Y^{-4}\left[\frac{1364}{5} + \frac{5032}{15}e^2 + \frac{263}{10}e^4\right], \end{aligned} \tag{3.2.14}$$

$$\frac{d\alpha}{dt} = 4\pi\nu\,\chi Z Y^{-3/2}. \tag{3.2.15}$$

The parameters $\Phi(t)$, $\nu(t)$, $\tilde{\gamma}(t)$, $e(t)$, $\alpha(t)$ are evolved forward in time, up to the point when the CO finally plunges into central mass. The plunge

occurs at $a_{\min} = M(6+2e)/(1-e^2)$ so the maximum of ν is set at

$$\nu_{max} = (2\pi M)^{-1}[(1-e^2)(6+2e)]^{3/2}. \tag{3.2.16}$$

The waveform is shut-off when ν reaches ν_{max}. To generate detector response we need to compute the polarization angle $\psi(t)$ which is given by

$$\psi(t) = \tan^{-1}\left(\frac{\cos\theta_S \sin\theta_L(t)\cos(\phi_S - \theta_L(t)) - \cos\theta_L(t)\sin\theta_S}{\sin\theta_L(t)\sin(\phi_S - \theta_L(t))}\right). \tag{3.2.17}$$

Finally, the polarization signals denoted by $h_+(t)$ and $h_\times(t)$ are given by

$$h_+(t) = A^+(t)\cos 2\psi(t) + A^\times(t)\sin 2\psi(t), \tag{3.2.18}$$

$$h_\times(t) = -A^+(t)\sin 2\psi(t) + A^\times(t)\cos 2\psi(t). \tag{3.2.19}$$

The polarization signals $h_+(t)$ and $h_\times(t)$ are then adjusted for time delay as will be discussed in later sections. The numerous Bessel functions that appear in Pater-Mathews expressions above are the most time consuming terms of this model. To generate a waveform of a given length, almost two-third of the total time is spent on the calculation of these Bessel functions. Instead of calculating them individually the familiar recursive relationship of Bessel functions can be used to reduce the total time by half.

3.2.1 Truncated EMRI Waveform Model

The truncated EMRI waveform model is a special version of the true AKW model introduced in [38]. In this model several simplifications are used in order to speed up the generation of waveforms. In these waveforms the number of harmonics of the orbital frequency, ν, is kept at $n \leq 5$ and that of the perihelion precession are kept at $l = 2$. For a further speed up the Bessel functions are expanded in the power of eccentricity using Taylor series up to order three only. The amplitude coefficients of the two

polarizations are computed using the following expressions.

$$\begin{aligned} A_n^+ &= -\frac{1}{2} n\, g_n(n,e)[A_{+1}\cos(n\Phi+2\gamma)+A_{+2}\sin(n\Phi+2\gamma)] \\ A_n^\times &= -n\, g_n(n,e)[A_{\times 1}\cos(n\Phi+2\gamma)+A_{\times 2}\sin(n\Phi+2\gamma)] \end{aligned} \quad (3.2.20)$$

where $g_n(n,e)$ is defined as

$$\begin{aligned} g_n(n,e) &= J_{n-2}(ne) - 2eJ_{n-1}(ne) + \frac{2}{n}J_n(ne) + 2eJ_{n+1}(ne) - J_{n+2}(ne) \\ &\quad + \sqrt{(1-e^2)}[J_{n-2}(ne) - 2J_n(ne) + J_{n+2}(ne)]. \end{aligned} \quad (3.2.21)$$

The Bessel functions appearing in the above equation are expanded by Taylor series up to the first three terms. For harmonics $n = 1,2,3,4,5$ the functions $g_n(n,e)$ are derived as

$$\begin{aligned} g_1(1,e) &= -3e + \frac{13}{8}e^3 + \frac{5}{192}e^5 \\ g_2(2,e) &= 2 - 5e^2 - \frac{23}{8}e^4 \\ g_3(3,e) &= 3e - \frac{57}{8}e^3 - \frac{321}{64}e^5 \\ g_4(4,e) &= 4e^2 - 10e^4 - \frac{101}{12}e^6 \\ g_5(5,e) &= \frac{125}{24}e^3 - \frac{5375}{384}e^5 - \frac{42125}{3072}e^7. \end{aligned} \quad (3.2.22)$$

Note that the functions $g_n(n,e)$ do not represent the individual Bessel functions themselves. Rather these are actually the solutions for $n = 1,2,3,4,5$ harmonics of the combined expression for Bessel functions as given in Equation (3.2.21). The amplitudes A_{+1}, A_{+2}, $A_{\times 1}$, and $A_{\times 2}$ are given in appendix (A.1). The PN-evolution equations for truncated EMRI model are the same as the full AKW model.

3.3 Parameter Ranges

The allowable ranges for different waveform parameters over which their prior distributions can be defined are those specified by MLDC. The ranges that are common to all types of EMRI sources are $\mu \in [9.5, 10.5]$, $e \in [0.15, 0.25]$, $\chi \in [0.5, 0.7]$, $(\phi_S, \tilde{\gamma}, \Phi, \alpha, \phi_K) \in [0, 2\pi]$, $(\theta_S, \lambda, \theta_K) \in [0, 2\pi]$, whereas the range of central mass parameter M is source dependent as given in Section 3.1. From the MLDC data it appears that the range of orbital frequency parameter ν is also source dependent but there is no specific range given. However, a range of $[10^{-5},\ 10^{-2}]$, which is a little narrower than LISA's actual sensitivity band $[10^{-5},\ 10^{-1}]$, seems reasonable as it covers the range of this parameter in all types of EMRI sources. The domain of luminosity distance is $\mathbb{R}^+$.

3.4 LISA Response

LISA does not observe the individual polarizations $h_+(t)$ and $h_\times(t)$ of the incoming signal directly, rather the observed signal is adjusted for the effects of different dynamics to which LISA will be exposed during its operation. Firstly, since the whole LISA constellation orbits about the Sun, its position with respect to a particular source will keep changing, which introduces a periodic Doppler shift, the magnitude and phase of which depends on the location of the source in the sky. Secondly, since each spacecraft has a different independent orbital plane, the LISA triangle will not be an equilateral one. Its arms will be continuously changing over time: the so-called "breathing" effect. Imagine that the LISA triangle is strictly equilateral like a classic Michelson interferometer in which one spacecraft plays the role of the central station where the laser is split and then sent to the other two spacecraft down the two arms. The two lasers are recombined after being transponded by the other two spacecraft. Suppose that before being split the initial laser develops a random fluctuation which travels through both arms and then returns back to the central spacecraft.

Since the arms are equal in length, the fluctuations (which in reality are the copies of the same fluctuation) along the two arms will arrive at the same time and hence effectively cancel out. However, if the arms are unequal then these fluctuations will not arrive at the same time and hence will not cancel out, resulting in laser phase noise which is expected to be several orders of magnitude larger than any instrumental noise and most of GW sources. The lasers used by LISA, like any other laser, are susceptible to frequency noise; specifically, there is a fluctuation in the operating frequency. This will be a significant noise source for LISA, especially since the interferometer arm lengths will not be exactly the same. To subtract the laser phase noise, time-delay-interferometry (TDI) scheme is used, in which the output time-series of the three spacecraft are carefully recombined to adjust the time-shifted phase measurements. The complete representation for TDI scheme given in [48] is rather complex and involves a lot of geometry and physics, therefore only a basic overview is presented here.

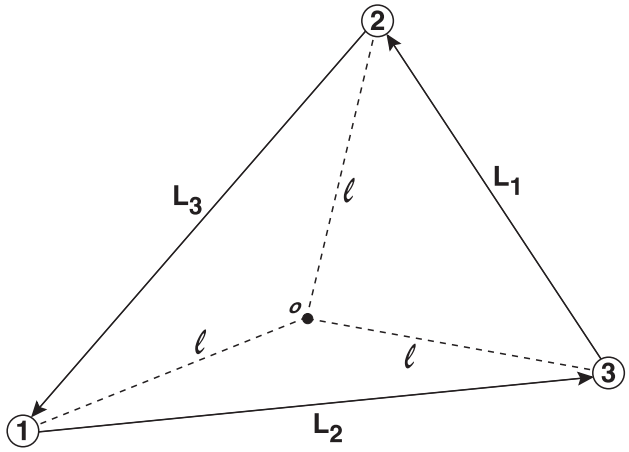

Figure 3.3: Schematic LISA configuration. The three spacecraft are labeled as {1,2,3} each at distance ℓ from point o in the plane of the spacecraft. The opposing arm-lengths are denoted by $\{L_1, L_2, L_3\}$. The point o is called the guiding centre.

The overall geometry of LISA constellation is shown in Figure 3.3. The vertices of the triangle denote the three spacecraft, labeled as {1,2,3} and the sides opposite to each spacecraft are labeled as $\{L_1, L_2, L_3\}$ respectively.

The point o is the guiding centre which is equidistant (ℓ) from each spacecraft. The distance ℓ is given by

$$\ell = \frac{L_1 L_2 L_3}{\sqrt{2L_1^2L_2^2 + 2L_2^2L_3^2 + 2L_3^2L_1^2 - L_1^4 - L_2^4 - L_3^4}}. \tag{3.4.1}$$

In basic Michelson interferometry, the data stream is constructed by measuring the phase differences between the laser beams in the two arms of the detector over the same time t, resulting in a trio of TDI observables associated with the three spacecraft. The time taken by a laser beam to travel along the arm L_i is given by $\tau_i = L_i/c$, where c is the speed of light. For an unequal-arm Michelson detector the TDI response associated with spacecraft 1 is traditionally denoted by $X(t)$ and is defined as

$$X(t) = s_1(t) - s_1(t - 2\tau_2) - s_2(t) + s_2(t - 2\tau_1) \tag{3.4.2}$$

where $s_i(t)$ is the actual incoming signal arriving at spacecraft i. The other two responses denoted as $Y(t)$ and $Z(t)$ can be derived using the simple permutation of the indices $1 \to 2 \to 3 \to 1$ given as following

$$Y(t) = s_2(t) - s_2(t - 2\tau_3) - s_3(t) + s_3(t - 2\tau_2) \tag{3.4.3}$$

$$Z(t) = s_3(t) - s_3(t - 2\tau_1) - s_1(t) + s_1(t - 2\tau_3). \tag{3.4.4}$$

Several other combinations of TDI observables can be found in literature, for example see [32, 48, 49] for more details. The above TDI observables tend to be highly correlated; therefore another linear transformation is applied to get another set of TDI observables traditionally denoted by A, E and T and defined as:

$$A = \frac{1}{3}(2X - Y - Z), \tag{3.4.5}$$

$$E = -\frac{1}{\sqrt{3}}(Z - Y), \tag{3.4.6}$$

$$T = -\frac{\sqrt{2}}{3}(X - Y - Z). \tag{3.4.7}$$

[49]. The correlation matrix of the above three combinations is a diagonal matrix which means that A, E and T are uncorrelated. The variable T is insensitive to GWs in low-frequency cases and therefore is ignored and only A and E are used to perform the searches and analyses.

3.4.1 Low Frequency Approximation

In practice, to compute LISA response in the form of three TDI variables as described in [50] requires huge memory and computing power particularly for longer waveforms. A fine and fast approximation called the low frequency approximation was proposed in [24, 29] in which the response from actual three arms LISA detector is approximated by envisioning it as two 2-arms detectors outputting two orthogonal signals. In this approximation the changes in the arm-length of LISA triangle due to LISA motions are ignored, i.e. the LISA constellation is assumed a static equal arm Michelson interferometer. This is still a good approximation to full LISA response for GWs with frequency $f \ll f_* \sim 10^{-2}$Hz (i.e. GWs with much larger wavelength than the LISA arms' length). The first two types of EMRIs, i.e. high mass and medium mass fall in this category. To describe this scheme, let us denote the two detectors as I and II with and the strain amplitudes that are measured by them by $h_I(t)$ and $h_{II}(t)$ respectively. Assuming that detector I is formed by arms 1 and 2 the strain amplitude is given by

$$h_I(t) = [\delta L_1(t) - \delta L_2(t)]/\bar{L} \tag{3.4.8}$$

where $\delta L_1(t) = L_1(t) - \bar{L}$ with $\bar{L}$ denoting the average arm length. The second orthogonal strain amplitude is given by

$$h_{II}(t) = \frac{1}{\sqrt{3}}[\delta L_1(t) + \delta L_2(t) - 2\delta L_3(t)]/\bar{L}. \tag{3.4.9}$$

The two strain amplitudes can be expressed as following,

$$h_I(t) = \frac{\sqrt{3}}{2}[F_I^+(t)A^+(t)F_I^\times(t)A^\times(t)] \quad (3.4.10)$$

$$h_{II}(t) = \frac{\sqrt{3}}{2}[F_{II}^+(t)A^+(t)F_{II}^\times(t)A^\times(t)] \quad (3.4.11)$$

where $A^+(t)$, $A^\times(t)$ are the same as given in Equations (3.2.2) and (3.2.20) and F_I^+, $F_I^\times$, F_{II}^+, $F_{II}^\times$ are detector beam patterns coefficients given by

$$\begin{aligned}
F_I^+ &= \frac{1}{2}(1+\cos^2\theta_S)\cos 2\phi_S \cos 2\psi - \cos\theta_S \sin 2\phi_S \sin 2\psi \\
F_I^\times &= \frac{1}{2}(1+\cos^2\theta_S)\cos 2\phi_S \sin 2\psi - \cos\theta_S \sin 2\phi_S \cos 2\psi \\
F_{II}^+ &= \frac{1}{2}(1+\cos^2\theta_S)\sin 2\phi_S \cos 2\psi - \cos\theta_S \cos 2\phi_S \sin 2\psi \\
F_{II}^\times &= \frac{1}{2}(1+\cos^2\theta_S)\sin 2\phi_S \sin 2\psi - \cos\theta_S \cos 2\phi_S \cos 2\psi \quad (3.4.12)
\end{aligned}$$

where (θ_S, ϕ_S) is the source sky location in the rotating detector-based coordinate system and ψ is the polarization angle defined by Equation (3.4.1) below. To express the above response in fixed ecliptic-based coordinate system the angles θ_S and ϕ_S are replaced by

$$\cos\theta(t) = \frac{1}{2}\cos\theta_S - \frac{\sqrt{3}}{2}\sin\theta_S[\bar{\phi}_0 + 2\pi(t/T) - \phi_S] \quad (3.4.13)$$

and

$$\begin{aligned}
\phi(t) &= \bar{\alpha}_0 + 2\pi(t/T) + \frac{3\pi}{4} - \lambda \quad (3.4.14) \\
&\quad + \tan^{-1}\left[\frac{\sqrt{3}\cos\theta_S + \sin\theta_S\cos[\bar{\phi}_0 + 2\pi(t/T) - \phi_S]}{2\sin\theta_S \sin[\bar{\phi}_0 + 2\pi(t/T) - \phi_S]}\right]
\end{aligned}$$

where $\lambda = 0,\ \pi/4$ are the orientations of the two detectors, $T = 1$ year and ϕ_0, α_0 are the initial orbital and rotational phase of the detector at $t = 0$. Note that this λ should not be confused with the one that appears in waveform model. The ecliptic-based polarization angle is given in terms of $(\theta_S,\ \phi_S)$ and $(\theta_L,\ \phi_L)$ as following:

$$\psi = \tan^{-1}\left[\frac{A}{B}\right] \tag{3.4.15}$$

where

$$\begin{aligned} A &= \Big[\frac{1}{2}\cos\theta_L - \frac{\sqrt{3}}{2}\sin\theta_L\cos[\bar{\phi_0} + 2\pi(t/T) - \phi_L] \\ &\quad - \cos\theta(t)[\cos\theta_L\cos\theta_S + \sin\theta_L\sin\theta_S\cos(\phi_L - \phi_S)]\Big] \end{aligned} \tag{3.4.16}$$

and

$$\begin{aligned} B &= \Big[\frac{1}{2}\sin\theta_L\sin\theta_S\sin(\phi_L - \phi_S) - \frac{\sqrt{3}}{2}\cos[\bar{\phi_0} + 2\pi(t/T) - \phi_L] \\ &\quad \times[\cos\theta_L\sin\theta_S\sin\phi_S - \cos\theta_S\sin\theta_L\sin\phi_L] \\ &\quad -\frac{\sqrt{3}}{2}\sin[\bar{\phi_0} + 2\pi(t/T)](\cos\theta_S\sin\theta_L\cos\theta_L \\ &\quad - \cos\theta_L\sin\theta_S\cos\phi_S)\Big]. \end{aligned} \tag{3.4.17}$$

In the low frequency approximation, the two polarizations ($+$ and $\times$) are weighted by detector beam patterns instead of being measured individually. In both types of derivation of the LISA response the complement of ecliptic latitude angle θ_S is used i.e. $\theta_S = \pi/2 - \theta_S$, which is then called ecliptic colatitude. The responses $h_I(t)$ and $h_{II}(t)$ are comparable to A and E respectively from full LISA response, however this equivalence is by no means obvious. In [29] it is claimed that "For GW wavelengths much larger than the LISA arm length, $h_I(t)$ and $h_{II}(t)$ coincide with the two 'Michelson variables' in [48], describing the responses of a pair of two-arm 90-degree detectors". However, it is ambiguous as Armstrong et al [48] do not discuss 90-degree equivalent interferometers. Probably Barack and Cutler meant to refer to Cutler's paper [24], or to Prince et al [49] (where A, E, and T are defined). Putting all these confusions aside we found this approximation computationally much cheaper than full LISA response and gives good results in detecting the low frequency EMRI signals and parameter estimation problems.

3.4.2 LISA Noise Modulation

The noise associated with the LISA data has three main components: LISA instrument noise, confusion noise from galactic binaries and confusion noise from extra-galactic binaries. The instrumental noise arises from the internal fluctuations of the detector equipment such as acceleration of test masses, photon shot noise, laser phase noise etc. This noise is unavoidable, i.e. it will be present in the data even if there were no signals at all. The noise from galactic binaries are actually "un-aimed at" signals from numerous NS and WD binaries in our milky way galaxy while extragalactic binaries (binary systems from other galaxies) are also expected to contribute, almost up to the same extent as galactic binaries but at diminished amplitude, to the total LISA noise. Different analytic fits have been proposed from time to time to model these noise sources. The LISA instrument noise is assumed to be stationary and Gaussian, whereas the noise produced by galactic and extragalactic binaries is assumed to be deterministic but coloured because of the evolution of these signals over time [24, 29, 51]. It is predicted that signals from galactic binaries will be dominating the LISA data, making it difficult to estimate an ideal noise spectrum to estimate the parameters of the weaker signals of EMRI sources. One could use "identify and subtract" type algorithm called "gCLEAN" in which first the signals from stronger sources are identified and then subtracted from the data individually [52]. However, the performance of this algorithm depends on the extent of matching between the true and the estimated signals. Subtraction of poorly matched signals may destroy the actual data which can result in changes to the characteristics of the other "sought for" signals. On the other hand, instead of subtraction, the Bayesian approach can be used to characterize the overall noise level [53, 54] by introducing a simple noise model based on the assumption of Gaussianity of the overall noise. Instead of specifying separate models to characterize the "un-accounted" for signals that are present in the data, one can use the estimated residuals based on the current estimates of parameters of the sought for signal to update the noise spectrum in a

Bayesian way. This is the approach which has been utilized here in this work and will be discussed in details in Section 4.5.

CHAPTER 4

METHODS

4.1 The Bayesian Approach

The Bayesian approach usually starts with the specification of a probability model $p(y|\theta)$ which is assumed to represent the process according to which the data $y = \{y_1, ..., y_n\}$ are generated as a function of some unknown parameters $\theta = \{\theta_1, ..., \theta_k\}$, sometimes referred to as the state of nature. Often some information about these unknown parameters is available in advance. One central characteristic of the Bayesian approach is its capability to incorporate this prior information into data analysis in a logical way to enhance the power of the inference. This is achieved by considering the unknown parameters θ as random quantities and thus defining probability densities $p(\theta)$ to describe the *prior information* about these unknown parameters θ, referred to as prior probability distributions or simply *prior distributions*. These prior distributions do not describe the possible variability of the underlying parameters, since θ will typically have definite values, rather they describe the degree of belief or the state of the available information associated with their actual values before the data were being observed [55]. Unlike a frequentist, who defines probability as a long run relative frequency of an event in a process which is repeatable under identical conditions, for a Bayesian the probability theory is simply a numerical quantification of one's state of knowledge or beliefs about the characteristics of the underlying phenomenon [2].

After observing the data y the information contained in it can be used to update the state of our beliefs about θ through the posterior distribution of θ given y using the Bayes' rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{4.1.1}$$

where $p(y|\theta)$ is the conditional density of the data y given the parameters θ or the likelihood function, which describes the probability that y would be the result if θ was the true value of the parameter. Here $p(y)$ represents the total probability of y for all possible values of θ, given by $p(y) = \int_\theta p(y|\theta)p(\theta)d\theta$. Since $p(y)$ does not depend on θ, it is referred to as a normalizing constant and in practice is often ignored in which case Equation (4.1.1) reduces to

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{4.1.2}$$

The posterior density $p(\theta|y)$ can be summarized by finding different statistical quantities such as means, standard deviations and confidence regions using the posterior expectations of the functions of θ. The posterior expectation of a function $h(\theta)$ can be expressed as

$$E[h(\theta)|y] = \frac{\int h(\theta)p(y|\theta)p(\theta)d\theta}{p(y)}. \tag{4.1.3}$$

In practice, apart from some most simple cases, the integrals involved in Equation (4.1.3) are impossible to evaluate analytically. Different analytical approximation methods such as Normal Approximation and Laplace's Method [56] or numerical approximation methods such as different iterative rules, e.g. Newton-Raphson method and Simpson's rule, or Monte Carlo integration methods, e.g. importance sampling and Markov chain Monte Carlo (MCMC) can be used to approximate these integrals. Comprehensive details about many of these methods can be found in any good text on Bayesian statistics or Monte Carlo methods e.g. see [57, 58].

4.2 Monte Carlo Integration

The methods described in the following three subsections, i.e. 4.2.1, 4.2.2 and 4.2.3, are very general. To avoid the unnecessarily Bayesian flavour we use general notations in these subsections, rather than the traditional Bayesian notations.

4.2.1 The Basics

The Monte Carlo approach uses random number generation to compute integrals. The basic idea behind this method is that anything can be learned about a distribution by drawing random samples from it and empirically summarizing those samples. A very basic description is given as follows.

Suppose the random variable X follows a distribution $\Phi(x)$ having probability density function $\pi(x)$ on some general space $\mathcal{X}$ (usually $\mathcal{X} \subseteq \mathbb{R}^d$) and we want to compute

$$E_\pi[h(X)] = \int_{\mathcal{X}} h(x)\pi(x)dx \tag{4.2.1}$$

where $h(X)$ is some function of X. For simplicity let us assume that $\Phi(x)$ is in some standard form so that we can directly draw independent random samples from it: $x_1, ..., x_N \sim \Phi(x)$, then we have

$$\bar{h}_m = \frac{1}{N}\sum_{i=1}^{N} h(x_i). \tag{4.2.2}$$

The strong law of large numbers states that $\bar{h}_m$ converges to $E_\pi[h(X)]$ almost surely as $N \to \infty$, provided $E_\pi[h(X)]$ is finite [59].

4.2.2 Importance Sampling

Importance sampling (see also e.g. [60]) is a general non-iterative Monte Carlo method in which indirect sampling is used for approximating integrals. In Section 4.2.1 we assume that $\Phi(x)$ is an easy-to-sample distribution. Suppose that direct sampling from the $\Phi(x)$ is not possible but another easy-to-sample distribution with density $q(x) \approx \pi(x)$ is available. By introducing $\frac{q(x)}{q(x)}$ into the right hand side part of Equation (4.2.1) we get the following expression,

$$\int_{\mathcal{X}} h(x)\pi(x)dx = \int_{\mathcal{X}} h(x)\frac{\pi(x)}{q(x)}q(x)dx. \tag{4.2.3}$$

By drawing independent random samples $x_1, x_2, ..., x_N \sim q(x)$ we can approximate the above integral by using the following expressions,

$$\bar{h}_m = \frac{\sum_{i=1}^{N} w(x_i)h(x_i)}{\sum_{i=1}^{N} w(x_i)}, \tag{4.2.4}$$

where the factors

$$w(x_i) = \frac{\pi(x_i)}{q(x_i)}$$

are the importance weights and $q(x)$ is the importance sampling density or proposal density . The motivation behind introducing the weight function is to correct the bias introduced by using a different but approximate density $q(x)$. The variance of the estimator $\bar{h}_m$ depends on the choice of $q(x)$. In order to get accurate results, $\pi(x)$ and $q(x)$ need to be reasonably similar, particularly $q(x)$ needs to have longer tails than $\pi(x)$. If $q(x)$ is close to $\pi(x)$ then the weights will be roughly equal and the variance of the estimator will be minimal. On the other hand a poor choice of $q(x)$, particularly with smaller tails than $\pi(x)$, will result in most of the weights being close to zero and giving too much weight to a few values, leading to wrong results.

4.2.3 Markov Chain Monte Carlo Methods

If $\Phi(x)$ does not have a standard closed form or cannot be completely specified or if X is a high dimensional vector, then drawing independent samples from $\Phi(x)$ is not always an easy task. However, for approximating $\Phi(x)$, the fundamental thing is not that we have independent samples from $\Phi(x)$ but that we are able to construct a large collection of dependent samples, $x_1, ..., x_N$, whose empirical distribution approximates $\Phi(x)$. That is, it is possible to relax the independence assumption and think of $x_1, ..., x_N$ as a time series in which the successive samples are dependent in a Markov chain fashion. Markov chains are a special kind of stochastic process that make use of the Markov property which states "given the current state (sample) the next state (sample) does not depend on earlier states (samples)". This implies that is there is at most one step dependency between successive samples. Monte Carlo sampling combined with Markov chains constitutes another class of Monte Carlo methods, for handling the high-dimensional integration problems, known as Markov chain Monte Carlo (MCMC). A Markov chain defined over a general state space $\mathcal{X}$, having a transition kernel density $P(x; y)$ that satisfies $\int P(x; y)dy = 1$, is said to have π as its stationary distribution if for all $x, y \in \mathcal{X}$,

$$\int_{x \in \mathcal{X}} \pi(x)P(x; y)dx = \pi(y). \tag{4.2.5}$$

In order to converge to a stationary distribution the Markov chain must have some properties. Firstly, it must be irreducible, that is regardless of its starting states it must reach any other state in a finite number of steps with a positive probability. Secondly, it must be aperiodic, that is the chain must not favour some particular states or sets of states. This prevents the chain from oscillating between different sets of states in some cycle of fixed length between these sets of states. Thirdly, it must be positive recurrent, that is if the chain has started from a state i, the probability of visiting this state again is 1 with the expected waiting time untill the first return being finite. That is once started from, the state i must certainly

be visited again and one does not have to wait forever. A key notion is the reversibility of the Markov chains. To construct an MCMC sampler, the underlying Markov chain must be reversible. A Markov chain with stationary distribution π, defined over $\mathcal{X}$, is reversible if for all $x, y \in \mathcal{X}$,

$$\pi(x)P(x;y) = \pi(y)P(y;x). \tag{4.2.6}$$

Any irreducible Markov chain with its stationary distribution π satisfying the above equation is positive recurrent and hence ergodic [61, 62]. For a Markov chain constructed in this manner Equation (4.2.2) still holds but now it is referred to as ergodic theorem rather than strong law of large numbers [60]. Equation (4.2.6) is called the detailed balance equation and it is the essential result that most MCMC samplers exploit. MCMC methods have contributed a lot to the vast development of Bayesian data analysis methods. One of the pioneer works in MCMC methods dates back to the paper by Metropolis et al [63], known as the Metropolis algorithm today and is the cornerstone of all MCMC methods. Since then many variants of the basic Metropolis algorithm have been developed in the statistical community [57, 58, 64]. The basic Metropolis algorithm and some of the other most widely used MCMC algorithms are explained in some details in the following sections. As stated before, these algorithms are very general but now we will study them in a Bayesian framework so the notations will be those that suit this framework.

4.2.4 The Metropolis Algorithm

The Metropolis algorithm uses an acceptance-rejection rule to simulate a target distribution by constructing a Markov chain from an auxiliary density which is referred to as proposal density or transition probability function. Let us denote the density of the target distribution by $p(\theta|y)$ and let $q(\cdot|\theta^{(t)})$ be the proposal density which depends on the current state $\theta^{(t)}$. For the basic Metropolis algorithm, $q(\cdot|\cdot)$ must be symmetric, i.e. $q(\theta|\theta') = q(\theta'|\theta)$. Being at current state $\theta^{(t)}$ the acceptance probability

for moving to a new state θ' is defined as:

$$\alpha = \min\left\{1, \frac{p(\theta'|y)}{p(\theta^{(t)}|y)}\right\}. \tag{4.2.7}$$

The Markov chain can then be constructed using the following steps:

Initialize θ_0; $t \leftarrow 0$
Repeat {
Sample $\theta' \sim q(\cdot|\theta^{(t)})$
Compute α using Equation (4.2.7)
Sample $u \sim \text{Uniform}(0, 1)$
If $u \leq \alpha$ then $\theta^{(t+1)} \leftarrow \theta'$
else $\theta^{(t+1)} \leftarrow \theta^{(t)}$
$t \leftarrow (t + 1)$
}

This will result in a Markov chain because the transition probabilities $q(\cdot|\cdot)$ of moving from $\theta^{(t)}$ to $\theta^{(t+1)}$ depend only on $\theta^{(t)}$ and not beyond it. Furthermore, the transition probabilities satisfy the so-called detailed balance equation with respect to the target distribution, which guarantees the convergence of an irreducible Markov chain to its stationary distribution. Having jogged for a sufficient number of iterations, the *burn-in* period, the chain will eventually forget its history and will converge to its stationary distribution which, theoretically, is the specified target distribution. The above algorithm can be used to generate virtual random samples from any target distribution known up to a normalizing constant, regardless of its dimensionality and its analytical complexity [64]. The steps' sizes are controlled by the scale of $q(\cdot|\theta^{(t)})$ and will be discussed in a later section.

4.2.5 The Metropolis-Hastings Algorithm

The Metropolis algorithm is useful when nothing is known about the target distribution since it uses a symmetric proposal distribution to gen-

erate the Markov chain [64]. However, if some information is available about the shape and orientation of the target distribution then one can set up the proposal distribution taking into account that information. The Metropolis-Hastings algorithm [65] is a generalization of the original Metropolis algorithm in which the proposal distribution does not need to be symmetric. The acceptance probability is then given by:

$$\alpha = \min\left\{1, \frac{p(\theta'|y)q(\theta^{(t)}|\theta')}{p(\theta^{(t)}|y)q(\theta'|\theta^{(t)})}\right\} \tag{4.2.8}$$

The detailed balance equation can be easily established for Metropolis-Hastings chain for almost any proposal density, thus the Markov chain generated this way has the specified target distribution as its stationary distribution. If the chain is irreducible and aperiodic, after a sufficient number of steps it will start delivering samples from the target distribution [60].

4.2.6 A Special Case: Random Walk Metropolis

For ease of implementation of the Metropolis algorithm, particularly in multi-parameter target density problems, it is useful to think of it as a special case of a random walk scheme whose evolution is controlled by the Metropolis acceptance-rejection rule. Being at the current state $\theta^{(t)}$ an increment $\epsilon^{(t)}$ is proposed so that the next state becomes $\theta' = \theta^{(t)} + \epsilon^{(t)}$, where $\epsilon^{(t)} \overset{iid}{\sim} q(0, \cdot)$ for different t. That is a random quantity $\epsilon^{(t)}$, drawn from $q(0, \cdot)$, is added to the current state $\theta^{(t)}$ to displace it to a new state θ'. The newly proposed state is then accepted or rejected according to the usual Metropolis rule. When nothing is known about the shape of the target distribution a natural choice for $q(0, \cdot)$ would be to use some symmetric distribution around zero such as Gaussian distribution $N(0, \Sigma)$ or a symmetric around zero uniform distribution $U(-c, c)$ [64, 66]. The algorithm can be outlined as follows:

Initialize θ^0; $t \leftarrow 0$
Repeat {
Sample $\epsilon' \sim q(0, \cdot)$
Set $\theta' = \theta^{(t)} + \epsilon'$
Compute α using Equation (4.2.7)
Sample $u \sim \text{Uniform}(0, 1)$
If $u \leq \alpha$ then $\theta^{(t+1)} \leftarrow \theta'$
else $\theta^{(t+1)} \leftarrow \theta^{(t)}$
$t \leftarrow (t + 1)$
}

The same modifications apply to the Metropolis-Hastings algorithm too where the proposal density $q(0, \cdot)$ is asymmetric.

4.2.7 The Gibbs Sampling

Another widely used MCMC algorithm is the Gibbs sampler [67], where the proposal distribution is defined in terms of the conditional distributions of the target distribution. The algorithm is defined for distributions depending on at least two parameters. Suppose the parameter vector θ can be decomposed into d $(d \geq 2)$ components or sub-vectors and the target density $p(\theta|y)$ can be factorized into full-conditional distributions $p(\theta_i|\theta_{-i})$, $i = 1, ..., d$ for each component in θ. For simplicity let us assume that all these full-conditional distributions are in some standard form that can be easily sampled from. Setting $p(\theta_i|\theta_{-i})$ as proposal distribution for θ_i, the general Gibbs sampling scheme can be constructed as follows.

Initialize $\theta^{(0)} = (\theta_1^{(0)}, ..., \theta_d^{(0)})$; $t \leftarrow 0$
Repeat {
Draw $\theta_1^{(t+1)} \sim p(\theta_1 | \theta_2^{(t)}, \theta_3^{(t)}, ..., \theta_d^{(t)}, y)$
Draw $\theta_2^{(t+1)} \sim p(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, ..., \theta_d^{(t)}, y)$
$\vdots$
Draw $\theta_d^{(t+1)} \sim p(\theta_d | \theta_1^{(t+1)}, \theta_2^{(t+1)}, ..., \theta_{d-1}^{(t+1)}, y)$
$t \leftarrow (t+1)$
}

Within each iteration there is a cycle of d steps and at each step a draw θ_i is proposed from its conditional distribution, given all other components of θ and thus each component of θ is updated using the latest values of the other components of θ drawn from their respective conditional distributions already updated in the previous steps. In cases where direct sampling from some of the conditional posterior distributions is not possible, Metropolis steps can be introduced to draw samples from those distributions. Thus some parameters are updated using the Gibbs sampler while the others are updated using Metropolis steps, leading to another algorithm namely *Metropolis-within-Gibbs* algorithm [66].

4.2.8 Implementation of Metropolis(-Hastings) Algorithm

In practice, there arise several issues when implementing the Metropolis(-Hastings) algorithm or any other MCMC method. The performance of any MCMC method is judged by the speed of its convergence to the underlying stationary distribution and the mixing of the resulting Markov chain. Some of the key issues and related remedies are discussed as follows.

4.2.8.1 The Choice of Proposal Distribution

The efficiency of a Metropolis(-Hastings) algorithm strongly depends on the choice of the proposal distribution $q(\cdot|\cdot)$. The first and the most

essential property of a good proposal distribution is that it should be easy to sample from. Although, theoretically any proposal distribution will eventually start sampling from the target distribution $p(\theta|y)$; however, in practice the functional complexity of $p(\theta|y)$ may have a potential impact on the convergence of the chain in a reasonable time [58]. An ideal proposal distribution would be the target distribution itself since then nearly-independent samples can be drawn with almost sure acceptance. Thus, the rate of convergence crucially depends on the similarity of the $q(\cdot|\cdot)$ and $p(\theta|y)$. Another issue related to the choice of proposal distribution is the mixing of the MCMC chain, which reflects the degree of inter-dependence or correlation between subsequent samples. Mixing, in general, depends on the scale (variance) of the proposal distribution. A good proposal distribution would lead to a small correlation of subsequent samples $\theta^{(t)}$ and $\theta^{(t+1)}$ with a reasonable acceptance rate. This correlation can occur in two ways. If the variance of the proposal distribution is set too small, the acceptance rate will be high; however, the subsequent samples of the resulting Markov chain will be highly correlated, as $\theta^{(t)} \approx \theta^{(t+1)}$. On the other hand, a large variance would lead to a very low acceptance rate and again a high correlation because then for most of the iterations, the chain will be stuck at $\theta^{(t)} = \theta^{(t+1)}$. The optimal choice of a proposal distribution is the one which establishes a compromise between these two competitive concepts: a good mixing to ensure low correlation between subsequent samples with a reasonable acceptance probability.

In the literature, there are different suggestions for the choice of proposal distributions. However, these should be regarded as rules of thumb only and may not be used as hardcore strategies in general. For example, when sampling from a uni-variate standard Normal target density $N(0,1)$, a normal proposal distribution with mean as the current state and variance ≈ 2.4, which is somewhat over-dispersed version of the target distribution, was found to give best results. These results were further general-

ized to d-dimensional Normal densities with known covariance matrix Σ, for which the proposal distribution with $\approx \frac{2.4}{\sqrt{d}}\Sigma$ was found to be an optimal choice [68]. In the same reference the optimal acceptance rates were also found to be $\approx 44\%$ for $d = 1$ which declines to $\approx 23\%$ as $d \to \infty$. Several other suggestions regarding the optimal scaling of the proposal distributions to ensure optimal acceptance rates can be found in literature; for example see [64, 66, 69]. However, all these choices are somehow case specific, based on some available information and thus cannot be argued for general applications when little is known about the target distribution.

4.2.8.2 Initialization of MCMC

For the efficient implementation of Metropolis(-Hastings) or any other MCMC sampler, another important issue is the choice of starting values. Though in principle, within the legitimate range (e.g. satisfying $p(\theta_0|y) \geq 0$), whatever the starting values are, the chain will eventually approach its stationarity. However, a poor choice of starting value can result in an extremely long burn-in period, typically impossible to finish in a finite time. For simpler models a good starting value can be a point near the distribution centre, such as the mode of the distribution which can be approximated by using maximum likelihood or multiple over-dispersed starting values can used to see as to which starting value results in faster convergence. In general, there is no hard rule available but several suggestions for choosing good starting values can be found in literature, for example see [70].

4.2.8.3 Re-parametrization

In multi-dimensional problems the structure of the target distribution often leads to highly dependent components of the parameter vector that can result in very poor mixing and slow convergence. Adequate reconstruction of the posterior by clever re-parametrization may produce near independence between components of parameter vector which may

improve convergence to a great extent. The re-parametrization is usually done by *transformation of random variable* techniques. If the inverse-transformation is unique, then the original variables can be recovered by back transforming the values of the MCMC output [58].

4.2.9 Convergence Assessment

The convergence of the MCMC chain can be assessed by different theoretical methods and approximations using the MCMC output. The simplest one is to just look at the history plot or trace plot of the MCMC samples. To see whether the chain is really converging to its true stationary distribution, in other words to determine an approximately true burn-in period, multiple chains with different starting distributions can be run either in parallel or individually, and see if all the chains converge to the same stationary distribution and which chain converges faster. More sophisticated methods such as Gelman and Rubin diagnostic which computes *potential scale reduction factor* (PSRF) for each chain to assess the individual convergence of different chains to the same stationary distribution. PRSF is an analysis of variance technique which uses the within chain variances and between chains variance to assess the convergence of a particular chain. For m MCMC chains each of length n and labeling the MCMC draws as θ_{ij} $(i = 1, ..., n; j = 1, ..., m)$, the PRSF denoted by $\widehat{R}$ is calculated as

$$\widehat{R} = \sqrt{\frac{\hat{\sigma}_+^2}{\sigma_W^2}} \tag{4.2.9}$$

where

$$\hat{\sigma}_+^2 = \frac{n-1}{n}\sigma_W^2 + \frac{1}{n}\sigma_B^2 \tag{4.2.10}$$

where σ_W^2 is within chain variance and σ_B^2 between chains variance. $\widehat{R}$ values larger than 1 indicate non-convergence. For multivariate problems an extended version of $\widehat{R}$, *multivariate potential scale reduction factor* (MPSRF); $\widehat{R}^p$ can be used [71]. There are several other diagnostic tools such as auto-

correlation plots, Geweke diagnostics [72] and Heidel diagnostics [73–75]. All these methods can be implemented using an R package called CODA (Convergence Diagnostics and Output Analysis) [76], which is available from the project website of R [77].

4.3 Convergence Acceleration

4.3.1 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 longer time before reaching the global mode. This not only results in poor mixing but also there are chances that the sampler does not even reach the global mode in an affordable sampling time or number of iterations. Some kind of bridging is needed to link parts of the target density that are separated by low probability regions [64]. One may think of proposing larger steps to help the MCMC chain to jump from one mode to the other easily, but it is not feasible since then the acceptance rate of the proposed steps can be too small. One way to deal with such situations is to use tempering methods, that actually use the principle of annealing: a physical process which is used to obtain perfect crystallization by heating and then cooling at a slow enough rate to give atoms sufficient time to attain the lowest energy state or the global optimum. Sufficiently high temperature causes atoms to move more freely with increased randomness and slow enough cooling causes improved equilibrium, that is the atoms get enough time to line themselves up in positions that corresponds to the global energy minimum of a perfect crystal. Analogous to this physical process, temperature is used to scale the target density in order to flatten local modes so that the MCMC sampler moves freely towards the global maximum without being attracted by lo-

cal modes. At a given temperature $T \geq 1$, samples are generated from a tempered version of the target density $p(\theta|y)$ defined as:

$$p_T(\theta|y) \propto p(\theta|y)^{\frac{1}{T}} \tag{4.3.1}$$

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

In the case of the Metropolis algorithm, the acceptance probability for the above tempered target distribution is defined as

$$\begin{aligned} \alpha_T &= \min\left\{1, \frac{p(\theta'|y)^{\frac{1}{T}}}{p(\theta^{(t)}|y)^{\frac{1}{T}}}\right\} \\ &= \min\left\{1, \left(\frac{p(\theta'|y)}{p(\theta^{(t)}|y)}\right)^{\frac{1}{T}}\right\} \end{aligned} \tag{4.3.2}$$

where $\theta^{(t)}$ and θ' are the current and proposed states respectively. Furthermore, one can also apply tempering only to the likelihood part of the posterior and leave the prior part unchanged, in which case the above equation changes to

$$\alpha_T = \min\left\{1, \frac{p(\theta')}{p(\theta^{(t)})}\left(\frac{p(y|\theta')}{p(y|\theta^{(t)})}\right)^{\frac{1}{T}}\right\}. \tag{4.3.3}$$

This convention is motivated by the fact that larger temperatures may destroy the properness property of the posterior distribution [78], because in its actual shape a posterior is proper as long as the prior is proper [57]. In this case, for higher temperatures the posterior becomes increasingly similar to the prior distribution and eventually for $T \to \infty$, the posterior and prior distributions become equal [54]. Some of the most popular temper-

ing methods are simulated annealing (SA) [79], simulated tempering (ST) [80, 81] and parallel tempering MCMC [82]. All these methods work in the same spirit, i.e. using modified target distributions (tempered version) but are different in their applications. SA and ST work by defining a series of distributions characterized by different temperatures varying between the cold one (the actual target distribution) and the hottest one. A single MCMC chain is run, which starts simulating the hottest distribution and gradually progresses towards the cold distribution moving through the tempered distributions in between either systematically (SA) or stochastically (ST). Parallel tempering MCMC is similar to ST but it simulates all the distributions (i.e. the cold one and tempered ones) simultaneously by running multiple inter-communicative MCMC chains in parallel. In the following sections we explain parallel tempering MCMC and its properties in details.

4.3.2 Metropolis-Coupled MCMC

The Metropolis coupled-MCMC (MCMCMC) [82], also known as parallel tempering MCMC (PTMCMC), is a powerful optimization of the simple Metropolis-(Hastings) algorithm which is very effective in improving the mixing of MCMC and in particular in escaping the local modes. The algorithm works by running multiple MCMC chains, each simulating a separate target density characterized by a different temperature, in parallel and occasionally attempting swaps of its current states with other chains. In principle, a high temperature chain sees the target density as more *flattened* relative to a low temperature chain, which means that high temperature chain can move more freely across the valleys of low probability regions in between of modes [64]. In order to make the low temperature chain get benefit from a high temperature chain that may be sampling near another mode, an exchange of information about the current states of the two chains is attempted by proposing a swap of the current states using an additional MCMC Metropolis acceptance-rejection

step. The algorithm is described as below:

Let there be I chains corresponding to I distant temperature levels $1 = T_1 < T_2 < ... < T_I$, where $T_1 = 1$ is the temperature level corresponding to the true or cold chain. For i^{th} chain, given the current state $\theta^{(t)}$ in the parameter space the Metropolis probability of acceptance of moving to a proposed state θ' is given by

$$\alpha_i = \min\left\{1, \left(\frac{p_i(\theta'|y)}{p_i(\theta^{(t)}|y)}\right)^{\frac{1}{T_i}}\right\}. \tag{4.3.4}$$

Further suppose that at a particular iteration t a swap between states of chains i and j, with $j = i + 1 \leq I$, is proposed then the Metropolis acceptance probability of this swap is

$$\alpha_{ij} = \min\left\{1, \left(\frac{p_j(\theta^{(t)}|y)}{p_i(\theta^{(t)}|y)}\right)^{\left(\frac{1}{T_i} - \frac{1}{T_j}\right)}\right\}. \tag{4.3.5}$$

The general form of the algorithm with a pre-defined temperature ladder with steps $1 = T_1 < T_2 < ... < T_I$, and the acceptance probabilities defined by Equations (4.3.4) and (4.3.5), is outlined as follows:

- Step 1: To generate the ith chain, $i \in (1, ..., I)$ the usual Metropolis sampler is applied with acceptance probability of a proposed move defined by (4.3.4).

  Initialize θ_i^0; $t \leftarrow 0$
  Repeat {
  Sample $\epsilon_i' \sim q(0, \cdot)$
  Set $\theta_i' = \theta_i^{(t)} + \epsilon_i'$
  Compute α using Equation (4.3.2)
  Sample $u_i \sim \text{Uniform}(0, 1)$
  If $u_i \leq \alpha_i$ then $\theta_i^{(t+1)} \leftarrow \theta_i'$

else $\theta_i^{(t+1)} \leftarrow \theta_i^{(t)}$
$t \leftarrow (t+1)$

- Step 2: After each chain is run for a given number of iterations, any two chains i and j are randomly selected to exchange states. The swap of states is accepted with probability α_{ij} (see Equation (4.3.5)). (Note that swapping can also be proposed at each iteration.)

- Step 3: Similar to a regular Metropolis step, a random number $u \sim \text{Uniform}(0,1)$ is generated and if $u \leq \alpha_{ij}$, then the proposed swap is accepted and chains i and j exchange states.

- Step 4: Go to "Repeat" in step 1.

It is easy to imagine that if a colder chain is stuck on some local mode, a hotter chain can pull it out by swapping its states. Thus the colder chain can now easily move through the low probability regions between modes with improved mixing [64]. Note that the same comments about applying temperature only to the likelihood part, as discussed at the end of the previous section, apply here as well. Moreover, since the temperature is applied to the Metropolis probability only, the actual proposals and posterior copies are unaffected; however, normally only the samples from the chain with $T = 1$ for the Metropolis probability are considered for the subsequent analysis and the samples from the other chains are discarded [58]. The local detailed balance equation holds for all chains (actual and tempered), thus if the chains are irreducible and aperiodic then each chain has the respective target distribution (actual and tempered) as its stationary distribution [83, 84]. In [85] it is shown that all the swapped states at a given iteration (regular Metropolis state) constitutes a Markov chain with its stationary distribution as the actual target distribution.

4.3.2.1 Computing Requirements

PTMCMC is computationally expensive and cannot be implemented efficiently using single processor/core machines since the amount of execution time increases linearly with the number of chains because all chains are independent Metropolis(-Hastings) samplers and each will require the same amount of computing power as all the others. However, the availability of high performance computers with multiple multi-core processors, sharing the same memory, has now solved this problem to a large extent. To parallelise a task on a multi-core machine or high-performance cluster, parallel computing techniques are required. There are several kinds of parallelization, one of which is to run multiple copies of the same code, though with slightly different parameters, simultaneously on independent but communicating processors. This is what a PTMCMC needs for its efficient implementation. For such a communicative parallelization a widely used protocol is Message Passing interface (MPI) [86, 87]. MPI can be implemented using a public domain software called *OpenMPI* [88], which specifies a library of functions and macros that implement the message passing model of the parallel computing.

4.3.2.2 Optimal Temperature Ladder

When constructing a temperature ladder some considerations need to be taken into account. For good mixing, one needs to make sure that ladder steps are such that there is sufficient overlap between any two neighbouring densities so that their respective chains can swap states at a reasonable rate. It depends on the difference or ratio between the temperatures of the two neighbouring densities. If this difference or ratio is too large then the two neighbouring densities will be quite different from each other and their respective chains will rarely swap states and vice versa. In both extreme cases, this defeats the purpose of heating. Also in general, the tar-

get density behaves differently at the two ends of the temperature ladder. At the start (low temperatures) even very small differences in the ladder steps can result in significant changes in the surface of the target density. These changes decrease as the temperature increases and at the end very large differences in the ladder steps can result in relatively lesser changes in the target density. For a linear temperature ladder the swap acceptance rates at the end of the ladder are generally too high as compared to the starting steps. Thus a good temperature ladder would be something like a geometric or power sequence to get approximately even swap acceptance rates throughout for all pairs of neighbouring chains. There are several suggestions for the construction of temperature ladders for PTMCMC algorithm. For example in his introductory paper Geyer [82] used an incremental rule defined as $T_i = 1 + \Delta T(i-1)$ to specify the temperature for i^{th} chain, where ΔT parameter is the incremental parameter. In some cases a ladder with geometric increments, i.e. $T_{i+1} = \frac{T_i}{c}$ can also be used, where $T_1 = 1$ and $c > 1$ is the tuning parameter. Another version of geometric ladder defined as $T_i = \lambda^{(i-1)}$, where $\lambda > 1$ is a tuning parameter, was suggested in [54, 89]. Regarding the optimal scaling of temperature steps it has been found that the optimal temperature steps would be those for which the swap acceptance rates for all individual pairs of neighbouring chains are $\approx 23\%$. The rules given in [54, 82, 89] do not in general guarantee a $\approx 23\%$ swap acceptance rates for all individual pairs of neighbouring chains, rather these rules can be used in way so that the average acceptance rate is $\approx$ 23% for all pairs. This is because this swap probability not only depends on the temperature differences but also on the current states of parameters θ in different chains, at regular Metropolis steps, that may result in a large difference between the target densities of two neighbouring chains and hence affecting the swap acceptance probability. Recently a Monte Carlo based swapping rule was proposed which maintains $\approx$ 23% swap acceptance rates for individual pairs; however, this rule is complicated and its applications to general problems are not yet explored [90]. The length of the ladder or the number of steps in the ladder depends on the size of available computing

resources.

4.3.2.3 Updating the Proposal Covariance Matrix in PTMCMC

When doing MCMC analysis, to improve efficiency it is a good practice to update the covariance matrix of the proposal distribution either by using the first few thousands MCMC samples of the parameters or by updating it after a specified number of iterations throughout the MCMC run. Some adaptive techniques for the optimal scaling of the proposal covariance matrix are explored in [69, 91–93]. However, one should be careful when employing these techniques in PTMCMC. In PTMCMC different chains explore different regions of the parameter space. The swaps of states between the neighbouring chains are accepted on the basis of their likelihood values no matter what the current parameters' values are. When a global maximum is not found yet, the likelihood values of two neighbouring chains can still be similar even if they are sampling from quite different regions of the parameter space. Thus it is possible that some of the corresponding parameters' values of the two chains are too far from each other but their respective chains would still exchange states because of their close likelihood values. This will improve mixing but on the other hand will result in a large variability in the MCMC samples of those parameters. If these highly dispersed MCMC samples are used to estimate the empirical proposal covariance matrix, the resulting covariance matrix will bring down the overall acceptance rate of the basic MCMC sampler to a very low level. The best way to do this is to halt the swaping mechanism for a specified of number of iterations, store those initial MCMC samples either internally (in memory) or externally (in a file), and then compute the empirical covariance matrix for each chain individually or just compute one from the first chain and use it for all chains.

4.4 Digital Signal Processing Methods

4.4.1 Fourier Transform

The Fourier transform (FT) defines a relationship between time and frequency domains. A physical process can be represented either in the time domain, as a function of time t, or else in the frequency domain, as a function of frequency f, and the bridge between the two domains is the FT. Let $y(t)$ be a real valued function of a real line $y : \mathbb{R} \to \mathbb{R}$ then the Fourier transform is a mapping $\tilde{y} : \mathbb{R} \to \mathbb{C}$ defined as:

$$\tilde{y}(f) = \int_{-\infty}^{\infty} y(t) \exp(-2\pi i f t) dt \tag{4.4.1}$$

where $i = \sqrt{-1}$. The integral transforms a function of time $y(t)$ to a function of frequency $\tilde{y}(f)$. If the FT $\tilde{y}(f)$ exists, the inverse Fourier transform takes us back to the time domain,

$$y(t) = \int_{-\infty}^{\infty} \tilde{y}(f) \exp(2\pi i f t) df. \tag{4.4.2}$$

The FT $\tilde{y}(f)$ exists if $y(t)$ is absolutely integrable that is $\int_{-\infty}^{\infty} y(t)dt < \infty$. If both $y(t)$ and $\tilde{y}(f)$ are continuous and absolutely integrable then the inverse FT of $\tilde{y}(f)$ is equal to $y(t)$. Equations (4.4.1) and (4.4.2) are called FT pair, denoted by $y(t) \overset{FT}{\rightleftharpoons} \tilde{y}(f)$. Two important properties of FT, that are relevant in later sections, are as follows:
Let $[x(t) \overset{FT}{\rightleftharpoons} \tilde{x}(f)]$ and $[y(t) \overset{FT}{\rightleftharpoons} \tilde{y}(f)]$ be the FT pairs and assume that the convolution of $x(t)$ and $y(t)$ (in time domain) exists:

$$(x * y)(t) = \int_{-\infty}^{+\infty} x(\tau) y(\tau - t) d\tau < \infty \tag{4.4.3}$$

then the following properties hold:

- **Linearity** : $ax(t) + by(t) \overset{FT}{\rightleftharpoons} a\tilde{x}(f) + b\tilde{y}(f)$

- **Convolution Theorem** : $(x * y)(t) \overset{FT}{\rightleftharpoons} \tilde{x}(f)\tilde{y}(f)$

Another important property is if $y(t)$ is real then $\tilde{y}(f) = \tilde{y}(-f)$ [94]. There are several other properties that can be found in any good textbook.

4.4.2 Discrete Fourier Transform

In practical applications, the FT is approximated by its discrete analogue called Discrete Fourier Transform (DFT). The DFT computes a discrete frequency function from a discrete time signal of finite length, i.e. here both the time and frequency domains are discretized.
Suppose $\{y(t_k)\}_{k=0,...,(N-1)}$ are the sampled values of $y(t)$ at equally spaced time points $t_k = k\Delta_t$ over a finite duration $N\Delta_t$. The DFT of $y(t_k)$ at discretized frequencies $f_n = \frac{n}{N\Delta_t}$, $n = 0, ..., N-1$, is defined by approximating the integral in Equation (4.4.2) by a discrete sum:

$$\begin{aligned} \tilde{y}(f_n) &= \int_{-\infty}^{\infty} y(t) \exp(-2\pi i t f_n) dt \\ &\approx \Delta_t \sum_{k=0}^{N-1} y(t_k) \exp(-2\pi i t_k f_n) \\ &= \Delta_t \sum_{k=0}^{N-1} y(t_k) \exp(-2\pi i k n / N). \end{aligned} \tag{4.4.4}$$

The summation term in the last equation is the DFT of $y(t_k)$. For simplicity in notations, let us denote it by $\tilde{y}_n$ thus we get

$$\tilde{y}_n = \sum_{k=0}^{N-1} y(t_k) \exp(-2\pi i k n / N) \tag{4.4.5}$$

this implies

$$\tilde{y}(f_n) = \Delta_t \tilde{y}_n \tag{4.4.6}$$

[4, 94]. When $y(t)$ is real then $\tilde{y}_n = \tilde{y}^*_{N-n}$, i.e. the complex conjugation symmetry holds. Furthermore, the element $\tilde{y}_0$ is always purely real and for even N the element $\tilde{y}_{N/2}$ is also real. When using the DFT for spectral

density estimation these extreme terms are often omitted for simplicity purposes and if included then they should be treated differently as will become clear in later sections. To back transform to time domain the formula is given as

$$y(t_k) = \frac{1}{N} \sum_{n=0}^{N-1} \tilde{y}_n \exp(2\pi i k n / N). \tag{4.4.7}$$

In practice, the DFT transforms are calculated using a Fast Fourier Transform (FFT) algorithm [94]. This is implemented through a freely available software package *Fast Fourier Transform in the West* (FFTW) [95].

4.4.3 Power Spectral Density

According to a result based on Parseval's theorem the total power in a square integrable signal is the same regardless of its domain of measurement. That is

$$\text{Total Power} = \int_{-\infty}^{\infty} |y(t)|^2 \, dt = \int_{-\infty}^{\infty} |\tilde{y}(f)|^2 \, df. \tag{4.4.8}$$

The above result leads to the following definition

$$S(f) = |\tilde{y}(f)|^2 + |\tilde{y}(-f)|^2 \quad -\infty < f < \infty. \tag{4.4.9}$$

Which is known as the *power spectral density* (PSD). Since for real $y(t)$ we have $\tilde{y}(f) = \tilde{y}(-f)$ (symmetry property), therefore Equation (4.4.9) can be rewritten as

$$S(f) = 2|\tilde{y}(f)|^2 \qquad 0 < f < \infty \tag{4.4.10}$$

which is called the *one sided* PSD. It is easier to understand that PSD is just the absolute square of the FT of the signal [94].
The discrete PSD can be defined using the discrete form of Parseval's the-

orem. The discrete form of Parseval's theorem is

$$\text{Total Power} = \sum_{k=0}^{N-1} |y(t_k)|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |\tilde{y}_n|^2 \tag{4.4.11}$$

Hence the one sided PSD is

$$\begin{aligned} S_n(f_0) &= \frac{1}{N}|\tilde{y}_0|^2 \\ S_n(f_n) &= \frac{1}{N}\left[|\tilde{y}_n|^2 + |\tilde{y}_{N-n}|^2\right] \quad \text{for } n = 1, ..., \frac{N}{2} - 1 \\ &= \frac{2}{N}|\tilde{y}_n|^2 \quad \because \text{ for real } y(t),\ \tilde{y}_n = \tilde{y}_{N-n} \\ S_n(f_{N/2}) &= \frac{1}{N}|\tilde{y}_{N/2}|^2. \end{aligned} \tag{4.4.12}$$

As can be seen in the above equations, the elements of $S_n(f_n)$ corresponding to end frequencies are different with respect to the multiplier because of the zero imaginary components.

4.4.3.1 The Welch's Method

In practice, to estimate the PSD of a very long signal, Welch's method [96] is used. Instead of using all data at once, Welch's method divides the data into shorter, possibly overlapping segments, applies a suitable window (see next section) to each segment before being Fourier transformed, averages the transforms, and takes the squared magnitude to produce a power spectrum. Welch's method has three advantages; it is faster, the averaging reduces the variance of the power spectrum and at the same time windowing avoids spectral leakage and smooths (see also Section 4.4.4) the estimate. Suppose the sampled data $\{y(t)\}_{t=0,...,N-1}$ can be divided into K segments each of length L (L even), such that $(K-1)D + L = N$ where D is number of data points by which two consecutive segments overlap, i.e. their starting points are D units apart. If this overlap is set to be 50% then $D = \frac{L}{2}$ in which case $(K+1)\frac{L}{2} = N$. Let us denote these segments by $\{y_j(t)\}_{j=1,...,K;\ t=0,...,L-1}$. This overlapping segmentation is illustrated in

Figure 4.1.

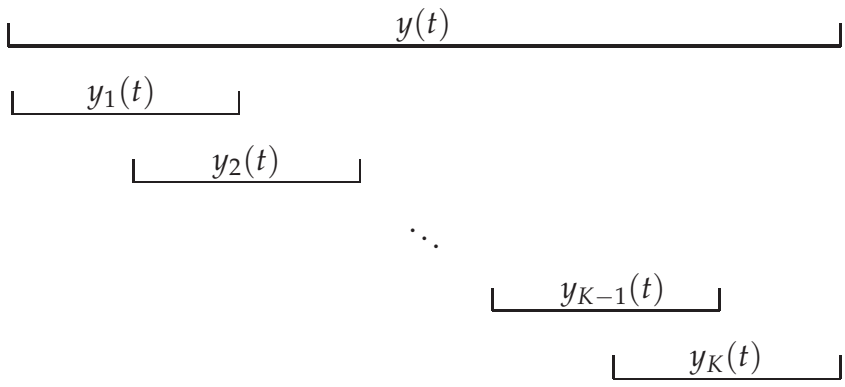

Figure 4.1: Illustration of overlapping segmentation.

With an appropriate data window $w(t)$ of length L, each segment $y_j(t)$ is windowed, i.e. $z_j(t) = y_j(t) \cdot w(t)$, before being Fourier transformed. For j^{th} windowed segment $z_j(t)$, the power spectral density is obtained as following.

$$\begin{aligned} S_j(f_0) &= \frac{1}{w_{ss}}|\tilde{z}_{j0})|^2 \\ S_j(f_n) &= \frac{1}{w_{ss}}\left[|\tilde{z}_{jn}|^2 + |\tilde{z}_{j(N-n)}|^2\right] \quad \text{for } n = 1, ..., \frac{N}{2} - 1 \\ &= \frac{2}{w_{ss}}|\tilde{z}_{jn}|^2 \quad \because \text{ for real } z_j(t),\ \tilde{z}_{jn} = \tilde{z}_{j(N-n)} \\ S_j(f_{N/2}) &= \frac{1}{w_{ss}}|\tilde{z}_{j(N/2)}|^2 \end{aligned} \tag{4.4.13}$$

where $w_{ss} = \frac{1}{L}\sum_{t=0}^{L-1} w^2(t)$. The additional subscript j is now running for jth segment. The final estimate is then just the average of these K spectral densities given as:

$$\widehat{S}_n(f) = \frac{1}{K}\sum_{j=1}^{K} S_j(f). \tag{4.4.14}$$

Note that the term N appearing in the denominator of Equation (4.4.12) is now replaced by w_{ss}, otherwise the formulas are the same [4, 94, 96].

4.4.4 Issues with DFT

Several issues arise when using the DFT to estimate the power spectral density. Two key issues are discussed as follows.

4.4.4.1 Nyquist Criteria and Aliasing

In order to correctly approximate the FT of a continuous function $y(t)$, which is sampled at a discrete time interval Δ_t one must sample at a frequency well below the Nyquist critical frequency, $f_c = \frac{1}{2\Delta_t}$, which is the highest frequency that can be resolved using Δ_t. If the continuous function $y(t)$ is band-limited to frequencies smaller than f_c, i.e. $\tilde{y}(f) = 0$ for all $|f| > f_c$, then the function $y(t)$ can be uniquely determined from its sampled values. If, on the other hand, the function is not band limited to less than Nyquist critical frequency then it will give rise to aliasing error, in which case its spectrum will have overlapping tails because the power outside $-f_c < f < f_c$ will move inside the range [94]. Thus to avoid aliasing one must sample at least two points per cycle.

4.4.4.2 Spectral Leakage and Windowing

In practice, a theoretically infinite signal can be measured for a finite time only. That is the actual signal is truncated to a finite stretch of data points. The DFT computation implicitly assumes that the observed signal is periodic, i.e. it repeats over and over for all time. If the number of periods in the measured signal is not an integer value then it will cause discontinuities at the end-points of the signal. When the DFT is applied to such a signal then the resulting frequency spectrum suffers from spectral leakage. Spectral leakage results in the signal energy spreading out over a wide frequency range in the DFT when it should be concentrated at only one frequency [97].

Theoretically the measured signal, $y(n)$, is equivalent to multiplying the

actual signal, $x(n)$, by a function called a *rectangular data window*, $w_R(n)$. That is,

$$y(n) = x(n) \cdot w_R(n) \tag{4.4.15}$$

where

$$w_R(n) = \begin{cases} 1 & \text{for} \quad 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \tag{4.4.16}$$

According to convolution theorem a product in time-domain is equal to convolution in frequency domain that is when we apply DFT to Equation (4.4.15) then we get

$$\tilde{y}(f) = \tilde{x}(f) * \tilde{w}_R(f) \tag{4.4.17}$$

where $\tilde{w}_R(f)$ is the DFT of rectangular window now called rectangular pulse and is well defined by the following expression.

$$\tilde{w}_R(f) = T\frac{\sin(\pi f N T)}{\sin(\pi f T)} \exp(-2\pi i f k n/N). \tag{4.4.18}$$

Thus in general, due to the convolution in Equation (4.4.17), the DFT of $y(n)$ is not equal to that of $x(n)$, which results in spectral leakage as explained above. In real applications spectral leakage is unavoidable; however, its effects can be reduced by using a different window function, whose value is 1 for middle point of the data and tappers to 0 at the end points $n = 0$ and $n = N-1$, hence avoiding discontinuities at the end points. However, there is a drawback of tapered windows too: the reduction in leakage comes at the costs of reduced resolution thus making it difficult to resolve spectral components that are close in frequency [98]. Thus in general a compromise between these two conflicting requirements is needed. Several different tapered windows can be found in literature, each one with its own merits and demerits relative to others. The choice of a particular window depends on the nature of the signal and the purpose of the analysis. Some windows are more effective in making the detection of the exact frequency of a peak in the spectrum easier. Some improve the amplitude accuracy to accurately indicate the level of the peak.

When estimating the spectral density of random signals, a widely used window is the Hann window, which is defined as:

$$w_H(n) = \begin{cases} \frac{1}{2}(1-\cos(\frac{n}{N}2\pi)) & \text{for} \quad 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \tag{4.4.19}$$

Another window, which is an amalgam of the rectangular and the Hann window, is the Tukey window, often called cosine-tapered window. This window sets the weights for a certain fraction of central data points to unity like a rectangular window while the data points at the tails are tapered to zero. The fraction is controlled by an additional parameter $\alpha \in [0,1]$. The window is defined as:

$$w_T(n) = \begin{cases} \frac{1}{2}\left(1-\cos(\pi\frac{n}{\frac{\alpha}{2}N})\right) & \text{for} \quad 0 \leq n \leq \frac{\alpha}{2}N \\ 1 & \text{for} \quad \frac{\alpha}{2}N \leq n \leq (1-\frac{\alpha}{2})N \\ \frac{1}{2}\left(1-\cos(\pi\frac{N-n}{\frac{\alpha}{2}N})\right) & \text{for} \quad (1-\frac{\alpha}{2})N \leq n \leq N-1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.4.20}$$

The window evolves from the rectangle to the Hann window as the parameter α varies from zero to unity and vice versa. This window has the advantage that the amount of tapering can be optimized according to the specific purpose of the spectral analysis [99].

4.5 Bayesian Spectrum Analysis

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

$$y(t) = s(t,\theta) + \epsilon(t) \qquad t = 1,...,N \tag{4.5.1}$$

where the deterministic component $s(t,\theta)$, is the true signal model and $\epsilon(t)$ are the random errors which are assumed to be i.i.d. $N(0,\sigma_\epsilon^2)$. The

Fourier series representation of the above model is given as:

$$y(t) = \sum_{n=0}^{N/2} [a_n \cos(2\pi nt/N) + b_n \sin(2\pi nt/N)] + \epsilon(t) \qquad t = 1, ..., N \tag{4.5.2}$$

where $a_n = R_n \cos\phi_n$ and $b_n = -R_n \sin\phi_n$ are Fourier coefficients, moreover $R_n = \sqrt{a_n^2 + b_n^2}$ is the amplitude and $\phi_n = \tan^{-1}(-b_n/a_n)$ is the phase of the nth harmonic. There is a common misconception about the Fourier coefficients a_n and b_n that they are random quantities. However, without the error term the above model is purely deterministic and thus all a_n and b_n will remain unchanged for each realization. The randomness of these coefficients is actually due to the noise variable which introduces uncertainty into this whole model which will become clear later in this section. Analogous to a multiple linear regression problem, using least square estimation method, a_n and b_n can be estimated by minimizing the residual sum of squares as shown below,

$$\sum_{t=1}^{N} \epsilon^2(t) = \sum_{t=1}^{N} \left[y(t) - \sum_{n=0}^{N/2} [a_n \cos(2\pi nt/N) + b_n \sin(2\pi nt/N)] \right]^2. \tag{4.5.3}$$

Differentiating with respect to a_n and b_n and equating to zero gives,

$$\hat{a}_0 = \bar{y} \tag{4.5.4}$$

$$\hat{a}_{N/2} = \sum_{t=1}^{N} \frac{(-1)^t}{N} \tag{4.5.5}$$

and for $n = 1, ..., \frac{N}{2} - 1$

$$\hat{a}_n = \frac{2}{N} \sum_{t=1}^{N} y(t) \cos(2\pi nt/N) \tag{4.5.6}$$

$$\hat{b}_n = \frac{2}{N} \sum_{t=1}^{N} y(t) \sin(2\pi nt/N) \tag{4.5.7}$$

and $\hat{b}_0 = \hat{b}_{N/2} = 0$. Hence the estimators for amplitude and phase are $\widehat{R}_n = \sqrt{\hat{a}_n^2 + \hat{b}_n^2}$ and $\hat{\phi}_n = \tan^{-1}(-\hat{b}_n/\hat{a}_n)$ respectively. Another important result that follows the estimation of a_n and b_n and hence R_n^2 is that for $n \neq 0, N/2$:

$$\begin{aligned}
\widehat{R}_n^2 &= \hat{a}_n^2 + \hat{b}_n^2 \\
&= \left(\frac{2}{N}\sum_{t=1}^{N} y(t)\cos(2\pi nt/N)\right)^2 + \left(\frac{2}{N}\sum_{t=1}^{N} y(t)\sin(2\pi nt/N)\right)^2 \\
&= \frac{4}{N^2}\left|\sum_{t=1}^{N} y(t)\exp(-2\pi int/N)\right|^2 \\
&= \frac{2}{N}S(f_n) \quad \text{from Equation (4.4.12).}
\end{aligned} \tag{4.5.8}$$

Rewriting the last line, we get,

$$S(f_n) \equiv N\widehat{R}_n^2/2 \qquad n = 1, ..., \frac{N}{2} - 1 \tag{4.5.9}$$

where $S(f_n)$ is the one sided power spectral density as defined by Equation (4.4.12), thus $\widehat{R}_n^2/2$ can be regarded as the contribution of nth harmonic to the power. Note that since for $n = 0, N/2$ the imaginary components in the Fourier representation are zero, they are required to be treated differently, same as in Equation (4.4.12). As stated above the noise variable $\epsilon(t)$ is assumed to be normally distributed with mean zero and variance σ_ϵ^2. In fact this *a priori* assumption of noise normality is based on the principle of maximum entropy which states that if the only available information about certain distribution is about its first and second order moments (mean and variance) and no additional information, which validates the use of some other sampling distribution, is available, the Normal sampling distribution is the best choice [4, 5]. Being the linear combinations of a normal random variable $y(t)$ (see Equation (4.5.2), (4.5.6) and (4.5.7)), each $\hat{a}_n$ and $\hat{b}_n$ are also i.i.d $\mathrm{N}(0, 2\sigma_\epsilon^2/N)$ for $n \neq 0,\ N/2$ and thus each $\hat{a}_n^2$ and $\hat{b}_n^2$ has a $\frac{2\sigma_\epsilon^2}{N}\chi^2_{(1\ df)}$ distribution respectively; thus a more general result

follows as:

$$\widehat{R}_n^2 = \hat{a}_n^2 + \hat{b}_n^2 \sim \frac{4\sigma_\epsilon^2}{N}\chi^2_{(2\text{ df})}. \tag{4.5.10}$$

When the noise variance σ_ϵ^2 is known then the squared amplitudes R_n^2 can be completely specified in terms of a_n and b_n, because their variances will also be known. On the other hand if σ_ϵ^2 is unknown then in frequentist framework it can be estimated using the estimated residuals that depend on the signal parameters θ and hence the variances of $\hat{a}_n$ and $\hat{b}_n$ will also become available which can then be used to find the distribution of R_n^2. In the Fourier domain, the estimation of these coefficients is rather simple because one just needs to take the DFT of the estimated residuals to get a sufficient statistic for R_n^2. Thus in the Bayesian paradigm, when sufficient information is available about the noise properties the conjugate prior for R_n^2 has a scaled inverse χ^2 distribution defined as:

$$p(R_n^2) = \text{Inv-}\chi^2(\nu_{n0},\, \sigma_{n0}^2) \tag{4.5.11}$$

where ν_{n0} is prior degrees of freedom and σ_{n0}^2 is the scale parameter. From Equation (4.5.9) it is clear that $\widehat{R}^2 = \hat{a}_n^2 + \hat{b}_n^2$ is a sufficient statistic for R_n^2. Therefore, with the above conjugate prior the posterior distribution for R_n^2, conditional upon other signal parameters θ through estimated residuals, is again a scaled inverse χ^2 distribution given as below.

$$p(R_n^2|\epsilon(\cdot),\theta) = \text{Inv-}\chi^2\left(\nu_{n0}+2,\, \frac{\nu_{n0}\sigma_{n0}^2 + \hat{a}_n + \hat{b}_n}{\nu_{n0}+2}\right). \tag{4.5.12}$$

If there are several independent (say K) time series, and hence independent noises, and if one assumes that the noise characteristics of different times series are different, one can use independent posteriors for each $R_{nj}^2,\ j = 1,...,K$, or if one assumes that the noise is the same in all time series then a common posterior for all R_{nj}^2 can be used. For the later case

the common posterior is defined as following,

$$p(R_n^2|\epsilon(\cdot),\theta) = \text{Inv-}\chi^2\left(v_{n0}+2K,\ \frac{v_{n0}\sigma_{n0}^2+\sum_{j=1}^{K}(\hat{a}_{nj}^2+\hat{b}_{nj}^2)}{v_{n0}+2K}\right) \tag{4.5.13}$$

where the additional subscript j now runs for jth time series, hence $\sum_{j=1}^{K}(\hat{a}_{nj}^2+\hat{b}_{nj}^2)$ is the sufficient statistics [54]. When using MCMC, the current residuals are used to update the above posterior distribution and the squared amplitudes are sampled from it using an additional Gibbs step [57]. It is also possible to maximize the noise variance (see next section) using a special type of likelihood function, so that to save the computational costs incured by this additional Gibbs step.

4.5.1 The Likelihood Function

After Fourier transform of the data the widely used likelihood function in the gravitational data analysis community is,

$$\mathfrak{L}(\theta) = K \times \exp\left[-2\int_0^\infty \frac{|\tilde{z}(f)-\tilde{s}(f,\theta)|^2}{S(f)}df\right] \tag{4.5.14}$$

Which for discretized data takes the form:

$$\mathfrak{L}(\theta) = K \times \exp\left[-\frac{1}{\nu}\sum_{j=1}^{\nu}\frac{|\tilde{z}(f_j)-\tilde{s}(f_j,\theta)|^2}{S(f_j)}\right] \tag{4.5.15}$$

where $\nu = \lfloor (N-1)/2 \rfloor$ is the greatest integer less than or equal to $(N-1)/2$, f_j are Fourier frequencies, $\tilde{z}(f_j)$ and $\tilde{s}(f_j,\theta)$ are Fourier transformed observables and model signal respectively; $S_n(f_j)$ is the one sided power spectral density and is assumed known here. Finally, K is the normalizing constant. When $S(f_j)$ is not known in advance or is assumed unknown then assuming that it is finite, one has to resort to another like-

lihood function which is of the form

$$\mathfrak{L}_w(\theta) = K \times \exp\left[-\frac{1}{\nu}\sum_{j=1}^{\nu}\left(\nu\log(S(f_j;\theta)) + \frac{|\tilde{z}(f_j) - \tilde{s}(f_j,\theta)|^2}{S(f_j;\theta)}\right)\right]. \tag{4.5.16}$$

In the literature Equation (4.5.17) is known as Whittle's approximation to the Gaussian likelihood or simply the Whittle's likelihood [54, 100]. In some references (see e.g. [101, 102]) Equation (4.5.16) is defined as follows:

$$\mathfrak{L}_w(\theta) = K \times \exp\left[-\frac{1}{\nu}\sum_{j=1}^{\nu}\left(\nu\log(\sigma_\epsilon^2 S(f_j;\theta)) + \frac{|\tilde{z}(f_j) - \tilde{s}(f_j,\theta)|^2}{\sigma_\epsilon^2 S(f_j;\theta)}\right)\right]. \tag{4.5.17}$$

The likelihood function defined by Equation (4.5.17) can be further simplified by maximizing over σ_ϵ^2 using the rule:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{\nu}\sum_{j=1}^{\nu}\frac{|\tilde{z}(f_j) - \tilde{s}(f_j,\theta)|^2}{S(f_j;\theta)}$$

[102] which results in the following concentrated or profile likelihood.

$$\mathfrak{L}_{wc}(\theta,\hat{\sigma}^2) \equiv K \times \exp\left[-\frac{1}{\nu}\left\{\nu\log\left(\frac{1}{\nu}\sum_{j=1}^{\nu}\frac{|\tilde{z}(f_j) - \tilde{s}(f_j,\theta)|^2}{S(f_j;\theta)}\right) + \sum_{j=1}^{\nu}\log(S(f_j;\theta))\right\}\right]. \tag{4.5.18}$$

The Whittle's likelihood, using the approximate properties of DFT, assumes that the discrete Fourier transform residuals are approximately complex normally distributed with mean zero and power spectrum $\sigma_\epsilon^2 S(f_j;\theta)$. In Equation (4.5.18) we see that σ_ϵ^2 is maximized and since $\hat{a}_j \sim N(0, \frac{2}{N}\sigma_\epsilon^2)$ and $\hat{b}_j \sim N(0, \frac{2}{N}\sigma_\epsilon^2)$ therefore their variances are also maximized. Furthermore, since $\widehat{R}_j^2 = \hat{a}_j^2 + \hat{b}_j^2 \sim \frac{4}{N}\sigma_\epsilon^2\chi^2_{2\,df}$ where again the the variance is maximized therefore the advantage of Equation (4.5.18) is that it does not need any additional specification of conditional posterior for the squared amplitudes, since the noise variance is effectively maximized

everywhere during the computation of the likelihood.

To limit the likelihood computation to a particular frequency range $[f_{j_L},\ f_{j_U}]$ within which the desired signal is believed to be more powerful, a rather more flexible and computationally efficient form is:

$$\mathfrak{L}_{wc}(\theta, \hat{\sigma}^2) \equiv K \times \exp\left[-\frac{1}{\nu_c}\left\{ \nu_c \log\left(\frac{1}{\nu_c} \sum_{j=j_L}^{j_U} \frac{|\tilde{z}(f_j) - \tilde{s}(f_j, \theta)|^2}{S(f_j;\theta)} \right) - \sum_{j=j_L}^{j_U} \log(S(f_j;\theta)) \right\} \right] \quad (4.5.19)$$

where f_{j_L} and f_{j_U} are the lower and upper bounds of the examined frequency range [54, 103] and ν_c is the number of frequencies within this range. The same modifications about limited frequency ranges apply to the other forms of the likelihood functions too.

4.5.2 Joint Likelihood

The joint likelihood function for multiple independent time series is just the product of their individual likelihood functions. Let $\mathfrak{L}^I$ be the likelihood function for I^{th} $(I = 1, ..., K)$ time series then we can write,

$$\mathfrak{L}(\theta) = \prod_{I}^{K} \mathfrak{L}^{(I)} \Leftrightarrow \log \mathfrak{L}(\theta) = \sum_{I}^{K} \log \mathfrak{L}^{(I)}. \quad (4.5.20)$$

4.6 Signal-to-Noise Ratio

Signal-to-noise ratio (SNR) is an important measure in signal processing which measures the strength of the detected signal to the background

noise. It is defined as:

$$SNR = \sqrt{4 \int_0^\infty \frac{|\tilde{s}(t,\theta)|^2}{S(f)} df} \tag{4.6.1}$$

[9] which for discretized frequencies becomes:

$$SNR = \sqrt{\frac{4}{N} \sum_{j=1}^{\nu} \frac{|\tilde{s}(f_j,\theta)|^2}{S(f_j)}} \tag{4.6.2}$$

and for a particular frequency range $[f_{j_L}, f_{j_U}]$ the definition becomes:

$$SNR = \sqrt{\frac{4}{N} \sum_{j=j_L}^{j_U} \frac{|\tilde{s}(f_j,\theta)|^2}{S(f_j)}}. \tag{4.6.3}$$

When searching a signal in several (say K) independent time series the combined SNR is just the square root of the sum of their individual squared SNRs defined as following:

$$SNR_K = \sqrt{\sum_{i=1}^{K} SNR_i^2} \tag{4.6.4}$$

[7]. SNR is a very useful measure as it tells us about the quality of the detected signal and hence the parameters' estimates either on the fly or at the end of MCMC estimation.

CHAPTER 5

APPLICATIONS AND RESULTS

5.1 Introduction

We used the Bayesian MCMC approach to address both the signal detection problem, and the estimation problem of estimating the parameters of a detected signal. The methods, discussed in Chapter 4, were collectively applied in different scenarios ranging from recovering an EMRI signal from the data containing a single EMRI source to recovering a particularly targeted EMRI signal from the data containing overlapping signals from multiple EMRI sources, with the noise level ranging from only instrumental noise to more complicated confusion noise created by millions of other GWs sources. Furthermore, two different approaches were used to derive the LISA response to GWs, i.e. the full LISA response and the low frequency approximation to full LISA response as discussed in Chapter 3.

Before proceeding to our applications and results a brief overview of the earlier MCMC applications for the detection and parameter estimation of EMRI signal is presented as follows:
The authors of [37] used a highly simplified version of Barack and Cutler waveform model in which the three phase parameters (Φ, $\tilde{\gamma}$, α) were fixed, i.e. only 11 parameters were used, and the PN-evolution equations of the orbital parameters such as v and e were truncated to the first few leading

order terms, and a low eccentricity expansion of the source orbit. To derive the LISA response they employed Cutler's one detector's response (h_I) in the LFA regime [24]. The resulting posterior density was explored by an automatic RJMCMC sampler which was effective in recovering a single EMRI signal buried in stationary Gaussian white noise. They did not explore the applications of their search algorithm to full LISA response data (i.e. MLDC data). The authors of [38] and [39] used simplified and somewhat similar waveform models, and MCMC algorithms but their LISA response approximations were different. That is, both used truncated versions of the full AKW model but to derive the LISA response, [38] used the LFA and [39] employed the full LISA response with a rotating but static LISA triangle, rather than the one with "breathing" effects (see also Section 3.4). Both employed the so-called MHMC algorithm [41], which generates a non-Markovian chain since it employs a variety of purposely directed proposal distributions that allow a range of jumps of different size and type in the parameter space, with simulated annealing and multiple parallel chains for harmonic identification as an initial step of the search. These initial searches were followed by MCMC searches using the high SNR parameters states found in the previous step as the starting distributions. In both cases the variance-covariance matrix of the main proposal distribution was updated by the eigendirections of the Fisher information matrix (FIM) with [38] imposing some predefined constraints on the FIM. The author of [40] also explored the same MHMC algorithm for EMRI detection, with several additional sophisticated techniques. His search algorithm uses a multi-stage hierarchical procedure that starts from searching shorter segments of the data and choose the best fit solutions to be utilized in the next stage of the search. The length of the segments increases at each stage of the search. It should be noted that all of the above searches were conducted to detect and characterize a single, bright EMRI signal buried in purely instrumental Gaussian noise. The approaches introduced in this thesis are strictly based on Markovian theory. Also, although we also use the truncated version of AKW model, namely TAKW of [38], but we use the full parameter models (both AKW and TAKW), i.e. we do not em-

ploy fixed values or maximize over any of the signal parameters, and use both (but in respective applications) the full LISA response and the LFA without any modifications to their proposed theoretical functions.

5.2 Data

The realistic data are expected to be available after 2022; thus all the results obtained and presented here are based on simulated data. The simulated data which we have been searching for EMRI signals are issued by the MLDC Task Force [104] in different rounds, called *challenge data*. To date, four rounds namely MLDC 1, MLDC 2, MLDC 1B (a repeat of MLDC 2) and MLDC 3 have been conducted at different times. We could not participate in any of these rounds as the last round (MLDC 3) was completed in April 2008, when our algorithm and the source codes were still under development. The current round named MLDC 4 is in progress with an expected deadline for submission of results in mid June 2011. It is planned to participate in this round and some preliminary results have already been presented at GWPAW (January 26-29, 2011, Milwaukee, Wisconsin, USA). In each of the previous rounds, there were different data sets for different sources. The complexity and purpose of these challenge data rises with each round. Some data sets contained only a single source buried in LISA's instrumental noise while some contained multiple superposed signals from the same type of sources but with different parameter values. EMRI signals were included in all rounds but MLDC 1B was the first round to see some success regarding the detection and estimation of parameters for some bright isolated EMRI signals. In this challenge there were five data sets each containing a single EMRI source buried in instrumental noise only. In MLDC 3, five overlapping EMRI signals were buried in instrumental noise in a single data set. The MLDC 4 is more complicated as it consists of a single data set called the "whole enchilada" which, besides the instrumental noise, contains millions of signals from all the other possible sources along with a few

EMRI signals.

All these data sets comprise four variables; i.e. a time variable and the three TDI observables X, Y and Z each recorded at a constant interval of 15 seconds. These data sets are encoded in XML (the eXtensible Markup Language) format and their contents can be retrieved using a routine known as *lisaXML* which is available in MLDC's community resources [43].

5.3 The Experimental Setup

The basic algorithm mainly consists of a waveforms model, a LISA response function and an MCMC sampler. The structure of the algorithm is illustrated by the diagram given in Figure 5.1. From this diagram it is clear that LISA response is required for the computation of the likelihood at each MCMC iteration. To compute the full LISA response, i.e. the X, Y, Z and then A, E TDI variables for the polarization signals $(h_+, h_\times)$, the actual LISA simulator software [105] was dissected and adapted to make it an active part of the MCMC algorithm. This adaptation was done by a research group called *Global LISA Inference Group (GLIG)* in 2006 [54]. In this current work some debugging was done to this adapted LISA simulator to correct some phase differences between it and the actual LISA simulator to make them deliver exactly similar LISA response, which improved the performance of the algorithm remarkably. The LISA simulator is not an MCMC friendly software; however, it is computationally quite expensive not only with respect to the processing speed but also with respect to memory. To compute the TDI response for a full length waveform, typically 2 years long, the LISA simulator takes several minutes. Furthermore, to avoid memory problems, the actual LISA simulator software divides the full length waveform into small segments, typically $1\frac{1}{2}$ months long, derives the TDI responses for all segments one by one, saves them externally in separate files and finally, recombines them into one file. The adapted LISA simulator included in our algorithm, however, does

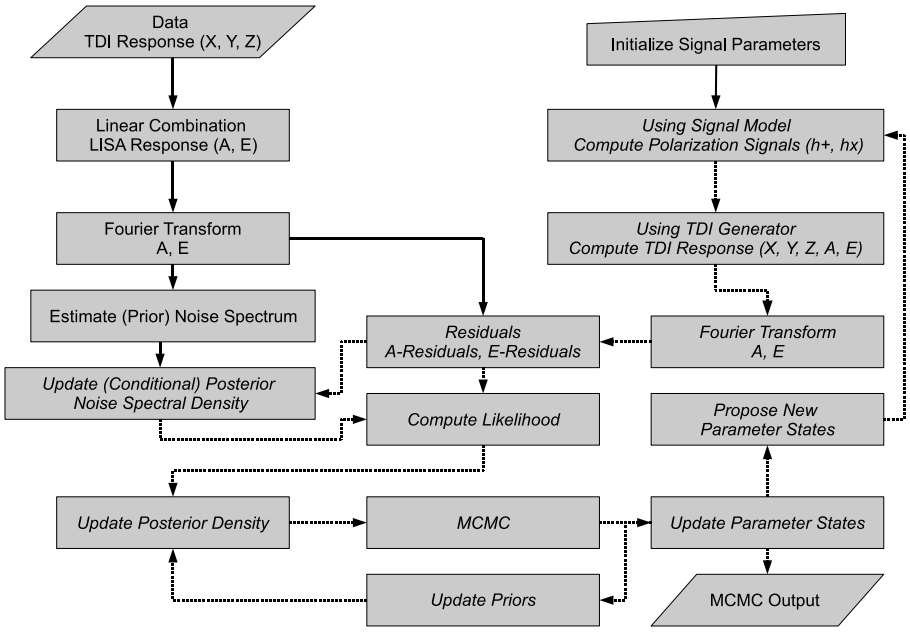

Figure 5.1: Illustration of the structure of algorithm.

not have this feature because it is not possible to repeat the same process within the MCMC sampler. This makes it quite ineffective for processing longer waveforms because it then needs huge memory to hold the entire operation at the same time. Moreover, the LISA simulator internally multiplies the length of the injected model signal by 4 which makes it considerably slower even for a shorter data segment. For example, if one is using a week long data segment then in the LISA simulator the injected model signal will be four weeks long. Since at each MCMC iteration, one needs to generate the model signals too, which is also computationally expensive and puts further burden on the speed and memory. Therefore at the end one has to use shorter segments, typically a few weeks long, to perform an efficient MCMC analysis within a reasonable time.

Due to the slow speed and memory hungry nature of the LISA simulator, it was decided to use the low frequency approximation for computing the LISA response [24, 29]. This improved the speed of our algorithm to a great extent which in turn enabled us to use more dispersed values as starting points for different parameters in our MCMC searches. The LISA simulator is too inflexible in the sense of prior specifications for signal parameters, i.e. one has to specify very narrow parameter prior to choose starting values from, i.e. values which are somewhere near the true modes of the parameters, otherwise the sampler, despite being a sophisticated one, would take prohibitively long time to converge to a true maximum.

In most of the initial MCMC searches the AKW model [29] was used to generate the model signals; however, here again because of the computationally complex nature of these waveforms it was decided to use the TAKW model [38] to save time, because TAKW is $\sim$ 3 times faster than AKW and has a very good agreement with AKW.

EMRI signals in the LISA data are expected to be very weak; the instantaneous amplitude is an order of magnitude below even the instrument noise level as can be seen in the bottom three boxes of Figure 2.4 of Chapter 2. These signals have a very complicated spectrum. The posterior surface contains several local maxima that are well separated from each other and some of them are as high as 85% of the global maximum. These local maxima arise due to several harmonics corresponding to different frequencies, particularly the orbital frequency which is characterized by a large number of harmonics that have a very strong influence on the overall spectrum of the EMRI signals. In both Figure 2.4 and Figure 3.1 several isolated peaks can be seen on the surface of the spectral densities of EMRI signals from different types of sources. These peaks correspond to different harmonics and the largest peak corresponds to the fundamental (true) frequencies. The sizes and spreads of some of these local peaks are almost equal to that of the true peak (the largest

one). This makes the posterior of these signals very complicated. The ordinary MCMC samplers fail to explore such a complicated multi-modal surface. The sampler quickly gets stuck at a local mode corresponding to a low strength harmonic. For these reasons a PTMCMC sampler was used, which is very effective in dealing with multi-modal target densities.

The entire algorithm was programmed in C programming language [106]. Several public domain software packages were used for doing different tasks. For example, the discrete Fourier transform operations were performed using a C subroutine library *FFTW* [95] and parallelization of the PTMCMC sampler and MPI was accomplished by using *OpenMPI* library [88]. All the distribution specific random numbers were generated using the C version of subroutine library called *RANDLIB* [107]. The MCMC output was processed using the R language [77].

Different tactics, as described in Section 4.4.4, were employed to enhance the precision of discrete Fourier transform estimation such as both the data and model signals were windowed in time domain using a Tukey window with a light intensity ($\alpha = 0.02$) to reduce the Fourier frequency noise arising from the incomplete cycles left at the endpoints of a segment. The spectral density was estimated using Welch's method for which a rather more rigorous window, the "Hann window" was used. When using LISA simulator, the conditional posterior spectral density was estimated directly from Fourier transformed residuals using Equation (4.5.8).

Within the MCMC sampler the proposals were generated from a multivariate Student t-distribution since it has longer tails than the Gaussian distribution which insures good coverage of target density and is also very useful for good mixing. The degrees of freedom was set as 3 as this is the smallest possible (integer) value for which the variance of the distribution is still finite.

Regarding the prior densities for waveform parameters, all the parameters were assumed to be independent. This enables us to specify independent prior distributions for all parameters. Thus all the waveform parameters were assumed to be uniform over their respective domains as defined by MLDC, except for some of the angles that are 'by definition' polar angles. For the polar angles which are θ_S (ecliptic latitude), θ_K (polar angle of spin) and λ (angle between spin and angular momentum), the prior density was defined as $f(\cdot) = \frac{1}{2}\sin(\cdot)$. The logic behind this prior density is that a given value of these angles corresponds to a "circle of latitude", whose circumference is proportional to $\sin(\cdot)$ [54]. Although for parameters ν (orbital frequency), μ (CO's mass), M (SMBH's mass) and D_L (luminosity distance) the use of uniform prior is straight forward, in the waveform model their logarithmic values are used [29]; therefore their prior distributions were transformed using the simple technique of transformation of random variables to yield a truncated exponential distribution. For the conditional posterior spectral density, the prior spectral density was estimated from a different segment of the data.

In general, when enough memory was available, the full covariance matrix of the proposal distribution was updated only once on the fly using iterations between 10,000 and 30,000. During these iterations the swapping mechanism of the sampler was kept at halt to avoid over estimation of the elements of the covariance matrix, as explained in Section 4.3.2.3. However, because of negligible correlations between different parameters ($< .50$), later on a fixed covariance matrix was used. It also depends on the theoretical expectations about the relationships of the model parameters to have some idea whether or how important the off-diagonal elements of the covariance matrix are. Updating the full covariance matrix manually or from an external file was never successful because of the numerical errors causing its inversion failure within the multivariate normal random number generator subroutine. It would not make sense to update the covariance matrix repeatedly while the chains

are swapping states as it would lead to very large variances for some of the parameters, particularly the angles that then frequently cross their boundaries during MCMC (see also Section 5.3).

Initially, due to lack of computing facilities and memory problems most of the MCMC computations were done on a very small scale using a local computing server, which has only 8 cores and at a time one could occupy up to 4 cores only. It was not until September 2009 when we started using the cluster of high performance computing servers, the BeSTGRID [12]. Running codes on BeSTGRID was not without challenges with the growing number of users from throughout the university. Later, the cluster was upgraded to include a few new fast servers thus it now consists of both fast and slow servers. To search a week long data the MCMC code, which uses LISA simulator to compute LISA response, would sometime run at a speed of $\sim$ 58 iterations per minute and sometimes at $\sim$ 16 iterations per minute. When running a large MCMC (usually more than 8 MCMC chains), it is now almost impossible to achieve a fast speed because if a single chain went to a slow speed core (and it almost always happens), then it will bring down the speed of the entire code because of the speed synchronization. There are storage problems too. The whole cluster had only 250 GB storage which is now upgraded to 500 GB, but it is still not enough considering the large number of users and the amount of their input/output data. Recently our department purchased five new compute servers each of which has eight cores. There are in total 40 fast cores but these servers are not linked like those in the BeSTGRID cluster, where different compute servers can communicate with each other, therefore one MPI program could only use up to eight cores (i.e. eight MCMC chains) at a time. Most of the MCMC searches on MLDC 4 blind data set were carried out on a supercomputer; ATLAS (Hannover, Germany) [108].

Out of Domain Angular Adjustment

As we have seen in Chapter 3, out of the 14 waveform parameters, 8 are angles. These angles are very tricky and need to be handled carefully. These angles are independent of each other as long as they remain within their defined range. In MCMC sampling it is possible that any one of them get out of their specified range. Some useful alternative equivalent relationships, which are used in case an angle cross its bounds, are presented in Table 5.1. At first glance these relationships may look complicated and,

Table 5.1: Adjustment of out of domain angles. The additional subscripts 'c' and 'n' indicates the current and new values respectively.

In case	Do this
$\theta_{Sc} > \pi$	$\phi_{Sn} = \phi_{Sc} + \pi$ and $\theta_{Sn} = 2\pi - \theta_{Sc}$
$\theta_{Sc} < 0$	$\phi_{Sn} = \phi_{Sc} + \pi$ and $\theta_{Sn} = -\theta_{Sc}$
$\phi_{Sc} > 2\pi$	$\phi_{Sn} = \phi_{Sc} - 2\pi$
$\phi_{Sc} < 0$	$\phi_{Sn} = \phi_{Sc} + 2\pi$
$\theta_{Kc} > \pi$	$\phi_{Kn} = \phi_{Kc} + \pi$ and $\theta_{Kn} = 2\pi - \theta_{Kc}$
$\theta_{Kc} < 0$	$\phi_{Kn} = \phi_{Kc} + \pi$ and $\theta_{Kn} = -\theta_{Kc}$
$\phi_{Kc} > 2\pi$	$\phi_{Kn} = \phi_{Kc} - 2\pi$
$\phi_{Kc} < 0$	$\phi_{Kn} = \phi_{Kc} + 2\pi$
$\lambda_c > \pi$	$\alpha_n = \alpha_c + \pi$ and $\lambda_n = 2\pi - \lambda_c$
$\lambda_c < 0$	$\alpha_n = \alpha_c + \pi$ and $\lambda_n = -\lambda_c$
$\alpha_c > 2\pi$	$\alpha_n = \alpha_c - 2\pi$
$\alpha_c < 0$	$\alpha_n = \alpha_c + 2\pi$
$\tilde{\gamma}_c > 2\pi$	$\tilde{\gamma}_n = \tilde{\gamma}_c - 2\pi$
$\tilde{\gamma}_c < 0$	$\tilde{\gamma}_n = \tilde{\gamma}_c + 2\pi$
$\Phi_c > 2\pi$	$\Phi_n = \Phi_c - 2\pi$
$\Phi_c < 0$	$\Phi_n = \Phi_c + 2\pi$

instead of using these relationships, one could think of just rejecting those proposals which push any one of these angles out of their domain but that would decrease the overall acceptance rate of the regular Metropolis proposals. These are actually equivalent cases as at these alternative positions the model signal remains the same. In fact these are very simple relationships. Let us arrange these angles in four pairs, e.g. $\{\theta_S, \phi_S\}$, $\{\theta_K, \phi_K\}$, $\{\lambda, \alpha\}$

and $\{\tilde{\gamma}, \Phi\}$. The first three pairs share the same feature, i.e. they represent the two angles of the position of a point in a spherical coordinate system, the first angle is the zenith or the polar angle whereas the second angle is the azimuth angle. The radius component of this point is ignored as we are not interested in how far the point is, rather we are interested in where the point is. Let us depict $\{\theta_K, \phi_K\}$ in a spherical coordinate system in which x-axis is playing the role of the North pole as shown in Figure 5.2. The plane (in light grey colour) passes through the mid point of y-axis, i.e.

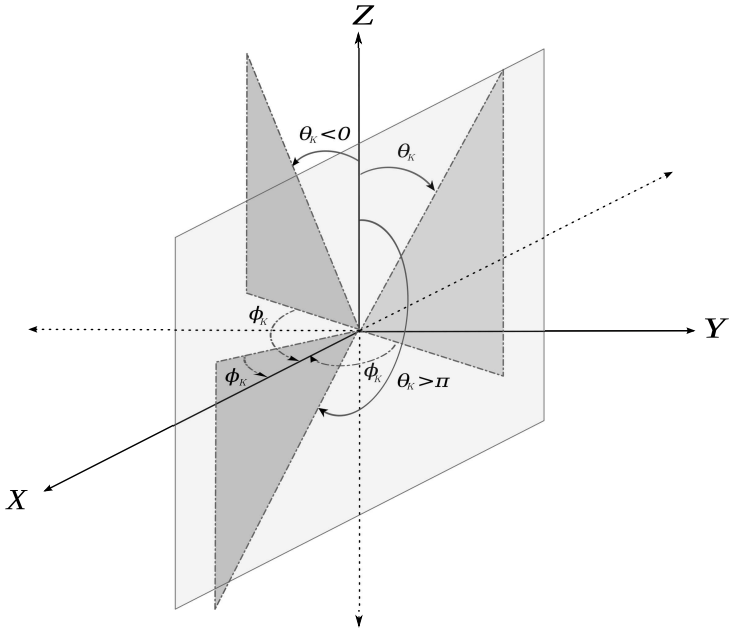

Figure 5.2: Spherical coordinates system to illustrate the effect of changes in θ_S on ϕ_S, due to out of domain angle magnitudes.

it divides the imaginary sphere into two equal hemispheres. Suppose the values of θ_K are valid only in the right hemisphere because it has a range $[0, \pi]$ which covers only the half part of the sphere. Now if it goes beyond this range (i.e. to the other hemisphere, in either direction) it will take along with it the ϕ_K angle too. For example, suppose its newly proposed value is negative, then it will be in the left hemisphere and if one makes it positive to bring it back to its actual right hemisphere then one would need to bring the second angle too by adding π into it. Similarly, if the newly proposed value of θ_K is greater than π then subtracting it from 2π

will place it at the same magnitude in its actual right hemisphere and at the same time one would again need to add π to ϕ_K so that it moves along with the new θ_K to keep the things correct. Furthermore, if the newly proposed values of ϕ_K turns out to be out of its domain then it can be brought back to its domain by simply subtracting or adding a 2π as shown in Table 5.1. The same comments apply to the other two pairs of angles too. The adjustment of angles α, $\tilde{\gamma}$ and Φ is performed in the same way as that of angle ϕ_K.

5.4 Results

The overall results obtained during this work can be mainly categorized into two parts: the results in which the LISA response to the waveform signals was computed using the adapted LISA simulator subroutines and the results in which we used the low frequency approximation to derive the LISA response. In both cases the other components of the algorithm were the same.

5.4.1 Full LISA Response Results

As stated before, in this setup the TDI response to the model signals was computed using the adapted LISA simulator. The code was first tested by applying it to some simulated data sets that were generated using LISA simulator software. The EMRI signals were generated according to the parameter keys corresponding to different EMRI sources that were issued by MLDC in previous rounds. Under this set up the code was applied to recover EMRI signals from two types of data sets: in one type there were single EMRI signals in different data sets and in the other type there were multiple overlapping EMRI signals in a single data set. In both types the noise source was only the instrumental noise while the confusion noise arising from galactic and extra-galactic binaries was ignored. In fact for the EMRI sources the confusion noise was introduced only in the current

round MLDC 4, which is still in progress. Because of the slow speed and memory problems of the overall code shorter segments of one to three weeks were used.

5.4.1.1 Single EMRI

The purpose of this example was two fold. Firstly to check whether our algorithm can recover an EMRI signal from LISA noise and secondly to evaluate the differences in the performance, with respect to signal detection and parameter estimation, of the two waveform models that were used in our algorithm, i.e. AKW and TAKW models. In this setup the simulated LISA data contained an EMRI signal from a medium mass ($M \in [4.75, 5.25] \times 10^6 M_\odot$), which was generated using AKW model with the parameters' values taken from MLDC 1B.3.2 training source. To recover this signal both the AKW and TAKW models were alternatively used. The AKW model is still very time consuming and despite using a week long data one MCMC iteration took $\sim$ 3 seconds on a 2.8 GHz *AMD Opteron(tm) Processor*, whereas by using TAKW, one MCMC iteration takes roughly 1.3 seconds. In both cases a thirty chain PTMCMC was run with starting values for some parameters in the vicinity of their true values. For the AKW model the estimated kernel densities and trace plots for the posterior samples for all the 14 parameters are shown in Figures 5.3 and 5.4 respectively. Similarly Figures 5.5 and 5.6 display the same results obtained with TAKW model. These results demonstrate that there is almost no difference between performance of the two models in recovering a signal, except the speed. On BeSTGRID the code using AKW model took $\sim$ 17 days to get half a million MCMC samples at an average speed of $\sim$ 20 iterations per minute while code using TAKW was running at $\sim$ 45 iterations per minute. Regarding the recovery of the signal, one can see that the modes of the estimated kernel densities of the marginal posterior samples for all parameters are very near their true values. The kernel density estimate of parameter ϕ_K exhibits an obvious bi-modality but in reality it is not. It is just because its true value is on

the very edge of the upper bound, 2π, of its domain and it is frequently crossing this bound during MCMC jumps. As explained in Section 5.3, when its proposed value is greater than π then subtracting this proposed valued from 2π brings it back to its domain. Both modes are equivalent and one can also see that the rest of all the parameters are unaffected by this bi-modality, particularly its companion angle; i.e. the parameter θ_K is still on its true mode as can be seen in the trace plots.

Although these were semi-blind searches for most of the key parameters, like ν, μ, e and χ and some of the angles, the starting values were set close to the true values. Otherwise because of the slow speed of LISA response computation it would take an extremely long time to complete convergence. Though different chains have different acceptance rates for the regular Metropolis jumps, the overall acceptance rate was $\sim$ 25%. Similarly the overall rate of swap acceptance between neighbouring chains was $\sim$ 30%. Using CODA package, the multivariate potential scale reduction factor, $\widehat{R}^p$, was found equal to 1.00, indicating that all the chains converged. The algorithm was tested on many other simulated as well as MLDC data sets and all cases it was observed that the two models (AKW and TAKW) gave almost similar results. Therefore in all of the subsequent MCMC searches the TAKW model was used. Table 5.2 shows some summary of the posterior estimates for all 14 parameters of EMRI source in actual MLDC 1B.3.2 training data set and Figure 5.7 shows their kernel density estimates. Different angles may be incorrectly estimated such a the phase angles $\tilde{\gamma}$ and Φ; however, these parameters are very difficult to estimate and are unimportant astrophysically at this initial stage of ongoing EMRI research [109]. Much emphasis is placed on the correct estimation of ν, M, e as these parameters affect the waveform length or time to plunge, which is a possible measure for the comparison of results.

The nature of EMRI signals is not yet well understood by researchers and there is no literature available which could put some light on the interac-

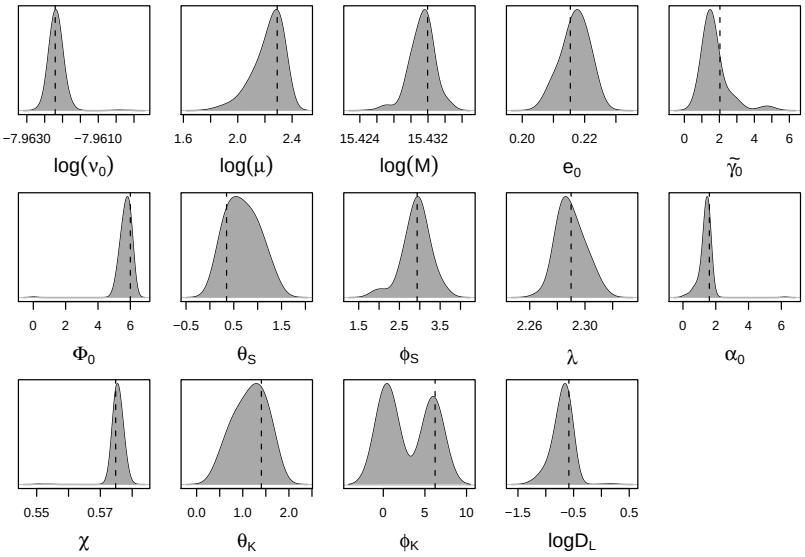

Figure 5.3: AKW model: Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI training source 1B.3.2. The dashed lines indicate the true values.

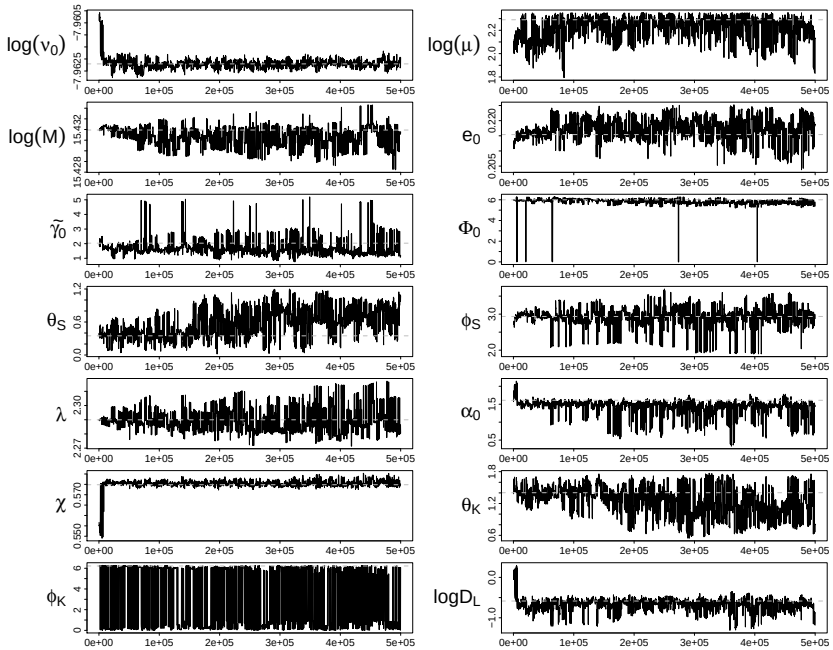

Figure 5.4: AKW model: Trace plots of the marginal posterior MCMC samples for all 14 parameters for the EMRI training source 1B.3.2. The grey dashed lines indicate the true values.

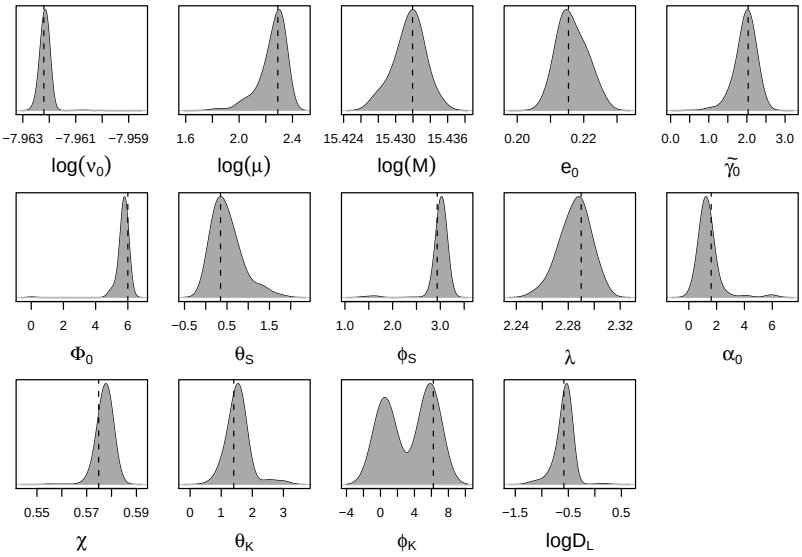

Figure 5.5: TAKW model: Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI training source 1B.3.2. The dashed lines indicate the true values.

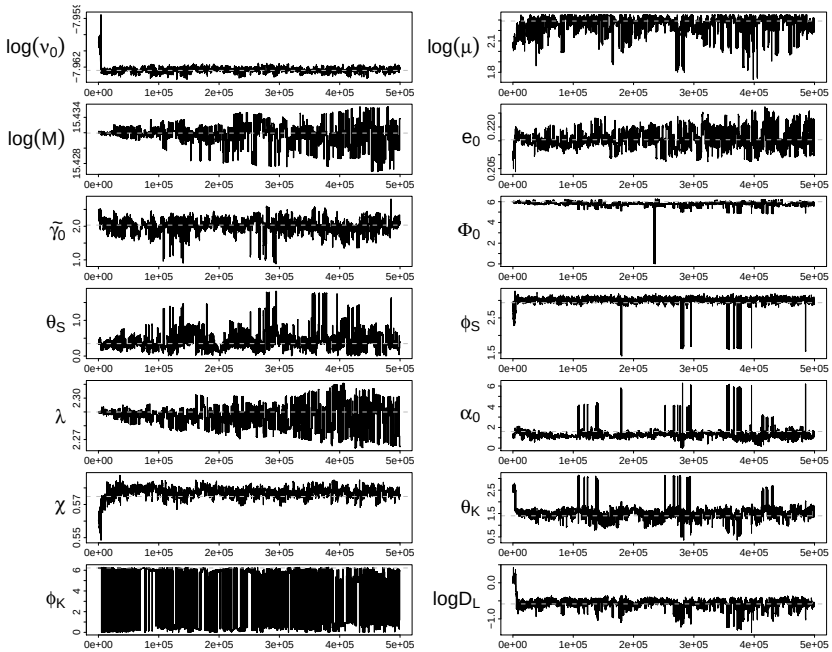

Figure 5.6: TAKW model: Trace plots of the marginal posterior MCMC samples for all 14 parameters for the EMRI training source 1B.3.2. The grey dashed lines indicate the true values.

Table 5.2: Posterior results and true parameter values for the recovered EMRI signal given in actual MLDC 1B.3.2 training data set.

Parameters	Mean	StdDev	Mode	95% BCI*	True Values
$\log(\nu_0)$	-7.9620378	0.0002130	-7.9620689	(-7.9623455, -7.9616517)	-7.9622049
$\log(\mu)$	2.2767335	0.0538632	2.2994283	(2.1787098, 2.3438539)	2.2902575
$\log(M)$	15.4347056	0.0043300	15.4315218	(15.429961, 15.4415627)	15.4319518
e_0	0.2206228	0.0092475	0.2113745	(0.2082393, 0.2363312)	0.2154005
$\tilde{\gamma}_0$	4.3606406	1.3283087	4.9727206	(1.3662998, 5.630289)	2.0332973
Φ_0	3.1230357	0.6555886	3.3730609	(1.9679207, 3.9275144)	5.9999000
θ_S	1.0140620	0.4358311	1.3024822	(0.3363044, 1.6545353)	0.3484659
ϕ_S	2.9175502	0.3106825	3.0264032	(2.1062912, 3.1686211)	2.9346250
λ	2.2562624	0.0159953	2.2528088	(2.2321323, 2.2861428)	2.2899505
α_0	2.7297661	1.7821384	1.6786134	(0.2875322, 5.9887385)	1.6092149
χ	0.5674923	0.0117862	0.5710465	(0.5319192, 0.5750127)	0.5748184
θ_K	2.0894894	0.3952991	2.0155767	(1.4146781, 2.7271543)	1.4034163
ϕ_K	3.7507945	2.5656960	5.7233197	(0.1093728, 6.1672813)	6.2231290
$\log(D_L)$	-0.8373054	0.2576184	-0.6801450	(-1.3078644, -0.5413055)	-0.5847797

*BIC: Bayesian Credible Intervals

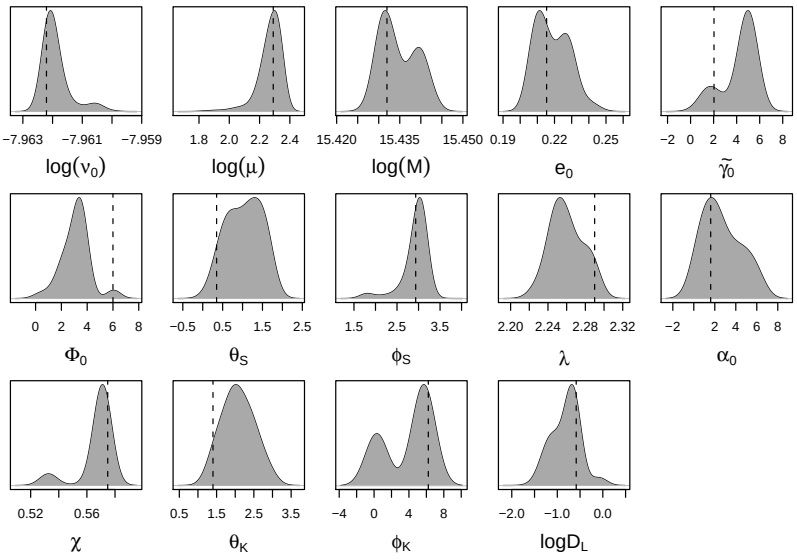

Figure 5.7: Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI source in actual MLDC 1B.3.2 training data set. The dashed lines indicate the true values.

tive behaviour and co-relationships of the EMRI parameters in general, as well as in MCMC applications. In these initial attempts no significant correlations among different pairs of posterior MCMC samples of different

parameters were found. However, it was realized that some of these 14 parameters are extremely sensitive to the specification of the proposal distributions for them. One must be careful about the step sizes of some of the key parameters such as ν, e, M and λ. Bold proposals for all these parameters significantly reduce the acceptance rates of the regular Metropolis jumps. Moreover, from Equation (3.2.16) it is clear that the time to plunge or in other words the waveform length is specified by the values of ν, e and M. Very large jumps in these parameters will not only reduce the acceptance rates of the regular Metropolis jumps but will also result in a very short or even zero plunge time (i.e. no signal at all) if the proposed values of these parameters do not correspond to an orbit which is inside the inner most stable orbit. Also since M appears in four of the five PN-evolution Equations (3.2.11–3.2.15), as a result of bold jumps in this parameter all the orbital parameters becomes highly correlated with M as well as with each other, leading to wrong results. In another MCMC analysis of EMRI signals [38] there were reports of sky location angles (θ_S, ϕ_S) getting locked at wrong positions; however, they were using low frequency approximation to derive LISA response instead of the full LISA simulator as well as a different MCMC sampler, MHMC, and sampling schemes.

5.4.1.2 Multiple EMRIs

For the recovery of an EMRI signal from a data set containing multiple EMRI sources the same strategy was used, i.e. first testing the algorithm on simulated data and then applying it to MLDC data, which were issued in MLDC 3. Since it was planned to participate in the MLDC 4 where there are multiple EMRI sources, to generate the simulated data, instead of using the parameter values given for MLDC 3 sources, the five EMRI sources (one high mass, two medium mass and two low mass) were generated with the parameter values corresponding to those given for the training sources in the first version of MLDC 4 (the data in this round were later revised by MLDC in October 2010). The confusion noises from other sources were ignored since the purpose of this attempt was to see how our

algorithm performs in recovering a signal from a data containing multiple overlapping EMRI sources. Different codes were run to recover these sources individually. Because of the long MCMC simulations, low speed and unavailability of enough slots on BeSTGRID for longer durations only two codes, one on 30 cores (on BeSTGRID) and the other on 4 cores (on local server), could be run in an attempt to recover only two of these five sources, named as source 4.2 and source 4.3. For the source 4.2, for which a 30 chains PTMCMC was run on BeSTGRID, the kernel density estimates are given in Figure 5.8. Figure 5.9 displays the traces plots of the marginal posterior samples and Table 5.3 shows summary statistics of the marginal posterior quantities. Figure 5.10 displays the log-likelihood values of the first 8 chains, which demonstrates the improvement of different chains towards the maximum by swapping states with each other. From Table 5.3 it is clear that almost all the parameters are estimated with great accuracy. Similarly Figure 5.11 and Table 5.4 shows the results for the source 4.3, for which a 4 chains PTMCMC was run on our local server. Since there were many signals in the same data therefore the likelihood surface was rather more complicated and hence we can see some differences in results for the second signal where some of the estimated angles are somewhat different from the true ones because of fewer chains and hence low temperature in PTMCMC run. The code was then applied to the actual MLDC 3 blind data set where there were five EMRI signals buried in LISA instrument noise. The target signal was that from a medium mass EMRI source; i.e. source 3.2 in MLDC 3 data. The true parameters for these signals were known as this round was completed in April 2008 when our algorithm and codes were still in development process. However, we used a completely blind search by first running several MCMC chains with random starting values in parallel and then set the modes of the posterior samples of the parameters, in those chains for which the SNR and likelihoods were higher than others, as the starting values for a new MCMC run. These results are given in Figures 5.12 and 5.13 and Table 5.5.

We attempted to search MLDC 4 data with low frequency approximation

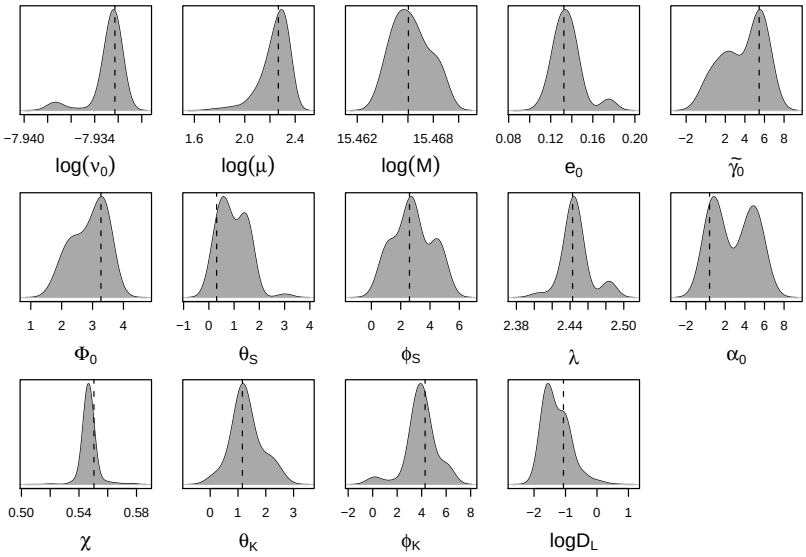

Figure 5.8: Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI source 4.2. The dashed lines indicate the true values.

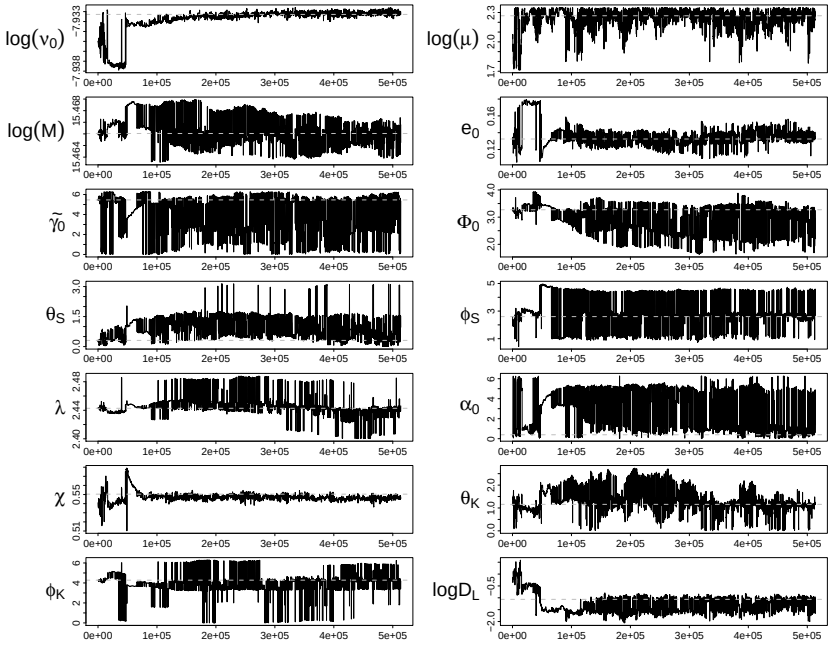

Figure 5.9: Trace plots of the marginal posterior MCMC samples for all 14 parameters for the EMRI source 4.2. The grey dashed lines indicate the true values.

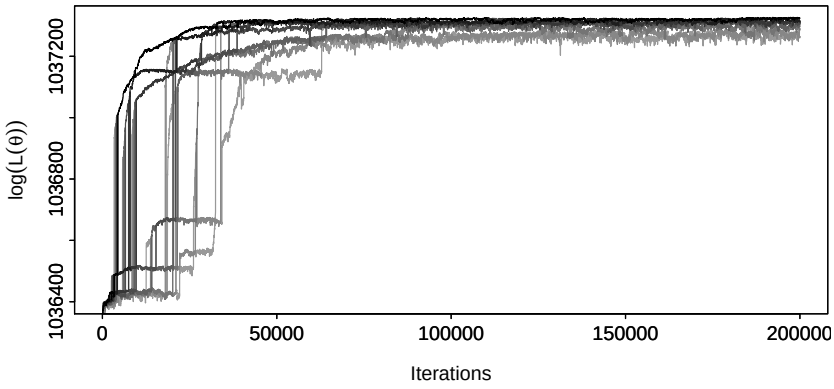

Figure 5.10: The log-likelihood values, obtained for the EMRI source 4.2, of the different chains are exchanging states to move towards the global maximum. The shades of the colours indicates the varying temperatures, i.e. the lighter the colours the higher the temperature. The darkest shade (black) indicates the true ($T = 1$) chain.

Table 5.3: Posterior results and true parameter values for Source 4.2 in multiple EMRI data set.

Parameters	Mean	StdDev	Mode	95% BCI	True Values
$\log(\nu_0)$	-7.9328188	0.0004434	-7.9323975	(-7.9327444, -7.9318485)	-7.9322820
$\log(\mu)$	2.2395396	0.0751304	2.2911850	(2.0606785, 2.34511)	2.2664567
$\log(M)$	15.4662469	0.0009966	15.4656986	(15.4640869, 15.4683055)	15.4660340
e_0	0.1358246	0.0059642	0.1336781	(0.1174296, 0.1475948)	0.1325064
$\tilde{\gamma_0}$	3.8447445	1.5515635	5.5369158	(0.2527217, 6.1285982)	5.4586812
Φ_0	2.8943139	0.1097466	3.286550	(1.9055349, 3.5590655)	3.2720448
θ_S	0.9507359	0.2539969	0.5841125	(0.1832194, 1.6954516)	0.3091959
ϕ_S	2.8264892	0.1904191	2.7220846	(1.0136832, 4.5198308)	2.6004854
λ	2.4467195	0.0087116	2.4442997	(2.4232488, 2.484018)	2.4427887
α_0	2.8256644	0.4742363	0.8437258	(0.2502399, 5.3935017)	0.3919867
χ	0.5471154	0.0014681	0.5466646	(0.5429923, 0.5511443)	0.5504105
θ_K	1.3120252	0.1068620	1.1761698	(0.3572251, 2.4022434)	1.1583793
ϕ_K	3.9613204	1.2991888	3.9093443	(0.4007337, 6.1953864)	4.2795129
$\log(D_L)$	-1.2820195	0.1965055	-1.5567342	(-1.7667382, -0.9134321)	-1.0651850

(see next section). Overall the PTMCMC algorithm performed well in recovering EMRI signals in different situations. However, using LISA simulator to derive the LISA response along with MCMC sampler was too time consuming. In order to implement it efficiently, one would either need to use faster computational resources, or on the other hand would need, to find some analytical approximations to the full LISA response.

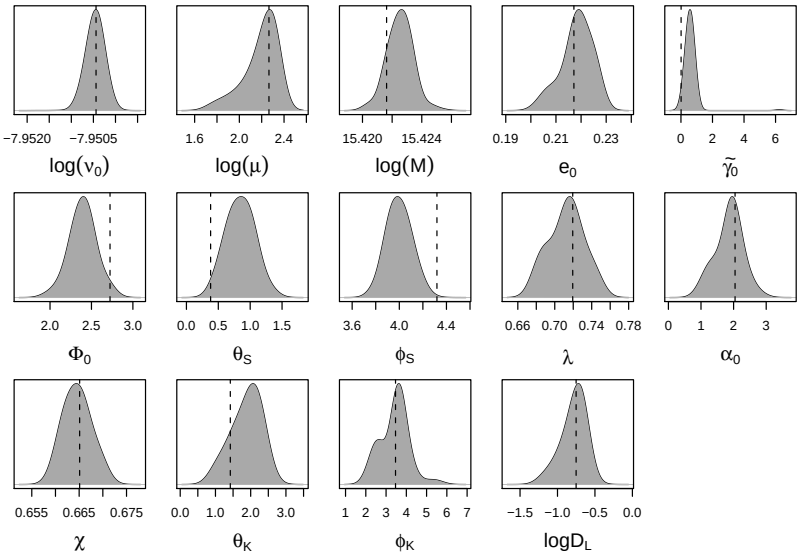

Figure 5.11: Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI source 4.3. The dashed lines indicate the true values.

Table 5.4: Posterior results and true parameter values for Source 4.3 in multiple EMRI data set.

Parameters	Mean	StdDev	Mode	95% BCI	True Values
$\log(\nu_0)$	-7.9504974	0.0002142	-7.9504998	(-7.9507933, -7.9501326)	-7.9504400
$\log(\mu)$	2.1904121	0.1383296	2.2681769	(1.8653783, 2.3426868)	2.2650997
$\log(M)$	15.4225400	0.0009074	15.4224313	(15.4212395, 15.4238013)	15.4216260
e_0	0.2162612	0.0044909	0.2182163	(0.2059087, 0.2267588)	0.2171434
$\tilde{\gamma}_0$	0.5690891	0.4191032	0.5371497	(0.2103864, 0.9056424)	0.0085699
Φ_0	2.4671885	0.1498988	2.4528525	(2.1165191, 2.6567066)	2.7240871
θ_S	0.7285653	0.1930421	0.6745416	(0.5072134, 1.2172905)	0.3796188
ϕ_S	4.0510350	0.1159312	4.0487354	(3.8301058, 4.165885)	4.3193568
λ	0.7198182	0.0126385	0.7142496	(0.6812118, 0.7435473)	0.7194546
α_0	2.7571544	0.2438051	2.6808929	(0.6812118, 0.7435473)	2.0452143
χ	0.6654901	0.0022146	0.6647366	(1.1124029, 2.5159715)	0.6651021
θ_K	0.7753960	0.3010134	0.8130486	(1.047814, 2.4606923)	1.4217963
ϕ_K	2.1553197	2.7038609	0.2541904	(2.305431, 4.5079389)	3.4671808
$\log(D_L)$	-0.7864942	0.1662228	-0.7286971	(-1.111278, -0.5682134)	-0.7514424

5.4.2 Low Frequency Approximations Results

In this setup the low frequency approximation was used, which is very useful to approximate the full LISA response if the source's GW has a longer wavelength than the LISA arms. The high and medium mass EMRI

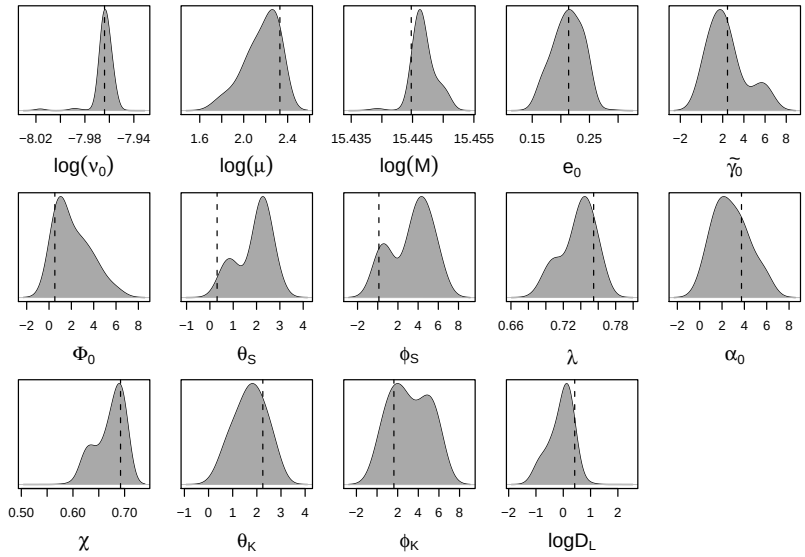

Figure 5.12: Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI source 3.2 in MLDC 3 data. The dashed lines indicates true values.

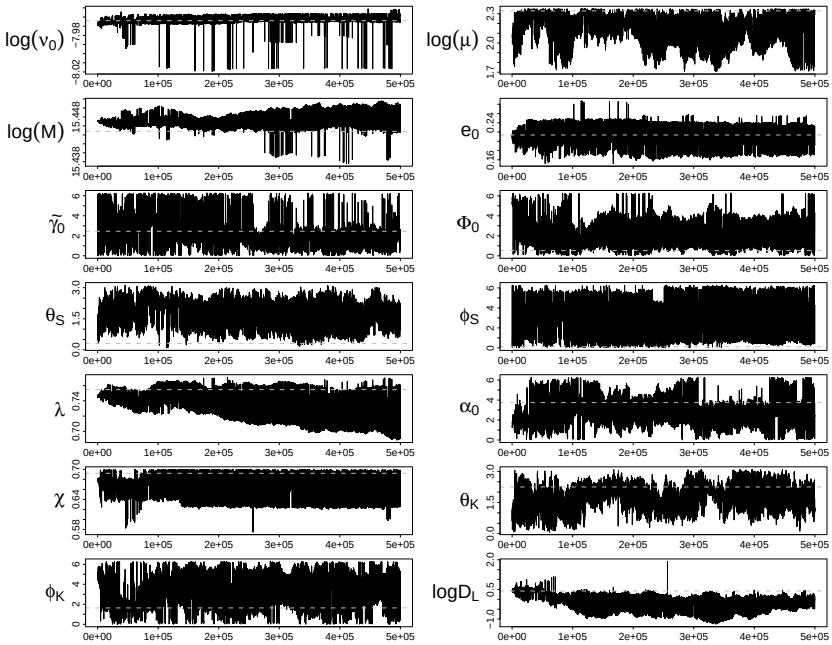

Figure 5.13: Trace plots of the marginal posterior MCMC samples for all 14 parameters for the EMRI source 3.2 in MLDC 3 data. The grey dashed lines indicates true values.

Table 5.5: Posterior summary and true parameter values for Source 3.2 in MLDC 3 blind data set.

Parameters	Mean	StdDev	Mode	95% BCI	True Values
$\log(\nu_0)$	-7.9632318	0.0074434	-7.9634255	(-7.9679815, -7.9565608)	-7.9639512
$\log(\mu)$	2.1400183	0.1637770	2.2610508	(1.8406894, 2.3410743)	2.3295377
$\log(M)$	15.4468872	0.0021757	15.4462224	(15.444836, 15.4504355)	15.4447901
e_0	0.2111625	0.0237776	0.2148893	(15.444836, 15.4504355)	0.2135734
$\tilde{\gamma}_0$	1.9696688	1.3582354	1.7592963	(0.1682269, 0.2462958)	2.4529800
Φ_0	1.9884147	1.4133054	1.0416850	(0.3080034, 5.1105686)	0.5211871
θ_S	1.8360713	0.7443911	2.2706380	(0.6016912, 2.8048535)	0.3093090
ϕ_S	3.4645194	1.9132735	4.3326998	(0.2865014, 5.9520283)	0.1149443
λ	0.7318439	0.0211169	0.7444001	(0.7006846, 0.763016)	0.7547361
α_0	2.6800513	1.4889402	2.1506454	(0.5643185, 5.771482)	3.7418401
χ	0.6727191	0.0279577	0.6896207	(0.6239665, 0.6991228)	0.6921341
θ_K	1.7524283	0.6284395	1.8150416	(0.6113763, 2.75775)	2.2481815
ϕ_K	3.2098031	1.9041195	1.9990239	(0.3246372, 5.9921598)	1.6346445
$\log(D_L)$	-0.2007830	0.3904250	0.1281864	(-0.8988743, 0.4504317)	0.4192122

systems fall in this category. While this approximation is not totally accurate to represent the full LISA response, it is quick and easy to compute. In our case this approximation is $\sim$ 13 times faster than the LISA simulator and has been found very good in estimating the signal parameters of different EMRI sources given in MLDC data. Thus in the new source codes the LISA simulator routines were replaced by the low frequency approximation routines. This new algorithm is fast enough to generate thousands of MCMC samples in a few hours, thus we were able to run our MCMC chains with highly dispersed starting values and with quick follow ups. A bold step that was taken under this setup was smoothing the conditional posterior spectrum of the noise on the fly. Under the previous setup the squared amplitudes or the noise parameters, R_n^2, were randomly drawn from the (conditional) posterior distribution given by Equation (4.5.13), for which the term $a_{nj}^2 + b_{nj}^2$ was estimated directly from the Fourier transformed residuals conditional on the current values of signal parameters, θ, using Equation (4.5.8). The conditional posterior spectrum estimated in this manner has a large variance. To reduce this variance the term $a_{nj}^2 + b_{nj}^2$ can be estimated by Welch's method instead of Equation (4.5.8). As an example let us consider the two conditional posterior spectra estimated for a week long segment (40320 data points) from EMRI 1B.3.2 data set,

which contains an EMRI signal buried in LISA's instrumental noise only, as are given in Figures 5.14 and 5.15. For the conditional posterior spectrum in Figure 5.14 the term $a_{nj}^2 + b_{nj}^2$ was estimated using Equation (4.5.8) and at iteration 500,000 (second box) it has the same large variance as it had at iteration 1 (first box). For the new conditional spectrum the term

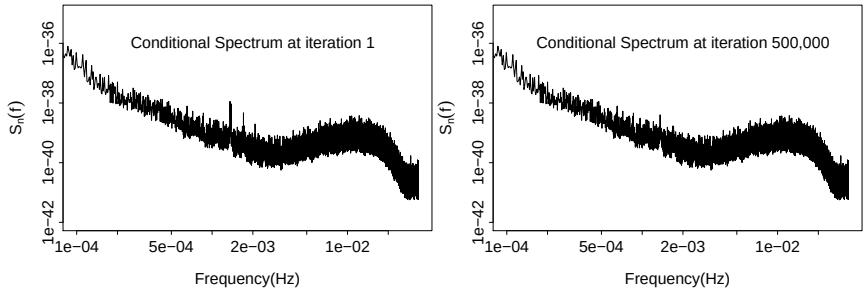

Figure 5.14: The plots of the conditional posterior spectra of noise at iteration 1 (left) and 500,000 (right).

$a_{nj}^2 + b_{nj}^2$ is estimated by Welch's method in which at each MCMC iteration the time domain residual vector is divided into a number of sub-segments of a suitable size, e.g. in this example there were 38 sub-segments each of size 2048 (see Section 4.4.3.1 for full details). The $a_{nj}^2 + b_{nj}^2$ are then used to update the conditional posterior distribution of R_n^2 as given by Equation (4.5.13). Figure 5.15 shows the smoothed conditional posterior spectrum at different iterations. One can see that as the MCMC progresses the spectrum becomes smoother and smoother. Looking at the first box (conditional spectrum at iteration 1) in Figure 5.15, at first glance one may think that the spectrum is still quite rough, but as we know this is the first iteration. Since both the signal parameters, θ, and the noise parameters, R_n^2, are not yet converged, the conditional posterior spectrum is still very rough. As the MCMC chains of different parameters progress towards the global maximum, at the same time the R_n^2 also move towards their best fits, thus the conditional spectrum gets improved. For example in this MCMC run most of the signal parameters as well as the noise parameters found their true stationary distributions within 50,000 iterations and thus conditional mean spectrum has become very smooth and stable. One may also think that this conditional posterior spectrum is now almost constant but

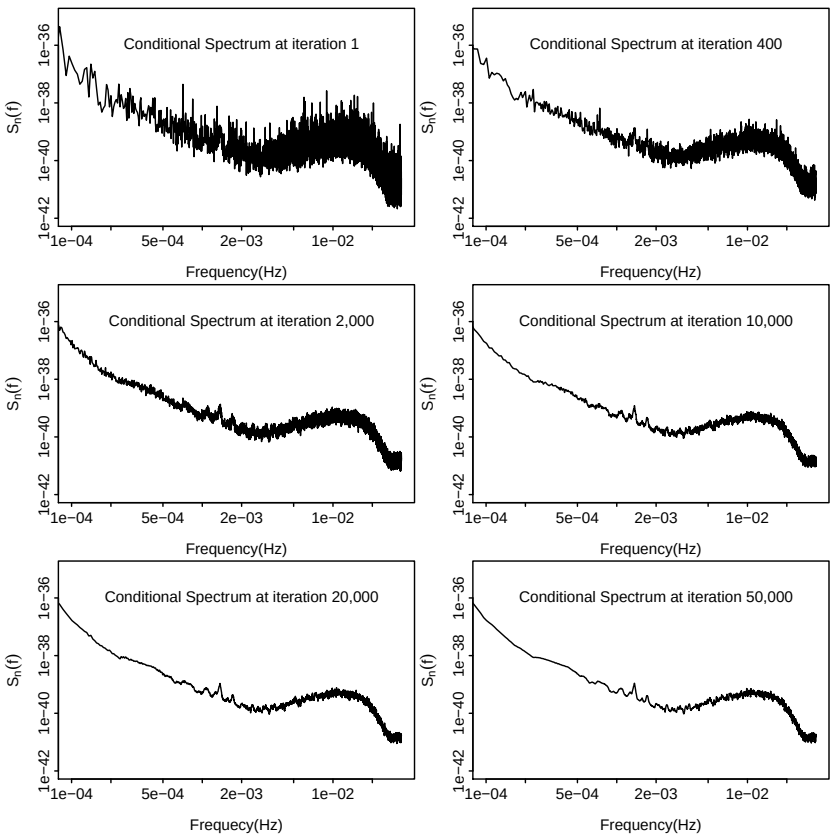

Figure 5.15: These plots reveal how nicely the conditional posterior spectrum is smoothed as MCMC progresses.

it is not the case. It is still being updated in the usual way but it remains stable. This randomness accounts for the unknown characteristics of the noise and it is observed that it also improves mixing. This method could not be utilized in our previous setup; i.e. the algorithm which used LISA simulator because it has slowed down the algorithm further. Using a week long data segment in this new setup, if the spectrum is not updated (i.e. assumed known), the MCMC sampler runs at a speed of ~ 650 iterations per minute, and when the spectrum is updated using Welch's method as explained above, the sampler runs at $\sim 500 - 550$ iterations per minute. This means that this spectrum smoothing comes at the cost of somewhat reduced speed.

In our previous setup, we first tested our algorithm on simulated data and

then applied it to actual MLDC data. But in this setup the algorithm was applied directly to actual MLDC data.

5.4.2.1 Single EMRI

The new algorithm was again applied to recover two different EMRI signals from their respective data sets. One was a high mass source and the other was a medium mass source. In these attempts 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 to test the performance of our algorithm. In both cases the blind searches were conducted in the traditional way, i.e. first several shorter MCMC chains were run in parallel, without 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 that of the others. The modes of these chains were then set as the starting points in a new MCMC run. This step was repeated for a few times to narrow the search range. In the following the posterior results of the parameters of the EMRI sources given in MLDC 1C.3.1 (high mass) and MLDC 1B.3.2 (medium mass) are given. Table 5.6 and Figures 5.16 and 5.16 show the results for source 1C.3.1, and Table 5.7 and Figures 5.18 and 5.19 show the results for source 1B.3.2. From these results it is clear that both signals were recovered to a

Table 5.6: Posterior results and true parameter values for the recovered EMRI signal given in actual MLDC 1C.3.1 training data set.

Parameters	Mean	StdDev	Mode	95% BCI	True Values
$\log(\nu_0)$	-8.5923072	0.0054073	-8.5910440	(-8.6043316, -8.58498)	-8.5914722
$\log(\mu)$	2.0268312	0.1857668	2.2094832	(1.7415931, 2.3226299)	2.3208773
$\log(M)$	16.1109182	0.0013308	16.1143473	(16.0907413, 16.1321731)	16.1193041
e_0	0.2054050	0.0146009	0.2048257	(0.1815489, 0.2266929)	0.1953372
$\tilde{\gamma}_0$	3.4258544	2.0117767	2.1168073	(0.189037, 6.1439493)	4.3815257
Φ_0	3.9488274	1.2730869	3.3662734	(0.9662915, 5.7219864)	3.4411844
θ_S	0.6163426	0.5378233	0.4068807	(0.0635606, 1.8238759)	1.2356774
ϕ_S	2.9447686	1.5804455	2.9152004	(0.8174902, 6.2240241)	4.0547847
λ	1.6522163	0.6050616	2.2917616	(0.7197503, 2.3529236)	2.3589634
α_0	2.6106968	1.8498715	1.2615516	(0.361536, 5.7977946)	2.1583558
χ	0.6126756	0.0518266	0.6509912	(0.5026822, 0.6625549)	0.6366439
θ_K	1.8875384	0.8714571	2.4647494	(0.4725882, 2.9745323)	2.0363598
ϕ_K	2.8270861	2.0937196	1.1248902	(0.1526174, 6.0819968)	1.2601279
$\log(D_L)$	-1.9435258	0.7637236	-1.7958534	(-3.8432567, -0.9393268)	-1.5180422

great accuracy. At first glance these results might look inferior to those obtained with LISA simulator, but in that setup we used very narrow ranges

Table 5.7: Posterior results and true parameter values for the recovered EMRI signal given in actual MLDC 1B.3.2 training data set.

Parameters	Mean	StdDev	Mode	95% BCI	True Values
$\log(\nu_0)$	-7.9618827	0.0015563	-7.9622387	(-7.9626445, -7.9594061)	-7.9622049
$\log(\mu)$	2.1081418	0.1379126	2.1226268	(1.8533068, 2.3236795)	2.2902575
$\log(M)$	15.4293237	0.0145646	15.4327324	(15.4013705, 15.4603173)	15.4319518
e_0	0.2083974	0.0158042	0.2104241	(0.1833913, 0.2294036)	0.2154005
$\tilde{\gamma_0}$	4.3846147	1.2294466	4.8456580	(1.2324418, 5.5431047)	2.0332973
Φ_0	4.6823412	0.6958241	5.0119812	(3.201359, 5.3400583)	5.9999000
θ_S	0.8621684	0.5310470	0.6496126	(0.4922331, 2.1144932)	1.2223304
ϕ_S	2.7107130	1.9050704	1.5804608	(1.0929966, 6.2370637)	2.9346250
λ	2.3318624	0.0680917	2.3028745	(2.2519453, 2.4630376)	2.2899505
α_0	2.8849141	2.4935911	0.6499724	(0.0878025, 6.2103565)	1.6092149
χ	0.5896674	0.0392056	0.5781922	(0.5157521, 0.659082)	0.5748184
θ_K	2.4782385	0.7730288	2.8709623	(0.5979172, 3.0963351)	1.4034163
ϕ_K	3.3736221	2.1551523	5.0501235	(0.2232368, 6.0421006)	6.2231290
$\log(D_L)$	-0.5383502	1.1281699	-0.1358916	(-2.7217512, 0.8325329)	-0.5847797

to choose starting values from, while in this attempt we used quite broad ranges. The width of different marginal posterior densities show that almost the whole prior range was searched before an eventual convergence.

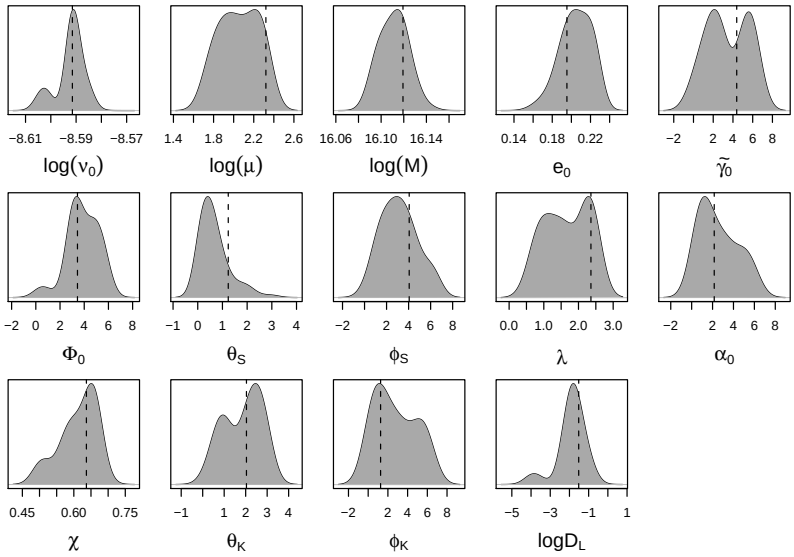

Figure 5.16: Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI source MLDC 1C.3.1 (high mass source). The dashed lines indicate the true values.

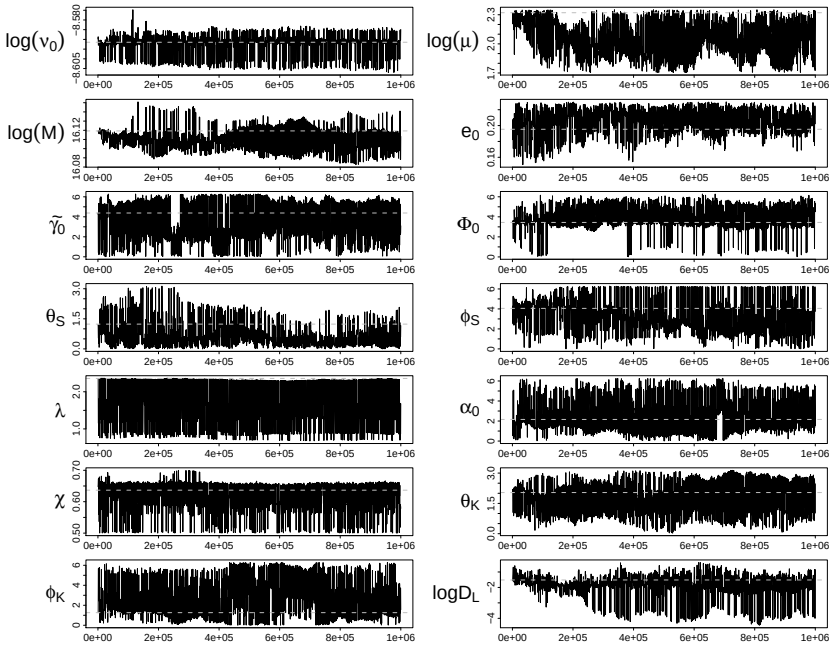

Figure 5.17: Trace plots of the marginal posterior MCMC samples for all 14 parameters for the EMRI source MLDC 1C.3.1 (high mass source). The grey dashed lines indicate the true values.

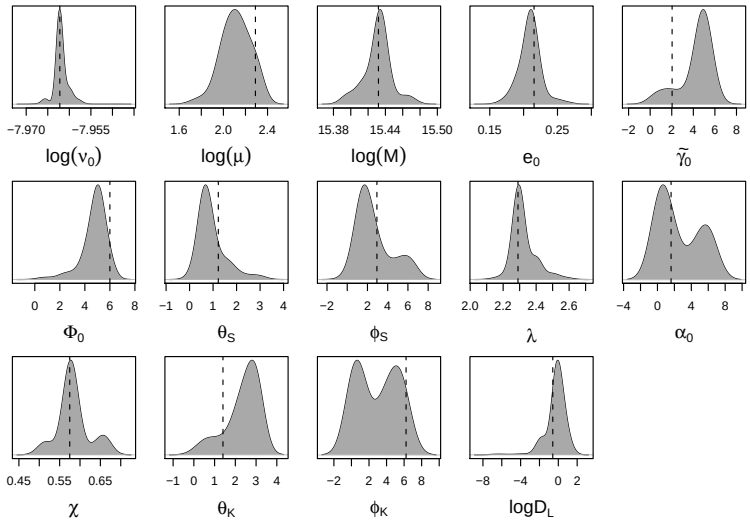

Figure 5.18: Kernel density estimates of the marginal posterior densities for all 14 parameters for the EMRI source MLDC 1B.3.2 (medium mass source). The dashed lines indicate the true values.

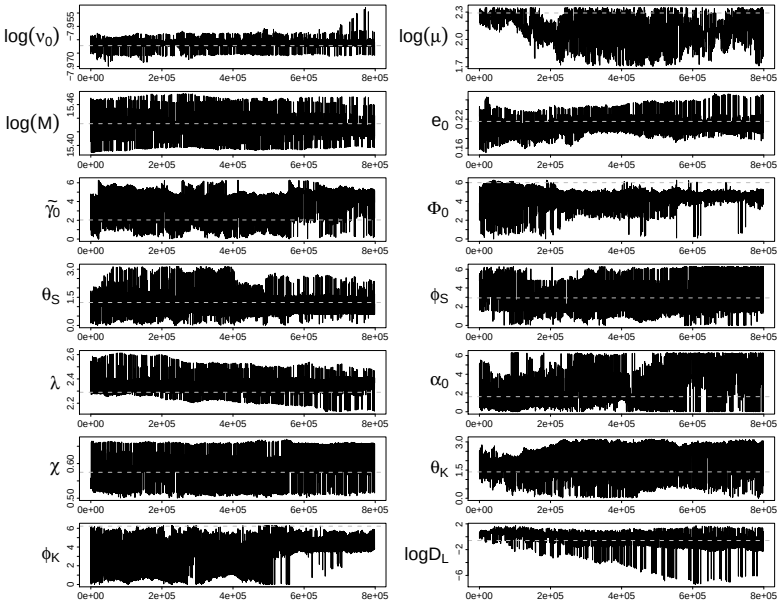

Figure 5.19: Trace plots of the marginal posterior MCMC samples for all 14 parameters for the EMRI source MLDC 1B.3.2 (medium mass source). The grey dashed lines indicate the true values.

5.4.2.2 Multiple EMRIs

MLDC 4 Preliminary Results

The low frequency approach was applied to detect signals generated by EMRI sources given in both training and blind datasets issued in the revised MLDC round 4. Looking at the amount of noise in these data, for now signals from only high mass EMRI systems were attempted to recover. Moreover, from the key files of training EMRI sources listed on MLDC 4 web page [110] it appears that there are no medium mass sources. These searches are still in progress as the deadline for submission of results for this round is June 2011. Some preliminary results were presented at GWPAW (January 26-29, 2011, Milwaukee, Wisconsin, USA). The first few searches on both training and blind data were carried out on local servers while most of the subsequent searches on blind data were conducted on ATLAS.

Training Data

The training data contains three high mass EMRIs sources which are somewhat similar to each other, therefore a joint MCMC search was conducted to recover them. In an 8 chain MCMC, three chains were started from the true parameters' values of the three signals while the rest of the chains were started from the values in the vicinity of the true parameters'. Figures 5.20 and 5.21 display 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 different signals. The solid lines refer to the true parameters' values of the first signal; similarly the dashed and dashed-dot-dashed lines refer to the true parameters' values of the second and third signal respectively. After running for a sufficiently large number of iterations ($\sim$ 4000,000) it was observed that the third signal (dashed-dot-dashed vertical lines) was dominating the other two as can be seen in Figure 5.20, 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 5.8 and Figures 5.22 and 5.23. 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. The distance parameter seems to be over estimated and the sky location (θ_S, ϕ_S) seems to have been locked at a different position. The wrong sky location can be attributed to the fact that now we are using low frequency approximation and are experiencing the same problem as was observed in [38], which also used the low frequency approximation. It is also evident that after some 800,000 iterations one of

Table 5.8: Posterior results for the third EMRI signal given in MLDC 4 training data set.

Parameters	Mean	StdDev	Mode	95% BCI	True Values
$\log(\nu_0)$	-8.6150574	0.0084353	-8.6180363	(-8.6249635, -8.5937387)	-8.6160956
$\log(\mu)$	2.2477959	0.5192954	2.1855669	(1.4335412, 3.218008)	2.3502807
$\log(M)$	16.1524864	0.0022851	16.1523191	(16.1487557, 16.1563142)	16.1533070
e_0	0.1846341	0.0194643	0.1950904	(0.1453335, 0.2032633)	0.1867124
$\tilde{\gamma_0}$	3.8304271	1.4072686	4.7348145	(1.0893358, 5.40843)	5.1388726
Φ_0	3.7702390	1.0793853	3.2443508	(2.1282663, 5.5913499)	2.0847788
θ_S	2.4558709	0.4340432	2.6112671	(1.3119373, 2.7558067)	2.3050326
ϕ_S	3.9017750	0.4852600	3.7807656	(3.5835136, 4.9489593)	4.7079276
λ	1.1157812	0.2241588	1.1800353	(0.5728021, 1.3951766)	1.1556773
α_0	1.9659821	1.3510073	1.3721568	(0.5215196, 5.0080447)	1.7088611
χ	0.5735728	0.0244604	0.5769022	(0.5402039, 0.6168362)	0.5797234
θ_K	1.2352639	0.6854020	1.1279509	(0.0967481, 2.7112745)	1.4297484
ϕ_K	1.7192452	2.2802323	0.3807332	(0.0880124, 5.8993531)	0.7452684
$\log(D_L)$	-1.2627411	0.5966728	-1.2166987	(-2.1874794, -0.1964734)	-2.2992912

the neighbouring chains, of the true ($T = 1$) chain, has found some other mode; however, overall the true chain is unaffected. This second mode can 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 can not be easily concluded; however, as to which case is true. 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.

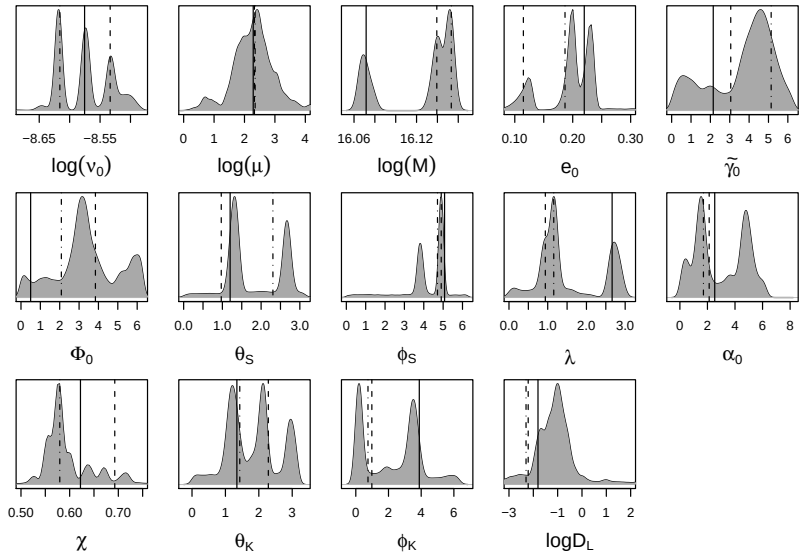

Figure 5.20: Kernel density estimates of the marginal posterior densities for all 14 parameters of the first three EMRI sources in MLDC 4 training set. Different vertical lines indicate the true parameter values of different signals.

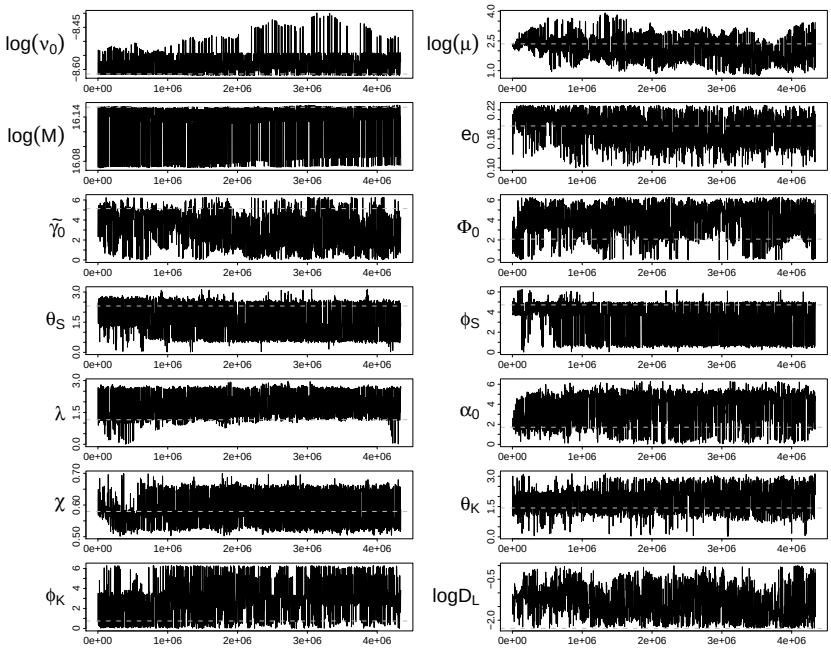

Figure 5.21: Trace plots of the 14 parameters of the first chain obtained for all the three signals in MLDC 4 training data. The dashed lines indicate the true parameter values of the third signal.

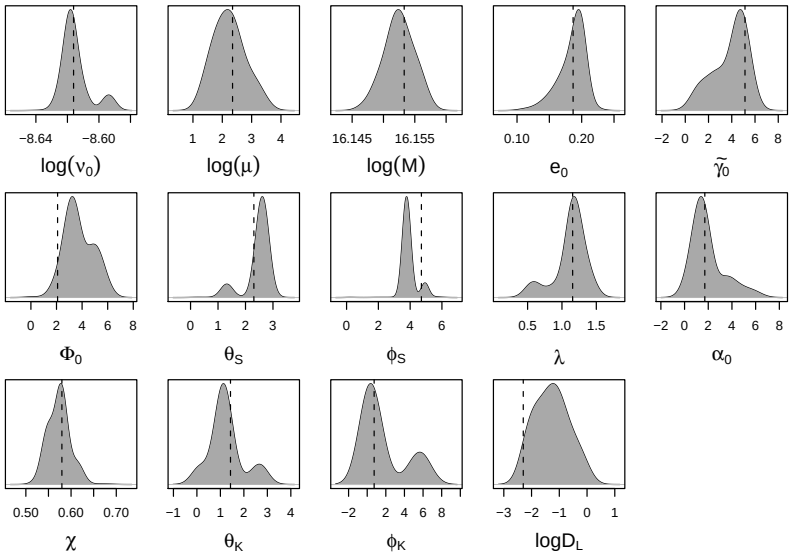

Figure 5.22: Kernel density estimates of the marginal posterior densities for all 14 parameters of the third EMRI signal in MLDC 4 Training data. The dashed lines indicate the true parameter values.

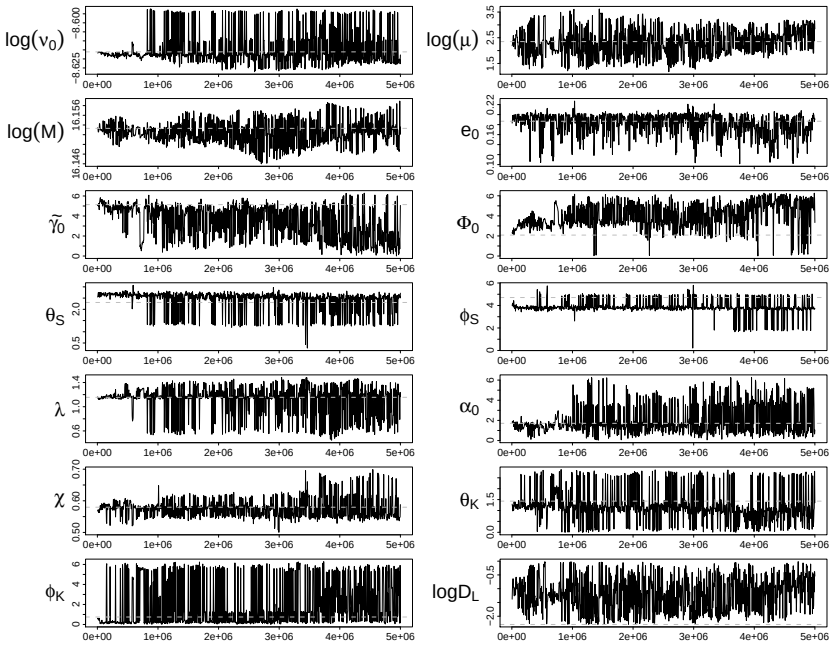

Figure 5.23: Trace plots of posterior samples for all 14 parameters of the first chain obtained for the third signal in MLDC 4 training data. The dashed lines indicate the true parameter values.

Blind Data

An 8 chain MCMC search was conducted on the first two weeks of the MLDC 4 blind set in which all the 8 chains were started from different values, chosen at random. Table 5.9 shows summary of the posterior estimates and Figures 5.24 and 5.25 show the kernel density estimates and trace plots respectively. In these plots the dashed lines indicate the *maximum a posteriori* (MAP) estimates. The trace plots show that the chains for all parameters show great stability except the chains for CO's mass μ and distance $\log D_L$ parameters that have somewhat oscillatory behaviour. The empirical correlation between these two parameters is ~ 0.92, which is quite high. This could not be figured out as to why this (high correlation) happened as such phenomenon was found neither in earlier nor in subsequent searches. On the other hand the SNR and log-likelihood plots as shown in Figure 5.26, show good stability and indicates that there is some signal being picked up. All the angles are also stable except α_0, which is vibrating between two different modes.

Table 5.9: Posterior results for the detected EMRI signal given in MLDC 4 blind data set.

Parameters	Mean	StdDev	Mode	95% BCI
$\log(\nu_0)$	-8.5734565	0.0030150	-8.5732594	(-8.5740346, -8.5723416)
$\log(\mu)$	1.4095992	0.6232590	1.1792795	(0.5891286, 2.2630964)
$\log(M)$	16.1330806	0.0081412	16.1349396	(16.1321258, 16.1368964)
e_0	0.1548801	0.0146666	0.1521146	(0.1391631, 0.1662986)
$\tilde{\gamma}_0$	1.5843590	0.7084836	1.4864924	(0.8599176, 2.2690120)
Φ_0	4.1974981	0.3362512	4.1475526	(3.8341652, 4.5926333)
θ_S	2.5343447	0.0981767	2.5401976	(2.5078450, 2.5794594)
ϕ_S	3.8224231	0.0978049	3.8098774	(3.7449672, 3.8869760)
λ	0.2981064	0.3654142	0.2499777	(0.1507341, 0.3472208)
α_0	3.0728843	2.4917998	5.7613603	(0.1502230, 6.1890025)
χ	0.4999016	0.0092534	0.5016475	(0.4881855, 0.5085584)
θ_K	2.2189922	0.0751287	2.2079507	(2.1613676, 2.2740995)
ϕ_K	1.0463941	0.3700402	0.9866329	(0.8704754, 1.1215218)
$\log(D_L)$	-0.9288239	0.8035098	-0.9860866	(-1.9740756, -0.0185142)

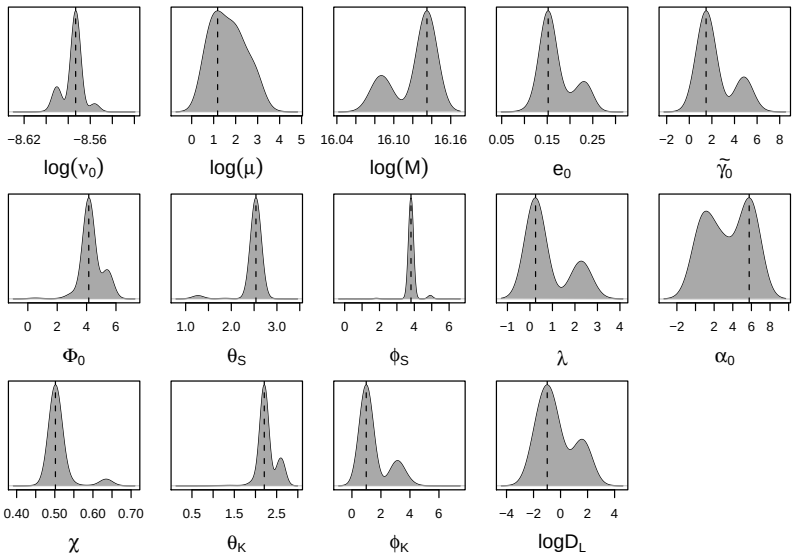

Figure 5.24: Kernel density estimates of the marginal posterior densities for all 14 parameters for the detected EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

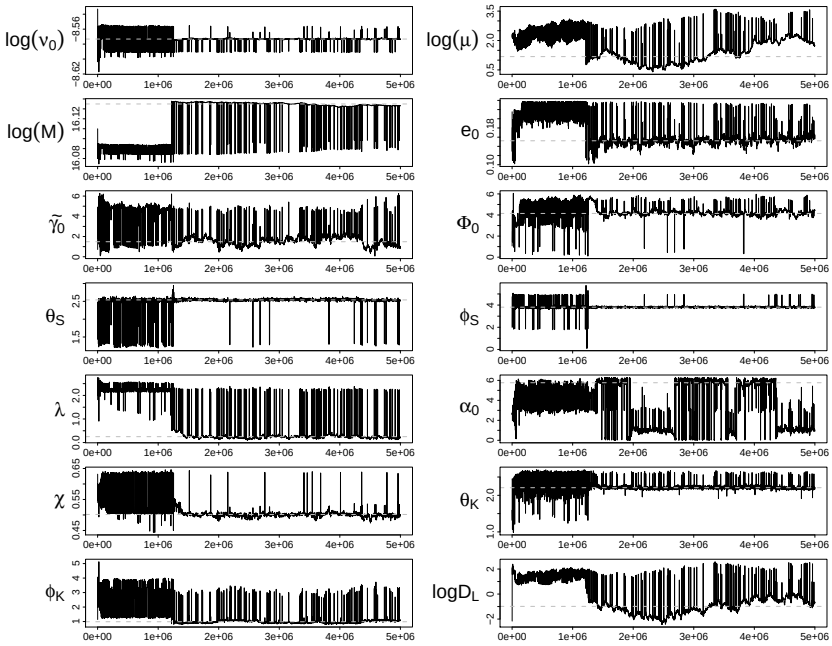

Figure 5.25: Trace plots of the 14 parameters of the first chain for the detected EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

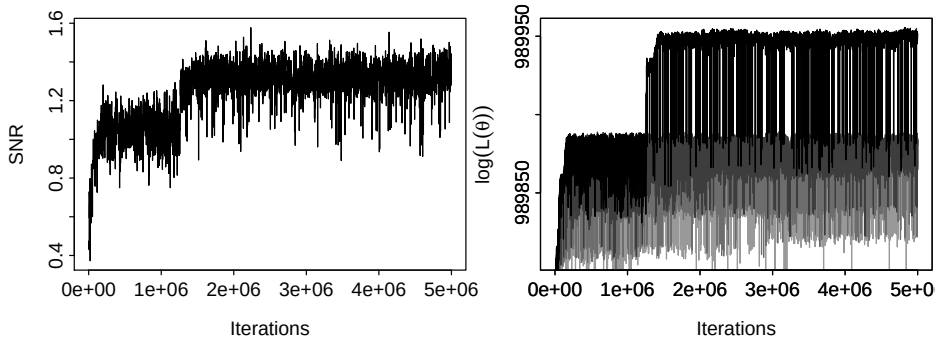

Figure 5.26: (Left) The SNR plot for the detected EMRI signal given in MLDC 4 blind data set. (Right) The Log-Likelihood plot for all 8 chains obtained for the MLDC 4 blind search. The intensity of the grey colour increases with decreasing temperatures. The darkest (black) shade denotes the Log-Likelihood values of the true (cold) chain.

In order to search the entire two years of the blind data several more MCMC searches were conducted on different time regions (data segments). Most of these MCMC searches were carried out on ATLAS. Unfortunately, running our codes on ATLAS was partly successful because of some unknown issues with the configuration of Condor (either locally on user's part or globally) to run MPI jobs. Condor is a software which is used to submit and manage jobs on a cluster of dedicated compute nodes [111, 112]. In most cases the code either crashes after some time or restarts automatically, particularly when the number of chains is large (usually >10). However, we were able to run some of our codes for a number of iterations ranging from half a million to 2 millions. Overall eight time regions were searched, these are the 1st, 2nd, 3rd, 16th, 17th, 53rd, 71st and 81st weeks of the two years data. The time region of the 3rd week was searched by running an 8 chains MCMC on a local server whereas the rest of the seven time regions were each searched by running a 10 chains MCMC on ATLAS. Like the BeSTGRID, different nodes of ATLAS have different processing speeds thus different MCMCs would run at different speeds, despite having similar source codes. The highest speed was $\sim$ 390 iterations per minute whereas the lowest speed was $\sim$ 180 iterations

per minute. Table 5.11 presents the posterior modes of all the 14 parameters obtained for different time regions. The kernel density estimates of the marginal posterior densities for each of the 14 parameters, obtained in these searches, are given in Figures 5.27, 5.28, 5.29, 5.30, 5.31, 5.32, 5.33 and 5.34 with the caption of each figure stating the respective time regions. The kernel density estimates of most of the parameters show multi-modality; however, in most cases the strongest modes can be clearly recognized. In all these searches the average regular Metropolis acceptance rates were in $20-40\%$ while the average swap acceptance rates were in $16\%-30\%$. An interesting result which was observed in these searches is that for all these time regions the joint plots of the sky location angles indicate a similar behaviour, though in the plots of the kernel density estimates of these two angles obtained for different time regions the (strongest) modes are different. For different time regions the joint plots of the two sky location angles are given in Figure 5.35. These plots indicate that there are four strong sky locations. The sky location of the third week search reveals that this MCMC has completely missed the other two sky locations although the code was the same except the searched time region and starting values, which can be attributed either to small number (8) of MCMC chains or that one of the signals is dominating the others. This search also has a low SNR as compared to others, see Figure 5.36 for SNRs obtained for different time regions, this is somewhat natural as we know when the angle θ_S increases, the direction of the incoming signal becomes more horizontal to the plane of LISA, which leads to lower the SNR. However, it can still be considered as a detection, because the estimated signal parameters show similarity with those obtained for the neighbouring time region: the second week. In all of these searches (except the one conducted on the third time region) the strongest sky positions, for which the SNR is also high, are the two that are located near the lower end of θ_S. The estimated strongest sky positions (see kernel density plots) for the 16th and 17th weeks are different from each other despite the two time regions are consecutive. However, in these two time regions most of the other estimated signal parameters are the same, except the orbital frequency parameter ν_0 and two of the angles,

i.e. α_0 and θ_K, that have different modes in the two searches. The multiple modes on the surface of the kernel density estimates for different parameters indicate that there are different EMRI signals of approximately the same strength, which would lead to confusion between any two different signals. Since the orbital parameters of the source evolve over time, for the same signal their values are slightly different in different time regions. The estimated values of these parameters obtained for different time regions, other than the first few weeks, can be projected backward using the true time to plunge once the MLDC task force reveals the true initial values of these parameters. At the moment it is difficult to say something exact about the estimated parameters of these detected signals as these strong modes may not necessarily belong to a single signal. However, for almost all time regions the SNR is large enough to claim a detection. The best estimates of the key parameters, i.e. ν, μ, M, e, (θ_S, ϕ_S), χ and D_L, for the four possible signals, that can be inferred from these results are given in Table 5.10. These estimates do not belong to a particular signal, rather these are the best estimates of these parameters for all the four possible signals given in MLDC blind data. The results of these searches can be used to do more refined searches by narrowing the search ranges around the MAP estimates and other stronger local modes. However, as stated before some degree of confusion between different EMRI signals will still be inevitable.

Table 5.10: Best fit parameter values for the four possible signals in MLDC blind data.

Parameters	Best fit values			
$\log(\nu_0)$	-8.57	-8.56	-8.54	
$\log(\mu)$	2.30	2.34		
$\log(M)$	16.07	16.10	16.13	
e_0	0.10	0.13	0.17	0.20
(θ_S, ϕ_S)	(0.58, 0.65)	(1.28, 4.98)	(1.88, 1.80)	(2.54, 3.72)
χ	0.53	0.57	0.65	
$\log(D_L)$	-2.11	-1.75	-1.44	-1.20
Plunge Time (in seconds)	46000000	54000000	59000000	

Table 5.11: Posterior modes obtained for the detected EMRI signals in different time regions of MLDC 4 blind data set.

Parameters	1st Week	2nd Week	3rd Week	16th Week	17th Week	53rd Week	71st Week	81st Week
$\log(\nu_0)$	-8.5624469	-8.5420324	-8.5469249	-8.5420113	-8.5341293	-8.5778874	-8.5695926	-8.6022630
$\log(\mu)$	2.3015909	2.3069223	2.3493141	2.3177382	2.2917084	2.3142853	2.3070914	2.3042908
$\log(M)$	16.1009479	16.1036831	16.1385529	16.0707744	16.0720299	16.1063878	16.1058961	16.0858657
e_0	0.1936360	0.1083228	0.1270810	0.1871121	0.1853625	0.2042374	0.1795685	0.1693156
$\tilde{\gamma}_0$	5.8030712	5.6343736	3.1625650	0.8702468	1.0640134	3.9272032	0.3987563	1.5926921
Φ_0	5.0889025	3.6028709	2.4193053	1.7215991	1.2743627	3.9369086	2.8742619	0.5113755
θ_S	1.8442185	1.8178380	2.5426859	1.8852653	0.5875816	1.8201771	1.8739365	1.2827079
ϕ_S	1.7715848	1.8022330	3.7290092	1.8271925	0.6564673	1.8014962	1.8344336	4.9810740
λ	1.9593972	2.0217854	0.5695286	1.9574667	1.6515278	2.2494780	2.0257953	0.9366183
α_0	0.9655049	0.8874571	1.3664558	1.6248148	4.7498712	4.3170717	1.0580947	2.0978468
χ	0.5716506	0.5779466	0.5753168	0.5578527	0.5645401	0.5500480	0.5632734	0.5717796
θ_K	1.5610164	2.8287502	1.3144399	2.1920370	0.6573303	2.5737013	0.9627854	1.2582156
ϕ_K	0.4157578	5.3910818	3.3135419	5.6287214	5.9317023	1.2340913	5.8382786	2.2728935
$\log(D_L)$	-1.2572572	-1.4181853	-1.1923971	-1.7557384	-2.1022674	-1.4752333	-2.1172415	-2.1300065

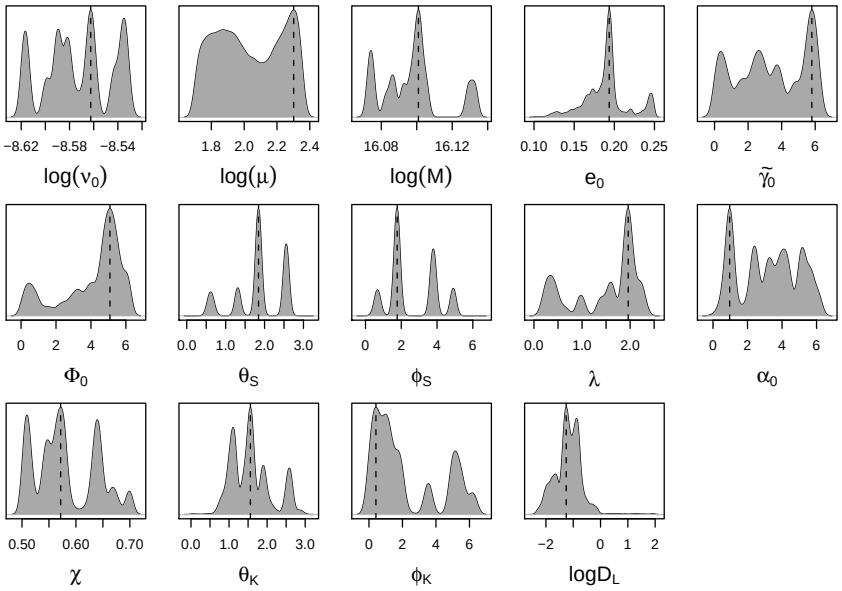

Figure 5.27: 1st Week: Kernel density estimates of the marginal posterior densities for all 14 parameters for an EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

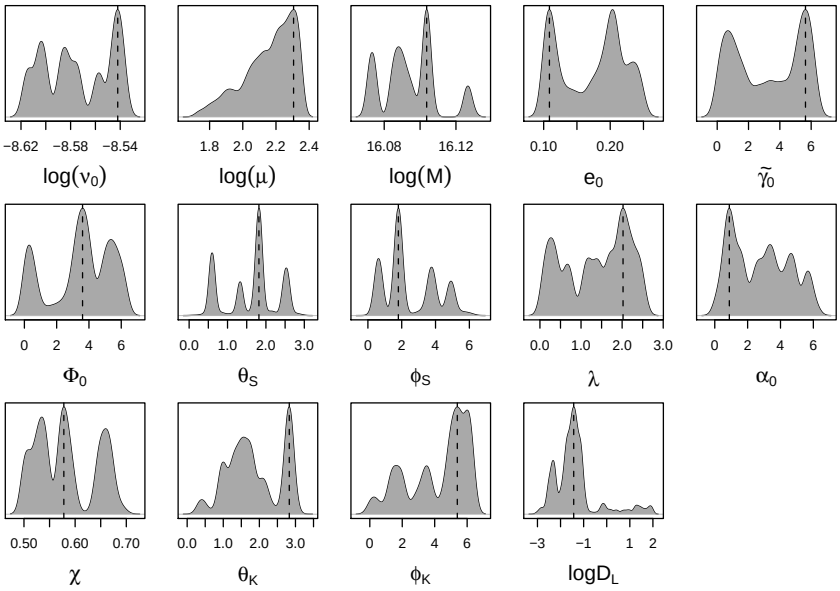

Figure 5.28: 2nd Week: Kernel density estimates of the marginal posterior densities for all 14 parameters for an EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

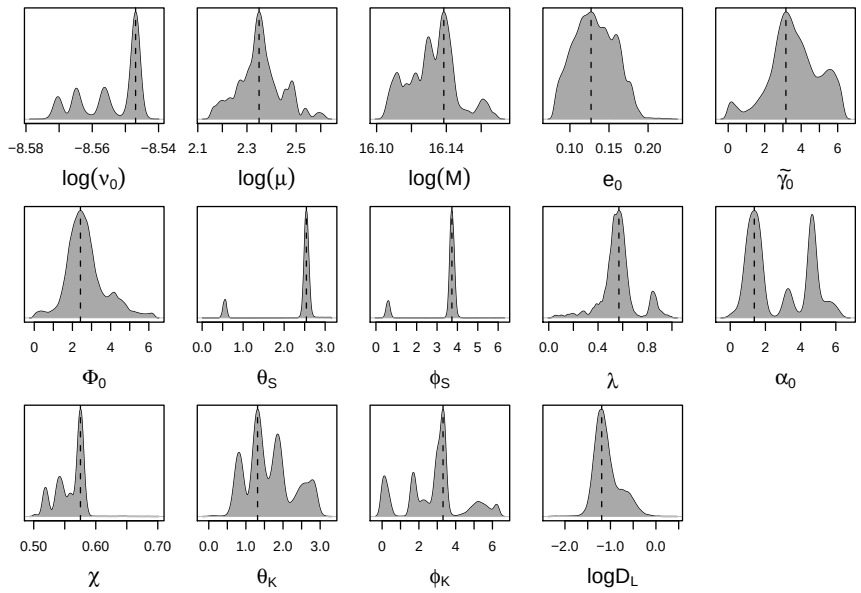

Figure 5.29: 3rd Week: Kernel density estimates of the marginal posterior densities for all 14 parameters for an EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

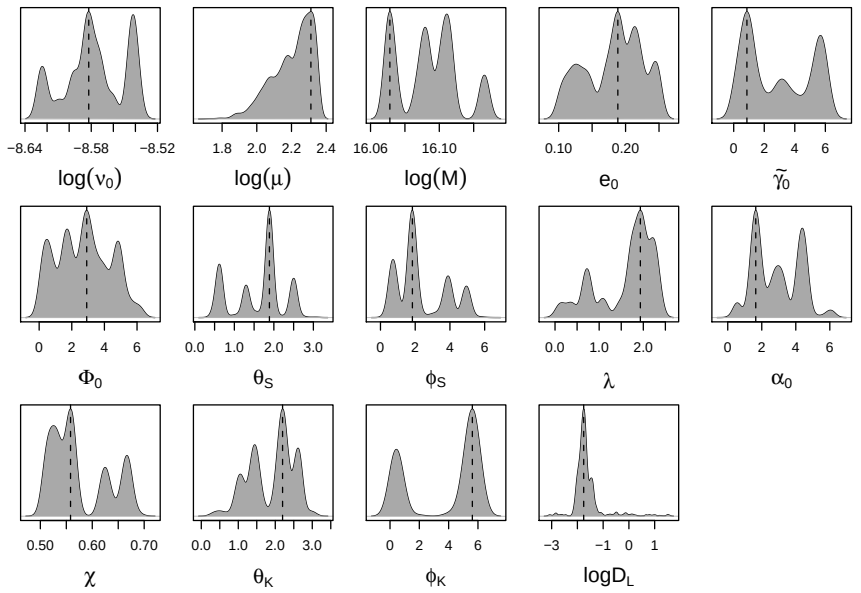

Figure 5.30: 16th week: Kernel density estimates of the marginal posterior densities for all 14 parameters for an EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

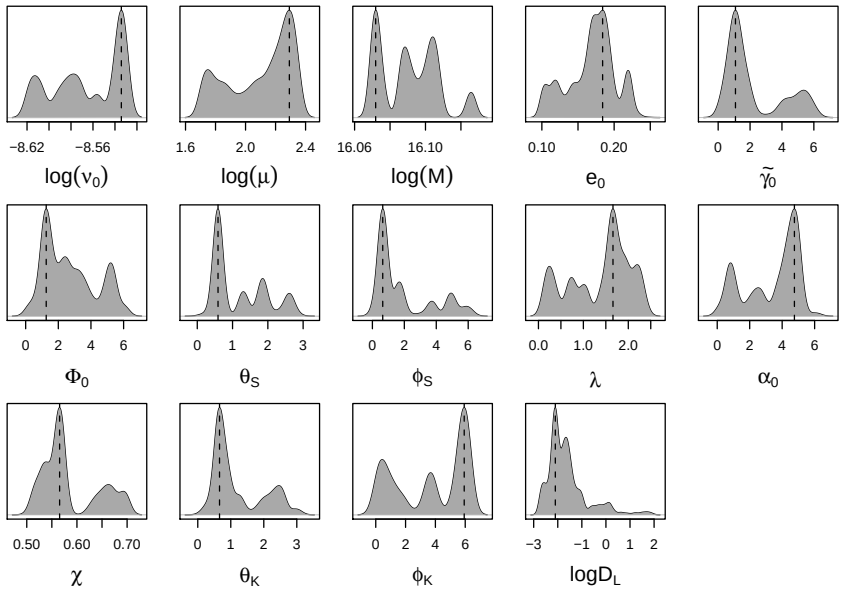

Figure 5.31: 17th week Week: Kernel density estimates of the marginal posterior densities for all 14 parameters for an EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

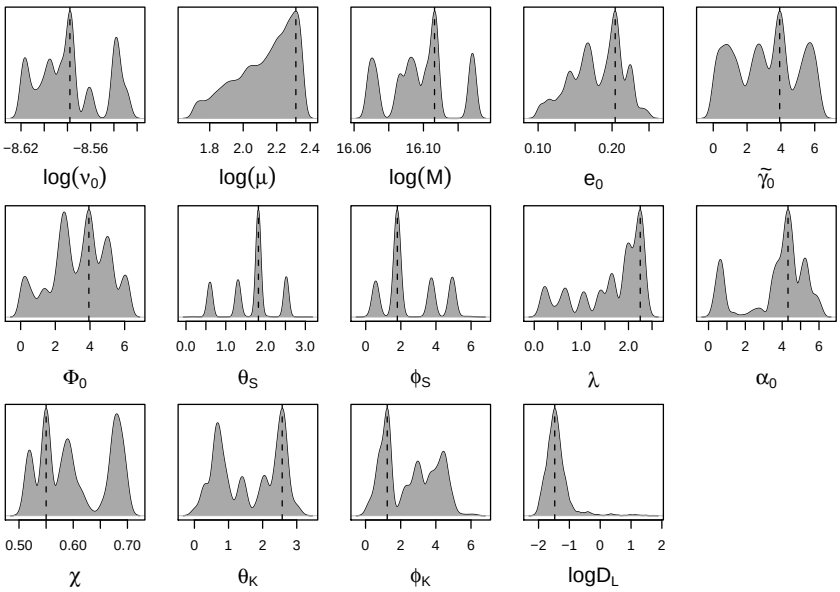

Figure 5.32: 53rd week: Kernel density estimates of the marginal posterior densities for all 14 parameters for an EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

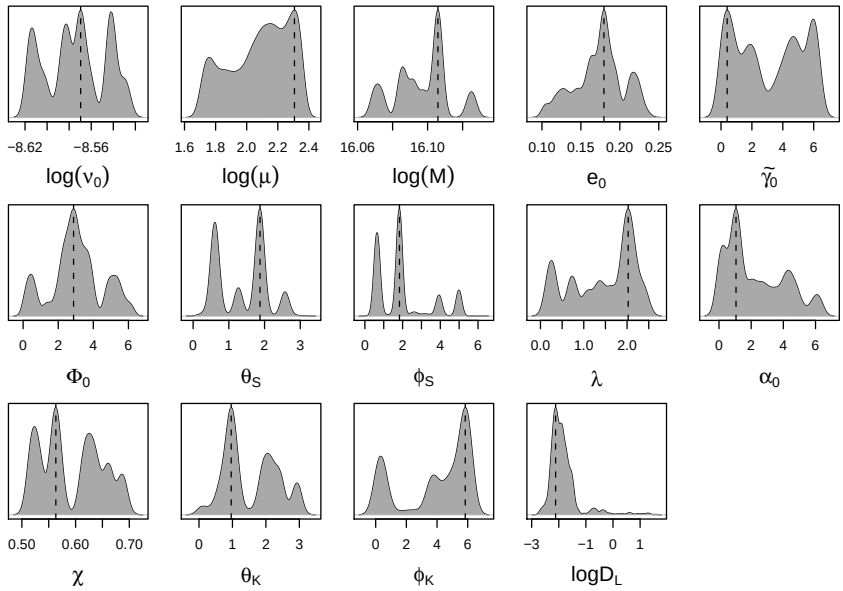

Figure 5.33: 71st Week: Kernel density estimates of the marginal posterior densities for all 14 parameters for an EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

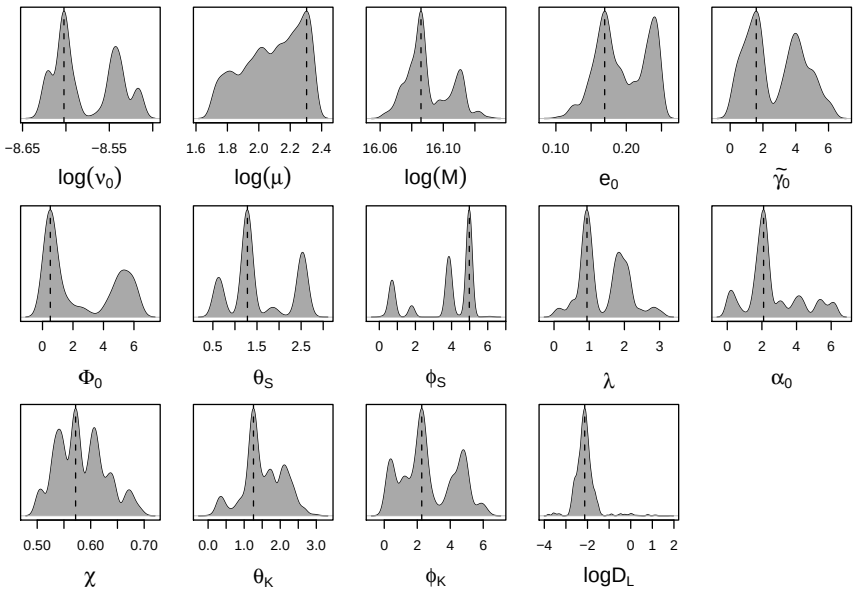

Figure 5.34: 81st Week: Kernel density estimates of the marginal posterior densities for all 14 parameters for an EMRI source in MLDC 4 blind data. The dashed lines indicates MAP values.

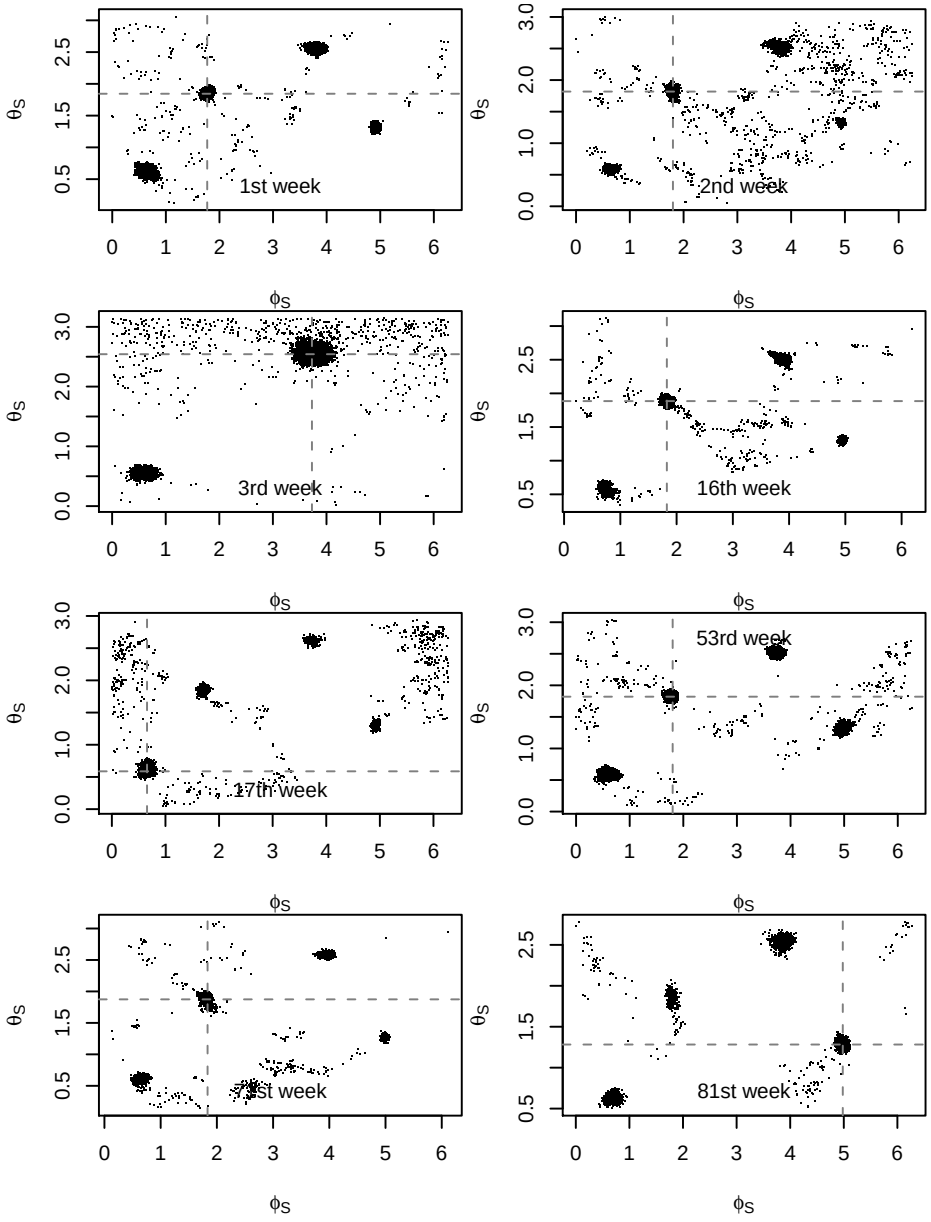

Figure 5.35: The joint plots of sky location angles for different time regions demonstrate that there are four most probable sky positions either of the same source or there are two or more EMRI sources located in different sky regions. The dashed lines indicate the MAP values.

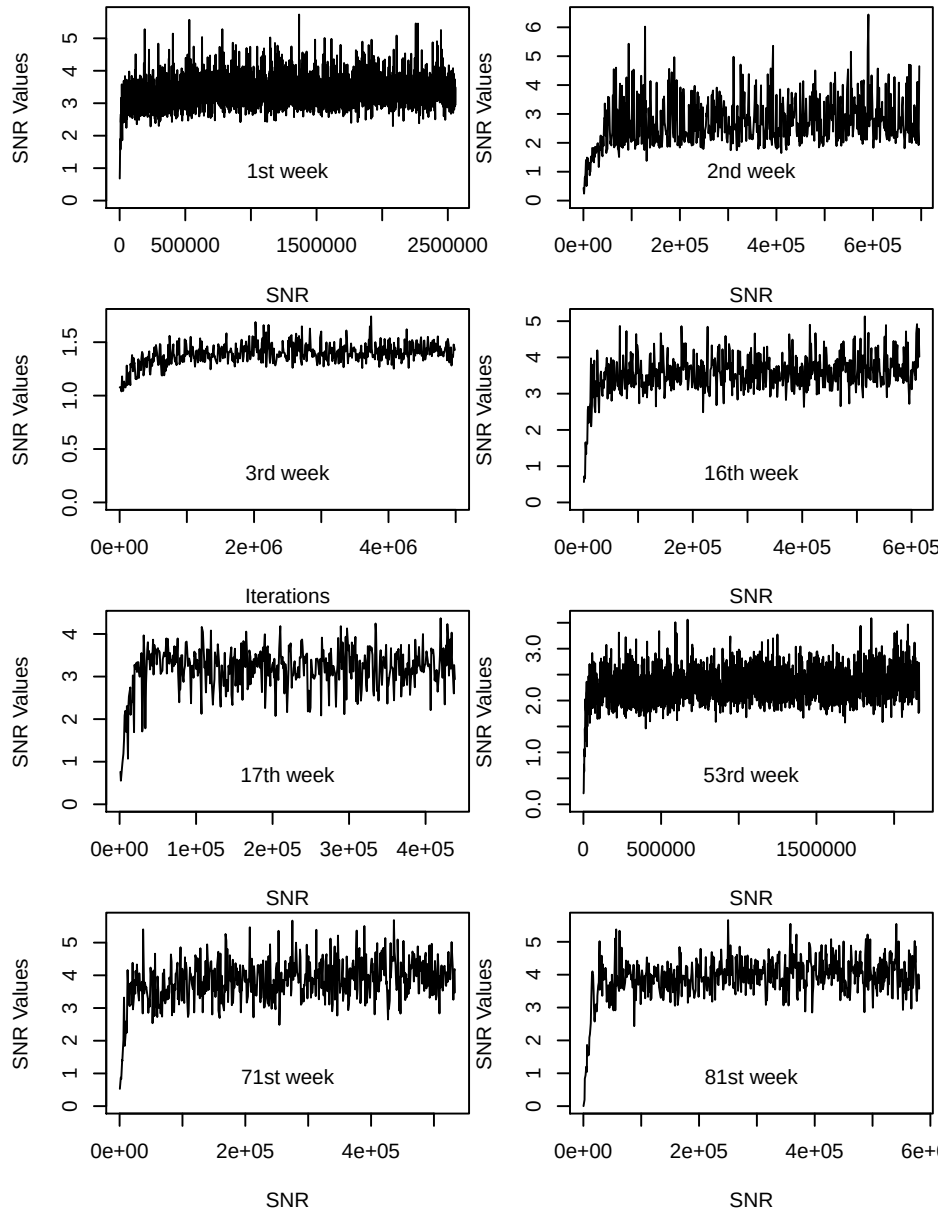

Figure 5.36: The SNR plots of the detected signals in different time regions of the blind data. The SNR obtained for the third week is low as compared to other searches most probably because of the corresponding sky location.

Chapter 6

Conclusions

We have explored the application of Bayesian methods to address both the signal detection problem and the estimation problem of the characteristics of the detected signals in the context of EMRI sources in LISA data. The methodology developed here was applied in different situations depending on the complexity of the underlying data.

The large parameter space and complicated multi-modal posterior surface of EMRI parameters make them an ideal testing ground for advanced MCMC methods such as PTMCMC. In our searches PTMCMC algorithm proved to be very effective in exploring the posterior distribution of EMRI parameters with great efficiency.

The waveform models that are currently being used for the realization of EMRI signals i.e. AKW and TAKW are theoretically the simplest and computationally fastest among others, however these are still time consuming considering the capabilities of the current computing resources and in order to search large data segments computationally faster models will be needed for the efficient analysis of these signals.

In principal, using the LISA simulator is a "near to realistic" option for the derivation of the full LISA response, however it is computationally

quite expensive and will need fast hardware with large memory in order to be effectively used in MCMC searches. On the other hand, the low frequency approximation, which is quicker and easier to implement, is a good alternative to approximate the full LISA response if the target signals belong to the family of long wavelength gravitational waves, i.e. the high and medium mass EMRI sources.

The Bayesian spectrum analysis methods, researched in this work, can be effectively used to characterize the unknown noise spectrum in a very logical way and along with this the use of the Whittle's likelihood is very sensible as it explicitly accommodates the noise spectrum as an unknown quantity to be simultaneously inferred along with the signal parameters.

Overall the algorithm performed very well in all situations and the results show that this algorithm is a very promising one for conducting Bayesian inference on EMRI signals and, in future research, signals from other GWs sources such as white dwarfs and black hole binaries by using their relevant waveform models, in realistic LISA data.

In general, the low frequency approach proved to be very effective in all situations and good results were obtained for the ongoing MLDC 4 searches. One of the three high mass EMRI signals given in the training data was recovered with a great accuracy. One probable signal was detected in the blind data set as well. These searches are still in progress and it is hoped that before the deadline of MLDC 4 results submission more signals will be detected, provided enough computing sources are available. In the presence of multiple overlapping EMRI signals some degree of confusion between different EMRI sources will be inevitable. This remains an area of future research.

APPENDIX A

APPENDIX

A.1 Amplitude Coefficients of Truncated EMRI Model

The amplitudes A_{+1}, A_{+2}, $A_{\times 1}$, and $A_{\times 2}$ that appear in TAKW model are derived as following. The notations are rather very non-standard because this model is not available in published form.

$$\begin{aligned} c_X &= \cos\theta_S \cos\theta_K + \sin\theta_S \sin\theta_K (\cos\phi_S \cos\phi_K + \sin\phi_S \sin\phi_K) \\ s_X &= s_{X_1} - s_{X_2} - s_{X_3} - s_{X_4} \end{aligned} \tag{A.1.1}$$

where

$$\begin{aligned} s_{X_1} &= [\sin\theta_S \cos\theta_K + \cos\theta_S \sin\theta_K + \sin\theta_S \cos\phi_S \sin\theta_K \sin\phi_K \\ &\quad + \sin\theta_S \sin\phi_S \sin\theta_K \cos\phi_K]^{1/2} \\ s_{X_2} &= 2\sin\theta_S \sin\phi_S \cos\theta_S \cos\theta_K \sin\theta_K \sin\phi_K \\ s_{X_3} &= 2\cos\theta_S \sin\theta_K \cos\phi_K \sin\theta_S \cos\phi_S \cos\theta_K \\ s_{X_4} &= 2\sin^2\theta_S \cos\phi_S \sin^2\theta_K \sin\phi_K \sin\phi_S \cos\phi_K \end{aligned}$$

$$
\begin{aligned}
A_{pc_1} &= (-\cos\theta_K \cos\phi_K \sin\theta_S \cos\phi_S - \cos\theta_K \sin\phi_K \sin\theta_S \sin\phi_S \\
&\quad + \sin\theta_K \cos\theta_S)/s_X \\
A_{ps_1} &= (\sin\phi_K \sin\theta_S \cos\phi_S - \cos\phi_K \sin\theta_S \sin\phi_S)/s_X \\
A_{pcn_1} &= \frac{\cos\lambda}{\sin\lambda}(\cos\theta_K \cos\theta_S + \sin\theta_K \cos\phi_K \sin\theta_S \cos\phi_S \\
&\quad + \sin\theta_K \sin\phi_K \sin\theta_S \sin\phi_S - c_X)/s_X
\end{aligned} \tag{A.1.2}
$$

$$
\begin{aligned}
A_{pc_2} &= (\sin\theta_S \cos\phi_S \sin\phi_K - \sin\theta_S \sin\phi_S \cos\phi_K)\cos\lambda/(s_X) \\
A_{ps_2} &= \cos\lambda(\cos\theta_K \cos\phi_K \sin\theta_S \cos\phi_S \\
&\quad + \cos\theta_K \sin\phi_K \sin\theta_S \sin\phi_S - \cos\theta_S \sin\theta_K)/(s_X) \\
A_{pcn_2} &= 0.
\end{aligned} \tag{A.1.3}
$$

$$
\begin{aligned}
A_{qc_1} &= (\sin\theta_S \cos\phi_S \sin\phi_K - \sin\theta_S \sin\phi_S \cos\phi_K)\frac{c_X}{s_X} \\
A_{qs_1} &= (cos\theta_K \cos\phi_K \sin\theta_S \cos\phi_S + \\
&\quad \cos\theta_K \sin\phi_K \sin\theta_S \sin\phi_S - \cos\theta_S \sin\theta_K)\frac{c_X}{s_X} \\
A_{qcn_1} &= 0.
\end{aligned} \tag{A.1.4}
$$

$$
\begin{aligned}
A_{qc_2} &= \cos\lambda(\cos\theta_K \cos\phi_K \sin\theta_S \cos\phi_S \\
&\quad + \cos\theta_K \sin\phi_K \sin\theta_S \sin\phi_S - \sin\theta_K \cos\theta_S)\frac{c_X}{s_X} \\
A_{qs_2} &= -\cos\lambda \sin\theta_S(\sin\phi_K \cos\phi_S - \cos\phi_K \sin\phi_S)\frac{c_X}{s_X} \\
A_{qcn_2} &= -(c_X \cos^2\lambda(\cos\theta_K \cos\theta_S + \\
&\quad \sin\theta_K \cos\phi_K \sin\theta_S \cos\phi_S + \sin\theta_K \sin\phi_K \sin\theta_S \sin\phi_S) \\
&\quad +1 - c_X^2 - \cos^2\lambda)/(s_X \sin\lambda)
\end{aligned} \tag{A.1.5}
$$

$$
\begin{aligned}
B_{p_1c_1} &= 2(A_{pc_1}A_{pcn_1} - A_{qc_1}A_{qcn_1} + A_{qc2}A_{qcn_2} - A_{pc_2}A_{pcn_2}) \\
B_{p1c2} &= \frac{1}{2}(A^2_{ps_2} - A^2_{qc_1} + A^2_{pcn_1} - A^2_{ps_1} + A^2_{qc_2} + A^2_{qs_1} - A^2_{pc_2} - A^2_{qs_2}) \\
B_{p_1s_1} &= 2(A_{qs_2}A_{qcn_2} - A_{ps_2}A_{pcn_2} - A_{qs_1}A_{qcn_1} + A_{ps_1}A_{pcn_1}) \\
B_{p_1s_2} &= (A_{pc_1}A_{ps_1} + A_{qc_2}A_{qs_2} - A_{pc_2}A_{ps_2} - A_{qc_1}A_{qs_1}) \\
B_{p_1cn} &= \frac{1}{2}(A^2_{pc_1} + A^2_{ps_1} - A^2_{qc_1} - A^2_{qs_1} - A^2_{pc_2} + A^2_{qc_2} + A^2_{qs_2} - A^2_{ps_2}) \\
&\quad + A^2_{qcn_2} - A^2_{qcn_1} + A^2_{pcn_1} - A^2_{pcn_2}
\end{aligned} \tag{A.1.6}
$$

$$
\begin{aligned}
B_{p_2c_1} &= (A_{pcn_1}A_{pc_2} + A_{pc_1}A_{pcn_2} - A_{qcn_1}A_{qc_2} - A_{qc_1}A_{qcn_2}) \\
B_{p_2c_2} &= \frac{1}{2}(A_{qs_1}A_{qs_2} - A_{ps_1}A_{ps_2} + A_{pc_1}A_{pc_2} - A_{qc_1}A_{qc_2}) \\
B_{p_2s_1} &= (A_{ps_1}A_{pcn_2} + A_{pcn_1}A_{ps_2} - A_{qcn_1}A_{qs_2} - A_{qs_1}A_{qcn_2}) \\
B_{p_2s_2} &= \frac{1}{2}(A_{pc_1}A_{ps_2} - A_{qc_1}A_{qs_2} + A_{ps_1}A_{pc_2} - A_{qs_1}A_{qc_2}) \\
B_{p_2cn} &= \frac{1}{2}(A_{ps_1}A_{ps_2} - A_{qs_1}A_{qs_2} - A_{qc_1}A_{qc_2} + A_{pc_1}A_{pc_2}) \\
&\quad - A_{qcn_1}A_{qcn_2} + A_{pcn_1}A_{pcn_2}
\end{aligned} \tag{A.1.7}
$$

$$
\begin{aligned}
B_{c_1c_1} &= (-A_{pc_2}A_{qcn_2} - A_{pcn_2}A_{qc_2} + A_{pc_1}A_{qcn_1} + A_{pcn_1}A_{qc_1}) \\
B_{c_1c_2} &= \frac{1}{2}(A_{pc_1}A_{qc_1} - A_{ps_1}A_{qs_1} - A_{pc_2}A_{qc_2} + A_{ps_2}A_{qs_2}) \\
B_{c_1s_1} &= (A_{pcn_1}A_{qs_1} - A_{ps_2}A_{qcn_2} + A_{ps_1}A_{qcn_1} - A_{pcn_2}A_{qs_2}) \\
B_{c_1s_2} &= \frac{1}{2}(-A_{pc_2}A_{qs_2} + A_{pc_1}A_{qs_1} - A_{ps_2}A_{qc_2} + A_{ps_1}A_{qc_1}) \\
B_{c_1cn} &= -A_{pcn_2}A_{qcn_2} + A_{pcn_1}A_{qcn_1} + \frac{1}{2}(A_{pc_1}A_{qc_1} \\
&\quad - A_{ps_2}A_{qs_2} + A_{ps_1}A_{qs_1} - A_{pc_2}A_{qc_2})
\end{aligned} \tag{A.1.8}
$$

$$
\begin{aligned}
B_{c_2c_1} &= (A_{qc_1}A_{pcn_2} + A_{qcn_1}A_{pc_2} + A_{pc_1}A_{qcn_2} + A_{pcn_1}A_{qc_2}) \\
B_{c_2c_2} &= \frac{1}{2}(A_{pc_1}A_{qc_2} - A_{ps_1}A_{qs_2} + A_{qc_1}A_{pc_2} - A_{qs_1}A_{ps_2}) \\
B_{c_2s_1} &= (A_{pcn_1}A_{qs_2} + A_{qs_1}A_{pcn_2} + A_{ps_1}A_{qcn_2} + A_{qcn_1}A_{ps_2}) \\
B_{c_2s_2} &= \frac{1}{2}(A_{qc_1}A_{ps_2} + A_{pc_1}A_{qs_2} + A_{qs_1}A_{pc_2} + A_{ps_1}A_{qc_2}) \\
B_{c_2cn} &= A_{qcn_1}A_{pcn_2} + A_{pcn_1}A_{qcn_2} + \frac{1}{2}(A_{pc_1}A_{qc_2} + \\
&\quad A_{qs_1}A_{ps_2} + A_{ps_1}A_{qs_2} + A_{qc_1}A_{pc_2})
\end{aligned}
\tag{A.1.9}
$$

Finally

$$
\begin{aligned}
A_{+1} &= B_{p_1c_1}\cos\alpha + B_{p_1c_2}\cos 2\alpha + B_{p_1s_1}\sin\alpha + B_{p_1s_2}\sin 2\alpha + B_{p_1cn} \\
A_{+2} &= B_{p_2c_1}\cos\alpha + B_{p_2c_2}\cos 2\alpha + B_{p_2s_1}\sin\alpha + B_{p_2s_2}\sin 2\alpha + B_{p_2cn} \\
A_{\times 1} &= B_{c_1c_1}\cos\alpha + B_{c_1c_2}\cos 2\alpha + B_{c_1s_1}\sin\alpha + B_{c_1s_2}\sin 2\alpha + B_{c_1cn} \\
A_{\times 2} &= B_{c_2c_1}\cos\alpha + B_{c_2c_2}\cos 2\alpha + B_{c_2s_1}\sin\alpha + B_{c_2s_2}\sin 2\alpha + B_{c_2cn}
\end{aligned}
\tag{A.1.10}
$$

INDEX

REFERENCES

[1] MLDC 4 website, . URL `http://astrogravs.nasa.gov/docs/mldc/round4/datasets.html`.

[2] T J Loredo. From laplace to supernova sn 1987a: Bayesian inference in astrophysics. In *Maximum-Entropy and Bayesian Methods*, pages 81–142. Dartmouth, 1989, ed. P. Fougere. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1990.

[3] E T Jaynes. *Probability Theory: The Logic of Science (Vol II)*. Cambridge University Press, April 2003.

[4] P Gregory. *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, New York, NY, USA, 2005. ISBN 052184150X.

[5] L G Bretthorst. *Bayesian Spectrum Analysis and Parameter Estimation, Lecture Notes in Statistics*. Number 48. Springer-Verlag, New York, 1988.

[6] S L Finn and D F Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Phys. Rev. D*, 47(6):2198–2219, Mar 1993.

[7] C Cutler and É E Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? *Phys. Rev. D*, 49(6):2658–2697, Mar 1994.

[8] N Christensen and R Meyer. Markov chain monte carlo methods for bayesian gravitational radiation data analysis. *Phys. Rev. D*, 58(8): 082001, Sep 1998.

[9] N Christensen and R Meyer. Using markov chain monte carlo methods for estimating parameters with gravitational radiation data. *Phys. Rev. D*, 64(2):022001, May 2001.

[10] C Röver, R Meyer, and N Christensen. Bayesian inference on compact binary inspiral gravitational radiation signals in interferometric data. *Classical and Quantum Gravity*, 23(15):4895, 2006.

[11] C Röver, R Meyer, and N Christensen. Coherent bayesian inference on compact binary inspirals using a network of interferometric gravitational wave detectors. *Phys. Rev. D*, 75(6):062004, Mar 2007.

[12] BeSTGRID webpage:. `www.bestgrid.org`, .

[13] J H Taylor and J M Weisberg. Further experimental tests of relativistic gravity using the binary pulsar psr 1913 + 16. *The Astrophysical Journal*, 345:434–450, October 1989.

[14] J M Weisberg, D J Nice, and J H Taylor. Timing measurements of the relativistic binary pulsar psr b1913+16. *The Astrophysical Journal*, 722 (2):1030, 2010.

[15] V Fafone. Resonant-mass detectors: status and perspectives. *Classical and Quantum Gravity*, 21(5):S377, 2004.

[16] O D Aguiar. Past, present and future of the resonant-mass gravitational wave detectors. *Research in Astronomy and Astrophysics*, 11(1): 1, 2011.

[17] J Hough and S Rowan. Laser interferometry for the detection of gravitational waves. *Journal of Optics A: Pure and Applied Optics*, 7(6): S257, 2005.

[18] F J Raab and the Ligo Scientific collaboration. The status of laser interferometer gravitational-wave detectors. *Journal of Physics: Conference Series*, 39(1):25, 2006.

[19] S Hild and the LIGO Scientific Collaboration. The status of geo 600. *Classical and Quantum Gravity*, 23(19):S643, 2006.

[20] D Sigg. Commissioning of ligo detectors. *Classical and Quantum Gravity*, 21(5):S409, 2004.

[21] F Acernese and the VIRGO Collaboration. Status of virgo. *Classical and Quantum Gravity*, 21(5):S385, 2004.

[22] R Takahashi and the TAMA Collaboration. Status of tama300. *Classical and Quantum Gravity*, 21(5):S403, 2004.

[23] A Vecchio. LISA observations of rapidly spinning massive black hole binary systems. *Phys. Rev. D*, 70(4):042001, Aug 2004.

[24] C Cutler. Angular resolution of the LISA gravitational wave detector. *Phys. Rev. D*, 57(12):7089–7102, Jun 1998.

[25] J A Edlund, M Tinto, A Królak, and G Nelemans. White-dwarf–white-dwarf galactic background in the LISA data. *Phys. Rev. D*, 71 (12):122003, Jun 2005.

[26] G Nelemans. The galactic gravitational wave foreground. *Classical and Quantum Gravity*, 26(9):094030, 2009.

[27] E Berti, V Cardoso, and C M Will. Gravitational-wave spectroscopy of massive black holes with the space interferometer LISA. *Phys. Rev. D*, 73(6):064030, Mar 2006.

[28] E Berti. LISA observations of massive black hole mergers: event rates and issues in waveform modelling. *Classical and Quantum Gravity*, 23(19):S785, 2006.

[29] L Barack and C Cutler. LISA capture sources: Approximate waveforms, signal-to-noise ratios, and parameter estimation accuracy. *Phys. Rev. D*, 69(8):082005, Apr 2004.

[30] N Christensen. Measuring the stochastic gravitational-radiation background with laser-interferometric antennas. *Phys. Rev. D*, 46 (12):5250–5266, Dec 1992.

[31] B Allen. The stochastic gravity-wave background: sources and detection. In *Relativistic gravitation and gravitational radiation : proceedings of the Les Houches School of Physics*, pages 373–418, 1997.

[32] S V Dhurandhar M Tinto. Time-delay interferometry. *Living Reviews in Relativity*, 8(4), 2005.

[33] M Vallisneri. A LISA data-analysis primer. *Classical and Quantum Gravity*, 26(9):094024, 2009.

[34] J R Gair, L Barack, T Creighton, C Cutler, S L Larson, E S Phinney, and M Vallisneri. Event rate estimates for LISA extreme mass ratio capture sources. *Classical and Quantum Gravity*, 21(20):S1595, 2004.

[35] L Wen and J R Gair. Detecting extreme mass ratio inspirals with LISA using time–frequency methods. *Classical and Quantum Gravity*, 22(10):S445, 2005.

[36] J R Gair, I Mandel, and L Wen. Time-frequency analysis of extreme-mass-ratio inspiral signals in Mock LISA Data. *Journal of Physics: Conference Series*, 122(1):012037, 2008.

[37] A Stroeer, J Gair, and A Vecchio. Automatic bayesian inference for LISA data analysis strategies. *AIP Conference Proceedings*, 873(1):444–451, 2006.

[38] J R Gair, E K Porter, S Babak, and L Barack. A constrained metropolis–hastings search for emris in the Mock LISA Data Challenge 1B. *Classical and Quantum Gravity*, 25(18):184030, 2008.

[39] S Babak, J R Gair, and E K Porter. An algorithm for the detection of extreme mass ratio inspirals in LISA data. *Classical and Quantum Gravity*, 26(13):135004, 2009.

[40] N J Cornish. Detection strategies for extreme mass ratio inspirals. arXiv:0804.3323v1.

[41] N J Cornish and E K Porter. Catching supermassive black hole binaries without a net. *Phys. Rev. D*, 75(2):021301, Jan 2007. doi: 10.1103/PhysRevD.75.021301.

[42] S A Hughes. (sort of) testing relativity with extreme mass ratio inspirals. *AIP Conference Proceedings*, 873(1):233–240, 2006.

[43] MLDC webpage:. `http://astrogravs.nasa.gov/docs/mldc`, .

[44] K A Arnaud, S Babak, J G Baker, M J Benacquista, N J Cornish, C Cutler, L S Finn, S L Larson, T Littenberg, E K Porter, M Vallisneri, A Vecchio, J-Y Vinet, and The Mock LISA Data Challenge Task Force. An overview of the second round of the Mock LISA Data Challenges. *Classical and Quantum Gravity*, 24(19):S551, 2007.

[45] S Babak, H Fang, J R Gair, K Glampedakis, and S A Hughes. "kludge" gravitational waveforms for a test-body orbiting a kerr black hole. *Phys. Rev. D*, 75(2):024005, Jan 2007.

[46] N Yunes, A Buonanno, S A Hughes, M C Miller, and Y Pan. Modeling extreme mass ratio inspirals within the effective-one-body approach. *Phys. Rev. Lett.*, 104(9):091102, Mar 2010.

[47] N Yunes, S A Hughes A Buonanno, Y Pan, E Barausse, M C Miller, and W Throwe. Extreme mass-ratio inspirals in the effective-one-body approach: Quasi-circular, equatorial orbits around a spinning black hole. 2010.

[48] J W Armstrong, F B Estabrook, and Massimo Tinto. Time-delay interferometry for space-based gravitational wave searches. *The Astrophysical Journal*, 527(2):814, 1999.

[49] T A Prince, M Tinto, S L Larson, and J W Armstrong. LISA optimal sensitivity. *Phys. Rev. D*, 66(12):122002, Dec 2002.

[50] N J Cornish and L J Rubbo. LISA response function. *Phys. Rev. D*, 67:022001, 2003.

[51] L Barack and C Cutler. Confusion noise from LISA capture sources. *Phys. Rev. D*, 70(12):122002, Dec 2004.

[52] N J Cornish and S L Larson. LISA data analysis: Source identification and subtraction. *Phys. Rev. D*, 67(10):103001, May 2003.

[53] R Umstätter, N Christensen, M Hendry, R Meyer, V Simha, J Veitch, S Vigeland, and G Woan. Bayesian modeling of source confusion in LISA data. *Phys. Rev. D*, 72(2):022001, Jul 2005.

[54] C Röver. *Bayesian inference on astrophysical binary inspirals based on gravitational wave measurements*. PhD thesis, The University of Auckland, 2007.

[55] J M Bernardo. Bayesian statistics. In R Viertl, editor, *Encyclopedia of Life Support Systems (EOLSS)*, chapter Probability and Statistics. Oxford, UK: UNESCO.

[56] L Tierney and J B Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986.

[57] A Gelman, J B Carlin, H S Stern, and D B Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2^{nd} edition, 2003. ISBN 158488388X.

[58] W R Gilks, S Richardson, and D J Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC, 1996. ISBN 0412055511.

[59] B P Carlin and T A Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC, 2^{nd} edition, 2000.

[60] C P Robert and G Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 1 edition, 1999.

[61] G O Robert. Markov chain concepts related to sampling algorithms. In W R Gilks, S Richerdson, and D J Spiegelhalter, editors, *Markov chain Monte Carlo in practice*, chapter 3, pages 45–58. Chapman & Hall/CRC, 1996.

[62] L Tierney. Introduction to general state-space markov chain theory. In W R Gilks, S Richerdson, and D J Spiegelhalter, editors, *Markov chain Monte Carlo in practice*, chapter 4, pages 59–74. Chapman & Hall/CRC, 1996.

[63] N Metropolis, A W Rosenbluth, M N Rosenbluth, A H Teller, and E Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[64] J S Liu. *Monte Carlo strategies in scientific computing*. Springer, 2001. ISBN 0387952306.

[65] W K Hastings. Monte carlo samping methods using markov chains and their applications. *Biometrika*, pages 97–109, 1970.

[66] D Gamerman and H F Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC, 2^{nd} edition, 2006.

[67] S Geman and D Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721–741, 1984.

[68] A Gelman, G O Roberts, and W R Gilks. Efficient Metropolis jumping rules. In *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., pages 599–607. Oxford Univ. Press, 1996.

[69] G O Roberts and J S Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.

[70] R E Kass, B P Carlin, A Gelman, and R M Neal. Markov chain monte carlo in practice: A roundtable discussion. *The American Statistician*, 52(2):93–100, 1998.

[71] S P Brooks and A Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

[72] J Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, 1991.

[73] P Heidelberger and P D Welch. A spectral method for confidence interval generation and run length control in simulations. *Commun. ACM*, 24:233–245, April 1981.

[74] P Heidelberger and P D Welch. Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144, 1983.

[75] L W Schruben. Detecting initialization bias in simulation output. *Operations Research*, 30(3):569–590, 1982.

[76] M Plummer, N Best, K Cowles, and K Vines. CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11, March 2006. URL `http://CRAN.R-project.org/doc/Rnews/`.

[77] R. Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. online, 2005. URL `http://www.R-project.org`.

[78] G Celeux, M Hurn, and C P Robert. Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970, 2000.

[79] S Kirkpatrick, C D Gelatt, and M P Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[80] E Marinari and G Parisi. Simulated tempering: A new monte carlo scheme. *EPL (Europhysics Letters)*, 19(6):451, 1992.

[81] C J Geyer and E A Thompson. Annealing markov chain monte carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90(431):909–920, 1995.

[82] C J Geyer. Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: Proc. 23rd Symp. Interface*, pages 156–163, 1991.

[83] D J Earl and M W Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910–3916, 2005.

[84] V I Manousiouthakis and M W Deem. Strict detailed balance is unnecessary in Monte Carlo simulation. *Journal of Chemical Physics*, 110 (6):2753–2756, 1999.

[85] D B Woodard, S C Schmidler, and M Huber. Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions. *The Annals of Applied Probability*, 19(2):617–640, 2009.

[86] Message Passing Interface Forum. MPI: A message-passing interface standard. Technical Report UT-CS-94-230, University of Tennessee, May 1994.

[87] W Gropp, S Huss-Lederman, A Lumsdaine, E Lusk, B Nitzberg, W Saphir, and M Snir. *MPI: The Complete Reference*. The MIT Press, 2nd edition, September 1998.

[88] Open MPI v 1.2.4, 2007. URL `http://www.open-mpi.org/software/ompi/v1.2/`.

[89] H Kozumi. Posterior analysis of latent competing risk models by parallel tempering. *Computational Statistics & Data Analysis*, 46(3): 441–458, 2004.

[90] Y Atchadé, G Roberts, and J Rosenthal. Towards optimal scaling of metropolis-coupled markov chain monte carlo. *Statistics and Computing*, pages 1–14, 2010.

[91] H Haario, E Saksman, and J Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[92] J S Rosenthal. Optimal proposal distributions and adaptive mcmc. In G Jones S Brooks, A Gelman and X L Meng, editors, *Handbook of MCMC*. Chapman & Hall/CRC Press, To appear.

[93] G O Roberts and J S Rosenthal. Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367, June 2009.

[94] W H Press, S A Teukolsky, W T Vetterling, and B P Flannery. *Numerical recipes in C (2nd ed.): the art of scientific computing*. Cambridge University Press, New York, NY, USA, 1992. ISBN 0-521-43108-5.

[95] M Frigo and S G Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[96] P D Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73, 1967.

[97] A Breitenbach. Against spectral leakage. *Measurement*, 25(2):135–142, 1999.

[98] S W Smith. *The Scientist & Engineer's Guide to Digital Signal Processing*. California Technical Pub.

[99] F J Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, Jan. 1978.

[100] R H Shumway and D S Stoffer. *Time Series Analysis and Its Applications*. Springer, 2000.

[101] P J Brockwell and R A Davis. *Introduction to Time Series and Forecasting*. Springer, 1996.

[102] M Calder and R A Davis. Introduction to whittle (1953): The analysis of multiple stationary time series. Technical Report 41, Colorado State University, 1996.

[103] N Christensen and R Meyer. Using markov chain monte carlo methods for estimating parameters with gravitational radiation data. *Phys. Rev. D*, 64(2):022001, May 2001.

[104] K A Arnaud and MLDC participants. Report on the first round of the Mock LISA Data Challenges. *Classical and Quantum Gravity*, 24 (19):S529, 2007.

[105] N J Cornish, L J Rubbo, and O Poujade. The LISA simulator. `www.physics.montana.edu/LISA`.

[106] B W Kernighan and D M Ritchie. *C Programming Language (2nd Edition)*. Prentice Hall, 2 edition, April 1988. ISBN 0131103628.

[107] B W Brown, J Lovato, K Russell, and J Venier. Library of c routines for random number generation. 1997. URL http://hpux.connect.org.uk/hppd/hpux/Maths/Misc/randlib-1.3/.

[108] ATLAS website, . URL https://n0.aei.uni-hannover.de/twiki/bin/view/ATLAS/UsefulItems.

[109] J R Gair and E K Porter. Emri parameter recovery for all round 2 training sources. online. http://www.tapir.caltech.edu.

[110] MLDC 4 Traing Data True Parameter Values. URL http://pandora.aei.mpg.de/~steffeng/MLDC4/FullDataset1/svn1159/challenge4.0-training-key-frequency.xml.

[111] *Condor-a hunter of idle workstations*, 1988.

[112] D Thain, T Tannenbaum, and M Livny. Distributed computing in practice: the Condor experience. *Concurrency and Computation: Practice and Experience*, 17(2-4):323–356, 2005.