

Eryn: a multipurpose sampler for Bayesian inference

Nikolaos Karnesis,¹★ Michael L. Katz,² Natalia Korsakova,³ Jonathan R. Gair² and Nikolaos Stergioulas¹

¹*Department of Physics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece*

²*Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany*

³*Astroparticule et Cosmologie, CNRS, Université Paris Cité, F-75013 Paris, France*

Accepted 2023 September 22. Received 2023 September 21; in original form 2023 March 14

ABSTRACT

In recent years, methods for Bayesian inference have been widely used in many different problems in physics where detection and characterization are necessary. Data analysis in gravitational-wave astronomy is a prime example of such a case. Bayesian inference has been very successful because this technique provides a representation of the parameters as a posterior probability distribution, with uncertainties informed by the precision of the experimental measurements. During the last couple of decades, many specific advances have been proposed and employed in order to solve a large variety of different problems. In this work, we present a Markov Chain Monte Carlo (MCMC) algorithm that integrates many of those concepts into a single MCMC package. For this purpose, we have built ERYN, a user-friendly and multipurpose toolbox for Bayesian inference, which can be utilized for solving parameter estimation and model selection problems, ranging from simple inference questions, to those with large-scale model variation requiring trans-dimensional MCMC methods, like the Laser Interferometer Space Antenna Global Fit problem. In this paper, we describe this sampler package and illustrate its capabilities on a variety of use cases.

Key words: gravitational waves – methods: data analysis – software: data analysis – software: development.

1 INTRODUCTION

In physics, and in science in general, one of the most encountered problems is the one of model calibration and comparison. We test our models of the physical world against the measured data, to estimate their parameters and to robustly determine the most suitable model that describes our observations. A crucial first step in this direction is to efficiently explore the posterior distribution of the parameters given the measured data. Markov Chain Monte Carlo (MCMC) algorithms have proven to be very successful in this regard (Metropolis et al. 1953; Hastings 1970; Gilks, Richardson & Spiegelhalter 1995; Hitchcock 2003), being one of the few methods which can efficiently perform Bayesian inference when the posterior is not analytically tractable and without solving exactly for the marginal likelihood. This is compared with, e.g. grid methods, which are often computationally unfeasible. This is especially true in the field of gravitational-wave (GW) astronomy, where MCMC methods have been extensively used in order to find physical parameters for signals buried in the data (see e.g. Shawhan & LIGO Scientific Collaboration 2003; Biwer et al. 2019; LIGO Collaboration et al. 2020; Abbott et al. 2021a, b; Ashton & Talbot 2021), as well as to hierarchically infer the properties of the underlying astrophysical populations (e.g. Ashton & Prix 2018; Thrane & Talbot 2019; Isi, Farr & Chatzioannou 2022). MCMC approaches can also compute the *marginal likelihood* or *evidence* (see Section 2), by using techniques such as *thermodynamic integration* (see Section 2.4). In a Bayesian framework, the evidence difference between two models

can be used to compute the Bayes factor, which is used to select between different models that could describe the observations.

Thermodynamic integration (or other approximations, see Section 2.5 or Cornish & Littenberg 2007 and Gelman, Roberts & Gilks 1996), is ideal for cases where the number of competing models is small. However, in situations where the number of potential models becomes too large, the task of iteratively and hierarchically computing the marginal likelihood can become computationally inefficient, or even practically unachievable. Such is the case for future signal-dominated GW observatories, such as the Laser Interferometer Space Antenna (LISA, Amaro-Seoane et al. 2017) or other proposed space-borne GW observatories (Kawamura et al. 2008; Luo et al. 2016; Ren et al. 2023). LISA will observe different types of GW sources, the most numerous of them being the ultra compact binaries (UCBs) within the Milky Way (Crowder & Cornish 2007; Amaro-Seoane et al. 2012, 2017; Littenberg et al. 2020, b; Karnesis et al. 2021; Zhang et al. 2021; Strub et al. 2022; Littenberg & Cornish 2023). Those are mostly comprised of a population of double white dwarfs, although a small fraction of neutron star (NS)–white dwarf or double NS binaries are expected (Nelemans, Yungelson & Portegies Zwart 2001; Breivik et al. 2020). In fact, LISA is going to detect GWs from the complete population of $\mathcal{O}(10^7)$ sources simultaneously, but only a small fraction of them is going to be individually resolvable ($\mathcal{O}(10^4)$). The large majority of signals will generate an anisotropic and non-stationary ‘confusion’ type of signal, which will dominate the LISA band between 0.05 and ~ 0.2 mHz.

In the aforementioned context, computing the marginal likelihood for such a large parameter space and for all possible numbers of events that could be in the population becomes computationally prohibitive. Instead, we can employ *dynamical trans-dimensional*

★ E-mail: karnesis@auth.gr

MCMC methods (Green 1995). This family of methods can be quite challenging to tune, but it has proven to yield satisfactory results, even for such demanding problems as the LISA data (Littenberg et al. 2020; Littenberg & Cornish 2023). There are also implementation challenges, which arise from technical aspects of the algorithm; one example being the *dimension matching* requirement when proposing moves between models with different dimensionality. In terms of algorithm efficiency, it is also crucial to choose proposal distributions that allow smooth transitions on a dynamical parameter space, a task which in many cases requires substantial effort. For these technical reasons, all the available software tools of this kind have been specifically developed for the particular problem they intend to solve.

In this work, we present ERYN, a reversible jump (RJ) MCMC algorithm, capable of *efficiently sampling dynamical parameter spaces, while remaining generic and usable by a large community*. We build upon various ideas from statistics, astronomy, etc., in order to develop an efficient statistical toolbox that can be applied to the majority of problems involving detection and characterization of signals. Our primary goal, however, is to utilize ERYN as a basic ingredient for a data analysis pipeline to perform the LISA *Global Fit* (Littenberg et al. 2020; Littenberg & Cornish 2023). The *Global Fit* is a data analysis strategy required to tackle the problem of multiple source detection, separation, and characterization in LISA data. For demonstration purposes in this work, we use ERYN to analyse a ‘reduced’ scenario of the LISA data in Section 4.

This paper is organized as follows: In Section 2, we begin with explaining the foundations of the MCMC algorithm, as well as some of the relevant methods that we have adopted for our implementation. In Section 3, we describe how the methods introduced in Section 2 are combined into the actual toolbox implemented in ERYN. In Section 3.3, we demonstrate the capabilities of ERYN through some toy examples, while in Section 4 we apply our machinery to more demanding applications in GW astronomy. In particular, we demonstrate ERYN by performing model selection on a simulated population of ultra compact Galactic binaries (UCBs) as measured by the future LISA observatory. Finally, in Section 5, we summarize our work and discuss future applications. We should state again here, that ERYN is available as open-source software at: <https://github.com/mikekatz04/Eryn>.

2 MCMC ALGORITHMS

Nowadays, MCMC methods are considered to be a cornerstone of Bayesian inference, being very effective in finding solutions to problems encountered across wide-ranging disciplines (e.g. Hobson et al. 2009; Sharma 2017; Hogg & Foreman-Mackey 2018; Biwer et al. 2019; Ashton & Talbot 2021). These include the sampling of the posterior densities of parameters of interest, the numerical marginalization over nuisance parameters, and providing a framework to compute the marginal posterior distributions (or evidences) that can be used for model selection. The Bayesian framework is based around Bayes’ Theorem:

$$p(\vec{\theta}|y, \mathcal{M}) = \frac{p(y|\vec{\theta}, \mathcal{M})p(\vec{\theta}|\mathcal{M})}{p(y|\mathcal{M})}, \quad (1)$$

where y is the measured data and \mathcal{M} our chosen model of analysis. The $p(\vec{\theta}|y)$ term is the posterior distribution of the parameter set $\vec{\theta}$, which is related to the likelihood function of the data $p(y|\vec{\theta}, \mathcal{M})$ and the prior densities of the parameters $p(\vec{\theta}|\mathcal{M})$. The evidence $p(y|\mathcal{M})$

is the marginal posterior over the parameter space $\vec{\theta} \in \Theta$:

$$\begin{aligned} \mathcal{Z} \equiv p(y|\mathcal{M}) &= \int_{\Theta} p(\vec{\theta}, y|\mathcal{M})d\vec{\theta} \\ &= \int_{\Theta} p(y|\vec{\theta}, \mathcal{M})p(\vec{\theta}|\mathcal{M})d\vec{\theta}. \end{aligned} \quad (2)$$

For parameter estimation purposes, the evidence acts as a normalization constant and can be ignored. However, it is really important if one wants to perform model selection over the measured data. We shall describe in detail how one can numerically approximate the integral of equation (2) in Section 2.5.

MCMC algorithms work by constructing a Markov Chain sequence, whose elements, $\vec{\theta}(t_i)$, for $i = 0, 1, \dots$, are (asymptotically) independent samples from the target distribution, $f(\vec{\theta})$. Under fairly general assumptions, the distribution of samples in the chain will converge to the target distribution provided the algorithm satisfies *detailed balance*:

$$f(\vec{\theta})p(\vec{\theta} \rightarrow \vec{\theta}') = f(\vec{\theta}')p(\vec{\theta}' \rightarrow \vec{\theta}). \quad (3)$$

Here $p(\vec{\theta} \rightarrow \vec{\theta}')$ is the probability that the Markov chain moves from point $\vec{\theta}$ to point $\vec{\theta}'$. The most widely used MCMC algorithm is Metropolis–Hastings (Metropolis et al. 1953; Hastings 1970), which is explained in algorithm box 1. The first step of the algorithm is to define an initial state, $\vec{\theta}(t_0)$. Then, at each subsequent step i , a new state is proposed, by randomly drawing from a given proposal distribution $q(\vec{\theta}'|\vec{\theta}(t_i))$. The newly proposed state is then accepted with a certain probability, given by equation (4). If the move is accepted we set $\vec{\theta}(t_{i+1}) = \vec{\theta}'$, otherwise we set $\vec{\theta}(t_{i+1}) = \vec{\theta}(t_i)$. Any reasonable choice of the proposal density will generate a Markov chain with the correct stationary distribution. However, a good choice of q is critical for its efficiency, i.e. achieving the convergence of the MCMC chains within a reasonable computational time. For the special case of a symmetric proposal distribution,¹ such as the widely used multivariate Gaussian distribution centred around $\vec{\theta}(t_i)$, the ratio of equation (4) in algorithm box 1 becomes simply the ratio of the target densities at the current $\vec{\theta}(t_n)$ and proposed $\vec{\theta}'$ points. For high-dimensional problems, the multivariate Gaussian proposal can be tuned during the burn-in period of sampling to improve efficiency (Gelman et al. 1996; Haario, Saksman & Tamminen 2001; Frenkel 2006; Andrieu & Thoms 2008; Roberts & Rosenthal 2009; Christensen & Meyer 2022), or even scaled according to the Cramer–Rao bound, estimated from the Information matrix (Vallisneri 2008).

Algorithm 1: The Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970). f is the target density to be sampled.

- 1: Set the initial state of chain: $\vec{\theta}(t_0) \equiv \vec{\theta}$.
- 2: At each subsequent step n , draw a new state $\vec{\theta}'$ from a *proposal distribution* $q(\vec{\theta}'|\vec{\theta}(t_{n-1}))$.
- 3: Take

$$\vec{\theta}(t_n) = \begin{cases} \vec{\theta}' & \text{with probability } \alpha(\vec{\theta}(t_{n-1}), \vec{\theta}') \\ \vec{\theta} & \text{with probability } 1 - \alpha(\vec{\theta}(t_{n-1}), \vec{\theta}') \end{cases},$$

where

$$\alpha(x, y) = 1 \wedge \left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)} \right\}. \quad (4)$$

- 4: Go to 2 and repeat until equilibrium is reached, and enough independent samples have been drawn from the target distribution.
-

Although the Metropolis–Hastings (MH) algorithm has been quite successful in tackling inference problems, there are practical imple-

¹That is, a proposal with the property $p(\vec{\theta} \rightarrow \vec{\theta}') = p(\vec{\theta}' \rightarrow \vec{\theta})$.

mentation issues to overcome. Improving acceptance rate is crucial for convergence, and sometimes improvements in the proposal distribution are not sufficient to efficiently sample the parameter space. To tackle these issues, various MCMC enhancements have been proposed. A prime example is the Hamiltonian Monte Carlo (HMC) algorithms that utilize local gradients in order to generate proposal points (Neal 2011; Betancourt & Girolami 2013). One variant of HMC is the No-U-Turn sampler which automates part of the required tuning of the HMC (Hoffman & Gelman 2014) sampler. Another alternative to the ‘standard’ MH is the Gibbs sampling algorithm, which is particularly useful if the conditional distributions of the parameters of the model are known (Muller 1991; Gilks & Wild 1992; Ritter & Tanner 1992). All of the above developments have been shown to be useful in various disciplines (Brooks 1998; Gelman et al. 2004; Kendall, Liang & Wang 2005; Sivia & Skilling 2006; Sorensen & Gianola 2007; Johannes & Polson 2010; Robert & Casella 2011; Baio 2012; Sharma 2017; Hogg & Foreman-Mackey 2018; Joseph 2019). Finally, there have recently been numerous proposals that aim to enhance sampling with machine learning techniques. At their core, many of these methods optimize the exploration of the likelihood surface, either by learning it directly (see e.g. Hermans, Begy & Louppe 2019) or by sampling it in a simpler latent space (e.g. Wu, Köhler & Noé 2020).

In this work, we introduce ERYN, which is built around the EMCEE package (Foreman-Mackey et al. 2013), enhanced with a variety of sampling mechanisms that allow us to perform inference on dynamical parameter spaces with minimal tuning. We expand on the most important features in the sections below.

2.1 Affine-invariant samplers

An affine transformation is one of the form $\vec{\theta} \rightarrow \vec{\zeta} = A\vec{\theta} + b$, where A and b are a constant matrix and vector, respectively. Under an affine transformation a probability density $p(\vec{\theta}|y)$ transforms to

$$p_{A,b}(\vec{\zeta}|y) = p(A^{-1}(\vec{\zeta} - b)|y)/\det(A). \quad (5)$$

Such transformations can help to transform difficult-to-sample distributions into easier-to-sample ones. A simple example is a multivariate normal distribution. If the dynamical range of the eigenvalues of the covariance matrix is very large, then sampling can be difficult, but any multivariate normal distribution can be transformed into a spherical distribution via an affine transformation.

Affine-invariant MCMC is a class of samplers that are designed to have equal sampling efficiency for all distributions that are related by an affine transformation (Goodman & Weare 2010; Foreman-Mackey et al. 2013). The sequence of samples in a Markov chain, $\{X(t)\}$, can be written deterministically as a function of a sequence of random variables, $\xi(t)$, which represent the random draws used to propose new points and evaluate the accept/reject decision. Specifically, we can always write

$$X(t+1) = R(X(t), \xi(t), p), \quad (6)$$

where p denotes the target density. An affine-invariant sampler has the property

$$R(AX(t) + b, \xi(t), p_{A,b}) = AR(X(t), \xi(t), p) + b, \quad (7)$$

i.e. the sequence of points visited when sampling an affine-transformed density are the affine transformations of the states visited when sampling the original density. If an affine transformation exists that maps the given target density to one which is more straightforward to sample from, an affine-invariant sampler should sample it as efficiently as it could the simpler distribution, so the convergence of

affine-invariant samplers is less affected by correlations between the parameters (Foreman-Mackey et al. 2013).

In practice, this goal is achieved by following an ensemble of points, called walkers, and basing proposed moves on the distribution of other points in the ensemble. In Foreman-Mackey et al. (2013), the primary update move is the so-called *stretch-move* proposal. Each walker at state $X_i(t)$ is updated by randomly selecting another walker j and proposing a new value $Y = X_j(t) + Z[X_i(t) - X_j(t)]$, where Z is a random variable drawn from the distribution (Goodman & Weare 2010)

$$g(z) \propto \begin{cases} \frac{1}{\sqrt{z}} z \in [\frac{1}{a}, a] \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

The parameter a can be tuned to improve convergence, but $a = 2$ works well in the majority of applications (Foreman-Mackey et al. 2013). The proposed point is accepted with probability

$$\alpha(X_i(t), Y) = 1 \wedge \left\{ z^{d-1} \frac{p(Y)}{p(X_i(t))} \right\}, \quad (9)$$

where p is the target density and d is the dimension of the parameter space. This acceptance probability is specific to the stretch proposal distribution given by equation (8). For other stretch proposals, the term z^{d-1} must be replaced by $z^{d-2}g(1/z)g(z)$. Following this scheme, detailed balance is maintained, and it can be proven that affine-invariant samplers converge faster to their target distribution (Foreman-Mackey et al. 2013). Below in Section 3, we discuss the extension of the stretch-move proposal to RJ MCMC methods. The benefits of running MCMC chains in parallel, combined with a proposal distribution that requires almost no tuning, have contributed to an increasing popularity of affine-invariant samplers. In particular, the EMCEE package (Foreman-Mackey et al. 2013), has been used widely in Astrophysics and Cosmology (e.g. McMillan 2017; De et al. 2018; Virtanen et al. 2020; Wong et al. 2020).

2.2 Delayed rejection

The delayed-rejection (DR) scheme of sampling was devised in order to improve two aspects of MCMC algorithms. First, it allows for improvements in the acceptance rate of the proposals, yielding ‘healthier’ parameter chains, with better mixing. Secondly, it is more robust against becoming trapped in local maxima of the posterior surface (Green & Mira 2001; Mira 2001; Haario et al. 2006; Trias, Vecchio & Veitch 2009). The strategy, as the name suggests, is, at each iteration, instead of immediately rejecting the newly proposed point based on Algorithm 1, we keep proposing new points while maintaining detailed balance by computing both the forward and backward transition probabilities. Suppose we are at a point $\vec{\theta}_0$ and use a proposal $q(\vec{\theta}_1|\vec{\theta}_0)$ to propose a new point $\vec{\theta}_1$. The usual acceptance probability, following the notation of equation (1), is

$$\alpha_1(\vec{\theta}_0, \vec{\theta}_1) = 1 \wedge \left\{ \frac{p(\vec{\theta}_1|y)q(\vec{\theta}_0|\vec{\theta}_1)}{p(\vec{\theta}_0|y)q(\vec{\theta}_1|\vec{\theta}_0)} \right\}, \quad (10)$$

as per equation (4). If $\vec{\theta}_1$ is rejected, then instead of going back to step 1 of Algorithm 1, we propose a new point, $\vec{\theta}_2$, drawn from a proposal distribution $q(\vec{\theta}_2|\vec{\theta}_1, \vec{\theta}_0)$. This proposal distribution may depend only on $\vec{\theta}_1$, but we write it more generally here to allow for the case that the proposal is adapted based on the sequence of steps that have been rejected. The acceptance probability for $\vec{\theta}_2$, $\alpha_2(\vec{\theta}_0, \vec{\theta}_1, \vec{\theta}_2)$, is

computed using

$$\alpha_2(\vec{\theta}_0, \vec{\theta}_1, \vec{\theta}_2) = 1 \wedge \left\{ \frac{p(\vec{\theta}_2|y)q(\vec{\theta}_1|\vec{\theta}_2)q(\vec{\theta}_0|\vec{\theta}_1, \vec{\theta}_2) \left[1 - \alpha_1(\vec{\theta}_2, \vec{\theta}_1) \right]}{p(\vec{\theta}_0|y)q(\vec{\theta}_1|\vec{\theta}_0)q(\vec{\theta}_2|\vec{\theta}_1, \vec{\theta}_0) \left[1 - \alpha_1(\vec{\theta}_0, \vec{\theta}_1) \right]} \right\}. \quad (11)$$

If $\vec{\theta}_2$ is rejected, further steps can be included and each step adds additional proposal and rejection-probability terms to the numerator and denominator of the acceptance probability. For example, the three step acceptance probability, $\alpha_3(\vec{\theta}_0, \vec{\theta}_1, \vec{\theta}_2, \vec{\theta}_3)$ is the minimum of one and

$$\frac{p(\vec{\theta}_3|y)q(\vec{\theta}_2|\vec{\theta}_3)q(\vec{\theta}_1|\vec{\theta}_2, \vec{\theta}_3)q(\vec{\theta}_0|\vec{\theta}_1, \vec{\theta}_2, \vec{\theta}_3)}{p(\vec{\theta}_0|y)q(\vec{\theta}_1|\vec{\theta}_0)q(\vec{\theta}_2|\vec{\theta}_1, \vec{\theta}_0)q(\vec{\theta}_3|\vec{\theta}_2, \vec{\theta}_1, \vec{\theta}_0)} \times \frac{\left[1 - \alpha_1(\vec{\theta}_3, \vec{\theta}_2) \right] \left[1 - \alpha_2(\vec{\theta}_3, \vec{\theta}_2, \vec{\theta}_1) \right]}{\left[1 - \alpha_1(\vec{\theta}_0, \vec{\theta}_1) \right] \left[1 - \alpha_2(\vec{\theta}_0, \vec{\theta}_1, \vec{\theta}_2) \right]}. \quad (12)$$

The proposal q can be different at each step, as long as the relevant proposal density is used in equation (11). For example, in Trias et al. (2009), the proposal is built upon a Gaussian mixture model that tries further points in the parameter space with the aim of efficiently exploring multiple modes of the posterior distribution. As the number of steps in the DR scheme becomes arbitrarily large, the acceptance probability slowly approaches zero. This algorithm is also limited in practice by high computational requirements, since at every DR step we need to evaluate a new likelihood and compute the backwards probability (the $\alpha_1(\vec{\theta}_2, \vec{\theta}_1)$ from equation 11). Nevertheless, the DR scheme offers many advantages, and despite the computational cost, it is very useful when the posterior surface exhibits high dimensionality, and when acceleration techniques are available. These, e.g. might include the use of graphical processing units (GPUs), and/or heterodyned likelihoods (Cornish 2021). In our implementation here, we follow closely the one in Trias et al. (2009), for improving the acceptance rate of the *between-model step* of the RJ algorithm (see Section 2.6). As already mentioned, the RJ MCMC allows for sampling dynamical parameter spaces. In the special case of nested models, such as the case of searching multiple signals in the LISA data, proposing the ‘birth’ of a signal out of a very wide prior can be very inefficient. A DR scheme alleviates this problem, by effectively performing a small search around the first set of rejections, increasing the chances of finding a good signal candidate, and thus improving the mixing of the chains.

2.3 Multiple try metropolis

The multiple try metropolis (MTM, Liu, Liang & Wong 2000; Martino, Del Olmo & Read 2012; Bédard, Douc & Moulines 2012; Martino 2018) is a subclass of the implementation of the MH algorithm, which is based on the idea of generating a number of proposals for each individual current state, and then selecting one of them based on their importance weight. In proposing a move from $\vec{\theta}_{t-1}$, a set of N possible new points, $\{y_k\}$, are drawn from a proposal distribution $q(y|\vec{\theta}_{t-1})$ and are assigned weights $w_i = w(y_i|\vec{\theta}_{t-1})$ using a weight function $w(\cdot|\vec{\theta}_{t-1})$. One of these proposed points, y_j , is selected with probability given by the normalized weight

$$\bar{w}_i = \frac{w_i}{\sum_{k=1}^N w_k}. \quad (13)$$

To compute the acceptance probability, we need to draw $N - 1$ points, $\{x_i, i = 1, \dots, N - 1\}$, for the reverse move from the proposal $q(x|y_j)$,

and assign weights $w(x|y_j)$. We then set $\vec{\theta}_t = y_j$ with probability

$$\alpha(\vec{\theta}_{t-1}, y_j) = 1 \wedge \left\{ \frac{w(y_j|\vec{\theta}_{t-1}) + \sum_{k=1, k \neq j}^N w(y_k|\vec{\theta}_{t-1})}{w(\vec{\theta}_{t-1}|y_j) + \sum_{k=1}^{N-1} w(x_k|y_j)} \right\}, \quad (14)$$

and set $\vec{\theta}_t = \vec{\theta}_{t-1}$ otherwise (Martino 2018). This procedure will satisfy detailed balance if the weight function is chosen such that

$$p(\vec{\theta}_0|y)q(\vec{\theta}_1|\vec{\theta}_0)w(\vec{\theta}_1|\vec{\theta}_0) = p(\vec{\theta}_1|y)q(\vec{\theta}_0|\vec{\theta}_1)w(\vec{\theta}_0|\vec{\theta}_1). \quad (15)$$

This will be satisfied by a weight function of the form

$$w(\vec{\theta}_t|\vec{\theta}_{t-1}) = p(\vec{\theta}_t|y)q(\vec{\theta}_{t-1}|\vec{\theta}_t)\xi(\vec{\theta}_{t-1}, \vec{\theta}_t), \quad (16)$$

where $\xi(\vec{\theta}_{t-1}, \vec{\theta}_t)$ is any symmetric function, i.e. $\xi(\vec{\theta}_{t-1}, \vec{\theta}_t) = \xi(\vec{\theta}_t, \vec{\theta}_{t-1})$, $\forall \vec{\theta}_t, \vec{\theta}_{t-1} \in \mathcal{D} \subseteq \mathbb{R}^d$, with d being the dimensionality of the problem at hand. The detailed balance condition can also be satisfied by a weight function of the form

$$w(\vec{\theta}_t|\vec{\theta}_{t-1}) = \frac{p(\vec{\theta}_t|y)}{q(\vec{\theta}_t|\vec{\theta}_{t-1})}. \quad (17)$$

Making this choice and additionally using a proposal function that is independent of the current point, $q(\vec{\theta}_t|\vec{\theta}_{t-1}) = q(\vec{\theta}_t)$ only, we obtain the *independent MTM* algorithm (Martino 2018). When using the independent MTM algorithm detailed balance is maintained when the same set of points is used for the reverse proposal as for the forward proposal, which saves the evaluation of $N - 1$ posterior densities.

The base MTM is currently implemented in ERYN with options for the independent MTM algorithm and symmetric proposals. For a symmetric proposal distribution, $q(\vec{\theta}_{t-1}|\vec{\theta}_t) = q(\vec{\theta}_t|\vec{\theta}_{t-1})$, equation (15) can be satisfied using the weight function $w(\vec{\theta}_1|\vec{\theta}_0) = p(\vec{\theta}_1|y)$. In this case, we still need to draw separate samples for the reverse step (unlike in the independent MTM case).

Generating a large number of candidate points yields certain advantages. As expected, the first advantage is the fact that there is usually very good coverage of the parameter space. The second is that the implementation of the MTM usually results in chain states with very low correlation between them. Nevertheless, as for DR, this algorithm requires increased computational resources, since multiple likelihoods/posterior densities have to be evaluated at each iteration of the chain. This cost can be offset in cases where the computations can be parallelized, e.g. using either Central Processing Unit (CPU) or GPU acceleration.

2.4 Adaptive parallel tempering

The concept of parallel tempering was introduced in order to efficiently sample surfaces with high multimodality (Swendsen & Wang 1986; Hukushima & Nemoto 1996; Voudsen, Farr & Mandel 2016). The idea is based on a transformation of the posterior density to a density with a different temperature, T , defined by

$$p_T(\vec{\theta}|y) \propto p(y|\vec{\theta})^{1/T} p(\vec{\theta}). \quad (18)$$

For $T = 1$ this is the target posterior density. In the limit $T \rightarrow \infty$, it is the prior density. Intermediate temperatures ‘smooth out’ the posterior by reducing the contrast between areas of high and low likelihood.

In parallel tempering, a set of Markov chains is constructed in parallel, each one sampling the transformed posterior for a different temperature T . These chains periodically exchange information. The idea is that the hottest chains explore the parameter space more widely, and information about areas of high likelihood that they encounter propagates to the colder chains. Information is exchanged

by proposing swaps of the states between the different chains. If two chains are sampling from target densities $p_1(\vec{\theta})$ and $p_2(\vec{\theta})$, respectively, then the transition probability for chain 1 in the swap is $p_1(\vec{\theta}_0 \rightarrow \vec{\theta}_1) = p_2(\vec{\theta}_1)\alpha(\vec{\theta}_0, \vec{\theta}_1)$. Detailed balance is thus maintained by accepting the swap with probability

$$\alpha(\vec{\theta}_0, \vec{\theta}_1) = 1 \wedge \left\{ \frac{p_1(\vec{\theta}_1)p_2(\vec{\theta}_0)}{p_1(\vec{\theta}_0)p_2(\vec{\theta}_1)} \right\}, \quad (19)$$

which for the specific case of swapping between two tempered chains i and j when doing parallel tempering is

$$\alpha_{i,j} = 1 \wedge \left\{ \left(\frac{p(y|\vec{\theta}_i)}{p(y|\vec{\theta}_j)} \right)^{\beta_j - \beta_i} \right\}, \quad (20)$$

with $\beta_i = 1/T_i$ being the inverse temperature, and $\vec{\theta}_i$ the given parameter state for the i th chain.

The temperature ladder T_i should be chosen in order to maximize the information flow between chains of different temperatures, so as to encourage the efficient exploration of the complete parameter space. Typically, this ladder can be static or dynamically adjusted during the sampling procedure. In ERYN, we have adopted the procedure of Vousden et al. (2016), which adapts the temperature ladder based on the swap acceptance rate calculated directly from the chains. Ideally, one should aim for equal acceptance ratio between every pair of neighbouring tempered chains, thus tuning their log-temperature-difference $S_i \equiv \log(T_i - T_{i-1})$, according to the swap acceptance rate from equation (20):

$$\frac{dS_i}{dt} = \kappa(t) [\alpha_{i,i-1}(t) - \alpha_{i+1,i}(t)], \quad (21)$$

where $\kappa(t)$ tunes the time-scale of the evolution of the temperatures. The function $\kappa(t)$ can be chosen depending on the desired behaviour of the procedure. In Vousden et al. (2016), a hyperbolic dependence on the t state is chosen, in order to suppress large dynamic adjustments on long time-scales. This set-up is the default option in ERYN, but it can be customized. This process is more straightforward for ensemble samplers, where multiple walkers are used, simply because one can get an estimate of the acceptance rate directly from the particular state of the walkers at any given time-step t . Otherwise, the acceptance rate is computed after iterating for a predefined number of steps, which can be chosen by the user for the given problem at hand. It can be proven (Vousden et al. 2016), that the temperature ladder will converge to a particular stable configuration. One should only use this scheme for the initial burn-in stage of sampling, and then continue with a stationary ladder for the rest of the analysis.

2.5 Marginal posterior calculation for model selection

One of the most frequently encountered problems in physics, and in science in general, is that of model or variable selection, i.e. identifying the model best supported by the observed data. Working in a Bayesian framework, comparison between different hypotheses may be done by computing their evidences or marginal posteriors (Gelman et al. 2004) and evaluating the Bayes factor:

$$B_{12} = \frac{p(\vec{y}|\mathcal{M}_1)p(\mathcal{M}_1)}{p(\vec{y}|\mathcal{M}_2)p(\mathcal{M}_2)}, \quad (22)$$

where the term $p(\mathcal{M}_i)$, is the prior probability assigned to the model \mathcal{M}_i .

The marginal posterior density, or evidence, is given by the integral of equation (2) and is in general quite challenging to compute. For some high signal-to-noise ratio (SNR) cases it can

be approximated if the covariance matrix Σ of the parameters for all candidate models \mathcal{M} are known. This approach is called the *Laplace approximation* (Kass & Raftery 1995; Gelman et al. 2004). However, this is only an approximation, and it sometimes fails for models with weak support from the data (Cornish & Littenberg 2007) (in particular when the posterior cannot be approximated by a multivariate Gaussian at $\vec{\theta}_{\text{MAP}}$).

When using parallel tempering (Section 2.4), it is possible to compute the evidence by a procedure known as *thermodynamic integration* (Lartillot & Philippe 2006). We define a continuous distribution of evidences based on the target distribution for a chain with inverse temperature $\beta = 1/T$ via

$$\mathcal{Z}_{i,\beta} = \int p(y|\vec{\theta}, \mathcal{M}_i)^\beta p(\vec{\theta}) d\vec{\theta}. \quad (23)$$

For $\beta = 0$, the chain is sampling the prior and therefore $\log \mathcal{Z}_{i,0} = 0$. For $\beta = 1$, we are sampling the target distribution and $\log \mathcal{Z}_{i,1} = \log \mathcal{Z}_i$. Additionally, we have

$$\begin{aligned} \frac{d \log \mathcal{Z}_\beta}{d\beta} &= \int \log[p(y|\vec{\theta}, \mathcal{M}_i)] p(y|\vec{\theta}, \mathcal{M}_i)^\beta p(\vec{\theta}) d\vec{\theta} \\ &\equiv \mathbb{E}_\beta[\log p(y|\vec{\theta}, \mathcal{M}_i)]. \end{aligned} \quad (24)$$

From this, we deduce

$$\log \mathcal{Z}_i = \int_0^1 \mathbb{E}_\beta [\log p(y|\vec{\theta}, \mathcal{M}_i)] d\beta. \quad (25)$$

The expectation value is over the distribution being sampled by the chain at temperature β and so can be computed by averaging over the posterior samples (Goggans & Chi 2004; Lartillot & Philippe 2006; Vousden et al. 2016). The integral can then be evaluated using standard methods, e.g. the trapezium rule, using the full ladder of temperatures. This approach generates reliable evidences, with accuracy limited only by the number of temperatures being sampled, and the efficiency of the sampling of the parameter space Θ by the chains. Since its first introduction, there have been many applications of this approach, and in particular, there is extensive usage in GW astronomy (Littenberg & Cornish 2009, 2010; Vousden et al. 2016; Maturana-Russel et al. 2019; Katz et al. 2022).

The thermodynamic integral in equation (25) can be thought of as computing the evidence as a sum of differences between the evidences at different temperatures. An alternative approach, called the *stepping-stone algorithm* (Xie et al. 2010), writes the evidence as a product of the ratios of evidences at different temperatures:

$$\mathcal{Z}_i \equiv \frac{\mathcal{Z}_{i,1}}{\mathcal{Z}_{i,0}} = \prod_{k=1}^{N_T-1} \frac{\mathcal{Z}_{i,\beta_{k+1}}}{\mathcal{Z}_{i,\beta_k}} \quad (26)$$

where β_k denotes the inverse temperature of chain k , N_T is the number of different temperatures being sampled, and we assume $\beta_1 = 0$ and $\beta_{N_T} = 1$. Each evidence ratio can be written as a posterior integral:

$$\begin{aligned} \frac{\mathcal{Z}_{i,\beta_{k+1}}}{\mathcal{Z}_{i,\beta_k}} &= \int \frac{p(y|\vec{\theta}, \mathcal{M}_i)^{\beta_{k+1}-\beta_k} p(y|\vec{\theta}, \mathcal{M}_i)^{\beta_k} p(\vec{\theta})}{\mathcal{Z}_{\beta_k}} d\vec{\theta} \\ &= \mathbb{E}_\beta \left[p(y|\vec{\theta}, \mathcal{M}_i)^{\beta_{k+1}-\beta_k} \right] \\ &\approx \frac{1}{n} \sum_{i=1}^n p(y|\vec{\theta}_k^i, \mathcal{M}_i)^{\beta_{k+1}-\beta_k}, \end{aligned} \quad (27)$$

where n is the number of posterior samples in each chain, and $\vec{\theta}_k^i$ denotes the i th sample at temperature β_k . This leads to the final

expression:

$$\log \mathcal{Z} = \sum_{k=1}^{N_T-1} \log \sum_{i=1}^n p(y|\vec{\theta}_k^i, \mathcal{M})^{\beta_{k+1}-\beta_k} - (N_T - 1) \log n. \quad (28)$$

In challenging situations, e.g. where the number of tempered chains is relatively small, the stepping-stone algorithm has been shown to produce more accurate estimates of the marginal likelihood than methods that use thermodynamic integration (Xie et al. 2010). This was also demonstrated with practical examples drawn from GW astronomy in Maturana-Russel et al. (2019).

2.6 Reversible jump

Another approach to the model selection problem is to follow a RJ MCMC strategy, which can dynamically estimate the most probable hypotheses given the data (Green 1995). The RJ MCMC is a generalization of the MH algorithm that allows trans-dimensional proposals. Thus, the model order is considered a free parameter which is fitted together with the parameters of the individual models. The most widespread variation of the algorithm uses a two-stage procedure. The first stage or *in-model step*, uses the standard MH algorithm to update all the parameters $\vec{\theta}_k$ for the given model k . The second stage or *between-model step* proposes to update the model state k to a new model state l . Parameters $\vec{\theta}_l$ for the new model are also proposed. The newly proposed state l , is accepted with a probability defined by Godsill (2001):

$$\alpha' = 1 \wedge \left\{ \frac{p(\vec{\theta}_l|l, y)g'(u_l)}{p(\vec{\theta}_k|k, y)g(u_k)} |\mathbf{J}| \right\}, \quad (29)$$

where

$$p(\vec{\theta}_k|k, y) = p(y|\vec{\theta}_k, k)p(\vec{\theta}_k, k)p(k) \quad (30)$$

with $p(y|\vec{\theta}_k, k)$ the likelihood for model k , $p(\vec{\theta}_k|k)$ the prior on the parameters $\vec{\theta}_k$ in model k , and $p(k)$ the prior for the model state k . The term $g'(u_l)|\mathbf{J}|/g(u_k)$ arises because of the need for *dimension matching* between the different model states. In general, we can define a move between model states in terms of a deterministic invertible mapping, $\vec{\theta}_k = q(\vec{\theta}_l, u_l)$ with inverse $\vec{\theta}_l = q'(\vec{\theta}_k, u_k)$, that is a function of the parameters and two sets of random variables, u_k and u_l , drawn from distributions $g(u_k)$ and $g'(u_l)$ (Green 1995; Godsill 2001). The term $|\mathbf{J}|$ is the Jacobian defined by this invertible mapping:

$$|\mathbf{J}| = \left| \frac{\partial(\vec{\theta}_l, u_l)}{\partial(\vec{\theta}_k, u_k)} \right| \quad (31)$$

and the term $g'(u_l)/g(u_k)$ plays the role of the proposal ratio in the standard MH acceptance probability. Dimension mapping means that $\dim(\vec{\theta}_l) + \dim(u_l) = \dim(\vec{\theta}_k) + \dim(u_k)$.

Using RJ MCMC introduces additional computational cost at each MCMC iteration, as well as technical challenges in implementation. Luckily, implementation can be easier when sampling *nested models*. This refers to problems where more complicated models contain their simpler counterparts. Examples of such cases are fitting polynomial models, which differ only in the highest order to be determined, or detection problems where multiple similar signals are potentially present the data. In such cases, the between-model step can always be formulated such that the Jacobian of equation (31) becomes unity, and equation (29) simplifies to the ratio of posteriors accounting for any differences in prior and proposal volumes (Dellaportas, Forster & Ntzoufras 2002; Lopes & West 2004; Littenberg et al. 2020).

After running RJ MCMC, the Bayes factor can be approximated by the ratio of the number of iterations spent within each model:

$$B_{12} = \frac{\# \text{ of iterations in model } \mathcal{M}_1}{\# \text{ of iterations in model } \mathcal{M}_2}. \quad (32)$$

This algorithm has proven to be robust for evaluating high-dimensional competing models, and has been quite successful in tackling data analysis problems in GW astronomy (Cornish & Littenberg 2007; Karnesis et al. 2014; Littenberg et al. 2020) as well as areas spanning physics and signal processing (e.g. Marrs 1997; Khan, Balch & Dellaert 2005; Yu et al. 2021). However, designing an efficient RJ MCMC algorithm can be quite challenging. The first challenge is to choose suitable proposal distributions, which can greatly affect the convergence of the algorithm. In situations where the models are nested, it is both tempting and convenient to take the proposal to be the same as the prior distribution of the parameters. As an illustrative example, we refer to Section 3.3.1, which describes a toy problem of searching for Gaussian pulses in noisy data. There, the parameters of the individual pulses are the amplitude and location of the pulse described by their (x, y) coordinates. In order to search for those signals, the prior on their location must be wide enough to include the complete data set (see Fig. 1a). A birth proposal based on the prior would inevitably be quite inefficient, simply because the chance of proposing a good source candidate is small, especially if the proposal distribution is flat across (x, y) . We treat the above problem as a motivation to adopt efficient proposals with minimal tuning in ERYN, which we further discuss in Section 3. The second major challenge, which of course depends on the given problem at hand, arises from the samplers' capability to explore a multimodal dynamical parameter space. We discuss our strategy to overcome that challenge in Section 3.

3 ERYN: GATHERING ALL THE PIECES TOGETHER

All the different algorithms described in previous sections can be extremely useful in tackling different kinds of problems that require sampling. In GW astronomy, we encounter such problems far too often, where dynamical parameter spaces require vast computational resources in order to be explored efficiently. Motivated by those problems, we have implemented a new toolbox that combines all these techniques to enhance the capabilities of an MCMC sampler. We have named this package ERYN,² borrowing the name from the Tolkien mythos (Tolkien Gateway 2023). The analogy has its basis in the idea of a forest: Within a forest you have trees which correspond to different walkers, also known as *Ents*, in an ensemble MCMC sampler. On each tree, there are branches that represent the various types of models used to fit the data. For example, in the case of GW global fitting for LISA, we can imagine using the Galactic binaries as one branch and massive black hole (BH) binaries as another branch. Each branch has leaves, which represent the individual instances of each model. In the LISA example, leaves would represent the individual Galactic binaries or massive BH binaries. And finally, to zoom out, when adding tempering capabilities, we can think of groups of walkers in each temperature taking the form of many forests (of walkers) located within different temperature climates.

²Eryn Vorn (Sindarin for Blackwood) was a wooded cape in Eriador, and a region of dark pine trees. Located in western Minhiriath, Eryn Vorn (likely named so by the Númenóreans) was originally part of the vast ancient treescape that covered most of north-western middle-Earth.

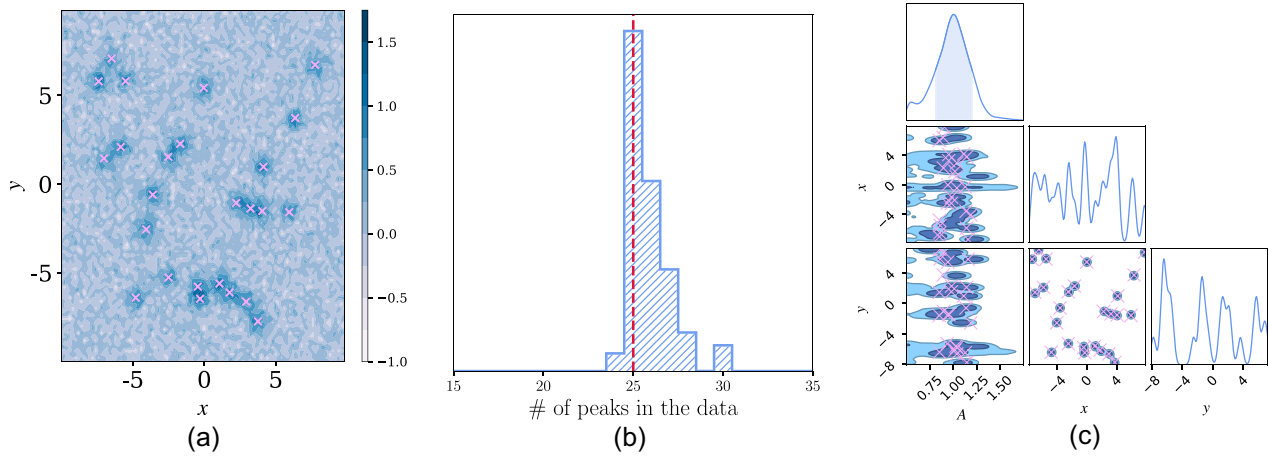


Figure 1. Searching for two-dimensional (2D) Gaussian pulses in the presence of Gaussian noise. (a) The simulated data, which consists of injections of 25 pulses in Gaussian noise with $\sigma_n = 0.2$. (b) The distribution of the model order, obtained by exploring the dynamical parameter space with ERYN. The true value is marked with a dashed red line. For this toy investigation, the correct number of simulated components is recovered. (c) The 2D posterior densities for the parameters of the k Gaussian peaks (see text for details).

We adopt the architecture of ensemble samplers, and in particular the one of EMCEE (Foreman-Mackey et al. 2013). Having multiple walkers running in parallel is ideal for efficiently sampling the parameter spaces using techniques such as parallel tempering, as described in Section 2.4 (also see Section 2.1). In this setting, we evolve n_w walkers per temperature T_i , where each walker follows a RJ MCMC (see Section 2.6), mapping a parameter space of altering dimensionality. In practice, walkers in higher temperatures sample the dynamic parameter space with fewer model components as the penalty from higher prior volume is not compensated by the smoother annealed likelihood. In other words higher temperatures have a sharper Occam’s razor: The data can be explained with models that are simpler, or lower dimensional. The highest temperature chain samples the prior on the model space (provided that $T_{\max} = \infty$). More details will be given in Section 3.3.

As already mentioned in Section 2.6, RJ algorithms are extremely challenging to tune, even for simpler classes of problems. One of the major challenges is the low acceptance rate for the between-model proposal, i.e. when we propose a new state where the parameter dimensionality differs. In cases of signal search and detection (which is a nested model situation), it is convenient to set the proposal corresponding to a ‘birth’ move to be the same as the prior distribution. In order to accommodate all possibilities for the signals present, the prior densities are usually quite wide, and thus accepting a new higher dimensional state becomes quite improbable. For that reason, within ERYN, we have implemented a DR scheme with the aim of improving this acceptance ratio. When proposing $\bar{\theta}_l$ for a higher dimensional model l , we do not reject immediately, but rather make new DR proposals around the first rejected point $\bar{\theta}_l$, using the given *in-model step* update proposal. This, in principle, allows the sampler to explore around $\bar{\theta}_l$ before rejecting the new state (Trias et al. 2009), which in turn improves the between-model step acceptance rate and produces healthier MCMC chains.

The DR scheme, as described in Section 2.2, requires a serial computation of the DR acceptance ratio for walkers where the newly proposed state has been rejected. This scheme of calculating costly likelihoods sequentially in a loop during the between-model step, can lead to a computational bottleneck of the MCMC process. This is especially true for the LISA Global Fit problem, where multiple

binary waveform signals are present in the data stream. Then, the computational time for each RJ MCMC iteration is significantly increased, since the progress will be halted until all walkers have gone through their respective DR process, which requires evaluation of new waveforms at each step. For the reasons summarized here, we have not used the DR scheme for our analysis in Section 4, and have resorted to the multiple try scheme. However, the DR scheme, as explained in Section 2.2, has been implemented in the ERYN package.

The multiple try scheme was essentially implemented in order to facilitate use of a parallelized likelihood framework. Parallelization is naturally compatible with multiple try MCMC as multiple proposals are made for each individual walker, which allows for the parallelized evaluation of proposal distributions, likelihood functions, and acceptance probabilities. Under these parallelized settings, one proposal can act as many when compared with the usual serial evaluation of proposals, allowing for better chain mixing in situations where proposals are infrequently accepted. That being said, it is still important to choose a good proposal distribution, for both the in-model and between-model RJ MCMC steps, which we discuss further ahead.

3.1 Choosing efficient proposal distributions

In the previous sections, we briefly discussed some of the challenges in choosing efficient proposal distributions for both the in-model and between-model steps of the RJ algorithm. For the in-model case, the challenge arises from the fact that it is sometimes impractical, or even unfeasible, to define a well-tuned proposal for each of the possible models that could represent the data. Using again the example of LISA data, one would need to tediously design an effective proposal distribution for the thousands of overlapping binary signals in the data. On the other hand, for the case of the between-model step, choosing proposals from the prior distribution, especially if it is highly uninformative, can be very inefficient for RJ sampling. For ERYN, in order to tackle those issues, we have implemented the *group proposals* explained below for addressing the within-model proposals in RJ, as well as a scheme to design an efficient proposal for birth moves during the between-model step, which is based on normalizing flows.

3.1.1 Group proposals

In Section 2.1, the stretch-move proposal was introduced and discussed. One of the obvious advantages of such a scheme of proposing new MCMC samples is that it requires minimal tuning (Foreman-Mackey et al. 2013). However, it does not extend well in its simplest form to the generalized RJ MCMC. The stretch proposal is based on the idea that the ensemble of points (X_j) is sitting on the same posterior mode as the current point (X_i). In a nested model situation where both the model count and the individual model parameters change, each point may lie on a different posterior mode representing a different set of leaves (sources) in the data. This can be alleviated by applying the stretch move to individual leaves within each branch of each walker, but there is still an issue of identifying leaves in different walkers that lie on the same posterior mode. The stretch proposal will technically still work when mixing leaves in different posterior modes, but the acceptance fraction will be negatively affected. However, within the stretch proposal formalism, the choice of X_j is customizable. The key to maintaining detailed balance is that X_j cannot depend on X_i , and X_j cannot be updated in the same iterative step as X_i (Goodman & Weare 2010).

We leverage this property to design a new type of stretch move that can handle RJ set-ups while maintaining a small number of tuning parameters. We call this proposal the ‘group’ proposal. The mathematics that governs the group proposal is equivalent to that of the original stretch proposal. The key difference is that the group proposal chooses X_j (see Section 2.1) from a stationary group that is fixed for many proposed updates. This is in contrast to the original stretch proposal that always uses the current set of points in the ensemble sampler to draw X_j . The stationary group is updated after a large number of sampler iterations and we make sure that detailed balance is maintained during the update. We update every leaf within every branch of every walker at each iteration and repeat many iterations between updates of the stationary group.

The appropriate stationary group varies from problem to problem. The goal is to set a group that resembles as best as possible the posterior modes of the current leaves and then draw from it strategically so that the drawn point is likely (but not guaranteed) to lie on the same mode as the leaf that is currently being updated, X_i . In the example of the LISA Galactic binaries analysis, we set our stationary group to the full set of leaves (binaries) across all walkers at a specific temperature of the sampler at the end of a given iteration. Then, at proposal time, we efficiently locate the $\sim n_w$ points in the stationary group that are closest to X_i from based on their initial frequency parameter. We then draw X_j from this group. The hope is that some percentage of the n_w drawn points will lie on the posterior mode on which X_i sits. The exact percentage will vary depending on how close the posterior modes are to each other and how many model instances exist in the sampler that include this specific mode. For low-SNR binaries, e.g. a source may exist in some walkers and not others, making it harder to map its posterior mode with the current group of stationary points.

The performance of group proposals is highly situation- and/or model-dependent. With individual source posterior modes that are well separated and easy to define in terms of separation, the performance will approach the performance of the base stretch proposal in non-RJ MCMC because the stationary group will well represent the specific posterior mode on which X_i is located. As the parameter space becomes more crowded and/or separation (distance) metrics become harder to define, the performance of group proposals will worsen.

3.1.2 Learning from the data

The second improvement concerns the between-model step of the RJ MCMC. As mentioned earlier, for the case of nested models, it is often convenient to draw ‘birth’ candidates directly from the prior distribution of parameters of the given model. This practice can be quite ineffective in terms of acceptance rate. As an example we can again use the LISA data set case. The UCBs are distributed within the Galactic disc, congregated mostly around its centre (Amaro-Seoane et al. 2023), therefore, adopting a proposal based on an uninformative uniform prior across the sky, would waste computational resources exploring a part of the parameter space with low probability mass. A proposed solution is to use an informative prior derived from the spatial distribution of binaries in the Galaxy (Littenberg et al. 2020). In our work here, we have chosen an alternative data-driven route, based on the actual residual data after a burn-in period of the RJ MCMC, which we describe below. After a sufficient number of RJ MCMC iterations, we can extract a subset of sources from nested models which are constantly present in almost all walkers of our cold chain. In other words, we can find and subtract the brightest sources from the data, and then allow for another burn-in period on the resulting residuals. This allows the sampler to explore the remaining parameter space more easily, thus providing a good initial estimate for the weak signals possibly buried in the noisy residual data. We can then use those samples to construct a proposal density which will help us search for good candidates for those weak signals, without excluding the rest of the parameter space. This can be accomplished by fitting the distribution to the residual data described above. The most efficient way to fit to the generic distribution, is to use an invertible transformation, such as a normalizing flow (e.g. Dinh, Sohl-Dickstein & Bengio 2016; Durkan et al. 2019). The method works in the following way: We sample from the base distribution (which is usually chosen to be normal $\mathcal{N}(\mathbf{z}; 0, 1)$) and transform samples to the desired distribution $p(\boldsymbol{\theta})$ by applying the change of variable equation:

$$p(\boldsymbol{\theta}) = N(f^{-1}(\boldsymbol{\theta}))|J_{f^{-1}}(\boldsymbol{\theta})|. \quad (33)$$

Here, function $f(\mathbf{z})$ is a bijection which we fit by optimizing the Kullback–Leibler divergence, $D_{\text{KL}}[p(f(\boldsymbol{\theta}))|\mathcal{N}(\mathbf{z}; 0, 1)]$, between a normal distribution and the inverse transform of the distribution that we want to estimate. After the fit has converged, we can draw samples from the normal distribution and transform these to samples from the distribution fitted to all residuals and use it as a proposal. We will cover this method in more detail in a separate paper which is being prepared.

3.1.3 Convergence diagnostics

Many standard approaches for assessing convergence of MCMC chains can be applied to the RJ MCMC chains generated by ERYN. Trace plots, both for individual parameters within models, and the indices labelling different models, will show that the chains are mixing well and that the initial burn-in phase has finished. Repeating runs with different numbers of walkers, different starting positions, different numbers of parallel chains, or different choices of temperatures and comparing results can be used to assess convergence. Similarly, posteriors produced from randomly selected subsets of chain points can be compared (using e.g. statistical tools such as the Jensen–Shannon divergence, Menéndez et al. 1997).

Additionally, ERYN computes and outputs the potential scale reduction factor (PSRF) for each tempered chain, which provides

a quantitative diagnostic of the convergence of the result (Gelman & Rubin 1992; Brooks & Gelman 1998). For a single parameter, the PSRF index is given by

$$\widehat{R} = \sqrt{\frac{(d+3)\widehat{V}}{(d+3)W}}, \quad (34)$$

where W the mean of the empirical variance within each chain, and \widehat{V} is the estimated variance of the all chains, assuming that the target distribution is Gaussian. The degrees of freedom are estimated by the method of moments as $d \approx 2\widehat{V}/\widehat{\text{var}}(\widehat{V})$ (Brooks & Gelman 1998).

The PSRF index takes values $\widehat{R} \geq 1$, with values close to unity indicating converged MCMC chains. In the end, the quantity that really matters is the PSRF value for the cold chain, but converged chains across for all temperatures are good indicators for healthy mixing and efficient temperature swaps.

3.2 Implementation

In this section, we discuss the main implementation details of ERYN. We refer the interested reader, or user, to the ERYN documentation for more exhaustive information and examples (Documentation).

The goal of ERYN is to produce a sampler that can handle all (or most) cases of MCMC sampling ranging from basic, non-tempered, single-model type, single-model instance posterior estimation to the full RJ MCMC with tempering, multiple model types, and adjustable model counts, as well as everywhere in between. In the basic case, ERYN aims to be a close replica of EMCEE trying to maintain as much simplicity as possible. At the complicated end of the spectrum, ERYN attempts to provide a common interface and underlying infrastructure for the variety of problems that may arise, allowing the user to maintain usage of the majority of the code from project to project, focusing on changing only the specific parts of the code that are difficult to implement or require special treatment for each specific problem. Since ERYN is effectively an enhanced version of the EMCEE package, the overall structure of EMCEE is strongly maintained. Like in EMCEE, ‘State’ objects move coordinate and likelihood information around the ensemble sampler, storing information in a similar back end object either in memory or HDF5 files. Additionally, the interface used for adding proposals has remained.

The various enhancements discussed in this work, including tempering, RJ moves, multiple try MCMC, etc., are all implemented within the EMCEE-like structure. This involved two main changes. First, the State objects have been scaled to hold information necessary for RJ MCMC: temperature information, prior information, and efficient and concise containers for multiple types of models with an adjustable number of individual model instances. Second, the RJ proposal has been added as a proposal base, similar to the use of the ‘MH’ or ‘RedBlue’ moves within EMCEE. Beyond these main enhancements, there are also a variety of smaller, but useful, additions to ERYN that help the user build a variety of analysis pipelines. These include stopping or convergence functions, functions to periodically update the sampler set-up while running, objects to carry special information through the sampler, and aids for coordinate transformation.

3.3 Toy examples

In this section, we present a series of working examples for ERYN. We begin with simple problems, such as searching for simple signals in noisy data, with the aim of demonstrating the performance of this toolbox in a dynamical parameter space. The impact of the different

enhancements discussed in Section 3 will be assessed and discussed. Finally, in Section 4, we will apply this machinery to more realistic problems in GW astronomy.

3.3.1 Searching for pulse signals in Gaussian noise

In this first example, we explore the capabilities of ERYN in a simplified application, commonly encountered in physical sciences. We perform a search for Gaussian pulses in a simulated 2D data set, in the presence of Gaussian noise with variance $\sigma_n = 0.2$. We generate 25 pulses randomly distributed on the $x - y$ plane with all pulses contained within $x, y \in [-10, 10]$ (see Fig. 1a), and amplitude uniformly drawn from $\mathcal{U}[0.7, 1.5]$. The amplitude A_k of each pulse, labelled by k , is considered a free parameter to be estimated, in addition to the Cartesian coordinates of their centres. The pulses’ width was kept fixed to $\sigma_p \times \delta_{ij}$, with $\sigma_p = 0.2$, for the sake of simplicity. Thus, we are required to estimate N_p , the total number of pulses in the data, and also estimate the parameters for each individual signal k : $\vec{\theta}_k = \{A_k, x_k, y_k\}$. The noise variance σ_n is estimated as part of the fit. The analysis of this problem is performed using the adaptive parallel tempering scheme of Section 2.4 and the RJ MCMC proposals (Section 2.6). The in-model proposals for each model component are Gaussians, with a diagonal covariance matrix $\Sigma = 10^{-4}\delta_{ij}$. This proposal is not tuned during sampling. The priors for the parameters are quite wide, covering the entire range of the data, while the prior on the number of pulses k is set to $k \sim \mathcal{U}[0, 50]$. With the above settings, we obtain the results summarized in Figs 1 and 2. In Fig. 1b, we plot the most probable number of Gaussian pulses present in the data, or in other words, the most probable model given this particular data set. It is clear that for the given level of noise, it is straightforward to recover the true number of signals. The noise variance is also estimated accurately as $\sigma_n = 0.2 \pm 2 \times 10^{-3}$. In Fig. 1c, we plot the posterior densities for the parameters of all pulses recovered, while we also mark the true injected values. Fig. 1c shows the trans-dimensional MCMC chains ‘stacked’ over all samples of both model order and model parameters. As already mentioned, in this simplified scenario all signals have similar value for the amplitude, thus the almost unimodal marginal on A_k . This illustrative example is useful as an introductory application to the more complicated case of detection in GW astronomy presented below, in Section 4.

In Fig. 2, three diagnostic quantities for this run are shown. In the top panel, the evolution of temperatures is presented. Following the recipe of Vouden et al. (2016), we control the distances between each temperature chain based on their in-between swap acceptance rate, computed from equation (21). The tuning term $\kappa(t)$ is set to $\kappa(t) = t_0/(v(t + t_0))$, with the *adaptation lag* $t_0 = 10^4$ and the *adaptation time* $v = 10^2$. The middle panel shows the evolution of the swap acceptance rate per number of walkers between the chains running at different temperatures. After $\sim 10^5$ sampler iterations, the system converges to an equilibrium, where the rate of swapping states reaches a single value across the temperature range. In the bottom panel, we show the acceptance rate for the in-model step of the algorithm, for all temperatures. As expected, after temperature equilibrium at $\sim 10^5$ samples, the acceptance rate converges to a (different) value for each temperature, which is higher for higher temperatures (smoother posterior surfaces are easier to explore).

Finally, it is interesting to investigate how the sampled dimensionality of the problem varies at different temperatures. In Fig. 3, we plot the posterior on the number of pulses at each temperature. As expected, higher temperature chains tend to favour lower model

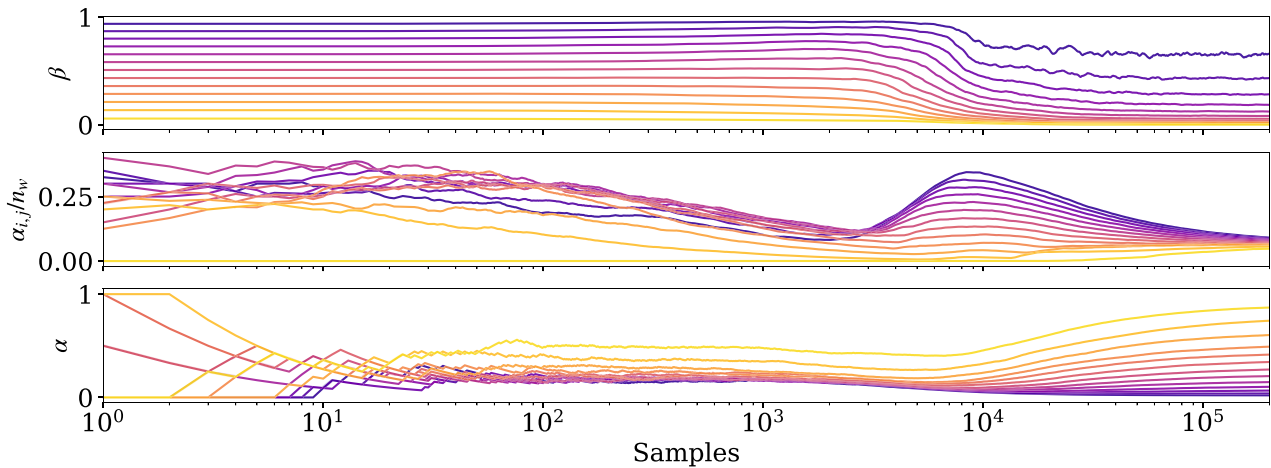


Figure 2. Top panel: The evolution of the temperature chains running in parallel for the toy problem of searching for 2D Gaussian pulses in Gaussian noise. The different colours indicate the initial temperature chain index. Following the parallel tempering recipe of Vouden et al. (2016), the temperature ladder is tuned according to equation (21), and the chains start to converge after $\sim 10^4$ iterations. Middle panel: The evolution of the swap acceptance rate $\alpha_{i,j}$ described in equation (20), per number of walkers n_w . For this run, we have used $n_w = 10$ walkers. After 10^5 iterations, the swap acceptance rate converges to a single (different) value for every temperature chain. Bottom panel: The ‘in-model’ acceptance rate per temperature chain, given by equation (4).

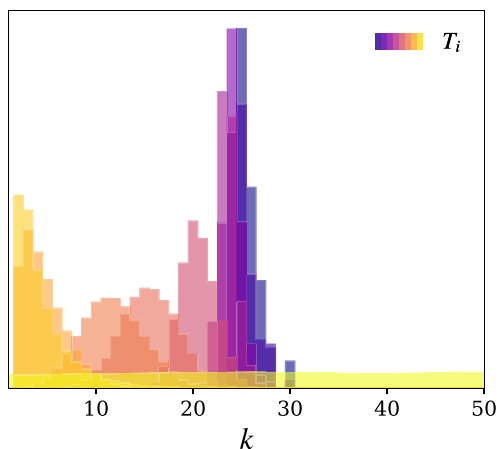


Figure 3. Posterior on the number of Gaussians, k , at each temperature T_i , for the toy problem of Section 3.3.1. The different colours indicate the initial temperature chain index. Darker colours correspond to colder chains and vice versa.

dimensionality and the T_∞ chain samples the prior on k . This can be attributed to the choice of priors and ‘birth proposal’ distributions for both the signal parameters and k . The likelihood is down-weighted at higher temperatures, making it harder to overcome the Occam penalty from including extra parameters in the model. This means quieter sources are less likely to be added and the preferred models have fewer sources.

3.3.2 Modelling power spectra: searching for the optimal number of B-spline knots

One of the most common problems in signal processing is the characterization of the spectra of the data. This is often done by adopting spectral models and fitting the spectra directly in the frequency domain. This methodology is used when the signal of interest has stochastic properties. Examples from GW astronomy,

include the measurement of stochastic signals with astrophysical, or cosmological origin (Abbott et al. 2021c; Auclair et al. 2023). There are many examples of possible stochastic signals for LISA (Amaro-Seoane et al. 2012, 2017, 2023; Auclair et al. 2023). Searching for signals with stochastic properties requires flexible spectral models, both for the observatory instrumental noise, and the measured stochastic signal. For these reasons, it is sometimes convenient to adopt a versatile model, such as one that is based on B-spline interpolation schemes.

B-splines are a geometrical modelling tool, and have proven to be very useful for modelling or generating smoother representations of data. They are piece wise polynomial curves with a certain number of continuous derivatives, and can be parametrized in various ways (Piegl & Tiller 1996). For this application, we follow Baghi et al. (2023), and choose to work with cubic-spline interpolation, using the corresponding SCIPY library (Virtanen et al. 2020). The procedure starts by selecting a number of control points, or *knots*, with a given position and amplitude, which the smooth polynomial curve crosses and at which there is a change in the first non-continuous derivative. One of the challenging problems using such methods, is to choose the optimal number of spline knots for fitting the data, without overfitting. This is a model selection problem that can be easily solved with dynamical algorithms such as the one presented here.

For our next example, we generate time-series data directly from a theoretical model Power Spectral Density (PSD). The simulated data are represented with the solid grey line in Fig. 4a. We then use the machinery of ERYN to find the optimal dimensionality for the problem, together with the best-fitting parameters for the knots. To ease the computational complexity, we compute the PSD of the time series using the methodology developed by Tröbs & Heinzl (2006). In more detail, we begin by choosing a new frequency grid, on which we compute the PSD using the *optimal* number of averaged segments for each given frequency. Following this method, we essentially split the time-series data at the maximum number of segments that the given choice of window and percentage of data overlap permits, which will allow us to estimate the PSD at each frequency bin with minimal variance. By carefully choosing

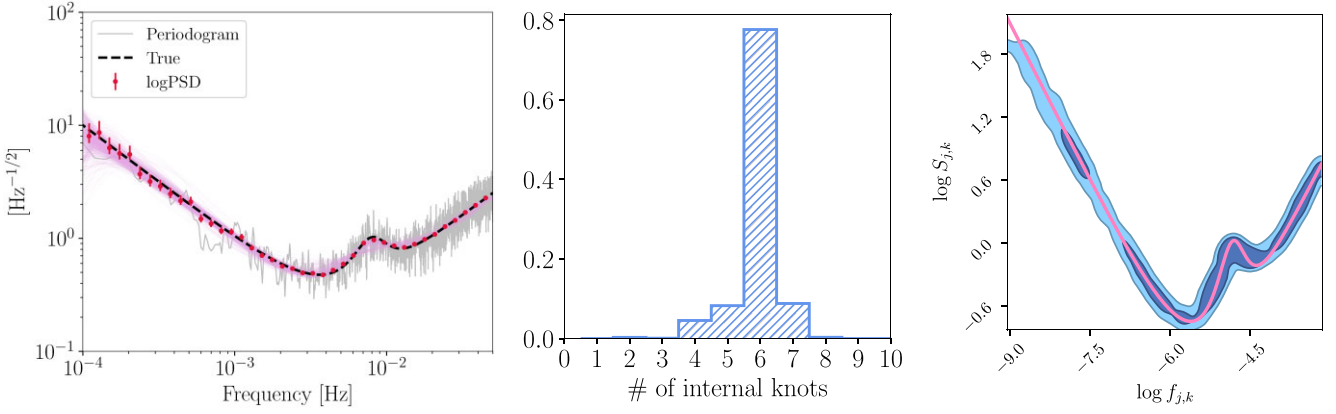


Figure 4. Results for power spectra modelling with a shape agnostic model. (a) The simulated data (grey), generated from the theoretical model (dashed black line). The PSD computed on an equally spaced logarithmic grid with the method of Tröbs & Heinzel (2006) and Armano et al. (2018), which is used for inference, is represented with the red data points. The pink solid lines represent models drawn randomly from the posterior chains. (b) The optimal B-spline knots estimated by the dynamical parameter estimation procedure. As shown from this histogram, the optimal interior knot count for these data converges to six, corresponding to eight total knots including the two edge knots. (c) Posterior slice for the knot parameters, $(\log f_{j,k}, \log S_{j,k})$, after stacking the MCMC chains across all model dimensions, k . This illustrates where the model prefers to place spline knots, which clearly corresponds to where the spectral density is changing most rapidly. It is also evident from this figure that we essentially ‘scan’ the true noise shape (pink solid line), by placing knots across the frequency range (see text for more details).

the window function and distance between the data points, one can compute a power spectrum with minimal correlations between frequencies. The estimated spectrum, D_i , at each frequency, f_i , is then used in the likelihood function given in equation (35) below (the spectrum D_i is represented by the red data points in Fig. 4). For more details about this method of computing the PSD, we refer the reader to Tröbs & Heinzel (2006) and Armano et al. (2018) and for a similar application to the work of Baghi et al. (2023).

Finally, we also keep two knots fixed at the edges of the spectra, allowing the sampler to estimate only their amplitude, while the rest of the knot parameters (and their number) are left to be estimated from the data. For the spline knot positions, $\{\log f_{j,k}\}$, and amplitudes, $\{\log S_{j,k}\}$, we adopt uniform priors that cover the complete parameter space. Here, the j index corresponds to the knot number for the given model order k . We also use a ladder of 10 temperatures, with 10 walkers each, while maintaining the same settings for the adaptivity of the temperatures as in Section 3.3.1. Each walker is initialized at a random point on the parameter space, after drawing the dimensionality k of the model from $k \sim \mathcal{U}[1, 20]$. We adopt a Gaussian likelihood, with its logarithm written as

$$\log p(D|\vec{\theta}_k) \propto -\frac{1}{2} \sum_i n_i \left(\frac{D_i}{\mathcal{N}_{i,k}(\vec{\theta}_k)} + \log \mathcal{N}_{i,k}(\vec{\theta}_k) \right), \quad (35)$$

where D_i is the PSD data value for the given frequency f_i , as computed by the method presented in Tröbs & Heinzel (2006) and Armano et al. (2018), using n_i averaged segments. The $\mathcal{N}_{i,k}(\vec{\theta}_k)$ is the spline noise model of order k evaluated at f_i , that depends on a parameter set

$$\vec{\theta}_k = \{\log f_{1,k}, \dots, \log f_{k,k}, \log S_0, \dots, \log S_{k,k}, \log S_{k+1}\}, \quad (36)$$

in which the $\log S_0$ and $\log S_{k+1}$ parameters refer to the logarithm of the PSD amplitude of the two fixed knots at the ‘edges’ of the spectrum. Those two parameters correspond to our zeroth model order ($k=0$), thus they are always being explored by the walkers of ERYN.

The results are shown in Fig. 4. In particular, in Fig. 4b, we show the histogram of the recovered number of knots for the particular data set. It is clear that eight spline knots are preferred, two of them being fixed at the edges of the spectrum, and the other six knots free

to take any position in the given frequency range. In Fig. 4c, we show the 2D sliced posteriors for the spline parameters, $\{\log S_{j,k}\}$ and $\{\log f_{j,k}\}$. In this figure, we again stack all the MCMC samples across model orders. The true spectrum is indicated by the orange solid line. There is an interesting outcome of this toy investigation; while there is a preferable dimensionality of the model, there is a weak constraint on the actual positions of the knots. We find that the sampler is virtually ‘scanning’ the PSD data, showing slightly higher preference for locations between -6 and -4 in \log -frequency, where the spectrum follows a more complicated shape. Finally, in Fig. 4a, the data (grey solid line and red data points), are shown together with model evaluations drawn from the posterior samples (pink solid lines).

4 EXAMPLES FROM GW ASTRONOMY

In recent years, we have witnessed the beginning of GW astronomy. Since the first detection (Abbott et al. 2016) dozens of waveform signatures have been measured with the current network of observatories. At the time of the writing of this paper, more than 90 events have been recorded (Abbott et al. 2021a), the vast majority of them are BH binary mergers, with a few of them being binary NS and BH–NS mergers. At the same time, detector networks are being improved (The LIGO Scientific Collaboration 2019; Abbott et al. 2020) and there are plans to expand them with the addition of new observatories, such as the *Einstein Telescope* (Punturo et al. 2010; Maggiore et al. 2020) or *Cosmic Explorer* (Abbott et al. 2017; Evans et al. 2021). Those detectors will unlock the sky to larger redshifts z , allowing access to a vast number of potential sources. In addition, space missions, such as LISA (Amaro-Seoane et al. 2012, 2017), are predicted to be signal-dominated observatories, with many types of sources populating their data streams. In fact, we expect that source confusion will be one of the primary challenges in future data analysis efforts in GW astronomy. In a typical data set, we expect an unknown number of signals, originating from sources that generate waveforms with different characteristics. Those range from the stellar-mass BH binaries now frequently observed by ground-based detectors, to the supermassive BH binaries, extreme mass ratio

inspirals, UCBs, and stochastic GW signals from both astrophysical and possibly cosmological origin (Amaro-Seoane et al. 2012, 2017; Auclair et al. 2023). For this final example, we will focus on the LISA mission, and in particular on the case of discriminating UCB signals.

4.1 Application to LISA data and the UCBs

LISA is going to measure GW signals in the mHz regime, accessing sources of all the aforementioned types. As already discussed, the most numerous of them are going to be the UCBs, which will be almost monochromatic in the LISA band. Out of the millions of sources, only $\sim \mathcal{O}(10^4)$ will be individually resolvable, and the rest will generate a confusion signal. As a consequence, for the duration of the mission, we will need to disentangle tens of thousands of sources which will be overlapping in both time and frequency domains. This is no trivial task, but various different strategies have already been proposed for analysing such challenging data sets. For example, Gaussian processes (Strub et al. 2022), swarm optimization techniques (Zhang et al. 2021), or hybrid swarm-based algorithms (Bouffanais & Porter 2016) can be employed. Pipelines based on MCMC methods have been tested extensively (Crowder & Cornish 2007; Littenberg 2011; Littenberg et al. 2020; Littenberg & Cornish 2023), and have been demonstrated to be able to tackle complex cases where signals are overlapping.

Here, we will focus on the same problem, employing ERYN to solve a down-scaled version of the UCB challenge. It is down-scaled because we focus only on a single narrow-frequency band, containing several overlapping signals, in the presence of instrumental noise only.³ In addition, we focus solely on demonstrating the capabilities of ERYN on dynamic parameter estimation for UCB type sources and no other types of signals are contained in the data (e.g. chirping signatures from supermassive BH binaries). At the same time, we have access to the level of instrumental noise, which is shown in both panels of Fig. 5. Searching for the UCB signals across the complete LISA band requires a more elaborate implementation of this simplified pipeline. This pipeline will be focusing on solving the complete second LISA Data Challenge (LDC2; LISA Data Challenges Working Group 2022), and is going to be presented in future work. We choose to work on the frequency segment between 3.997 and 4 mHz, which contains 10 UCB objects, drawn directly from the LDC2 catalogue (LISA Data Challenges Working Group 2022). Those are shown in the top panel of Fig. 5 which shows the power spectrum of the A data channel of LISA. We use the two noise-orthogonal A , E , and T time delay interferometry (TDI) variables (Prince et al. 2002; Tinto & Dhurandhar 2005; Baghi et al. 2021), which are linear combinations of the LISA relative frequency TDI Michelson measurements X , Y , and Z as:

$$\begin{aligned} A &= \frac{1}{\sqrt{2}}(Z - X), & E &= \frac{1}{\sqrt{6}}(X - 2Y + Z), \\ T &= \frac{1}{\sqrt{3}}(X + Y + Z). \end{aligned} \quad (37)$$

In ideal conditions (equal noises across spacecrafts, and equal LISA arms), the noise in A and E is independent, while the T data stream can be used as a signal-insensitive *null* channel, useful for instrument noise calibration. Since we perform analysis on a noise-free injection,

³No confusion signal from other unresolved UCBs is considered in this investigation.

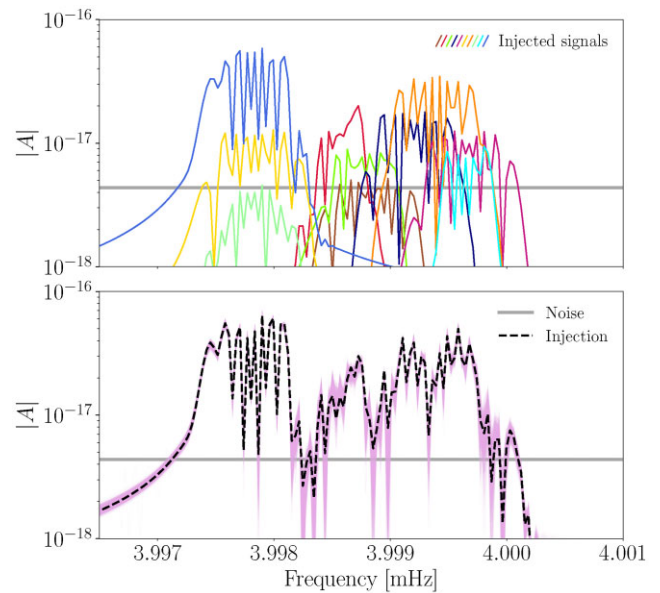


Figure 5. Top panel: The simulated data used for demonstrating the capabilities of ERYN in tackling a high-dimensional problem. A total of 10 UCBs in the vicinity of our Galaxy emitting GW signals at the mHz range. Here, we plot the power spectral density of the A data channel of LISA. The catalogue of sources is taken from the LDC2 (LISA Data Challenges Working Group 2022). Each signal is represented by a different colour. Bottom panel: The same data set, now comparing the injected signal against the solution yielded by ERYN (see text for more details). We have plotted the shaded area by sampling the joint posterior on model order k and the corresponding parameters.

Table 1. The optimal SNR ρ_{opt} for each of the 10 injected sources, computed for the given duration of the mission (see equation 42). The dominant emission frequency f_{gw} is also given for reference.

#	f_{gw} (mHz)	ρ_{opt}
1	3.99780	9.98
2	3.99781	46.70
3	3.99784	4.55
4	3.99854	39.45
5	3.99873	13.02
6	3.99882	8.47
7	3.99919	10.88
8	3.99939	19.07
9	3.99964	20.00
10	3.99965	7.99

we will be neglecting the T channel altogether. We simulated the injection data for an observation time of $T_{\text{obs}} = 1$ yr.

The *optimal* SNR for each injected source, ρ_{opt} , is given in Table 1. The ρ_{opt} quantity refers to the SNR of each source in *isolation*, with respect only to the instrumental noise, and can be calculated as

$$\rho_{\text{opt}}^2 = \sum_C (h_C | h_C)_C, \quad (38)$$

with $C \in \{A, E\}$ the noise-orthogonal TDI channels of equation (37), while the $(\cdot | \cdot)$ notation represents the noise-weighted inner product expressed for two time series a and b as

$$(a|b) = 2 \int_0^\infty df [\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)] / \tilde{S}_n(f). \quad (39)$$

The tilde represents the data in the Fourier frequency domain, and the asterisk indicates complex conjugate. The $\tilde{S}_n(f)$ is the one-sided PSD of the noise for a given TDI channel. Under our assumptions $S_{n,A}(f) = S_{n,E}(f)$.

For our investigation, we chose to analyse noiseless data (no noise realization), while in the likelihood we are using the PSD noise levels taken from the LISA design studies (LISA Science Study Team 2018). For the signals, we utilize the fast frequency-domain UCB waveform model of Cornish & Littenberg (2007). Then, the two polarizations of an emitting UCB can be written as

$$\begin{aligned} h_+(t) &= \frac{2\mathcal{M}}{D_L} (\pi f_{\text{gw}}(t))^{2/3} (1 + \cos^2 \iota) \cos \phi(t), \\ h_\times(t) &= -\frac{4\mathcal{M}}{D_L} (\pi f_{\text{gw}}(t))^{2/3} \cos \iota \sin \phi(t), \end{aligned} \quad (40)$$

where \mathcal{M} is the chirp mass, f_{gw} is the instantaneous GW frequency, D_L is the luminosity distance, ι is the inclination of the binary orbit, and $\phi(t)$ is the GW phase. The phase ϕ can be expressed as $\phi = \phi_0 + 2\pi \int^t f_{\text{gw}}(t') dt'$, with ϕ_0 being an initial arbitrary phase shift. The LISA constellation is assumed to be rigid and with equal arms, while the spacecrafts are assumed to follow analytic Keplerian orbits (Cornish & Rubbo 2003). Under these assumptions, it is straightforward to compute its response to the almost monochromatic waveforms of equation (40) (Cornish & Rubbo 2003; Babak, Petiteau & Hewitson 2021). For more details about the waveform model, the response of the instrument, and the orbits of the constellation, we refer the reader to Cornish & Littenberg (2007), Robson et al. (2018), Katz et al. (2022), Cornish & Rubbo (2003), and Babak et al. (2021).

In our simplified scenario, each binary signal in the Solar system barycentre is determined by a set of eight parameters. Those are the $\vec{\theta} = \{\mathcal{A}, f_{\text{gw}} [\text{mHz}], \dot{f}_{\text{gw}} [\text{Hz s}^{-1}], \phi_0, \cos \iota, \psi, \lambda, \sin \beta\}$, where \mathcal{A} is the overall amplitude, \dot{f}_{gw} is the first derivative of the GW frequency, ψ the polarization, λ is the ecliptic longitude, and β the ecliptic latitude of the binary. The amplitude of the signal is calculated as

$$\mathcal{A} = \left(2\mathcal{M}^{5/3} \pi^{2/3} \dot{f}_{\text{gw}}^{2/3} \right) / D_L, \quad (41)$$

which can be used to obtain a rough SNR estimate, via (Littenberg et al. 2020)

$$\rho^2 = \frac{\mathcal{A}^2 T_{\text{obs}} \sin^2(f_{\text{gw}}/f_*)}{4S_n(f_{\text{gw}})}, \quad (42)$$

with $S_n(f_{\text{gw}})$ being the instrumental noise power spectral density at frequency f_{gw} , and $f_* = 1/(2\pi L)$, where L the LISA arm length. Given equation (41) and (42), we find it convenient to directly sample on ρ instead of \mathcal{A} , which also yields a more illustrative measure of the amplitude of each binary. Then, if d is the measured TDI data and h the given GW signal after applying the response of the instrument, the logarithm of the likelihood for an arbitrary number k of UCB signals can be written as

$$\log p(d|\vec{\theta}_k) \propto (d|h_k) - \frac{1}{2} (h_k|h_k), \quad (43)$$

where for the sake of convenience, we have defined $h_k = \sum_k h(\vec{\theta}_k)$.

We use wide uniform priors for all the rest of the binary parameters, covering essentially the complete parameter space. The exception is again the amplitude (SNR), ρ , where we adopt a prior which was first introduced in Cornish & Littenberg (2015) and then adapted in Littenberg et al. (2020). The prior density can be expressed as

$$p(\rho) = \frac{3\rho}{4\rho_* (1 + \rho/(4\rho_*))}, \quad (44)$$

where ρ_* is a given constant that specifies the peak of the above distribution. This distribution is designed to prevent the proposal of sources with very small SNR in the model, as it drops sharply for $\rho \rightarrow 0$. Those weak sources do not significantly affect the likelihood, and so their inclusion must be penalized by the prior.⁴ This prior choice forces the sampler to explore only potential sources with non-zero SNR, avoiding populating the chains with numerous undetectable signals. This prior performs adequately in this problem, but there are other solutions one could adopt in order to keep control of the number of very weak sources. This discussion, which sets the grounds for a Global Fit analysis pipeline for the LISA data (Littenberg et al. 2020), is out of the scope of this paper, but a more detailed description will be presented in a future work.

4.1.1 Search phase

Before parameter estimation, we initiate a *search phase* of our analysis, with the aim of getting the walkers to a better starting point on the posterior surface. This phase consists of an iterative brute force procedure, based on drawing a very large number of proposals, then maximizing the likelihood over the initial phase ϕ_0 , and finally perform a rapid MCMC sampling over the parameter space, using only a one-source model (therefore there are no dynamical parameter spaces). In particular, we draw 5×10^6 points in the parameter space, and after phase maximization, we use them as starting points to a parallel tempered MCMC run with $N_T = 10$ temperatures, each running with $n_w = 500$ walkers. When this step concludes, we keep the 100 best samples in terms of likelihood value and use their corresponding parameter estimates as starting points for the parameter estimation portion of the analysis. We also use the maximum likelihood solution to subtract the source found from the data. We then use the residual data to search for another source, and this process repeats until there is no signal found with SNR $\rho > 5$. In between successive iterations of the single-binary search, we run another MCMC over all sources found so far in order to adjust the parameters to account for correlations and overlap between sources. After convergence, we found eight sources in our data set with an optimal SNR > 5 . We take these found sources and add them to all walkers in the sampler at the beginning of the full MCMC run described below.

4.1.2 Parameter estimation

During this step, we perform hybrid MCMC sampling, where we both update the found sources (in-model) and dynamically search for new and weaker sources in the data employing RJ sampling. For the number of signals k , we adopt a uniform prior $k \sim \mathcal{U}[6, 20]$. For the sake of convenience in this simple application, we keep the six loudest binaries found during the search phase as *fixed*. This means that we still sample their waveform parameters, but they are not allowed to be removed by the RJ process. We chose this set-up in order to accelerate the convergence of the algorithm, being confident that these sources are part of the solution. In future work, this will be adjusted to deal with the much larger complexity of the full problem.

Concerning the sampler settings, we use the adaptive parallel tempering scheme of Section 2.4, building a temperature ladder of $N_T = 10$ temperatures, with 100 walkers for each temperature. For this run, we have also utilized the MTM algorithm (see Section 2.3) in order to improve the acceptance rate in the RJ proposal. We have

⁴We remind that the prior is also used as our ‘birth’ proposal here.

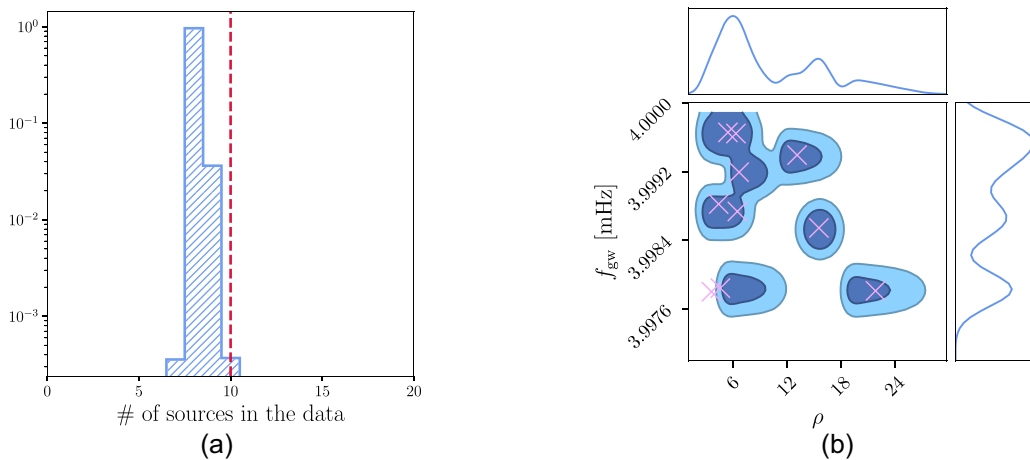


Figure 6. Left panel: In this figure, we show the posterior on the number of UCB sources in the data. The true injected number is shown with the red dashed line. It is clear that, for the given measurement duration of the particular data set, we manage to confidently resolve eight binaries out of a total of 10. Right panel: Corner plot for two of the eight parameters characterizing each UCB source. These are the amplitude, expressed as an SNR ρ , and the dominant emission (or initial) frequency, f_{gw} (mHz) (see text for more details). The violet crosses represent the injected parameter values. A corner plot for more parameters is shown in Fig. A1 in the Appendix.

also tried the DR scheme which is implemented for ERYN, but we found that the multiple try strategy yields more efficient sampling. Finally, we have utilized the basic stretch and group stretch proposals that were described in Section 2.1.

After convergence, the result is shown in Figs 5 and 6. In Fig. 6a, the sampled posterior on the number of sources k is presented. In this histogram, we have added the six fixed binaries to the actual number of signals being sampled via the RJ algorithm. It is fairly obvious that we have managed to confidently resolve eight out of the 10 injected binary signals. This fact that we do not favour 10 sources can be explained partly by the low SNR of the signals (see Table 1) and partly by confusion from source overlap (also shown in Fig. 5). Additionally, the result of Fig. 6a depends on the given observation duration. The greater the T_{obs} , the better our ability to resolve the confused sources. Thus, in that case, we should expect more RJ iterations across the higher dimensional models.

On the right panel, in Fig. 6b, the ensemble 2D posterior slice is shown, for two selected parameters. We call it ensemble because we are again ‘stacking’ all the chains for these two parameters for all sources for all model orders k . We chose to show only the amplitude (the ρ parameter explained in equation 42) and the dominant emission frequency f_{gw} , which illustrates the number of sources resolved, and how they overlap in frequency. A corner plot for more parameters is shown in Fig. A1 in Appendix A. We also show the true injection values, marked as crosses, on top of the 2D posterior. From this plot alone, one can see that the sampler is exploring efficiently the parameter space, converging to the true values of the resolvable binaries that were injected.

5 DISCUSSION

We have implemented ERYN, a Bayesian sampling package capable of performing efficient trans-dimensional inference, by employing different techniques that improve its acceptance rate. These techniques are the affine invariant sampling, the adaptive parallel tempering, the DR, and MTM, in combination with the construction of informative proposal distributions for the parameters of the models. The structure of ERYN is based on the widely used software EMCEE (Foreman-Mackey et al. 2013), enhanced with the ability of

performing RJs (Green 1995) between different model spaces. The sampler capabilities have been demonstrated with toy models that are commonly encountered in different data analysis problems. We have begun with an application to signal detection, and in particular to searching for simple signals in the form of Gaussian pulse signals in the presence of Gaussian noise (see Section 3.3.1).

In Section 3.3.2, we applied our algorithm to a problem of modelling power spectra with arbitrary shapes in frequency domain. In such cases, it is convenient to define models based on B-splines, which are able to faithfully capture the shape of any spectral data. However, in order to avoid overfitting situations, the optimal order of the model (i.e. the optimal number of spline knots), needs to be estimated from the data. This can be done either sequentially, by trying models of different dimensionality and then comparing their performance, or dynamically, by using trans-dimensional algorithms such as ERYN. This class of problems is often encountered in cosmology (Planck Collaboration I 2020a; Planck Collaboration V 2020b), where the signal of interest is stochastic in nature, and sometimes the prior knowledge on its shape is very limited. As already discussed, this is especially true for future GW observatories, which open the possibility of detecting such signals from both astrophysical and cosmological origin (Amaro-Seoane et al. 2017; Auclair et al. 2023; Baghi et al. 2023). The different theoretical models produce spectra with distinct shapes, increasing the need for shape-agnostic spectral models, such as the B-spline used here.

Finally, in Section 4.1, we demonstrated ERYN in a more complicated problem, that of the analysis of UCB signals measured by the future LISA detector. These objects are going to produce the majority of the signals in the LISA data, each emitting almost monochromatic radiation. Their vast number will generate a confusion foreground, while only a few thousand of them will be resolvable from the data. We employ our tools described in this work, together with a search phase that is based on iteratively running the sampler on ‘static models’ (no trans-dimensional moves) with phase-maximized likelihoods. We do these runs on the residuals of each iteration, with the aim of extracting all bright sources. In order to achieve faster convergence of our parameter estimation run, we choose to keep the brightest sources found during search as fixed (minimum number of model order is $k = 6$), while the RJ algorithm is used to search for

weaker signals in the data. This is purely a choice that allows quick convergence in this fully controlled and simplified LISA data set.

We perform this analysis for a mission duration of $T_{\text{obs}} = 1$ yr and only on a single narrow-frequency band around 4 mHz, which contains a total of 10 binary signals. It is worth noting here, that the synthetic data were produced assuming idealized conditions. This means that we do not consider any data irregularities, such as data gaps and glitches and spectral lines, or any other contamination originating from the mixing of signals of different types (such as supermassive BH binaries). In the end, as shown in Fig. 6a, we manage to recover eight out of ten injected signals. This result makes sense given the relative strength of the injections, and their waveform overlap. Many of the injected sources have an optimal SNR in isolation which is rather low (see Table 1), so these are more susceptible to deterioration when we account for signal overlap.

The above investigations demonstrate that the dynamical parameter estimation capabilities of ERYN are suitable for these types of problems. This feature is missing from already existing libraries such as BILBY (Ashton & Talbot 2021), which are used by the GW community. BILBY offers a wide selection of tools which are necessary for the Bayesian analysis of GW data. These include implementations of likelihood and prior functions, instrument response models and spectral densities, waveform models, and a number of samplers to choose from (both MCMC and nested samplers). On the other hand, ERYN is based on parallel tempering MCMC enhanced with RJ, which allows Bayesian analyses for a wider set of problems, in which sampling of a dynamical parameter space is needed.

ERYN has already been used in several works that have been already published (Katz 2022; Katz et al. 2022; Baghi et al. 2023; Sasli, Karnesis & Stergioulas 2023), or are going to appear soon. The work presented in this paper is the initial part of our efforts towards implementing a data analysis pipeline for LISA data. This pipeline will be demonstrated on the LDC2 data set (LISA Data Challenges Working Group 2022), which contains multiple types of signals overlapping in both time and frequency domains. That being said, ERYN is a generic and versatile sampler, which can be used in any investigation that requires RJ sampling, and to our knowledge is one of the very few statistical tools of this kind that is not specialized to a single type of analysis (see discussion in Section 2.6).

ACKNOWLEDGEMENTS

The ERYN is available as open-source software at: <https://github.com/mikekatz04/Erlyn>. We wish to thank S. Babak, M. Le Jeune, S. Marsat, T. Littenberg, and N. Cornish for their useful comments and very fruitful discussions. NS and NK acknowledge support from the Gr-PRODEX 2019 funding program (PEA 4000132310).

DATA AVAILABILITY

The sampler code can be found at: <https://github.com/mikekatz04/Erlyn>. The code for the toy examples of Section 3.3 can be found in the documentation of ERYN (<https://mikekatz04.github.io/Erlyn/html/index.html#>). For the analysis of Section 4, we used the GBGPU package (<https://github.com/mikekatz04/GBGPU>), which contains an implementation of the UCB waveforms of the LDC software (<https://lisa-ldc.lal.in2p3.fr/code>).

REFERENCES

- Abbott B. P. et al., 2016, *Phys. Rev. Lett.*, 116, 061102
 Abbott B. P. et al., 2017, *Class. Quantum Gravity*, 34, 044001

- Abbott B. P. et al., 2020, *Living Rev. Relativ.*, 23, 1
 Abbott R. et al., 2021a, preprint (arXiv:2111.03606)
 Abbott R. et al., 2021b, *SoftwareX*, 13, 100658
 Abbott R. et al., 2021c, *Phys. Rev. D*, 104, 022004
 Amaro-Seoane P. et al., 2012, *Class. Quantum Gravity*, 29, 124016
 Amaro-Seoane P. et al., 2017, preprint (arXiv:1702.00786)
 Amaro-Seoane P. et al., 2023, *Living Rev. Relativ.*, 26, 2
 Andrieu C., Thoms J., 2008, *Stat. Comput.*, 18, 343
 Armano M. et al., 2018, *Phys. Rev. Lett.*, 120, 061101
 Ashton G., Prix R., 2018, *Phys. Rev. D*, 97, 103020
 Ashton G., Talbot C., 2021, *MNRAS*, 507, 2037
 Auclair P. et al., 2023, *Living Rev. Relativ.*, 26, 5
 Babak S., Petiteau A., Hewitson M., 2021, preprint (arXiv:2108.01167)
 Baghi Q., Thorpe J. I., Slutsky J., Baker J., 2021, *Phys. Rev. D*, 103, 042006
 Baghi Q., Karnesis N., Bayle J.-B., Besançon M., Inchauspé H., 2023, *J. Cosmol. Astropart. Phys.*, 2023, 066
 Baio G., 2012, in *Bayesian Methods in Health Economics*. Chapman & Hall/CRC, New York
 Bédard M., Douc R., Moulines E., 2012, *Stoch. Process. Appl.*, 122, 758
 Betancourt M. J., Girolami M., 2013, preprint (arXiv:1312.0906)
 Biwer C. M., Capano C. D., De S., Cabero M., Brown D. A., Nitz A. H., Raymond V., 2019, *PASP*, 131, 024503
 Bouffanais Y., Porter E. K., 2016, *Phys. Rev. D*, 93, 064020
 Breivik K. et al., 2020, *ApJ*, 898, 71
 Brooks S. P., 1998, *J. R. Stat. Soc. D*, 47, 69
 Brooks S., Gelman A., 1998, *J. Comput. Graph. Stat.*, 7, 434
 Christensen N., Meyer R., 2022, *Rev. Mod. Phys.*, 94, 025001
 Cornish N. J., 2021, *Phys. Rev. D*, 104, 104054
 Cornish N. J., Littenberg T. B., 2007, *Phys. Rev. D*, 76, 083006
 Cornish N. J., Littenberg T. B., 2015, *Class. Quantum Gravity*, 32, 135012
 Cornish N. J., Rubbo L. J., 2003, *Phys. Rev. D*, 67, 022001
 Crowder J., Cornish N. J., 2007, *Phys. Rev. D*, 75, 043008
 De S., Finstad D., Lattimer J. M., Brown D. A., Berger E., Biwer C. M., 2018, *Phys. Rev. Lett.*, 121, 091102
 Dellaportas P., Forster J. J., Ntzoufras I., 2002, *Stat. Comput.*, 12, 27
 Dinh L., Sohl-Dickstein J., Bengio S., 2016, preprint (arXiv:1605.08803)
 Documentation, Eryn., Eryn Documentation. mikekatz04.github.io/Erlyn/
 Durkan C., Bekasov A., Murray I., Papamakarios G., 2019, preprint (arXiv:1906.04032)
 Evans M. et al., 2021, preprint (arXiv:2109.09882)
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
 Frenkel D., 2006, in *Waste-Recycling Monte Carlo*. Springer, Berlin, p. 127
 Gelman A., Rubin D. B., 1992, *Stat. Sci.*, 7, 457
 Gelman A., Roberts G. O., Gilks W. R., 1996, in Bernardo J. M., Berger J. O., Dawid A. P., Smith A. F. M., eds, *Bayesian Statistics*. Oxford Univ. Press, Oxford, p. 599
 Gelman A., Carlin J. B., Stern H. S., Rubin D. B., 2004, in *Bayesian Data Analysis*, 2nd edn. Chapman and Hall/CRC, New York
 Gilks W. R., Wild P., 1992, *J. R. Stat. Soc. C*, 41, 337
 Gilks W., Richardson S., Spiegelhalter D., 1995, in *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, New York
 Godsill S. J., 2001, *J. Comput. Graph. Stat.*, 10, 230
 Goggans P. M., Chi Y., 2004, in Erickson G. J., Zhai Y., eds, in *AIP Conf. Proc. Vol. 707, Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Am. Inst. Phys., New York, p. 59
 Goodman J., Weare J., 2010, *Commun. Appl. Math. Comput. Sci.*, 5, 65
 Green P. J., 1995, *Biometrika*, 82, 711
 Green P. J., Mira A., 2001, *Biometrika*, 88, 1035
 Gateway Tolkien, Tolkien Gateway. 2023, <http://tolkiengateway.net/wiki/Erlyn-Vorn>
 Haario H., Saksman E., Tamminen J., 2001, *Bernoulli*, 7, 223
 Haario H., Laine M., Mira A., Saksman E., 2006, *Stat. Comp.*, 16, 339
 Hastings W. K., 1970, *Biometrika*, 57, 97
 Hermans J., Begy V., Louppe G., 2019, preprint (arXiv:1903.04057)
 Hitchcock D. B., 2003, *Am. Stat.*, 57, 254
 Hobson M. P., Jaffe A. H., Liddle A. R., Mukherjee P., Parkinson D., 2009, in *Bayesian Methods in Cosmology*. Cambridge Univ. Press, Cambridge

- Hoffman M. D., Gelman A., 2014, *J. Mach. Learn. Res.*, 15, 1593
- Hogg D. W., Foreman-Mackey D., 2018, *ApJS*, 236, 11
- Hukushima K., Nemoto K., 1996, *J. Phys. Soc. Japan*, 65, 1604
- Isi M., Farr W. M., Chatziioannou K., 2022, *Phys. Rev. D*, 106, 024048
- Johannes M., Polson N., 2010, in Ait-Sahalia Y., Hansen L. P., eds, *Handbooks in Finance*, Vol. 2, *Handbook of Financial Econometrics: Applications*. Elsevier, San Diego, p. 1
- Joseph A., 2020, in *Markov Chain Monte Carlo Methods in Quantum Field Theories*. Springer International Publishing, Switzerland
- Karnesis N. et al., 2014, *Phys. Rev. D*, 89, 062001
- Karnesis N., Babak S., Pironi M., Cornish N., Littenberg T., 2021, *Phys. Rev. D*, 104, 043019
- Kass R. E., Raftery A. E., 1995, *J. Am. Stat. Assoc.*, 90, 773
- Katz M. L., 2022, *Phys. Rev. D*, 105, 044055
- Katz M. L., Danielski C., Karnesis N., Korol V., Tamanini N., Cornish N. J., Littenberg T. B., 2022, *MNRAS*, 517, 697
- Kawamura S. et al., 2008, *J. Phys. Conf. Ser.*, 122, 012006
- Kendall W., Liang F., Wang J., 2005, in *Markov Chain Monte Carlo: Innovations and Applications*, Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore. World Scientific Press, Singapore
- Khan Z., Balch T., Dellaert F., 2005, *IEEE Trans. Pattern Anal. Mach. Intell.*, 27, 1805
- Lartillot N., Philippe H., 2006, *Syst. Biol.*, 55, 195
- LIGO Collaboration et al., 2020, *Class. Quantum Gravity*, 37, 055002
- LISA Data Challenges Working Group, 2022, *LISA Data Challenges*. <https://lisa-ldc.lal.in2p3.fr/>
- LISA Science Study Team, 2018, in *Technical Report ESA-L3-EST-SCI-RS-001*, LISA Science Requirements Document. ESA, Noordwijk, <https://www.cosmos.esa.int/documents/678316/1700384/SciRD.pdf>
- Littenberg T. B., 2011, *Phys. Rev. D*, 84, 063009
- Littenberg T. B., Cornish N. J., 2009, *Phys. Rev. D*, 80, 063007
- Littenberg T. B., Cornish N. J., 2010, *Phys. Rev. D*, 82, 103007
- Littenberg T. B., Cornish N. J., 2023, *Phys. Rev. D*, 107, 063004
- Littenberg T., Cornish N., Lackeos K., Robson T., 2020, *Phys. Rev. D*, 101, 123021
- Liu J. S., Liang F., Wong W. H., 2000, *J. Am. Stat. Assoc.*, 95, 121
- Lopes H. F., West M., 2004, *Stat. Sinica*, 14, 41
- Luo J. et al., 2016, *Class. Quantum Gravity*, 33, 035010
- McMillan P. J., 2017, *MNRAS*, 465, 76
- Maggiore M. et al., 2020, *J. Cosmol. Astropart. Phys.*, 03, 050
- Marrs A., 1997, in Jordan M., Kearns M., Solla S., eds, *Advances in Neural Information Processing Systems*, Vol. 10. MIT Press, <https://proceedings.neurips.cc/paper/1997/file/0ed9422357395a0d4879191c66f4faa2-Paper.pdf>
- Martino L., 2018, *Digit. Signal Process.*, 75, 134
- Martino L., Del Olmo V. P., Read J., 2012, *Stat. Probab. Lett.*, 82, 1445
- Maturana-Russel P., Meyer R., Veitch J., Christensen N., 2019, *Phys. Rev. D*, 99, 084006
- Menéndez M., Pardo J., Pardo L., Pardo M., 1997, *J. Franklin Inst.*, 334, 307
- Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, *J. Chem. Phys.*, 21, 1087
- Mira A., 2001, *Metron - Int. J. Stat.*, LIX, 231
- Muller P., 1991, in *Technical Report 91-09*, A Generic Approach to Posterior Integration and Gibbs Sampling. Purdue Univ.
- Neal R., 2011, in *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, New York, p. 113
- Nelemans G., Yungelson L. R., Portegies Zwart S. F., 2001, *A&A*, 375, 890
- Piegl L., Tiller W., 1996, in *The NURBS Book*, 2nd edn. Springer, New York, NY
- Planck Collaboration I, 2020a, *A&A*, 641, A1
- Planck Collaboration V, 2020b, *A&A*, 641, A5
- Prince T. A., Tinto M., Larson S. L., Armstrong J. W., 2002, *Phys. Rev. D*, 66, 122002
- Punturo M. et al., 2010, *Class. Quantum Gravity*, 27, 194002
- Ren Z. et al., 2023, *Front. Phys.*, 18, 64302
- Ritter C., Tanner M., 1992, *J. Am. Stat. Assoc.*, 87, 861
- Robert C., Casella G., 2011, *Stat. Sci.*, 26, 102
- Roberts G. O., Rosenthal J. S., 2009, *J. Comput. Graph. Stat.*, 18, 349
- Robson T., Cornish N. J., Tamanini N., Toonen S., 2018, *Phys. Rev. D*, 98, 064012
- Sasli A., Karnesis N., Stergioulas N., 2023, preprint ([arXiv:2305.04709](https://arxiv.org/abs/2305.04709))
- Sharma P. S., 2017, *ARA&A*, 55, 213
- Shawhan P. S., LIGO Scientific Collaboration, 2003, *Nucl. Instrum. Methods Phys. Res. A*, 502, 396
- Sivia D., Skilling J., 2006, in *Data Analysis: A Bayesian Tutorial*. Oxford Univ. Press, Oxford
- Sorensen D., Gianola D., 2007, in *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Statistics for Biology and Health. Springer, New York
- Strub S. H., Ferraioli L., Schmelzbach C., Stähler S. C., Giardini D., 2022, *Phys. Rev. D*, 106, 062003
- Swendsen R. H., Wang J.-S., 1986, *Phys. Rev. Lett.*, 57, 2607
- The LIGO Scientific Collaboration, 2019, preprint ([arXiv:1904.03187](https://arxiv.org/abs/1904.03187))
- Thrane E., Talbot C., 2019, *Publ. Astron. Soc. Aust.*, 36, e010
- Tinto M., Dhurandhar S. V., 2005, *Living Rev. Relativ.*, 8, 4
- Trias M., Vecchio A., Veitch J., 2009, preprint ([arXiv:0904.2207](https://arxiv.org/abs/0904.2207))
- Tröbs M., Heinzl G., 2006, *Measurement*, 39, 120
- Vallisneri M., 2008, *Phys. Rev. D*, 77, 042001
- Virtanen P. et al., 2020, *Nature Methods*, 17, 261
- Vousden W., Farr W., Mandel I., 2016, *MNRAS*, 455, 1919
- Wong K. C. et al., 2020, *MNRAS*, 498, 1420
- Wu H., Köhler J., Noé F., 2020, preprint ([arXiv:2002.06707](https://arxiv.org/abs/2002.06707))
- Xie W., Lewis P. O., Fan Y., Kuo L., Chen M.-H., 2010, *Syst. Biol.*, 60, 150
- Yu X., Wang X., Lu C., Wang Y., 2021, *IOP Conf. Ser.: Earth Environ. Sci.*, 660, 012124
- Zhang X.-H. et al., 2021, *Phys. Rev. D*, 104, 024023

APPENDIX A:

In Fig. A1, we show the triangle plot of the stacked posterior points as sampled by ERYN, for the investigation of Section 4. The difference to Fig. 6b is that here we include more parameters of the sources, but we still do not include all parameters for the sake of clarity.

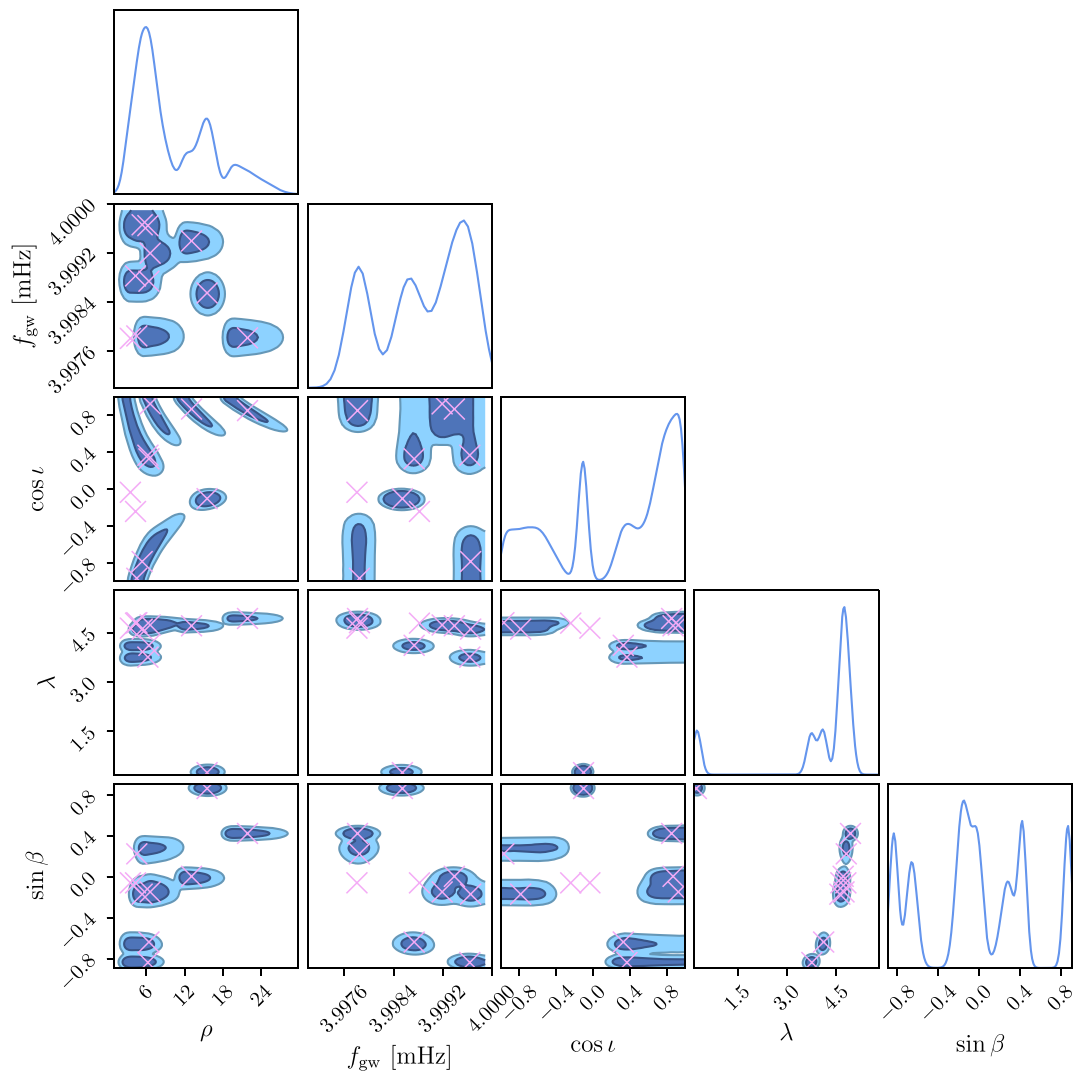


Figure A1. A triangle plot showing the 2D posterior slices for the application of Section 4.1, but for a greater selection of parameters than Fig. 6b. The rest of the parameters, if plotted stacked in the same manner, result in surfaces that cannot be so easily interpreted, and therefore have been left out. The true injected values are marked with crosses.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.