

# Rapid determination of *LISA* sensitivity to extreme mass ratio inspirals with machine learning

Christian E. A. Chapman-Bird<sup>1</sup>,<sup>2</sup> Christopher P. L. Berry<sup>1</sup> and Graham Woan

*SUPA, University of Glasgow, Glasgow G12 8QQ, UK*

Accepted 2023 May 5. Received 2023 May 4; in original form 2023 January 26

## ABSTRACT

Gravitational wave observations of the inspiral of stellar-mass compact objects into massive black holes (MBHs), extreme mass ratio inspirals (EMRIs), enable precision measurements of parameters such as the MBH mass and spin. The *Laser Interferometer Space Antenna* is expected to detect sufficient EMRIs to probe the underlying source population, testing theories of the formation and evolution of MBHs and their environments. Population studies are subject to selection effects that vary across the EMRI parameter space, which bias inference results if unaccounted for. This bias can be corrected, but evaluating the detectability of many EMRI signals is computationally expensive. We mitigate this cost by (i) constructing a rapid and accurate neural network interpolator capable of predicting the signal-to-noise ratio of an EMRI from its parameters, and (ii) further accelerating detectability estimation with a neural network that learns the selection function, leveraging our first neural network for data generation. The resulting framework rapidly estimates the selection function, enabling a full treatment of EMRI detectability in population inference analyses. We apply our method to an astrophysically motivated EMRI population model, demonstrating the potential selection biases and subsequently correcting for them. Accounting for selection effects, we predict that with 116 EMRI detections *LISA* will measure the MBH mass function slope to a precision of 8.8 per cent, the CO mass function slope to a precision of 4.6 per cent, the width of the MBH spin magnitude distribution to a precision of 10 per cent, and the event rate to a precision of 12 per cent with EMRIs at redshifts below  $z = 6$ .

**Key words:** gravitational waves – methods: statistical – software: data analysis – stars: black holes – black hole mergers.

## 1 INTRODUCTION

The *Laser Interferometer Space Antenna* (*LISA*; Amaro-Seoane et al. 2017) is a planned space-based observatory capable of observing gravitational waves (GWs) with frequencies  $\sim 10^{-5}$ – $10^{-1}$  Hz. Promising sources of GWs in this band are extreme mass ratio inspiral (EMRI) systems, comprising a compact object (CO) orbiting, and gradually inspiralling towards, a massive black hole (MBH; Amaro-Seoane et al. 2007, 2023). EMRI systems typically complete  $\sim 10^4$ – $10^5$  orbital cycles in the *LISA* band and generate GWs with an intricate frequency evolution. Relativistic effects, including Lense–Thirring precession and pericentre precession, generate many distinct sideband modes in the signal spectrum (Barack & Cutler 2004; Hughes et al. 2021). The amplitude and phase evolution of these modes is strongly dependent on the parameters of the MBH–CO system, enabling precise measurements of these parameters (Babak et al. 2017). In particular, the (redshifted) MBH and CO masses, MBH spin, and orbital eccentricity may be determined to accuracies of  $\sim 10^{-3}$  per cent, and the sources localized in space to better than 10 per cent relative precision (Amaro-Seoane et al. 2007; Berry et al. 2019). This level of measurement precision for MBHs surpasses both existing electromagnetic techniques (Daly 2011) and *LISA* observations of MBH binaries (Klein et al. 2016).

The number of EMRIs that will be detected is uncertain, largely due to poorly constrained astrophysical parameters in current formation channel theories, but the detection rate is likely to be of order  $10^1$ – $10^3$  yr $^{-1}$  (Babak et al. 2017; Pan, Lyu & Yang 2021; Vázquez-Aceves et al. 2022; Amaro-Seoane et al. 2023). During *LISA*’s proposed 4 yr mission, we therefore expect to attain a sufficiently large catalogue of EMRIs (each of which providing excellent measurements of their parameters) to probe and resolve features of the MBH mass and spin populations. For example, we expect to match the precision of current observational estimates of an MBH mass power-law spectral index, even for pessimistic EMRI detection rate predictions (Gair, Tang & Volonteri 2010). Such a catalogue also enables the testing of the wide array of EMRI formation channel theories. Several EMRI formation channels have been suggested, including loss-cone scattering of COs into inspiral orbits (Alexander 2017), radial migration of COs by dynamical interaction with an accretion disc (Pan et al. 2021), capture via the Kozai–Lidov mechanism due to the presence of a binary MBH system (Naoz et al. 2022), and the tidal disruption of main-sequence or helium stars on highly eccentric orbital trajectories around MBHs (Bortolas & Mapelli 2019). For a given formation channel, the rate of EMRI production depends on astrophysical parameters (e.g. Broggi et al. 2022, for loss-cone or TDE channels) that are poorly constrained by existing observations. Determining the relative contributions of the formation channels to the overall EMRI rate therefore places constraints on these parameters; this could be estimated from a cata-

\* E-mail: [c.chapman-bird.1@research.gla.ac.uk](mailto:c.chapman-bird.1@research.gla.ac.uk)

logue of EMRI observations in a similar manner to how ground-based GW observations are used to constrain their source astrophysics (e.g. Zevin et al. 2021).

Extracting information about the astrophysical population requires a hierarchical inference where the parameters of each EMRI in the catalogue (and their associated uncertainties) are collectively used to constrain the parameters of a chosen population model. However, the catalogue only contains sources loud enough to cross a detection threshold, and these are generally not representative of the underlying population: if these selection effects are ignored, the inferred population parameters will be biased. EMRIs that are intrinsically fainter (depending on their masses, spins, or orbital parameters) and more distant are less likely to be detected than their nearer and louder counterparts. In practice, one may correct for this by determining the fraction of the proposed population that is detectable, and re-weighting the population likelihood accordingly (Mandel, Farr & Gair 2019; Alsing et al. 2023).

In the absence of a standard EMRI detection pipeline, the detectability of a given EMRI is typically assessed with respect to the signal-to-noise ratio (SNR) of the EMRI waveform (Gair et al. 2004, 2010; Babak et al. 2017). The detectable fraction of a proposed population (the selection function) may therefore be estimated by randomly drawing EMRI events from the population and computing their SNRs to determine the fraction of these samples that are detectable. These selection function estimates may then be used to re-weight the population likelihood and obtain unbiased inferences of the population parameters.

The function that maps EMRI parameters to the waveform SNR is complicated, so bias correction is computationally expensive. This high cost comes from both the generation of complex long-duration waveforms and the manipulation of these large data sets. Even exploiting graphics processing unit (GPU) acceleration and vectorization, each SNR takes of order 1 s to compute for a 4 yr data-stream duration. Using conservative estimates, if one draws  $10^5$  EMRIs for each candidate population, and a population inference sampling run consists of  $10^5$  candidate populations, full selection bias correction would require  $10^{10}$  GPU s. This is too costly for analyses including selection bias correction in this manner to be conducted in a reasonable amount of time.

Previous studies have addressed the issue of computational cost by indirectly approximating the behaviour of the selection function via: proxy distance thresholds (Laghi et al. 2021); a reduction in EMRI parameter space complexity, for instance by neglecting the dependence of eccentricity or inclination on detectability (Gair et al. 2010); a reduction of waveform complexity by only computing a small number of sideband modes with faster, less accurate waveform models (Chua & Cutler 2022). These approaches permit the rapid computation of the selection function, but do not account for more complex correlations between EMRI parameters and may introduce systematic biases due to the approximations made. For instance, the evolution of orbital eccentricity and inclination during an inspiral is correlated with mass ratio, and the mode amplitudes (and therefore the overall SNR of the waveform) are correlated with both eccentricity and inclination evolution. Therefore, even for parameters not directly of interest to a given population study, the correlations between these parameters and event detectability must still be taken into account to avoid biases in the results obtained.

In this work, we propose an alternative approach that leverages the speed of the recently developed GPU-accelerated EMRI waveform package FastEMRIWaveforms (FEW; Katz et al. 2021) and machine-learning techniques to interpolate the EMRI SNR function, directly correcting for an SNR-based selection bias without the need for

major systematic approximations or simplifications. In Section 2, we outline the Bayesian population inference framework employed, including the correction for selection biases. In Section 3, we introduce our approach for estimating the selection bias with machine learning, which we achieve by replacing the bottleneck in the selection bias calculation (the EMRI SNR function) with a neural network interpolator. The effectiveness of our framework is demonstrated in Section 4 for a typical EMRI population, presenting clearly the manifestation of the selection bias in the obtained results and how this is corrected for in practice; the corrected results provide an unbiased estimate for the precision to which *LISA* observations could constrain the astrophysical EMRI population. Finally, in Section 5, we perform a global posterior consistency check to verify the analysis.

The method we describe in this paper is implemented in our open-source code package *POPLAR* (Chapman-Bird 2023), which we plan to use for (and develop alongside) future EMRI population studies.

## 2 HIERARCHICAL BAYESIAN INFERENCE FRAMEWORK

Our goal is to infer the properties of an EMRI population model using a catalogue of many EMRI observations.

For each EMRI in the catalogue, information about its parameters  $\theta_i$  is encoded in the data,  $\mathbf{d}$ , where the subscript  $i \in [1, N_{\text{obs}}]$  identifies the particular EMRI in the catalogue of  $N_{\text{obs}}$  detections. The posterior distribution for  $\theta_i$  given the data is

$$p(\theta_i | \mathbf{d}) = \frac{\pi(\theta_i) \mathcal{L}(\mathbf{d} | \theta_i)}{\mathcal{Z}(\mathbf{d})}, \quad (1)$$

where  $\pi(\theta_i)$  is the prior distribution on  $\theta_i$ ,  $\mathcal{L}(\mathbf{d} | \theta_i)$  is the likelihood of observing the data given a set of source parameters, and  $\mathcal{Z}(\mathbf{d})$  is the evidence (marginalized likelihood). We estimate the parameters of the EMRI by stochastically sampling the posterior distribution, obtaining a set of posterior samples  $\{^k\theta_i\}$  (Christensen & Meyer 2022), where the superscript  $k \in [1, S_i]$  denotes each posterior sample for a given event. The posterior  $p(\theta_i | \mathbf{d})$  provides information about a single EMRI source; by combining together the properties of the catalogue of sources, we can constrain a population model.

The population model  $p_{\text{pop}}(\theta | \lambda)$  describes the astrophysical distribution of EMRI source parameters. It is described by a set of hyperparameters  $\lambda$  that determine the shape of the population, and a Poissonian mean event rate  $\mathcal{R}$  that parametrizes how often EMRIs occur. We use the *hyper* prefix to differentiate these population-level (hyper)parameters from the event-level EMRI parameters. By estimating the hyperparameters, we constrain the relative probabilities of different population shapes and event rates in accordance with the contents of the catalogue. We perform this hyperparameter estimation in a hierarchical Bayesian inference framework (Mandel et al. 2019).

To obtain an estimate of  $\lambda$  and  $\mathcal{R}$ , we sample the hyperparameter posterior distribution

$$p(\lambda, \mathcal{R} | \{\theta\}) = \frac{\pi(\lambda) \pi(\mathcal{R}) \mathcal{L}(\{\theta\} | \lambda) \mathcal{L}(\{\theta\} | \mathcal{R})}{\mathcal{Z}(\{\theta\})}, \quad (2)$$

where  $\pi(\lambda)$  and  $\pi(\mathcal{R})$  are hyperprior distributions, and  $\mathcal{Z}(\{\theta\})$  is the hyperevidence. The hyperparameter likelihood  $\mathcal{L}(\{\theta\} | \lambda)$  is defined as

$$\mathcal{L}(\{\theta\} | \lambda) = \prod_{i=1}^{N_{\text{obs}}} \frac{1}{S_i \alpha(\lambda)} \sum_{j=1}^{S_i} \frac{p_{\text{pop}}(^j\theta_i | \lambda)}{\pi(\theta_i)}, \quad (3)$$

where, in general, the population probability of each posterior sample must be re-weighted by the prior used in the EMRI parameter estimation step (Mandel et al. 2019). In our case, we adopt uniform

priors on all EMRI parameters and this re-weighting simplifies to a proportionality constant. The rate likelihood  $\mathcal{L}(\{\theta\}|\mathcal{R})$  is

$$\mathcal{L}(\{\theta\}|\mathcal{R}) = \exp[-\mathcal{R}\alpha(\lambda)] [\mathcal{R}\alpha(\lambda)]^{N_{\text{obs}}}. \quad (4)$$

Here, the selection function  $\alpha(\lambda)$  is a corrective factor applied to account for the presence of selection bias on the observations: of the (unknown) number of events that occurred, only a subset  $N_{\text{obs}}$  were detected. It may be written as

$$\alpha(\lambda) = \int p_{\text{det}}(\theta) p_{\text{pop}}(\theta|\lambda) d\theta, \quad (5)$$

for some detection probability  $p_{\text{det}}(\theta)$ , and represents the fraction of a population (described by a particular set of hyperparameters  $\lambda$ ) that is detectable. Performing an inference including the selection effects should produce results unbiased by detectability (Mandel et al. 2019).

To use this inference framework, one must specify:

- (i) A method for obtaining  $\{\theta\}$ ;
- (ii) The form of  $p_{\text{pop}}(\theta|\lambda)$ ;
- (iii) The selection effects for the observations.

Once these three ingredients have been formally defined, we can construct (and ultimately, sample) the hyperposterior, equation (2). We introduce the approach we use to obtain posterior samples for each detected EMRI in Section 2.1; we detail the form of our population model in Section 2.2, and we outline our treatment of detectability and the modelling of selection effects in Section 2.3.

## 2.1 EMRI parameter estimation

Despite recent reductions in EMRI waveform computation time to the sub-second level (Katz et al. 2021), standard Bayesian parameter estimation techniques are too costly for the event posteriors to be sampled directly en-masse as is required in population studies. We instead opt to approximate the EMRI likelihood (and by extension, posterior) with a Fisher matrix approach, operating under the linear signal approximation (LSA; Cutler & Flanagan 1994), in which the likelihood is approximated by a multivariate normal distribution,

$$p(d|\theta) \approx \mathcal{N}(\theta, \Gamma^{-1}), \quad (6)$$

where the covariance matrix of the distribution is the inverse of the Fisher information matrix (FIM)  $\Gamma$  of the EMRI waveform. The LSA is only valid in the high-SNR limit (which may lie far above the detection threshold), which should be verified before it is used to approximate likelihoods (Vallisneri 2008). The FIM is given by

$$\Gamma_{\ell m} = \langle \partial_\ell h | \partial_m h \rangle, \quad (7)$$

where  $\partial_\ell h$  refers to the derivative of the waveform strain  $h(t)$  with respect to the  $\ell$ -th parameter of  $\theta$ , evaluated at  $\theta$ . The noise-weighted inner product is defined as

$$\langle x | y \rangle = 4\Re \left[ \int_0^\infty \frac{\tilde{x}^*(f)\tilde{y}(f)}{S_n(f)} df \right], \quad (8)$$

where  $\tilde{x}(f)$  is the Fourier transform of a time-domain strain  $x(t)$ ,  $\Re$  refers to the real part and  $S_n(f)$  is the one-sided power spectral density (PSD) of the detector (Maggiore 2008). We adopt the analytic fit to the *LISA* PSD derived in Robson, Cornish & Liu (2019). Using the FIM, we can rapidly produce posterior distributions for a catalogue of EMRIs.

Specifics of our EMRI event catalogue simulation pipeline, including both waveform generation and FIM computation, are discussed in Appendix A.

**Table 1.** The functions, free parameters, and limits of the sub-population distributions  $p_x(x|\lambda_x)$ , the product of which is the EMRI population chosen. The hyperparameters  $\lambda_x$  are estimated via population inference. The upper limit for  $t_{\text{plunge}}$  is reduced to 2 yr for our validation analysis (Section 5).

$x$	$p_x(x \lambda_x)$	$\lambda_x$	$[x_{\text{min}}, x_{\text{max}}]$
$M$	Power law	$\lambda_M, M_{\text{min}}, M_{\text{max}}$	$[M_{\text{min}}, M_{\text{max}}]$
$\mu$	Power law	$\lambda_\mu, \mu_{\text{min}}, \mu_{\text{max}}$	$[\mu_{\text{min}}, \mu_{\text{max}}]$
$a$	Trunc. Normal	$\mu_a, \sigma_a^2$	$[0.001, 0.999]$
$e_0$	Uniform	—	$[0.1, 0.5]$
$\cos i_0$	Uniform	—	$[0, \pi/3]$
$\sin \theta_S$	Uniform	—	$[0, \pi]$
$\sin \theta_K$	Uniform	—	$[0, \pi]$
$\Delta\phi$	Uniform	—	$[0, 2\pi]$
$t_{\text{plunge}}$	Uniform	—	$[0, (2, 10)] \text{ yr}$
$z$	$p_z(z)$	—	$[0, 6]$

## 2.2 Population model

For simplicity, we choose a population model that is a product of independent univariate subpopulations, such that

$$p_{\text{pop}}(\theta|\lambda) = \prod_{x \in \theta} p_x(x|\lambda_x), \quad (9)$$

where  $x$  denote EMRI parameters and  $\lambda_x$  the corresponding hyperparameters that describe the shape of the subpopulation. The mathematical form of these subpopulations is summarized in Table 1, and in detail:

(i) Mass functions for both MBHs and stellar-mass black holes are well approximated by power laws, albeit with additional substructure present when examined in detail (Shankar 2013; Abbott et al. 2023). We therefore model the MBH and CO mass distributions  $p_M(M|\lambda_M, M_{\text{min}}, M_{\text{max}})$  and  $p_\mu(\mu|\lambda_\mu, \mu_{\text{min}}, \mu_{\text{max}})$  as power laws, with index  $\lambda_x$  and limits  $[x_{\text{min}}, x_{\text{max}}]$ , which have the form

$$p_x(x|\lambda_x, x_{\text{min}}, x_{\text{max}}) = \frac{1 + \lambda_x}{x_{\text{max}}^{1+\lambda_x} - x_{\text{min}}^{1+\lambda_x}} x^{\lambda_x}. \quad (10)$$

(ii) The form of the MBH spin magnitude distribution  $p_a(a|\mu_a, \sigma_a)$  is dependent on a number of astrophysical processes during the formation and evolution of MBHs and their host galaxies (Volonteri 2010; Sesana et al. 2014). Incorporating these into our population model and characterizing their impact on inference results is beyond the scope of this study. For simplicity, we instead choose a truncated normal distribution with mean  $\mu_a$  and variance  $\sigma_a^2$  as has been done in previous analyses of the stellar-mass binary black hole (BBH) mergers (Roulet & Zaldarriaga 2019; Miller, Callister & Farr 2020). This is written as

$$p_x(x|\mu_x, \sigma_x) = \frac{1}{\sigma} \frac{\psi[(x - \mu_x)/\sigma_x]}{\Psi[(B - \mu_x)/\sigma_x] - \Psi[(A - \mu_x)/\sigma_x]}, \quad (11)$$

where  $\psi(x)$  and  $\Psi(x)$  are the probability density and cumulative distribution functions (CDF) of the standard normal distribution, respectively. The limits  $[A, B]$  are chosen to be  $[0.001, 0.999]$  as waveform generation is unstable at extremal spins beyond these limits.

(iii) High initial orbital eccentricities ( $>0.99$ ) are expected for EMRIs formed by relaxation mechanisms, but significant orbital eccentricity will be lost before the GW emission of the system enters the *LISA* band (Peters & Mathews 1963), broadening the distribution and shifting it to lower eccentricities (Amaro-Seoane 2020). To reflect this behaviour, we choose a uniform eccentricity distribution  $e_0 \in [0.1, 0.5]$ , with upper limit chosen to reflect that the waveform model is a series expansion in eccentricity and should

therefore not be trusted for high eccentricities (Fujita & Shibata 2020; Isayama et al. 2022). The waveform model also consists of a system of ordinary differential equations (ODEs) that must be solved (Katz et al. 2021). The lower limit of the eccentricity distribution is chosen due to increasing stiffness in this ODE system at lower eccentricities leading to high computational cost (Burden & Faires 1993). We do not anticipate the validity of our approach to be affected by this lower eccentricity cut-off.

(iv) Orbital inclination  $i_0$  is similarly truncated due to ODE stiffness issues, but is otherwise distributed uniformly on the unit sphere along with other angular parameters.

(v) We choose a redshift distribution that is uniform in comoving volume and in comoving time (Hogg 1999); this has the form

$$p_z(z) \propto \frac{1}{(1+z)E(z)} \left( \int_0^z \frac{dz'}{E(z')} \right)^2, \quad (12)$$

where

$$E(z) = \sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}, \quad (13)$$

and we assume a standard cosmology with  $\Omega_M = 0.3$  and  $\Omega_\Lambda = 1 - \Omega_M = 0.7$ . The upper redshift limit for this distribution is chosen to be  $z = 6$  such that the detectable region of parameter space is not significantly truncated (otherwise, selection effects will be artificially suppressed). Increasing the redshift limit leads to high computational costs as the event rate  $\mathcal{R}$  must also be increased accordingly, as the event rate density has remained constant but the comoving volume over which we are distributing events has grown.

The chosen form of these subpopulations is motivated primarily by computational simplicity. However, our approach is flexible and can be applied to the hierarchical inference of any population model.

### 2.3 Selection effects

In the absence of a specific EMRI search pipeline, we model the detection probability as a binary SNR threshold, as is typical for EMRI studies (Gair et al. 2010; Babak et al. 2017; Bonetti & Sesana 2020). This may be written as

$$p_{\text{det}}(\theta) = \mathcal{H}(\rho_n - \rho_t), \quad (14)$$

where  $\rho_n$  is a (noise-realized) SNR,  $\mathcal{H}(x)$  is the Heaviside step function, and  $\rho_t$  is a chosen threshold SNR. We obtain  $\rho_n^2$  by drawing a sample from a non-central  $\chi^2$  distribution with two degrees of freedom and non-centrality parameter  $\rho_{\text{opt}}^2$  (Maggiore 2008),

$$p(\rho_n^2 | \rho_{\text{opt}}^2) = \frac{1}{2} \exp\left(-\frac{\rho_n^2 + \rho_{\text{opt}}^2}{2}\right) I_0(\rho_n \rho_{\text{opt}}), \quad (15)$$

where  $I_0(x)$  is a modified Bessel function of the first kind (Abramowitz & Stegun 1964), and  $\rho_{\text{opt}}^2$  is the square of the optimal matched-filter SNR

$$\rho_{\text{opt}}^2 = \langle h | h \rangle. \quad (16)$$

We assume that  $\rho_n$  is the positive square root of  $\rho_n^2$ ; while noise fluctuations can lead to negative values, this is not expected for large values around our detection threshold. One may analytically compute the mean detection probability in equation (14) over all noise realizations by directly computing the non-central  $\chi^2$  CDF  $p(\rho_n^2 > \rho_t^2 | \rho_{\text{opt}}^2)$ , such that

$$\overline{p}_{\text{det}}(\theta) = 1 - p(\rho_n^2 > \rho_t^2 | \rho_{\text{opt}}^2), \quad (17)$$

where an overline denotes the mean.

We approximate the selection function equation (5) by evaluating the Monte Carlo sum

$$\alpha(\lambda) \approx \frac{1}{N_t} \sum_{k=0}^{N_t} \overline{p}_{\text{det}}(\theta_k), \quad (18)$$

where  $\{\theta_k\}$  are sampled from  $p_{\text{pop}}(\theta | \lambda)$ . As the variance on this approximation scales inversely with  $N_t$ , one must compute  $\overline{p}_{\text{det}}(\theta)$  (and therefore  $\rho_{\text{opt}}$ ) of the order of  $10^5$  times for each computation of  $\alpha(\lambda)$  to achieve per cent-level accuracy; even with parallelization, this would be prohibitively expensive with typical computing resources (taking of the order of minutes) for use in a typical sampling run, in which  $\alpha(\lambda)$  must be computed once per hyperlikelihood call. We address this problem by replacing the SNR function with an accurate and rapid interpolator, allowing for equation (18) to be evaluated in parallel at a sufficiently low computational cost to be practical for use in inference problems.

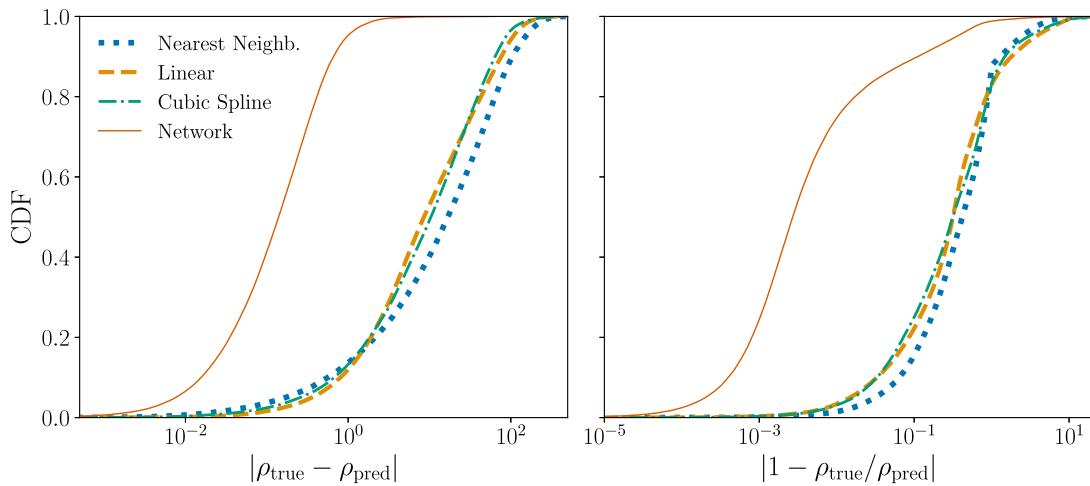
## 3 INTERPOLATING OVER SNR

The principal requirements for our SNR interpolator are that it must be accurate and unbiased across the EMRI parameter space: inaccuracies may bias the results of our population inference. It must also be sufficiently fast as to not bottleneck the sampling process, capable of estimating SNRs for  $10^5$  sets of EMRI parameters in  $< 1$  s. These constraints are particularly challenging to meet due to the high dimensionality of the EMRI parameter space (13 dimensions, as defined in Appendix A).

Fortunately, we can reduce the number of parameters that we need to interpolate over by considering how the SNR of an EMRI waveform depends on each parameter. We can ignore the orbital phase parameters ( $\Phi_r$ ,  $\Phi_\theta$ ,  $\Phi_\phi$ ) due to their negligible correlation with SNR, as the initial phase becomes relatively unimportant for an inspiral with  $\sim 10^4$  orbital cycles. Additionally, as SNR scales inversely with luminosity distance  $d_L$ , we may further reduce the dimensionality of the parameter space by fixing  $d_L$  in training data and applying this scaling post-interpolation: we use  $d_L = 1$  Gpc for convenience.

Despite eliminating four dimensions of the EMRI parameter space, we are still in a regime where standard interpolation schemes are ineffective. As a representative example, we consider spline interpolation schemes with piecewise polynomials of zeroth, first, and third order: these are more commonly known as nearest neighbour, linear, and cubic spline interpolation, respectively (Piegl & Tiller 1987). Our requirement for a fast interpolator prevents us from interpolating over points randomly distributed in the parameter space, as the complexity of the algorithms used for this scales quadratically with the number of basis points (Barber, Dobkin & Huhdanpaa 2013) and the computational cost of these methods quickly becomes impractical. Instead, we may use grid-based versions of these techniques. However, these methods suffer from the curse of dimensionality: the Euclidean distances between neighbouring grid vertices grows as the dimensionality of the space increases, which leads to poor interpolation accuracy.

To demonstrate the unsuitability of linear interpolation in practice, we generate  $\rho_{\text{opt}}$  on a regular grid with  $10^6$  total grid points and construct the aforementioned spline interpolators with this grid as a basis. We then compute a testing set of  $10^6$  SNRs from randomly sampled sets of EMRI parameters and compare the interpolator output at these points by calculating the discrepancies between the true and predicted SNRs, denoted  $\rho_{\text{true}}$  and  $\rho_{\text{pred}}$ , respectively. The cumulative distribution of the (i) absolute and (ii) relative differences



**Figure 1.** Cumulative distribution functions for the (left) absolute and (right) fractional differences between the SNR predictions of (linear, nearest neighbour, cubic spline, and neural network) interpolation methods and the true values. The former three grid-based interpolation methods perform poorly, with at least 47 per cent of SNRs inaccurate by at least 10, and at least 75 per cent predicted to less than 10 per cent accuracy. Conversely, the latter method precisely estimates SNRs: 95.3 per cent of SNRs are determined to within 1, and 89.6 per cent within 10 per cent of the true values.

between prediction and truth for the three spline interpolation schemes we consider are shown in Fig. 1. As expected, the grid-based interpolator performance is poor regardless of the order of the piecewise polynomial used. While linear or spline interpolation offers marginal improvement over nearest neighbour interpolation, the low spatial resolution of the grid limits the improvements. The majority of the interpolated SNRs are inaccurate by at least 50 per cent, with absolute errors typically exceeding 10 (or even as high as 100 in extreme cases). As we will demonstrate in Section 5, this performance is inadequate for unbiased population inference.

### 3.1 Interpolation with neural networks

Neural networks are highly flexible mathematical tools that are capable of learning complex relationships in high-dimensional spaces (Goodfellow, Bengio & Courville 2016). For our purposes, we need a neural network that takes a vector as an input (the EMRI parameters) and produces a scalar output (the SNR estimate). We opt for the multilayer perceptron (MLP) algorithm (Hastie, Tibshirani & Friedman 2001) as it fits this specification. MLPs are fast and capable of high accuracy, satisfying our requirements well. The design and training of this MLP are discussed in Appendix B. The trained network achieves two orders of magnitude of improvement in accuracy compared to other interpolation approaches, as shown in Fig. 1; the majority of the test data are predicted to per cent-level accuracy. This network is capable of producing  $10^5$  SNR estimates in  $<0.1$  s, which is six orders of magnitude faster than calculating the SNR directly.

As the MLP was trained with an L1 loss function, which minimizes the absolute difference between the prediction and truth (Goodfellow et al. 2016), it does not perform as well in terms of fractional error for  $\rho_{\text{opt}} \ll 1$ . This manifests as a larger upper tail in the relative CDF. However, this does not translate to a reduction in performance, as these signals are too weak to be detectable across the majority of the luminosity distance distribution. If adequate performance across all SNR scales is required, this may be achieved with the appropriate choice of loss function, for instance by training on the log of the SNR. By choosing not to train on log SNRs, we prioritize the regions of parameter space corresponding to larger SNRs in the data set

(at the fiducial luminosity distance of 1 Gpc). As these SNRs will be pushed towards the detection threshold at larger distances, and the majority of our luminosity distance distribution is above 1 Gpc, estimating these larger SNRs well has the greatest impact on accurate detectability estimates.

### 3.2 Interpolating over the selection function

Using our trained interpolator network to produce SNR estimates, we are now able to approximate  $\alpha(\lambda)$  via equation (18) with the interpolated SNR as an input. However, two issues still remain that must be addressed before this may be applied in practice. First, this set-up still requires that the parameter set  $\{\theta\}$  is drawn from  $p_{\text{pop}}(\theta|\lambda)$  for each hyperlikelihood call; this is a slow operation, even for the relatively simple population models in Table 1. It is also inefficient, because  $\alpha(\lambda)$  will typically not vary significantly across the high-probability region of the posterior, where the majority of samples are drawn. Second, the stochastic nature of the Monte Carlo selection function estimates itself presents challenges in sampling: the hyperlikelihood surface becomes noisy, which can be problematic for the reliable convergence of sampling algorithms. These issues prohibit the use of our stochastic selection function estimates in hyperposterior sampling and must be solved.

To address these problems and further accelerate our bias-corrected likelihood, we extend the idea of interpolating over high-dimensional spaces with MLPs further. A second MLP trained prior to sampling can be used to interpolate directly over  $\alpha(\lambda)$ . The architecture and training settings for this MLP are discussed in Appendix B. Using this second MLP step greatly reduces the time per likelihood call, achieving a further order of magnitude of speedup with respect to stochastic estimation of the selection function. For our chosen population model, the resulting computational cost of the numerator and denominator in equation (2) become roughly equivalent. Further speedup is achieved with vectorized evaluation of the hyperlikelihood, as the selection function MLP is capable of handling many sets of hyperparameters at no additional cost provided that sufficient GPU memory is available. This vectorization would not be practical if one were to estimate the selection function values with the stochastic approach. With the main limitations of our method

addressed, we are now able to produce selection function estimates usable in sampling at low computational cost.

## 4 UNBIASED POPULATION INFERENCE RESULTS

With our strategy for selection bias correction formulated, we are now properly equipped to tackle an EMRI population inference problem. We begin by simulating a catalogue of EMRI observations. To estimate how well *LISA* will resolve the shape of the EMRI population in a realistic scenario, we choose hyperparameter values supported by recent black hole population studies:

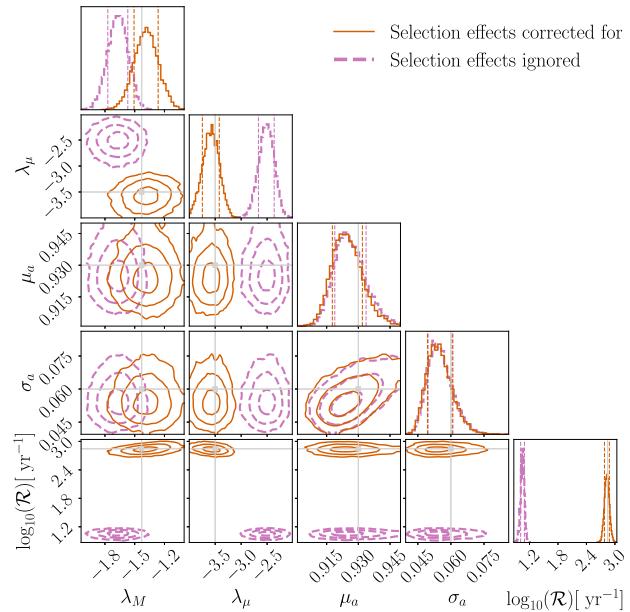
(i) While constraints have been placed on the slope of the MBH mass function by current observations (Shankar 2013), the mass function for MBHs hosting EMRIs is subject to additional selection effects that are poorly understood at present (Babak *et al.* 2017). Recent work has estimated that the spectral index of this power law after the inclusion of selection effects is  $\lambda_M \approx -1.43$  in the mass range  $[M_{\min}, M_{\max}] = [10^5, 10^7] M_\odot$  (Babak *et al.* 2017). We assume that the slope of the CO mass function is equal to the median value observed in stellar-mass BBH mergers of  $\lambda_\mu \approx -3.50$  in the mass range  $[3, 90] M_\odot$  (Abbott *et al.* 2023), assuming that the progenitors of these mergers are representative of the universal stellar-mass BH population. The selection effects that translate this into the EMRI CO mass function are poorly understood and would require a dedicated set of *N*-body simulations of stellar cusps to properly quantify (Babak *et al.* 2017; Broggi *et al.* 2022), so we do not consider them in this study. Despite this caveat, the slope observed via BBH mergers is the strongest constraint placed on the mass function for black holes in this mass range available and is a reasonable starting point for estimating *LISA*'s ability to resolve the CO mass function with EMRI observations.

(ii) The MBH spin magnitude distribution is also poorly constrained by observational data. Current measurements are limited to MBHs in active galactic nuclei (Daly 2011), which may not be representative of the full MBH spin magnitude population as different formation channels will yield different MBH spin magnitude distributions (Amaro-Seoane *et al.* 2023). Self-consistent simulations of MBH growth with cosmic evolution predict that most MBHs have spins greater than 0.9 in the MBH mass range quoted above, with a fairly narrow spread below  $10^7 M_\odot$  (Sesana *et al.* 2014). We include this characteristic shape of a narrow spin distribution above  $a = 0.9$  in our population by choosing  $[\mu_a, \sigma_a] = [0.93, 0.06]$ .

(iii) The time that each EMRI plunges with respect to the start of the observational data  $t_{\text{plunge}}$  is randomly distributed in the range  $[0, 10]$  yr. In line with the planned *LISA* mission duration, we assume a 4-yr observational window (Amaro-Seoane *et al.* 2017): some EMRIs will not plunge until after the end of our observational data, but may still be detectable if they are bright enough. We assume that EMRIs occur at a rate of  $\mathcal{R} = 700 \text{ yr}^{-1}$ , which is conservative (considering our redshift cut-off of  $z = 6$ ) when compared with EMRI rate estimates from astrophysical modelling (Babak *et al.* 2017; Broggi *et al.* 2022; Vázquez-Aceves *et al.* 2022).

After discarding the signals too faint to be detected, we obtain a catalogue of 116 EMRIs.

To demonstrate the selection biases present, we perform two sampling runs: one in which selection biases are corrected for with our interpolation scheme, and another in which selection effects are not accounted for, i.e. replacing  $\alpha(\lambda)$  with 1 in equation (3) and equation (4). We sample the hyperposterior equation (2) with the *nessai* nested sampler (Williams 2021; Williams, Veitch &



**Figure 2.** Recovered hyperposterior distributions for our example scenario with selection effects corrected for (solid) or ignored (dashed). The true values of each hyperparameter are indicated by the cross-hairs. Significant bias is present in the recovery of  $\lambda_\mu$ , with minor bias also visible in the recovery of  $\lambda_M$ . The rate  $\mathcal{R}$  is accurately recovered with the inclusion of selection effect correction.

Messenger 2021), using default settings. The convergence of all sampling runs with these settings was verified by examining the results of internal consistency checks built into *nessai*. The hyperposterior distributions obtained from these sampling runs for a subset of hyperparameters are shown in Fig. 2. The full hyperposterior is shown in Appendix C, which demonstrates minor discrepancies between the two hyperposterior distributions for the other hyperparameters, with marginal posteriors that are too narrow (overconstrained) but otherwise fairly consistent with the set hyperparameter values.

The bias that results from ignoring selection effects on the observations is visible here as a discrepancy between the credible interval contours of the two hyperposterior distributions at the 99 per cent level. For  $\lambda_M$ , the uncorrected posterior is inconsistent with the true value at the 99 per cent credible level; the marginal posteriors for  $\lambda_\mu$  that include or exclude selection bias correction disagree to an even greater extent. There is also a clear difference between the marginal posteriors on  $\mathcal{R}$ : this is symptomatic of the presence of selection effects, as it indicates that predicting the overall event rate solely from the size of the detection catalogue will result in a significant underestimate of the actual event rate.

After accounting for selection effects, our results serve to probe how well *LISA* can resolve the form of this EMRI population. We estimate (quoting the median and the 90 per cent credible interval) that  $\lambda_M = -1.39^{+0.12}_{-0.12}$  and  $\lambda_\mu = -3.58^{+0.16}_{-0.17}$ , corresponding to precisions of 8.8 per cent and 4.6 per cent, respectively. The MBH spin distribution is well recovered, with  $\mu_a = 0.924^{+0.008}_{-0.007}$  and  $\sigma_a = 0.054^{+0.006}_{-0.005}$ ; these hyperparameters are recovered to within 0.87 per cent and 10 per cent, respectively. The EMRI rate is estimated with 12 per cent precision to be  $\mathcal{R} = 678^{+88}_{-75} \text{ yr}^{-1}$ . The precision achievable by *LISA* will roughly scale with the square root of the true event rate, which we have assumed a conservative value for in this study: for the most optimistic scenarios, this could improve by as much as an order of magnitude (Babak *et al.* 2017).

**Table 2.** Hyperprior distributions chosen for all sampling runs. As the range of plunge times is reduced by a factor of 5 for our probability–probability plot analysis (Section 5), our prior bounds on the EMRI rate are adjusted accordingly: this is indicated by (\*).

Parameter	Distribution	Limits
$\lambda_M$	Uniform	[-4, -1]
$M_{\min}$	Uniform	$[5, 50] \times 10^4 M_\odot$
$M_{\max}$	Uniform	$[5, 50] \times 10^6 M_\odot$
$\lambda_\mu$	Uniform	[-4, 1]
$\mu_{\min}$	Uniform	$[1, 5] M_\odot$
$\mu_{\max}$	Uniform	$[80, 100] M_\odot$
$\mu_a$	Uniform	[0.05, 0.95]
$\sigma_a$	Uniform	$[10^{-3}, 2]$
$\mathcal{R}$	Log-uniform	$[350, 1050] ([75, 150]^*) \text{ yr}^{-1}$

The number of detected events depends on the underlying population. For example, if the number of events is skewed to high redshift with respect to our assumed distribution of uniform in comoving volume and comoving time, then the number of detected events will decrease accordingly. Our choice of redshift distribution is equivalent to assuming that the probability of an EMRI occurring for a given MBH is constant across cosmic time. In reality, we expect that the physics of EMRI formation, such as cusp erosion (Babak et al. 2017), will lead to deviations away from this. Similarly, we expect that the distributions of MBH and CO masses will differ in reality from our assumptions. Hence, the results presented here should only be considered illustrative. A comprehensive study of how population inference results vary with the underlying population (which would require a computationally efficient method, such as ours) is necessary to fully map out how well *LISA* could measure the EMRI source population.

## 5 VERIFYING THE ACCURACY OF RESULTS

In the previous section, we demonstrated the capability of our approach for a single example. However, this is not sufficient to ensure that the selection function estimates output by our MLP are sufficiently accurate and unbiased that population inference will return hyperposterior that are consistent with the truth across the hyperparameter space.

To assess whether this is the case, we opt for the probability–probability (P–P) plot test (Cook, Gelman & Rubin 2006). First, we draw  $N$  sets of hyperparameters from the hyperpriors described in Table 2, and generate the corresponding population catalogues. We then perform hyperposterior sampling runs to produce estimates of the hyperparameters in each case, and determine the confidence interval  $q_\lambda$  of the true hyperparameters with respect to the posterior obtained. Last, we plot the CDF of  $q_\lambda$ . When the trial sets of hyperparameters are drawn from the hyperprior, we expect that the true value of a hyperparameter will fall within the  $x$  per cent credible interval in  $x$  per cent of realizations (i.e. a plot of  $q_\lambda$  against its CDF will be diagonal) if our hyperposterior are consistent with the true values in all cases. We test the accuracy of our inference framework by comparing the calculated CDF with the expected diagonal trend.

Some variation of each CDF from the diagonal due to small-number statistics is expected. For the P–P plot to be meaningful  $N$  needs to be large, so we modify our population to reduce the computational cost of waveform generation by reducing the length of the observational window from 4 to 2 yr and the range of EMRI plunge times from [0, 10] to [0, 2] yr. To further reduce the cost of generating each population, we lower the overall event rate by

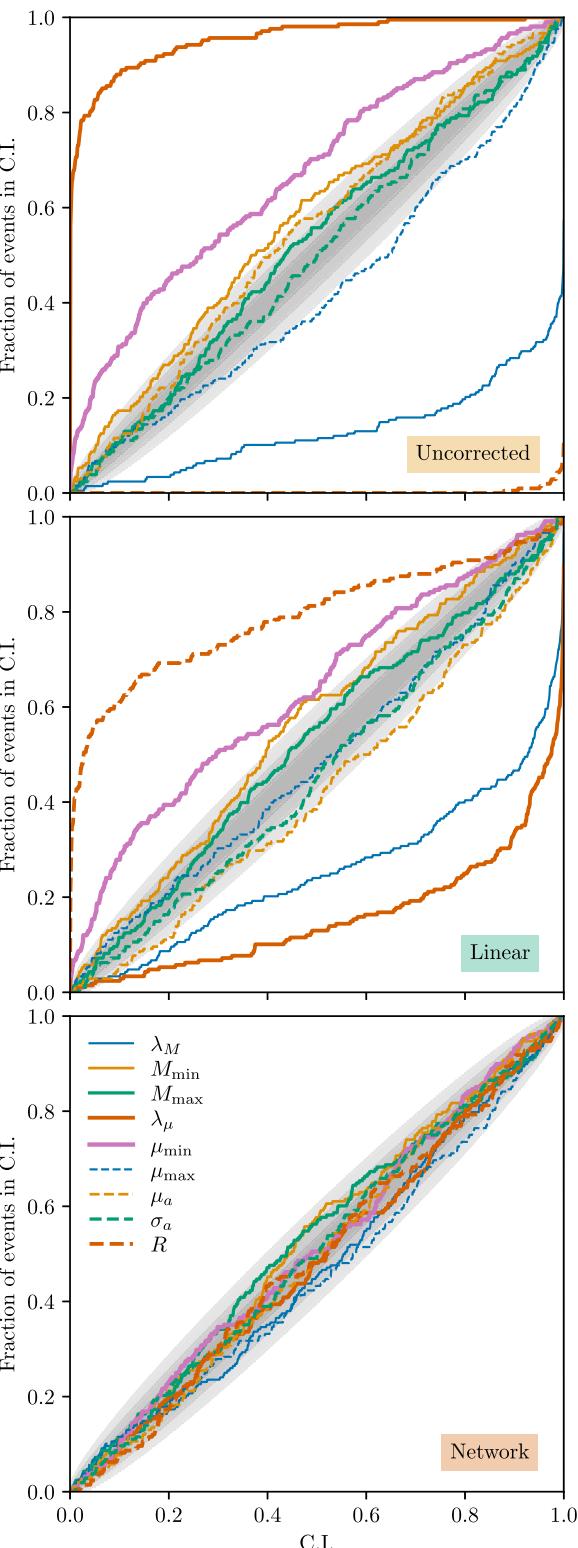
limiting our population to a maximum redshift of  $z = 1$ . Adjusting the event rate to account for these reductions in both duration and sensitive volume, the number of expected EMRIs for each population decreases by a factor of 35. The simulation and analysis configuration remains otherwise unchanged from the analysis described in Section 4. We perform three analyses to compare their results: first, we exclude selection effect correction; second, we include selection effect correction by means of a linear interpolation scheme, and last, we include selection effect correction with our neural network interpolation scheme. The P–P plots obtained from these analyses are shown in Fig. 3, broken down by hyperparameter and compared to the expected 68 per cent, 90 per cent, and 99 per cent deviations for the  $N = 208$  sets of drawn hyperparameters (Ibe 2013).

We first examine the P–P plot for the hyperposterior obtained when selection effects are excluded, setting  $\alpha(\lambda) = 1$ , as shown in the top panel of Fig. 3. The resulting posteriors exhibit strong biases with respect to the true values, and the P–P test fails; the majority of the hyperparameter CDFs deviate beyond the 99 per cent confidence interval.

The P–P plot is a useful diagnostic for how the posterior deviates from what is expected on a dimension-by-dimension basis. We can identify that  $\lambda_\mu$ ,  $m_{\min}$ ,  $\mu_{\min}$ , and  $\mu_a$  are typically overestimated when selection effects are ignored, and conversely that  $\mu_{\max}$  and  $\lambda_M$  are underestimated. This can be understood by considering how the SNR of an EMRI waveform is correlated with the parameters associated with these hyperparameters. The SNR is positively correlated with  $\mu$ , so the power-law slope is flattened off by the suppression of low CO mass events; similarly, SNR is higher for larger MBH spins, so the mean of the truncated Gaussian shifts to the right due to the suppression of the lower spin events. For  $\lambda_M$ , the opposite is true: the number of detectable higher MBH mass EMRIs (which are at lower frequencies) is suppressed due to the shape of the *LISA* sensitivity curve and so the observed power-law steepens. The parameters that pass the P–P test,  $M_{\max}$  and  $\sigma_a$ , are not strongly influenced by selection bias. The mean of the MBH spin truncated Gaussian may shift, but the change in the width of the distribution will be proportionally smaller, and is therefore less sensitive to this selection effect. Likewise, as the high MBH mass EMRIs are typically unobserved and occupy a small fraction of the overall EMRI population, small changes to the maximum MBH mass do not strongly affect the detectable fraction of the population. This is not the case for the CO mass distribution: the high mass events in the upper end of the power law are also the brightest events in the population, so adjusting the upper limit of the power law leads to larger changes in the fraction of events expected to be detectable. As the observed deviations from consistent hyperposterior align with our expectations, we are confident they are the result of selection effects.

To demonstrate the impact of the low interpolation accuracy seen in Fig. 1, on the resulting selection function estimates, we repeat this analysis with an MLP selection function trained on SNR estimates produced by the linear interpolator described in Section 3. The resulting P–P plot, shown in the middle panel of Fig. 3, demonstrates that although modest correction is achieved in this case, it is still far too biased and inaccurate to result in consistent hyperposterior. In the case of  $\lambda_\mu$ , this even results in an overcorrection of the selection bias when compared to the uncorrected case. The danger of overcorrection implies that it is not sufficient to include a selection function term in the population likelihood: the selection function must also be accurately calculated to obtain good results.

Finally, we present the P–P plot obtained for this analysis with our MLP selection function estimator in the bottom panel of Fig. 3. In



**Figure 3.** Probability–probability plots showing the fraction of hyperparameters within a given credible interval (CI) for  $N = 208$  hyperprior draws. Leaving selection biases uncorrected, setting  $\alpha(\lambda) = 1$ , demonstrates the presence of significant selection biases in the population inference (top). These biases are successfully rectified with our approach (bottom), whereas standard interpolation techniques fail to produce unbiased results (middle). The expected 68 per cent, 90 per cent, and 99 per cent confidence intervals are shown in grey.

stark contrast to the previous two plots, the hyperparameter CDFs are fully consistent with the expected confidence intervals. This conclusion is supported by the results of Kolmogorov–Smirnov (Dodge 2008) tests for each hyperparameter. Combining the  $p$ -values from each test with Fisher’s method (Mosteller & Fisher 1948) returned a combined  $p$ -value of 0.3. This indicates that all hyperparameters (including the EMRI rate  $\mathcal{R}$ ) are consistently recovered: the MLP is capable of producing selection function estimates that are sufficiently accurate for consistent posteriors to be obtained. This result verifies the application of our method in the treatment of selection biases in population inference.

## 6 CONCLUSIONS

Population inference with EMRIs has the potential to probe the evolution of both MBHs and their galactic neighbourhoods to unprecedented precision. However, the computational cost of components of this analysis is prohibitively high. Estimating selection biases in EMRI populations is computationally expensive due to a combination of the need for costly waveform models and the resources required to perform SNR calculations for long-duration data. As the SNR calculation is the computational bottleneck, we substitute it for an interpolation over pre-computed SNRs. We find that commonly employed interpolation schemes are not sufficiently fast or accurate for this problem, so we instead use machine-learning techniques. Using a neural network trained on a data set of SNRs distributed uniformly in the EMRI parameter space, we achieve a speedup of six orders of magnitude over direct SNR evaluation. We then replace the stochastic estimation of the selection function obtained via Monte Carlo integration with a second neural network that is trained on these stochastic estimates distributed uniformly in the hyperparameter space. This further improves the speed of hyperlikelihood evaluation by an order of magnitude, and enables vectorized estimation of the selection function for further reductions in computational cost. To verify the robustness of our approach against systematic biases, we globally evaluate hyperposterior consistency by simulating 208 EMRI populations and checking the results of selection bias-corrected hyperparameter estimation with a P–P test. This test confirmed that (i) the presence of selection effects significantly biased inferences that did not correct for them appropriately, and (ii) our approach successfully corrected for selection effects to produce unbiased results.

We apply our method to the inference of an astrophysically motivated EMRI population (assuming sources distributed uniformly in comoving volume and time) to study *LISA*’s ability to probe the structure of such populations. We estimate that  $\lambda_M = -1.39^{+0.12}_{-0.12}$  (a precision of 8.8 per cent) and that  $\lambda_\mu = -3.58^{+0.16}_{-0.17}$  (a precision of 4.6 per cent). For the MBH spin magnitude distribution, we find that  $\mu_a = 0.924^{+0.008}_{-0.007}$  and that  $\sigma_a = 0.054^{+0.006}_{-0.005}$ , resolving the width of the MBH spin magnitude distribution to within 10 per cent. The event rate is estimated to be  $\mathcal{R} = 678^{+88}_{-75} \text{ yr}^{-1}$  (a precision of 12.0 per cent).

The capability of our approach for treating selection effects in the case of a simple population model, which excludes substructure or correlations due to astrophysical effects, paves the way for future work to investigate more complex EMRI population models. The EMRI population we expect to be present in reality is multifaceted (Babak *et al.* 2017), and by introducing these features systematically to the population inference problem we can begin to characterize their measurability with space-based detectors such as *LISA*. As population inference is intimately tied to other hierarchical inference problems, including cosmological inference (MacLeod & Hogan

2008; Laghi et al. 2021) and tests of general relativity (Chua et al. 2018), proper treatment of selection effects for EMRIs has direct implications for these analyses as well. Ultimately, a joint hierarchical inference over this problem space may be required, of which our approach can be an integral part.

As our method is not predicated on a particular population model, it is applicable to a wide variety of population inference problems. Similarly, as any waveform model may be used, this approach is capable of accommodating future changes to EMRI waveform models with little tuning required. While our approach specifically targets the EMRI population inference problem, it may be generalized to any problem with an SNR-threshold selection bias due to this model-agnostic nature. The reduction in computational cost achieved by employing our method will be most pronounced in cases where the SNR function is expensive to compute and of high dimensionality, but the ability to form a vectorized SNR approximant will still offer a notable speedup for waveform models that are not easily parallelizable or vectorizable (e.g. due to memory constraints).

Our code package `POPLAR`, containing the tools used in this paper, has been released as an open-source package (Chapman-Bird 2023) for use in future population studies.

## ACKNOWLEDGEMENTS

We thank Michael Katz and Alvin Chua for useful discussions regarding EMRI waveforms and detection, and Alvin Chua for additional careful review of the manuscript. CEAC-B thanks Michael Williams for useful discussions on parameter estimation, sampling, and good coding practices. This work was performed using `NUMPY` (Harris et al. 2020), `CUPY` (Okuta et al. 2017), and `PYTORCH` (Paszke et al. 2019). Figures were produced using `MATPLOTLIB` (Hunter 2007), `SEABORN` (Waskom 2021), and `CORNER` (Foreman-Mackey 2016). This work uses the Black Hole Perturbation Toolkit.<sup>1</sup> CEAC-B is supported by STFC studentship 2446638. This work was supported in part by STFC grant ST/V005634/1.

## DATA AVAILABILITY

This paper is accompanied by a tagged data release (Chapman-Bird 2022), consisting of: (i) the data and scripts required to reproduce the figures presented; (ii) the trained neural networks used to produce the results presented, along with the code required to use them; (iii) scripts for generating the data used to train the neural networks.

## REFERENCES

Abbott R. et al., 2023, *Phys. Rev. X.*, 13, 011048  
 Abramowitz M., Stegun I. A., 1964, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications Inc., New York, p. 375  
 Alexander T., 2017, *J. Phys. Conf. Ser.*, 840, 012019  
 Alsing J., Peiris H., Mortlock D., Leja J., Leistedt B., 2023, *ApJS*, 264, 29  
 Amaro-Seoane P., 2021, The Gravitational Capture of Compact Objects by Massive Black Holes, Handbook of Gravitational Wave Astronomy. Springer, Singapore  
 Amaro-Seoane P., Gair J. R., Freitag M., Miller M. C., Mandel I., Cutler C. J., Babak S., 2007, *Class. Quantum Gravity*, 24, R113  
 Amaro-Seoane P. et al., 2017, preprint ([arXiv:1702.00786](https://arxiv.org/abs/1702.00786))  
 Amaro-Seoane P. et al., 2023, *Living Rev. Relativ.*, 26, 2  
 Babak S. et al., 2017, *Phys. Rev. D*, 95, 103012  
 Barack L., Cutler C., 2004, *Phys. Rev. D*, 69, 082005  
 Barber C. B., Dobkin D. P., Huhdanpaa H., 2013, Astrophysics Source Code Library, record ascl: 1304.016  
 Ben-Israel A., Greville T. N. E., 2003, Generalized Inverses: Theory and Applications. Springer, New York  
 Berry C. et al., 2019, *BAAS*, 51, 42  
 Bonetti M., Sesana A., 2020, *Phys. Rev. D*, 102, 103023  
 Bortolas E., Mapelli M., 2019, *MNRAS*, 485, 2125  
 Broggi L., Bortolas E., Bonetti M., Sesana A., Dotti M., 2022, *MNRAS*, 514, 3270  
 Burden R. L., Faires J. D., 1993, Numerical Analysis. Brooks Cole, Boston  
 Chapman-Bird C., 2022, *CC Chapmanbird/emri\_selection\_biases*, Chapman-Bird C., 2023, *poplar*,  
 Christensen N., Meyer R., 2022, *Rev. Mod. Phys.*, 94, 025001  
 Chua A. J. K., Cutler C. J., 2022, *Phys. Rev. D*, 106, 124046  
 Chua A. J. K., Moore C. J., Gair J. R., 2017, *Phys. Rev. D*, 96, 044005  
 Chua A. J. K., Hee S., Handley W. J., Higson E., Moore C. J., Gair J. R., Hobson M. P., Lasenby A. N., 2018, *MNRAS*, 478, 28  
 Cook S., Gelman A., Rubin D., 2006, *J. Comput. Graph. Stat.*, 15, 675  
 Cutler C., Flanagan E. E., 1994, *Phys. Rev. D*, 49, 2658  
 Daly R. A., 2011, *MNRAS*, 414, 1253  
 Dodge Y., 2008, The Concise Encyclopedia of Statistics. Springer New York, New York, NY  
 Feurer M., Hutter F., 2019, Hyperparameter Optimization. Springer International Publishing, Cham, p. 3  
 Foreman-Mackey D., 2016, *J. Open Source Softw.*, 1, 24  
 Fujita R., Shibata M., 2020, *Phys. Rev. D*, 102, 064005  
 Gair J. R., Barack L., Creighton T., Cutler C., Larson S. L., Phinney E. S., Vallisneri M., 2004, *Class. Quantum Gravity*, 21, S1595  
 Gair J. R., Tang C., Volonteri M., 2010, *Phys. Rev. D*, 81, 104014  
 Goodfellow I. J., Bengio Y., Courville A., 2016, Deep Learning. MIT Press, Cambridge, MA, USA  
 Harris C. R. et al., 2020, *Nature*, 585, 357  
 Hastie T., Tibshirani R., Friedman J., 2001, The Elements of Statistical Learning. Springer New York Inc., New York, NY, USA  
 Hogg D. W., 1999, preprint ([arXiv:astro-ph/9905116](https://arxiv.org/abs/astro-ph/9905116))  
 Huerta E. A., Gair J. R., 2011, *Phys. Rev. D*, 84, 064023  
 Hughes S. A., Warburton N., Khanna G., Chua A. J. K., Katz M. L., 2021, *Phys. Rev. D*, 103, 104014  
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90  
 Ibe O., 2013, Markov Processes for Stochastic Modeling. Elsevier, Amsterdam  
 Isoyama S., Fujita R., Chua A. J. K., Nakano H., Pound A., Sago N., 2022, *Phys. Rev. Lett.*, 128, 231101  
 Johansson F. et al., 2013, mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18)  
 Katz M. L., Chua A. J. K., Speri L., Warburton N., Hughes S. A., 2021, *Phys. Rev. D*, 104, 064047  
 Kingma D. P., Ba J., 2014, preprint ([arXiv:1412.6980](https://arxiv.org/abs/1412.6980))  
 Klein A. et al., 2016, *Phys. Rev. D*, 93, 024003  
 Krolak A., Schutz B. F., 1987, *Gen. Relativ. Gravit.*, 19, 1163  
 Laghi D., Tamanini N., Del Pozzo W., Sesana A., Gair J., Babak S., Izquierdo-Villalba D., 2021, *MNRAS*, 508, 4512  
 MacLeod C. L., Hogan C. J., 2008, *Phys. Rev. D*, 77, 043512  
 Maggiore M., 2008, Gravitational Waves Volume 1: Theory and Experiments. Oxford University Press, Oxford, p. 352  
 Mandel I., Farr W. M., Gair J. R., 2019, *MNRAS*, 486, 1086  
 Miller S., Callister T. A., Farr W. M., 2020, *ApJ*, 895, 128  
 Mosteller F., Fisher R. A., 1948, *Am. Stat.*, 2, 30  
 Naoz S., Rose S. C., Michaely E., Melchor D., Ramirez-Ruiz E., Mockler B., Schnittman J. D., 2022, *ApJ*, 927, L18  
 Okuta R., Unno Y., Nishino D., Hido S., Loomis C., 2017, Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS). Curran Associates Inc., New York  
 Pan Z., Lyu Z., Yang H., 2021, *Phys. Rev. D*, 104, 063007

<sup>1</sup>[bhptoolkit.org/](https://bhptoolkit.org/)

Paszke A. et al., 2019, in Wallach H., Larochelle H., Beygelzimer A., dAlché Buc F., Fox E., Garnett R., eds, *Advances in Neural Information Processing Systems*. Curran Associates, Inc., New York

Peters P. C., Mathews J., 1963, *Phys. Rev.*, 131, 435

Pieg L., Tiller W., 1987, *Comput. Aided Des.*, 19, 485

Robson T., Cornish N. J., Liu C., 2019, *Class. Quantum Gravity*, 36, 105011

Roulet J., Zaldarriaga M., 2019, *MNRAS*, 484, 4216

Sauer T., 2012, *Numerical Analysis*. Pearson, Michigan

Sesana A., Barausse E., Dotti M., Rossi E. M., 2014, *ApJ*, 794, 104

Shankar F., 2013, *Class. Quantum Gravity*, 30, 244001

Stein L. C., Warburton N., 2020, *Phys. Rev. D*, 101, 064007

Stevenson S., Berry C. P. L., Mandel I., 2017, *MNRAS*, 471, 2801

Tinto M., Dhurandhar S. V., 2021, *Living Rev. Relativ.*, 24, 1

Vallisneri M., 2008, *Phys. Rev. D*, 77, 042001

Vázquez-Aceves V., Zwick L., Bortolas E., Capelo P. R., Amaro Seoane P., Mayer L., Chen X., 2022, *MNRAS*, 510, 2379

Volonteri M., 2010, *A&AR*, 18, 279

Waskom M. L., 2021, *J. Open Source Softw.*, 6, 3021

Williams M. J., 2021, *nessai: Nested Sampling with Artificial Intelligence*, Zenodo

Williams M. J., Veitch J., Messenger C., 2021, *Phys. Rev. D*, 103, 103006

Zevin M. et al., 2021, *ApJ*, 910, 152

## APPENDIX A: EMRI PARAMETER ESTIMATION AND CATALOGUE GENERATION

The generation of a catalogue of EMRI observations is multifaceted. We first describe the generation of EMRI waveforms, including our treatment of initial conditions and our choices regarding waveform model and detector response, in Section A1. In Section A2, we outline our procedure for obtaining posterior samples for EMRIs that pass the detection threshold, along with consistency checks we perform to ensure the approximations made in this process are justifiable. These two steps are performed for each set of EMRI parameters drawn from the population to construct a catalogue of EMRI detections and their corresponding posterior samples.

### A1 Parametric conventions and waveform generation

The EMRI parameter space is complicated, consisting of 18 parameters:

(i) The intrinsic parameters of the EMRI describe the properties of the two objects and their initial orbital configuration. The primary MBH is described by its mass  $M$  and spin vector  $\mathbf{a}$ , and the secondary CO similarly by mass  $\mu$  and spin vector  $\mathbf{a}_{\text{CO}}$ . As  $\mathbf{a}_{\text{CO}}$  is not predicted to have a significant effect on EMRI detectability (Huerta & Gair 2011), it is not currently included in state-of-the-art waveform models, including the model used in this study (Katz et al. 2021). The inspiral orbit is described by the initial eccentricity  $e_0$ , semilatus rectum  $p_0$ , and orbital inclination  $i_0$ , along with three orbital phases  $\Phi_r$ ,  $\Phi_\theta$ , and  $\Phi_\phi$  (Fujita & Shibata 2020).

(ii) The extrinsic parameters describe the orientation of the system and its location with respect to the detector. The position of the system is described by the luminosity distance vector  $\mathbf{d}_L$ , the magnitude of which may be described by a redshift  $z$  via the relation

$$d_L = (1+z) \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}, \quad (\text{A1})$$

where  $H_0$  is the Hubble constant (Hogg 1999). We separate out vectors into their magnitudes and angular components, such that  $\mathbf{d}_L = \{d_L, \theta_S, \phi_S\}$ . Similarly, we decompose the primary spin magnitude vector  $\mathbf{a} = \{a, \theta_K, \phi_K\}$ . The sets of angles describe the orientation of the MBH spin-angular momentum vector and the sky

position vector, respectively, with  $\theta$  and  $\phi$  referring to polar and azimuthal angles, respectively.

(iii) The plunge time of the system with respect to the start of the observation is described by the parameter  $t_{\text{plunge}}$ . However, the concept of initial conditions on a population level is not well defined unless a common reference point in the waveform is set. Therefore, we adjust  $p_0$  such that the EMRI waveform will plunge after 10 yr (Stein & Warburton 2020). For our fiducial example outlined in Section 4, we assume a *LISA* observing window of 4yr in line with the current mission proposal (Amaro-Seoane et al. 2017). Our choice to allow for EMRI plunges to occur up to 10 yr after the beginning of *LISA* observation is made to accommodate the presence of EMRIs in the data that plunge after the end of the observation window. Neglecting these events as being undetectable is not typically a reasonable approximation. In many cases,  $\rho_{\text{opt}} > \rho_t$  even for  $t_{\text{plunge}} = 10$  yr, so some detectable events in the data are ignored despite this extension of EMRI plunge times to the post-window regime. By excluding these events that plunge after 10 yr, the results of our population inference will be conservative. We include this effect to demonstrate that our approach is capable of accommodating a post-window cut on  $t_{\text{plunge}}$ , but acknowledge that the tuning of such a cut-off point with respect to the detectability of the excluded signals is an issue that warrants further investigation in future work.

Waveform generation also includes some additional considerations to transform from the source frame to the detector frame. Prior to waveform generation, we convert source-frame masses to detector-frame masses with the mapping  $M_{\text{det}} = (1+z)M$  (Krolak & Schutz 1987). For convenience, we do not include a detector response in our EMRI waveform modelling and instead work in terms of the waveform strain [as opposed to the TDI combinations that the *LISA* detector outputs will be used to construct (Tinto & Dhurandhar 2021)]. This choice leads to a degeneracy between  $\phi_K$  and  $\phi_S$ , which we navigate by defining a new parameter  $\Delta\phi = \phi_S - \phi_K$ . Our approach can incorporate a chosen *LISA* response by reverting back to the separate angles and including the response function in waveform generation. As the addition of a response function does not significantly alter EMRI SNRs, we do not expect its exclusion to affect the validity of our approach.

For our waveform model, we choose the fifth-order post-Newtonian Augmented Analytic Kludge recently implemented in the FEW package (Chua, Moore & Gair 2017; Katz et al. 2021). The validity of our population inference framework should not depend strongly on this choice, as the EMRI SNR function should remain well behaved and smooth for any reasonable choice of waveform model, although the specific numerical results may vary for different waveforms.

To accommodate our choice of initial conditions, we generate 10 yr EMRI waveforms in the time domain with a sampling rate of 0.1Hz, and crop them according to their (randomly sampled)  $t_{\text{plunge}}$  values. We calculate the waveform's  $\rho_{\text{opt}}$  value via equation (16) and produce a noise-realized SNR estimate  $\rho_n$  by drawing a sample from equation (15). Detection is evaluated via equation (14).

### A2 Parameter estimation

For waveforms that pass the detection threshold, we proceed to draw samples from the posterior distribution on  $\theta$ . Operating under the LSA, we determine the FIM  $\Gamma$  from numerical waveform derivatives computed using the five-point stencil method (Sauer 2012). Appropriate step sizes that produce accurate (and stable)

numerical derivatives were determined empirically by computing  $\langle \partial_\ell h | \partial_\ell h \rangle$  (the FIM diagonal terms) on a grid and identifying regions of step-size space for which this converged. This stability was then verified across the EMRI parameter space. With  $\Gamma$  computed, we then invert it to obtain the covariance matrix of the LSA likelihood. FIMs for EMRI waveforms typically have large condition numbers, which can cause issues when performing matrix inversion; we mitigate this by employing singular value decomposition (SVD) to compute the pseudoinverse of  $\Gamma$  (Ben-Israel & Greville 2003). The numerical stability of this inversion can also be problematic, even for double precision; to alleviate this, we perform the SVD with 500-point decimal precision using the `MPMATH` package (Johansson et al. 2013). We also perform some additional verification of the validity of the LSA likelihood as a substitute for the full likelihood as recommended in Vallisneri (2008). This check is performed by examining the ratio between the LSA and full likelihood, which we denote  $r$ . By sampling the 1 standard deviation isoprobability contour in the LSA likelihood and computing  $|\log_{10} r|$  for each of these samples, we can examine whether the high-probability regions of the LSA and the full likelihood are consistent. Following Vallisneri (2008), we accept the LSA likelihood as a suitable approximation if and only if 90 per cent of the samples satisfy  $|\log_{10} r| < 0.1$ . We found that for waveforms with  $\rho_{\text{opt}} > 10$  that this condition was satisfied in all cases, which is well below our chosen SNR threshold of 20.

Our goal is to obtain posterior samples via the LSA that are representative of those that would be obtained from sampling the true posterior. In the absence of noise, the likelihood will peak on the true values provided there are no degeneracies in the parameter space. For EMRIs, this is not generally true, but the degeneracies are non-local and therefore do not affect the morphology of the posterior near the true values (Chua & Cutler 2022). However, when noise is included, the likelihood shifts in a random direction in parameter space such that the true values are no longer at the maximum likelihood point. To simulate this measurement effect, we draw one sample from the posterior (which, assuming broad uniform priors, is equivalent to the event likelihood) and recompute the FIM at this new point (Stevenson, Berry & Mandel 2017). Finally, with our second FIM computed, we invert it and draw  $S = 10^4$  samples from the corresponding multivariate normal distribution. The posterior samples in  $d_L$  are converted to samples in  $z$  assuming a standard cosmology, and detector-frame masses then converted back to the source frame by dividing through by  $(1 + z)$  on a per-sample basis. This modified set of posterior samples is the final product of the individual EMRI event simulation.

## APPENDIX B: NEURAL NETWORK DESIGN AND TRAINING

MLPs are tuneable mappings between input and output vectors (of specified lengths) that consist of a layered structure of matrix multiplications which are passed through non-linear functions. The non-linearity between each layer, combined with a large number of tuneable parameters in each matrix multiplication, enables the resulting neural network to mimic complex mappings between high-dimensional spaces (Goodfellow et al. 2016). This tuning is performed in a process known as training, in which the performance of the neural network is maximized with respect to a pre-computed data set.

**Table B1.** Architecture and training settings chosen for our two MLP neural networks introduced in Section 3.

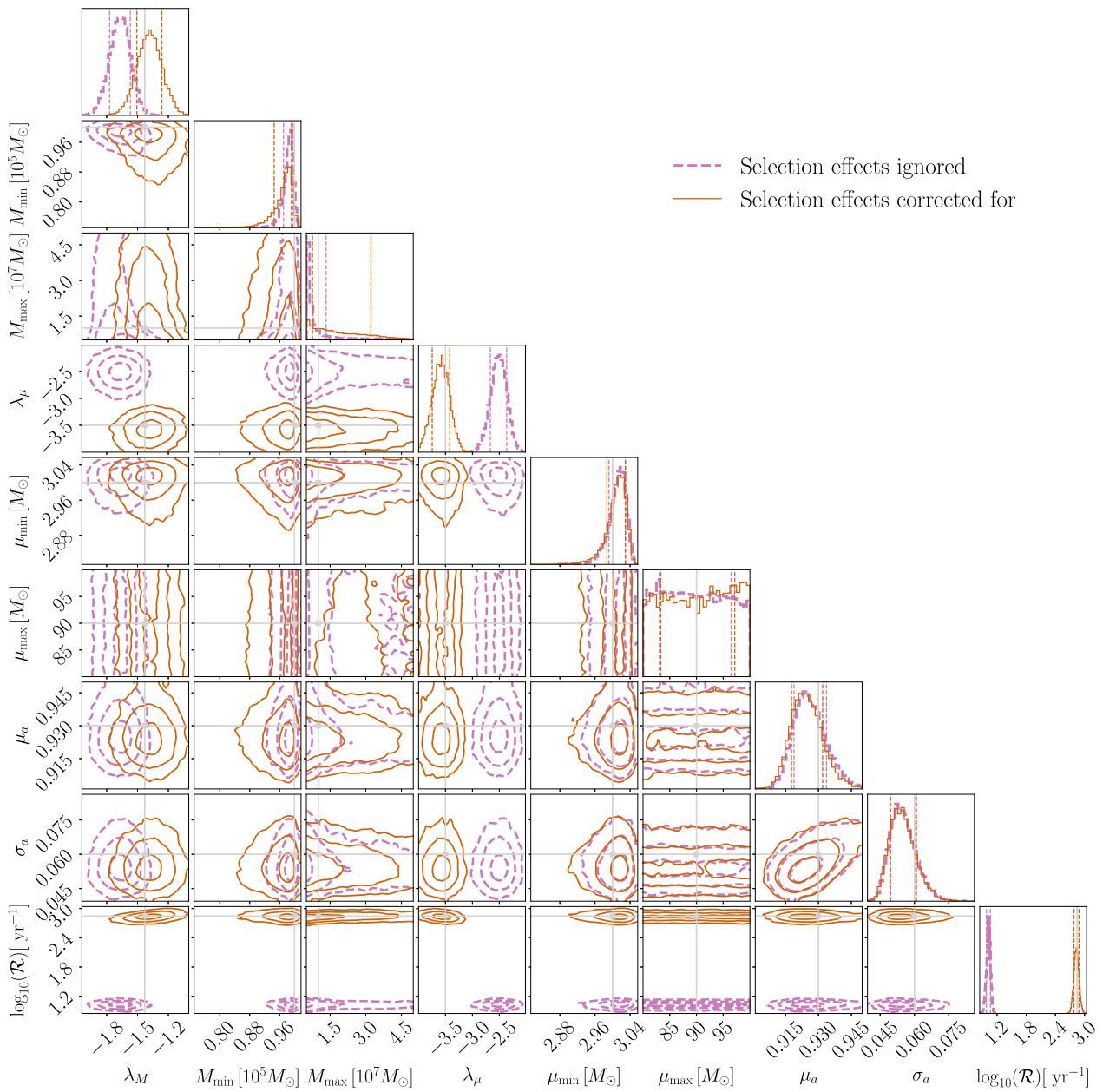
Setting	SNR MLP	$\alpha(\lambda)$ MLP
Number of (hidden) layers	10	8
Neurons per layer	128	128
Activation function	SiLU	SiLU
Rescaling	Unit normal	Unit normal
Optimizer	Adam	Adam
Learning rate	$5 \times 10^{-4}$	$5 \times 10^{-4}$
Batch size	$10^4$	$10^5$
Max epochs	$10^5$	$10^3$
Loss function	L1	L1

The number of required neurons and layers (which describe the dimensions of each matrix multiplication) in an MLP depends on the complexity of the function to be interpolated and the number of interpolation dimensions. Due to the stochastic nature of training neural networks, tuning of the learning rate, batch size, and number of iterations employed during training is required to maximize performance. For complicated problems, optimization techniques may be employed to explore the space of network settings and identify a sensible configuration (Feurer & Hutter 2019). In our case, the problem is sufficiently low-dimensional that we were able to obtain effective MLPs through the manual tuning of network settings. Network complexity was gradually increased through the addition of neurons or layers until overfitting (Goodfellow et al. 2016) was observed. This is characterized by the performance of the network on testing data degrading despite continued improvement in performance on training data. At this point, training settings were adjusted to minimize this overfitting. The choice to rescale training data to that of a unit normal distribution, and to use the Adam optimization algorithm (Kingma & Ba 2014), was made following current best practices (Goodfellow et al. 2016): use of other optimization or rescaling functions was not found to significantly affect network performance.

The resulting network settings chosen for the SNR and selection function MLPs, summarized in Table B1, are almost identical. Two extra hidden layers are added for the SNR MLP, which is to be expected given the higher dimensionality of the EMRI parameter space in comparison to the hyperparameter space.

## APPENDIX C: FULL HYPERPOSTERIOR OBTAINED FROM 4-YR SCENARIO

The full hyperposterior from the population inference in Section 4 are shown in Fig. C1. By including the mass range parameters, we can observe a more subtle consequence of the presence of selection biases: overconstrained hyperposterior. Neglecting selection effects leads to an underestimation of the error on parameters: this is reflected in the P–P plot analysis of Section 5, where it is demonstrated that these effects lead to globally inconsistent hyperposterior effects at a statistically significant level.



**Figure C1.** Full recovered hyperposterior distributions for our example scenario with selection effects corrected for (solid) or ignored (dashed). The true values of each hyperparameter are indicated by cross-hairs. Significant bias is present for  $\lambda_\mu$ , with more minor bias for  $\lambda_M$ . The rate  $\mathcal{R}$  is accurately recovered with the inclusion of selection effect correction. For other hyperparameters, the dashed hyperposterior distributions are overconstrained when compared to the corrected hyperposterior distributions.

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