Unlocking New Paths for Efficient Analysis of Gravitational Waves from Extreme-Mass-Ratio Inspirals with Machine Learning

Bo Liang[1,2,3†], Hong Guo[4†], Tianyu Zhao[1], He Wang[5,3], Herik Evangelinelis[6], Yuxiang Xu[1,2,3], Chang Liu[1,7], Manjia Liang[1], Xiaotong Wei[1], Yong Yuan[1], Minghui Du[1*], Peng Xu[1,8,3,9*], Weiliang Qian[4], and Ziren Luo[1,8,3]

[1]*Center for Gravitational Wave Experiment, National Microgravity Laboratory, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China*
[3]*Taiji Laboratory for Gravitational Wave Universe (Beijing/Hangzhou), University of Chinese Academy of Sciences, Beijing 100049, China*
[4]*Escola de Engenharia de Lorena, Universidade de São Paulo, 12602-810 Lorena, Brazil*
[5]*International Centre for Theoretical Physics Asia-Pacific, University of Chinese Academy of Sciences, Beijing 100049, China*
[6]*Escola de Engenharia, Universidade Federal Fluminense, 24210-240, Niterói, Brazil*
[7]*National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China*
[8]*Key Laboratory of Gravitational Wave Precision Measurement of Zhejiang Province, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China*
[9]*Lanzhou Center of Theoretical Physics, Lanzhou University, Lanzhou 730000, China*

(Received 11 June 2025; accepted manuscript online 16 June 2025)

Extreme-mass-ratio inspiral (EMRI) signals pose significant challenges to gravitational wave (GW) data analysis, mainly owing to their highly complex waveforms and high-dimensional parameter space. Given their extended timescales of months to years and low signal-to-noise ratios, detecting and analyzing EMRIs with confidence generally relies on long-term observations. Besides the length of data, parameter estimation is particularly challenging due to non-local parameter degeneracies, arising from multiple local maxima, as well as flat regions and ridges inherent in the likelihood function. These factors lead to exceptionally high time complexity for parameter analysis based on traditional matched filtering and random sampling methods. To address these challenges, the present study explores a machine learning approach to Bayesian posterior estimation of EMRI signals, leveraging the recently developed flow matching technique based on ordinary differential equation neural networks. To our knowledge, this is also the first instance of applying continuous normalizing flows to EMRI analysis. Our approach demonstrates an increase in computational efficiency by several orders of magnitude compared to the traditional Markov chain Monte Carlo (MCMC) methods, while preserving the unbiasedness of results. However, we note that the posterior distributions generated by FMPE may exhibit broader uncertainty ranges than those obtained through full Bayesian sampling, requiring subsequent refinement via methods such as MCMC. Notably, when searching from large priors, our model rapidly approaches the true values while MCMC struggles to converge to the global maximum. Our findings highlight that machine learning has the potential to efficiently handle the vast EMRI parameter space of up to seventeen dimensions, offering new perspectives for advancing space-based GW detection and GW astronomy.

DOI: 10.1088/0256-307X/42/8/081101 CSTR: 32039.14.0256-307X.42.8.081101

1. Introduction. Gravitational wave (GW) astronomy and cosmology, emerging in recent decades as a transformative branch of physics, hold great promise to revolutionize our understanding of the universe. In 2015, the LIGO Scientific Collaboration marked a milestone with the first detection of GWs originating from the merger of two black holes.[1] Since this discovery, ground-based GW observatories, including LIGO and VIRGO, have cataloged more than one hundred GW events from various binary systems, such as binary black holes, binary neutron stars, and neutron star-black hole pairs.[2–7] Constrained by seismic and shot noise, the sensitive frequency ranges of these ground-based detectors span from 10 Hz to 10^3 Hz,[7,8] corresponding to binary systems with a total mass of up to several hundred solar masses.

Future space-based GW detectors like the commissioned LISA project,[9,10] as well as Taiji[11,12] and TianQin,[12,13] are engineered to detect GW signals in

†These authors contributed equally to this work.
*Corresponding authors. Email: duminghui@imech.ac.cn; xupeng@imech.ac.cn
© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

the 0.1 mHz to 1 Hz bands. These bands contain a diverse array of sources that are often challenging to observe but are rich in scientific potential. Such sources typically involve strong-field dynamics and complex interactions within extreme environments, potentially carrying vast amounts of untapped information. These systems include double white dwarfs,[14–17] coalescing massive black hole binaries,[16,18–20] extreme-mass-ratio inspirals (EMRIs),[16,21,22] stochastic GW backgrounds,[16,23–25] and other unmodeled sources. Among these, EMRIs are particularly significant as primary detection targets for these observatories.[26] In an EMRI system, a stellar-mass compact object (the secondary) is captured by and gradually spirals into a central supermassive black hole (the primary) over more than 10^5 orbits, emitting low-frequency gravitational radiation in the millihertz band. Due to the cumulative effects of thousands to millions of orbital cycles and the prolonged orbital evolution in strong-field and extreme environments that last several months to years, EMRIs provide an exceptional platform for probing strong gravitational fields and near-horizon physics, as well as for precise measurements of the theoretical parameters.[27,28]

In the field of EMRI parameter estimation, several key challenges significantly hinder the effectiveness of traditional methods like Markov chain Monte Carlo (MCMC). Firstly, the long-term nature of EMRI signals requires extremely precise and computationally intensive waveform modeling, demanding sustained analytical efforts over periods that can extend for months or years.[29,30] Secondly, the high-dimensional parameter space inherent to EMRIs, typically consisting of more than a dozen interrelated variables, adds considerable complexity to the analysis.[31–33] Lastly, analysis of EMRI signals is often plagued by significant **degeneracy**, making it challenging to distinguish between different parameters of a given model,[34–36] as well as **confusion** arising from the difficulty in discriminating between different GW models.[35,37,38] Both factors lead to significant challenges in identifying the model as well as its underlying physical parameters. Specifically, the nonlocal parameter degeneracies can be largely attributed to multiple local maxima, as well as flat regions and ridges inherent in the high-dimensional likelihood function. Traditional methods, particularly MCMC,[36,39,40] struggle with these challenges due to their inherent limitations. MCMC is not only computationally expensive, typically requiring the generation of approximately $\sim 10^{40}$ waveform templates,[31] but also inefficient in exploring vast parameter spaces. In practice, only two to four parameters can be accurately estimated, and this often requires setting the initial estimates near the true values.[41–43] This limitation undermines the ability to thoroughly explore the parameter landscape, frequently missing critical global maxima necessary for accurate parameter estimation. In response, several new tools are under development.[35,43–45] However, at present, most methods are still incapable of meeting the stringent demands of EMRI parameter estimation, highlighting the need for more efficient and robust deep learning techniques.

Deep learning has rapidly become an essential tool in scientific discovery, yielding significant advancements across a wide range of fields, from physics to biology.[46,47] Additionally, the scope of machine learning techniques has broadened to include EMRI data analysis. Current research has made strides in denoising,[48] detecting,[49,50] and performing point estimation of EMRI parameters,[51] which are critical in interpreting these complex signals. However, despite these advancements, a significant challenge remains in the Bayesian posterior estimation of EMRI signals. This difficulty stems largely from the inherent complexity of EMRI data, characterized by protracted signal durations and intricate waveform structures, which continue to test the limits of existing machine learning methodologies. Overcoming this challenge requires a concerted effort to develop more sophisticated algorithms that can harness the full potential of deep learning for comprehensive EMRI analysis, pushing the boundaries of what is currently achievable in GW astronomy.

In this study, we develop a deep learning approach specifically tailored for the posterior estimation of EMRI signals by employing the recently developed flow matching (FM) technique based on ordinary differential equation neural networks. Our method not only enhances computational efficiency, but also overcomes the issue of MCMC methods getting easily trapped in local optima during parameter estimation. Such an approach exhibits computational efficiency that surpasses traditional MCMC methods by several orders of magnitude, all while maintaining unbiased parameter estimation. It is demonstrated that the machine learning technology employed is capable of effectively managing the extensive parameter space, comprising as many as eleven variables, inherent to EMRI signals. Moreover, as the method can be adapted to a broader prior range to achieve unbiased parameter estimates, it significantly narrows the parameter space of Bayesian inference. As a result, our findings potentially provide an innovative solution for dealing with real EMRI signals in the future. Specifically, one first narrows down the parameter space of EMRI using the proposed machine learning procedure, and then performs accelerated MCMC sampling using the machine learning posterior as a prior or proposal distribution. Such an approach combines the efficiency of machine learning with the precision of traditional Bayesian inference, offering an innovative solution for the analysis of EMRI signals.

The remainder of the paper is organized as follows. In the next section, we discuss the data generation and preprocessing, laying the foundation for the training and validation of the model. Subsequently, we elaborate on the machine learning algorithms specifically tailored for the Bayesian posterior estimation of EMRIs. The numerical results are presented in the following section, where our results are compared against the standard MCMC approach. Concluding remarks and future perspectives are given in the last section.

2. Methodological Framework. In this section, we give a detailed account of the dataset preparation for model

Table 1. The prior distributions of EMRI source parameters used in this work. For each parameter, the table presents its lower and upper bounds, and a uniform distribution is assumed between these bounds.

Parameter	Description	Prior lower bound	Prior upper bound
$\mathcal{M}[M_\odot]$	Primary mass	$9\times10^5 M_\odot$	$1.1\times10^6 M_\odot$
μ	Secondary mass	$9M_\odot$	$11M_\odot$
a	Primary spin	0.8	0.99
p_0	Semi-latus rectum	9.1	9.3
e_0	Orbital eccentricity	0.1	0.3
Y_0	Orbital inclination parameter	0.59	0.79
θ_S [rad]	Polar angle of source orientation	0 rad	π rad
ϕ_S [rad]	Azimuthal angle of source orientation	0 rad	2π rad
θ_K [rad]	Polar angle of orbital angular direction	0 rad	π rad
ϕ_K [rad]	Azimuthal angle of orbital angular direction	0 rad	2π rad

training and validation, which consists of the generation of EMRI waveforms, projection to detector responses, and a preprocessing step.

A prominent challenge in the analysis of EMRI signal lies in the accurate and high-fidelity modeling of waveform template, which relies on the theories of gravitational self-force and black hole perturbation, and other related theories.[29,30] To date, the most advanced computational framework named `FastEMRIWaveforms` (`FEW`)[32,33] can rapidly compute fully relativistic EMRI waveforms in the time domain. As the first step of data preparation, we employ the augmented analytic kludge (AAK) model implemented in `FEW` to generate the polarizations of EMRI waveforms. For the second step, we adopt the generic time-domain response function `FastLISAResponse`[52] to project the polarizations into the detector responses of LISA, and the resulting EMRI signals are output in the form of second-generation time delay interferometry (TDI) variables.[53] Each sample in the dataset includes the TDI-$\{A, E\}$ variables with a duration of 2 months, calculated at a sampling frequency of 0.2 Hz. We employ the graphics processing units (GPUs) to accelerate the computation of waveforms and Taiji response functions,[52] achieving the generation of each EMRI signal in less than one second.[42] The whole dataset consists of 3080000 samples, with the source parameters randomly drawn from the priors specified by Table 1. In our research, we have come to deeply appreciate the importance of the effectiveness, reliability, and scalability of machine learning algorithms in achieving long-term research objectives. To ensure the realization of these goals, we have adopted a robust, step-by-step approach, continuously improving and optimizing based on existing research. In this process, we have carefully adjusted and compromised our prior assumptions. In particular, to enhance the accuracy of posterior estimation, we have reasonably restricted the prior range of some intrinsic parameters of EMRI signals.

To reasonably reduce the dimensionality of data, we further incorporate a preprocessing step. Frequency domain analysis[42,54,55] proves to be an exceptionally effective method for estimating parameters in EMRIs. Therefore, the time-domain samples are converted to the frequency domain using the fast Fourier transform (FFT) algorithm. The data of EMRIs are inherently challenging to process due to their vast sizes,[56] and this issue is even more pronounced for machine learning, for which the models are trained on batched data. After FFT, the length of each sample is up to millions, necessitating methods such as downsampling or pooling. To reduce the complexity of the data while still retaining key information, we adopted a max-pooling strategy.[57] We implemented a max-pooling layer with a window size of 512 and a stride of 512 along the temporal dimension of the input signals. Each window processes 512 consecutive data points and retains only the maximum value within that segment. Prior to training, 500000 samples are stored on the hard disk, establishing a robust data foundation for later calculations.

3. Machine Learning Framework. We begin this section by reviewing our methodology, focusing on the use of continuous normalizing flows (CNFs) for fast and accurate EMRI parameter estimation. Currently, many successful attempts aimed at enhancing parameter estimation through machine learning have adopted the neural posterior estimation (NPE) approach[58–62] using discrete normalizing flows (DNFs).[63,64] In theory, the application of NPE to EMRIs is feasible. However, recent research in simulation-based inference (SBI) suggests that FM[65] with CNFs[66] is a more promising approach. This method is known as flow matching posterior estimation (FMPE).[65,67] FMPE offers a more rapid training process and superior accuracy compared to NPE.[60,62,67]

The emergence of FM provides a new perspective for training CNFs. Dax *et al.*[67] have applied the FM to SBI, offering a more efficient training objective that helps address the computational challenges inherent in traditional methods. The loss function of FMPE is given by the following equation:

$$L_{\text{FMPE}} = E_{t\sim p(t),\theta_1\sim p(\theta),x\sim p(x|\theta_1),\theta_t\sim p_t(\theta_t|\theta_1)} \times \| v_{t,x}(\theta_t) - u_t(\theta_t|\theta_1) \|^2 . \quad (1)$$

In the FMPE approach, the vector field $v_{t,x}(\theta_t)$ is responsible for generating the path to the target probability distribution, while $u_t(\theta_t|\theta_1)$ adjusts this path based on the conditions of the sample. By defining a Gaussian path with time-dependent parameters, FMPE ensures that at $t = 0$ and $t = 1$, the probability path closely approximates the base and target distributions.

In order to accurately estimate parameters for EMRIs, we have conducted customized training on our model. Our model is based on CNF. It is optimized using FMPE technology. We trained the model for 100 epochs, with a batch size of 1024 samples per epoch. The initial learning rate was set to 0.00005, and we employed a cosine annealing strategy to gradually reduce the learning rate to zero as training progressed. The training was conducted on an NVIDIA 4090 GPU, the entire training process took approximately 1 hour. Once our model is trained, the inference process for a single EMRI signal can be completed in just tens of seconds.

The development of the CNFs for EMRI parameter estimation task employed the deep learning libraries PaddlePaddle and PyTorch, supported by the data analysis libraries Pandas and NumPy. This work model has two parts: (1) The first part is a linear compression network N_x, designed to effectively capture the intrinsic structure of the signal by compressing the data x into a 1024-dimensional feature space. This process not only simplifies the data processing workflow but also enhances the model's ability to capture signal features. (2) The second part of the model is a flow network composed of 56 residual blocks with decreasing sizes, ranging from 4096 dimensions down to 10 dimensions. It receives the embedded tuple (x, t, θ_t) and predicts the vector field $v_{t,x}(\theta_t)$.

The Bayesian statistical inference framework along with the MCMC sampling technique have been widely adopted in GW source detection and parameter estimation, and is nowadays one of the benchmark approaches in this area, especially for the global analysis of multiple sources.[68,69] Since FMPE is designed to simulate the Bayesian posterior distribution of parameters, its performance could be more clearly demonstrated if the Bayesian results for the same event are presented for comparison. The framework of Bayesian statistical inference is built on the Bayes' theorem, which, neglecting the normalization factor,[42] can be expressed in logarithmic form as:

$$\log p(\theta|d) = \log p(d|\theta) + \log p(\theta). \tag{2}$$

The first term on the right hand side $\log p(d|\theta)$ represents the likelihood function. The likelihood function describes the probability of observing the data stream d given the parameters θ, which takes the Whittle form under the assumption of Gaussian and stationary noise.[70] The second term $\log p(\theta)$ is the prior probability distribution of the parameters θ given in Table 1. In the parameter estimation of EMRI signals, we face challenges caused by nonlocal parameter degeneracies, which lead to multiple local maxima on the likelihood surface. These local maxima make the EMRI signal extremely sensitive to small variations in parameters, thus accurately identifying the global maximum in the likelihood function becomes particularly difficult. Current MCMC-based methods[42,71,72] typically require either searching near true parameter values or estimating only a subset of EMRI parameters. In comparison to our machine learning approach, we implement conventional Bayesian inference using `Eryn`,[72] an advanced MCMC sampler based on `Emcee`,[73] to sample the posterior distribution. The `Eryn` has become one of the widely used tools in the current space-based GW data analysis. We performed Bayesian inference on the injected EMRI signals to test the performances of both MCMC and our machine learning methods. The MCMC tests are conducted based on two settings differing in the choices of starting points:[41]

1. MCMC (1): The starting points are confined within a tight neighborhood around the true values (within a range of $\times 10^{-7}$ times the true values);
2. MCMC (2): The starting points are randomly drawn according to the priors.

4. Numerical Results. By running MCMC with setting (1), we obtain the posterior distributions of parameters shown in Fig. 1, where the contours from inside to outside represent the 1 σ, 2 σ, and 3 σ confidence intervals, respectively, and the black lines indicate the true parameters underlying the injected signal.

Unsurprisingly, all the median values of parameters are perfectly coincident with the truths, demonstrating that when the starting points are close to the true values, MCMC can efficiently and correctly identify the source parameters. While given that in practical data analysis we do not have prior access to the true parameters, we explored the capability of MCMC based on the setting (2), which involves randomly drawing starting points from the prior distributions. Results of the second setup are shown in Fig. 2. For this case, MCMC failed to recover the true parameters, with most of the "posteriors" clustered near the boundaries set by the priors. Also shown in Fig. 2 are the trajectories of the Markov chains, indicating that MCMC failed to find the global optimum,[31,36,56] with some of the parameters showing evidence of being trapped in a local optimum.

Both MCMC configurations employed GPU acceleration, yet still required approximately two days to complete the posterior estimation of EMRI signals. Moreover, to the best of the authors' knowledge, there is currently no research that has shown MCMC methods can yield the correct posterior distribution of EMRI when starting from a relatively broad prior range. To evaluate our machine learning model, we first tested it on 1000 EMRI signals randomly generated according to the priors. The P–P plot in Fig. 3, constructed from our model's posterior samples, shows excellent agreement with the expected diagonal line for all parameters' cumulative distribution functions (CDFs), thereby validating the unbiasedness of our estimations over the whole parameter range in consideration.

In comparison to the MCMC method, we also performed parameter estimation on the same EMRI signal as shown in Table 2. Since our model cannot exploit the advantage of knowing the true parameters priorly, its application scenario is essentially identical to MCMC (2) (i.e., the more realistic one). As shown in Fig. 4, the posterior distribution of our model successfully encompasses the true parameters within the 1 σ (or at most 2 σ) ranges. This is in stark contrast to the result of MCMC (2) (Fig. 2),

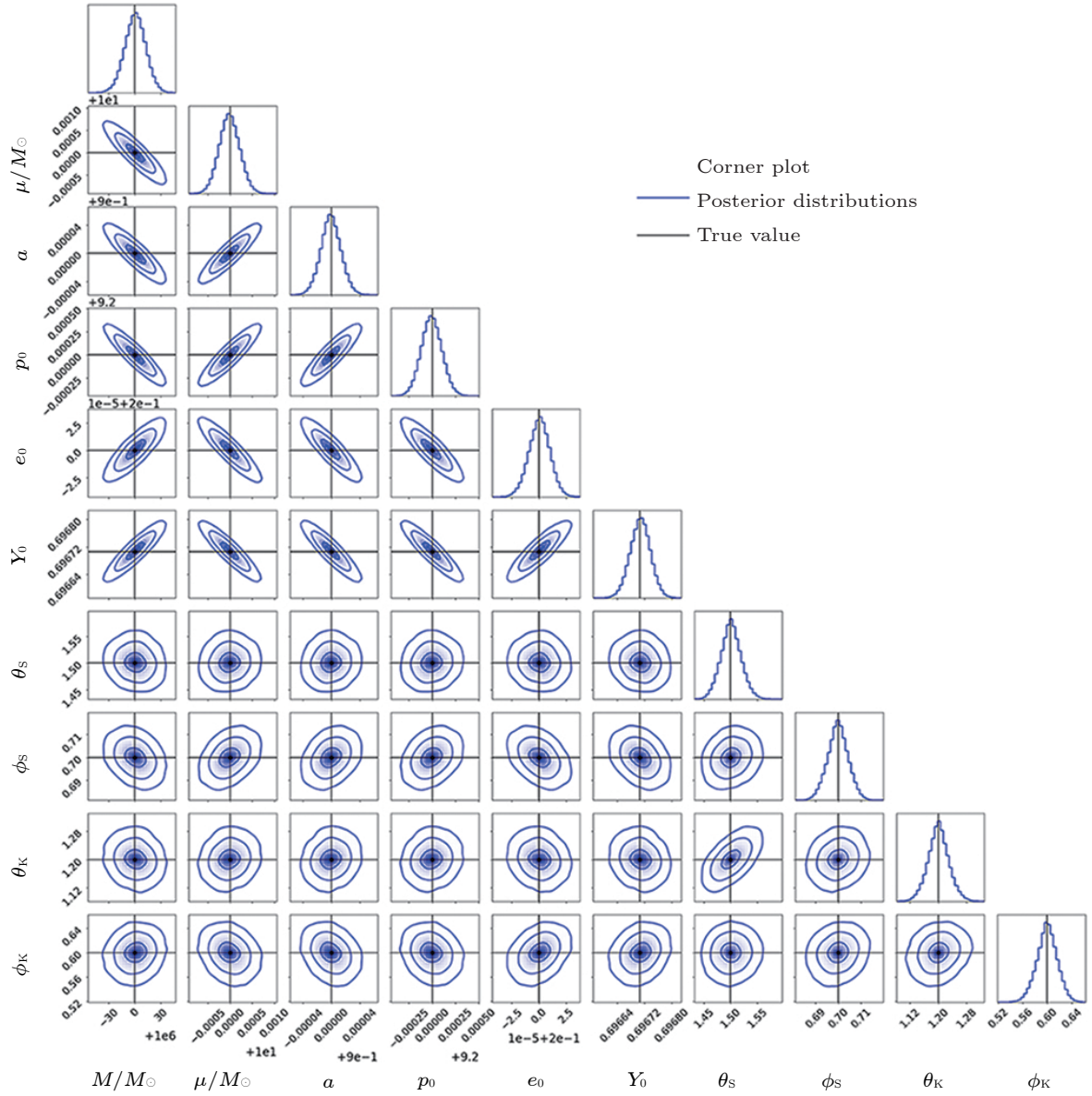

Fig. 1. MCMC with narrow initialization. Posterior distributions obtained by initializing MCMC chains near true parameters (black vertical lines). The analyzed EMRI signal has a signal-to-noise ratio of approximately 67 and is based on two years of observational data. In the figure, the blue line represents the posterior probability distribution of the EMRI parameters, while the black line indicates the true parameter values.

where the estimations are severely biased. This comparison provided us with an intuitive overview of the performance of both methods. FMPE exhibits a strong capability to escape from local maxima, and this is achieved in a very short time scale. To some extent, this suggests that FMPE has the potential to become a viable methodology for EMRI posterior estimation, in terms of both efficiency and accuracy.

To further assess the reliability of our model in terms of reconstructing posterior distributions, it is helpful to compare it with MCMC (1), whose results might be regarded as a benchmark. Our machine learning process typically generates posterior distributions with a broader range than those of `Eryn`; however it has effectively narrowed down the parameter ranges compared to the priors, especially for the sky position parameters (as can be seen in the first peaks of θ_S and ϕ_S), making it possible to perform rapid localization of EMRIs via our model. Through the above comparisons, by integrating the features of MCMC and machine learning, we can gain insightful ideas to further refine the results of both methods. Looking ahead, in dealing with real EMRI signals, we can adopt the following strategy: first, use the machine learning process to quickly narrow down the range of the parameter space, and then apply the MCMC method for more refined sampling, using the posterior distribution of the machine learning model as a prior. This approach combines efficiency and ability to break local maxima of machine learning with the precision of Bayesian inference, providing an innovative solution for analyzing EMRI signals.

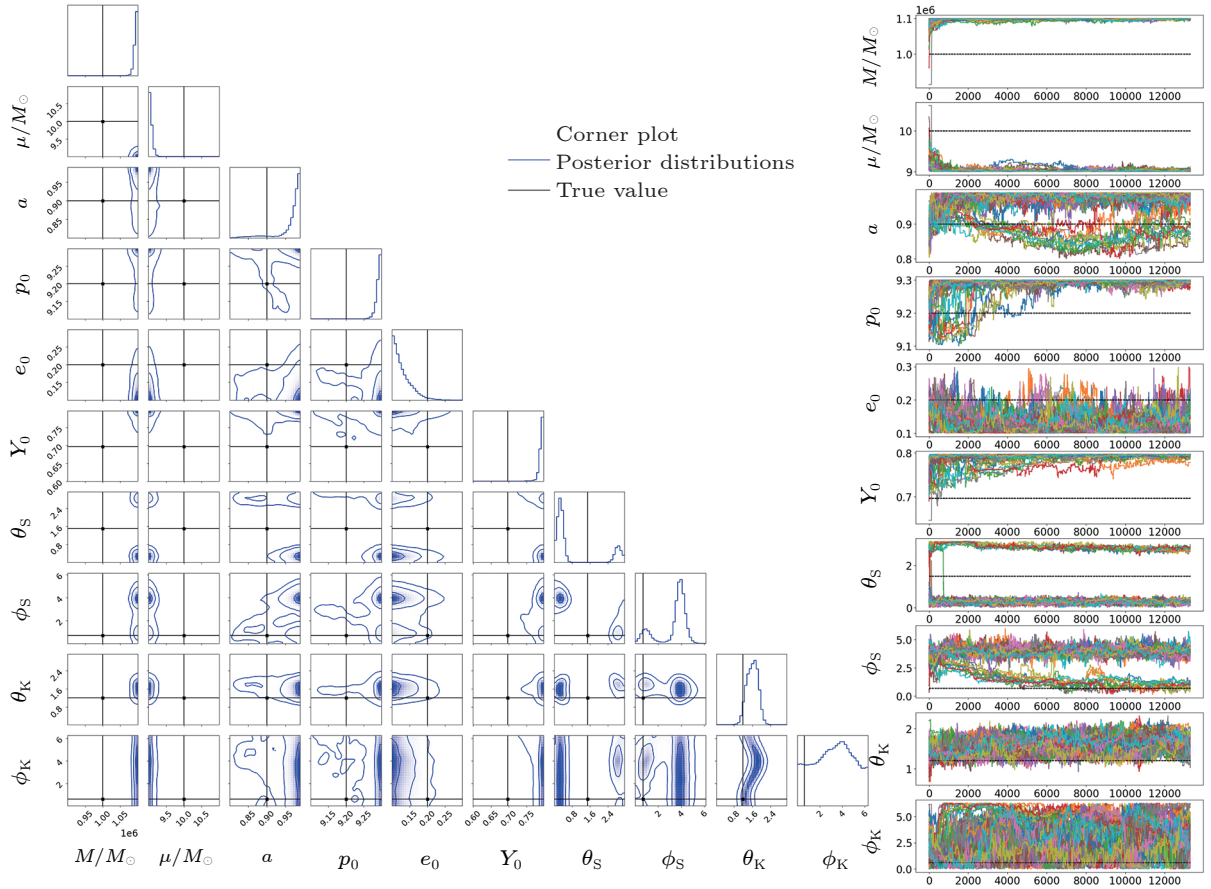

Fig. 2. MCMC with broad initialization. (Left) Posterior distributions from prior-sampled initialization. (Right) Chain convergence plot showing trapping in local optima. The starting points for the MCMC were obtained by random sampling from the prior distribution. The analyzed EMRI signal has a signal-to-noise ratio of approximately 67 and is based on two years of observational data. In the figure, the blue line represents the posterior probability distribution of the EMRI parameters, while the black line indicates the true parameter values.

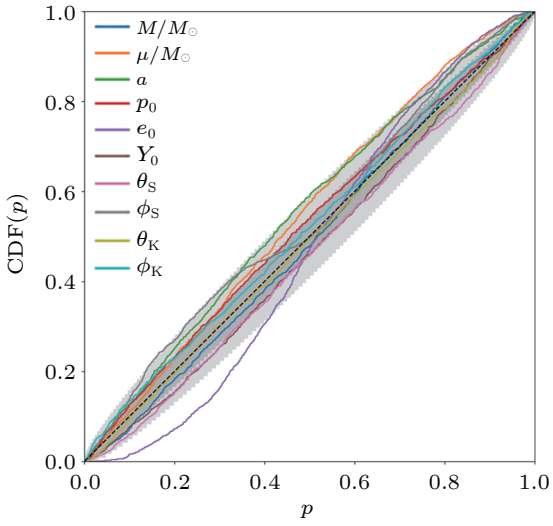

Fig. 3. P–P plot for a set of 1,000 injections, as analyzed by our machine learning model. We calculate the true percentile from the marginalized posterior and plot the CDF.

5. Discussion. The present study applies machine learning algorithms to EMRI posterior estimation, offering a novel approach for Bayesian parameter estimation of EMRI signals. Our method not only provides a comprehensive statistical framework for analyzing EMRIs but also establishes a new perspective in space-based GW astronomy. We have demonstrated the potential of machine learning in handling the low-frequency characteristics and vast parameter space of EMRI signals. Compared to traditional MCMC methods, our machine learning approach efficiently explores a broader parameter space, including regions that are challenging for conventional methods to reach. In terms of computational speed, our method delivers an order-of-magnitude improvement over MCMC techniques. This work provides new insights into enhancing existing parameter estimation methods, particularly in addressing complex signals characterized by non-local parameter degeneracies and multi-harmonic overlap. As machine learning technology continues to advance, we anticipate it will play an increasingly critical role in space-based GW detection, driving further progress in GW astronomy. Additionally, our findings suggest a promising approach for processing future real EMRI signals. Specifically, machine learning could be used to narrow the parameter space, followed by MCMC methods where the prior likelihood distribution is informed by the machine learning-derived posterior. While MCMC methods could theoretically explore extrema in machine learning-reduced spaces, there

Table 2. The table compares the parameters injected a prior in the real case with those recovered by FMPE and emcee. The recovered values are accompanied by their 68% confidence regions.

Parameter	Description	Injected value	FMPE	Eryn
$\log_{10}\mathcal{M}[M_\odot]$	Primary mass	6	$5.998^{+0.005}_{-0.005}$	$6.041^{+0.000}_{-0.001}$
$\mu[M_\odot]$	Secondary mass	10	$9.788^{+0.353}_{-0.313}$	$9.017^{+0.031}_{-0.013}$
a	Primary spin	0.9	$0.863^{+0.065}_{-0.042}$	$0.979^{+0.008}_{-0.028}$
p_0	Semi-latus rectum	9.2	$9.205^{+0.006}_{-0.067}$	$9.296^{+0.003}_{-0.008}$
e_0	Orbital eccentricity	0.2	$0.196^{+0.014}_{-0.014}$	$0.121^{+0.032}_{-0.016}$
Y_0	Orbital inclination parameter	0.7	$0.702^{+0.054}_{-0.055}$	$0.795^{+0.001}_{-0.004}$
θ_S[rad]	Polar angle of source orientation	1.5	$1.536^{+0.155}_{-0.141}$	$0.308^{+2.489}_{-0.104}$
ϕ_S[rad]	Azimuthal angle of source orientation	0.7	$1.200^{+2.665}_{-0.518}$	$3.824^{+0.385}_{-2.559}$
θ_K[rad]	Polar angle of orbital angular direction	1.2	$1.651^{+0.498}_{-0.495}$	$1.632^{+0.208}_{-0.265}$
ϕ_K[rad]	Azimuthal angle of orbital angular direction	0.6	$3.061^{+2.088}_{-1.982}$	$3.31^{+1.774}_{-2.115}$

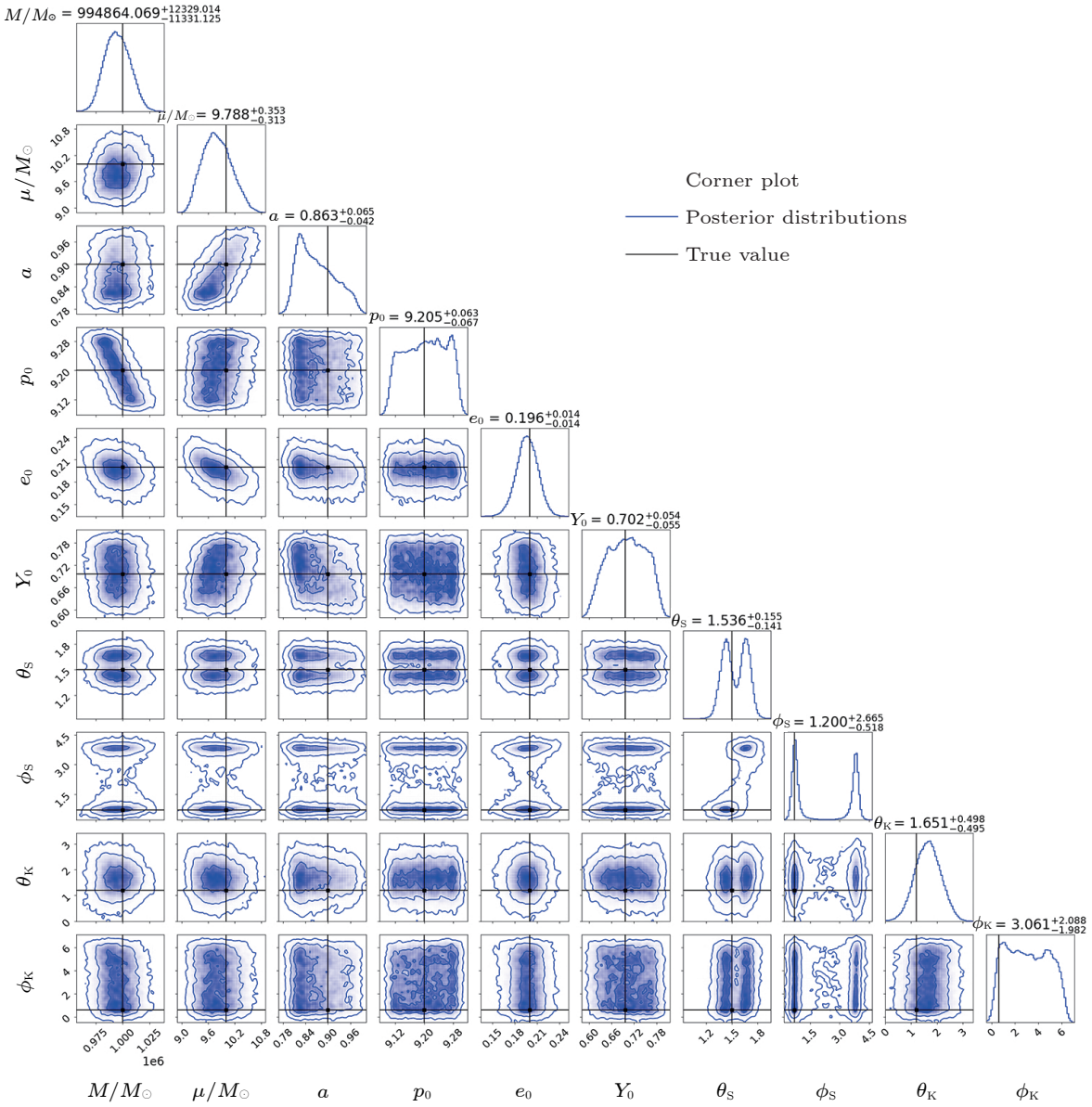

Fig. 4. FMPE posterior estimation broad initialization. The analyzed EMRI signal has a signal-to-noise ratio of approximately 67 and is based on two years of observational data. In the figure, the blue line represents the posterior probability distribution of the EMRI parameters, while the black line indicates the true parameter values.

are currently no published studies demonstrating successful MCMC sampling within our proposed prior range constraints. This will be the focus of our next-step research.

Acknowledgements. This work was supported by the National Key Research and Development Program of China (Grant Nos. 2021YFC2201901, 2021YFC2203004, 2020YFC2200100 and 2021YFC2201903), International Partnership Program of the Chinese Academy of Sciences (Grant No. 025GJHZ2023106GC). We also gratefully acknowledge the financial support from Brazilian agencies Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

[1] Aasi J, Abbott B P, Abbott R *et al.* 2015 *Class. Quant. Grav.* **32** 074001
[2] Abbott B P, Abbott R, Abbott T D *et al.* 2016 *Phys. Rev. Lett.* **116** 061102
[3] Abbott B P, Abbott R, Abbott T D *et al.* 2016 *Phys. Rev. Lett.* **116** 241103
[4] Abbott B P, Abbott R, Abbott T D *et al.* 2019 *Phys. Rev. X* **9** 031040
[5] Abbott R, Abbott T D, Abraham S *et al.* 2021 *Phys. Rev. X* **11** 021053
[6] Acernese F, Agathos M, Agatsuma K *et al.* 2015 *Class. Quant. Grav.* **32** 024001
[7] Abbott R, Abbott T D, Acernese F *et al.* 2023 *Phys. Rev. X* **13** 041039
[8] Kaiser A R and McWilliams S T 2021 *Class. Quant. Grav.* **38** 055009
[9] Amaro-Seoane P, Audley H, Babak S *et al.* 2017 arXiv:1702.00786 [astro-ph.IM]
[10] Danzmann K 1997 *Class. Quant. Grav.* **14** 1399
[11] Hu W R and Wu Y L 2017 *Natl Sci. Rev.* **4** 685
[12] Gong Y, Luo J, and Wang B 2021 *Nature Astron.* **5** 881
[13] Luo J, Chen L S, Duan H Z *et al.* 2016 *Class. Quant. Grav.* **33** 035010
[14] Korol V, Rossi E M, and Groot P J 2017 *ASP Conf. Ser.* **509** 529
[15] Kang Y, Liu C, and Shao L 2021 *Astron. J.* **162** 247
[16] Ren Z, Zhao T, Cao Z, Guo Z K, Han W B, Jin H B, and Wu Y L 2023 *Front. Phys.* **18** 64302
[17] Huang S J, Hu Y M, Korol V, Li P C, Liang Z C, Lu Y, Wang H T, Yu S, and Mei J 2020 *Phys. Rev. D* **102** 063021
[18] Klein A, Barausse E, Sesana A *et al.* 2016 *Phys. Rev. D* **93** 024003
[19] Ruan W H, Liu C, Guo Z K, Wu Y L, and Cai R G 2021 *Research* **2021** 6014164
[20] Wang H T, Jiang Z, Sesana A *et al.* 2019 *Phys. Rev. D* **100** 043003
[21] Babak S, Gair J, Sesana A, Barausse E, Sopuerta C F, Berry C P L, Berti E, Amaro-Seoane P, Petiteau A, and Klein A 2017 *Phys. Rev. D* **95** 103012
[22] Fan H M, Hu Y M, Barausse E, Sesana A, Zhang J d, Zhang X, Zi T G, and Mei J 2020 *Phys. Rev. D* **102** 063016
[23] Chen Z C, Huang F, and Huang Q G 2019 *Astrophys. J.* **871** 97
[24] Chen Z C, Huang Q G, Liu C, Liu L, Liu X J, Wu Y, Wu Y M, Yi Z, and You Z Q 2024 *JCAP* **2024** 022
[25] Liang Z C, Hu Y M, Jiang Y, Cheng J, Zhang J d, and Mei J 2022 *Phys. Rev. D* **105** 022001
[26] Berry C P L, Hughes S A, Sopuerta C F, Chua A J K, Heffernan A, Holley-Bockelmann K, Mihaylov D P, Miller M C, and Sesana A 2019 arXiv:1903.03686 [astro-ph.HE]
[27] Laghi D, Tamanini N, Del Pozzo W, Sesana A, Gair J, Babak S, and Izquierdo-Villalba D 2021 *Mon. Not. Roy. Astron. Soc.* **508** 4512
[28] Gair J R, Vallisneri M, Larson S L, and Baker J G 2013 *Living Rev. Rel.* **16** 7
[29] Barack L and Pound A 2019 *Rept. Prog. Phys.* **82** 016904
[30] Pound A and Wardell B 2021 arXiv:2101.04592 [gr-qc]
[31] Gair J R, Barack L, Creighton T, Cutler C, Larson S L, Phinney E S, and Vallisneri M 2004 *Class. Quant. Grav.* **21** S1595
[32] Chua A J K, Katz M L, Warburton N, and Hughes S A 2021 *Phys. Rev. Lett.* **126** 051102
[33] Katz M L, Chua A J K, Speri L, Warburton N, and Hughes S A 2021 *Phys. Rev. D* **104** 064047
[34] Babak S, Baker J G , Benacquista M J *et al.* 2010 *Class. Quant. Grav.* **27** 084009
[35] Chua A J K and Cutler C J 2022 *Phys. Rev. D* **106** 124046
[36] Babak S, Gair J R, and Porter E K 2009 *Class. Quant. Grav.* **26** 135004
[37] Glampedakis K and Babak S 2006 *Class. Quant. Grav.* **23** 4167
[38] Qiao X, Xia Z W, Pan Q, Guo H, Qian W L, and Jing J 2024 arXiv:2408.10022 [gr-qc]
[39] Gair J R, Porter E K, Babak S, and Barack L 2008 *Class. Quant. Grav.* **25** 184030
[40] Ali A, Christensen N, Meyer R, and Rover C 2012 *Class. Quant. Grav.* **29** 145014
[41] Ollie B, Gabriel Andres P, Niels W, Philip L, Lorenzo S, Chris K, Barry W, Adam P, Leanne D, and Jeremy M 2025 *Phys. Rev. D* **109** 124048
[42] Burke O, Piovano G A, Warburton N, Lynch P, Speri L, Kavanagh C, Wardell B, Pound A, Durkan L, and Miller J 2024 arXiv:2310.08927 [gr-qc]
[43] Saltas I D and Oliveri R 2023 arXiv:2311.17174 [gr-qc]
[44] Chua A J K 2022 *Phys. Rev. D* **106** 104051
[45] Burke O, Piovano G A, Warburton N, Lynch P, Speri L, Kavanagh C, Wardell B, Pound A, Durkan L, and Miller J 2024 *Phys. Rev. D* **109** 124048
[46] Huerta E A and Zhao Z 2021 *Advances in Machine and Deep Learning for Modeling and Real-Time Detection of Multi-messenger Sources, in Handbook of Gravitational Wave Astronomy*, Edited by Bambi C, Katsanevas S, and Kokkotas K D (Singapore: Springer Singapore) p. 1
[47] Zhao T, Shi R, Zhou Y, Cao Z, and Ren Z 2023 arXiv:2311.15585 [gr-qc]
[48] Zhao T, Lyu R, Wang H, Cao Z, and Ren Z 2023 *Commun. Phys.* **6** 212
[49] Zhang X T, Messenger C, Korsakova N, Chan M L, Hu Y M, and Zhang J D 2022 *Phys. Rev. D* **105** 123027
[50] Zhao T, Zhou Y, Shi R, Cao Z, and Ren Z 2024 *Phys. Rev. D* **109** 084054
[51] Yun Q, Han W B, Guo Y Y, Wang H, and Du M 2023 arXiv:2309.06694 [gr-qc]
[52] Katz M L, Bayle J B, Chua A J, and Vallisneri M 2022 *Phys. Rev. D* 106
[53] Martens W and Joffre E 2021 *J. Astronaut. Sci.* **68** 402
[54] Bayle J B, Bonga B, Caprini C, Doneva D, Muratore M, Petiteau A, Rossi E, and Shao L 2022 *Nat. Astron.* **6** 1334
[55] Speri L, Karnesis N, Renzini A I, and Gair J R 2022 *Nat. Astron.* **6** 1356
[56] Babak S, Gair J R, and Cole R H 2015 *Extreme Mass Ratio Inspirals: Perspectives for Their Detection*, (Springer International Publishing) p. 783

[57] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, and Chintala S 2019 arXiv:1912.01703 [cs.LG]
[58] Papamakarios G and Murray I 2018 arXiv:1605.06376 [stat.ML]
[59] Du M, Liang B, Wang H, Xu P, Luo Z, and Wu Y 2024 *Sci. China Phys. Mech. Astron.* **67** 230412
[60] Liang B, Du M, Wang H, Xu Y, Liu C, Wei X, Xu P, Qiang L, and Luo Z 2024 arXiv:2407.07125 [gr-qc]
[61] Dax M, Green S R, Gair J, Macke J. H, Buonanno A, and Schölkopf B 2021 *Phys. Rev. Lett.* 127
[62] Gebhard T D, Wildberger J, Dax M, Angerhausen D, Quanz S P, and Schölkopf B 2023 arXiv:2312.08295 [astro-ph.IM]
[63] Rezende D and Mohamed S 2015 *Proceedings of the 32nd International Conference on Machine Learning*, PMLR **37** 1530
[64] Papamakarios G, Nalisnick E, Rezende D J, Mohamed S, and Lakshminarayanan B 2021 *JMLR* **22** 1
[65] Lipman Y, Chen R T, Ben-Hamu H, Nickel M, and Le M 2022 arXiv:2210.02747 [cs.LG]
[66] Chen R T Q, Rubanova Y, Bettencourt J, and Duvenaud D 2019 arXiv:1806.07366 [cs.LG]
[67] Dax M, Wildberger J, Buchholz S, Green S R, Macke J H, and Scholkopf B 2023 arXiv:2305.17161
[68] Littenberg T B and Cornish N J 2023 *Phys. Rev. D* **107** 063004
[69] Katz M L, Karnesis N, Korsakova N, Gair J R, and Stergioulas N 2024 arXiv:2405.04690 [gr-qc]
[70] Whittle P 1957 *J. R. Stat. Soc. Ser. B (Methodol.)* **19** 38
[71] Ali A, Christensen N, Meyer R, and Röver C 2012 *Class. Quantum Grav.* **29** 145014
[72] Karnesis N, Katz M L, Korsakova N, Gair J R, and Stergioulas N 2023 *Mon. Not. R. Astron. Soc.* **526** 4814
[73] Foreman-Mackey D, Hogg D W, Lang D, and Goodman J 2013 *PASP* **125** 306