Decoding the Effects of Varying Frequency Bands on GW Source Characterization

Sena Kalabalık
Boğaziçi University, İstanbul, Turkey

Mentors: Rhiannon Udall, Lucy M. Thomas and Derek Davis
LIGO, California Institute of Technology, Pasadena

Final Report
LIGO Caltech SURF Program 2024

Gravitational wave (GW) science has opened new avenues for understanding astrophysical phenomena, with precise signal characterization being essential for interpreting these cosmic events. However, short-duration terrestrial noise transients, known as "glitches", complicate this task. A common strategy for mitigating the impact of glitches involves restricting the analysis to a reduced frequency band. In our work, we perform parameter estimation on artificial data, which includes both simulated waveforms and noise realizations. By examining different frequency ranges, we aim to characterize the typical effects of the frequency restrictions on the inference of the source parameters. Building on these findings, we generate a statistical test to systemize the application of these restrictions to glitch-affected real events. Additionally, we focus on evaluating the use of a DL model in the GW parameter estimation pipeline. By comparing results obtained with the DL model and the conventional pipeline, we explore the effectiveness of the DL approach in improving analysis speed and accuracy.

I. INTRODUCTION

Gravitational waves originate predominantly from the accelerated motion of massive objects, such as the orbital movement of black holes and neutron stars. This motion disturbs the fabric of space-time, leading to the propagation of waves in all directions from the source. As gravitational waves travel through space, they induce tiny expansions and contractions in the spatial dimensions they traverse. The measurements of this strain are detected by ground-based instruments operated by the Laser Interferometer Gravitational Wave Observatory (LIGO), Virgo, and KAGRA (LVK) Collaboration. These detectors are advanced Michelson interferometers, simply consisting of two perpendicular arms of equal length, each equipped with mirrors at their ends. A laser beam is split and sent down each arm, reflecting off the mirrors and then recombining at a central chamber. When a gravitational wave passes through, it causes the lengths of the arms to change slightly. This alters the interference pattern when the laser beams recombine, allowing the detection of the gravitational waves in the form of weak signals.

One of the major sources of gravitational waves are binary black hole mergers and neutron star mergers, which fall under the category of compact binary coalescences (CBC).The characterization of these waves is crucial since they carry promising information about the nature of the compact objects. Accurate source characterization depends on the assumption that detector noise behaves as stationary Gaussian noise, which is violated when there are short-duration noise transients, called "glitches" present in the signal [1]. Glitches can arise from various reasons, such as instrumental artifacts or environmental disturbances, and their sources are not typically directly identifiable. Therefore, it is common for glitches to occur unpredictably in the signal. In [2], researchers reported that during the third observing run (O3) [3, 4], the median rate of glitches in the LIGO and Virgo detectors exceeded one per minute for much of the run, suggesting that coincidences between glitches and gravitational wave signals are likely to occur relatively frequently. Since these glitches corrupt the signal and invalidate the noise assumptions for typical gravitational wave source characterization processes, their identification and mitigation are crucial for accurate analysis.

The process of identifying and subtracting glitches is not straightforward, leading to the adoption of various strategies. One of the most common and sophisticated methods is the BayesWave algorithm [5]. This algorithm assumes that the strain data comprises Gaussian noise, a gravitational wave signal, and a glitch, which is modeled as a sum of sine-Gaussian wavelets. As the result of the algorithm, a posterior distribution of time series of the glitch is obtained, and the mitigation is conducted by randomly selecting a sample from the posterior, and subtracting it from the data [2]. Despite the common use of the BayesWave algorithm, various studies, including [6] and [3], have demonstrated that the method may potentially leave residual artifacts within the signal, which is likely due to the probability of a randomly drawn sample wavelet accurately capturing the precise characteristics of the glitch is low.

Another approach for glitch mitigation is gwsubstract algorithm [7], which uses information from auxiliary channels. However, the accuracy of this subtraction method depends on the accuracy of the auxiliary sensor and the transfer function estimate [2, 6]. The mentioned systematic and statistical uncertainties of these two methods poses the risk that even after BayesWave or gwsubtract algorithm is used, the data can still be

undersubstracted, as shown in [3, 6].

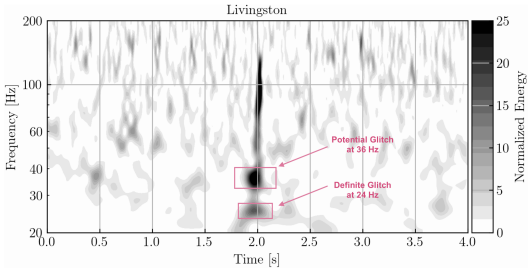


FIG. 1: The spectrogram demonstrates the GW191109 event data for Livingston in the frequency domain, showing features that indicate the data is glitch-affected. It highlights the definite glitch at 24 Hz and the potential glitch at 36 Hz, marked by pink annotations, as noted in [8].

Although the glitches are commonly dealt with by eliminating data in the time domain, it is possible to come across cases where the glitch only affects a specific frequency range, as shown in Figure 1. In such instances, addressing data quality issues might involve restricting the analysis to a narrower frequency range [9]. We refer to these restrictions as "frequency cuts".

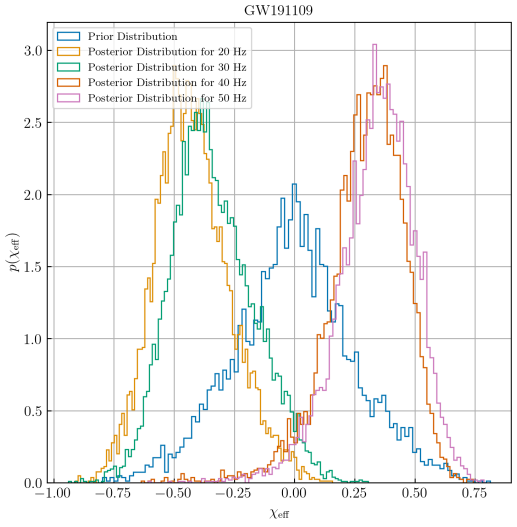


FIG. 2: The plot shows posteriors for $\chi_{\rm eff}$ at 20, 30, 40, and 50 Hz, along with the prior. The transition from 20 Hz and 30 Hz to 40 Hz and 50 Hz is particularly noteworthy due to the drastic shift observed. This result was first presented in [2].

A notable example of frequency cuts for glitch mitigation is the case of GW191109. As illustrated in Figure 2, the posterior distributions for $\chi_{\rm eff}$ parameter shift significantly when the frequency cut at 40 Hz is applied. This behavior is unusual because losing data due to a frequency cut typically results in posteriors becoming less informative and more prior-like, rather than showing such a drastic change. Since frequency cuts are not applied through a systematic process, it is difficult to determine whether the analysis at 20 Hz (full-band) or 40 Hz is more reliable.

Therefore, our work proposes the development of a robust test to systematize the application of frequency cuts, determining their necessity based on the impact on parameter estimation (PE). A crucial point to emphasize is that we only need to subtract the glitch—by applying frequency cuts—if it directly affects the parameter estimation process. If the glitch does not introduce bias or alter the posterior distributions, the frequency cuts are deemed unnecessary. In our analyses, preserving information is critical, and we aim to avoid sacrificing valuable data unless it is clear that the glitch introduces bias.

To conduct such a diagnostic test, an effective method is to use a reference distribution as a benchmark for real cases. We propose utilizing generated injections, which consist of a pure gravitational wave signal combined with Gaussian noise, in line with typical assumptions. As the reference distributions become more reliable with an increasing number of generated injections, the computational cost of inferring these injections will rise accordingly. Therefore, we seek a parameter estimation method that is faster than our conventional tools, such as Bilby [10], to manage this increased computational demand efficiently. To address this, we employ DINGO [11], a simulation-based inference model that utilizes a likelihood-free approach. DINGO is a neural posterior estimator (NPE) network, which has previously demonstrated the capability to produce inferences rapidly—typically within seconds—while maintaining accuracy comparable to Bilby and taking several days for training [11].

With this context in mind, the structure of the report is as follows: In Section II, we describe Bayesian inference for CBC using the two methods: Bilby and DINGO. Additionally, we introduce and characterize the κ_D statistical test. In Section III, we apply the Bilby method to 50 injections drawn from GW191109, inspecting the posterior evolutions with frequency cuts and generating reference distributions for our statistical test. Our findings indicate that GW191109 falls significantly outside the expected distributions for 40 and 50 Hz, suggesting anomalies in the data below 40 Hz. In Section IV, we utilize the DINGO method to replicate the described methodology and compare the results to those obtained from Bilby. We find that while DINGO has notable difficulties in accurately representing the posteriors, its statistical test results exhibit a consistency with those derived from Bilby for GW191109.

II. METHODOLOGY

A. Bayesian Inference: Using Bilby

In order to perform source characterization for CBC signals, we conduct Bayesian inference with the Bilby pipeline. Bayesian inference applies Bayes' theorem, which allows for estimating the posterior distributions of the source parameters by combining the observed data d, the likelihood function $p(d|\theta)$, and the prior distribution $p(\theta)$ for the parameters. Bayes' theorem is expressed as:

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}, \tag{1}$$

where $p(\theta|d)$ is the posterior distribution, $p(d|\theta)$ is the likelihood, $p(\theta)$ is the prior distribution, and $p(d)$ is the evidence, which acts as a normalizing factor. The observed data d consists of both the gravitational wave signal and the detector noise.

In typical parameter estimation pipelines for gravitational waves, the detector noise is assumed to be stationary and Gaussian. This assumption simplifies the likelihood function, as it implies that the difference between the data and the waveform should be Gaussian noise. The noise is modeled by the power spectral density (PSD), which characterizes the distribution of noise across different frequencies at the time of detection. The waveform model, $h(\theta)$, and the noise model forms the basis of the likelihood function.

The likelihood function in the frequency domain is typically given by:

$$\ln p(d|\theta) = -\frac{1}{2}\sum_k \left(\frac{|d_k - h_k(\theta)|^2}{\sigma_k^2} + \ln\left(2\pi\sigma_k^2\right) \right),$$

where k represents the frequency bin index, σ_k is the noise amplitude spectral density, and $h(\theta)$ is the waveform model [10]. This likelihood function is crucial in evaluating how well a given set of parameters θ explains the observed data, taking into account the noise characteristics modeled by the PSD.

Since the evaluation of the likelihood function requires the generation and comparison of numerous waveform models across a large parameter space, the process is computationally expensive and time-consuming. A vast number of waveform evaluations are needed to sample from the posterior distributions, significantly increasing the computational cost.

For this analysis, we focus on the effective spin parameter, $\chi_{\rm eff}$ which represents the mass-weighted projection of the individual component spins along the orbital angular momentum vector, and is defined as:

$$\chi_{\rm eff} = \frac{(m_1\vec{a}_1 + m_2\vec{a}_2)\cdot\hat{L}}{m_1 + m_2},$$

where m_1 and m_2 are the component masses of the binary, $\vec{a}_1$ and $\vec{a}_2$ are their respective dimensionless spin vectors, and $\hat{L}$ is the unit vector along the orbital angular momentum. A non-zero value of $\chi_{\rm eff}$ can indicate aligned or anti-aligned spins with the orbital angular momentum. $\chi_{\rm eff}$ is a crucial parameter in astrophysics as it provides insights into the binary's formation history. Aligned spins often suggest isolated binary evolution, whereas anti-aligned spins may indicate a dynamical formation origin [8].

In this study, we model waveforms using the IMRPhenomXPHM approximant [12] and employ prior distributions aligned with the GW191109 event. The waveform model is conditioned on these prior distributions, which integrate our prior understanding of the source parameters. Specifically, the chirp mass is uniformly distributed between 35 and 175 $M_\odot$:

$$35 \leq \mathcal{M}_c \leq 175$$

The mass ratio q ranges from 0.25 to 1:

$$0.25 \leq q \leq 1$$

The spin magnitudes a_1 and a_2 are uniformly distributed between 0 and 0.99:

$$0 \leq a_1, a_2 \leq 0.99$$

The tilt angles θ_1 and θ_2 follow a sine distribution. The luminosity distance d_L is uniformly sampled between 100 and 7000 Mpc:

$$100 \leq d_L \leq 7000$$

For DINGO, we use a different prior at the inference time, sampling from a uniform comoving volume distribution between 100 and 2000 Mpc:

$$100 \leq d_c \leq 2000$$

The PSD used is also specific to GW191109.

For generating injections, Bilby is used with the same priors and PSD, along with a maximum frequency of 512 Hz and a duration of 4 seconds.

B. Bayesian Inference: Using DINGO

Unlike the classical approaches, DINGO trains a neural network to approximate the Bayesian posterior distribution over source parameters using simulated data rather than direct likelihood evaluations. The process involves generating a multitude of simulated datasets, each with its corresponding parameters, which are used to train a

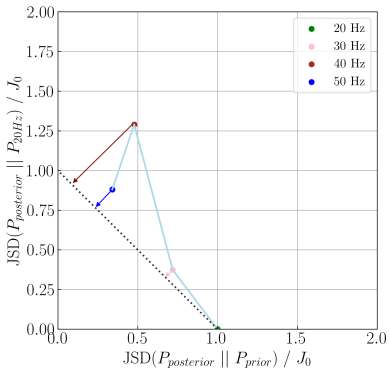


FIG. 3: The figure illustrates the κ_D values for the 30 Hz (pink), 40 Hz (red), and 50 Hz (blue) posteriors. The dotted line represents the imaginary line $J_0 = x + y$, showing the expected evolution of the JS distances as the frequency cutoff increases. This line reflects the progression from the full-band analysis with a 20 Hz cutoff to the most extreme case, where the posterior includes no additional information and matches the prior. In this scenario, the JS distance between the posterior and the 20 Hz posterior is equal to J_0. The light blue curve tracks the evolution of the JS distances of a single injection as the frequency cutoff is raised.

specific type of neural network known as a normalizing flow. This approach integrates the prior by sampling parameters and incorporates the likelihood through data simulation. [11]

With DINGO, we also incorporate importance sampling as a key component to enhance our results. In DINGO's importance sampling strategy, DINGO-IS [13], the approach of reweighting allows samples drawn from one posterior distribution to be used for estimating a different posterior distribution. Given a set of n samples $\theta_i \sim q(\theta|d)$ (the proposal distribution), each sample is assigned an importance weight:

$$w_i = \frac{p(d|\theta_i)p(\theta_i)}{q(\theta_i|d)},$$

where $p(\theta_i|d)$ represents the target posterior distribution. The accuracy of this sampling process is evaluated using the "effective number of samples", denoted as n_{eff}, which is estimated by:

$$n_{\text{eff}} \approx \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2} = n \left(\frac{1}{1 + \left(\frac{\sigma_w}{\bar{w}}\right)^2} \right).$$

And the sample efficiency defined as:

$$\epsilon = \frac{n_{\text{eff}}}{n} \in (0, 1].$$

High individual weights relative to the average weight $\bar{w}$ increase the variance σ_w^2, which reduces n_{eff}. [13, 14]

Low efficiency can often be caused by the overlap between regions of low posterior for the approximate model and regions of high posterior for the target model. In such cases, this mismatch results in disproportionately high weights for certain samples [14]. Consequently, the final distribution becomes dominated by a small subset of high-weight samples, while other samples contribute minimally. This effect reduces n_{eff} and means fewer samples are effectively representing the target distribution.

C. Statistical Analysis: κ_D Test

To evaluate whether applying frequency cuts is necessary, we develop a statistical test called "κ_D". The purpose of this test is to examine how posterior distributions evolve with different frequency limits and assess whether frequency cuts mitigate bias in the PE process.

The process begins by using the preferred PE method to infer the injections and obtain posterior distributions for the χ_{eff} parameter across different frequency cut-offs, specifically at 20, 30, 40, and 50 Hz. We then compare these distributions in two key ways:

1. **Jensen-Shannon (JS) Distance Between Posteriors and Prior:** We compute the JS distance between each of the posterior distributions (from 20, 30, 40, 50 Hz) and the prior distribution for χ_{eff}. The JS distance is a measure of similarity between two probability distributions, providing a symmetric and finite divergence score. Mathematically, for two distributions P and Q, the JS distance is defined as:

$$JSD(P||Q) = \frac{1}{2}\left(D_{\text{KL}}(P||M) + D_{\text{KL}}(Q||M)\right)$$

where M is the point-wise mean of P and Q, and D_{KL} is the Kullback-Leibler divergence. We computed the JS distance using the Scipy implementation [15]. We use this distance to quantify how

much the posterior shifts away from the prior as we change the frequency cut.

2. **JS Distance Between Posteriors and the 20 Hz Posterior:** The second step is to compute the JS distance between each posterior (30, 40, 50 Hz) and the 20 Hz posterior. This helps identify how the posteriors change as we increase the frequency cut-off.

Next, we plot these JS distances on a two-dimensional graph, where the x-axis represents the JS distance between the posterior and the prior, and the y-axis represents the JS distance between the posterior and the 20 Hz posterior, as shown in Figure 3. We then draw an imaginary line, $x + y = J_0$, where $J_0 = JSD(\text{posterior}_{20}, \text{prior})$. This line represents the boundary that posteriors of unbiased data should generally follow, while allowing reasonable deviations.

1. *Computation of κ_D*

To assess the behavior of posteriors at different f_{min} values, we compute the distance of each point (x, y) on the plot (where x and y are the JS distances) from the J_0 line, as illustrated in the same Figure 3. This distance measures how far the posteriors at each frequency cut deviate from the expected bias-free evolution. We normalize this distance by dividing it by J_0, producing the 'κ_D' statistic for each case.

2. *Interpretation of κ_D*

Under the null hypothesis—where the injections provide a reference distribution containing only CBC signals and Gaussian noise without any glitches—the κ_D values should follow a predictable distribution. If the test case exhibits no bias in the posteriors, the κ_D values will align with this expected distribution. However, if the bias in the posteriors is reduced by applying a frequency cut at a particular f_{min} value, the κ_D value will diverge significantly from the null distribution because the posteriors are changing drastically when the bias affecting the PE process is reduced.

Thus, if the κ_D test shows significant divergence, it suggests that the frequency cut was beneficial for reducing bias. Conversely, if the κ_D values remain close to the expected distribution, this implies that the frequency cut was unnecessary, as the glitch was likely not affecting the parameter estimation process. The overarching goal is to avoid unnecessary loss of information by applying frequency cuts only when they provide a clear benefit in reducing bias.

III. APPLICATION OF κ_D TEST WITH BILBY

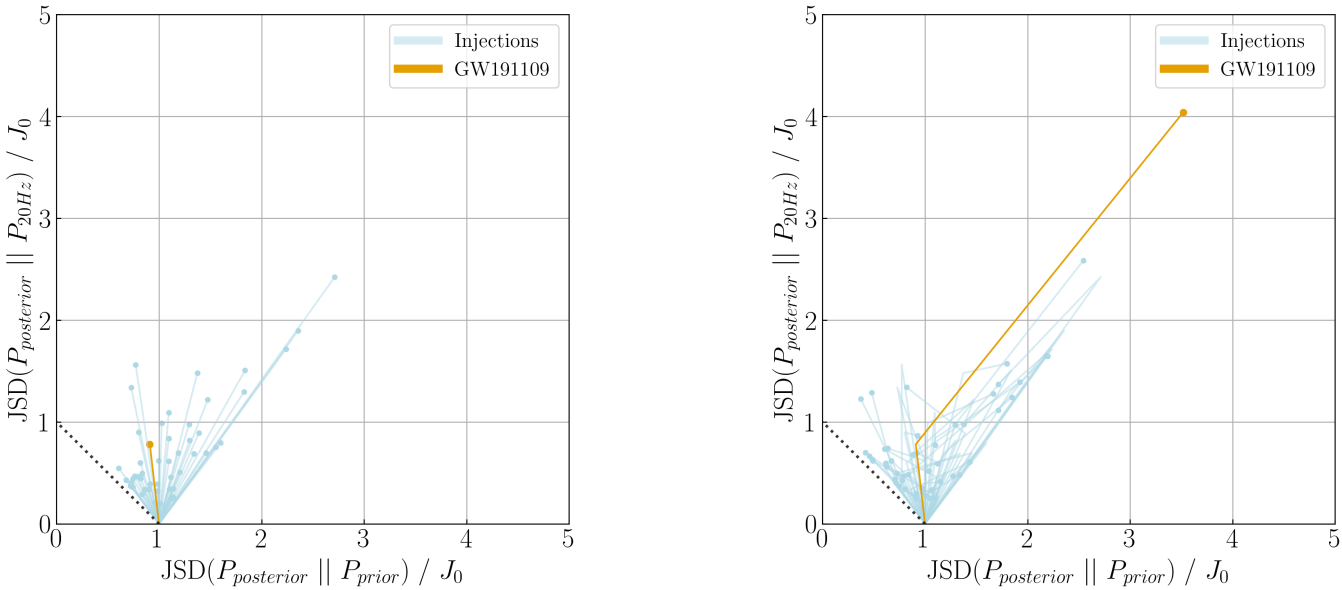
For the application, we evaluate the GW191109 event and compare it against reference distributions of κ_D, created using 50 injections. To accomplish this, we perform inference on the GW191109 event with Bilby and compute the JS distances for the real data. These computed JS distances are then compared with those for the 50 injections, as outlined in Section II C. The traces of these JS distances are visualized in Figure 4. Subsequently, we compute κ_D values for the real data and compare them against the reference distributions by examining histograms, as shown in Figure 5.

The results suggest that the frequency range that coincides with the potential glitch shown in Figure 1, which is the 30-40 Hz range, may be introducing bias, as evidenced by the observed significant divergence in κ_D values when the frequency cut is changed from 30 Hz to 40 Hz. The κ_D values at 30 Hz fall within the reference distribution, indicating consistency between the data at $f_{\text{min}} = 20$ Hz and $f_{\text{min}} = 30$ Hz. This implies that either both cases are unbiased or they exhibit a similar degree of bias. Likewise, the close correspondence between the JS distance and κ_D results for 40 Hz and 50 Hz suggests that any variation between these intervals is negligible and primarily due to information loss, rather than the introduction or elimination of bias within this interval. Although the disruption in consistency reveals clear issues in the 30-40 Hz frequency range, we still cannot assert with confidence that using a 40 Hz cut is more reliable than full-band analysis. This investigation is still in its early stages, and further evidence is required to assess the reliability of these analyses fully. Our findings are consistent with previous tests suggesting potential anomalies in the 30-40 Hz range, as noted in [8], highlighting the need for continued scrutiny to better understand the data and its implications.

IV. APPLICATION OF κ_D TEST WITH DINGO

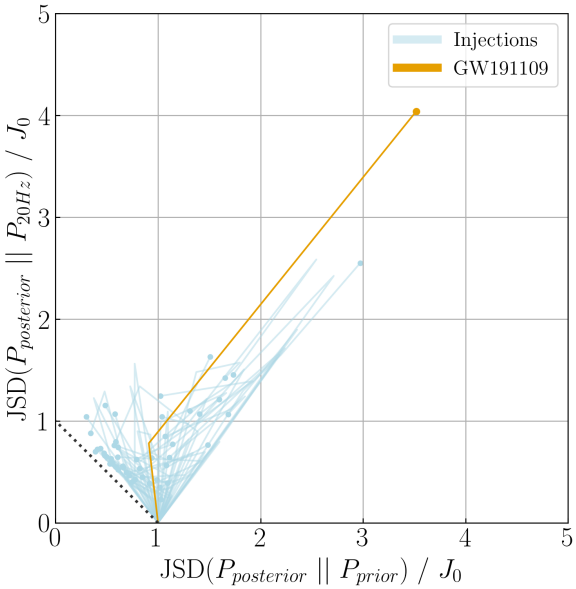
To utilize DINGO as our rapid parameter estimation pipeline, we trained the network with a dataset consisting of 4 million waveforms. The training incorporated frequency masking using both DINGO and DINGO-BNS [16]. In this setup, frequency cuts were applied to the waveforms before they were processed by the embedding network. During the training phase, random frequency masking was employed between the specified upper f_{min} value and the actual minimum frequency of the data, enabling the network to adapt to variations within this frequency range.

Importance sampling was also employed to enhance the network's performance. The results demonstrate that all injections have an efficiency of less than 0.15%. This suggests that the current model is not sufficiently accurate in capturing the target posterior distributions. Fig. 8 exemplifies the efficiency and IS weights for a single injec-



(a) 30 Hz

(b) 40 Hz



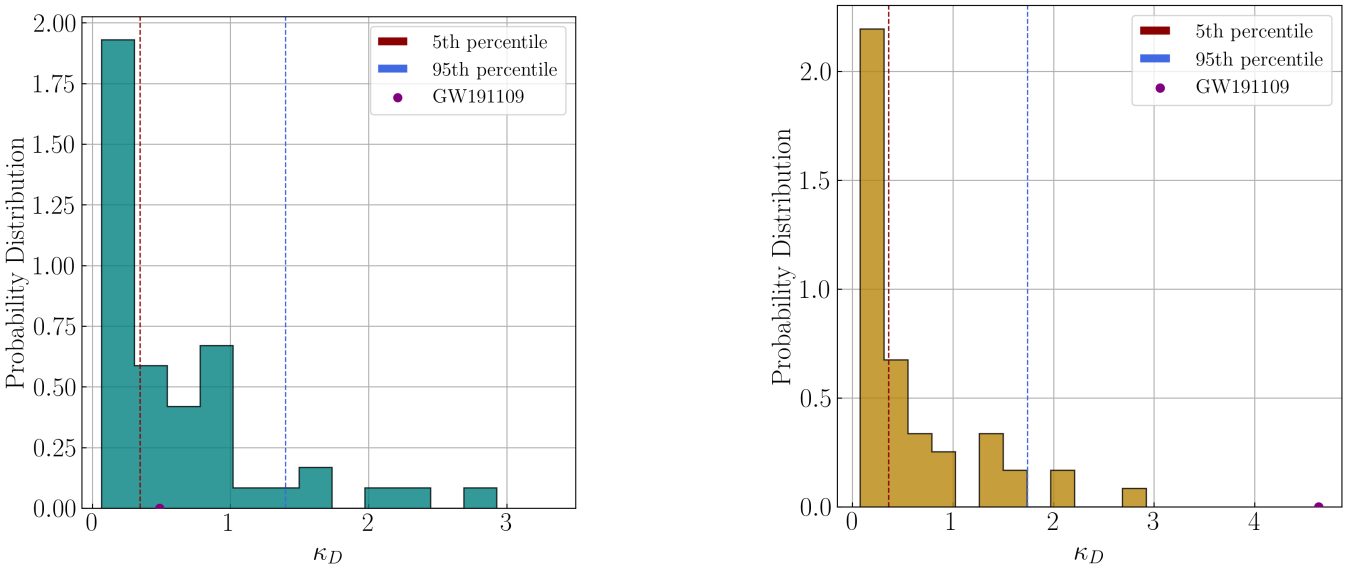
(c) 50 Hz

FIG. 4: Light blue lines illustrate the evolution of JS distances for the injections inferred with Bilby, while the golden line represents the GW191109 event. The significant deviation of the golden line from the injections at 40 and 50 Hz suggests that the posterior evolution with these frequency cut-offs diverges notably from the expected behavior of unbiased data.

tion. The posterior distributions obtained for the single injection are also shown in Fig. 9.

The low efficiency observed necessitates the use of a large number of samples to achieve meaningful and interpretable results. However, this approach demands considerable computational effort, which negates the potential computational advantages of DINGO compared to Bilby. As such, despite its intended purpose of providing faster PE, DINGO did not show a significant computational cost benefit over Bilby for our analysis.

Subsequently, we computed the JS distances and κ_D statistics using the injections inferred with the DINGO network, employing the same methodology as used with Bilby. The results of these computations are visualized



(a) 30 Hz

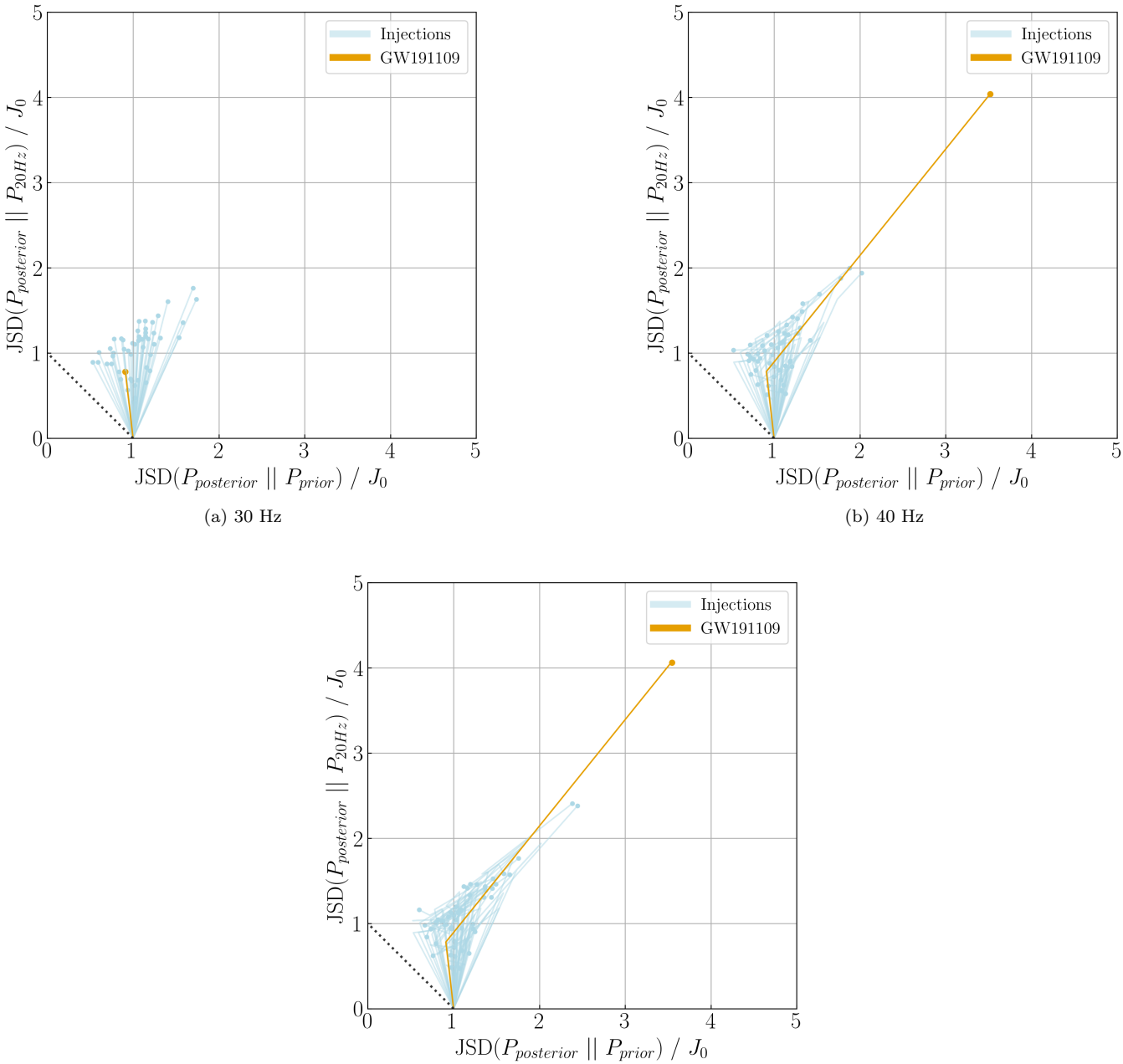
(b) 40 Hz

(c) 50 Hz

FIG. 5: The histograms illustrate the distributions of κ_D values for injections with f_{min} set at 30, 40, and 50 Hz, along with the κ_D value for GW191109. Vertical lines represent the 5th and 95th percentiles. For 30 Hz, the κ_D value falls within the reference distribution, suggesting that the data is consistent within the 20–30 Hz range. In contrast, the κ_D value for 40 Hz is significantly outside the reference distribution, indicating a marked change in the posteriors when applying a 40 Hz frequency cut. The κ_D value for 50 Hz is similar to that for 40 Hz, suggesting that the posteriors do not exhibit significant changes between the 40 and 50 Hz intervals.

in Figs. 6 and 7, respectively.

As observed in the visualizations, the JS results for the injections, while not closely resembling those of Bilby, demonstrate an agreement in detecting the divergence of GW191109 from the injections in our analysis. The reference distributions generated for κ_D are crucial for the reliability of this test. The fact that DINGO's distributions do not align well with those produced by Bilby is a notable concern, as this misalignment suggests discrepancies in how DINGO models the underlying data. However, despite these differences, both pipelines classify the real data consistently, indicating that the overall



(a) 30 Hz

(b) 40 Hz

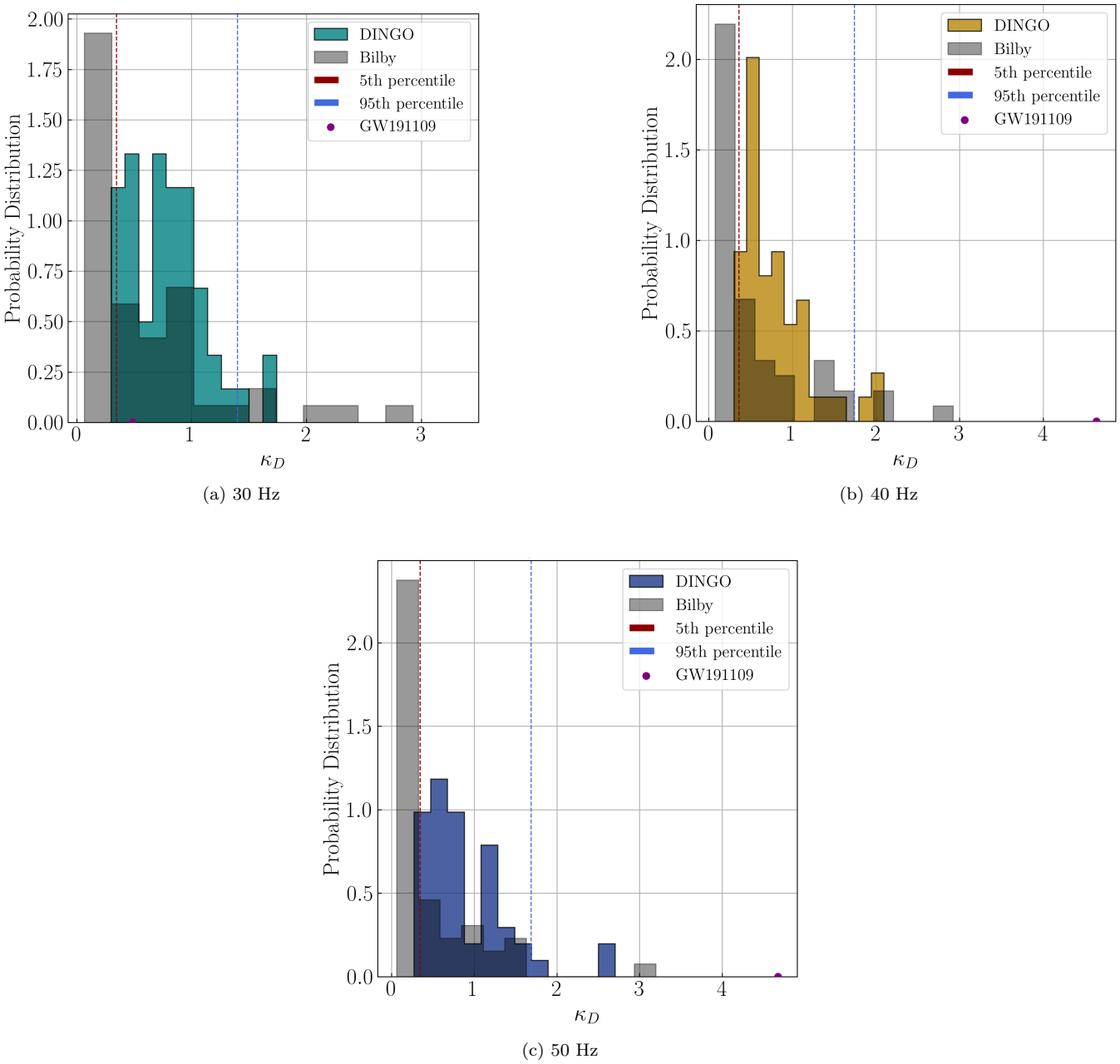
(c) 50 Hz

FIG. 6: Light blue lines illustrate the evolution of JS distances for the injections inferred with DINGO, while the golden line represents the GW191109 event. Although the JS distance evolution of the posteriors of the individual injections does not align closely with the results from Bilby, the golden line shows a similar significant divergence from the injections, consistent with the trend observed in the Bilby analysis.

test outcome remains unchanged for comparative analysis. This suggests a tentative possibility that DINGO may exhibit lower performance for individual inferences; however, it appears to show a tendency toward correctness when applied to a population of data. Nevertheless, this notion remains speculative and lacks sufficient evidence to draw firm conclusions.

V. CONCLUSIONS

In this study, we generated simulated data and conducted PE using two methods: Bilby and DINGO. The DINGO model was trained on a dataset constructed in accordance with the priors and noise PSD from the



(a) 30 Hz

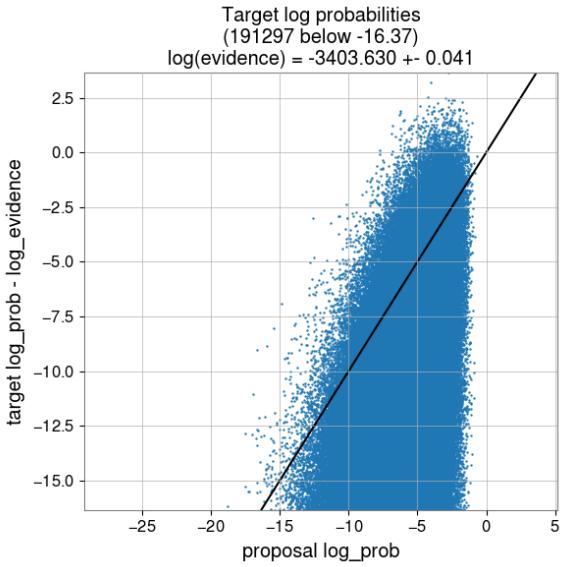
(b) 40 Hz

(c) 50 Hz

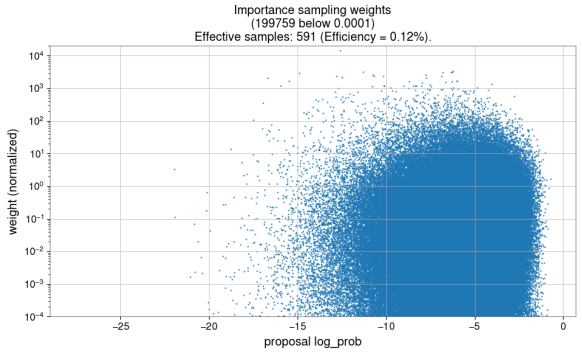
FIG. 7: The histograms display the distributions of κ_D values for injections with f_{min} set at 30, 40, and 50 Hz, alongside the κ_D value for GW191109. The gray histograms represent Bilby's results. Although DINGO's distributions do not align closely with Bilby's, both pipelines yield consistent classifications regarding the placement of the real data within or outside the distributions across different f_{min} values. This suggests that, despite differences, the comparative analysis provides similar outcomes for real data classification.

GW191109 event. We compared the performance of both pipelines in the inference of parameters for our simulated injections and observed that DINGO demonstrated notably low efficiency. This result suggests that further refinement of the network is necessary to achieve accurate and reliable analysis.

We further created and applied a statistical test to systematically evaluate the impact of frequency cuts on parameter estimation. By applying both inference methods independently, we identified that the inconsistency in the results primarily arises within the 30-40 Hz frequency range. This suggests a possible shift in the behavior of the parameter estimation process in this range, leading to a disruption in consistency. While these findings high-



(a) Plot illustrating the relationship between the target and proposal distributions. The plot shows that the samples are widely scattered, whereas, for accurate inference, the points should ideally align closely with the reference line.



(b) Plot illustrating the importance sampling weights for the posterior distribution of the selected injection. The plot reveals that most samples have low weights, while a few samples exhibit significantly higher weights, identifying them as the effective samples. Low sampling efficiency reflects the suboptimal performance of the network.

FIG. 8: The figure corresponds to the statistics for the inference of a single injection at the minimum frequency limit of 20 Hz; however, similar features are observed across all frequency limits.

light the need for further investigation, the current results are insufficient to definitively determine whether a full-band analysis or the application of a 40 Hz frequency cut offers a more reliable approach for analyzing the GW191109 event.

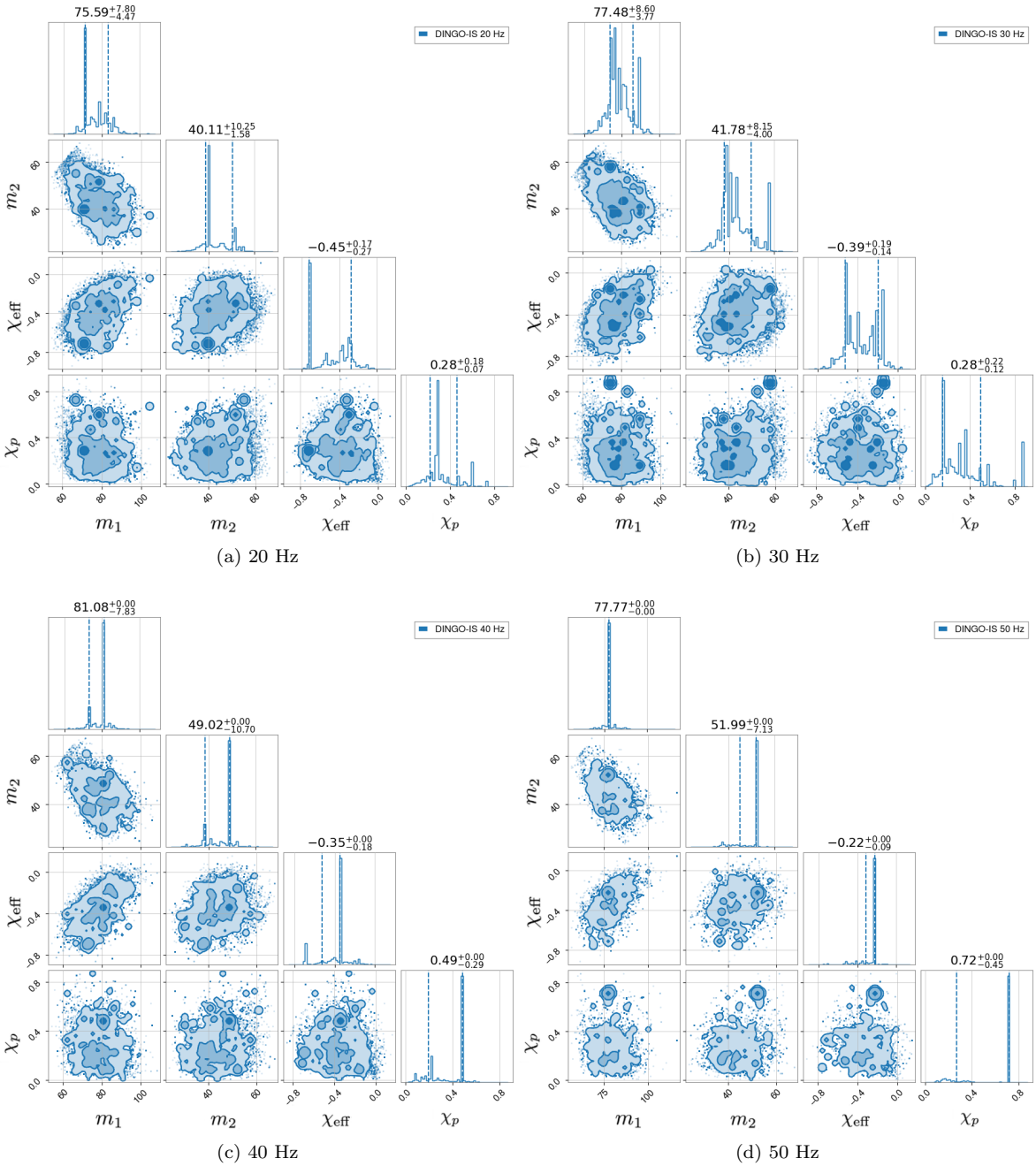
Our results highlight the complexities involved in applying frequency cuts for mitigating glitches in gravitational wave signals. Systematizing this process is of paramount importance, as the availability of high-quality, credible data is essential for making precise inferences about the source parameters—especially the effective spin parameter χ_{eff}, which plays a critical role in determining the formation origin of the binary system.

Moreover, while our analysis was based on 50 injections, we recognize that a larger number of injections may be required to establish more robust reference distributions for each event. This reinforces the continued need for a fast and user-friendly PE pipeline, both for the purposes of our study and for similar research efforts, such as those referenced in [8]. Although DINGO shows considerable promise, our findings suggest that its current implementation did not deliver optimal efficiency in the context of our analysis.

Despite being in its preliminary stages, our work presents a promising framework for systematically evaluating the frequency-cutting process in cases involving complex and uncertain glitches. With further investigation and the use of a more efficient pipeline, we anticipate that this method will yield more definitive findings, particularly in upcoming observing runs of the LVK Collaboration. The continued refinement of this framework will be crucial in enhancing its capacity to address the challenges posed by the subtle intricacies of glitch mitigation.

VI. ACKNOWLEDGEMENTS

I would like to extend my thanks to my mentors, Rhiannon Udall, Lucy M. Thomas, and Derek Davis, for their guidance, support, and valuable input throughout this process. I'm also grateful to the LIGO Laboratory team at Caltech for their valuable contributions and for fostering a collaborative and inclusive environment. I appreciate the opportunities provided by the Caltech Student-Faculty Program and the LIGO Summer Undergraduate Research Fellowship, which made this experience possible. Additionally, I acknowledge the DINGO Developers for their helpful contributions. This research was supported by the NSF Research Experience for Undergraduates (REU), the LIGO Summer Undergraduate Research Fellowship, and Caltech's Student-Faculty Programs.



(a) 20 Hz (b) 30 Hz

(c) 40 Hz (d) 50 Hz

FIG. 9: Plots illustrating the posterior distributions obtained using DINGO-IS for the m_1, m_2, χ_{eff}, and χ_p parameters. While a slight shift in χ_{eff} is observed as f_{min} is increased, the results are not sufficiently reliable, indicating that the network did not accurately model the target posterior distributions.

[1] Chris Pankow et al. Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817. *Phys. Rev. D*, 98(8):084016, 2018.

[2] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton. Subtracting glitches from gravitational-wave detector data during the third ligo-virgo observing run. *Classical and Quantum Gravity*, 39(24):245013, 2022.

[3] R. Abbott et al. GWTC-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run. *Phys. Rev. X*, 11:021053, 2021.

[4] R. Abbott et al. Gwtc-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run. *Phys. Rev. X*, 13:041039, 2023.

[5] Neil J. Cornish, Tyson B. Littenberg, Bence Bécsy, Katerina Chatziioannou, James A. Clark, Sudarshan Ghonge, and Margaret Millhouse. Bayeswave analysis pipeline in the era of gravitational wave observations. *Phys. Rev. D*, 103:044006, Feb 2021.

[6] Ethan Payne, Sophie Hourihane, Jacob Golomb, Rhiannon Udall, Derek Davis, and Katerina Chatziioannou. Curious case of GW200129: Interplay between spin-precession inference and data-quality issues. *Phys. Rev. D*, 106:104017, 2022.

[7] Derek Davis, Thomas Massinger, Andrew Lundgren, Jennifer C Driggers, Alex L Urban, and Laura Nuttall. Improving the sensitivity of advanced ligo using noise subtraction. *Classical and Quantum Gravity*, 36(5):055011, 2019.

[8] Rhiannon Udall, Sophie Hourihane, Simona Miller, Derek Davis, Katerina Chatziioannou, Max Isi, and Howard Deshong. The anti-aligned spin of gw191109: glitch mitigation and its implications. *arXiv:2409.03912*, page 19, 2024.

[9] Derek Davis and Marissa Walker. Detector characterization and mitigation of noise in ground-based gravitational-wave interferometers. *Galaxies*, 10(1):12, 2022.

[10] Gregory Ashton, Moritz Hübner, Paul D. Lasky, Colm Talbot, Kendall Ackley, Sylvia Biscoveanu, Qi Chu, Atul Divakarla, Paul J. Easter, and Boris Goncharov. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241(2):27, 2019. Published 2019 April 1. © 2019. The American Astronomical Society. All rights reserved.

[11] Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-time gravitational wave science with neural posterior estimation. *Phys. Rev. Lett.*, 127:241103, 2021.

[12] Geraint Pratten, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, Maite Mateu-Lucena, Rafel Jaume, Maria Haney, David Keitel, Jonathan E. Thompson, and Sascha Husa. Computationally efficient models for the dominant and sub-dominant harmonic modes of precessing binary black holes. *Physical Review D*, 103(10):104056, 2021.

[13] Maximilian Dax, Stephen R. Green, Jonathan Gair, Michael Pürrer, Jonas Wildberger, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Neural importance sampling for rapid and reliable gravitational-wave inference. *Physical Review Letters*, 130(17):171403, 2023.

[14] Sophie Hourihane, Patrick Meyers, Aaron Johnson, Katerina Chatziioannou, and Michele Vallisneri. Accurate characterization of the stochastic gravitational-wave background with pulsar timing arrays by likelihood reweighting. *arXiv preprint*, 2022.

[15] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[16] Maximilian Dax, Stephen R. Green, Jonathan Gair, Nihar Gupte, Michael Pürrer, Vivien Raymond, Jonas Wildberger, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-time gravitational-wave inference for binary neutron stars using machine learning. *arXiv preprint arXiv:2407.09602*, 2024. Submitted on 12 Jul 2024.