Exploring the Impact of Frequency Cuts on Gravitational-Wave Parameter Estimation

Sena Kalabalık
Boğaziçi University, Turkey

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

Interim Report II
LIGO Caltech SURF Program 2024

Gravitational-wave science has empowered a new era in astrophysical exploration, with the characterization of gravitational-wave signals holding crucial importance. However, accurately characterizing these signals poses challenges, especially in the presence of short-duration noise transients, known as "glitches." A specific strategy employed to address this issue and mitigate the effects of glitches involves cutting a particular frequency range off the signal. In this study, we aim to delve into the critical role that frequency cuts play in gravitational-wave analysis. Through systematic analysis and simulation-based experiments, incorporating data injections and utilizing a neural posterior estimator, we will investigate the effects of different frequency cut configurations on our interpretations of parameter posterior distributions across a diverse array of characteristic waves.

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. Traveling at the speed of light, these cosmic ripples convey valuable information about their source's characteristics and provide insights into the fundamental nature of gravity. As gravitational waves travel through space, they induce tiny expansions and contractions in the spatial dimensions they traverse. The measurements of this strain are made possible by the Laser Interferometer Gravitational-Wave Observatory(LIGO), Virgo and KAGRA Collaboration (LVK). The LVK currently operates three detectors: two LIGO detectors located in the United States and one Virgo detector in Italy. The detectors are specialized versions of a Michelson interferometer, which simply consists of two equally long, perpendicular arms 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 scientists to detect 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 especially crucial since they carry promising information about the nature of compact objects. When a signal is detected, the source characterization is made possible by employing Bayesian inference, which relies on having established models for both the signals and the detector noise. In the context of gravitational waves, these signal models are represented by waveform predictions $\mathrm{h}(\theta)$, which are contingent upon various source parameters θ, such as masses and locations of the objects. Meanwhile, the detector noise is typically assumed to be stationary and Gaussian, characterized by a certain spectrum that can be empirically estimated. Collectively, these models yield the likelihood $p(d|\theta)$ for the observed strain data d, presumed to comprise both signal and noise components. By selecting a prior $p(\theta)$ over the parameters, the posterior distribution is determined through Bayes' theorem as:

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

where $p(d)$ acts as a normalizing factor termed the evidence. This posterior distribution encapsulates our beliefs regarding the source parameters, given the observed data.[1]

In order to accomplish the inference, we have chosen to utilize the DINGO method [1] over other conventional methods commonly used in LVC, such as LALInference [2] and Bilby [3]. The primary reason for this choice is the imperative need for speed in conducting numerous trials of parameter estimations. Traditional methods such as LALInference and Bilby employ stochastic algorithms like Markov chain Monte Carlo (MCMC) and nested sampling to characterize the posterior distribution by drawing samples from it. However, these algorithms are computationally intensive, demanding numerous likelihood evaluations for each independent posterior sample. Each likelihood evaluation necessitates a waveform simulation, making the entire inference process time-consuming. However, DINGO is a neural posterior estimation (NPE) model, and has shown to achieve both significantly reduced analysis time and high accuracy. As a likelihood-free and simulation-based inference model, the simple principle of DINGO is to generate numerous simulated datasets, each with its corresponding parameters, and utilize these datasets to train a specific type

of neural network called a normalizing flow, which is employed to approximate the posterior distribution. Once the network is trained, it can rapidly produce new posterior samples following a detection.[1]

Accurate source characterization depends on the specific assumptions about the behavior of the detector noise, which are violated when there are short-duration noise transients, called "glitches" present in the signal [4]. 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. Throughout their third observing run (O3) [5, 6], the median rate of glitches in the LIGO and Virgo detectors has been reported to have surpassed 1 per minute for the majority of the duration, which indicates that the coincidence of the glitches with gravitational-wave signals are expected quite commonly [7]. 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 [8], which models the glitch and subtracts it by only using the strain data. The algorithm assumes that the strain data consists of Gaussian noise, a gravitational-wave signal, and a glitch. It models the glitch as a sum of sine-Gaussian wavelets, and these wavelets are marginalized over the parameters using a trans-dimensional Markov chain Monte Carlo (MCMC). 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 [7]. Despite the common use of the BayesWave algorithm, various studies, including [9] and [5], have demonstrated that the method may potentially leave residual artifacts within the signal, which is not unexpected since 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 [10], which uses information from auxiliary channels. The algorithm assumes that the measured strain is a linear combination of time series from different sources, where one of these sources can be modeled as the convolution of a witness time series and an unknown transfer function. In this approach, the transfer function between the auxiliary sensor and the strain data channel is determined and used to estimate the contribution of the noise source to the strain data. However, the accuracy of this subtraction method depends on the accuracy of the auxiliary sensor and the transfer function estimate [7, 9]. The mentioned systematic and statistical uncertainties of these two methods poses the risk that even after BayesWave or subtract algorithm is used, the data can still be undersubstracted, as shown in [5, 9].

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. In such instances, addressing data quality issues might involve employing a narrower frequency range during the analysis [11]. This strategy has been adopted in the cases of [5, 9]. In [5], it has been reported that after the application of these methods, the identified glitch is considered unmitigated if the data surrounding the event are inconsistent with Gaussian noise. In such cases, they evaluated the SNR lost by restricting the frequency range of data considered in parameter inference to fully remove the glitch. If the SNR loss is below 10%, they used the reduced frequency range in the analyses. Otherwise, they have used the nominal frequency range. Similarly in [9], to further investigate the relation between the potentially under-substractive glitch mitigation strategies and the estimated spin-precession posteriors, they have limited the frequency range above a progressively increasing lower limit, and evaluated the SNR values as well as posteriors of various parameters for each lower limit. The results showed that even though SNR loss was small, the posterior for the spin-precession parameter $\chi_{\rm p}$ became less informative when a more increased lower limit was used, possibly indicating that the glitch remnants were causing a misleading estimation of the posterior.

While previous studies, including [9] and [5], have explored the benefits of constraining the frequency range of data primarily for mitigating the remnants of glitches, our investigation suggests that frequency cuts may have overlooked implications. An example of this concern can be found in [7]. In this study, all three methods described have been employed and compared by their SNR values and posterior distributions. When the lower frequency limit was raised, the posteriors for spin parameters became less informative, and more influenced by the prior. Expectedly, the SNR loss was significant when higher limits were placed on the lower frequency. The results concluded that the glitch subtraction strategies narrowed the posterior distributions, and only caused a little change in SNR, therefore indicating their superiority over frequency cuts. However, of particular interest is the unexpected revelation that different frequency cut configurations led to distinctly different estimations for the same parameter, implying a complex scenario. The figures obtained demonstrated a significant differentiation in the posteriors for the $\chi_{\rm eff}$ parameter, with noticeable shifts across the plot, when different frequency range limits were applied. This demonstration implies a sophisticated relation between frequency cuts and source characterization for gravitational-wave signals.

II. OBJECTIVES

The primary objective of this research is to analyze the impact of different frequency cuts on parameter es-

timation results for compact binary coalescences. This involves systematically examining how variations in frequency cuts influence the posterior distributions of gravitational wave signals. Given that current frequency cut strategies for noise mitigation lack a systematic approach, this study aims to establish a more educated and reliable framework, potentially reducing incorrect estimations in PE results.

Additionally, this research seeks to explore the feasibility and effectiveness of utilizing a deep learning (DL) model, DINGO, for the PE pipeline. By comparing the results obtained from the conventional PE pipeline, Bilby, with those derived from DINGO, the study aims to evaluate DINGO's reliability and efficiency for large-scale analysis. This investigation is not only significant for our analysis, but also for the ongoing observing run, O4, where traditional methods may be limited by the large amount of computations needed. A successful integration of DL algorithms could enhance the speed and the efficiency of parameter estimation, thereby improving overall analysis capabilities.

III. CURRENT PROGRESS

A. Preliminary Work

To achieve my research objectives, I first aimed to comprehend the DINGO pipeline to prepare for utilizing it in subsequent steps of my research. I analyzed the event GW150914 by following an official tutorial from the DINGO documentations [12]. This analysis involved four steps: generating a waveform dataset, generating an ASD dataset, training the network, and performing inference.

In the first step, the configuration file included settings for the generated waveforms, such as priors, waveform generator specifications, and domain specifications. The second step required setting another configuration file to generate the ASD dataset, including attributes like sampling frequency, segment duration, and window type. These settings defined how the ASD was estimated using segments of noise data from detectors H1 and L1 during the O1 observing run. For computational efficiency in this simple model, only one ASD was generated and used for noise realization.

For network training, it was important to specify the features of the neural network architecture, which consist of two components: a normalizing flow (neural spline flow) and an embedding network. These settings included the number of flow transforms, hidden dimensions, epochs and optimizer choices. In the final step, inference was performed using the `dingo_pipe` command, allowing for the analysis and manipulation of the results. The resulting posterior distributions for the 11 parameters of GW150914 are shown in Figure 1.

Another milestone in preparing for my analysis involved becoming proficient with Bilby and mastering essential tools for data manipulation. I utilized the `prior` class to construct, evaluate, and sample from priors, while also creating insightful plots. Additionally, working with the `WaveformGenerator` and `Interferometer` classes enabled me to bridge theoretical waveforms with realistic detector observations, highlighting key detector features affecting waveform projections and noise realizations.

One of the pivotal concepts I explored was noise realization. This included loading the actual PSD of the event GW191109 into the detector simulation, using this ASD to accurately set the strain data and simulate realistic noise conditions. As part of these processes, I worked in both time domain (TD) and frequency domain (FD), performing effective Fourier and Inverse Fourier transforms using Bilby. I also employed strategies such as whitening to enhance plot readability and interpretability. Finally, I gained proficiency in injecting signals into detector simulations and analyzing residuals by subtracting the injected waveforms.

Weekly seminars also provided valuable insights into CBC parameter estimation methodologies, modeling and characterization of the detectors, solving equations numerically with Python, and testing general relativity (GR). Additionally, the visit to the Hanford LIGO site enhanced my understanding of detector operations and the complexities of gravitational wave detection.

B. Methods

1. *Generating Injections*

A crucial step in our analysis involves comparing DINGO and Bilby. To achieve this, we aimed to generate identical injections for inference with both pipelines. We used the same set of parameters, ASD, domain settings, and the seed to ensure that the randomness would be consistent and reproducible for both sets of injections. This included utilizing the ASD from the GW191109 event and aligning the priors accordingly.

Following this plan, I initially used DINGO's tools and created 50 injections per detector. However, we discovered that the noise realization processes for DINGO and Bilby are not identical, leading to differences in the resulting injections, as illustrated in 2.

To address this discrepancy, we decided to use injections from a single framework for both pipelines, specifically those generated by Bilby. This ensures consistency and comparability in our analysis.

2. *Preparing the Frequency Cuts Network*

To apply frequency masking for both training and inference, we utilized parts of the code from [13] and integrated them into our network. This integration of frequency masking is implemented mainly as a transform applied to the input before it passes to the SVD layer,

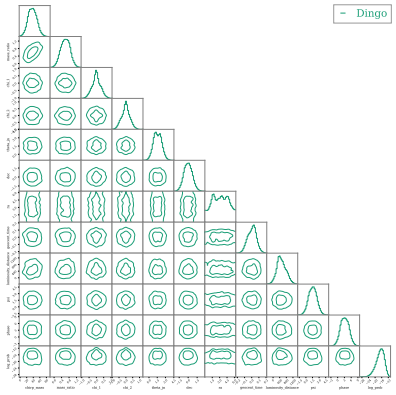


FIG. 1: This figure illustrates the posterior distribution results for 11 parameters of GW150914. The less promising results were anticipated due to the simplified workload for this analysis.

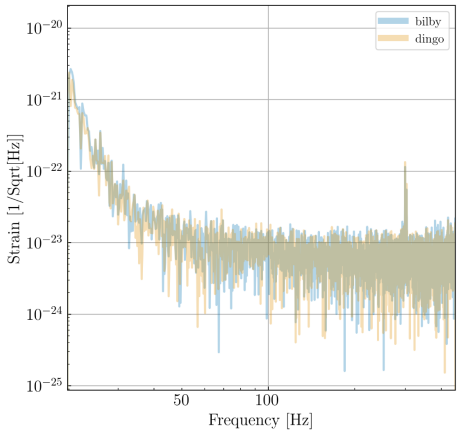


FIG. 2: The figure shows the comparison of one selected injection between DINGO and Bilby. Although they use the same PSD, the signals do not appear identical due to differences in the noise realization process.

which is the linear and first stage of the embedding network in the DINGO architecture.

The application is straightforward: when a minimum frequency upper limit is used during training, the network applies masking between the actual minimum frequency of the signals (20 Hz) and the upper limit. This choice is random, allowing the model to be trained on the dataset with various frequency masking variations between these two values. For inference, the masking, with the specified upper limit, is directly applied to the input as a waveform transform.

We then trained a small network to test if frequency masking was being applied as expected. Our networks, similarly to the injections, align with the GW191109 event in terms of the chosen ASD, priors, and domain settings. Although the network ran without errors in its final form, it gave insufficient results, possibly indicating that the network did not learn to interpret the frequency masking. This is expected to be due to the network being small with few layers, trained for a few epochs, and the dataset being small (5000 samples).

3. *Improving the Network*

To enhance the network's capability, we expanded the training dataset to 100000 samples. We also introduced two stages of training, increased the hidden dimensions for both the embedding network and the normalizing flow, extended the number of epochs and made adjustments to the learning rates accordingly. Given the larger dataset, compression was applied to manage the data more efficiently.

After the training, we selected an injection and inferred it with different frequency cutting limits to observe how the application works. The selected injection and associated ASD, which together comprise our input to the network, are plotted in 3. For the inference of this network, we increased the number of samples and enabled impor-

tance sampling to obtain more sophisticated results.

The posterior distributions of the injection, as obtained from the network, demonstrated that it was able to differentiate between the various frequency limits applied and also able to provide more detailed insights than the toy model, as shown in 5. However, these posteriors did not exhibit a high level of clarity and were less detailed than anticipated.

The observed poor performance in the posterior plots is primarily due to the sampling process inadequately capturing the characteristics of the true distribution, leading to undersampling. This issue arises because the network is not effectively modeling the true underlying distribution of the parameters.

In DINGO's importance sampling strategy, the approach of reweighting allows samples drawn from one posterior distribution to be used for estimating a different posterior distribution [14, 15]. 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:

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

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}.

Low efficiency is often caused by the overlap between regions of low posterior for the approximate model and regions of high posterior for the target model. This mismatch results in disproportionately high weights for certain samples [15]. 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. Figure 4 illustrates this issue in our network. The observed low sample efficiency suggests that the network is not performing as expected, and we anticipate this value should be significantly higher for more accurate results.

Our current focus is on diagnosing the causes of the network's poor performance and developing solutions to address this issue.

IV. NEXT STEPS

The next steps we have outlined for the remainder of the summer are as follows:

- Conduct inference with both DINGO and Bilby for each frequency cut range, and compare the results to assess DINGO's reliability in large-scale analysis.
- Train a larger DINGO network using multiple ASDs for more robust results.
- Continue inference with the enhanced DINGO network to derive posterior distributions for the injections at each frequency cut.
- Analyze the outcomes to develop systematic conclusions.
- Aim to utilize DINGO for analyzing O4 data.

V. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation Research Experience for Undergraduates (NSF REU) program, the LIGO Laboratory Summer Undergraduate Research Fellowship program (NSF LIGO), and the California Institute of Technology Student-Faculty Programs.

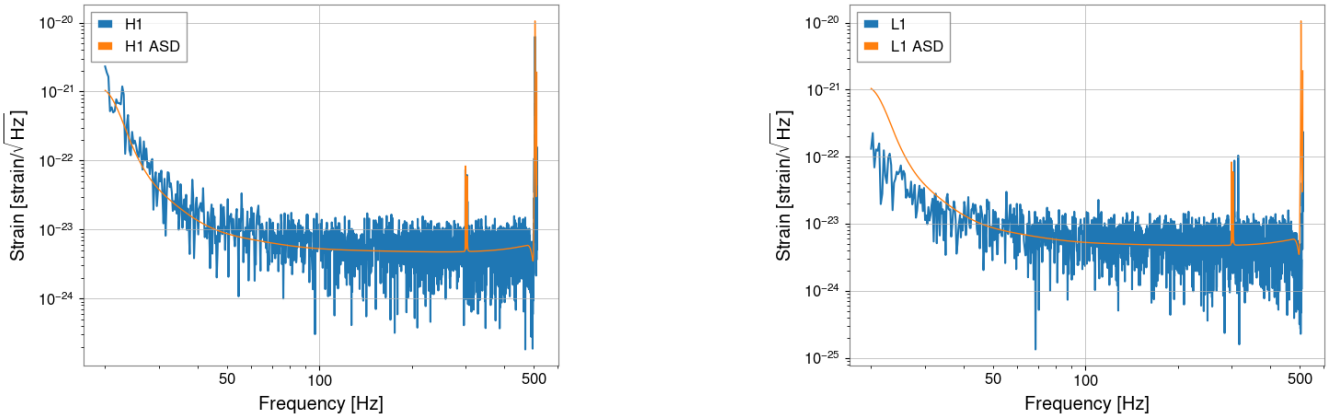
[1] Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-time gravitational wave science with neural posterior estimation. *Phys. Rev. Lett.*, 127:241103, 2021.

[2] J. Veitch et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library. *Phys. Rev. D*, 91:042003, 2015. Published 6 February 2015.

[3] 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.

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

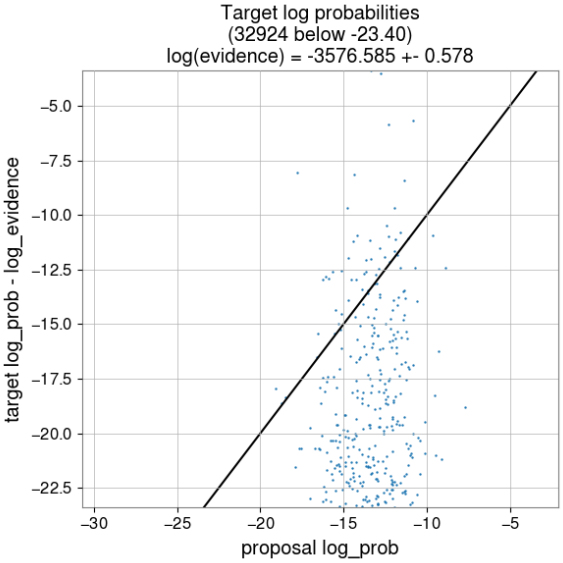
[5] R. Abbott et al. GWTC-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run. *Phys. Rev. X*, 11:021053, 2021.



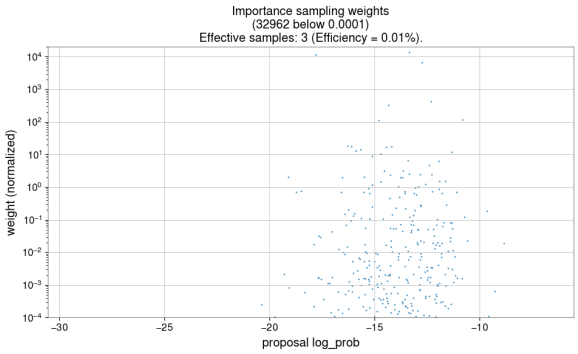
(a) H1

(b) L1

FIG. 3: The plot illustrates the injection per detector alongside the corresponding ASDs, representing the combined input data for the network.



(a) 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.



(b) 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.

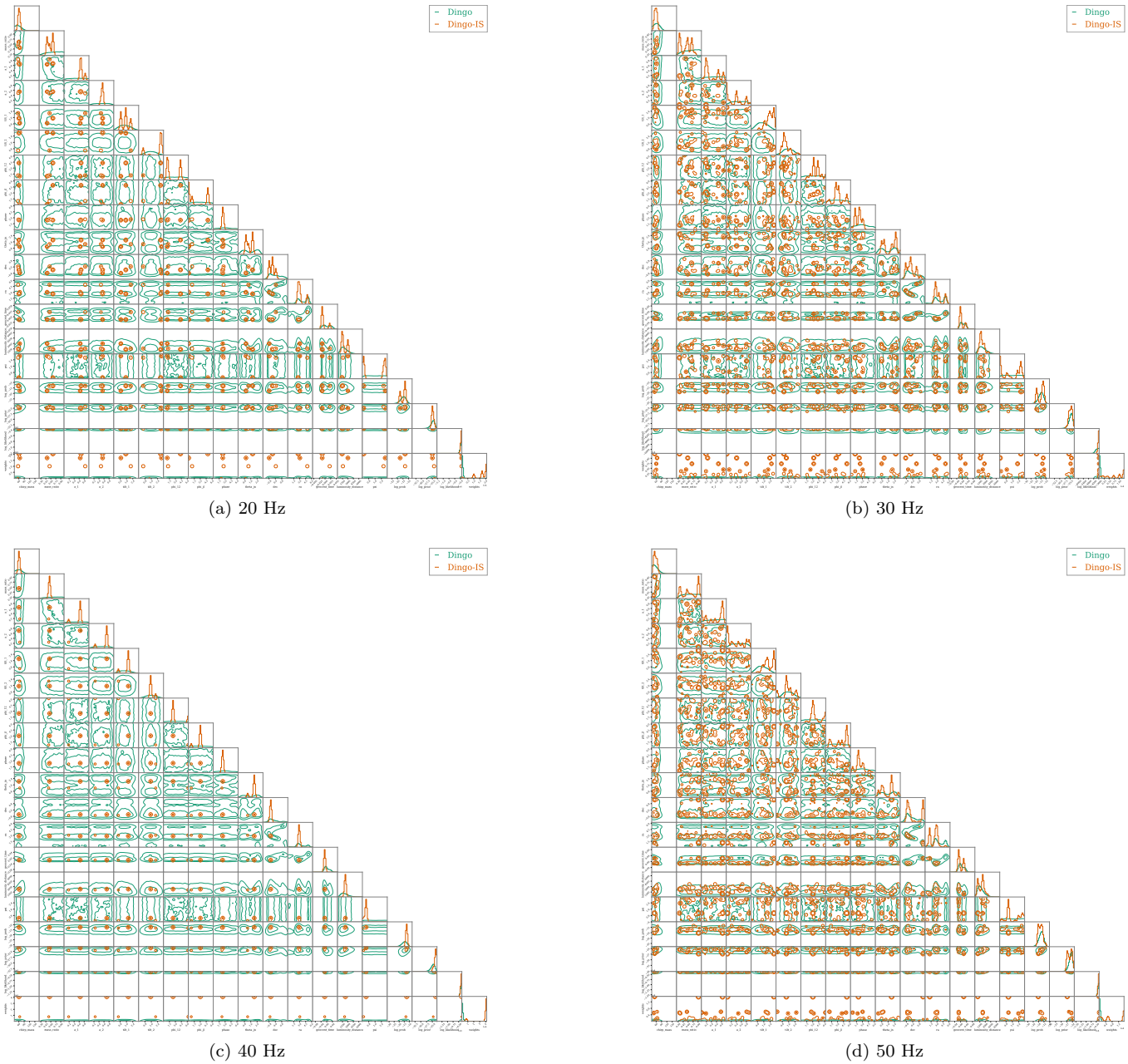
FIG. 4: The figure corresponds to the minimum frequency limit of 20 Hz; however, similar features are observed across all frequency limits.

[6] R. Abbott, Virgo Collaboration others (LIGO Scientific Collaboration, and KAGRA Collaboration). 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.

[7] 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.

[8] 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.

[9] Ethan Payne, Sophie Hourihane, Jacob Golomb, Rhiannon Udall, Derek Davis, and Katerina Chatziioannou. Curious case of GW200129: Interplay between spin-



(a) 20 Hz

(b) 30 Hz

(c) 40 Hz

(d) 50 Hz

FIG. 5: The plots display the posterior distributions resulting from applying different frequency cuts to a single injection, with importance sampling used. These distributions demonstrate that the frequency masking is being applied as expected. However, the posteriors lack the level of clarity and detail that was anticipated.

precession inference and data-quality issues. *Phys. Rev. D*, 106:104017, 2022.

[10] 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.

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

[12] DINGO Collaboration. DINGO: Example toy npe model. https://dingo-gw.readthedocs.io/en/latest/example_toy_npe_model.html, Accessed: 2024-07-15. Accessed on: 2024-07-15.

[13] 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.

[14] 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.

[15] 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.