Robust inference of gravitational wave source parameters in the presence of noise transients using normalizing flows

Chun-Yu Xiong,[1] Tian-Yang Sun,[1] Jing-Fei Zhang,[1] and Xin Zhang[1, 2, 3, *]

[1]*Liaoning Key Laboratory of Cosmology and Astrophysics,*
College of Sciences, Northeastern University, Shenyang 110819, China
[2]*MOE Key Laboratory of Data Analytics and Optimization for Smart Industry,*
Northeastern University, Shenyang 110819, China
[3]*National Frontiers Science Center for Industrial Intelligence and Systems Optimization,*
Northeastern University, Shenyang 110819, China

Gravitational wave (GW) detection is of paramount importance in fundamental physics and GW astronomy, yet it presents formidable challenges. One significant challenge is the removal of noise transient artifacts known as "glitches", which greatly impact the search and identification of GWs. Recent research has achieved remarkable results in data denoising, often using effective modeling methods to remove glitches. However, for glitches from uncertain or unknown sources, current methods cannot completely eliminate them from the GW signal. In this work, we leverage the inherent robustness of machine learning to obtain reliable posterior parameter distributions directly from GW data contaminated by glitches. Our network model provides reasonable and rapid parameter inference even in the presence of glitches, without needing to remove them. We also investigate various factors affecting the rationality of parameter inference in our normalizing flow network, including glitch and GW parameters. The results demonstrate that the normalizing flow can reasonably infer the source parameters of GWs even with unknown contamination. We find that the nature of the glitch itself is the only factor that can affect the rationality of the inferred results. With improvements to our model, we anticipate accelerating the localization of electromagnetic counterparts and providing priors for more accurate deglitching, thereby speeding up subsequent data processing procedures.

I. INTRODUCTION

Approximately a hundred years ago, Einstein predicted the existence of gravitational waves (GWs) within the framework of general relativity. It wasn't until 2015 that the Advanced LIGO detectors achieved the first direct detection of a GW signal from the merger of a binary black hole (BBH) system, known as GW150914 [1]. This discovery marked a monumental advance in the field.

GW research is pivotal for testing general relativity [2–9], exploring astrophysical and cosmological phenomena, and accurately measuring cosmological parameters [10–28]. These activities significantly enhance our understanding of fundamental physics and the evolution of the universe.

However, the data received by detectors is highly susceptible to interference from noise transient, resulting in deviation in the analysis process [29–32]. The LIGO-Virgo-KAGRA collaboration has revealed that non-Gaussian features in the data, such as noise transient, or called "glitch" [33–35], can significantly impact the detector's ability to search for GW signals [33].

Current observations have characterized dozens of different glitches by Gravity Spy [36, 37]. These glitches can exhibit variations in terms of their frequency, amplitude, and duration, making their identification and mitigation complex tasks. They are categorized according to time, frequency, signal-to-noise ratio (SNR), and duration [38, 39], and there are many kinds of glitches, among which the common glitches being blip glitches, scattered-light glitches and thunderstorm glitches [40–42]. Indeed, glitches have significant impacts on GW data and subsequent data analysis [43–45]. These non-Gaussian non-stationary features can mimic or mask the signature of GWs, making it more difficult to distinguish real signals from noise [46, 47]. For example, they can affect the search phase, causing detectors to generate false alarms [29]. The presence of glitches causes incorrect sky positioning in GW data analysis, affects the tracking of electromagnetic (EM) counterparts [48], and introduces biases in the inference of cosmological parameters [49, 50].

The interference caused by glitches in GW data cannot be ignored, and many efforts are being made to mitigate their disturbance [41, 51–57]. The gating approach uses a windowing function to directly smooth the portion affected by the glitch for events whose time-frequency can be separated [58]. Gwsubtract [32, 59, 60] relies on the auxiliary channel to remove glitches, but it has a limited impact on removing glitches that do not appear in auxiliary channels. BayesWave reconstructs the Gaussian noise, glitches, and GW signals using wavelet transforms and inter-correlations between detectors to remove glitches [61–63]. However, this high-dimensional Bayesian analysis is extremely computationally intensive, and individual events cannot be analyzed in a timely and efficient manner at expected future event rates. Therefore, the development of an effective framework for rapidly mitigating the impacts of glitches and inferring

GW parameters has become an urgent need.

The development of Artificial Intelligence offers an excellent way to facilitate GW research by overcoming the problems of high computational costs and handling huge amounts of data [64]. Currently, most state-of-the-art machine learning algorithms remove non-Gaussian noise components [65–71] by fitting their parameters. Moreover, machine learning techniques [72–85] utilize the normalizing flow [86–90] for the posterior distribution of source parameters [91–94]. While these methods can effectively enhance data quality [95] in certain aspects, they cannot ensure that the noise remaining after deglitching consists solely of Gaussian noise. Moreover, other approaches have not involved the removal or subtraction of non-Gaussian noise [89, 96].

In our previous work, we employed the normalizing flow to perform parameter inference on GW signals contaminated by several common glitches [97]. However, glitches come in different types with varying frequencies of occurrence, and not all types of glitches can be well modeled or have sufficient numbers for training. We are seeking a method that does not rely on modeling to remove glitches and enable rapid parameter inference. Recent studies have demonstrated the robustness of machine learning [98–102]. The robustness of machine learning methods can aid in performing parameter inference reasonably, even without explicitly modeling and removing glitches. It enables the effective mitigation of non-Gaussian noise interference while inferring GW parameters.

In this work, our primary aim is to investigate the impact of noise transients on the parameter inference obtained from the normalizing flow network. We use Gaussian noise for training to infer signal parameters contaminated by non-Gaussian noise. Specifically, we utilize simulated sine-Gaussian glitches, encompassing a wide range of morphologies found in advanced detector noise transients [103, 104]. We discuss the effect of the number of detectors on robust inference in the normalizing flow network. We find that our model provides a relatively better posterior distribution compared to the Bayesian inference method when dealing with glitch-contaminated GW data. In this work, we utilize the `bilby dynesty` [105, 106] to implement Bayesian inference. We have also explored the impact of the GW signal intensity on the robustness of network inference. The results indicate that the SNR of the GW signal has varying degrees of impact on the network's ability to perform source parameter inference. Furthermore, we consider the influence of glitches' intensity on the robustness of source parameter inference when utilizing the normalizing flow network. Meanwhile, we conduct a detailed study on the impact of three key features of non-stationary noise [107], on the robustness inference using the normalizing flow. We find that it is possible to determine the feasibility of obtaining reasonable results by assessing the robustness of the inference considering the characteristics of non-Gaussian noise.

The results demonstrate that the normalizing flow network can provide reliable posterior parameter inference for GWs contaminated by unknown glitches without undergoing deglitching. When GW data are contaminated by unknown noise, our model is capable of delivering swift and reasonable parameter inference, albeit with a difference in magnitude. This provides a reasonable prior for subsequent more accurate processing. When extended to binary neutron star (BNS) mergers, this method is also expected to enable the rapid localization of EM counterparts.

This paper is structured as follows. In Section II, the methodology used in the work is described in detail. Some of the work's conclusions and discussions are presented in Section III. Finally, we present the main conclusions of our work in Section IV.

II. METHODOLOGY

A. Normalizing flow network architecture

The normalizing flow is a deep learning-based generative model for learning complex probability distributions [108]. Its core idea is to transform a simple probability distribution into a complex probability distribution through a series of reversible transformations. Neural posterior estimation [109] involves training a neural network to approximate the transformation process of this complex posterior distribution, thereby achieving an approximate estimation of the posterior distribution. In the inference process, the normalizing flow generates samples by applying a series of reversible transformations to the base distribution. The normalizing flow network learns the parameters of these transformations to map the base distribution to a posterior distribution corresponding to the real data distribution.

The normalizing flow network consists of a series of coupling transformations, each of which is transformed according to the input data and the output of the previous layer of the neural network, allowing the predicted distribution to gradually approximate the posterior distribution. We also employed neural spline flow [110], which utilizes monotonic rational quadratic spline coupling transformations. Between coupled transformations, sample components were inverted to ensure that the transformation sequence specifically and completely transformed all components.

The normalizing flow [108, 111–113] can be expressed by the following formula

$$q(\theta|s) = \pi\left(f_s^{-1}(\theta)\right)\left|\det J_{f_s}^{-1}\right|, \qquad (1)$$

where f_s represents the invertible transformation of the normalizing flow [114], and s represents the input data strain to the network. We select the base distribution $\pi(u)$ in a way that facilitates easy sampling and evaluation of its density [77]. In our case, we consistently choose

TABLE I. The key hyperparameters of the normalizing flow.

Hyperparameter	Value
Flow Steps	9
Hidden Layer Size	4096
Transform Blocks	7
Activation	ReLU
Batch Norm	False
Bins	32

the standard multivariate normal distribution with the dimension D matching that of the sample space. We approximate the posterior distribution of GW $p(\theta|s)$, by utilizing the normalizing flow $q(\theta|s)$, aiming to obtain the posterior distribution from the parameter θ.

Our architecture for the normalizing flow is quite similar to the one described in Ref. [97], except for data fusion. Our network architecture combines a one-dimensional ResNet-50 network [115] for feature extraction with the normalizing flow network for estimating the parameters of the posterior distribution. The overall structure of the working principle is illustrated in Fig. 1. The data strain input to the network first undergoes the ResNet neural network, then the feature vectors extracted from the ResNet model. Finally, the feature vectors are processed by the normalizing flow to obtain the GW waveform as well as the parameter inference.

Unlike the parameters used in Ref. [97], the network parameters in this work are shown in Table I. For networks corresponding to different numbers of detectors, their network parameters remain the same, except that the number of channels in the network corresponds to the number of detectors. We also experimented with model complexity reduction and regularization techniques to enhance the robustness of the network but found that they had minimal impact on the results.

During the training process, we utilize the AdamW optimizer [116] with a learning rate set to 0.0001, a batch size of 200, and a learning rate decay factor of 0.99. To ensure the validity of the normalizing flow network predictions and prevent overfitting, we regenerate 10,000 new data points at the beginning of each epoch and employ 50 cycles of early stopping. The training of this network was conducted on an NVIDIA RTX A6000 GPU with 48 GB of memory.

B. Data set

In terms of dataset preparation, the data used during the training phase consists of simulated Gaussian noise and simulated GW signals. The signals used are from the GW signals simulated by the SEOBNRv4_opt model [117], and they are selected in the range of SNR between

TABLE II. Priors of simulated GW waveform parameters. Note that other parameters not mentioned are set to zero for simplicity.

Parameter	Uniform distribution
Chirp mass	$\mathcal{M}_c \in [25.0, 62.5]\ M_\odot$
Mass ratio	$q \in [0.5, 1]$
Right ascension	$\alpha \in [0, 2\pi]$ rad
Declination	$\delta \in [-\pi/2, \pi/2]$ rad
Polarization angle	$\psi \in [0, 2\pi]$ rad
Luminosity distance	$d_L \in [500, 5000]$ Mpc
Signal to noise ratio	SNR $\in [8, 32]$

8 and 32. The noise used is a simulated Gaussian noise generated using `Pycbc` [118, 119] with a frequency range of $20 - 4096$ Hz, which represents the high-pass frequency and sampling frequency respectively. This also implies a Nyquist frequency of 2048 Hz. To obtain the noise sensitivity curves $S_{n,k}(f)$ for different detectors, we utilize the aLIGOZeroDetHighPower, AdvVirgo, and KAGRA methods from the pycbc.psd library [120, 121]. After generating Gaussian noise, we inject a signal between 0.7 s and 0.9 s. The specific parameters distribution of the signal are shown in Table II, and the simulated GW observation data are whitened. Finally, a 2-second dataset is generated.

Here, for a signal h, the SNR is defined as

$$\rho = (\tilde{\boldsymbol{h}} \mid \tilde{\boldsymbol{h}})^{1/2}, \qquad (2)$$

where $\tilde{\boldsymbol{h}}$ is the frequency domain GW waveform considering the detector network including N independent detectors and can be written as $\tilde{\boldsymbol{h}} = \left[\tilde{h}_1, \tilde{h}_2, \cdots, \tilde{h}_k, \cdots, \tilde{h}_N\right]$. Here $\tilde{h}_k$ is the frequency domain GW waveform of the k-th detector. The inner product is defined as [26]

$$(\tilde{\boldsymbol{h}} \mid \tilde{\boldsymbol{h}}) = \sum_{k=1}^{N} (\tilde{h}_k \mid \tilde{h}_k) = \sum_{k=1}^{N} 4 \int_{f_{\min}}^{f_{\max}} \frac{\tilde{h}_k(f)\tilde{h}_k^*(f)}{S_{n,k}(f)} \mathrm{d}f, \qquad (3)$$

and $S_{n,k}(f)$ is the sensitivity curve function of the k-th detector [122]. In the above expression, $f_{\min}$ is the high-pass frequency, and $f_{\max}$ is the Nyquist frequency. Also, we calculate the glitch SNR in the same way.

In the test dataset, we utilize simulated Gaussian noise and simulated GW signals, injecting simulated glitches into them, as illustrated in Fig. 2. The glitches we use are modeled using sine-Gaussian functions [123], which cover most of the types of glitches currently observed by the LIGO, Virgo, and KAGRA networks. Note that when injecting glitches into detectors, we selected the LIGO Livingston detector due to its higher sensitivity and higher occurrence frequency of glitches, as random glitches typically occur in only one detector. We examine
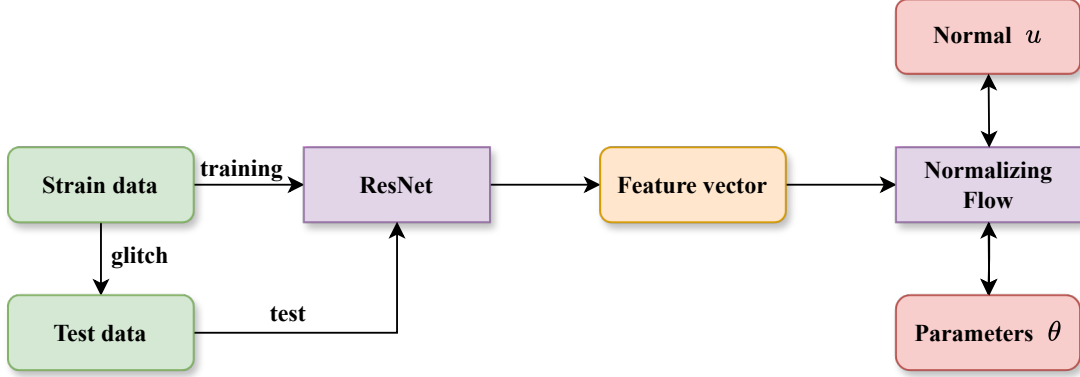
FIG. 1. The workflow of the normalizing flow network. The input strain data contain simulated GW signals and Gaussian noise. The strain data is passed through an identical residual net (ResNet) model, and then the feature vectors are extracted from the ResNet model. Then it proceeds through the normalizing flow to obtain the parameter posterior distributions. The above steps are repeated by injecting glitches into the strain data as a test set.
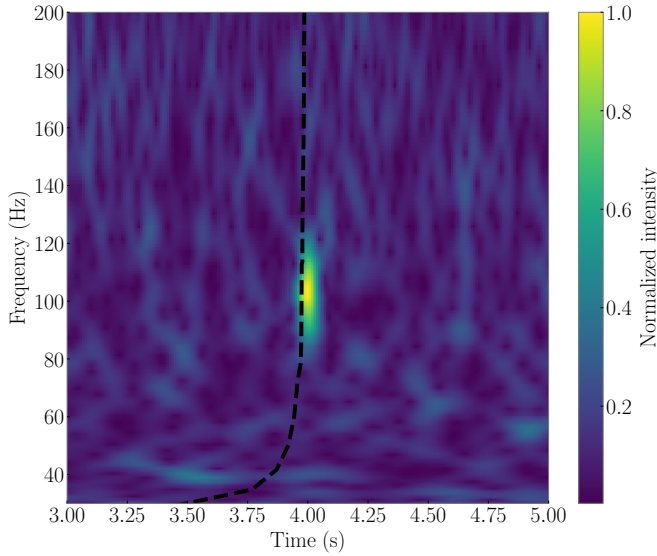

FIG. 2. Time-frequency spectrogram for a simulated GW signal coincident with a glitch. The black dashed track shows the inferred time-frequency evolution of the GW signal. The light spot is the approximate location of the simulated glitch.

a sine-Gaussian glitch [103], which is defined as

$$h(t) = h_0 \sin(2\pi f_0(t - t_0)) \exp(-\frac{(t - t_0)^2}{2\tau^2}), \quad (4)$$

where $\tau = Q/\sqrt{2}\pi f_0$ is the time duration. Here $h_0 = $ hrss/$\sqrt{\tau}$, where hrss is the root sum squared amplitude of the glitch, Q is the dimensionless quality factor, t_0 and f_0 are the location of the sine-Gaussian in time and frequency respectively. We set $t_0 = 0.8$ s to simulate the scenario where the signal and glitch overlap (the under-

TABLE III. Parameters used to prepare the simulated glitches set.

Parameter	Value
f_0	$50 - 300$ Hz
Q	$2 - 20$
h_0	$4 \times 10^{-22} - 1 \times 10^{-21}$

lying rationale is provided in Appendix A). The selected parameters range of glitches are shown in Table III.

C. Analytical approach

The Kolmogorov-Smirnov (KS) test [124] is a statistical method that can be used to test the goodness of fit between a sample distribution and a reference distribution. In the context of GW parameter estimation, the KS test can be employed to assess the agreement between the estimated parameter distribution and a theoretical or expected distribution [118, 125–127]. The KS test first calculates the cumulative distribution functions (CDFs) of the two distributions. Then it evaluates the distance D_n between the two distributions based on the CDFs, which can be written as

$$D_n = \max | F_n(x) - F(x) |, \quad (5)$$

where $F_n(x)$ is a specified continuous distribution function from a random sample that a set of random variables X_i, $i = 1, ..., n$. $F_n(x)$ and $F(x)$ respectively represent the CDFs of the two distributions under validation. The resulting D_n is also known as the KS statistic. The p-value, calculated based on the KS statistic, is commonly

employed as a straightforward measure to assess the similarity or dissimilarity between two distributions. The p-value is commonly compared to the significance level. If the p-value is less than the significance level, it indicates that the two distributions are considered statistically different. Here, the significance level has taken a value of 0.05. If the p-value obtained from the KS test is less than 0.05, it indicates that the two distributions are statistically different, and the hypothesis of their similarity can be rejected.

In this work, we compare two distributions: the ideal distribution and the one-dimensional distribution generated by the normalizing flow network. Since the percentile should be uniformly distributed between 0 and 1, the ideal distribution is uniform. The other distribution consists of the probabilities corresponding to the confidence intervals in which the true value of each sample lies among the simulated 200 samples. Our initial hypothesis assumes that the two distributions are the same distribution, which means the posterior inference of the parameters given by the normalizing flow network is reasonable. Then the p-value is compared with the significance level of 0.05; if the p-value is less than 0.05, we consider the original hypothesis invalid.

The Jensen–Shannon divergence (JSD) is used to compare the difference between two probability distributions and measure the similarity of two distributions [128, 129]. It is achieved by taking a weighted average of the probability density functions of the two distributions, given by

$$\text{JSD}(p \parallel q) = \frac{1}{2}\text{KLD}(p \parallel \frac{p+q}{2}) + \frac{1}{2}\text{KLD}(q \parallel \frac{p+q}{2}), \tag{6}$$

where Kullback-Leibler divergence (KLD) [130, 131] is defined as

$$\text{KLD}(p \parallel q) = \int_{-\infty}^{+\infty} p_{(x)} \ln \frac{p_{(x)}}{q_{(x)}} \mathrm{d}x. \tag{7}$$

The KLD can also be used to measure the difference between two probability distributions, but the KLD is asymmetric, so calculating the KLD between two distributions p and q may yield different results than calculating the KLD between q and p. To overcome the asymmetry of KLD, JSD is more symmetrical and smooth. JSD values range from 0 to 1, making explaining and comparing differences between different distributions easier. The JSD value ranges from 0 to 1, where a JSD value of 0 indicates identical distributions and a JSD value of 1 represents the maximum difference between the distributions [132]. This work uses the JSD criterion to measure the difference between the parameter distribution inferred by the normalizing flow network after being contaminated by glitches and the ideal parameter distribution.

III. RESULTS AND DISCUSSION

Before proceeding with further analysis, we conducted tests using the normalizing flow to assess the rationality of parameter inference in the presence of glitches. We injected glitches into the simulated GW data and performed parameter inference using the normalizing flow and `bilby dynesty` respectively. The results are shown in Fig. 3. It can be seen that the parameter posterior distribution obtained by the normalizing flow is closer to the injection values of parameters than the results inferred by the `bilby dynesty`. This shows that the normalizing flow is superior to the conventional `bilby dynesty` for processing GW data contaminated by glitches, which is the basis of our subsequent work.

We simulate 200 events to validate the robustness of the normalizing flow network's inference for arbitrarily selected parameters. By injecting 200 simulated GW waveforms, along with the added glitches, we measured the frequency of the true parameter over a range of confidence levels. We also simulated 200 events without glitches. We calculate the cumulative probability of each parameter posterior's percentile at the injection values of parameters, then plot the CDF for the joint distribution of the source parameters in Fig. 4. Ideally, the injected CDF of the parameters should exhibit a diagonal pattern, as the CDF represents the cumulative sum of individual event distributions. The proximity of the curve to the diagonal indicates the accuracy of parameter inference. A curve that closely aligns with the diagonal signifies superior performance in accurately estimating the parameters. The legend displays p-values obtained from the posterior distributions for each method.

Based on our assumption that p-values less than 0.05 are deemed unreasonable for the model to support true posterior inference, the p-value provide a quantitative measure of the agreement between the normalizing flow model cumulative distribution and the expected theoretical distribution. With only Gaussian noise, both the normalizing flow and `bilby dynesty` methods perform well. But the normalizing flow employed in our study demonstrates superior performance compared to the `bilby dynesty` when it comes to accurately learning the true posterior distribution, particularly in scenarios involving glitch-contaminated GW data.

From Figs. 3 and 4, it can be observed that our normalizing flow model outperforms the `bilby dynesty` in posterior parameter inference. We investigated the reasons behind the superior performance of the neural network, which are demonstrated in Fig. 5. To generate different sets of data, we generate different Gaussian noises by changing the random seed and then injecting glitches and signals into them. We selected five different sets of Gaussian noise data, with each set having glitch SNR ranging from 0 to 19. Each point on the same line consists of 200 simulated GW data, where these events have the same GW signal and Gaussian noise, but different injected glitches. We plotted the differences between the
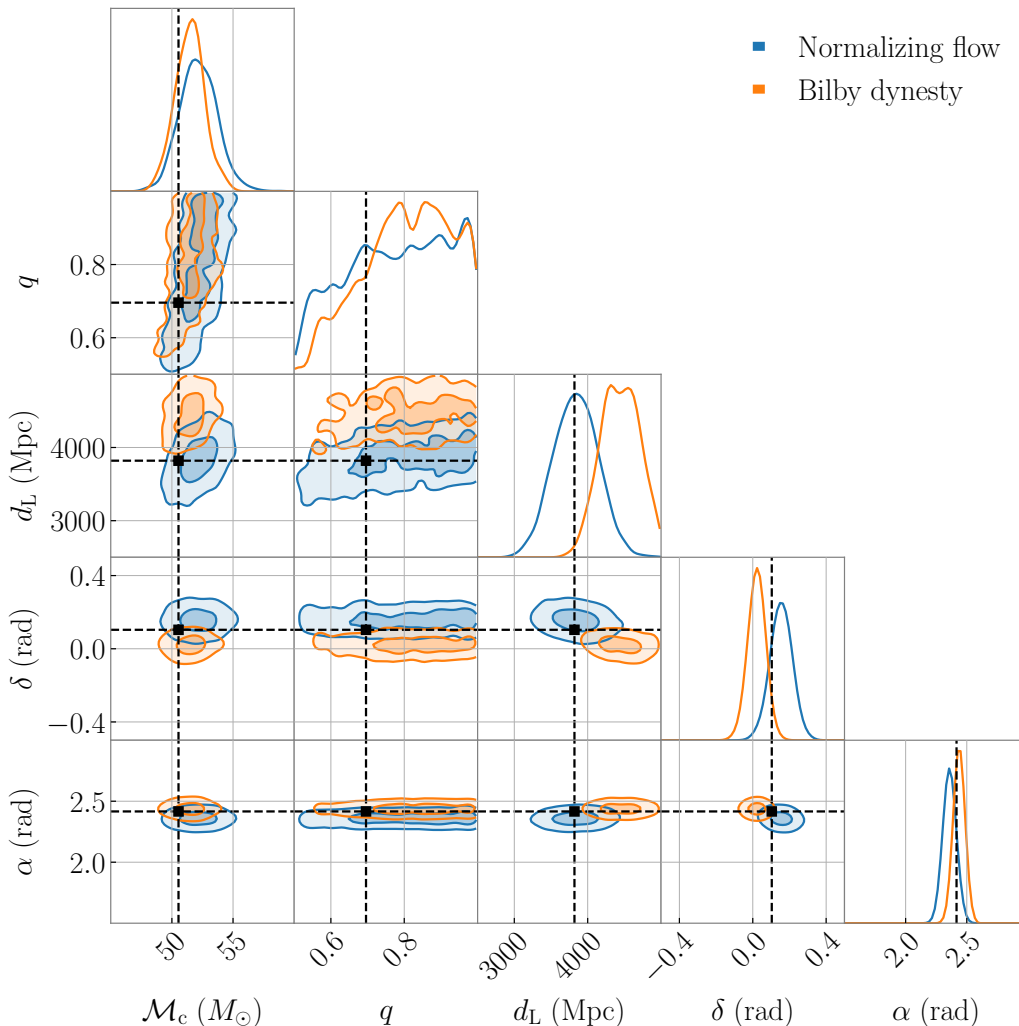
FIG. 3. Parameter posterior corner plot of injected signal in Gaussian noise with glitch SNR of 8. Here the blue and orange lines show the inference of the joint posterior distribution of the parameters by the normalizing flow and `bilby dynesty`, respectively. One-dimensional histograms of the posterior distribution of each parameter in both methods are plotted along the diagonal. The injection values of the simulated signal are indicated by the black vertical and horizontal lines.

mean and standard deviation of the fitted posterior distributions of parameters obtained by the normalizing flow network and the Bayesian inference method for different glitch SNRs. These differences were calculated by subtracting the results obtained at a glitch SNR of 0.

We selected the d_{L} parameter for demonstration purposes. As the glitch SNR increases, the mean and standard deviation obtained by the normalizing flow tend to be centered around the injected values when the glitch SNR is relatively low. However, they gradually deviate from the injected values as the glitch SNR increases. In contrast, the `bilby dynesty` continues to deviate from the injected values persistently as the glitch SNR increases. This also demonstrates the results above, specifically in Fig. 3, where the normalizing flow provides a wider range of posterior distributions as the injected val-

ues increase, while the `bilby dynesty` shows more deviation from the injected values. Indeed, this also highlights the robustness of the normalizing flow in handling glitches. The normalizing flow exhibits a certain level of stability and resilience in the presence of glitches, as evidenced by its ability to provide posterior distributions that are closer to the true values compared to the `bilby dynesty`.

A. Different numbers of detectors

Indeed, with the increase in the number of future GW detectors and improvements in their sensitivity, it is expected that the number of events contaminated by glitches will also significantly increase. This will pose
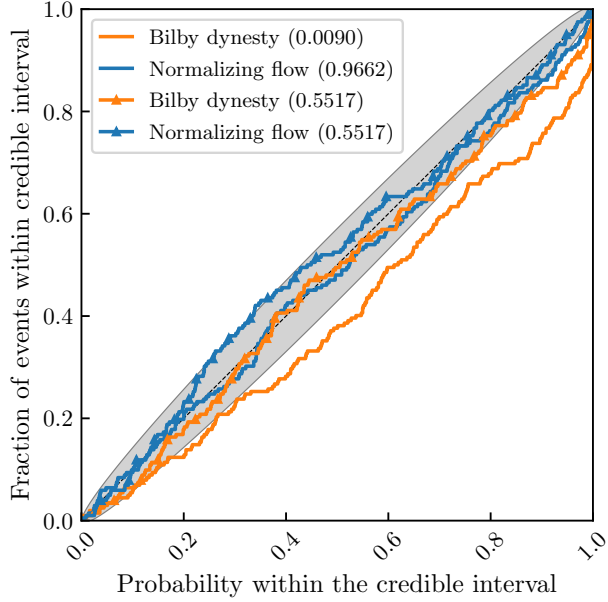
FIG. 4. Cumulative distribution of the quantile of which the true value lies for marginalized distribution. The curves were constructed using the 200 samples and the dashed black diagonal line indicates the ideal result. Shaded areas indicate 95% confidence intervals from the ideal distribution with the same number of events as the injection samples. The legend shows p-values obtained by the normalizing flow and `bilby dynesty`. The triangular line represents p-values without glitches.
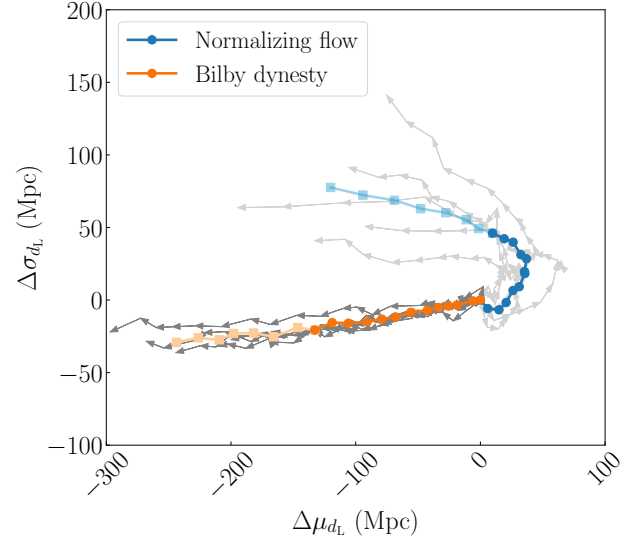

FIG. 5. Variation of the difference in the mean and standard deviation of inferred GW source parameter d_L with increasing glitch SNR. The dark and light gray lines represent the results from five independent experiments for the normalizing flow and the `bilby dynesty`, respectively. The arrows represent the direction of increasing glitch SNR. The blue and orange lines represent the average results from five independent experiments. Circles and squares indicate glitch SNR below and above 8, respectively.

even greater challenges for GW data processing. In principle, increasing the number of detectors will result in a lower proportion of contaminated data considering only one detector data contaminated by glitches. Indeed, in the case of the same SNR of the detector networks, according to Eq. (3), the SNR for each detector will decrease. Therefore, the intensity of the detected GW signals by a single detector will also decrease, resulting in a higher relative intensity of glitches compared to the signal, as compared to a network with fewer detectors. In this section, we investigate the impact of the number of detectors on our normalizing flow network capability to infer parameters.

Figure 6 illustrates p-values obtained parameter inference when glitches were injected into the GW data from different numbers of detectors, respectively. It depicts the distribution of p-values error bars for 20 sets of 200 GW data with varying detectors as the injected glitch SNR increases. These error bars are derived from the 1σ confidence intervals of the values provided by each group. The curves of the detectors exhibit an upward trend and then a downward trend. The normalizing flow can exhibit overconfident posterior approximations, leading to underestimated parameter variances [134]. However, as shown in Fig. 5, it can be observed that when
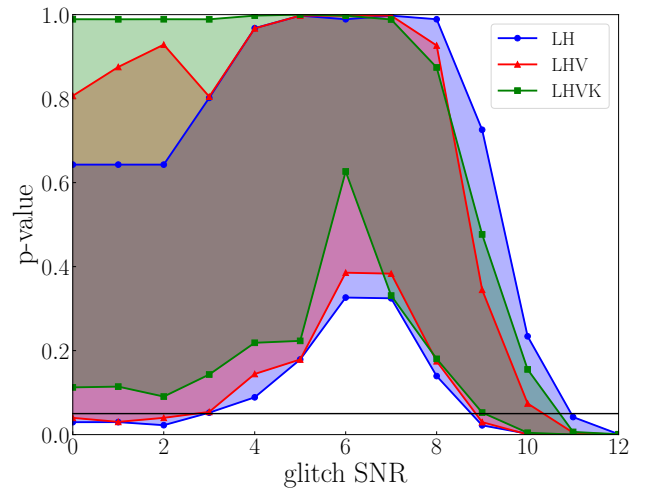

FIG. 6. P-values of different glitch intensities with different number of detectors. The legend provides information about the colors and labels associated with the curves representing different detectors. Specifically, "LH" represents the LIGO network, "LHV" represents the LIGO and Virgo network, and "LHVK" represents the LIGO, Virgo, and KAGRA network [133]. The black line represents a p-value of 0.05.

the glitch SNR is low, the mean value of the parameters does not change significantly, while the standard deviation increases noticeably. This behavior partially offsets the excessive confidence in the normalizing flow, leading to an increase in p-values in parameter inference calculations when glitches are absent. Therefore, when the glitch is small, the p-value will tend to increase. Since the p-values are bounded at 1, the expansion of the upper bounds of the error bars is constrained, while a more pronounced rise is observed in the lower bounds. This interplay between the inherent overconfidence of the network and the additional variance introduced by glitches leads to the tightening effect in Fig. 6 when the glitch SNR reaches 6.

Notably, the lower and upper limits of the error curves for the four detectors demonstrate a smoother downward trend compared to the error curves of the two or three detectors. The intersection point of the lower limit of error for these curves with a p-value of 0.05 occurs at a glitch SNR of approximately 9. This observation highlights that beyond a glitch SNR of 9, the inference results of the network's parameter estimation become unreliable for any number of detectors. The presence of glitches severely affects the accuracy of the parameter inference in such cases. This result indicates that as the intensity of the glitch increases, the influence of the glitch on different detector numbers gradually becomes more significant. Although there may be slight differences, overall, the impact is relatively consistent across the detectors.

B. The SNR of gravitational wave signal

In this section, we investigate the impact of signal SNR on the normalizing flow network's capability to facilitate GW parameter inference. By examining these metrics, we aim to gain insights into the normalizing flow network's performance in accurately estimating GW parameters.

We explore the effect of the SNR of signal on the robustness of the network, using a glitch with 8. For this purpose, we selected several parameters, namely, $\mathcal{M}_c$, q, d_L, δ, α. These parameters have a significant impact on GW signal simulation, parameter estimation, and sky localization. By considering these parameters in our analysis, we aim to comprehensively analyze their effects and gain a deeper understanding of their implications in GW studies. For quantitative comparison, we computed the JSD of the marginal distributions of the GW parameters between glitch SNR of 8 and without glitch, shown in Fig. 7.

The blue portion depicts the distribution characteristics of JSD values for the marginal distributions of parameters among the 2000 simulated GW events. To highlight the trend more explicitly, an orange line is plotted to showcase the overall variation in the number of events with JSD values less than 0.05. From the orange line, it is evident that the five parameters of GW exhibit varying

trends in terms of their JSD values. While the parameter q demonstrates a relatively flat trend, the remaining four parameters $\mathcal{M}_c$, d_L, δ, α show an increasing trend in the percentage of values with JSD values below 0.05, relative to the total number of values.

The black line indicates a JSD value of 0.05. Based on the assumption mentioned in the section, a JSD value below 0.05 indicates a difference between the two distributions that cannot be ignored. It also implies that the parameter inferences provided by the normalizing flow network are no longer reasonable. As signal SNR increases, the $\mathcal{M}_c$, d_L, δ, α parameter inference results are affected by glitches, and the greater SNR, the less likely these four parameter inference results are to be affected by glitches. In other words, a stronger SNR enhances the robustness of the parameter inference process and reduces the impact of glitches on the accuracy of the inferred values for $\mathcal{M}_c$, d_L, δ, α. The reasons for the trend of q parameter inference results are given in [128]. Due to the poor limiting accuracy of q, it shows that JSD itself is smaller in the same case with a large variance.

C. The SNR of the glitch

In this section, we explore the effect of the glitch's parameters and intensity on the robustness of the normalizing flow network. Despite the identification of sources for some glitches, a significant proportion of glitches remain of unknown origin. At the same time, due to the significant uncertainties associated with fault models, there are limitations to eliminating glitches through modeling techniques [70]. Here, we investigate the impact of glitches on parameter estimation. We aim to figure out the robustness of the normalizing flow network against unknown glitches, sequentially inferring parameters more quickly and accurately.

Here we investigate how the SNR of the glitches affects the ability of the network to infer GW parameters. Between the glitch-contaminated signals and Gaussian noise, we computed the JSD values for the marginal distributions of GW parameters at various SNRs of glitches. The results are presented in Fig. 8, providing a graphical representation of the JSD values and enabling a comparative analysis of the parameter fringes under different glitch SNRs. The black line on the graph represents the JSD threshold of 0.05. Except for the q parameter, whose JSD values are all below 0.05, the intersection of the error bars of the $\mathcal{M}_c$, d_L, δ, α parameters with a JSD of 0.05 lies between glitch SNR between 4 and 7. Among the parameters, the intersection point between the error bar of parameter $\mathcal{M}_c$ and JSD equal to 5 occurs at a glitch SNR between 4 and 5. In comparison to the other parameters, the posterior distribution of parameter $\mathcal{M}_c$ is more susceptible to contamination from glitches.

The parameters of glitch that we explore are those used in the simulation of glitch: f_0, h_0 and Q. Figure 9 shows the p-values plot of the normalizing flow network affected
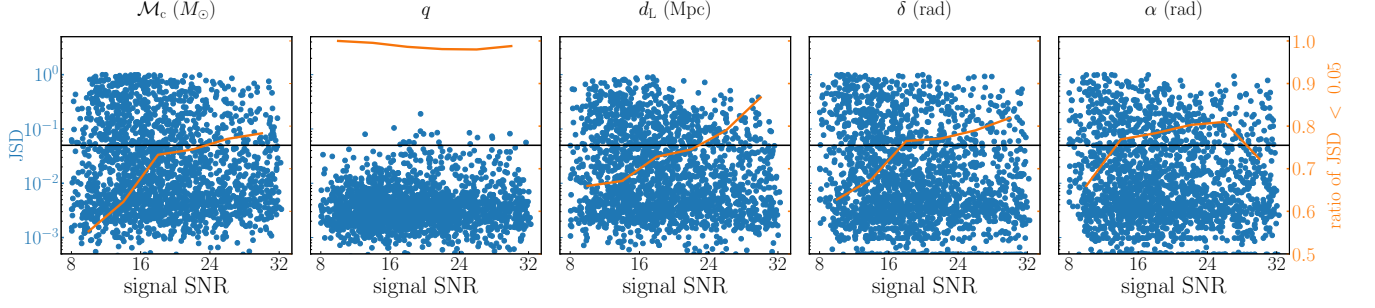
FIG. 7. JSD of GW parameter marginal distributions as a function of signal SNR between glitch SNR of 8 and without glitch. The blue scatter plot represents the distribution of JSD values for the marginal distributions of the parameters of the 2000 simulated GW events. The black line represents a JSD value of 0.05. The orange line represents the ratio of the number of events with JSD values less than 0.05 to the total number of events, for different signal SNR.
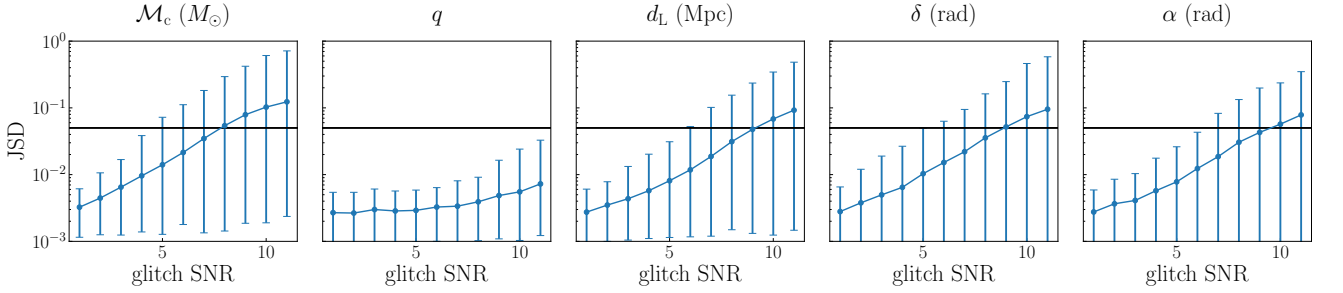

FIG. 8. JSD for GW parameter marginal distributions as different SNR of glitch between glitch-contaminated signals and Gaussian noise. The error bars on JSD represent the range of 90% confidence interval. The black line represents a JSD value of 0.05.
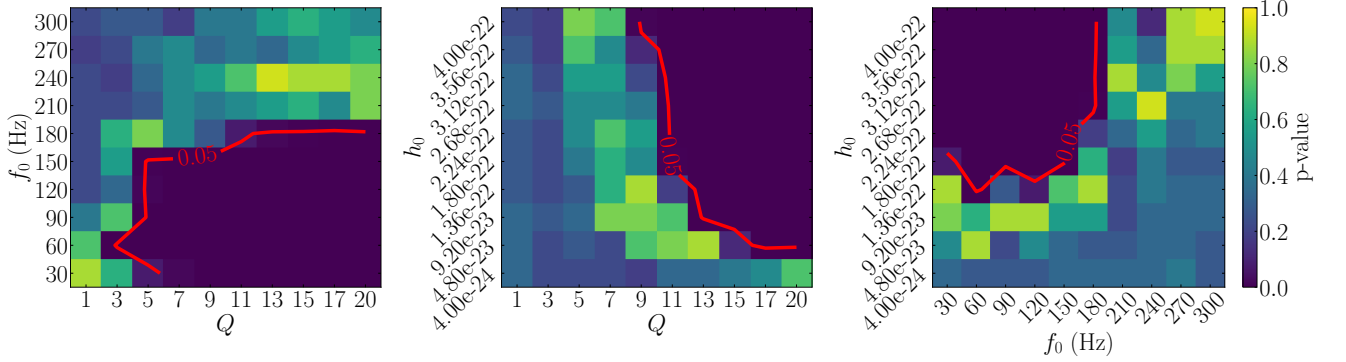

FIG. 9. The effect of glitch parameters on the inferred parameters of GW. Separate plots display the p-value distributions of the two-by-two combinatorial inference maps for the three fault parameters. The red line corresponds to a p-value of 0.05.

by the glitch parameter, drawn according to the KS test. We vary the glitch's frequency, strength, and duration and ensure they are within a reasonable range. The horizontal and vertical coordinates on the graph are the two combinations of the glitch's parameters. According to our assumption that the p-value is less than 0.05, the inference of network parameters is unreasonable. It can

be seen in Fig. 9 that each parameter of glitch has an impact on the accuracy of the network's inference of GW parameters. Specifically, as the duration of the glitch increases, along with a decrease in frequency and an increase in amplitude, the adverse effect on the rationality of parameter inference becomes more significant. Taking into account the upper limit of the GW signals frequency

range, which is 200 to 400 Hz, and the lower limit of 20 Hz, it is evident that the impact of glitches on the inference of parameter inference becomes increasingly significant as more glitches overlap in both time and frequency with the GW signals.

IV.   CONCLUSION

When analyzing and processing GW signals, glitches are a significant factor affecting the inference of GW parameters and source inversion. Several methods exist to model and remove glitches, but these approaches are more effective when the glitches are strong. For glitches with a SNR below 7, identifying them based on their morphology in images is challenging [135]. Consequently, removing these low-SNR glitches is difficult. To address this, we explore the removal of glitches by leveraging the inherent robustness of the normalizing flow network, without the need for explicit glitch modeling.

When GW data are contaminated by unknown non-Gaussian noise, it is possible to directly perform parameter inference using a normalizing flow network trained only on Gaussian noise. For unknown and unmodeled glitches, if the glitch SNR is within an appropriate range, the robustness of the normalizing flow network ensures reasonable parameter inferences. Additionally, we investigate the reasons behind the differences in the posterior distribution of source parameters observed between the normalizing flow and Bayesian inference methods, providing detailed explanations.

We also find that the parameters of the glitches significantly impact the robustness of parameter inference using the normalizing flow network, compared to the GW parameters. This highlights the importance of accurately estimating the characteristic parameters of glitches when dealing with data contaminated by them. When glitches and GW signals overlap in time and frequency, it can significantly affect the parameter inference of the normalizing flow network, reducing the network's robustness and making accurate inference challenging.

This study is the first to investigate the robustness of the normalizing flow network in the presence of non-stationary noise, demonstrating its practical feasibility for addressing unknown interference. Moreover, machine learning methods can provide prior support to traditional methods, accelerating the inference of GW signal parameters and improving accuracy to some extent. With future model improvements, we aim to extend this method to BNS systems, accelerating the search process for electromagnetic counterparts.
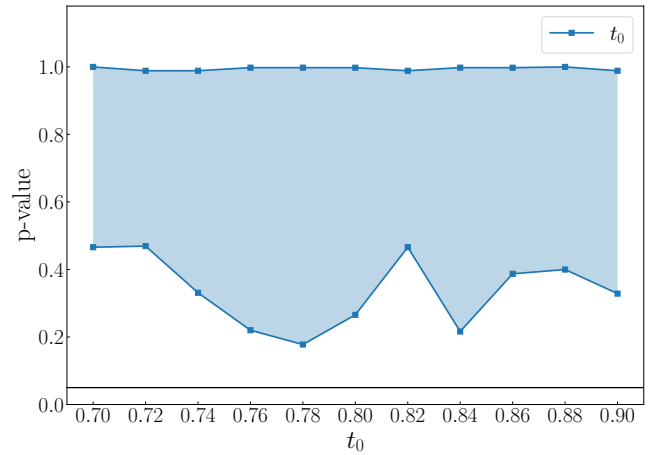
ACKNOWLEDGMENTS

FIG. 10. The effect of t_0 on the inference of network parameters. The variation of p-values over different t_0. The blue error bar represents the 1σ confidence intervals for 20 sets of results. The black line represents a p-value of 0.05.
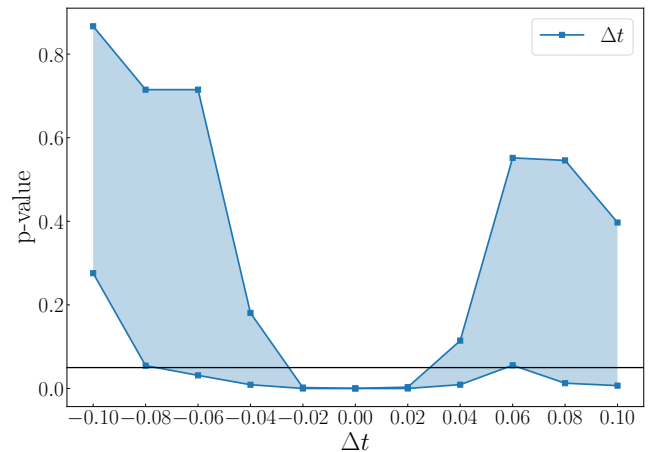

FIG. 11. The effect of Δt on the inference of network parameters. The p-values of parameter inference under different Δt. The blue curves illustrate the 1σ confidence bounds. The black line represents a p-value of 0.05.

Appendix A: Glitch time parameter

In this Appendix, we examine the impact of glitch central time t_0 and its interval with the GW signal Δt on the inference of network parameters.

To mitigate the impact of Δt, t_0 is uniformly distributed in $(0.7,0.9)$ s. Simultaneously, the GW signal is uniformly varied within a range of 0.1 s around, with adjustments ensuring its time remains between 0.7 and 0.9 s. The results are presented in Fig. 10. Figure 10 illustrates the variation in p-values for parameter inference using the normalizing flow network as t_0 changes. It depicts error bar for the 1σ confidence intervals for 20 sets of results. It shows that with the change in t_0, the variation of p-values exhibits a relatively smooth trend overall, without significant upward or downward trends. This suggests that a single change in t_0 does not significantly affect parameter inference robustness.

To study Δt between the glitch and GW signal, we varied Δt in $(-0.1, 0.1)$ s. Glitch times were generated from the GW signal time and Δt to assess its impact on parameter robustness inference. The results are shown in Fig. 11. Figure 11 displays the p-value for network parameter inference as Δt varies. The curves in Fig. 11 also represent error bar curves for 20 sets of data. It can be observed that when Δt is relatively large, indicating glitches occurring either before or after the signal, parameter inference remains within a reasonable range. However, when Δt is 0 s, indicating overlapping glitches and GW signals, the network parameter inference results are poorer, which is consistent with Ref. [49]. This indicates that Δt influences network parameter inference, with a significant impact when they overlap.

Since a single change in t_0 does not affect network parameter inference and the overlap between glitches and GW signals has a significant impact, we fixed the glitch time at 0.8 s in this work. This ensures that Δt is uniformly distributed in $(-0.1, 0.1)$ s, mitigating its influence on subsequent discussions.

[1] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].

[2] E. Berti *et al.*, Class. Quant. Grav. **32**, 243001 (2015), arXiv:1501.07274 [gr-qc].

[3] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **116**, 221101 (2016), [Erratum: Phys.Rev.Lett. 121, 129902 (2018)], arXiv:1602.03841 [gr-qc].

[4] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **123**, 011102 (2019), arXiv:1811.00364 [gr-qc].

[5] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **100**, 104036 (2019), arXiv:1903.04467 [gr-qc].

[6] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **103**, 122002 (2021), arXiv:2010.14529 [gr-qc].

[7] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2112.06861 [gr-qc].

[8] C. Gong, T. Zhu, R. Niu, Q. Wu, J.-L. Cui, X. Zhang, W. Zhao, and A. Wang, Phys. Rev. D **105**, 044034 (2022), arXiv:2112.06446 [gr-qc].

[9] C. Gong, T. Zhu, R. Niu, Q. Wu, J.-L. Cui, X. Zhang, W. Zhao, and A. Wang, Phys. Rev. D **107**, 124015 (2023), arXiv:2302.05077 [gr-qc].

[10] B. P. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER), Nature **551**, 85 (2017), arXiv:1710.05835 [astro-ph.CO].

[11] J. Sakstein and B. Jain, Phys. Rev. Lett. **119**, 251303 (2017), arXiv:1710.05893 [astro-ph.CO].

[12] H.-Y. Chen, M. Fishbach, and D. E. Holz, Nature **562**, 545 (2018), arXiv:1712.06531 [astro-ph.CO].

[13] M. Soares-Santos *et al.* (DES, LIGO Scientific, Virgo), Astrophys. J. Lett. **876**, L7 (2019), arXiv:1901.01540 [astro-ph.CO].

[14] L. Verde, T. Treu, and A. G. Riess, Nature Astron. **3**, 891 (2019), arXiv:1907.10625 [astro-ph.CO].

[15] S.-J. Jin, D.-Z. He, Y. Xu, J.-F. Zhang, and X. Zhang, JCAP **03**, 051 (2020), arXiv:2001.05393 [astro-ph.CO].

[16] A. Palmese *et al.* (DES), Astrophys. J. Lett. **900**, L33 (2020), arXiv:2006.14961 [astro-ph.CO].

[17] B. P. Abbott *et al.* (LIGO Scientific, Virgo, VIRGO), Astrophys. J. **909**, 218 (2021), arXiv:1908.06060 [astro-ph.CO].

[18] S.-J. Jin, L.-F. Wang, P.-J. Wu, J.-F. Zhang, and X. Zhang, Phys. Rev. D **104**, 103507 (2021), arXiv:2106.01859 [astro-ph.CO].

[19] L.-F. Wang, S.-J. Jin, J.-F. Zhang, and X. Zhang, Sci. China Phys. Mech. Astron. **65**, 210411 (2022), arXiv:2101.11882 [gr-qc].

[20] Z.-K. Guo, Science China. Physics, Mechanics & Astronomy **65**, 210431 (2022).

[21] S.-J. Jin, R.-Q. Zhu, L.-F. Wang, H.-L. Li, J.-F. Zhang, and X. Zhang, Commun. Theor. Phys. **74**, 105404 (2022), arXiv:2204.04689 [astro-ph.CO].

[22] R. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA), Astrophys. J. **949**, 76 (2023), arXiv:2111.03604 [astro-ph.CO].

[23] A. Palmese, C. R. Bom, S. Mucesh, and W. G. Hartley, Astrophys. J. **943**, 56 (2023), arXiv:2111.06445 [astro-ph.CO].

[24] J.-Y. Song, L.-F. Wang, Y. Li, Z.-W. Zhao, J.-F. Zhang, W. Zhao, and X. Zhang, Sci. China Phys. Mech. Astron. **67**, 230411 (2024), arXiv:2212.00531 [astro-ph.CO].

[25] T.-N. Li, S.-J. Jin, H.-L. Li, J.-F. Zhang, and X. Zhang, Astrophys. J. **963**, 52 (2024), arXiv:2310.15879 [astro-ph.CO].

[26] S.-J. Jin, Y.-Z. Zhang, J.-Y. Song, J.-F. Zhang, and X. Zhang, Sci. China Phys. Mech. Astron. **67**, 220412 (2024), arXiv:2305.19714 [astro-ph.CO].

[27] T. Han, S.-J. Jin, J.-F. Zhang, and X. Zhang, Eur. Phys. J. C **84**, 663 (2024), arXiv:2309.14965 [astro-ph.CO].

[28] S.-J. Jin, R.-Q. Zhu, J.-Y. Song, T. Han, J.-F. Zhang, and X. Zhang, JCAP **08**, 050 (2024), arXiv:2309.11900 [astro-ph.CO].

[29] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Class. Quant. Grav. **35**, 065010 (2018), arXiv:1710.02185 [gr-qc].

[30] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. X **9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[31] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. X **11**, 021053 (2021), arXiv:2010.14527 [gr-qc].

[32] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), Phys. Rev. X **13**, 041039 (2023), arXiv:2111.03606 [gr-

[33] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Class. Quant. Grav. **37**, 055002 (2020), arXiv:1908.11170 [gr-qc].

[34] D. Davis *et al.* (LIGO), Class. Quant. Grav. **38**, 135014 (2021), arXiv:2101.11673 [astro-ph.IM].

[35] F. Acernese *et al.* (Virgo), Class. Quant. Grav. **40**, 185006 (2023), arXiv:2210.15633 [gr-qc].

[36] M. Zevin *et al.*, Class. Quant. Grav. **34**, 064003 (2017), arXiv:1611.04596 [gr-qc].

[37] S. Soni *et al.*, Class. Quant. Grav. **38**, 195016 (2021), arXiv:2103.12104 [gr-qc].

[38] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, SoftwareX **12**, 100620 (2020), arXiv:2007.11374 [astro-ph.IM].

[39] R. Lynch, S. Vitale, R. Essick, E. Katsavounidis, and F. Robinet, Phys. Rev. D **95**, 104046 (2017), arXiv:1511.05955 [gr-qc].

[40] M. Razzano and E. Cuoco, Class. Quant. Grav. **35**, 095016 (2018), arXiv:1803.09933 [gr-qc].

[41] S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, and N. Cornish, Phys. Rev. D **106**, 042006 (2022), arXiv:2205.13580 [gr-qc].

[42] M. Zevin *et al.*, Eur. Phys. J. Plus **139**, 100 (2024), arXiv:2308.15530 [gr-qc].

[43] J. Aasi *et al.* (LIGO Scientific, VIRGO, NINJA-2), Class. Quant. Grav. **31**, 115004 (2014), arXiv:1401.0939 [gr-qc].

[44] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **116**, 241102 (2016), arXiv:1602.03840 [gr-qc].

[45] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. X **6**, 041015 (2016), [Erratum: Phys.Rev.X 8, 039903 (2018)], arXiv:1606.04856 [gr-qc].

[46] A. H. Nitz, Class. Quant. Grav. **35**, 035016 (2018), arXiv:1709.08974 [gr-qc].

[47] M. Cabero *et al.*, Class. Quant. Grav. **36**, 15 (2019), arXiv:1901.05093 [physics.ins-det].

[48] C. Pankow *et al.*, Phys. Rev. D **98**, 084016 (2018), arXiv:1808.03619 [gr-qc].

[49] R. Macas, J. Pooley, L. K. Nuttall, D. Davis, M. J. Dyer, Y. Lecoeuche, J. D. Lyman, J. McIver, and K. Rink, Phys. Rev. D **105**, 103021 (2022), arXiv:2202.00344 [astro-ph.HE].

[50] L.-G. Zhu and X. Chen, Astrophys. J. **948**, 26 (2023), arXiv:2302.10621 [astro-ph.CO].

[51] J. Powell, Class. Quant. Grav. **35**, 155017 (2018), arXiv:1803.11346 [astro-ph.IM].

[52] S. Soni *et al.* (LIGO), Class. Quant. Grav. **38**, 025016 (2020), arXiv:2007.14876 [astro-ph.IM].

[53] N. J. Cornish, Phys. Rev. D **103**, 104057 (2021), arXiv:2101.01188 [gr-qc].

[54] C. Talbot, E. Thrane, S. Biscoveanu, and R. Smith, Phys. Rev. Res. **3**, 043049 (2021), arXiv:2106.13785 [astro-ph.IM].

[55] B. Steltner, M. A. Papa, and H.-B. Eggenstein, Phys. Rev. D **105**, 022005 (2022), arXiv:2105.09933 [gr-qc].

[56] R. Udall and D. Davis, Appl. Phys. Lett. **122**, 094103 (2023), arXiv:2211.15867 [astro-ph.IM].

[57] S. D. Mohanty and M. A. T. Chowdhury, Class. Quant. Grav. **40**, 125001 (2023), arXiv:2301.02398 [gr-qc].

[58] S. A. Usman *et al.*, Class. Quant. Grav. **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].

[59] D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall, Class. Quant. Grav. **36**, 055011 (2019), arXiv:1809.05348 [astro-ph.IM].

[60] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton, Class. Quant. Grav. **39**, 245013 (2022), arXiv:2207.03429 [astro-ph.IM].

[61] N. J. Cornish and T. B. Littenberg, Class. Quant. Grav. **32**, 135012 (2015), arXiv:1410.3835 [gr-qc].

[62] T. B. Littenberg, J. B. Kanner, N. J. Cornish, and M. Millhouse, Phys. Rev. D **94**, 044050 (2016), arXiv:1511.08752 [gr-qc].

[63] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, Phys. Rev. D **103**, 044006 (2021), arXiv:2011.09494 [gr-qc].

[64] E. A. Huerta *et al.*, Nature Astron. **5**, 1062 (2021), arXiv:2012.08545 [gr-qc].

[65] R. Biswas *et al.*, Phys. Rev. D **88**, 062003 (2013), arXiv:1303.6984 [astro-ph.IM].

[66] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepanczyk, and S. Vitale, Phys. Rev. D **101**, 042003 (2020), arXiv:1911.09083 [gr-qc].

[67] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, Phys. Rev. Res. **2**, 033066 (2020), arXiv:2005.06534 [astro-ph.IM].

[68] J. Merritt, B. Farr, R. Hur, B. Edelman, and Z. Doctor, Phys. Rev. D **104**, 102004 (2021), arXiv:2108.12044 [gr-qc].

[69] H. Yu and R. X. Adhikari, Front. Artif. Intell. **5**, 811563 (2022), arXiv:2111.03295 [astro-ph.IM].

[70] G. Ashton, Mon. Not. Roy. Astron. Soc. **520**, 2983 (2023), arXiv:2209.15547 [gr-qc].

[71] S. Bini, G. Vedovato, M. Drago, F. Salemi, and G. A. Prodi, Class. Quant. Grav. **40**, 135008 (2023), arXiv:2303.05986 [gr-qc].

[72] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Phys. Rev. Lett. **120**, 141103 (2018), arXiv:1712.06041 [astro-ph.IM].

[73] A. J. K. Chua, C. R. Galley, and M. Vallisneri, Phys. Rev. Lett. **122**, 211101 (2019), arXiv:1811.05491 [astro-ph.IM].

[74] H. Wang, S. Wu, Z. Cao, X. Liu, and J.-Y. Zhu, Phys. Rev. D **101**, 104003 (2020), arXiv:1909.13442 [astro-ph.IM].

[75] M. Chen, Y. Zhong, Y. Feng, D. Li, and J. Li, Sci. China Phys. Mech. Astron. **63**, 129511 (2020), arXiv:2003.13928 [astro-ph.IM].

[76] M. Cabero, A. Mahabal, and J. McIver, Astrophys. J. Lett. **904**, L9 (2020), arXiv:2010.11829 [gr-qc].

[77] S. R. Green and J. Gair, Mach. Learn. Sci. Tech. **2**, 03LT01 (2021), arXiv:2008.03312 [astro-ph.IM].

[78] A. Gunny, D. Rankin, J. Krupa, M. Saleem, T. Nguyen, M. Coughlin, P. Harris, E. Katsavounidis, S. Timm, and B. Holzman, Nature Astron. **6**, 529 (2022), arXiv:2108.12430 [gr-qc].

[79] W.-H. Ruan, H. Wang, C. Liu, and Z.-K. Guo, Phys. Lett. B **841**, 137904 (2023), arXiv:2111.14546 [astro-ph.IM].

[80] T. Zhao, R. Lyu, H. Wang, Z. Cao, and Z. Ren, Commun. Phys. **6**, 212 (2023), arXiv:2207.07414 [gr-qc].

[81] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **130**, 171403 (2023), arXiv:2210.05686 [gr-qc].

[82] W. Ruan, H. Wang, C. Liu, and Z. Guo, Universe **9**, 407 (2023), arXiv:2307.14844 [astro-ph.IM].

[83] Q. Yun, W.-B. Han, Y.-Y. Guo, H. Wang, and M. Du, (2023), arXiv:2311.18640 [gr-qc].

[84] C. Ma, S. Wang, W. Wang, and Z. Cao, Phys. Rev. D **109**, 043009 (2024), arXiv:2309.02030 [astro-ph.IM].

[85] T. Zhao, Y. Zhou, R. Shi, Z. Cao, and Z. Ren, Phys. Rev. D **109**, 084054 (2024), arXiv:2308.16422 [astro-ph.IM].

[86] M. J. Williams, J. Veitch, and C. Messenger, Phys. Rev. D **103**, 103006 (2021), arXiv:2102.11056 [gr-qc].

[87] D. Ruhe, K. Wong, M. Cranmer, and P. Forré, (2022), arXiv:2211.09008 [astro-ph.IM].

[88] J. Langendorff, A. Kolmus, J. Janquart, and C. Van Den Broeck, Phys. Rev. Lett. **130**, 171402 (2023), arXiv:2211.15097 [gr-qc].

[89] M. Du, B. Liang, H. Wang, P. Xu, Z. Luo, and Y. Wu, Sci. China Phys. Mech. Astron. **67**, 230412 (2024), arXiv:2308.05510 [astro-ph.IM].

[90] T.-Y. Sun, Y. Shao, Y. Li, Y. Xu, and X. Zhang, (2024), arXiv:2407.14298 [astro-ph.CO].

[91] X. Fan, J. Li, X. Li, Y. Zhong, and J. Cao, Sci. China Phys. Mech. Astron. **62**, 969512 (2019), arXiv:1811.01380 [astro-ph.IM].

[92] A. J. K. Chua and M. Vallisneri, Phys. Rev. Lett. **124**, 041102 (2020), arXiv:1909.05966 [gr-qc].

[93] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **127**, 241103 (2021), arXiv:2106.12594 [gr-qc].

[94] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Nature Phys. **18**, 112 (2022), arXiv:1909.06296 [astro-ph.IM].

[95] T. Zhao, Y. Zhou, R. Shi, Z. Cao, and Z. Ren, (2024), arXiv:2402.02825 [gr-qc].

[96] M. Sun and J. Li, (2023), arXiv:2312.07834 [gr-qc].

[97] T.-Y. Sun, C.-Y. Xiong, S.-J. Jin, Y.-X. Wang, J.-F. Zhang, and X. Zhang, Chin. Phys. C **48**, 045108 (2024), arXiv:2312.08122 [gr-qc].

[98] W. Wei and E. A. Huerta, Phys. Lett. B **800**, 135081 (2020), arXiv:1901.00869 [gr-qc].

[99] C. Chatterjee, L. Wen, F. Diakogiannis, and K. Vinsen, Phys. Rev. D **104**, 064046 (2021), arXiv:2105.03073 [gr-qc].

[100] H. Wang, Y. Zhou, Z. Cao, Z. Guo, and Z. Ren, Mach. Learn. Sci. Tech. **5**, 015046 (2024), arXiv:2212.14283 [gr-qc].

[101] Y.-X. Wang, S.-J. Jin, T.-Y. Sun, J.-F. Zhang, and X. Zhang, Chin. Phys. C **48**, 125107 (2024), arXiv:2305.19003 [gr-qc].

[102] Y. Xu, M. Du, P. Xu, B. Liang, and H. Wang, Phys. Lett. B **858**, 139016 (2024), arXiv:2402.13091 [gr-qc].

[103] J. Powell, D. Trifirò, E. Cuoco, I. S. Heng, and M. Cavaglià, Class. Quant. Grav. **32**, 215012 (2015), arXiv:1505.01299 [astro-ph.IM].

[104] S. Bose, S. Dhurandhar, A. Gupta, and A. Lundgren, Phys. Rev. D **94**, 122004 (2016), arXiv:1606.06096 [gr-qc].

[105] G. Ashton *et al.*, Astrophys. J. Suppl. **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[106] J. S. Speagle, Mon. Not. Roy. Astron. Soc. **493**, 3132 (2020), arXiv:1904.02180 [astro-ph.IM].

[107] T. D. Canton, S. Bhagwat, S. V. Dhurandhar, and A. Lundgren, Class. Quant. Grav. **31**, 015016 (2014), arXiv:1304.0008 [gr-qc].

[108] D. Rezende and S. Mohamed, in *International conference on machine learning* (PMLR, 2015) pp. 1530–1538.

[109] G. Papamakarios and I. Murray, Advances in neural information processing systems **29** (2016).

[110] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, Advances in neural information processing systems **32** (2019), 1906.04032.

[111] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, Advances in neural information processing systems **29** (2016).

[112] X. Chen, D. P. Kingma, T. Salimans, Y. Duan, P. Dhariwal, J. Schulman, I. Sutskever, and P. Abbeel, in *International Conference on Learning Representations* (2016).

[113] G. Papamakarios, T. Pavlakou, and I. Murray, Advances in neural information processing systems **30** (2017).

[114] D. J. Rezende, S. Mohamed, and D. Wierstra, in *International conference on machine learning* (PMLR, 2014) pp. 1278–1286.

[115] K. He, X. Zhang, S. Ren, and J. Sun, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 770–778.

[116] I. Loshchilov and F. Hutter, in *International Conference on Learning Representations* (2018).

[117] A. Bohé *et al.*, Phys. Rev. D **95**, 044028 (2017), arXiv:1611.03703 [gr-qc].

[118] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, Publ. Astron. Soc. Pac. **131**, 024503 (2019), arXiv:1807.10312 [astro-ph.IM].

[119] T. Dal Canton *et al.*, Phys. Rev. D **90**, 082004 (2014), arXiv:1405.6731 [gr-qc].

[120] A. Nitz and othes, "gwastro/pycbc: v2.0.5 release of pycbc," (2022).

[121] L. Sellers, A. Bobrick, G. Martire, M. Andrews, and M. Paulini, (2022), arXiv:2212.02065 [astro-ph.IM].

[122] B. J. Owen and B. S. Sathyaprakash, Phys. Rev. D **60**, 022002 (1999), arXiv:gr-qc/9808076.

[123] L. Blackburn *et al.*, Class. Quant. Grav. **25**, 184004 (2008), arXiv:0804.0800 [gr-qc].

[124] R. H. C. Lopes, "Kolmogorov-smirnov test," in *International Encyclopedia of Statistical Science*, edited by M. Lovric (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011) pp. 718–720.

[125] T. Sidery *et al.*, Phys. Rev. D **89**, 084060 (2014), arXiv:1312.6013 [astro-ph.IM].

[126] J. Veitch *et al.*, Phys. Rev. D **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].

[127] E. Thrane and C. Talbot, Publications of the Astronomical Society of Australia **37**, e036 (2020).

[128] J. Lin, IEEE Transactions on Information Theory **37**, 145 (1991).

[129] C. R. Rao, Differential geometry in statistical inference **10**, 217 (1987).

[130] I. Csiszar, The Annals of Probability **3**, 146 (1975).

[131] S. Kullback and R. A. Leibler, The Annals of Mathematical Statistics **22**, 79 (1951).

[132] I. M. Romero-Shaw *et al.*, Mon. Not. Roy. Astron. Soc. **499**, 3295 (2020), arXiv:2006.00714 [astro-ph.IM].

[133] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, Virgo, VIRGO), Living Rev. Rel. **21**, 3 (2018), arXiv:1304.0670 [gr-qc].

[134] J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel,

V. Begy, and G. Louppe, "A trust crisis in simulation-based inference? your posterior approximations can be unfaithful," (2022), arXiv:2110.06581 [stat.ML].

[135] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, J. R. Smith, V. Kalogera, and A. Katsaggelos, Information Sciences **444**, 172 (2018).