

Joint inference for gravitational wave signals and glitches using a data-informed glitch model

Ann-Kristin Malz 

*School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom
and Centre for Particle Physics and Astronomy, Royal Holloway University of London,
Egham TW20 0EQ, United Kingdom*

John Veitch 

School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom



(Received 2 May 2025; accepted 25 June 2025; published 30 July 2025)

Gravitational wave data are often contaminated by non-Gaussian noise transients, “glitches,” which can bias the inference of astrophysical signal parameters. Traditional approaches either subtract glitches in a preprocessing step, or a glitch model can be included from an agnostic wavelet basis (e.g., BayesWave). In this work, we introduce a machine-learning-based approach to build a parametrized model of glitches. We train a normalizing flow on known glitches from the Gravity Spy catalog, constructing an informative prior on the glitch model. By incorporating this model into the Bayesian inference analysis with Bilby, we estimate glitch and signal parameters simultaneously. We demonstrate the performance of our method through bias reduction, glitch identification, and Bayesian model selection on real glitches. Our results show that this approach effectively removes glitches from the data, significantly improving source parameter estimation and reducing bias.

DOI: [10.1103/fp4b-mvzx](https://doi.org/10.1103/fp4b-mvzx)

I. INTRODUCTION

Since the first detection [1], the field of gravitational wave (GW) astronomy has now progressed toward regular observations. During the first three observing runs, around 100 signals were detected [2] by the interferometers of the LIGO-Virgo-KAGRA (LVK) Collaboration [3–5]. The sensitivity of the detectors is limited by background noise, comparable in amplitude to a GW signal from an astrophysical source, approximated as quasistationary colored Gaussian noise, as well as non-Gaussian noise transients known as glitches [6]. Much of the background noise can be filtered from the data, for example, by excluding frequencies of known noise sources such as the electrical power grid [6] and filtering frequencies outside the expected range of astrophysical signals.

Glitches, however, are not filtered from the data by these generic methods and require special attention. They are troublesome, as they can be mistaken for astrophysical signals or bias the parameter estimation results of detected signals. Additionally, they often have unknown physical

origins and/or are difficult to mitigate in the detectors [7], while occurring at a rate of approximately one per minute [8].

Glitches are found in the data by Omicron [9,10], which searches the GW strain data of each detector for excess power and characterizes properties such as frequency, duration, and amplitude of each glitch. There are a variety of different glitch classes, with different underlying causes and distinct features in the data. The classification of glitches depending on features in the time-frequency domain is addressed by the Gravity Spy project for LIGO [11,12] and GWitchHunters for Virgo [13].

In the third observing run of LIGO and Virgo, around 20% of the GW signals detected contained glitches that had to be mitigated [14]. To improve the detection of GWs, the causes of the glitches must be identified and eliminated or, alternatively, the glitches removed from the data. It has been shown that bias introduced to signal parameter estimation due to glitches can be mitigated by removing the glitches from the data preanalysis [15].

Glitch mitigation is an area of increasing interest as the number of detected gravitational wave signals grows. The third observing run detected several signals [2,16,17] where glitches had to be removed before source parameter inference [14]. The earliest and clearest example of a glitch impacting analysis was GW170817, the first binary neutron star detection [18]. In this case, the glitch was

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

sufficiently simple that it could be subtracted by a manual procedure from the data, followed by regular inference of the source parameters [19]. This iterative subtraction procedure works well if there is little interference between the signal and the glitch (i.e., a low inner product between the glitch and the signal waveform), and the glitch is easily modeled; but in the general case, one would rather perform simultaneous fitting of signal and glitch, accompanied by statistical uncertainty on the parameters of both. Examples of signals where the glitch removal was less clear cut are GW191109 [2,14,20] and GW200129 [2,14,21,22].

When using a traditional stochastic sampler (Markov chain Monte Carlo or nested sampling) for performing inference, one requires a parametric model of the glitches that can be included as part of the model of the data, possibly in addition to a signal. Such a model could be based on a very agnostic, phenomenological description, such as wavelets, or on a physical understanding of the glitch generation process (e.g., Scattered Light). From a Bayesian point of view, these correspond to different strengths of prior information assumed in the model.

In this work, we develop a model that lies between these extremes. By examining known glitches from the Gravity Spy dataset [23], we can find their dominant components and train a normalizing flow (NF) to describe the joint prior distribution of these components. In this way, our method incorporates physical information from previous glitches while retaining flexibility in its parametrization. We demonstrate that this model can successfully mitigate parameter estimation bias caused by glitches.

A. Related work

There has been much work on analyzing and mitigating glitches in GW data, using a variety of tools and methods. Some of these focus on modeling only the glitch in order to detect and flag these times, while others aim to include glitch models as part of general inference. Below, we review similar works in the field and their relationship to our own method.

BayesWave [24] is the most established glitch removal algorithm in use in the LVK Collaboration and is routinely used when glitches contaminate data surrounding a detected signal. BayesWave is a Bayesian inference-based algorithm where non-Gaussian features (glitches and signals) are modeled as a sum of Morlet-Gabor (sine-Gaussian) wavelets. The total model can contain a variable number of wavelet components, whose parameters are sampled using the transdimensional jump Markov chain Monte Carlo method, giving it great flexibility at identifying generic transients. BayesWave has thus proven very useful at modeling and removing glitches from the strain data, however, there are limitations. One

disadvantage is that the glitch model constructed from the wavelets does not “know” anything about real glitches, but simply models the patterns observed in the data. Davis *et al.* [14] discuss this further and highlight the importance and need for different glitch mitigation methods suitable for the wide range of possible glitch and signal scenarios. Plunkett *et al.* [25] simultaneously estimate compact binary and noise parameters using the BayesLine algorithm for power spectral density (PSD) estimation [26], as well as the BayesWave sine-Gaussian wavelet glitch model. Simultaneous glitch and signal modeling with BayesWave is described in Refs. [27,28].

The other primary glitch removal method used in the third observing run is GWSUBTRACT [29]. The algorithm uses information from an auxiliary witness channel (correlated with, but independent of the strain data) to model and subtract glitches.

Ashton [30] uses a Gaussian process to model long-duration glitches and perform inference on a combined signal + glitch model by including the Gaussian process in the Bayesian noise likelihood, although the glitch model is not informed by data. Merritt *et al.* [31] developed a glitch model (“Glitschen”) based on probabilistic principal component analysis for dimensional reduction, followed by a multivariate normal model for the resulting components. This allows the glitch prior to be informed by the set of observed glitches, similar to our approach; however, the prior is limited to those that can be captured by the multivariate normal distribution. Unlike our work, it does not integrate a signal model in the analysis. Mohanty and Chowdhury [32] make use of the adaptive spline fitting method SHAPES to subtract broadband, short-duration glitches (Blip, Koi Fish, and Tomte [11]). They inject a binary neutron star signal overlapping the glitch to imitate a GW170817-like case and also apply their model to GW170817 directly. Udall *et al.* [20] implement a physically motivated model for Scattered Light glitches and perform joint Bayesian inference of the glitch and signal, extending the work done in [33].

The advent of machine learning (ML) methods has led to a range of novel approaches to glitch identification and mitigation. Gravity Spy [11,12] is a citizen-science project that leverages ML to classify glitches by their morphological features. The algorithm consists of a convolutional neural network [34–36], which is trained on time-frequency-energy plots, known as omega scans or Q transforms [37], that have been manually classified by experts and volunteers. Both citizen scientists and the trained ML algorithm contribute to the classification of new glitches. In our work, we use the classifications made by Gravity Spy to build a training dataset of glitches.

Vajente *et al.* [38] developed a ML method to model noise transients in the strain channel based on information from auxiliary channels and demonstrated a reduction in

parameter estimation bias when glitches were subtracted before inference. Wang *et al.* [39] developed a deep neural network approach, WaveFormer, for noise and glitch suppression in GW data, resulting in a cleaned data stream for further analyses. Macas *et al.* [22] used a neural network to subtract broadband noise in the Livingston detector observed around the signal GW200129. Their network was trained to model the strain data $h(t)$ as a function of three auxiliary channels. Lopez *et al.* [40] used generative adversarial networks (GANs), to learn the distribution of Blip glitches and to generate artificial populations. Similarly, Dooney *et al.* [41] used a GAN (conditional derivative GAN) to create a model that can be used to generate Blip and Tomte glitches and binary black hole signals. Li *et al.* [42] created an unsupervised model using an autoencoder, CTSAE, to cluster glitches and identify glitch classes similar to Gravity Spy. Bondarescu *et al.* [43] present a quasiphysical, four-parameter, glitch model, Antigitch, to model and thus subtract Blip, low-frequency Blip, Tomte, and Koi Fish glitches. Sun *et al.* [44] used a normalizing flow for likelihood-free inference of GW source parameters, based on training data that included Blip and Scattered Light glitches. In Xiong *et al.* [45], the authors then extended the work to not rely on glitch modeling by training on signals in Gaussian noise. Applying it to signals contaminated by injected glitches modeled as sine-Gaussians, they showed that the posterior from their normalizing flow is less biased compared to the standard inference code Bilby [46,47]. Legin *et al.* [48] used a score-based model instead of the traditional Whittle likelihood to relax the assumption of Gaussian noise in doing parameter estimation.

Our approach has a similar motivation to some of the above works—the desire to remove glitches from the data and thus improve inference—but we offer some key differences. Primarily, we implement our glitch model into the existing, well-established signal analysis code Bilby. Except for Ashton [30] and Udall *et al.* [20], the above-mentioned papers developed independent workflows—instead building on the established code simplifies the usage of the method for downstream users. Furthermore, by analyzing glitch and signal simultaneously with Bilby, our work can be easily integrated into LVK analysis workflows. Second, by training on real glitches, our model can capture the variety of features present within a glitch class, and the result of the training is a proper Bayesian prior on the glitch parameters. Third, while we mainly focus on demonstrating the robustness of our model on Blip glitches in this paper, the method is applicable to any type of glitch, provided a new model is trained.

This paper is structured as follows: Sec. II briefly introduces NFs and Bayesian inference, Sec. III details our glitch-model-and-subtracting method, and in Sec. IV we perform various analysis to test our model. Section V discusses our conclusions.

II. BACKGROUND

A. Normalizing flows

Normalizing flows [49–51] are a type of generative ML algorithm that can explicitly learn the probability density function of given data. A NF maps a complex distribution $p_x(x)$, such as a set of data, to a simple latent distribution $p_z(z)$, such as a uniform or Gaussian distribution, by applying one or multiple transformation functions. These functions f are bijections, and the inverse can thus be used to generate a realization of the data, such that $z = f(x)$ and $x = f^{-1}(z)$. Since the transformation is built by composing many layers of simple transformations, it is possible to evaluate its Jacobian determinant, so that the probability distribution of the data can be defined as

$$p_x(x) = p_z(f(x)) \left| \det \left(\frac{\partial f(x)}{\partial x} \right) \right|, \quad (1)$$

where $\partial f(x)/\partial x$ is the Jacobian of the transformation function f . With many layers of transformation, a complex distribution can be approximated by a simple one. The NF is trained to learn the transformation function f and the Jacobian determinant, which can be done by minimizing the negative log-likelihood (Kullback-Leibler divergence) of the data. Applying the inverse transform f^{-1} of the trained flow to the latent space p_z , samples of the data distribution p_x can be obtained. There are different types of NFs, with different properties, as described in [49]. Autoregressive flows are generally more expressive but computationally expensive to train, while coupling flows are less flexible but computationally efficient. In this work, we choose the latter for their efficiency.

B. Bayesian inference

In this work, we adopt a Bayesian approach to data analysis. Bayesian inference describes how prior knowledge can be updated when observations provide new evidence through Bayes' theorem,

$$p(\theta|d, H) = \frac{p(d|\theta, H)p(\theta|H)}{P(d|H)}, \quad (2)$$

which defines the probability density as a function of parameters θ , given the observed data d and postulated model H . Here the posterior $P(\theta|d, H)$ depends on the likelihood $P(d|\theta, H)$, the prior $P(\theta|H)$, and the evidence $P(d|H) \equiv \mathcal{Z}$.

The likelihood describes the probability of the data for the model given the parameters and is chosen based on the problem, as it depends on the noise model. For GW analysis, Gaussian noise is usually assumed [52]. In this project, we build a glitch model and implement it as the Bayesian prior distribution over the glitches. The likelihood can then be calculated from the data when modeled

as including a glitch and, optionally, a signal. If a signal $s(t)$ is present, the data $d(t)$ can be expressed as $d(t) = s(t) + n(t) + g(t)$, where $n(t)$ is the Gaussian noise and $g(t)$ the glitch.

In addition to parameter estimation, Bayesian inference can also be used for model selection. The Bayesian evidence $\mathcal{Z}$ is the normalization factor in Bayes' theorem and can be written as

$$\mathcal{Z} = \int p(d|\theta, H) p(\theta|H) d\theta. \quad (3)$$

Comparing the evidence for different models indicates which model is statistically preferred. The Bayes factor is the evidence ratio of two different models and can thus be used to compare, for example, a signal model to a Gaussian noise model to determine if there is a signal present. The log of the Bayes factor, $\log \text{BF} = \log(\mathcal{Z}_1) - \log(\mathcal{Z}_2)$, is used for convenience in this paper.

III. METHOD

The aim of this work is to model and remove glitches from the LIGO GW data simultaneously with the signal analysis being performed. We accomplish this by training a NF on glitches in the Gravity Spy O1 training dataset [53], and implementing the trained NF model into the Bayesian GW inference library Bilby [46] as a prior for the glitches. Thus, we can perform a single analysis on a GW event, where the glitch is removed and the signal is analyzed.

The work consists of two parts, training and analysis. During training, we train a NF to build the glitch model. The NF can be trained to learn the shape of the glitches and the transformation between the glitch model and a Gaussian distribution. The Gravity Spy project provides data such as the time and duration of individual glitches for many different glitch classes. For this initial study, we primarily focus on Blip glitches, characterized by their short time duration (~ 10 ms) and large frequency bandwidth (~ 100 Hz). A typical Blip glitch is shown in Figs. 3 (time domain) and 4 (time-frequency spectra). Blip glitches occur in both LIGO detectors, with an average of two Blips per hour, and are of high interest due to their shape resembling the signal of high-mass compact binary mergers and due to their origins being largely unknown [54]. The Blip glitches were chosen as the first glitch type to model, due to their short duration and characteristic shape, simplifying the analysis. Our method is straightforward to apply to other glitch classes, and a brief investigation shows similar performance (see Sec. V). However, we will leave a full investigation of other glitch classes to future work.

A. Training data

The GW strain data are publicly available from the Gravitational Wave Open Science Center (GWOSC) [55]

and were accessed using the GWpy Python package [56]. Using GWOSC data, we create a training dataset consisting of 1 s of strain data around each of the Blip glitches in the Gravity Spy training set with durations < 1 s (a total of 1785 glitches). The data are filtered to remove noise from outlying frequencies, using a bandpass filter, and whitened to normalize the power at all frequencies so that the glitches become clearer. Next, we cut the strain data to the desired length, and then we apply a Hann window to ensure the model tapers smoothly to zero at the limits of its time stretch. See Sec. III B for further discussion.

Having compiled our training dataset, we use singular value decomposition (SVD) to reduce its dimensionality, making it simpler to describe with the NF. SVD factorizes the dataset into three separate matrices such that $A_{m \times n} = U_{m \times m} S_{m \times n} V_{n \times n}^T$, where U and V^T contain the orthonormal eigenvectors of AA^T and $A^T A$, respectively, and S is a diagonal matrix consisting of the singular values (the square roots of the eigenvalues of $A^T A$) along the main diagonal. Combining U and S such that $T = US$, the training dataset of glitches, G , can be described by $G = TV^T$, where V^T is the basis matrix containing the time series of all the glitch components, and T represents the weights, or glitch amplitudes, assigned to each basis. Thus, the NF can be trained on T only, and the glitch is later reassembled using the saved V^T . The dimensions of T can be reduced by only keeping the most significant eigenvectors. The cutoff can be obtained by limiting the power loss allowed. Here the cutoff was chosen so that 97% of the total power is contained in the remaining bases so that each glitch can be described by around ten parameters, depending on the filtering and frequency bands used.

B. Choosing the model

Glitches from the Gravity Spy training set [53] from the first observing run, O1, were used as the training data for this paper. Glitches identified by Gravity Spy in O1, O2, O3a, and O3b [23, 57] were used to test the model. As the model is easy to retrain, other training datasets could also be explored.

The data were filtered to remove noise from outlying frequencies, using a bandpass filter between 20 and 400 Hz. The bandpass was chosen as the majority (92%) of the glitches in the training set have peak frequencies within this range, and we wish to exclude higher frequencies where lines in the noise spectrum due to, e.g., violin modes produce additional features that are imperfectly removed by whitening with a representative spectrum (see below). With further data treatment, the frequency range may be adapted in the future to include a different range of frequencies depending on the aim of the analysis and the properties of the targeted glitches. However, the number of bases needed to model a glitch increases with a broader bandwidth, and an increased number of parameters in the glitch model entails longer run times when applying the model to data.

We find that this range is suitable for the glitches considered in this initial study.

The data are then whitened before training the normalizing flow, using the O1 representative amplitude spectral density (ASD) for each of the LIGO detectors [55]. This normalizes the power at all frequencies, reducing the noise level and scaling the data to values of order 1. The latter is important to improve the performance of the machine learning algorithm. Since the glitch model is thus whitened, we need to unwhiten it before applying it to a glitch. This is done using the ASD calculated from the real data used in the inference analysis, so that the glitch model matches the data properties.

Another consideration is the length in time of the model, as a longer model also leads to more bases. For this project, Blip glitches with durations up to 1 s (according to the Omicron [9,10] values reported in the Gravity Spy dataset [53]) were chosen. However, the actual peaks of the glitches are significantly shorter than this duration, generally around 10 ms long. Thus, the data chunks can be shortened further to 1/8 s, reducing the number of model parameters and decreasing run times. After applying the Hann window to these reduced-length data chunks, this still leaves the model long enough to effectively capture the Blip glitches considered here. Choosing the time and frequency constraints as discussed, we obtain a glitch model that is described by 12 parameters.

Figure 1 shows the components of the V^T matrix, the basis functions from which the glitches are built, in both the time and frequency domain, illustrating the most important features for reconstructing a Blip glitch as given in our model.

C. Training the flow

We train a neural spline flow (CouplingNSF [58]) on the reduced glitch amplitudes T , using the Python library GLASFLOW [59]. The trained flow describes the prior distribution of the glitch parameters T and can be easily sampled by drawing latent parameters z from a Gaussian distribution and passing them through the inverse flow to reconstruct T values corresponding to glitches. Recombining these with the basis matrix V^T then gives the parametrized glitch model we use in our analysis. This process is illustrated in Fig. 2.

D. Applying the glitch model

Once we have a trained glitch model, we can apply it with Bilby to remove glitches from GW data. We define a Bilby prior class to contain the trained NF, and a likelihood class, which can simulate the data containing the glitch along with the GW signal. The nested sampler in Bilby (we use Nessai [60]) thus fits the glitch model to the glitch in the data as well as estimates the posterior of the signal.

The glitch is handled by subtracting its waveform from the data. First, the amplitudes T are sampled from the saved

NF in the priors and passed to the likelihood as parameters for the glitch model. Within the likelihood, the glitch is then reconstructed using the basis matrix V^T , and the residual (the glitch-subtracted data) is calculated. The log-likelihood used for the signal posterior inference is thus determined from the residual, rather than the glitch-contaminated data.

We introduce two additional unknown parameters at inference time to give the glitch model greater flexibility. A time shift is added to ensure that the glitch model can be accurately fitted to the data even if the trigger time is slightly offset, within a Gaussian prior with standard deviation 10 ms. We also add a scaling factor A to the glitch model, with a prior $\propto A^{-1}$ within the range $10^{-3} - 10^3$. This allows us to adjust the overall scale of the glitch while maintaining the shape as encoded by the relative size of the T parameters.

IV. ANALYSIS

We performed a thorough investigation of the glitch model's performance, looking at both model selection and parameter estimation results, with and without a signal present. In Sec. IV A, we test the performance of our model in identifying glitches via Bayesian model selection. We analyze interferometer data with known glitches identified by Gravity Spy, as well as simulated Gaussian noise, and compare the distribution of log Bayes factors for the glitch model and the Gaussian noise model. In Sec. IV B, we use the glitch model to analyze data containing signals, but no glitches, and check that the model does not misidentify signals as glitches (i.e., that it is not preferred over the signal model itself). In Sec. IV C, we test the performance of our model in the scenario where both a signal and a glitch may be present in the data. We use model selection to distinguish between signal-only, glitch-only, and signal + glitch models, with a signal injected into data containing glitches. We also investigate the effect of varying the time of the glitch relative to the peak of the signal. In Sec. IV D, we investigate the bias in the recovered parameters of the injected signal when using the glitch model and compare it to the bias in the case where no glitch model is used. Finally, in Sec. IV E we experiment with using the Blip glitch model to remove Tomte glitches.

A. Glitch-only fitting

We first wish to establish that the glitch model is functioning well and can distinguish glitches from Gaussian data. We investigate the recovery of glitches from the training dataset, as well as model unseen glitches. To determine if the glitch model is favored or disfavored for given data, we compute the Bayes factor between the glitch model and the Gaussian noise model. A log Bayes factor greater than 0 implies that the glitch model is favored. Ideally, we would like a clean separation of results so that

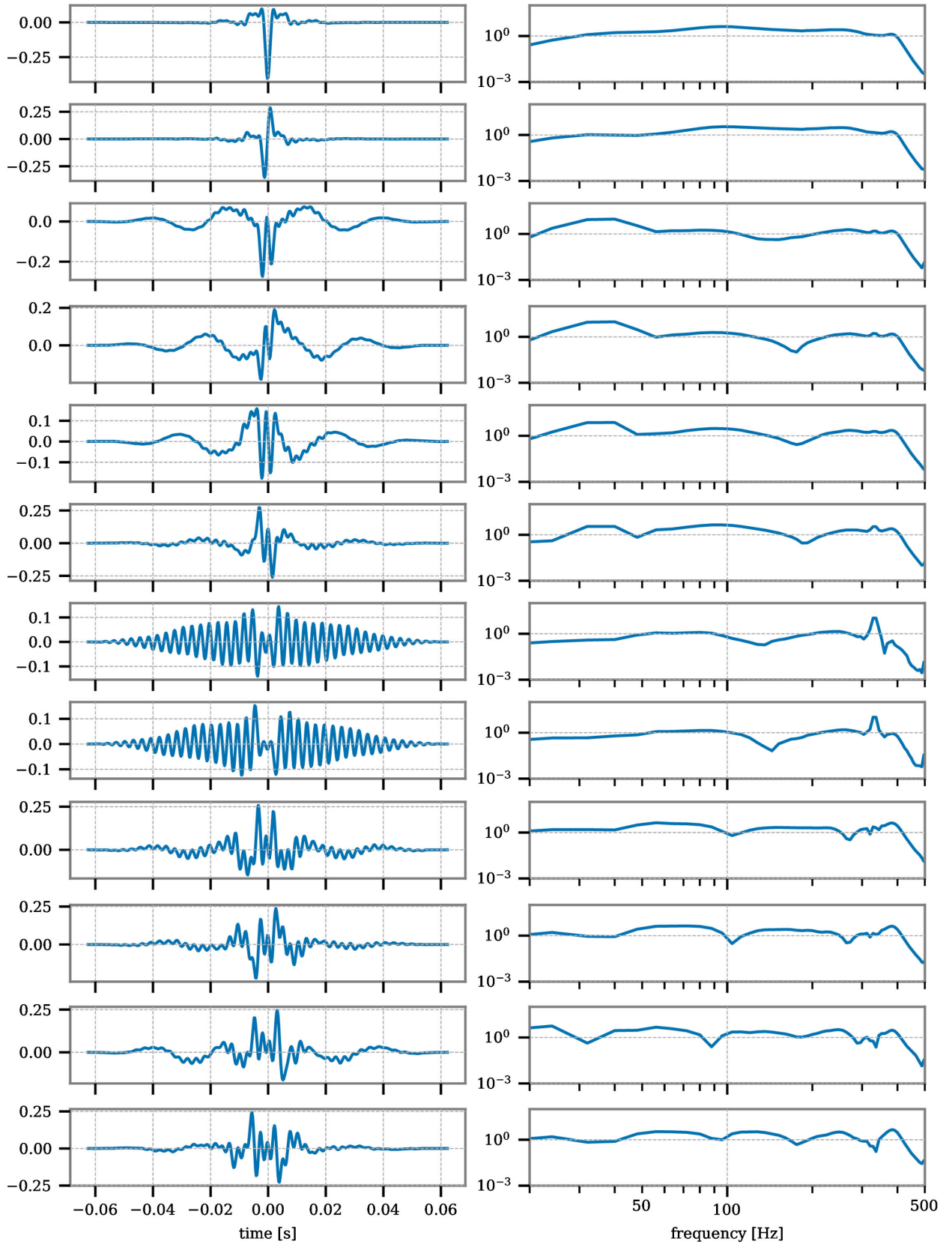


FIG. 1. The 12 components of the V^T matrix used to model the glitch, in descending order of significance, in the time domain (left) and their amplitude spectral density (right).

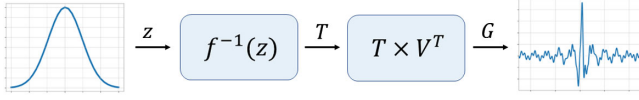


FIG. 2. Illustration of glitch reconstruction. Latent parameters z from a Gaussian distribution are passed through the inverse (trained) flow f^{-1} , and the obtained glitch amplitudes T are then combined with the basis matrix V^T to construct a glitch G . The glitch pictogram on the right is a representation of a typical Blip glitch in the time domain.

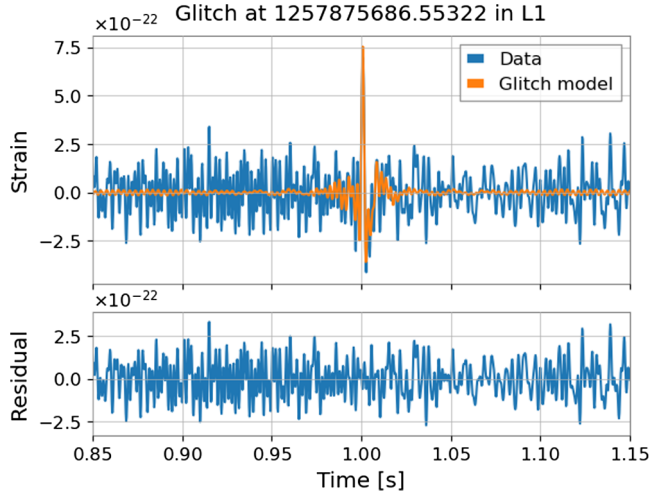


FIG. 3. Upper: plot of an O3 Blip glitch (GPS time given in title) in the LIGO-Livingston (L1) data. The filtered strain data are represented in blue, and the maximum-likelihood fitted glitch model is in orange. Lower: the residual after subtraction of the glitch, showing no visible trace of the glitch remaining.

no Gaussian data are identified as containing a glitch, and vice versa.

1. With known glitches

To test the glitch model for real glitches, we select several known glitch times randomly from the Gravity Spy

dataset and run the analysis on the strain data from the relevant LIGO detector around these times. This was done both for glitches present in the training data (O1) and for unseen glitches from the O2, O3a, and O3b runs.

Figure 3 shows an example of a glitch occurring in the LIGO-Livingston detector during the third observing run, O3. The plot shows the filtered strain data, overlaid with the maximum-likelihood glitch model fitted to the data by Bilby (the full result also contains uncertainty on the glitch parameters). The residual is plotted underneath and used as a first reference to show that the glitch has been successfully removed. Furthermore, we can also use time-frequency-energy spectrogram plots (so-called Q scans) to verify if a glitch has been removed fully. In Fig. 4, the Q scan of the same example glitch from Fig. 3 is shown, before and after glitch removal. The Blip glitch is clearly visible in the center of the left-hand plot, while no visible trace of it is left after subtraction in the plot on the right-hand side. While not quantified, this demonstrates the successful removal of all of the glitch power in a view familiar to GW analysts.

To quantitatively determine if the glitch model is favored or disfavored for given data, we investigate the log Bayes factors for the glitch model. To summarize the results, we provide a histogram of all log Bayes factors obtained from fitting the glitch model to glitches in O1 and O3 data, respectively, in Figs. 5 and 6. Almost all of the log Bayes factors are greater than zero, demonstrating that the glitch model is favored over a Gaussian noise model, thus showing that there is a glitch in the data, well described by the glitch model. There is one (of 786) glitch for O3 where the log Bayes factor is less than zero (-0.22), thus indicating that a Gaussian noise model is slightly preferred over the glitch model for this glitch. Investigating this further, we note that the outlier glitch appears louder and broader in the time-frequency spectrogram than most Blip glitches, and it is possible that this is the reason for the glitch model not being preferred, as it looks different from the Blip glitches the model is trained on (see Sec. IV E for further discussion).

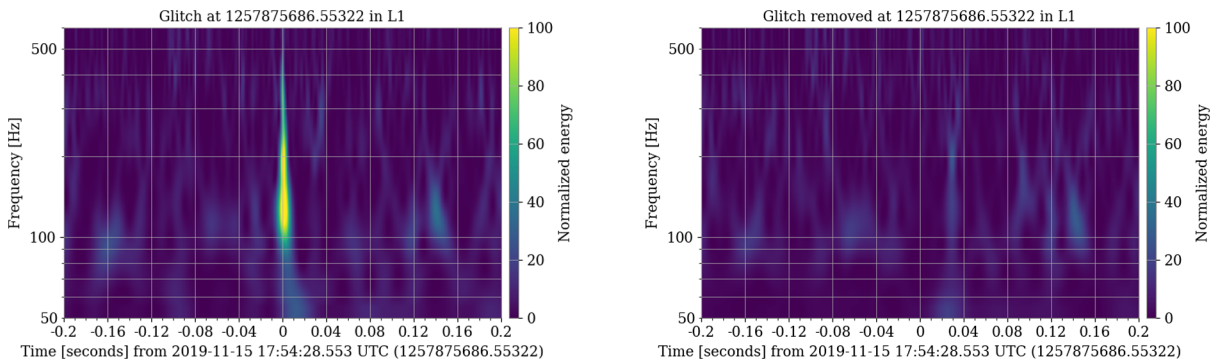


FIG. 4. Q-scan plots of the O3 Blip glitch shown in Fig. 3. The plot on the left shows the data pre-glitch-removal, while in the plot on the right, the glitch model has been applied to remove the glitch, leaving no visible trace.

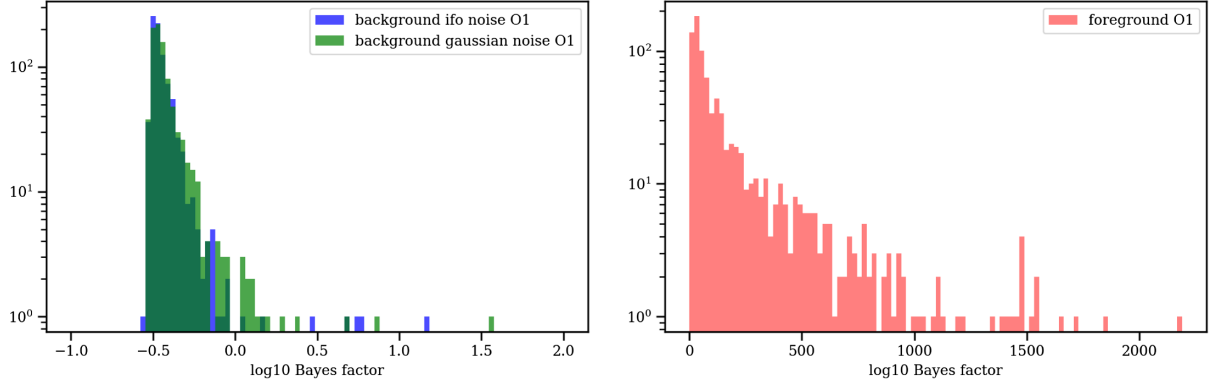


FIG. 5. Background: histogram of log Bayes factors obtained from applying the glitch model to glitch-free O1 interferometer data and Gaussian noise data. The log Bayes factors are mostly below zero, except for 7 of 860 runs on detector noise and 15 of 881 runs on Gaussian noise. Foreground: histogram of log Bayes factors from applying the glitch model to glitches in the O1 strain data. The log Bayes factors are all above zero.

2. On Gaussian noise

To check that the glitch model is disfavored when applied to data not containing a glitch, we apply it to Gaussian noise generated according to a known PSD, as well as to LIGO strain data at times when no glitch (or signal) is present (according to Gravity Spy, using an Omicron SNR threshold of 7.5 [23]). The test runs returned log Bayes factors of values less than zero for the majority of cases. The mostly negative log Bayes factors demonstrate that the glitch model is disfavored as expected when no glitch is present and hence will not give false alarms. The resulting background Bayes factor histograms are shown in Figs. 5 and 6. Comparing the background histograms for O1 and O3, we note that the distributions look very similar between runs on detector noise and between the Gaussian noise and the detector data. We find that 7 of 860 runs on detector noise and 15 of 881 runs on Gaussian noise have log Bayes factors greater than zero for O1, while there are 7 outliers each for O3 (out of 849 runs each on detector and Gaussian

noise). From a brief investigation of the outliers, we note that, visually, there appears to be a small amount of excess power in the strain data plots. It is worth highlighting that almost all of the outliers have log Bayes factors smaller than 0.5, indicating that the evidence in favor of the glitch model is “not worth more than a bare mention” according to the commonly used table by Kass and Raftery [61]. Furthermore, none of the outliers have log Bayes factors above 2, which by the above classification indicates the evidence is not decisive. The good separation between our background and foreground Bayes factor distributions indicates that naive Bayesian model selection with equal prior odds for each hypothesis would result in a posterior odds ratio that serves as a reasonable detection statistic. A more sophisticated analysis could take into account the rate of glitches (and later signals) or target an expected false alarm probability by adjusting the prior probability for each hypothesis considered, but we find this unnecessary in the present work [62].

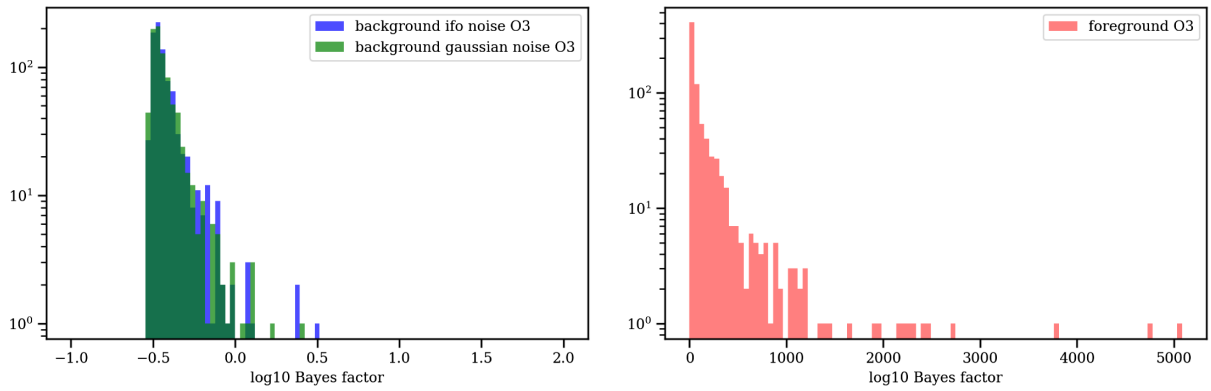


FIG. 6. Background: histogram of Bayes factors obtained from applying the glitch model to glitch-free O3 interferometer data and PSD noise data. The Bayes factors are mostly below zero, implying that the glitch model is disfavored, except for 7 of 849 runs each on detector and Gaussian noise. Foreground: histogram of log Bayes factors from applying the glitch model to glitches in the O3 strain data. The log Bayes factors are all above zero, except for one of the 786 glitches (with a log Bayes factor value -0.22).

Overall, these results indicate that the model is able to cleanly separate the populations of glitches and nonglitchy data with little overlap in the Bayes factor statistic.

B. Signal only

Next, we test the glitch model in the presence of an injected signal of a binary black hole merger. It is important to first confirm that the glitch model does not affect the results and that the glitch model is not favored over the signal model if no glitch is present in the data.

To show that the glitch model is not favored, we inject a signal into noisy data from the interferometers at times without glitches and apply both the glitch and signal models. Figure 7 shows a ternary plot of the log Bayes factors for the three cases: glitch and signal model, glitch model only, and signal model only. The dashed lines indicate the boundaries between preferred models. We repeat the experiment for detector noise from observing runs O1 and O3. In both cases, we find that the model containing the signal only is preferred, and we have thus shown that the glitch model is not preferred when there is no glitch in the data. We find that in all cases the signal-only model is quite strongly preferred over the glitch-only model, and that it is also preferred less strongly over the glitch + signal model. As expected, the glitch + signal model is only slightly disfavored, as this glitch model can contain low amplitude glitches consistent with the data, and the lower Bayes factor comes from the Occam factor penalizing the model for unnecessary complexity.

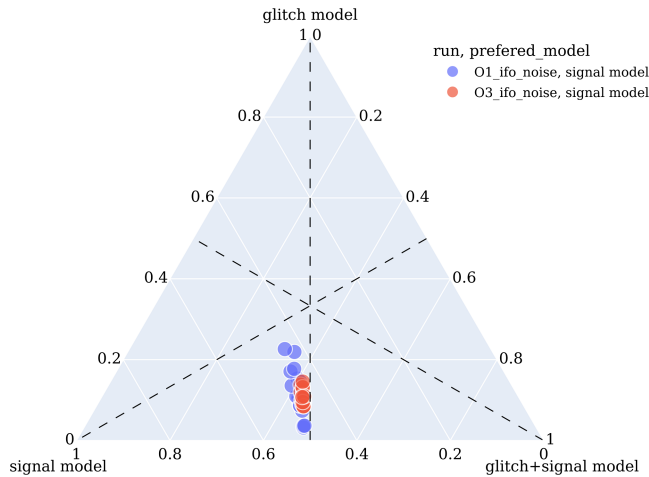


FIG. 7. The ternary plot shows the model comparison results, as calculated on detector data with an injected signal and no glitches, when applying both the glitch and signal model, and when applying only the glitch and only the signal model, respectively. The dashed lines represent the boundaries between preferred models. Events are colored by the observing run the data originate from, and the shapes are determined by which model is preferred by the log Bayes factor for each test glitch.

C. Injection tests

To test our model in the presence of a signal and a glitch, we now inject GW signals into glitch-contaminated data. By comparing the injected values to the obtained posteriors, we can determine if the results are improved by applying our glitch model to remove the glitch prior to the parameter evaluation. For the comparison, 6 of the 15 parameters typically used to describe a binary black hole merger were chosen as the parameters of interest: mass ratio q , chirp mass $\mathcal{M}$, inclination θ_{JN} , luminosity distance, right ascension, and declination. We chose to only present these parameters for clarity and because these are the most interesting and important for how they affect the signal model.

In Fig. 8 we show an example of the glitch model being fitted to strain data containing both a glitch and an injected signal. The waveform of a source at a luminosity distance of 750 Mpc was injected with the merger happening at the time of the glitch, thus overlapping the signal waveform with the glitch. The plot clearly demonstrates how, once the glitch is removed from the data, the signal becomes clearer. The waveform of the injected signal is plotted together with the residual to highlight this further.

We can again make use of the Bayes factors to compare models and determine how well the glitch model is doing. Running on data containing both a glitch and an injected signal, we compare the case where we have both the glitch and signal model to when there is only the glitch and only the signal model. We expect larger Bayes factors for the model preferred by the data. In Fig. 9, the log Bayes factors for these three cases are plotted against each other in a ternary plot. The figure contains example glitches from all three observing runs. We observe that the case where both models are applied is preferred for the majority of our test

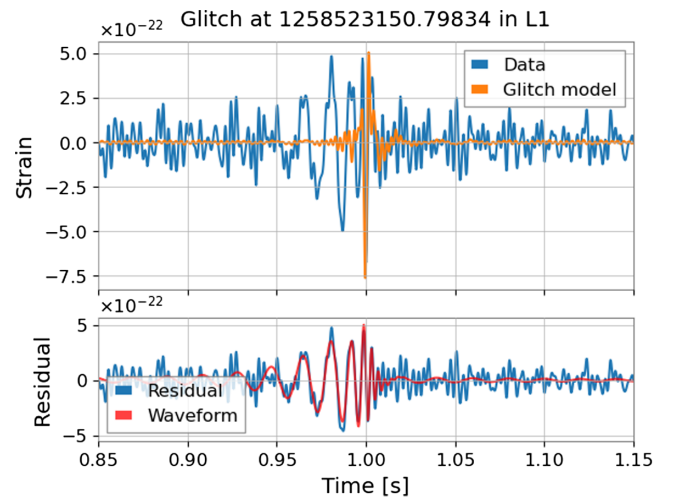


FIG. 8. Plot of data containing a glitch during an injected binary black hole signal, together with the fitted glitch model. The residual in the lower panel is overlaid by the waveform model for the injected signal.

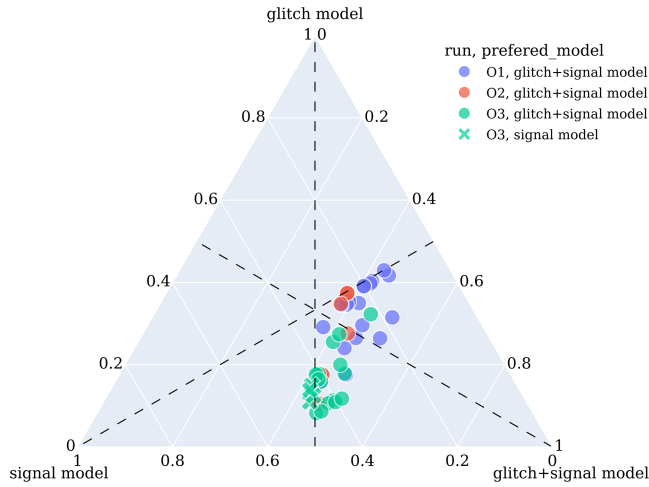


FIG. 9. The ternary plot shows the log Bayes factors (as calculated on data containing both an injected signal and a glitch) when applying both the glitch and signal model, and when applying only the glitch and only the signal model, respectively. The dashed lines represent the boundaries between preferred models. Events are colored by the observing run the glitchy data originate from, and the shapes are determined by which model is preferred by the log Bayes factor for each test glitch.

glitches. However, in 4 of the 25 O3 test glitches, the signal-only model is slightly favored.

A possible explanation could be that the background is not well measured. For example, if the noise is nonstationary around the time of the glitch, the detector output varies throughout the duration of the data analyzed, affecting the results. Alternatively, this could be explained by these glitches all being very weak and having a slightly less distinct shape than typical for Blip glitches. Thus, applying the glitch model does not improve the signal analysis sufficiently for this model to be favored. Both the Q scans and residual plots only show a minimal difference before and after the glitch has been removed. Looking at the filtered time data, these glitches appear to have an amplitude comparable to the injected signal and could, by eye, be mistaken for part of the signal. The Q scans before glitch removal also do not appear to have a visible glitch in the signal, and the power only shifts minimally when the glitch is removed. From these observations, it is comprehensible why the signal-only model is slightly favored over the glitch-and-signal model case. However, investigating these glitches further, we find that the signal posteriors still appear to be improved when the glitch model is applied. Studying the bias for these four glitches, we note that, for all parameters, the bias is comparable between the glitch-contaminated and glitch-removed cases. Hence, removing the glitch does not reduce the bias much, if at all, for these cases.

Furthermore, we briefly investigate the effect of the glitch being placed before, during, or after the signal merger. So far, all experiments have considered a glitch

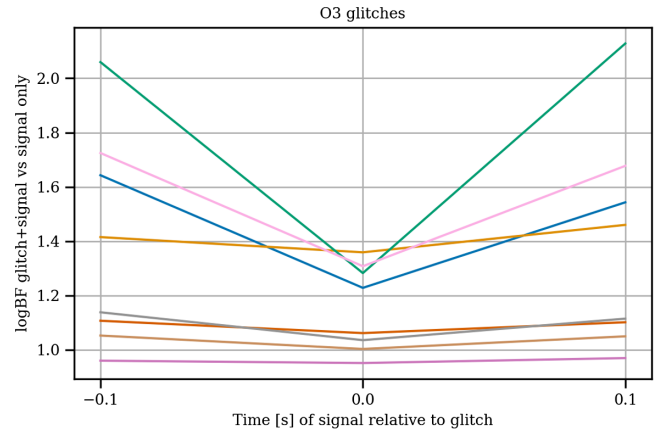


FIG. 10. The plot shows how the Bayesian model preference changes depending on if the signal is injected before, during or after the glitch. The x axis represents the three relative times when the signal is injected, and the y axis shows the ratio of log Bayes factors for the glitch + signal model versus the signal model only. Each line in the plot represents a specific test glitch. The uncertainty on the log Bayes factors is around 0.1 for all test glitches and thus is not shown in the plot.

overlapping with the merger. We now perform the model selection experiment, as above, for signals injected 0.1 s before and after the glitch. Running on data containing both a glitch and an injected signal, we compare the case where we have the glitch and signal model to when there is only the signal model. In Fig. 10, the log Bayes factors between these two cases are plotted on the y axis for a few of the test glitches from the third observing runs. The x axis shows the three signal injection times considered: before, during, and after the glitch.

We observe that the log Bayes factors vary for the same glitch, depending on where the signal is injected in relation to the glitch. The log Bayes factor has similar values when the signal is injected before and after the glitch, indicating that the model preference is almost the same between these two cases. Meanwhile, when the signal is injected during the glitch, compared to the before and after cases, the glitch + signal model is somewhat less favored for most glitches. This is because it is easier for the signal model to absorb some of the glitch power if the signal is injected during the glitch, thus reducing the need for the glitch model. Because of the signal model being time constrained, it is difficult for additional glitch power to be absorbed when the glitch and signal do not overlap.

Furthermore, all three cases for each glitch are either above or below 1, which is the limit separating which model is preferred. Thus, we find the four glitches for which the signal-only model is favored in Fig. 9 have values below 1 for all three cases of the signal being injected before, during, and after the glitch (only one of these examples is shown in Fig. 10). For no other test glitches, of any lag, is the signal-only model preferred over

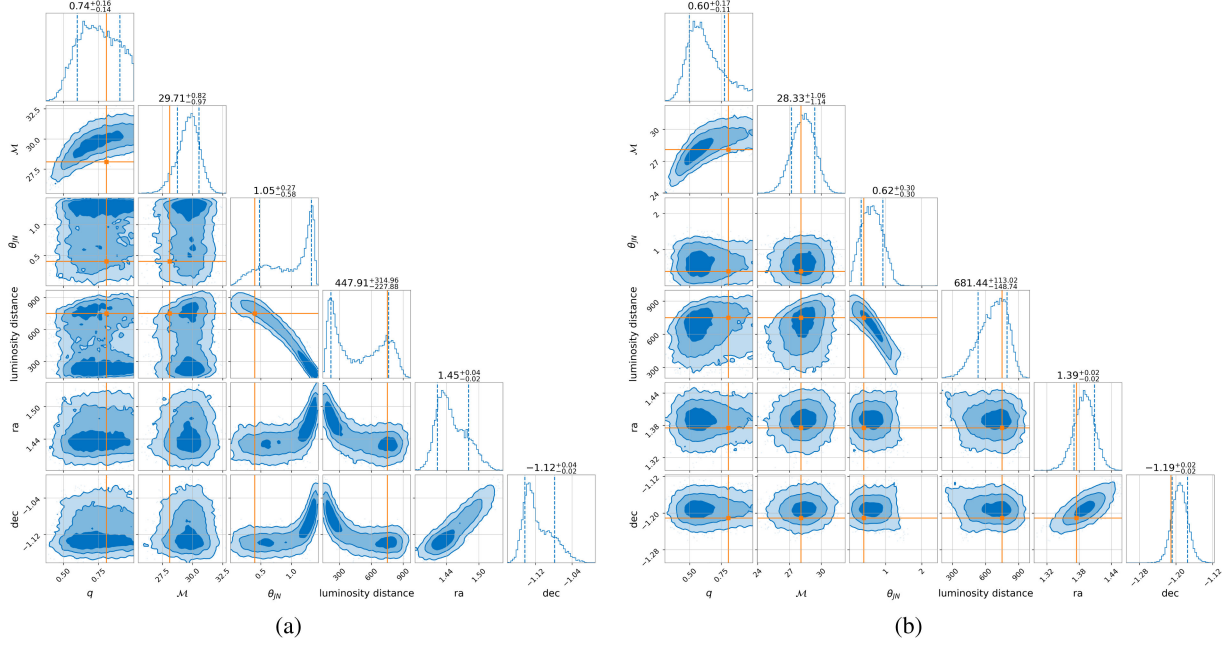


FIG. 11. Corner plots of signal posteriors, inferred (a) before and (b) after the glitch was removed from the data, respectively. For clarity, only mass ratio q , chirp mass $\mathcal{M}$, inclination θ_{JN} , luminosity distance, right ascension, and declination are shown. The blue contours show the 16% and 84% percentiles of the posterior distributions, and the crosshairs in each subplot indicate the injected values. This is the posterior for the same glitch also shown in Fig. 8. (a) Posterior for glitch-contaminated signal. (b) Posterior for glitch-removed signal.

the glitch + signal model. Hence, the quietness and properties of the glitch itself must be the reason for the glitch + signal model not being preferred, rather than how the glitch affects the signal.

D. Bias tests

As a bias test, we inject a signal into glitch-contaminated data and analyze the signal before and after removing the glitch. We can then compare the inference of the posterior values of the injected signal.

To quantify the bias, we compute the “standard accuracy,” which is defined for parameter x as

$$\Sigma_x = \frac{|x_{\text{maxL}} - x_{\text{true}}|}{\sigma_x}, \quad (4)$$

where x_{maxL} is the maximum-likelihood posterior value of x , x_{true} is the injected value, and σ_x is the standard deviation of the posterior of x . For an unbiased estimate of the posterior, we would expect the standard accuracy to be around 1 due to statistical fluctuations only.

1. On Blip glitches

The two corner plots in Fig. 11 show the signal posteriors as inferred from data with or without the glitch being removed, respectively. The corner plots in Fig. 11 demonstrate how the inference of the signal posterior improves

significantly if the glitch is removed from the data prior to analysis.

The mean of the standard accuracy of a few parameters for 25 O1 test glitches are shown in Fig. 12 and for O3 in Fig. 13. The plots show a lower mean standard accuracy for the case where the glitch has been removed from the data for all parameters. We can thus conclude that applying the glitch model improves the signal parameter estimation by reducing the bias. It is also interesting to note that the mass parameters seem to be less affected by the glitch

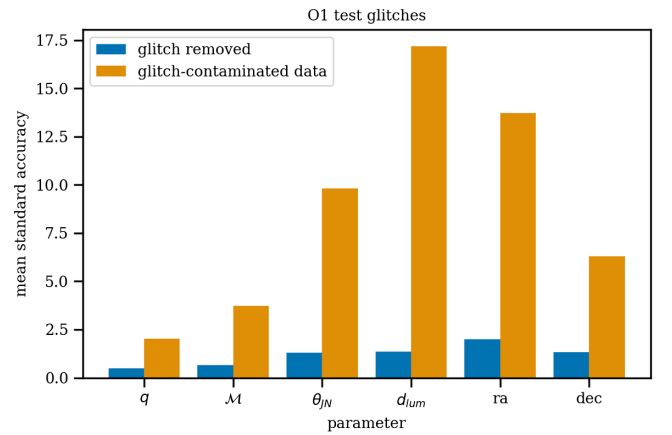


FIG. 12. Mean standard accuracy over 25 O1 test glitches, for a few parameters: mass ratio q , chirp mass $\mathcal{M}$, inclination θ_{JN} , luminosity distance d_{lum} , right ascension, and declination.

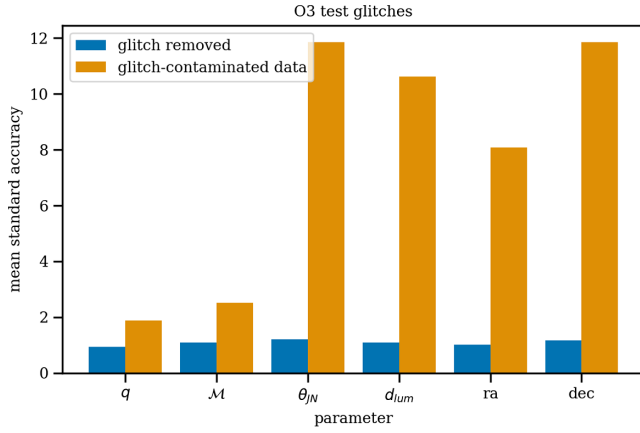


FIG. 13. Mean standard accuracy over 25 O3 test glitches, for a few parameters: mass ratio q , chirp mass $\mathcal{M}$, inclination θ_{JN} , luminosity distance d_{lum} , right ascension, and declination.

contamination than the other parameters shown in the plots. This is likely because the presence of a glitch in only one detector cannot create as large a deviation in the phase as it can in the relative amplitude between detectors that more strongly informs the extrinsic parameters.

In the above analysis, the signal is injected during the glitch. We also investigate how the bias standard accuracy changes if the signal is injected before or after the glitch. We find that the mean bias standard accuracy for the glitch-contaminated data is significantly higher (for all parameters except the mass ratio q) when the signal is injected during the glitch compared to before or after. There is little to no difference between the latter two cases. For the glitch-removed data, the bias standard accuracy is around 1 for all parameters, regardless of where in time the signal is injected. This shows that, although the bias is higher before glitch removal when the signal coincides with the glitch, it is still successfully reduced when considering the glitch-removed case.

The results are similar for both the O1 and O3 test glitches, although the variations between injections before or after the glitch are slightly larger in O1, with the signal being injected before the glitch having slightly larger biases for most parameters.

E. On Tomte glitches

To explore how our glitch model, trained on Blip glitches, behaves when applied to a different type of glitch, we perform the same tests for Tomte glitches. Tomte glitches have a similar morphology to Blip glitches and are hence interesting candidates for this experiment.

First, we plot a foreground histogram of log Bayes factors for 69 Tomte glitches found in O3 data, see Fig. 14. Similar to the case with Blip glitches in the data, we find that the glitch model is preferred over the noise model for all except 15 of the 69 glitches. This demonstrates that, although the model is trained on Blip glitches, it can also

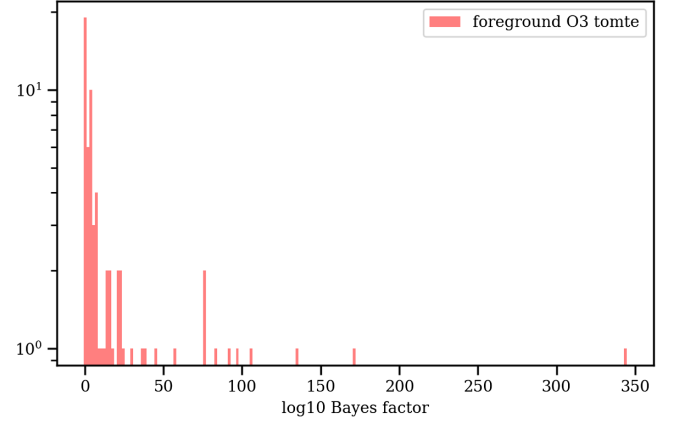


FIG. 14. Foreground histogram of log Bayes factors from applying the glitch model (trained on Blip glitches) to Tomte glitches in the O3 strain data. The log Bayes factors are mostly above zero, but 15 of the 69 test glitches have values below zero.

successfully remove Tomte glitches from the data in certain cases, but the success rate is significantly lower than for removing Blip glitches.

Second, we inject signals in data contaminated by Tomte glitches and compare the analyses when the glitch model is applied and not. The resulting log Bayes factor scatter plot is shown in Fig. 15. We note that almost all points are below 1, indicating that including the (Blip-trained) glitch model in the analysis is not preferred. However, there are 3 (of 20) points above 1, likely due to these Tomte glitches being particularly similar to Blip glitches. Some of the glitches are very close to 1, showing that the signal-only model is only slightly preferred by the data.

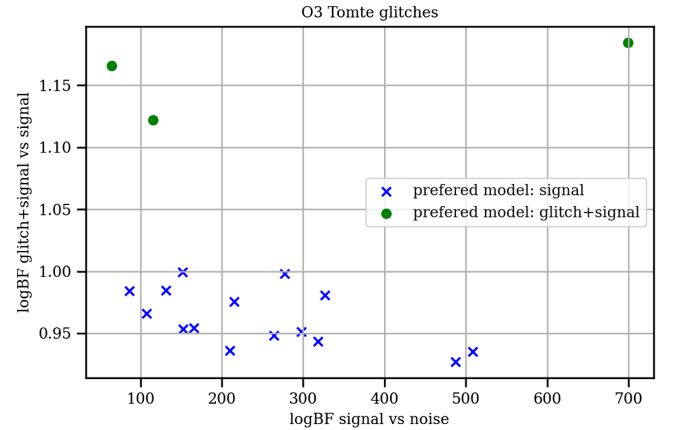


FIG. 15. The plot shows the log Bayes factors when applying the (Blip-trained) glitch and signal model versus the signal model only to O3 data containing Tomte glitches and signal injections. The x axis shows the signal-only model versus noise log Bayes factors, and the y axis shows the ratio between the log Bayes factors of the glitch + signal model and the signal-only model. The uncertainty on the log Bayes factors is around 0.1 for all test glitches and thus is not shown in the plot.

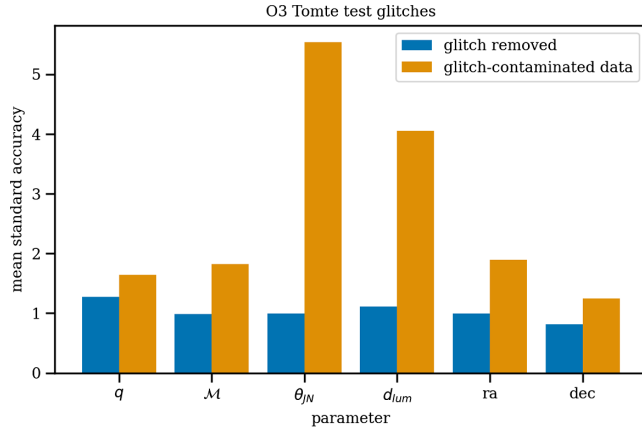


FIG. 16. Mean standard accuracy for 20 O3 Tomte test glitches, for signal parameters mass ratio q , chirp mass $\mathcal{M}$, inclination θ_{JN} , luminosity distance d_{lum} , right ascension, and declination.

Last, we repeat the standard accuracy calculations for the signals injected into data containing Tomte glitches. The results are shown in Fig. 16. The plot shows that all parameters have a lower mean standard accuracy when the glitch has been removed, even though the glitch model was trained on a different type of glitch than what is being removed. We thus show that removing the glitch, even with an incorrect model, reduces the bias in the signal. Remembering Figs. 12 and 13, however, the bias was reduced significantly more between the glitch-contaminated and glitch-removed cases for the Blip glitches than for the Tomte glitches in Fig. 16. The standard accuracy for the glitch-removed cases is, on average, around 1 for both Blip and Tomte glitches. However, the bias goes up to an average of 5 for Tomte and up to about 17/15 for Blip glitches for the glitch-contaminated cases.

V. DISCUSSION AND CONCLUSIONS

In this work we have shown that modeling Blip glitches using normalizing flows can successfully remove glitches from the data, thus reducing bias in the GW signal parameter estimation. We have shown that our glitch model successfully removes the glitches we have applied it to, without significantly affecting signals also present in the data.

First, we explored the case of glitch removal only and found that our glitch model reliably removes all test glitches. Applying the model to a variety of Blip glitches in both O1 and O3 data, we obtained 0% and 0.1% false dismissal rates, respectively. Similarly, applying the glitch model to nonglitchy detector noise and Gaussian noise data, we find false alarm rates of 0.8% and 1.7%, respectively, for O1 and 0.8% for both backgrounds in O3. Thus, we have shown that our model can cleanly separate the populations of Gaussian noise and glitches with little overlap. Second, we injected signals onto real glitches and tested the performance of our glitch model for

joint signal and glitch inference. We found that the joint signal + glitch model was favored for the majority of our test glitches when applying Bayesian model selection to data containing a glitch and a signal. Meanwhile, applying the glitch model to data containing a signal but no glitch, we found that the signal-only model is preferred over the glitch + signal model for all test glitches. Finally, we studied the bias in the signal posterior parameters induced by the presence of a glitch. We found that removing the glitch with our model significantly reduces the bias and improves the parameter estimation of the signal.

To train the model, we make use of the Blip glitches identified in Gravity Spy’s O1 training data to train our glitch model. Hence, the model is only knowledgeable about these types of glitches as they appeared during O1. If glitch models of this kind are to be applied in future observing runs, and to other glitches, it might be relevant to train models on more recent data as the detectors are continuously improved. Some glitch classes change, and new ones might appear between detector upgrades [23,35,63]. For optimal performance, the glitches in the training data should be as similar as possible to the glitches that the model is then applied to.

The computational cost of the analysis increases when including the glitch model. We run the code on four CPU cores using the Nessai sampler. On average, the sampling time for fitting only the glitch model to data containing a glitch is approximately 1 h. Fitting the glitch-only, signal-only, and glitch + signal models to data containing both a glitch and a signal, we get median sampling wall times (over our 25 test cases) of 6, 0.5, and 30 h, respectively. There is scope for further optimization of our likelihood function and normalizing flow prior within Bilby, which we will investigate in the future. However, the run time is not prohibitive for an off-line analysis. Furthermore, training a NF to create the glitch model only takes a few minutes.

This paper has focused on exploring the performance of a model trained on Blip glitches, but the methodology is applicable to any glitch class or, indeed, any temporally isolated feature in the data. A similar approach may, for example, be useful for signal model discovery where no clean mapping to physical parameters is known, such as, e.g., burst signals from supernovae. We briefly investigate how our method performs for other, more complex, glitch types; we repeat the training of the model as before, now for Koi Fish and Power Line glitches, and then run the glitch-only analysis as before. We find that the glitch model removes the glitches as expected and returns log Bayes factors favoring the glitch model. We do not repeat all the extensive tests for these glitch classes, but leave it to future work to explore these and other classes fully. However, our brief investigation still shows the potential of our method being applicable to other glitch classes as well.

From the studies on removing Tomte glitches with a Blip-trained model in Sec. IV E, we observe that, although

the model removes the glitch, the model selection results are biased. The foreground analysis on Tomte glitches shows a 22% false alarm rate. Injecting signals onto Tomte glitches, we find that the joint glitch + signal model is only preferred over the signal-only model in 15% of our test glitches. However, our study on Tomte glitches also shows that using the Blip-trained model improves the signal parameter estimation and reduces the bias in the posteriors. We thus show that, although the model selection tests indicate that the Blip model is not a good fit for Tomte glitches, it still improves the signal analysis. This is probably only true due to the similarity in morphology between Blip and Tomte glitches. Further investigations would be needed to draw more conclusions regarding the limits of the glitch model. However, it is straightforward and not computationally expensive to train a new glitch model, and thus, training one model for each glitch class would be the best approach. We thus draw the conclusion that, although a model trained on a different glitch class can be of use, it is better to use a glitch model trained on the same glitch class it will be applied to. If, however, a new type of glitch appears, or there are too few glitches of a certain type to train a new glitch model, making use of a glitch model trained on glitches with similar features would be the next best option.

Another future analysis is to apply the glitch model to real GW events contaminated by glitches. This work has focused on proof of concept, and we use injected signals only to demonstrate robustness, since we know their true parameters. The next step would be to analyze real events that are known to be problematic due to glitches and investigate if our model can compare to (and possibly improve) state-of-the-art results. To perform this analysis fully, different glitch models would need to be trained for each class of glitches we want to remove. Some signals of particular interest for glitch subtraction include, for example, GW191109 [20] and GW200129 [22].

In conclusion, we have demonstrated the potential of a normalizing-flow-trained glitch model to improve gravitational wave signal analysis and reduce bias through joint inference of signal and glitch. We have also discussed the developments needed to scale our analysis to real events, and future work would include analysis of glitch-contaminated events in the gravitational wave catalog.

The code produced and used for this work, including the trained glitch model, is available from [64].

ACKNOWLEDGMENTS

We are grateful for feedback on this work from members of the LVK Collaboration and, in particular, Christopher Berry, Laura Nuttall, Cailin Plunkett, and Rhiannon Udall. We are also grateful for useful discussions and technical advice from Michael Williams and Daniel Williams. J. V. was supported by STFC Grant No. ST/V005634/1. We are grateful for the computational resources provided by Cardiff University and funded by STFC awards supporting UK Involvement in the Operation of Advanced LIGO. This research has made use of data or software obtained from the Gravitational Wave Open Science Center [65], a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN), and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

DATA AVAILABILITY

The data and code that support the findings of this article are openly available [53,57,64], embargo periods may apply.

[1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, and O. D. Aguiar (LIGO Scientific and Virgo Collaborations), Observation of gravitational

waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).

[2] R. Abbott *et al.* (KAGRA, Virgo, and LIGO Scientific Collaborations), 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).
- [3] J. Aasi, B. Abbott, R. Abbott, T. Abbott, M. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari *et al.*, Advanced LIGO, *Classical Quantum Gravity* **32**, 074001 (2015).
- [4] F. a. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin *et al.*, Advanced Virgo: A second-generation interferometric gravitational wave detector, *Classical Quantum Gravity* **32**, 024001 (2014).
- [5] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, H. Yamamoto (KAGRA Collaboration), Interferometer design of the KAGRA gravitational wave detector, *Phys. Rev. D* **88**, 043007 (2013).
- [6] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith *et al.* (LIGO Scientific and Virgo Collaborations), A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, *Classical Quantum Gravity* **37**, 055002 (2020).
- [7] D. Davis and M. Walker, Detector characterization and mitigation of noise in ground-based gravitational-wave interferometers, *Galaxies* **10**, 12 (2022).
- [8] R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. Adhikari, V. Adya, C. Affeldt, D. Agarwal *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), 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).
- [9] F. Robinet, Omicron: An algorithm to detect and characterize transient events in gravitational-wave detectors, Technical Report No. VIR-0545B-14, 2016.
- [10] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, Omicron: A tool to characterize transient noise in gravitational-wave detectors, *SoftwareX* **12**, 100620 (2020).
- [11] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, T. K. Lee, C. Lintott, T. B. Littenberg, A. Lundgren, C. Østerlund, J. R. Smith, L. Trouille, and V. Kalogera, Gravity Spy: Integrating advanced LIGO detector characterization, machine learning, and citizen science, *Classical Quantum Gravity* **34**, 064003 (2017).
- [12] M. Zevin, C. B. Jackson, Z. Doctor, Y. Wu, C. Østerlund, L. C. Johnson, C. P. Berry, K. Crowston, S. B. Coughlin, V. Kalogera *et al.*, Gravity Spy: Lessons learned and a path forward, *Eur. Phys. J. Plus* **139**, 100 (2024).
- [13] M. Razzano, F. Di Renzo, F. Fidicaro, G. Hemming, and S. Katsanevas, GWitchHunters: Machine learning and citizen science to improve the performance of gravitational wave detector, *Nucl. Instrum. Methods Phys. Res., Sect. A* **1048**, 167959 (2023).
- [14] 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 Quantum Gravity* **39**, 245013 (2022).
- [15] S. Ghonge, J. Brandt, J. M. Sullivan, M. Millhouse, K. Chatziioannou, J. A. Clark, T. Littenberg, N. Cornish, S. Hourihane, and L. Cadonati, Assessing and mitigating the impact of glitches on gravitational-wave parameter estimation: A model agnostic approach, *Phys. Rev. D* **110**, 122002 (2024).
- [16] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), 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).
- [17] R. Abbott *et al.* (LIGO Scientific and VIRGO Collaborations), GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, *Phys. Rev. D* **109**, 022001 (2024).
- [18] B. P. Abbott, R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, V. B. Adya *et al.*, GW170817: Observation of gravitational waves from a binary neutron star inspiral, *Phys. Rev. Lett.* **119**, 161101 (2017).
- [19] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Properties of the binary neutron star merger GW170817, *Phys. Rev. X* **9**, 011001 (2019).
- [20] R. Udall, S. Hourihane, S. Miller, D. Davis, K. Chatziioannou, M. Isi, and H. Deshong, Antialigned spin of GW191109: Glitch mitigation and its implications, *Phys. Rev. D* **111**, 024046 (2025).
- [21] E. Payne, S. Hourihane, J. Golomb, R. Udall, R. Udall, D. Davis, and K. Chatziioannou, Curious case of GW200129: Interplay between spin-precession inference and data-quality issues, *Phys. Rev. D* **106**, 104017 (2022).
- [22] R. Macas, A. Lundgren, and G. Ashton, Revisiting the evidence for precession in GW200129 with machine learning noise mitigation, *Phys. Rev. D* **109**, 062006 (2024).
- [23] J. Glanzer, S. Banagiri, S. Coughlin, S. Soni, M. Zevin, C. P. L. Berry, O. Patane, S. Bahaadini, N. Rohani, K. Crowston *et al.*, Data quality up to the third observing run of advanced LIGO: Gravity Spy glitch classifications, *Classical Quantum Gravity* **40**, 065004 (2023).
- [24] N. J. Cornish and T. B. Littenberg, BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches, *Classical Quantum Gravity* **32**, 135012 (2015).
- [25] C. Plunkett, S. Hourihane, and K. Chatziioannou, Concurrent estimation of noise and compact-binary signal parameters in gravitational-wave data, *Phys. Rev. D* **106**, 104021 (2022).
- [26] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, *Phys. Rev. D* **91**, 084034 (2015).
- [27] K. Chatziioannou, N. Cornish, M. Wijngaarden, and T. B. Littenberg, Modeling compact binary signals and instrumental glitches in gravitational wave data, *Phys. Rev. D* **103**, 044013 (2021).
- [28] S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, and N. Cornish, Accurate modeling and mitigation of overlapping signals and glitches in

- gravitational-wave data, *Phys. Rev. D* **106**, 042006 (2022).
- [29] D. Davis, T. Massinger, A. Lundgren, J. C. Driggers, A. L. Urban, and L. Nuttall, Improving the sensitivity of advanced LIGO using noise subtraction, *Classical Quantum Gravity* **36**, 055011 (2019).
- [30] G. Ashton, Gaussian processes for glitch-robust gravitational-wave astronomy, *Mon. Not. R. Astron. Soc.* **520**, 2983 (2023).
- [31] J. Merritt, B. Farr, R. Hur, B. Edelman, and Z. Doctor, Transient glitch mitigation in advanced LIGO data, *Phys. Rev. D* **104**, 102004 (2021).
- [32] S. D. Mohanty and M. A. T. Chowdhury, Glitch subtraction from gravitational wave data using adaptive spline fitting, *Classical Quantum Gravity* **40**, 125001 (2023).
- [33] R. Udall and D. Davis, Bayesian modeling of scattered light in the LIGO interferometers, *Appl. Phys. Lett.* **122**, 094103 (2023).
- [34] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, J. R. Smith, V. Kalogera, and A. Katsaggelos, Machine learning for Gravity Spy: Glitch classification and dataset, *Inf. Sci. (N.Y.)* **444**, 172 (2018).
- [35] S. Soni *et al.*, Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning, *Classical Quantum Gravity* **38**, 195016 (2021).
- [36] Y. Wu, M. Zevin, C. P. Berry, K. Crowston, C. Østerlund, Z. Doctor, S. Banagiri, C. B. Jackson, V. Kalogera, and A. K. Katsaggelos, Advancing glitch classification in Gravity Spy: Multi-view fusion with attention-based machine learning for advanced LIGO's fourth observing run, [arXiv:2401.12913](https://arxiv.org/abs/2401.12913).
- [37] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis, Multiresolution techniques for the detection of gravitational-wave bursts, *Classical Quantum Gravity* **21**, S1809 (2004).
- [38] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepanczyk, and S. Vitale, Machine-learning non-stationary noise out of gravitational-wave detectors, *Phys. Rev. D* **101**, 042003 (2020).
- [39] H. Wang, Y. Zhou, Z. Cao, Z. Guo, and Z. Ren, WaveFormer: transformer-based denoising method for gravitational-wave data, *Mach. Learn.* **5**, 015046 (2024).
- [40] M. Lopez, V. Boudart, K. Buijsman, A. Reza, and S. Caudill, Simulating transient noise bursts in LIGO with generative adversarial networks, *Phys. Rev. D* **106**, 023027 (2022).
- [41] T. Dooney, R. L. Curier, D. S. Tan, M. Lopez, C. Van Den Broeck, and S. Bromuri, One flexible model for multiclass gravitational wave signal and glitch generation, *Phys. Rev. D* **110**, 022004 (2024).
- [42] Y. Li, Y. Wu, and A. K. Katsaggelos, Cross-temporal spectrogram autoencoder (CTSAE): Unsupervised dimensionality reduction for clustering gravitational wave glitches, [arXiv:2404.15552](https://arxiv.org/abs/2404.15552).
- [43] R. Bondarescu, A. Lundgren, and R. Macas, Quasiphsical model for removing short glitches from LIGO and Virgo data, *Phys. Rev. D* **108**, 122004 (2023).
- [44] T.-Y. Sun, C.-Y. Xiong, S.-J. Jin, Y.-X. Wang, J.-F. Zhang, and X. Zhang, Efficient parameter inference for gravitational wave signals in the presence of transient noises using temporal and time-spectral fusion normalizing flow, *Chin. Phys. C* **48**, 045108 (2024).
- [45] C.-Y. Xiong, T.-Y. Sun, J.-F. Zhang, and X. Zhang, Robust inference of gravitational wave source parameters in the presence of noise transients using normalizing flows, *Phys. Rev. D* **111**, 024019 (2025).
- [46] G. Ashton *et al.*, Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [47] I. M. Romero-Shaw, C. Talbot, S. Biscoveanu, V. D'emilio, G. Ashton, C. Berry, S. Coughlin, S. Galaudage, C. Hoy, M. Hübner *et al.*, Bayesian inference for compact binary coalescences with Bilby: Validation and application to the first LIGO–Virgo gravitational-wave transient catalogue, *Mon. Not. R. Astron. Soc.* **499**, 3295 (2020).
- [48] R. Legin, M. Isi, K. W. K. Wong, Y. Hezaveh, and L. Perreault-Levasseur, Gravitational-wave parameter estimation in non-Gaussian noise using score-based likelihood characterization, *Astrophys. J. Lett.* **985**, L46 (2025).
- [49] I. Kobayev, S. J. Prince, and M. A. Brubaker, Normalizing flows: An introduction and review of current methods, *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3964 (2020).
- [50] D. Jimenez Rezende and S. Mohamed, Variational inference with normalizing flows, [arXiv:1505.05770](https://arxiv.org/abs/1505.05770).
- [51] G. Papamakarios, E. T. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, *J. Mach. Learn. Res.* **22**, 1 (2021), <https://www.jmlr.org/papers/v22/19-1028.html>.
- [52] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models—Corrigendum, *Publ. Astron. Soc. Aust.* **37**, e036 (2020).
- [53] S. Coughlin, Gravity Spy training set, Zenodo, [10.5281/zenodo.1486046](https://zenodo.org/record/1486046) (2018).
- [54] M. Cabero, A. Lundgren, A. H. Nitz, T. Dent, D. Barker, E. Goetz, J. S. Kissel, L. K. Nuttall, P. Schale, R. Schofield *et al.*, Blip glitches in advanced LIGO data, *Classical Quantum Gravity* **36**, 155010 (2019).
- [55] R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, and C. L. S. C. Adams (Virgo Collaboration), Open data from the first and second observing runs of advanced LIGO and advanced Virgo, *SoftwareX* **13**, 100658 (2021).
- [56] D. M. Macleod, J. S. Areeda, S. B. Coughlin, T. J. Massinger, and A. L. Urban, GWpy: A Python package for gravitational-wave astrophysics, *SoftwareX* **13**, 100657 (2021).
- [57] J. Glanzer, S. Banagari, S. Coughlin, M. Zevin, S. Bahaadini, N. Rohani, S. Allen, C. Berry, K. Crowston, M. Harandi, C. Jackson, V. Kalogera, A. Katsaggelos, V. Noroozi, C. Osterlund, O. Patane, J. Smith, S. Soni, and L. Trouille, Gravity Spy machine learning classifications of LIGO glitches from observing Runs O1, O2, O3a, and O3b, Zenodo, [10.5281/zenodo.5649212](https://zenodo.org/record/5649212) (2021).
- [58] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, Neural spline flows, [arXiv:1906.04032](https://arxiv.org/abs/1906.04032).

- [59] M. J. Williams, jmcginn, federicostak, and J. Veitch, UOF-GRAVITY/GLASFLOW: V0.2.0, Zenodo, [10.5281/zenodo.7598678](https://zenodo.org/record/7598678) (2023).
- [60] M. J. Williams, J. Veitch, and C. Messenger, Nested sampling with normalizing flows for gravitational-wave inference, *Phys. Rev. D* **103**, 103006 (2021).
- [61] R. E. Kass and A. E. Raftery, Bayes factors, *J. Am. Stat. Assoc.* **90**, 773 (1995).
- [62] J. Veitch and A. Vecchio, Assigning confidence to inspiral gravitational wave candidates with Bayesian model selection, *Classical Quantum Gravity* **25**, 184010 (2008).
- [63] S. Soni, B. K. Berger, D. Davis, F. Di Renzo, A. Effler, T. Ferreira, J. Glanzer, E. Goetz, G. Gonzalez, A. Helmling-Cornell *et al.* (LIGO Collaboration), LIGO detector characterization in the first half of the fourth observing run, *Classical Quantum Gravity* **42**, 085016 (2025).
- [64] A.-K. Malz and J. Veitch, Code for paper: Joint inference for gravitational wave signals and glitches using a data-informed glitch model, Zenodo, [10.5281/zenodo.15316399](https://zenodo.org/record/15316399) (2025).
- [65] <https://gwosc.org>