

Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run

D Davis¹, T B Littenberg², I M Romero-Shaw^{3,4,5}, M Millhouse⁶, J McIver⁷, F Di Renzo^{8,9}, G Ashton¹⁰

¹LIGO, California Institute of Technology, Pasadena, CA 91125, USA

²NASA Marshall Space Flight Center, Huntsville, AL 35811, USA

³OzGrav, School of Physics & Astronomy, Monash University, Clayton 3800, Victoria, Australia

⁴Monash Astrophysics, School of Physics and Astronomy, Monash University, VIC 3800, Australia

⁵Department of Applied Mathematics and Theoretical Physics, Cambridge CB3 0WA, United Kingdom

⁶OzGrav, University of Melbourne, Parkville, Victoria 3010, Australia

⁷University of British Columbia, Vancouver, BC V6T 1Z4, Canada

⁸Università di Pisa, I-56127 Pisa, Italy

⁹INFN, Sezione di Pisa, I-56127 Pisa, Italy

¹⁰Department of Physics, Royal Holloway, University of London, TW20 0EX, United Kingdom

Abstract. Data from ground-based gravitational-wave detectors contains numerous short-duration instrumental artifacts, called “glitches.” The high rate of these artifacts in turn results in a significant fraction of gravitational-wave signals from compact binary coalescences overlapping glitches. In LIGO-Virgo’s third observing run, $\approx 20\%$ of gravitational-wave source candidates required some form of mitigation due to glitches. This was the first observing run where glitch subtraction was included as a part of LIGO-Virgo-KAGRA data analysis methods for a large fraction of detected gravitational-wave events. This work describes the methods to identify glitches, the decision process for deciding if mitigation was necessary, and the two algorithms, `BayesWave` and `gwssubtract`, that were used to model and subtract glitches. Through case studies of two events, GW190424_180648 and GW200129_065458, we evaluate the effectiveness of the glitch subtraction, compare the statistical uncertainties in the relevant glitch models, and identify potential limitations in these glitch subtraction methods. We finally outline the lessons learned from this first-of-its-kind effort for future observing runs.

1. Introduction

Data analysis of gravitational-wave signals from compact binary coalescences (CBCs) [1–4] requires precise understanding of the expected astrophysical waveforms as well as models of the noise in gravitational-wave detectors [5]. These analyses, and the identification of times containing gravitational waves, are limited by the amplitude of background noise, referred to as the sensitivity of the detectors. Numerous noise sources limit the sensitivity of ground-based gravitational-wave detectors [6–8], including fundamental noise sources that are due to the physical limitations of the design and materials of the detectors, and technical noise sources, that are due to the imperfections in the current detectors [9–11]. There are also short bursts of excess noise in the detectors, referred to as “glitches,” that further pollute the gravitational-wave strain data [10, 12, 13].

Typical Bayesian methods in gravitational-wave data analysis assume that the noise is stationary and Gaussian over the time period analyzed [5]. These assumptions are violated, however, when glitches are present in the analyzed data. When not accounted for as a part of the analyses, the presence of glitches can bias estimates of gravitational-wave properties [14–17]. There are some techniques proposed in which these assumptions have been relaxed [18–21], but these methods have not yet been used in an analysis by the LIGO-Virgo-KAGRA (LVK) collaboration.

Rather than being a rare edge case, it is quite common for glitches to be present near a gravitational-wave signal. The median rate of glitches in the LIGO and Virgo detectors during their third observing run (O3) [2, 3] were above 1 per minute for the majority of the run. At this rate, the probability a glitch to overlap a signal from a binary black hole is $\mathcal{O}(10\%)$ and it is almost certain that a glitch will overlap a signal from a binary neutron star. These pessimistic estimates have held true for recent catalogs. In total, 18 of the events detected in O3 were deemed to require glitch subtraction before further analysis [2–4], including all confidently detected gravitational-wave signals from binary neutron stars [22, 23] and neutron star – black hole binaries [24, 25]. Since typical methods used to estimate gravitational-wave source properties do not account for these glitches, it is therefore important to develop alternate mitigation methods to reduce any potential biases.

Three main categories of glitch mitigation have been used in analyses of CBCs: removing the entire stretch of data containing a glitch; modelling glitches using the strain data and subtracting this model; and using an additional time series that witnesses the source of the glitch to subtract the relevant excess noise. The first approach is commonly used in searches for CBCs [26, 27], but requires additional modifications to the analysis techniques to be used in parameter estimation without introducing biases [20, 28]. For this reason, only the latter two techniques have been used in analyses by the LVK.

A wide variety of similar methods have been developed to model and subtract glitches. Many techniques have been developed to model glitches using gravitational-wave strain data alone, including modeled Bayesian approaches [19, 28–32] and some based on machine learning [33, 34]. Likewise, techniques that utilize the large number of additional auxiliary sensors at each observatory to model sources of noise with analytic methods [35–44] and machine-learning approaches [45–49] have been proposed.

Of these tools, the only algorithms that have been used to model and subtract glitches for LVK analyses [1–4] are `BayesWave` [28–30] and `gwsubtract` [35, 40].

`BayesWave` models glitches based on only the gravitational-wave strain data using wavelets. It has been used to subtract glitches in LVK analyses since the second observing run (O2) [22, 28] and is the most commonly used algorithm to date. Conversely, `gwsubtract` was used to subtract glitches in an LVK analysis for the first time in O3 [3]. However, this method was previously used for broadband noise subtraction in O2 [40]. `gwsubtract` models glitches using information from an additional auxiliary witness that is correlated with, but independent of, the strain data.

This work focuses on describing the versions of these tools that were used in O3, the methods for determining if glitch subtraction was needed, and the potential limitations of current glitch subtraction methods. We first provide background on these two methods in Section 2, highlighting how the specific configuration of the tools used in O3 differs from previous or more recent versions of the tools. Section 2 also explains the methods used to evaluate the need for and results of glitch subtraction. We then focus on two events, GW190424_180648 and GW200129_065458 (which we will refer to as GW190424 and GW200129), in Section 3 and investigate the result of glitch subtraction in each case and estimate the additional statistical errors introduced from glitch modelling. Finally, in Section 4, we describe potential updates to these glitch mitigation methods for the next observing run and the expected challenges for glitch subtraction with a higher detection rate.

2. Methods

In the third observing run, two glitch subtraction methods were utilized as a pre-processing step before parameter estimation analyses were completed [2–4]. The first algorithm, `BayesWave` [28–30], was used for 15 events in O3 [2–4]. The second algorithm, `gwsubtract` [40], was used for only 1 event [3]. This section outlines the basic methods both algorithms use to model the noise contributions from a glitch. We also discuss the metrics we used to evaluate if glitch subtraction was needed and if the glitch subtraction was successful at removing the excess power from any identified glitches.

2.1. Glitch subtraction using wavelets

The `BayesWave` algorithm assumes that in each interferometer, the detector data $h(t)$ can be expressed as

$$h(t) = n(t) + s(t) + g(t) \quad (1)$$

where $n(t)$ is Gaussian noise, $s(t)$ is an astrophysical signal, and $g(t)$ is a non-Gaussian instrumental glitch. `BayesWave` models non-Gaussian features (both signal and glitch) as a sum of sine-Gaussian (also called Morlet-Gabor) wavelets. In the time domain, the wavelets are given by

$$\Psi(t; \vec{\theta}) = Ae^{-(t-t_0)^2/\tau^2} \cos(2\pi f_0(t-t_0) + \phi_0) \quad (2)$$

where $\vec{\theta} = \{t_0, f_0, Q, A, \phi_0\}$ are the parameters of each wavelet, and $\tau = Q/(2\pi f_0)$. The wavelet parameters, as well as the total number of wavelets used, is marginalised over via a transdimensional Markov chain Monte Carlo.

For the glitch model, the wavelet parameters in each detector are independent. For the signal model, there is one common set of wavelets which are projected onto

the detectors using a set of extrinsic parameters that describe the sky location and polarisation content. The power spectral density is modeled using [50], which also uses a transdimensional MCMC to model the noise as a combination of a cubic spline and Lorentzians. Full details can be found in [29] and [30].

The transient features in the data can be reconstructed using three different model assumptions:

- I. The data contain Gaussian noise with an astrophysical signal.
- II. The data contain Gaussian noise with an instrumental glitch.
- III. The data contain Gaussian noise with both a signal *and* a glitch.

To identify and remove glitches from detector data, we use either the glitch-only model (case II above), or signal-plus-glitch model (case III). The glitch-only approach is used when the glitch is sufficiently separated in time and/or frequency from the part of the signal that `BayesWave` can reconstruct. When using the glitch-only approach, we only need to consider data from the single detector in which the glitch appears. If the signal and glitch overlap significantly, we use the signal-plus-glitch model. The signal-plus-glitch model requires data from multiple detectors in order to separate the coherent signal power and the incoherent glitch power. By simultaneously fitting the for signals and glitches, we ensure that we do not subtract any signal power during glitch mitigation.

As `BayesWave` uses an MCMC to marginalise over the parameters, the end result is a posterior distribution of time series of the glitch, $g(t)$. To actually perform the glitch mitigation, we select a fair draw from the posterior [31] by randomly selecting a single sample from the posterior, and subtract it from the detector data. The timescale to model glitches and produce glitch-subtracted data using `BayesWave` is up to multiple days depending on the type of glitch that is being analyzed [30]. The code for both `BayesWave` and the `BayesWave` glitch subtraction methods can be found at [51].

2.2. Linear subtraction with a witness

The codebase used to subtract glitches using an auxiliary witness, `gwsbtract`, is the same linear subtraction algorithm as was presented in [40]. In this section, we provide a short description of this method, and note how the use case considered in this work differs from the original use case presented in [40]. The code associated with `gwsbtract` is publicly available at [52].

In order to subtract noise, we first assume that the measured strain in the gravitational-wave detector, $h(t)$, is a linear combination of time series from different sources, $n_i(t)$. We further assume that at least one of these noise sources can be modelled as the convolution of a witness time series, $a(t)$, and an unknown transfer function $c_{ah}(t)$. In the time domain, the gravitational-wave strain can therefore be written as

$$\begin{aligned} h(t) &= n_1(t) + n_2(t) + \dots + n_N(t) \\ &= n_1(t) + n_2(t) + \dots + a(t) * c_{ah}(t) \\ &= h'(t) + a(t) * c_{ah}(t) \end{aligned} \quad (3)$$

where all noise sources except one are included in $h'(t)$. This can be conveniently cast into the frequency domain as

$$\tilde{h}(f) = \tilde{h}'(f) + \tilde{a}(f) \cdot \tilde{c}_{ah}(f) \quad (4)$$

Following the derivation from [35, 40], we can estimate the unknown transfer function by using the discrete Fourier transform of the strain data, $\tilde{Y}_h(f)$, and the witness time series, $\tilde{Y}_a(f)$. A single value is calculated for the frequency band $[f_1, f_2]$ via

$$\tilde{c}_{ah}(f') = \frac{df}{f_2 - f_1} \sum_{f=f_1}^{f_2} \tilde{Y}_h(f) \tilde{Y}_a^*(f), \quad (5)$$

where df is the frequency resolution of the data.

By averaging nearby frequencies when calculating the transfer function, contributions to the measurement from chance correlations are reduced. Specifically, the fractional measurement error from this averaging method is expected to be $\sqrt{\frac{df}{f_2 - f_1}}$ if both $h(t)$ and $a(t)$ are stationary Gaussian processes.

Although the basic description of the process for glitch subtraction is quite similar to [35, 40], the process has a few practical differences. In [35, 40], multiple witness time series were used to subtract contributions from a single noise source, requiring that correlations between each of the witness time series were also calculated. In this work, only a single witness time series is considered. As the glitch subtraction can be completed by `gwsbtract` with a few straightforward calculations, the timescale to produce this data for one event is generally less than multiple hours.

An additional key difference is the stationary nature of the time series. Data that contains glitches is, by definition, not a stationary Gaussian time series. Regardless of the use case, if the transfer function is still stationary and linear, then theoretically this should not prevent these linear subtraction methods from being used. However, it is possible that the true transfer functions may have some non-linear features that are not captured by this method. It is also possible that some of the methods used in [40] to reduce the impact of false correlations could prevent the transfer function from being measured accurately. For these reasons, the ideal type of glitch to subtract with these methods is one that occurs at a high rate over a short time period. This increases the probability that the transfer function will be accurately measured.

Due to these practical differences between glitch subtraction and the use case that [40] was originally designed for, we completed additional validation tests to confirm that the methods of [40] were able to be used for glitch subtractions. These tests are summarized in [Appendix A](#).

2.3. Estimating the significance of residuals

One of the key steps in the glitch subtraction process, after deciding whether to subtract excess power in the data, is determining whether this excess power was successfully subtracted. For these purposes, we employ a residual test inspired by [53] to evaluate the statistical significance of any non-Gaussian features in the data. This test relies upon comparing the power spectral density (PSD) of the data during the time of interest against the PSD measured over 512 seconds. Any positive fluctuations in the short-duration PSD are considered excess power.

Following the description in [53], we aim to measure the PSD variation, ρ , which is

$$\rho = N \int_{t_0 - \Delta t}^{t_0} |F(t) * s(t)|^2 dt \quad (6)$$

for some data, $s(t)$, at a given time, t_0 . Δt is the timescale on which the PSD variation is measured, $F(t)$ is a filter function to weight different frequencies that are being

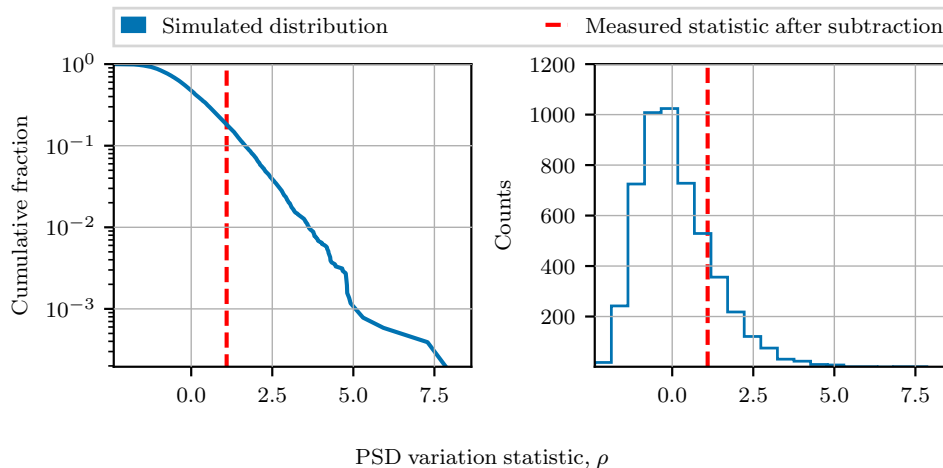


Figure 1: An example of the distribution of the PSD variation statistic, ρ that is used to calculate the p-value of any excess power identified in the data. This example uses data from around GW190425 [23] from LIGO Livingston. The distribution of the statistic over the simulated background noise is shown in blue and the measured statistic after glitch mitigation is marked in red. The left panel shows the cumulative distribution of the statistic in the simulated data, while the right panel shows a histogram of the measured statistic values. Before subtraction, the statistic is much larger than any trial in the simulated background (and is not shown on this plot). After subtraction, the measured statistic is consistent with the background.

analyzed, and N is a normalizing constant. In the frequency domain with discrete samples, this is

$$\rho = N \sum_{f=f_{\text{low}}}^{f_{\text{high}}} |F(f)s(f)|^2 \Delta f \quad (7)$$

We approximate the squared frequency spectrum of the data, $|s(f)|^2$, as the PSD of the data measured using `GWpy` [54] over a short duration that overlaps the glitch. We refer to this as $S_G(f)$. Rather than designing a filter function based on a gravitational-wave inspiral signal (as was done in [53]), we choose to use a filter function that is weighted by the sensitivity of the detector. Specifically, $F(f) = 1/S_L(f)$, where $S_L(f)$ is the PSD of the data over 512 seconds of data measured using the “median method” [55] with windows the same size as the short duration PSD and 50% overlap. With these substitutions, the variation in the PSD, ρ is

$$\rho = \sum_f S_L(f) \times \sum_f \frac{S_G(f)}{S_L^2(f)}. \quad (8)$$

where the normalizing constant, N , is

$$N = \sum_f \frac{S_L(f)}{\Delta f}. \quad (9)$$

However, the variance of this statistic is dependent on both the time window used to estimate $S_G(f)$, Δt , and the “effective bandwidth” considered, Δf_{eff} [53], making it unsuitable for comparisons where these two quantities differ. Specifically, the variance is

$$\text{Var}(\rho) = (\Delta t \Delta f_{\text{eff}})^{-1}. \quad (10)$$

The effective bandwidth is dependent on both the bandwidth considered, $\Delta f = f_{\text{high}} - f_{\text{low}}$, and the shape of the PSD. This effective bandwidth is calculated via

$$\Delta f_{\text{eff}} \approx \int_{f_{\text{low}}}^{f_{\text{high}}} \frac{\max[S_L(f)]}{S_L(f)} df. \quad (11)$$

To account for this dependence on Δt and Δf , we normalize the statistic by the standard deviation for the parameters we use in each instance and subtract the mean of the distribution. Hence our new statistic, $\tilde{\rho}$ is

$$\tilde{\rho} = (\rho - 1) \times \sqrt{\Delta t \Delta f_{\text{eff}}}. \quad (12)$$

This new statistic has a mean of 0 and a variance of 1 for all choices of Δt and Δf .

Although this new statistic is less susceptible to analysis choices for the reasons listed above, the true variance of the statistic may be different due to practical limitations in the measurement of the PSD. For example, it is not possible to calculate this statistic when $\Delta f < \Delta t^{-1}$ as there will be no points in the measured PSD in the bandwidth of interest. When $\Delta f \approx \Delta t^{-1}$, we also see additional variance in the statistic that is not accounted for in the normalization due to the large frequency resolution of the PSD. If the frequency resolution of the PSD is sufficiently coarse, only a small number of data points are the bandwidth of interest, inflating the measured variance.

To address this issue, we calculate a p-value for the measured value of $\tilde{\rho}$ for a specific choice of Δt and Δf using simulated data. In each case, we generate 2048 seconds of stationary, Gaussian data with the measured long-duration PSD using Bilby [56, 57]. We then repeat the same procedure over each sequential short-duration time window for the whole time period simulated. These results are used to generate the distribution from which the p-value is estimated.

An example distribution of this statistic based on data around GW190425 [23] is shown in Figure 1. Before subtraction, the measured statistic value ρ was ≈ 180 , much larger than any trial in the simulated background. After subtraction, the statistic dropped to 1.4, which was higher than 18% of the simulated noise, corresponding to a p-value of 0.18. This can be seen from the cumulative distribution of the simulated noise, shown on the left, and a histogram of the measured statistics in the noise, shown on the right. The plotted histogram also demonstrates that the peak of this distribution is indeed close to 0. The distribution of statistics is asymmetric, with a longer tail towards larger values. This asymmetry, even for cases that should be well-behaved, motivates the use of the p-value rather than a threshold on the measured statistic based on an assumed variance.

2.4. Glitch subtraction validation procedures

In order to ensure that the glitch-subtracted data used in downstream analyses do not introduce biases, a series of glitch-subtraction review tests were completed as a

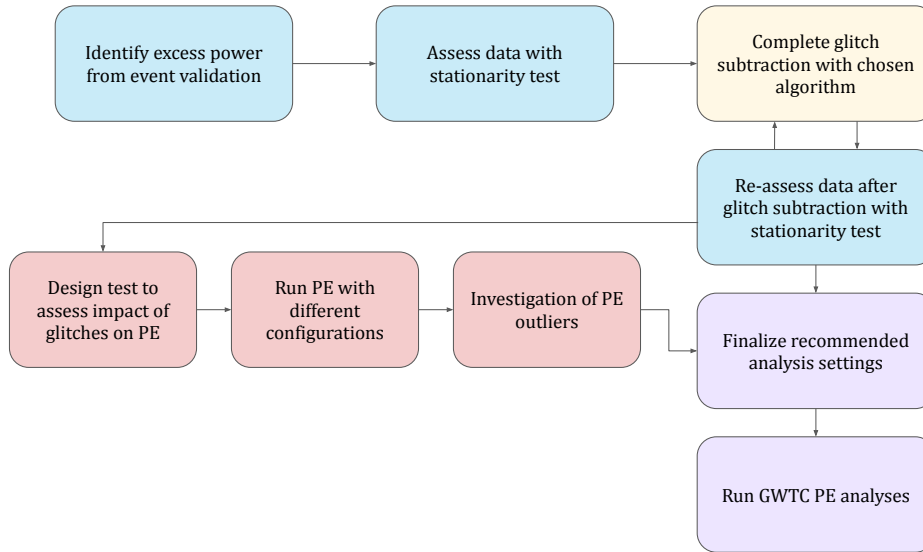


Figure 2: Flowchart of the glitch subtraction and parameter estimation (PE) procedures discussed in section 2.4. Note that the the two steps involving “Complete glitch subtraction with chosen algorithm” and “Re-assess data after glitch subtraction with stationarity test” may be completed multiple times if the first attempt at glitch subtraction does not pass the stationarity test. The coloring corresponds to the general theme of each part of the process: blue is data evaluation steps, beige is modelling and subtracting identified glitches, red is tests with parameter estimation analyses, and purple is coordinating and completing parameter estimation analyses for GWTC analyses [2–4].

part of the O3 LVK analyses [2–4]. These tests focus on ensuring that either the data was consistent with Gaussian noise or that any residual instrumental excess power does not bias estimates of an event’s source properties. Previous investigations into the potential biases on parameter estimation when glitches overlap or are nearby signals [14–17] have not yet covered the entire parameter space of glitch types and where in the signal the glitch overlaps. We therefore use conservative thresholds for each test to minimize potential biases.

In order to identify if data surrounding gravitational-wave events required glitch subtraction, we first utilize the results of event validation procedures completed for each event. These procedures for O3 are detailed in [10, 58]. The outcome of event validation specifically indicates if any glitches were nearby a gravitational-wave signal. Events that do not have any glitches flagged by these procedures are deemed not to require glitch subtraction.

For flagged events, the residual test described in Section 2.3 is used to determine if the data is inconsistent with Gaussian noise. The time-frequency bounds to use for the residual test are determined by visual inspection of the data around events using spectrograms produced with the Q-transform [59]. Cases where the residual p-value was greater than 0.1 are deemed consistent with Gaussian noise and no glitch subtraction is attempted, cases with p-values less than 0.001 are passed on for glitch subtraction, and cases with p-values between these two thresholds require

further follow up before a decision was reached. In the cases when the result is inconclusive, the overlap of the excess power and signal is used to determine if glitch subtraction is needed; glitch subtraction is generally recommended in cases where excess power directed overlapped the signal. As previously described, glitch subtraction was completed with `BayesWave` in all but 1 case where `gwsubtract` was used. The choice of algorithm for each event is made based on the suitability of each algorithm to subtract the glitches in glitches. Technical considerations are not a part of this choice as the time to completion and computing usage are not a limiting factor in the generation of reviewed, glitch-subtracted data.

Once the glitch subtraction was completed, the data is reevaluated to determine if the excess power is sufficiently subtracted. The residual metric is recomputed for the glitch-subtracted data using the same time-frequency regions. The same p-value thresholds are also used to decide if the data is now consistent with Gaussian noise.

In addition to the residual tests, comparisons of parameter estimation results using different mitigation methods are used to evaluate if further glitch subtraction is needed or if a different glitch mitigation method should be used[‡]. To test if excess power in the data could bias analyses, we estimate the source properties of events using either the nominal lower frequency limit (20 Hz) or a higher lower limit that excludes the data containing glitches. In cases where the glitch subtraction is able to improve the stationarity of the data, analyses using the original and glitch subtracted data are also compared. If the glitch-subtracted data does not pass the residual test, different frequency bounds that exclude the flagged time-frequency regions are recommended if the different frequency bound changed the measured chirp mass[§] posterior by more than 1σ and that the signal-to-noise ratio (SNR) is reduced by less than 10%. Other parameters, such as χ_p and χ_{eff} [§] are also compared between analyses. Although not the deciding factor in determining which mitigation method is recommended, any differences in these parameters between analyses are flagged. Additional details about these tests can be found in [Appendix B](#).

Once either the data around an event passed the residual test or the identified glitches are deemed to not bias parameter estimation, the data is approved for use in downstream analyses. These procedures generally are completed for each event within a few days (for events not requiring glitch subtraction) to weeks (for events that require glitch subtraction). The timeline for this process is limited by the need for human input at multiple steps; in future observing runs, additional automation will be needed to significantly speed up these procedures.

3. Results with real events

In this section, we consider the data around two events, GW190424 [2] and GW200129 [3]. Both events directly overlapped glitches that were mitigated using the glitch subtraction procedures discussed in this work. Data around GW190424 was processed with `BayesWave` and data around GW200129 was processed using `gwsubtract`. GW190424 was one of the more challenging cases of glitches that were subtracted with `BayesWave` while GW200129 is the only event in O3 processed with `gwsubtract`. In both cases, however, there are well-known witnesses to the source of

[‡] These parameter estimation comparison tests are completed when practical for all events that fail the initial residual test, regardless if the glitch-subtracted data passes the residual test.

[§] See [57] and references therein for definitions of these parameters.

the glitches near each event. We are therefore confident that the power subtracted is due to glitches as opposed to an unmodelled astrophysical signal.

GW190424 directly overlapped two different types of glitches; one glitch was due to a mechanical camera shutter that was inadvertently left running on a timer [58] and one was due to scattered light that is reflected off a moving surface [60]. A data quality flag was developed that removed time segments containing glitches from the astrophysical searches due to these camera shutter glitches [61]. GW190424 is an event of interest as one of the signals in O3 that was detected using only the data from a single gravitational-wave detector. There was therefore no data from a different detector available to help disentangle coherent power (that is astrophysical) from incoherent power (that is instrumental). Instead, glitch power was identified using auxiliary witnesses and then modelled using `BayesWave`. For events in O3 that were detected in more than one detector, it was instead possible to disentangle incoherent and coherent power using the multiple strain time series. Although these noise sources were correlated with auxiliary witnesses, `gwssubtract` was not feasible to be used in this case as scattered light is not linearly correlated with any witnesses and the camera shutter glitches were too infrequent and short to accurately measure the relevant transfer function.

GW200129 directly overlapped glitches that were due to a fault in the 45 MHz electro-optic modulator driver system at LIGO Livingston [62]. This type of glitch was repeating for many minutes around the event and was well-witnessed by the modulation control system. Furthermore, the broadband nature of this glitching is challenging to model with the sine-Gaussian wavelets used in `BayesWave`. For these reasons, this glitch type was addressed using linear subtraction. These glitches were also vetoed by a data quality flag in O3 [61]. GW200129 is of interest astrophysically as the gravitational-wave event with potentially the strongest evidence to date in favor of spin-induced precession in a CBC [3, 63–65]. The gravitational-wave strain from a spin-precessing binary is lower at specific frequencies as compared to the corresponding non-precessing signal. As the effect of orbital precession could be mimicked by the presence of non-Gaussian noise, it is essential to the analysis of any signal that any glitches present are well-modelled and subtracted. Conversely, the uncertainties on glitch modelling could introduce additional uncertainties in the parameter estimation results [65].

Including these two events, a total of 16 events required glitch subtraction in O3. Table 1 lists all of these events, the interferometer from which the glitch was subtracted, and the time-frequency regions that were targeted for glitch subtraction. Note that due to the nature of the glitch subtraction algorithms used, glitch power may also have been subtracted from outside the listed time-frequency regions. However, these time-frequency bounds correspond to the time windows and bandwidth targeted for glitch subtraction and used for residual tests. An additional 10 events^{||} were assessed with these review processes but were determined to not require glitch subtraction. This was due to the identified time-frequency regions passing the residual test or being located outside the time window used in parameter estimation analyses. All glitch-subtracted data is publicly available [66–68].

^{||} GW190720_000836, GW190828_065509, GW190910_112807, GW190915_235702, GW191204_171526, GW200208_130117, GW200219_094415, GW200220_124850, GW200311_115853, and GW200220_124850

Event	GPS time	IFO	Time (s)	Frequency (Hz)
GW190413_134308	1239198206.74	L	[-1.7, 0.7]	[20, 35]
GW190424_180648	1240164426.14	L	[-.3, 0.1]	[60, 100]
			[-1.5, 0.5]	[20, 40]
GW190425	1240215503.02	L	[-61.57, -59.97]	[20, 50]
GW190503_185404	1240944862.30	L	[-1.1, 0.5]	[20, 35]
		L	[-3.9, -1.9]	[20, 40]
GW190513_205428	1241816086.75	L	[-5.65, 0.15]	[15, 40]
GW190514_065416	1241852074.89	L	[-1.3, 2.7]	[15, 30]
		L	[-5.3, 0.3]	[15, 30]
GW190701_203306	1246048404.80	L	[-1.28, 1.72]	[20,55]
GW190924_021846	1253326744.85	L	[-5.0, -3.0]	[18,110]
GW191105_143521	1256999739.93	V	[-7.05, -5.65]	[25, 50]
			[0.1, 7.1]	[15, 30]
			[-4.35, -3.15]	[16, 22]
			[-6.2, -5.0]	[18, 25]
GW191109_010717	1257296855.22	H	[-3.1, -0.1]	[25, 45]
		L	[-1.5, 1.5]	[20, 32]
GW191113_071753	1257664691.84	H	[-1.75, 1.75]	[25, 60]
			[-4.4, -3.6]	[25, 55]
GW191127_050227	1258866165.55	H	[-2.125, 0.375]	[25, 40]
			[0.0, 3.0]	[20, 45]
GW191219_163120	1260808298.45	H	[-7.85, -4.35]	[20, 1000]
		L	[4.5, 7.0]	[20, 25]
GW200105_162426	1262276684.06	L	[-4.25, 0.25]	[18, 22]
GW200115_042309	1263097407.74	L	[-70, 30]	[20, 25]
GW200129_065458	1264316116.44	L	[-0.525, 0.275]	[30, 60]
			[0.6, 1.6]	[30, 55]

Table 1: A list of all of the events from O3 LVK analyses [2–4] which required glitch subtraction. For each event, we list: the GPS time of the event; the interferometer(s) which required glitch subtraction; and the time window (relative to the event time) and frequency range targeted for glitch subtraction. These time-frequency regions were used for the residual tests of glitch-subtracted data. Note that additional excess power may have been subtracted outside of these time-frequency bounds. Glitches around all events were subtracted with `BayesWave` except for glitches around GW200129_065458, which were subtracted with `gwsubtract`.

3.1. Glitch models for events in O3

The process of glitch subtraction can be broken into two stages: glitch modelling and glitch subtraction. The subtraction process is done simply by finding the difference between the glitch model and the original data. However, as described in Section 2, the methods used by the tools described in this work to estimate the glitch model differ significantly. These differences are highlighted in the two representative events we have considered in this section.

The first data visualizations we will consider for our comparisons are spectrograms

generated using the Q-transform [59]. Spectrograms of the data around GW190424 and GW200129 from LIGO Livingston can be seen in Figure 3. The panels of this figure show the data before subtraction, after subtraction, and the difference of the two spectrograms. The power subtracted near GW190424 is confined to two distinct burst of power: one corresponding to power from the camera shutter glitch at ≈ 70 Hz, and the other one from the arches in the nearby scattered light glitch. These bursts of power are not directly overlapping the signal, but are near enough that they might bias parameter estimation results. Conversely, the power subtracted near GW200129 is more broadband, with visible glitch power below 70 Hz subtracted for multiple seconds that directly overlaps the signal.

A comparison of the original strain data, the deglitched data, and the glitch models can be seen in Figure 4. Both the glitch models and the strain data have been whitened using `GWpy` [54]. The same types of features that were visible in the Figure 3 can be seen in the whitened time series. The glitch model for GW190424 is confined to two short bursts, while the glitch model for GW200129 is non-zero over the entire time considered.

These results also demonstrate limitations of these two methods. `BayesWave` was not able to successfully subtract all of the excess power from the scattering glitches, while the linear subtraction was not able to subtract any power above 70 Hz. Scattered light glitches are a known challenge for `BayesWave`. Recent investigations [31] have shown that in order to fully model scattered light glitches with `BayesWave`, low-SNR wavelets must be used. These investigations also find that wavelets at the required SNRs are strongly disfavored when using priors similar to those used by the analyses discussed in this work. The linear subtraction method is also limited by the input data used to witness the glitch power. In the case of GW200129, the auxiliary sensor used in the subtraction was negligibly coherent with the gravitational-wave strain above 70 Hz. In other cases, such as the broadband noise subtraction performed in O2 [40], the sampling rate of the auxiliary sensors can limit the range of frequencies where subtraction is effective. Ultimately the effectiveness of the linear subtraction method relies upon how faithfully the relevant auxiliary sensors are able to witness the noise source of interest.

3.2. Estimated uncertainties

As previously stated, the uncertainty in the model of a glitch overlapping a gravitational-wave signal introduces another potential source of uncertainty in the parameter estimation results. Although it is possible to sample over the glitch model as an additional component in the parameter estimation (as is possible with recent updates to `BayesWave` [19, 31]), this was not done in the O3 analyses. The impact of accounting for the statistical uncertainties in the `BayesWave` glitch model while performing gravitational-wave parameter estimation is explored in [31]. Instead of these methods, a single glitch model was subtracted from the data as a preprocessing step. Although this means that any glitch modelling errors are not included in the analysis, this is not expected to significantly bias the results if the chosen glitch model is at least representative of the true glitch model. In this subsection, we discuss how we assess errors in glitch modelling for both `BayesWave` and `gwsubtract`, and demonstrate that statistical errors in the chosen glitch models are qualitatively low.

As discussed in Section 2.1, the Bayesian nature of `BayesWave` results in a posterior distribution of the glitch time series. Previous studies of both simulated

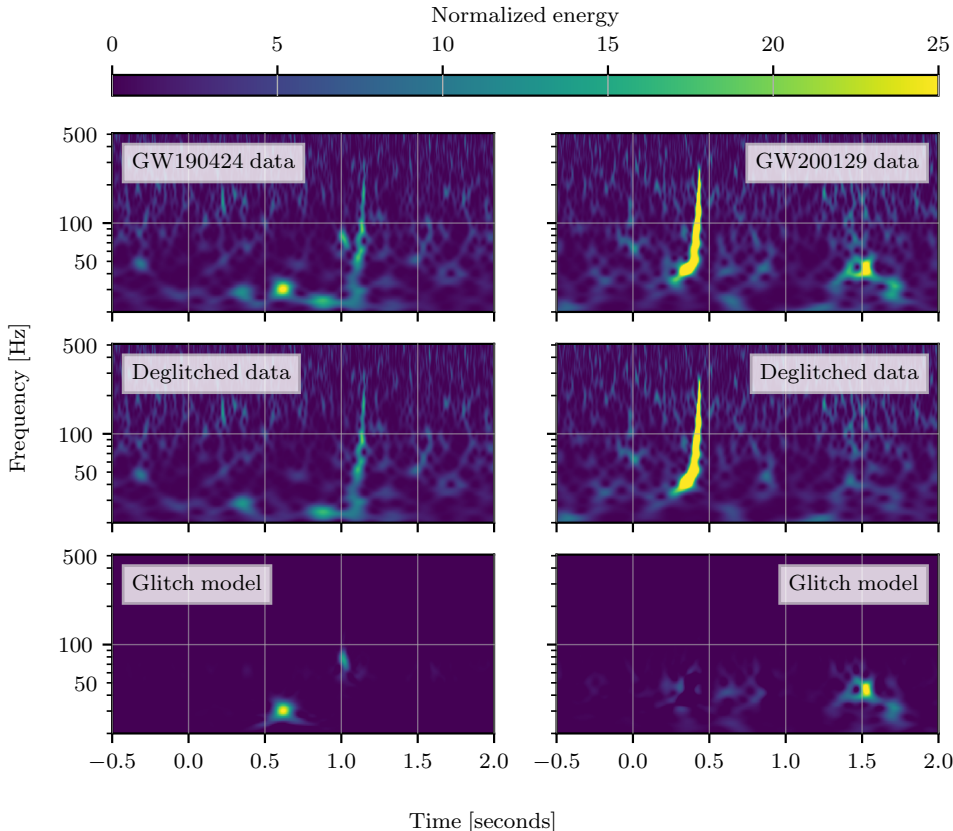


Figure 3: Spectrograms of the data around GW190424 and GW200129 before glitch subtraction (Top), the data after glitch subtraction (Middle), and the residual between the two which corresponds to the subtracted glitch model (Bottom). Data around GW190424 is shown on the left and data around GW200129 is shown on the right. All spectrograms are generated with GWpy [54] using the Q-transform [59]. The bottom panels highlight the time-frequency regions that were modified by glitch subtraction. Glitches around GW190424 were processed using **BayesWave**, resulting to the subtraction distinct time-frequency regions. Glitches around GW200129 were conversely processed with **gwsbtract**, resulting in broadband subtraction below 70 Hz.

signals [69] and real glitches [31] have shown that the accuracy of **BayesWave** is related to the duration of the data analyzed; **BayesWave** more accurately reconstructs short-duration ($\lesssim 1$ sec) events. To produce the glitch-subtracted data, we use a fair draw from the posterior distribution. An alternative choice is to use the median of the glitch time series posterior, however in some cases this can lead to oversubtraction [31]. An example of the 90% credible interval of the glitch time series posterior for GW190424, along with the whitened glitch model (a fair draw from the posterior), is shown in Figure 5. We see that although the fair draw is a good representation of a possible glitch model, we could get slightly different results depending on which glitch time

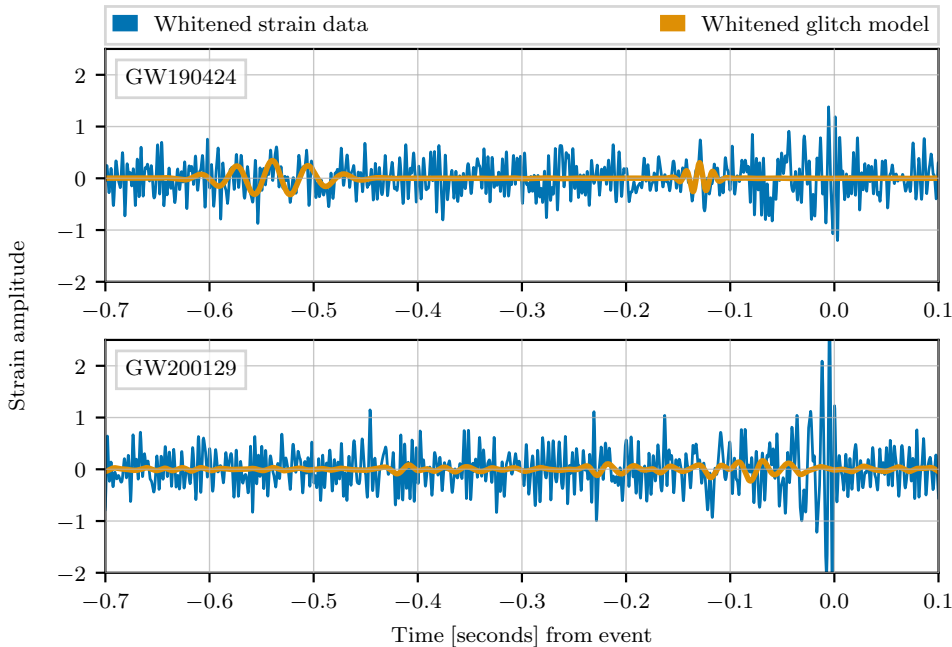


Figure 4: A plot of the whitened strain data from LIGO Livingston before glitch subtraction and the whitened glitch model that was subtracted near GW190424 and GW200129. The chirps corresponding to GW190424 and GW200129 are visible in the whitened strain data merging at $t = 0$. In both the amplitude and frequency of the glitches near these events are comparable to amplitude and frequency of the signal. The glitch model for GW190424 is two distinct wavelets, while the glitch model for GW200129 is always non-zero over the plotted time segment.

series we happen to draw.

Statistical errors in the glitch model from `gwssubtract` come from both chance correlations between the two data streams considered and errors in the transfer function measurement itself. As described in Section 2.2, the errors from chance correlations can be calculated directly from parameters used in the glitch modelling. In the case of GW200129 the 1σ statistical errors from chance correlations should be $\frac{1}{\sqrt{2048}} \approx 0.022$. The transfer function measurement errors, however, are not directly estimated as part of the glitch modelling process. However, if we assume that the transfer function is stationary over a timescale longer than used to measure the transfer function, we can take multiple measurements to estimate this source of error.

In order to qualitatively estimate the significance of measurement errors, we recalculate the model of the glitches overlapping GW200129 after changing the parameters used in `gwssubtract`. Specifically, we shift the time window used to estimate the transfer function by $\{-1024, -512, 512, 1024\}$ seconds with respect to original run. The glitch model estimated for these four additional attempts is shown in Figure 5 compared against the original model. The statistical errors on the original glitch model due to chance correlations is represented by the orange shaded region.

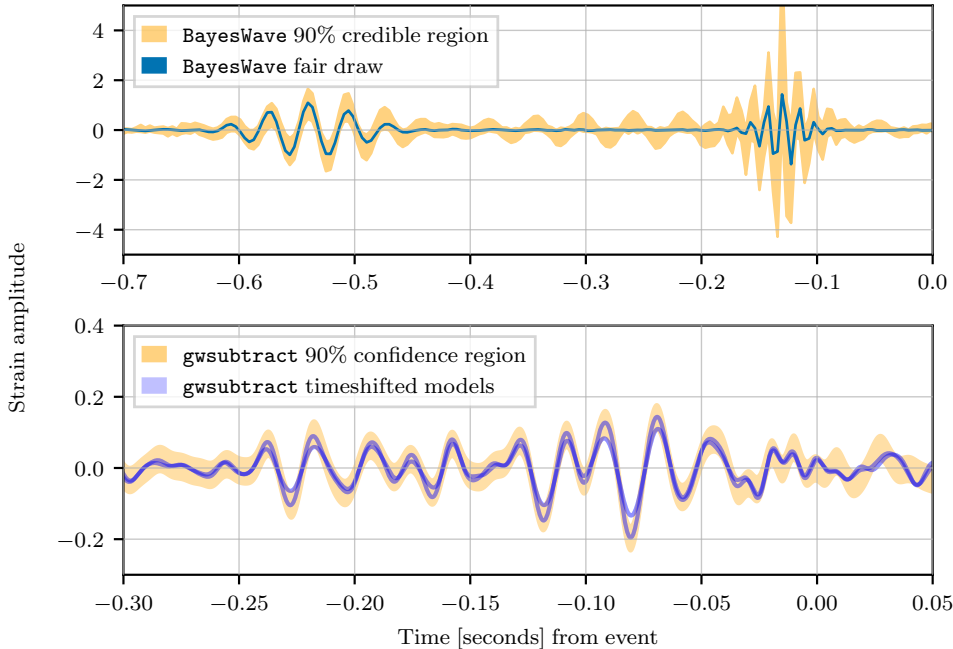


Figure 5: Top: The 90% credible interval of the whitened glitch time series posterior from `BayesWave`, and a fair draw of a single whitened waveform from the posterior (blue). Bottom: A comparison of the 90% confidence interval of the whitened glitch time series from `gsubtract` (orange) and the whitened glitch waveform subtracted when measuring the transfer function over different time periods (blue). The different time series are consistent in amplitude and phase. Note the differing axis scales for each panel. The `BayesWave` 90% credible is much larger than the plotted `gsubtract` 90% confidence interval due to the different methods used by each algorithm to estimate the errors. Due to the differing error estimation methods, the relative amplitude of the shown credible/confidence regions should not be used to compare the precision of the two methods.

All five runs show qualitative agreement in the amplitude and phase of the glitch, within the expected errors from chance correlations. We therefore conclude that the statistical errors from the glitch modelling procedure are low and should not introduce biases in the analysis.

It is important to note that these estimates of the errors do account for systematic errors due to modelling assumptions. Although the wavelet model used by `BayesWave` aims to be fully generic, this model, along with specific choices on the priors used in the analysis, is not able to fully model all glitch features for all glitches. Furthermore, specific glitch classes are known to be more problematic for `BayesWave` than others [31]. Likewise, `gsubtract` is limited by the assumption that the witness sensor used to subtract the excess noise is a perfect witness for the noise source and that the transfer function between the witness sensor and the gravitational-wave strain is linear over the amplitudes of interest. These types of systematic errors are mitigated by investigations

into the limitations of each glitch subtraction algorithm and the choice of which algorithm to apply in each scenario. This type of choice was one of the reasons that `gwsubtract` was used for one of the events in O3. Other problematic glitches impacting events in O3 were investigated using `gwsubtract`, but only the glitches around GW200129 were successfully mitigated with this algorithm.

The potential impact of unaccounted systemic error sources on parameter estimation analyses is not addressed in this work. However, studies of `BayesWave`-based parameter estimation suggest that systematic errors are low relative to other sources of error in estimation of gravitational-wave event source properties [31]. Systematic uncertainties in the case of `gwsubtract` are more difficult to probe, as these will vary with each noise source considered. Previous investigations into the use of `gwsubtract` and other tools for broadband subtraction did not identify biases [40, 70], but it is not clear if these studies are fully applicable in this case due to this differences explained in Section 2.2. As the sensitivity of the detectors increases, these systematic uncertainties may have a larger impact on the measured source properties. Recent investigations into GW200129 have already shown that the evidence of orbital precession from this event may be impacted by systematic errors in the glitch modeling [65].

4. Future outlook

The methods described in this work have been used to subtract glitches nearby 16 of the gravitational-wave events identified by LVK analyses in O3, representing $\approx 20\%$ the total population to date. The high rate of glitches and gravitational-wave signals means that many yet-to-be identified signals will overlap glitches as well. The need for this type of mitigation will only grow as the rate of discovery increases in future observing runs [71]. It is therefore important to consider glitch mitigation and subtraction as an inherent component of gravitational-wave data analysis.

As the detection rate grows further, there will be a need for additional automation in both the glitch subtraction algorithms and the process for deciding on their use and configuration. O3 represents the first time that procedures to identify the need for glitch subtraction were utilized due to the high rate of both glitches and signals. The need for analysis of more events in O4 will only reduce the amount of person power available on a per-event basis. One of the biggest foreseen challenges for future observing runs is reducing human input required to generate glitch-subtracted data. The procedures and methods described in this work will form a basis for the evaluation of events in future observing runs. This effort will also be significantly helped by investigations into the impact of glitches on specific gravitational-wave analyses and additional automation.

One key area of research that is required to be confident in the astrophysical implications of gravitational-wave signals overlapping glitches is the systematic uncertainties and limitations of different glitch subtraction methods. Further complicating these efforts is the potential variability between glitches of the same source, which makes it difficult to apply investigations from past observing runs to new gravitational-wave events. Although challenging to estimate, these systematic uncertainties may have significant impact on the conclusions derived from individual events as the sensitivity of the detectors increase. We therefore urge caution in interpreting results that are dependent on glitch subtraction algorithms when the potential systematics are not fully understood. Although this work and previous

studies have indicated that such systematics are qualitatively small for tools similar to those used in O3 [31, 39, 45], this may not be true for all analyses [65]. It is also important to note, however, that while these previous studies have shown that these tools have statistically unbiased results, this does not mean that analyses of individual events will not be impacted.

The case studies considered in this work also demonstrate the benefit of multiple glitch subtraction tools. In addition to the manpower limitations created by relying upon a single tool, the wide variety of glitches and glitch-signal overlaps that are possible mean that it is unlikely that a single glitch subtraction method will be best suited for every scenario. Even tools that are limited in scope can be useful, as was demonstrated by GW200129. Additional glitch subtraction methods also could be used as a potential cross-check to understand potential systematic due to specific glitch subtraction tools, as was done in [65]. The authors therefore encourage additional investigation into alternative, targeted glitch subtraction methods. However, new tools are only optimally beneficial when their effectiveness is compared to current methods to understand the specific cases in which one tool is better suited than another.

As previously described, there are a wide variety of additional glitch subtraction methods that are already available [32, 36–39, 41–49] or may be ready for use by the next observing run. There are also updates already implemented to `BayesWave` [19, 31] that were not available for use in O3. However, the overall landscape for the data quality of the detectors and the need for glitch subtraction is expected to remain unchanged, as the sources of numerous classes of glitch remain unknown [10, 12, 13]. With the next observing run planned to start in the coming year [71], now is the time for identifying lessons learned from the O3 glitch subtraction experience and developing new methods to address current limitations.

5. Acknowledgments

The authors thank the LIGO-Virgo-KAGRA Detector Characterization and Parameter Estimation groups for their input and suggestions during the development of this work. We thank Sophie Hourihane and Katerina Chatziioannou for discussions about `BayesWave`. We also thank Ronaldas Macas for their comments during internal review of this paper. DD is supported by the NSF as a part of the LIGO Laboratory. IMR-S acknowledges support received from the Herchel Smith Postdoctoral Fellowship Fund.

This material is based upon work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation, and operates under cooperative agreement PHY-1764464. Advanced LIGO was built under award PHY-0823459. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. This work carries LIGO document number P2200192.

Parts of this research are supported by the Australian Research Council (ARC) Centre of Excellence for Gravitational Wave Discovery (OzGrav) (project number CE170100004) and ARC Discovery Project DP170103625.

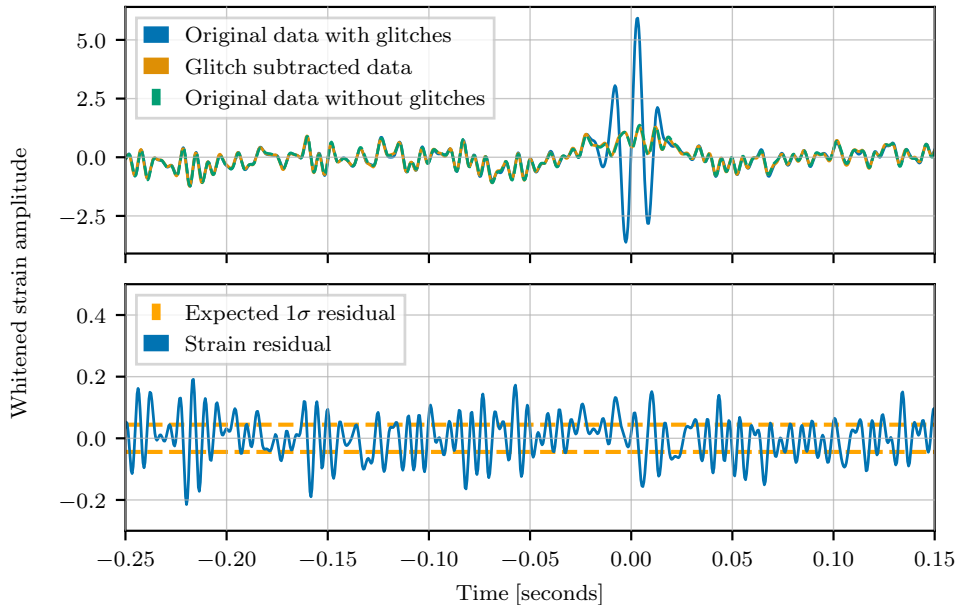


Figure A1: A comparison of whitened time series for simulated data (Top) with a glitch in blue, simulated data with a glitch subtracted in orange, and simulated data that does not contain the glitch in green. The similarity of the three time series for times far away from the glitch show that there was correctly only minimal subtraction done for these time periods. The residual between the simulated data with a glitch subtracted and the simulated data that does not contain the glitch (Bottom) is also shown in blue. The expected standard deviation of the residual is ≈ 0.05 , which roughly agrees with the measured standard deviation of 0.068.

Appendix A. Linear subtraction tests with simulated data

As `gwssubtract` has not been previously used to intentionally subtract glitches from gravitational-wave strain data, it is important to confirm that the algorithm can be used in this scenario. While glitches should theoretically be subtracted using the procedures described in Section 2.2, the practical differences listed in this section may make glitch subtraction difficult. Cases where glitches were subtracted were identified in investigations of the data produced in [40], but the glitch subtraction effectiveness was not assessed. To confirm that `gwssubtract` could be used to subtract glitches in a controlled setting, we test the algorithm with simulated data.

For this test, we first generate a time series with sine-Gaussian wavelets and a low amplitude background (i.e. an auxiliary witness) and a time series that includes both Gaussian noise colored to match the spectrum of LIGO Livingston in O3 and the same-Gaussian wavelets (i.e. the gravitational-wave strain). The sine-Gaussian wavelets in each channel are related by a known transfer function. The additional Gaussian noise in the witness channel is not correlated with the strain data and requires the algorithm

to disentangle correlated and uncorrelated noise in the witness. We then process the simulated strain data with `gwsubtract` to generate a glitch-subtracted time series. A comparison of the original strain data with glitches and the glitch-subtracted data is shown in the top panel Figure A1. Alongside these time series, we also plot the original data without the sine-Gaussian glitches included.

Ideally, the glitch-subtracted data should match the original data without glitches, up to expected statistical uncertainties. The difference between these two time series is shown in the bottom panel of Figure A1. With the settings used in this analysis, the residual should have standard deviation of $\frac{1}{\sqrt{512}} \approx 0.05$. The true standard deviation of the residual is 0.068, which qualitatively agrees with expectations.

Appendix B. PE studies of glitch subtraction

For events with relatively low-frequency glitches, there are two potential data mitigation strategies: (a) using only data above a certain cut-off frequency, below which the glitch resides; or (b) modelling and subtracting the glitch. To converge on the best approach to use for an event, we perform Bayesian inference on real data with various mitigation strategies applied. For each event, we then visually compare the posterior probability distributions for the network and per-detector signal-to-noise ratios (SNRs) and other parameters of interest to inform our choice of mitigation strategy. In every case for which these PE studies were performed, glitch subtraction was chosen as the preferred mitigation strategy.

We use `Bilby` and `bilby_pipe` [56, 57] to perform our parameter estimation and obtain posterior probability distributions on binary parameters. Aside from changing the data or lower frequency limits, the remainder of the settings are identical between runs on the same event. We use the `IMRPhenomPv2` waveform approximant [72, 73], a sampling frequency of 2048 Hz, a maximum frequency of 1024 Hz, and a reference frequency of 20 Hz. Unless altered as part of a mitigation strategy, the default lower frequency of analysis is 20 Hz. We employ marginalisation in phase, time and distance.

Here we focus on the results of two case studies: GW191109_010717 (Henceforth referred to as GW191109) and GW200129.¶ For both of these events, we perform analyses on 4s of data. Data from LIGO Hanford and LIGO Livingston was used for both events, while Virgo data was included for runs with GW200129 (Virgo was offline at the time of GW191109). We use `bilby` default priors in both cases: the `high_mass` prior for GW191109, and the 4s prior for GW200129 (see [57] for details of these priors).

GW191109 is one of the most likely events in O3 to exhibit signs of negatively-aligned spins. For GW191109, the data sets that we test each employ of the following mitigation strategies:

- (i) **“H45, L20”**: Limiting the frequency of Hanford data to above 45 Hz;
- (ii) **“H20, L45”**: Limiting the frequency of Livingston data to above 40 Hz;
- (iii) **“H45, L40”**: Limiting the frequency of Hanford and Livingston data to above 45 Hz and 40 Hz respectively;
- (iv) **“BW dg”**: BayesWave glitch subtraction.

¶ The other events that were subject to PE comparison studies were GW190413_134308, GW190424_180648, GW190513_205428, GW190514_065416, GW190727_060333, GW190828_065509, GW190910_112807, GW190915_235702, GW191105_143521, GW191127_050227, and GW200115_042309.

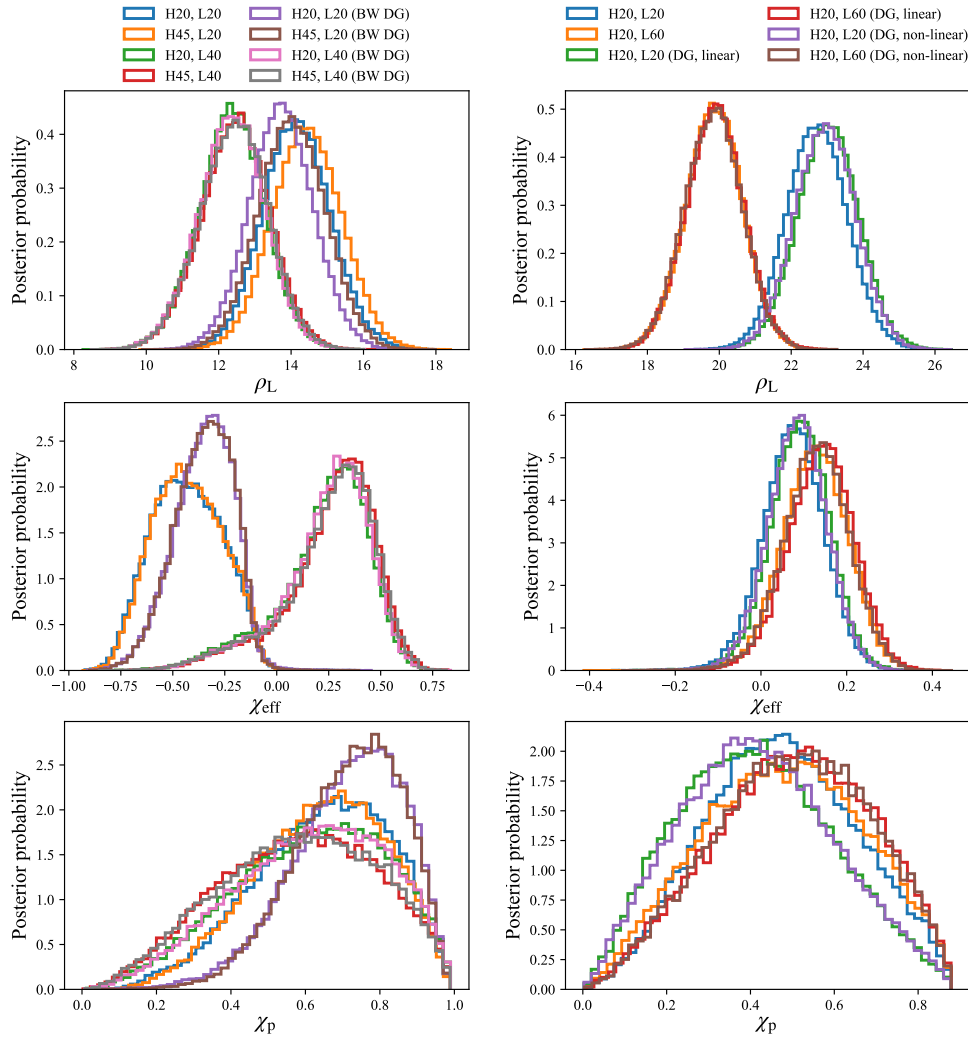


Figure B1: Plots showing the posteriors on matched-filter SNR in the LIGO Livingston detector ρ_L (Top), effective spin parameter χ_{eff} (Middle), and effective precession parameter χ_p (Bottom) under different mitigation strategies for GW191109 (Left) and GW200129 (Right). The unmitigated posterior is shown in blue in all plots. The “successful” posterior (representing the chosen mitigation strategy) is shown in purple for GW191109 (Left; “H20, L20 (BW dg)”) and green for GW200129 (Right; “H20, L20 (dg, linear)”). The Livingston SNR is highest when the nominal lower frequency of 20 Hz is used for the Livingston data. For GW191109, *BayesWave* deglitching preserves the support for negative χ_{eff} , while raising the lower frequency at Livingston to 40 Hz pushes the majority of the support to positive values. In both cases, the most informative posteriors come from deglitched data with the nominal lower frequency limit of 20 Hz.

GW200129 is one of the most likely events in O3 to contain evidence for spin-induced orbital precession. For GW200129, the compared data sets each individually have one of the following mitigation strategies applied:

- (i) **“H20, L60”**: Limiting the frequency of Livingston data to above 60 Hz;
- (ii) **“dg, non-linear”**: Linear glitch subtraction on data previously cleaned with broadband non-linear 60 Hz subtraction;
- (iii) **“dg, linear”**: Linear glitch subtraction on data previously cleaned with broadband linear 60 Hz subtraction.

We also compare these results against results obtained using unmitigated data (“H20, L20”).

Results are shown for matched-filter Livingston SNR, effective aligned spin χ_{eff} , and effective precession spin χ_{p} in Figure B1 for GW191109 and GW200129. We find that when we increase the lower-frequency limit of analysis for LIGO Livingston data, the Livingston optimal SNR is reduced, as expected. Meanwhile, there is little change to the SNR when the various glitch subtraction strategies are employed. There is also negligible difference between posteriors obtained with the linearly and non-linearly cleaned data for GW200129.

The posteriors on spin parameters χ_{eff} and χ_{p} are dependent on our choice of data mitigation strategy. When the lower frequency limit is raised above 20 Hz, information about the spins of the system can be lost, as less of the inspiral track is contained within the reduced extent of the data. Therefore, posteriors on spin parameters with limited low-frequency data are more influenced by the prior and are less well-constrained. In both cases, deglitching the data narrows the posteriors on χ_{eff} and χ_{p} relative to the posteriors computed from unmitigated data and reduced low-frequency data.

Low frequencies contain most of the information about χ_{p} , so the red histogram in both plots on the final row roughly traces the prior (uninformative) distribution on χ_{p} ; both events have more informative χ_{p} posteriors when deglitched than in the unmitigated case. In the case of GW191109, it appears that Hanford contains very little information about spin parameters: the χ_{eff} and χ_{p} posteriors with the Hanford lower frequency limit at 45 Hz and no deglitching (orange) follow the posteriors with no data mitigation strategy applied (blue). Since the SNR loss was significant when higher limits were placed on the lower frequency, our choice for both events demonstrated here was to deglitch the data and maintain a lower frequency limit of 20 Hz.

References

- [1] B. P. Abbott et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Phys. Rev. X*, (3):031040, 2019.
- [2] R. Abbott et al. GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *Phys. Rev. X*, 11:021053, 2021.
- [3] R. Abbott et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. 11 2021. arXiv:2111.03606.
- [4] R. Abbott et al. GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. 8 2021. arXiv:2108.01045.
- [5] B. P. Abbott et al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Class. Quant. Grav.*, 37(5):055002, 2020.
- [6] J. Aasi et al. Advanced LIGO. *Class. Quantum Grav.*, 32:074001, 2015.
- [7] F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quantum Grav.*, 32(2):024001, 2015.

- [8] T. Akutsu et al. KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector. *Nature Astron.*, 3(1):35–40, 2019.
- [9] A. Buikema et al. Sensitivity and performance of the Advanced LIGO detectors in the third observing run. *Phys. Rev. D*, 102(6):062003, 2020.
- [10] F. Acernese et al. Virgo Detector Characterization and Data Quality during the O3 run. 5 2022. arXiv:2205.01555.
- [11] H. Abe et al. Performance of the KAGRA detector during the first joint observation with GEO 600 (O3GK). 3 2022. arXiv:2203.07011.
- [12] D. Davis et al. LIGO detector characterization in the second and third observing runs. *Class. Quant. Grav.*, 38(13):135014, 2021.
- [13] T. Akutsu et al. Overview of KAGRA: Calibration, detector characterization, physical environmental monitors, and the geophysics interferometer. *PTEP*, 2021(5):05A102, 2021.
- [14] J. Powell. Parameter Estimation and Model Selection of Gravitational Wave Signals Contaminated by Transient Detector Noise Glitches. *Class. Quant. Grav.*, 35(15):155017, 2018.
- [15] J. Y. L. Kwok, R_l K. L. Lo, A. J. Weinstein, and T. G. F. Li. Investigation of the effects of non-Gaussian noise transients and their mitigation in parameterized gravitational-wave tests of general relativity. *Phys. Rev. D*, 105(2):024066, 2022.
- [16] S. Mozzon, G. Ashton, L. K. Nuttall, and A. R. Williamson. Does non-stationary noise in LIGO and Virgo affect the estimation of H_0 ? 10 2021. arXiv:2110.11731.
- [17] R. Macas, J. Pooley, L. K. Nuttall, D. Davis, M. J. Dyer, Y. Lecoecue, J. D. Lyman, J. McIver, and K. Rink. Impact of noise transients on low latency gravitational-wave event localization. *Phys. Rev. D*, 105(10):103021, 2022.
- [18] C. Talbot, E. Thrane, S. Biscoveanu, and R. Smith. Inference with finite time series: Observing the gravitational Universe through windows. *Phys. Rev. Res.*, 3(4):043049, 2021.
- [19] 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(4):044013, 2021.
- [20] C. D. Capano, M. Cabero, J. Westerweck, J. Abedi, S. Kastha, A. H. Nitz, A. B. Nielsen, and B. Krishnan. Observation of a multimode quasi-normal spectrum from a perturbed black hole. 5 2021. arXiv:2105.05238.
- [21] S. Kumar, A. H. Nitz, and X. J. Forteza. Parameter estimation with non stationary noise in gravitational waves data. 2 2022. arXiv:2202.12762.
- [22] B. P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017.
- [23] B. P. Abbott et al. GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4M_{\odot}$. *Astrophys. J. Lett.*, 892(1):L3, 2020.
- [24] R. Abbott et al. GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object. *Astrophys. J. Lett.*, 896(2):L44, 2020.
- [25] R. Abbott et al. Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences. *Astrophys. J. Lett.*, 915(1):L5, 2021.
- [26] S. A. Usman et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.*, 33(21):215004, 2016.
- [27] B. Zackay, T. Venumadhav, JavJ.ier Roulet, L. Dai, and M. Zaldarriaga. Detecting gravitational waves in data with non-stationary and non-Gaussian noise. *Phys. Rev. D*, 104(6):063034, 2021.
- [28] C. Pankow et al. Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817. *Phys. Rev.*, D98(8):084016, 2018.
- [29] N. J. Cornish and T. B. Littenberg. BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches. *Class. Quant. Grav.*, 32(13):135012, 2015.
- [30] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse. BayesWave analysis pipeline in the era of gravitational wave observations. *Phys. Rev. D*, 103(4):044006, 2021.
- [31] 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. 5 2022. arXiv:2205.13580.
- [32] J. Merritt, B. Farr, R. Hur, . Edelman, and Z. Doctor. Transient glitch mitigation in Advanced LIGO data. *Phys. Rev. D*, 104(10):102004, 2021.
- [33] W. Wei and E. A. Huerta. Gravitational Wave Denoising of Binary Black Hole Mergers with Deep Learning. *Phys. Lett. B*, 800:135081, 2020.
- [34] A. Torres-Forné, E. Cuoco, J. A. Font, and A. Marquina. Application of dictionary learning to

- denoise LIGO's blip noise transients. *Phys. Rev. D*, 102(2):023011, 2020.
- [35] B. Allen et al. Automatic cross-talk removal from multi-channel data. 1999. arXiv:gr-qc/9909083.
- [36] J. C. Driggers, M. Evans, K. Pepper, and R. X. Adhikari. Active noise cancellation in a suspended interferometer. *Review of Scientific Instruments*, 83(2):024501, 2012.
- [37] G. D. Meadors, K. Kawabe, and K. Riles. Increasing LIGO sensitivity by feedforward subtraction of auxiliary length control noise. *Classical and Quantum Gravity*, 31(10):105014, 2014.
- [38] V. Tiwari, M. Drago, V. Frolov, S. Klimenko, G. Mitselmakher, V. Necula, G. Prodi, V. Re, F. Salemi, G. Vedovato, and I Yakushin. Regression of environmental noise in LIGO data. *Classical and Quantum Gravity*, 32(16):165014, 2015.
- [39] J C Driggers et al. Offline noise subtraction for Advanced LIGO. *Technical document LIGO-P1700260*, 2017.
- [40] D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall. Improving the sensitivity of advanced ligo using noise subtraction. *Class. Quant. Grav.*, 36(5):055011, 2019.
- [41] M. Was, R. Gouaty, and R. Bonnand. End benches scattered light modelling and subtraction in advanced Virgo. *Class. Quant. Grav.*, 38(7):075020, 2021.
- [42] A. Viets and M. Wade. Subtracting narrow-band noise from ligo strain data in the third observing run. Technical Report T2100058, LIGO DCC, 2021.
- [43] F. Acernese et al. Calibration of Advanced Virgo and reconstruction of detector strain $h(t)$ during the Observing Run O3. *Class. Quant. Grav.*, 39(4):045006, 2022.
- [44] H. Abe et al. Noise subtraction from KAGRA O3GK data using Independent Component Analysis. 6 2022. arXiv:2206.05785.
- [45] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale. Machine-learning nonstationary noise out of gravitational-wave detectors. *Phys. Rev. D*, 101:042003, Feb 2020.
- [46] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis. Noise Reduction in Gravitational-wave Data via Deep Learning. *Phys. Rev. Res.*, 2(3):033066, 2020.
- [47] N. Mukund et al. Bilinear noise subtraction at the GEO 600 observatory. *Phys. Rev. D*, 101(10):102006, 2020.
- [48] H. Yu and R. X. Adhikari. Nonlinear Noise Cleaning in Gravitational-Wave Detectors With Convolutional Neural Networks. *Front. Artif. Intell.*, 5:811563, 2022.
- [49] K. Mogushi. Reduction of transient noise artifacts in gravitational-wave data using deep learning. 5 2021. arXiv:2105.10522.
- [50] T. B. Littenberg and N. J. Cornish. Bayesian inference for spectral estimation of gravitational wave detector noise. *Phys.Rev.*, D91(8):084034, 2015.
- [51] BayesWave. <https://git.ligo.org/lscsoft/bayeswave>.
- [52] gwsubtract repository. <https://git.ligo.org/derek.davis/gwsubtract>.
- [53] S. Mozzon, L. K. Nuttall, A. Lundgren, T. Dent, S. Kumar, and A. H. Nitz. Dynamic Normalization for Compact Binary Coalescence Searches in Non-Stationary Noise. *Class. Quant. Grav.*, 37(21):215014, 2020.
- [54] 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, January 2021.
- [55] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton. FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries. *Phys. Rev. D*, 85:122006, 2012.
- [56] G. Ashton et al. BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy. *Astrophys. J. Suppl.*, 241(2):27, 2019.
- [57] I. M. Romero-Shaw et al. Bayesian inference for compact binary coalescences with bilby: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue. *Mon. Not. Roy. Astron. Soc.*, 499(3):3295–3319, 2020.
- [58] D. Davis, L. V. White, and P. R. Saulson. Utilizing aLIGO Glitch Classifications to Validate Gravitational-Wave Candidates. *Class. Quant. Grav.*, 37(14):145001, 2020.
- [59] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis. Multiresolution techniques for the detection of gravitational-wave bursts. *Class. Quantum Grav.*, 21:S1809–S1818, 2004.
- [60] S. Soni et al. Reducing scattered light in LIGO's third observing run. *Class. Quant. Grav.*, 38(2):025016, 2020.
- [61] D. Davis et al. Data Quality Vetoes Applied to the Analysis of LIGO Data from the Third Observing Run. Technical Report DCC-T2100045, LIGO, 2021.
- [62] B. P. Abbott et al. Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Class. Quant. Grav.*, 33(13):134001, 2016.

- [63] M. Hannam, C. Hoy, J. E. Thompson, S. Fairhurst, and V. Raymond. Measurement of general-relativistic precession in a black-hole binary. 12 2021. arXiv:2112.11300.
- [64] V. Varma, S. Biscoveanu, T. Islam, F. H. Shaik, C. Haster, M. Isi, W. M. Farr, S. E. Field, and S. Vitale. Evidence of Large Recoil Velocity from a Black Hole Merger Signal. *Phys. Rev. Lett.*, 128(19):191102, 2022.
- [65] E. Payne et al. The curious case of gw200129: interplay between spin-precession inference and data-quality issues. 2022. arXiv:2206.11932.
- [66] LIGO Scientific Collaboration and Virgo collaboration. Glitch model for O3a catalog events. Technical Report T2000470-v2, LIGO DCC, 2020.
- [67] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Glitch modelling for events. <https://doi.org/10.5281/zenodo.5546680>, November 2021.
- [68] LIGO Scientific Collaboration and Virgo Collaboration. GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run - Glitch modelling for events. <https://doi.org/10.5281/zenodo.6477076>, April 2022.
- [69] S. Ghonge, K. Chatziioannou, J. A. Clark, Tyson Littenberg, M. Millhouse, L. Cadonati, and N. Cornish. Reconstructing gravitational wave signals from binary black hole mergers with minimal assumptions. *Physical Review D*, 102(6):064056, September 2020.
- [70] J.C. Driggers et al. Improving astrophysical parameter estimation via offline noise subtraction for Advanced LIGO. *Phys. Rev. D*, 99(4):042001, 2019.
- [71] B. P. Abbott et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo. *Living Rev. Relat.*, 19:1, 2016.
- [72] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *Phys. Rev. D*, 93(4):044006, 2016.
- [73] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé. Frequency-domain gravitational waves from non-precessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93:044007, 2016.