



Lopez, D., Gayathri, V., Pai, A., Heng, I. S. , Messenger, C. and Gupta, S. K. (2022) Utilizing Gaussian mixture models in all-sky searches for short-duration gravitational wave bursts. *Physical Review D*, 105(6), 063024.

(doi: [10.1103/PhysRevD.105.063024](https://doi.org/10.1103/PhysRevD.105.063024))

This is the Author Accepted Manuscript.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<https://eprints.gla.ac.uk/279349/>

Deposited on: 14 September 2022

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Utilizing Gaussian mixture models in all-sky searches for short-duration gravitational wave bursts

Dixeena Lopez,¹ V. Gayathri,² Archana Pai,³ Ik Siong Heng,⁴ Chris Messenger,⁴ and Sagar Kumar Gupta³

¹*Physik-Institut, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*

²*Department of Physics, University of Florida, PO Box 118440, Gainesville, FL 32611-8440, USA*

³*Department of Physics, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India*

⁴*SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom*

Coherent WaveBurst is a generic, multidetector gravitational wave burst search based on the excess power approach. The coherent WaveBurst algorithm currently employed in the all-sky short-duration gravitational wave burst search uses a conditional approach on selected attributes in the multidimensional event attribute space to distinguish between noisy events from that of astrophysical origin. We have been developing a supervised machine learning approach based on the Gaussian mixture modeling to model the attribute space for signals as well as noise events to enhance the probability of burst detection [1]. We further extend the GMM approach to the all-sky short-duration coherent WaveBurst search as a postprocessing step on events from the first half of the third observing run (O3a). We show an improvement in sensitivity to generic gravitational wave burst signal morphologies as well as the astrophysical source such as core-collapse supernova models due to the application of our Gaussian mixture model approach to coherent WaveBurst triggers. The Gaussian mixture model method recovers the gravitational wave signals from massive compact binary coalescences identified by coherent WaveBurst targeted for binary black holes in GWTC-2, with better significance than the all-sky coherent WaveBurst search. No additional significant gravitational wave bursts are observed.

I. INTRODUCTION

The Gravitational wave (GW) catalogs of signals from the compact binary coalescence (CBC), present in the first half of the third observing run (O3a), had been recently released [2, 3].¹ This includes intermediate black hole (IMBH) system GW190521 [6], the merger of two compact objects with the unequal mass ratio GW190814 [7], the inspiral of a binary neutron star (BNS) system, GW190425 [8]. Further, the first confident observation of GW signal from a neutron star-black hole (NSBH) binary coalescence was made during O3b run [9, 10].

Besides CBC signals, the current ground-based detectors are also sensitive to detecting GW short-duration bursts, such as signals from the core-collapse supernova (CCSN), cosmic strings, non-linear memory effects and isolated Neutron stars [11–15]. The LIGO-Virgo-Kagra collaboration has been actively searching for GW bursts signals since the fifth science run. The all-sky burst search for short-duration transient includes the burst signal duration of up to a few seconds [16–20]. This search is identified as the unmodeled search, sensitive to detect GW signals of different morphologies, including GW signals from the CBC system.

The morphology independent coherent WaveBurst (cWB) short-duration algorithm identifies triggers with short-duration excess power in the time-frequency domain [21]. The cWB made a significant contribution in

the detection of the first GW signal from the binary black hole (BBH) merger GW150914 [22] and the intermediate-mass black hole (IMBH) merger event [23]. In cWB, each trigger is associated with multiple attributes. The cWB adopts an empirical classification of noise-based glitches based on various threshold values applied to different attributes. The threshold values, though *ad-hoc*, are chosen based on the simulation exercise. These threshold choices vary based on the detector sensitivity during each run, the network combination, and the type of the transient signal.

Machine learning (ML) techniques are the possible option to address such problems as they offer powerful tools for classification between signal and noisy transients [24–33]. More generically, ML ideas have the notable potential to improve the detection sensitivity of unmodeled GW signals [1, 34–37]. ML techniques have also been used to enhance searches and classification of GW signals from CCSN [38].

In an earlier work [1], the authors had proposed an alternative ML based approach to address the *ad-hoc* thresholding on the attribute set. In this Gaussian mixture model (GMM)-based supervised machine learning approach, we modeled the cWB trigger attributes in the multidimensional attribute space. Thus, it provides an alternative approach to thresholds applied on multidimensional space to the detection problem under the scalar log-likelihood ratio. In this work, we continue to use the Gaussian mixture models in obtaining the model for the attributes in the multidimensional parameter set. Here, we explore this approach by carefully choosing the attribute set and investigating the dependence on the search sensitivity with the appropriate usage of the re-parametrization approach. We present the search sen-

¹ The third observing run (O3) of the Advanced LIGO [4] and Advanced Virgo [5] (April 1, 2019, to March 27, 2020) detectors was divided into two segments as O3a run (April 1, 2019 - October 1, 2019,) and O3b (November 1, 2019 - March 27, 2020).

sitivity for the generic short-duration gravitational-wave transients using the data from the O3a run with the low-frequency (16–1024 Hz) analysis.

The paper is organized as follows. Section II describes the unmodeled search algorithm cWB and Gaussian mixture modeling for cWB triggers. Section III discusses the dataset used to study the GMM approach to all-sky search, Gaussian mixture model generation, and sensitivity improvement to generic signal morphologies and core-collapse supernova waveforms. Section IV discusses the GMM method results on cWB triggers of coincident events from the O3a run of Advanced LIGO detectors. Finally, in Section V, we summarize the significance of the GMM method on an all-sky search for short-duration transient signals using a minimal model approach.

II. METHODOLOGY

A. Coherent WaveBurst algorithm

The cWB algorithm is a morphology-independent, multidetector GW signal detection algorithm that coherently maps the multidetector’s data into the multiresolution time-frequency scale domain using the Wilson-Daubechiers-Meyer wavelet transformation [39]. It is based on the constraint maximum-likelihood ratio approach applied to the strain data in the time-frequency domain. The clusters of excess energy pixels are selected above the noise level of the detector and labeled as triggers if they exceed the thresholds on coherent energy (E_c) and network correlation coefficient (c_c) [40, 41]. For each trigger, the cWB computes a collection of attributes that characterize signal as well as noisy transients properties.

We consider the following attribute sets for this study, which are generic for the short-duration transient signals. The network coherent energy E_c , the effective correlated signal-to-noise η_c , the network correlation coefficients c_{c0} and c_{c2} , network energy short-duration N_{ED} , the ratio between the reconstructed energy and the total energy N_{norm} , the residual noise energy measure χ^2 and attributes pertaining to the noise vetoes like Q_{veto} and L_{veto} . We list this attribute set in Table I [42]. In standard cWB-based analysis, the threshold on the multidimensional attribute space is placed to distinguish the noise based triggers from the astrophysical GW event.

B. Gaussian mixture modeling in postproduction

In [1], we proposed the Gaussian mixture modeling approach [43] to construct two distinct GMM models in the multidimensional trigger attribute set for astrophysical GW signals and noise glitches. We used the Bayesian information criterion (BIC) method to obtain the optimum number of Gaussians and the GMM model parameters (mean, covariance and weights). For each model (signal and noise), we had obtained the maximum log-likelihood

statistics $W = \ln(\hat{\mathcal{L}})|_{\hat{K}}$, where $\hat{\mathcal{L}}$ is the maximum value of the likelihood function with the optimum number of Gaussians \hat{K} [1, 44, 45]. Then, we defined the detection statistic, T , as

$$T = W_s - W_n. \quad (1)$$

where the subscript s and n stand for signal and noise. We had used this quantity as a ranking statistic for each cWB trigger instead of applying threshold values on the multidimensional attribute set. We suggest the reader to follow [1] for more details.

A well behaved multidimensional Gaussian distribution is preferred to obtain GMM. When the distribution of a selected subset of attributes is not well behaved (highly peaked), the GMM overestimates the required number of Gaussians. Thus, we either precondition the data by reparameterizing some of the cWB attributes or combining two attributes in a single attribute.

We reparameterize the subset of attributes η_c , c_{c0} , c_{c2} , N_{ED} , E_c , Q_{veto} and L_{veto} , which shows the non-Gaussian distribution by either applying a logarithmic or inverse sigmoid transformation. Additionally, we define a new parameter $L_{\text{ratio}} : L_{\text{veto1}}/L_{\text{veto0}}$ instead of using L_{veto0} and L_{veto1} as two different attributes. The distribution of attributes after the reparameterization is well behaved and better conditioned than before reparameterization. This shows a direct implication on the number of Gaussians in the GMM model. For example, for the data set considered here, the number of optimum Gaussians in the original attribute set for the signal and noise model is 113 and 115, whereas the optimum number of Gaussians is reduced to 90 and 82 for the reparameterized attribute set. More details about the dataset is in Sec. III. More details on reparameterization can be found in Appendix.

III. O3a ANALYSIS WITH GMM

This study uses publicly available O3a data [46, 47] from the LIGO Hanford and LIGO Livingston (HL) network in the low-frequency range (16-1024 Hz)². The total amount of coincidence data between the two detectors used for this analysis is 104.9 days. The distribution of accidental triggers is calculated by time-shifting the data of one LIGO detector with respect to the other LIGO detector by an amount that breaks any correlation between the detectors for an actual signal. For the cWB O3a analysis, 1000 years of background data were generated.

² As we focus on maximization of the detection efficiency, we consider the HL network for this study rather than HLV because of the similar reasoning mentioned in [20].

Attribute	Definition
E_c	Cross correlation of reconstructed waveforms between the detector pairs.
η_c	$\eta_c = \sqrt{\frac{E_c}{(K-1)}}$, where K is the number of detector.
c_{c0}	$c_{c0} = \frac{E_c}{(E_c +E_n)}$, where E_n is the energy of the residual noise.
c_{c2}	$c_{c2} = \frac{E_c \times c_{c0}}{(E_c +E_n)}$
N_{ED}	Energy disbalance of the trigger between the detectors.
N_{norm}	Ratio between the reconstructed energy and the total energy.
χ^2	$\chi^2 = \frac{E_n}{N}$, where N is the number of independent wavelet amplitude of the event.
Q_{veto0}	Energy distribution of an event over different time segments.
Q_{veto1}	An estimate of quality factor assuming the signal to be a CosGaussian (Q factor).
L_{veto0}	Central frequency of the reconstructed signal, to identify narrow band glitches.
L_{veto1}	Root mean square frequency of the reconstructed signal.
L_{veto2}	Energy ratio between pixel energy and total energy of the event.

TABLE I: Summary of the selected attributes associated with the short-duration burst transients in cWB and used for the GMM-based analysis.

Attribute	LF1	LF2	LF3
c_{c0}	> 0.8	> 0.8	> 0.8
c_{c2}	> 0.8	> 0.8	> 0.8
N_{norm}	> 2.5	> 2.5	> 2.5
Q_{veto0}	= 0	≠ 0	
Q_{veto1}	≤ 3	≤ 3	> 3

TABLE II: Definitions of the search bins used for the O3a low-frequency short-duration burst analysis. The different thresholds are applied to attributes c_{c0} , c_{c2} , N_{norm} , Q_{veto0} and Q_{veto1} to classify the triggers into different bins based on background trigger distribution.

A. The standard O3a cWB all-sky analysis

In standard cWB O3a analysis, triggers are required to pass frequencies above 24 Hz and high network correlation with c_c above 0.8 for the HL detectors. These triggers are classified into three different bins as LF1, LF2, and LF3 based on background trigger morphologies to isolate the triggers to a small part of the parameter space [20]. LF1 contains a population of glitches dominated by very short and loud blip-type glitches, with negligible energy outside the single oscillation [48]. LF2 contains triggers resembling blip glitches with Q factor ≤ 3 , and LF3 contains remaining high Q factor triggers. Table II lists the thresholds applied to cWB attributes for classification into different bins used in O3a analysis. After applying thresholds on the cWB attributes, the events are ranked based on inverse false alarm rate (iFAR). A trials factor of three is applied to iFAR values corresponding to three different background bins, and a threshold of 100 years on iFAR is used for the significant detection [20]. By contrast, GMM based postprocessing does not require the creation of multiple analysis bins

(see below).

B. GMM model generation

For O3a GMM analysis, we choose most of the attributes required to characterize the generic short-duration transient signal from standard cWB (Table I) to generate the GMM model. We exclude the attributes reconstructed signal's central frequency, duration, and strain from this analysis, as these attributes are strongly affected by the priors on the injection parameters for the foreground model [1, 42]. Since we do not want the GMM to develop a bias towards the choice of the signal population used in the simulation. For GMM analysis, we consider those triggers with $\eta_c > 5.5$ and $c_c > 0.5$ for the HL network. The cWB uses different thresholds on attributes to generate gravitational wave candidates and veto out the noise triggers in its standard configuration.

We consider 70% of noise background as well as simulated burst signals for training and the rest for testing the GMM method. We use the equal percentage of generic waveforms (Gaussian Pulse, sine-Gaussian wavelets, and white noise burst) as the training set aimed to improve the detection efficiency of all injected waveforms. We construct the GMM for the noise and signal triggers using this simulation [1]. More specifics of the models can be found in Appendix.

C. Data simulation

A set of generic short-duration burst waveforms (typically used in the all-sky short-duration bursts search [20]) are injected in the HL O3a noise to estimate the search sensitivity using the cWB plus GMM method. They are sine-Gaussian wavelets (SG), Gaussian pulses (GA),

and band-limited white-noise bursts (WNB) signals injected over a range of amplitude expressed in terms of the root-sum-squared strain amplitude (h_{rss}). The SG waveforms are characterized by mean frequency f_0 and quality factor Q , the GA waveforms are described by the duration τ , and the WNB signals are characterized by lower frequency bound f_{low} , frequency bandwidth Δf , and duration τ . The SG and GA signals are injected as $h_{\text{rss}} = (\sqrt{3})^N 5 \times 10^{-23} \text{Hz}^{-1/2}$ over a grid of maximum strain values with N ranges from 0 to 8, and WNB waveforms are injected uniformly as the square of the signal's distance. Linearly polarized signals are used for GA, and sine-Gaussian wavelets use both elliptical (SGE) and linearly (SG) polarized waveforms [17, 20]. Elliptical waveforms are uniform in the cosine of the source inclination angle, which is defined as the angle between total angular momentum and the line of sight. The WNB represents isotropic emission at the source and carries an equal amplitude from both polarization of GW strain at the detector [49]. Table III lists the simulated signals and their characteristic parameter values used in this analysis.

Sine-Gaussian Burst (SGW)			
No.	f_0 (Hz)	Q	
1	70	3	
2	70	9	
3	70	100	
4	100	9	
5	153	9	
6	235	3	
7	235	9	
8	235	100	
9	361	9	
10	554	9	
11	849	3	
12	849	9	
13	849	100	
White-Noise Burst (WNB)			
	f_{low} (Hz)	Δf (Hz)	τ (s)
14	150	100	0.1
15	300	100	0.1
16	750	100	0.1
Gaussian Pulse (GP)			τ (s)
17			0.1
18			1
19			2.5
20			4

TABLE III: List of generic burst waveforms and their characteristic parameters.

The low-frequency, generic all-sky short-duration burst search targets the frequency range of GW burst from most of the core-collapse supernovae (CCSN). To demon-

strate the robustness of the cWB plus GMM approach against different signal morphologies, here, we analyze the search sensitivity to CCSN waveforms used in all-sky short-duration transient analyses during O3a [20]. The waveforms are generated from five different 3D CCSN simulations models like s18 [50], m20 [51], s9 [52], m39 [53], and 35OC [54]. The waveforms are distributed uniformly in distance with maximum distance for s18, m20, s9, m39, and 35OC CCSN models set to 25 kpc, 5kpc, 5kpc, 70kpc, and 70 kpc, respectively [20].

D. Sensitivity improvement for generic signal morphologies

We use the receiver-operating-characteristic (ROC) curve to characterize the sensitivity improvement. The ROC curve of the standard cWB and the cWB plus GMM relates the detection efficiency against the false alarm rate (FAR) is shown in Fig. 1. The detection efficiency is the fraction of the recovered injections by a given algorithm. The standard cWB uses η_c , and cWB plus GMM uses T as a detection statistic. We show that the cWB plus GMM algorithm significantly enhances the detection efficiency for all the 20 waveforms compared to the standard cWB.

In order to understand this improvement, we show the two-dimensional distribution of attributes c_{c0} , N_{norm} , $Q_{\text{veto}0}$ and $Q_{\text{veto}1}$ of the signal triggers with $\text{iFAR} \geq 100$ years in cWB plus GMM (by purple color) and standard cWB (by gray color) analysis in Fig. 2. From panels (a) and (c), it is evident that with GMM, we have improved the signal/noise classification ability that we can now recover events below c_{c0} of 0.8.

The most significant improvement after using cWB plus GMM is shown in the recovery of GA morphology (see Fig. 1(a)). The GA signal mostly resembles blip glitches and lies mainly in bin LF1. During the O3 run, the search sensitivity of standard cWB to the GA waveforms has worsened compared to its sensitivity O2. This is primarily due to an abundance of blip glitches that resemble GA signals. However, with GMM based postproduction approach appears to have modeled them well, which is also evident in the recovery of GA with $Q_{\text{veto}1} < 3$ (in Fig 2 panels (b) and (d)) those otherwise would get classified as bin LF1 events in the standard cWB. In standard cWB analysis, detection efficiency for triggers falls in each bin estimates separately. However, the signal waveforms in bin LF1 are mostly buried in background glitches. As a result, cWB plus GMM has shown an enhanced sensitivity for Gaussian pulses with an improvement of greater than 20% at an iFAR of 100 years. The signal waveforms lying in bin LF1 show comparatively lower significance to other morphologies in both cWB plus GMM and standard cWB method at the given signal strength because of excess glitches.

The cWB plus GMM approach marginally improves the SG class of injections (Fig. 1(b)-(e)). This indicates that our approach can recover other burst mor-

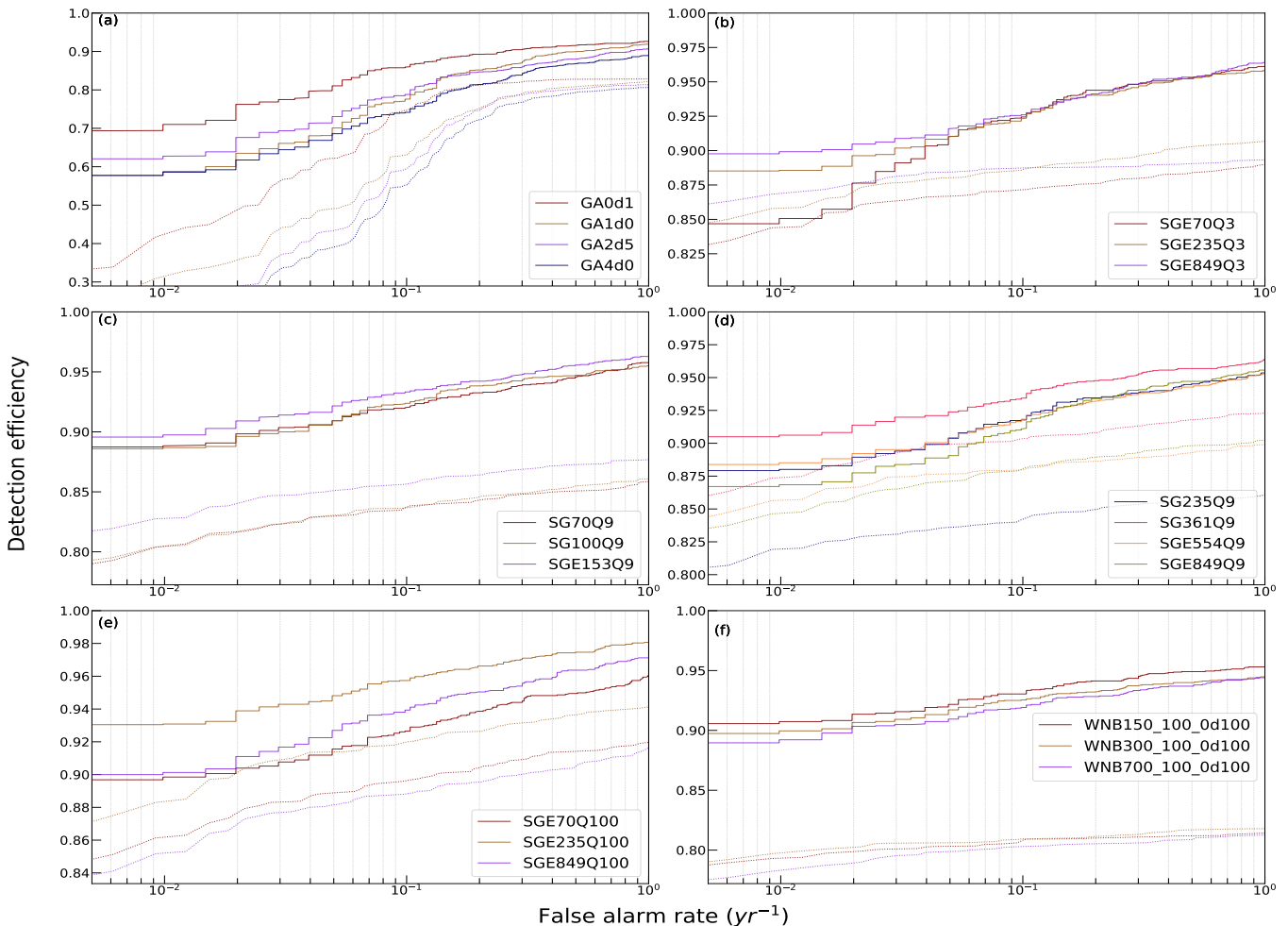


FIG. 1: ROC curves for the cWB plus GMM search (solid lines) compared to the standard cWB search (dotted lines) for the simulated waveforms. Panel (a) shows GA class waveforms with duration 0.1, 1, 2.5 and, 4 s. (b) shows elliptically polarized SG waveforms with a quality factor $Q = 3$ and frequency at 70, 235, and 849 Hz. Panel (c) and (d) show the SG waveforms with $Q = 9$ and 70-900 Hz frequency range. (e) shows SG with $Q = 100$ and frequency at 70, 235, and 849 Hz. (f) WNB is characterized by f_{low} at 150, 300, and 750 Hz with frequency bandwidth and duration at 100 and 0.1 s.

phologies considerably better than standard cWB. For WNB, the cWB plus GMM approach consistently improves the search sensitivity with respect to standard cWB (Fig. 1(f)). WNB falls in the bin LF3, the cleanest part of the background distribution and cWB plus GMM has shown improvement of around 10% throughout the iFAR threshold.

It should also be noted that categorizing the events into various bins in standard cWB analysis iFAR is penalized by a trials factor of three. Whereas in cWB plus GMM based analysis, there is no such bin classification. This can also be attributed to some improvement in detection efficiency.

E. Robustness test to CCSN

To test the robustness against the different morphologies of waveforms and distribution, we applied the GMM model on the core-collapse supernovae (CCSN) injections used in [20] that are not included in the training data set. Fig. 3 shows the ROC curves for injections with different CCSN waveforms. The models 35OC and m39 describe the explosion driven by rapid rotation of massive progenitor star with zero age main sequence (ZAMS) mass of $35M_{\odot}$ and $39M_{\odot}$, which creates strong GW emission [20]. Both standard cWB and cWB plus GMM show comparable high efficiency to CCSN model 35OC at low FAR because GW signal occurs frequency above 100 Hz, which is optimally sensitive to the cWB low-frequency search method. However, the CCSN model m39, cWB plus GMM enhances the detection efficiency compared

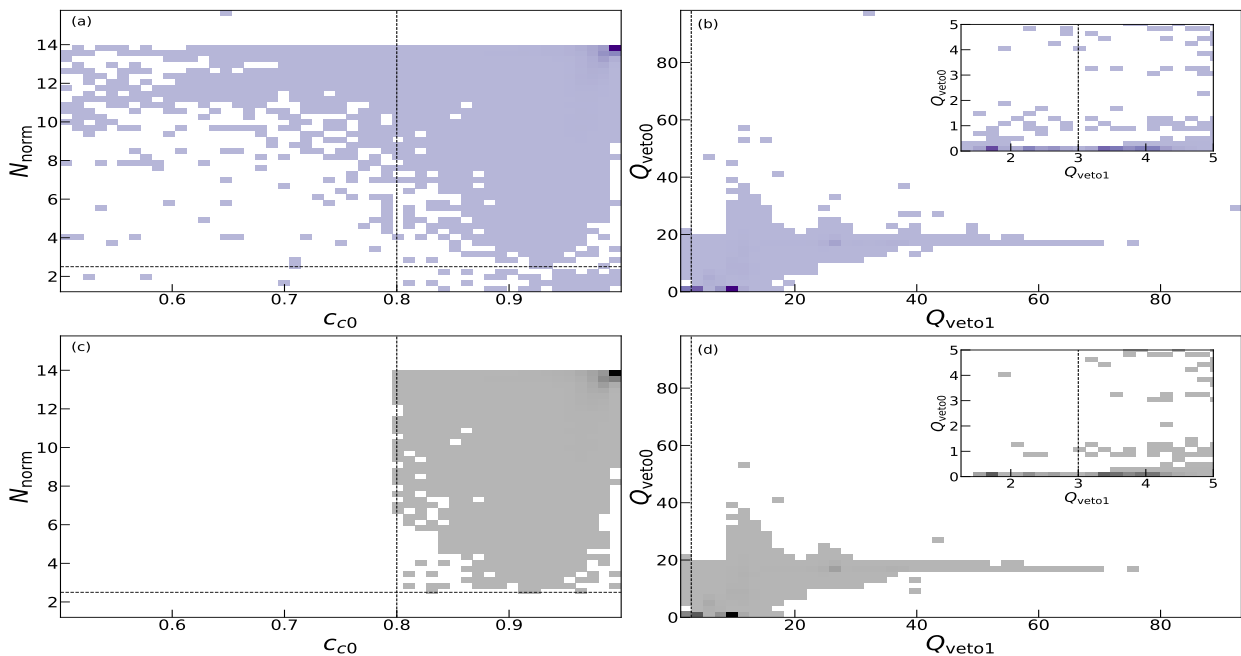


FIG. 2: The two-dimensional distribution of signal trigger attributes c_{c0} vs N_{norm} ((a) and (c)) and Q_{veto1} vs Q_{veto0} ((b) and (d)) for triggers recovered by cWB plus GMM (purple) and standard cWB (gray) at $\text{iFAR} \geq 100$ years.

The inset shows a zoom-in of the 2D distribution of Q_{veto1} vs Q_{veto0} . The black dashed lines show the postproduction threshold applied to these attributes for the standard cWB all-sky search, which classify the events into various bins as shown in Table II.

to the standard cWB. Since GW amplitude peaks at a frequency ~ 750 Hz, at the edge of the postproduction threshold applied for standard cWB. The standard cWB and cWB plus GMM give comparatively low efficiency for all the other models such as m20, s18 and s9, which describes GW emission from nonrotating progenitors with ZAMS mass ranges from $9M_{\odot}$ to $20M_{\odot}$. Among them, cWB plus GMM performs better than standard cWB at all FAR threshold values. The ML based cWM plus GMM method does not focus on one class but shows good improvement for a broad class of waveforms at detectable iFAR levels.

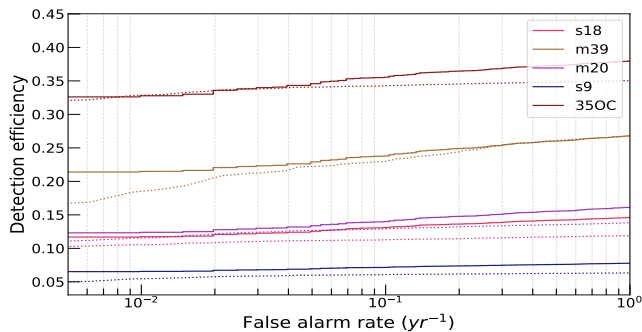


FIG. 3: ROC curves for CCSN waveforms. As before, the cWB plus GMM search ROC curves are represented by solid lines, while the standard cWB search ROC curves are plotted as dotted lines.

IV. SEARCH RESULTS

Following the sensitivity study, we applied the cWB plus GMM on the O3a data from LIGO Hanford and LIGO Livingston. The maximum background for cWB plus GMM analyses is 200 years since we used 20% of the total O3a background as test data. The results of this search from the O3a low-frequency region are shown in Fig. 4. We plot the cumulative number of recovered events by standard cWB (by green color) and cWB plus GMM (by purple color) in terms of their iFAR values.

The cWB plus GMM method recovers all CBC events identified by the targeted cWB search for CBC in GWTC-2 and GWTC-2.1 [2, 3, 55]. No new candidates were found with high statistical significance in this search. The most significant non-CBC event observed at UTC 2019-09-30 23:46:52 with an iFAR of 0.3 years in the cWB plus GMM method shows 0.01 years of iFAR in the standard cWB search. The second most significant event at UTC 2019-05-11 04:12:15 was with an iFAR of 0.15 years in cWB plus GMM, which had shown iFAR of 0.002 years in standard cWB. We note that the two loudest non-CBC events in standard cWB all-sky search are at UTC 2019-09-28 02:11:45 and UTC 2019-08-04 08:35:43, with an iFAR of 0.53 years, and 0.19 years [20], showed less significance in the cWB plus GMM search with an iFAR of 0.006 and 0.05, respectively. After excluding the CBC signals, the distribution of coincident events is close to the predicted background for the analyzed time.

Table IV lists the CBC events from GWTC-2 and GWTC-2.1 that are also observed by standard cWB all-sky burst search [20]. We list their attribute values, T statistic value, total source mass, and the corresponding iFAR values for both standard all-sky burst cWB search [20] and cWB plus GMM search³. We observe that the cWB plus GMM method is less sensitive to low mass BBH systems (with total source mass $< 60 M_{\odot}$) since these events are more significant in the standard cWB all-sky search. The exception is GW190412, which has significantly asymmetric component masses and shows the contribution from higher modes⁴ [56]. The IMBH system GW190521 [6] of total mass $163M_{\odot}$ was observed with high significance in cWB plus GMM with an iFAR of greater than 200 years, which shows an iFAR for 65 years in all-sky burst cWB search [20]⁵. A similar trend is observed for other events in Table IV, with total source mass $M > 100M_{\odot}$. The CBC event count acts as a verification step just to review the sensitivity of cWB plus GMM search against standard cWB all-sky search.

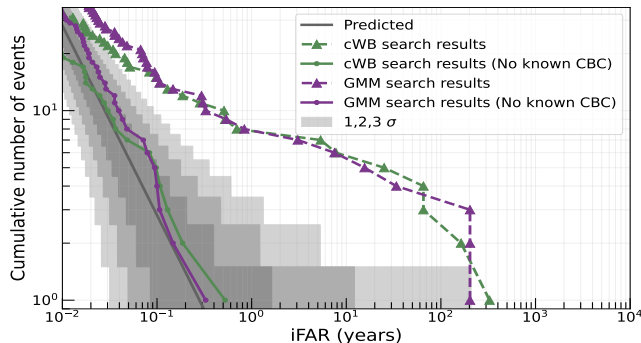


FIG. 4: The cWB plus GMM (purple) and standard cWB (green) observed the cumulative number of events versus iFAR in the low-frequency range. The iFAR distribution for all coincident data is plotted using triangular markers, while the line with circular markers shows the distribution after the removal of known CBC events. The expected mean value of background shows as a solid gray line with 1, 2 and 3 σ Poisson uncertainty (shaded regions).

³ The significance values for standard cWB shown in this table are different from those listed in GWTC-2 and GWTC-2.1 because the significance estimates in these publications are based on a version of cWB optimized for CBC events. Since the cWB plus GMM approach is applied in the all-sky search context, we find it fairer to compare the event significance obtained for all-sky searches.

⁴ The component masses for GW190412 system are $m_1 = 30.1^{+4.6}_{-5.3} M_{\odot}$ and $m_2 = 8.3^{+1.6}_{-0.9} M_{\odot}$.

⁵ GW190521 reported with an iFAR greater than 4900 years using targeted CBC searches [6].

V. DISCUSSIONS

This paper provides a robust method of direct usage of a supervised machine learning approach at the post-production stage of a well-established model-independent cWB detection algorithm. The work employs a Gaussian mixture modeling approach to model cWB triggers (signal as well as noise) in a multidimensional attribute set. The GMM modeling helps to classify background glitches from the transient GW signals in a log likelihood-based test statistic. While this approach was initially conceptualized in [1] here, we extend it further in conditioning the attribute set by first taking an appropriate subset targeted for a given signal class. Then introduce appropriate reparametrization of the attributes ensuring the well-behaved distributions for the construction of GMM.

We use the cWB triggers set from the all-sky short-duration burst search during O3a coincident data from the two Advanced LIGO detectors. We consider the signals from generic morphology and specific CCSN models used in [20] for sensitivity study. For generic morphology, there is a definite improvement in sensitivity for a fixed value of false alarm rate for all waveforms using the cWB plus GMM method over the standard cWB. We observed notable improvement for the Gaussian pulse, which falls in the bin LF1 containing a very short and loud (blip-type) population of glitches. Sensitivity improvement to CCSN models clearly demonstrates the robustness of the cWB plus GMM method against the variation in morphology as well as distribution of waveforms as CCSN waveforms were not part of the training set.

The application of cWB plus GMM on the coincident data search is consistent with the all-sky short-duration burst results using the standard cWB search as published in [20]. No additional, significant events are observed. The search recovered the GWs signals from the BBH merger as found by the targeted search for CBC signals. We observed higher significance for massive CBC events compared to that reported in the all-sky burst [20]. This clearly shows the promise of the cWB plus GMM approach at the postproduction stage of the burst algorithm as it gives a better handle in distinguishing the short-duration bursts from that of the short-duration noisy glitches. We plan to further extend this work to Generalized mixture modeling [57], which allows us to model more complex distributions for the fourth and fifth observing runs of ground-based GW detectors [58].

ACKNOWLEDGMENTS

The authors thank Shubhanshu Tiwari for valuable discussion and suggestions. D.L. acknowledges support from Swiss National Science Foundation (SNSF) grant number 200020-182047. A.P. acknowledges the SERB Matrics grant MTR/2019/001096 and SERB-Power-fellowship grant SPF/2021/000036 of Department of Science and Technology, India for support. The authors are thankful

Event	η_c	Q_{veto}	c_{c0}	χ^2	N_{norm}	T	$M(M_\odot)$	iFAR in years	
								cWB+GMM	cWB
GW190408_181802	8.59	0.92	0.96	0.13	5.09	-0.41	$43.0^{+4.2}_{-3.0}$	0.30	25.14
GW190412	11.69	4.16	0.95	0.06	5.4	13.21	$38.4^{+3.8}_{-3.7}$	15.62	14.86
GW190421_213856	6.46	0.31	0.97	-0.07	4.41	-0.38	$72.9^{+13.4}_{-9.2}$	0.30	0.04
GW190426_190642	5.52	0.45	0.88	0.08	4.07	-4.85	$184.4^{+41.7}_{-36.6}$	0.02	0.01
GW190503_185404	7.34	0.34	0.93	-0.02	4.76	1.65	$71.7^{+9.4}_{-8.3}$	0.84	0.70
GW190513_205428	7.05	1.67	0.86	0.15	3.77	-2.99	$53.9^{+8.6}_{-5.9}$	0.07	0.28
GW190517_055101	6.08	0.19	0.88	-0.15	3.05	-2.79	$63.5^{+9.6}_{-9.6}$	0.08	0.01
GW190519_153544	10.13	0.53	0.89	0.01	7.63	18.04	$106.6^{+13.5}_{-14.8}$	33.83	7.78
GW190521	9.24	0.60	0.92	-0.16	10.53	32.45	$163.9^{+39.2}_{-23.5}$	> 200	65.38
GW190521_074359	14.19	0.56	0.96	-0.08	8.44	72.77	$74.7^{+7.0}_{-4.8}$	> 200	326.88
GW190602_175927	7.25	0.43	0.95	-0.13	6.5	0.73	$116.3^{+19.0}_{-15.6}$	0.54	0.51
GW190706_222641	9.29	0.79	0.83	-0.10	7.36	24.93	$104.1^{+20.2}_{-13.9}$	> 200	65.38
GW190727_060333	5.86	0.35	0.96	0.17	4.96	-2.94	$67.1^{+11.7}_{-8.0}$	0.07	0.006
GW190728_064510	6.50	3.94	0.87	-0.13	2.55	-4.93	$20.6^{+4.5}_{-1.3}$	0.02	0.051
GW190828_063405	10.27	0.84	0.82	0.10	5.01	8.78	$58.0^{+7.7}_{-4.8}$	7.52	163.44
GW190915_235702	8.07	0.42	0.95	0.06	4.29	5.29	$59.9^{+7.5}_{-6.4}$	3.07	5.36
GW190929_012149	5.97	0.22	0.85	0.103	3.44	-6.20	$104.3^{+34.9}_{-25.2}$	0.01	0.009

TABLE IV: Table of CBC candidate events identified by cWB in GWTC-2 and GWTC-2.1. The columns show the effective correlated signal-to-noise η_c , energy distribution of the event over different time segments $Q_{\text{veto}0}$, network correlation coefficients C_{c0} , residual noise energy measure χ^2 , the ratio between the reconstructed energy and the total energy N_{norm} and source total mass M reported in GWTC-2 [2] and GWTC-2.1 [3]. The two rightmost columns report event significances for the cWB plus GMM search and the standard cWB all-sky search.

to LIGO-Virgo-Kagra collaboration for provisions of the cWB attribute data used for the all-sky short-duration O3a analysis. The authors are grateful for the computational resources and data provided by the LIGO Laboratory and supported by National Science Foundation Grants No.PHY-0757058 and No.PHY-0823459. C.M. and I.S.H. are supported by the Science and Technology Research Council (grant No. ST/V005634/1) and the European Cooperation in Science and Technology (COST) action CA17137. The open data is available in the Gravitational Wave Open Science Center [59], a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. The authors also acknowledge the use of the LDG clusters for computational/numerical work.

APPENDIX: REPARAMETRIZATION OF ATTRIBUTES

Given the cWB trigger attributes E_c , η_c , c_{c0} , c_{c2} , N_{ED} , N_{norm} , χ^2 , $Q_{\text{veto}0}$, $Q_{\text{veto}1}$, $L_{\text{veto}0}$, $L_{\text{veto}1}$, and $L_{\text{veto}2}$. We reparametrize some of the attributes which do not follow a well behaved Gaussian distribution. In Table V, we show the reparametrized attributes along with their original form. We further show the reduction in the number of Gaussians after the reparameterization of the attributes. The distribution of reparametrized cWB trigger attributes for the signal initially chosen for GMM analysis and after the reparameterization is shown in Fig. 5.

TABLE V: Details of reparametrization.

Original attribute set	Reparametrized attribute set
η_c	$\log_{10}(\eta_c)$
c_{c0}	$\text{logit}(c_{c0})$
c_{c2}	$\text{logit}(c_{c2})$
N_{ED}	$\log_{10}(N_{ED} + 10^3)$
E_c	$\log_{10}(E_c)$
N_{norm}	N_{norm}
χ^2	χ^2
$Q_{\text{veto}0}$	$\log_{10}(Q_{\text{veto}0} + 1)$
$Q_{\text{veto}1}$	$\log_{10}(Q_{\text{veto}1})$
L_{ratio}	$\text{logit}(L_{\text{ratio}})$
$L_{\text{veto}2}$	$\text{logit}(L_{\text{veto}2} \times 0.99)$

Number of optimum Gaussians in GMM		
Model	Original attribute set	Reparametrized attribute set
Signal	113	90
Noise	115	82

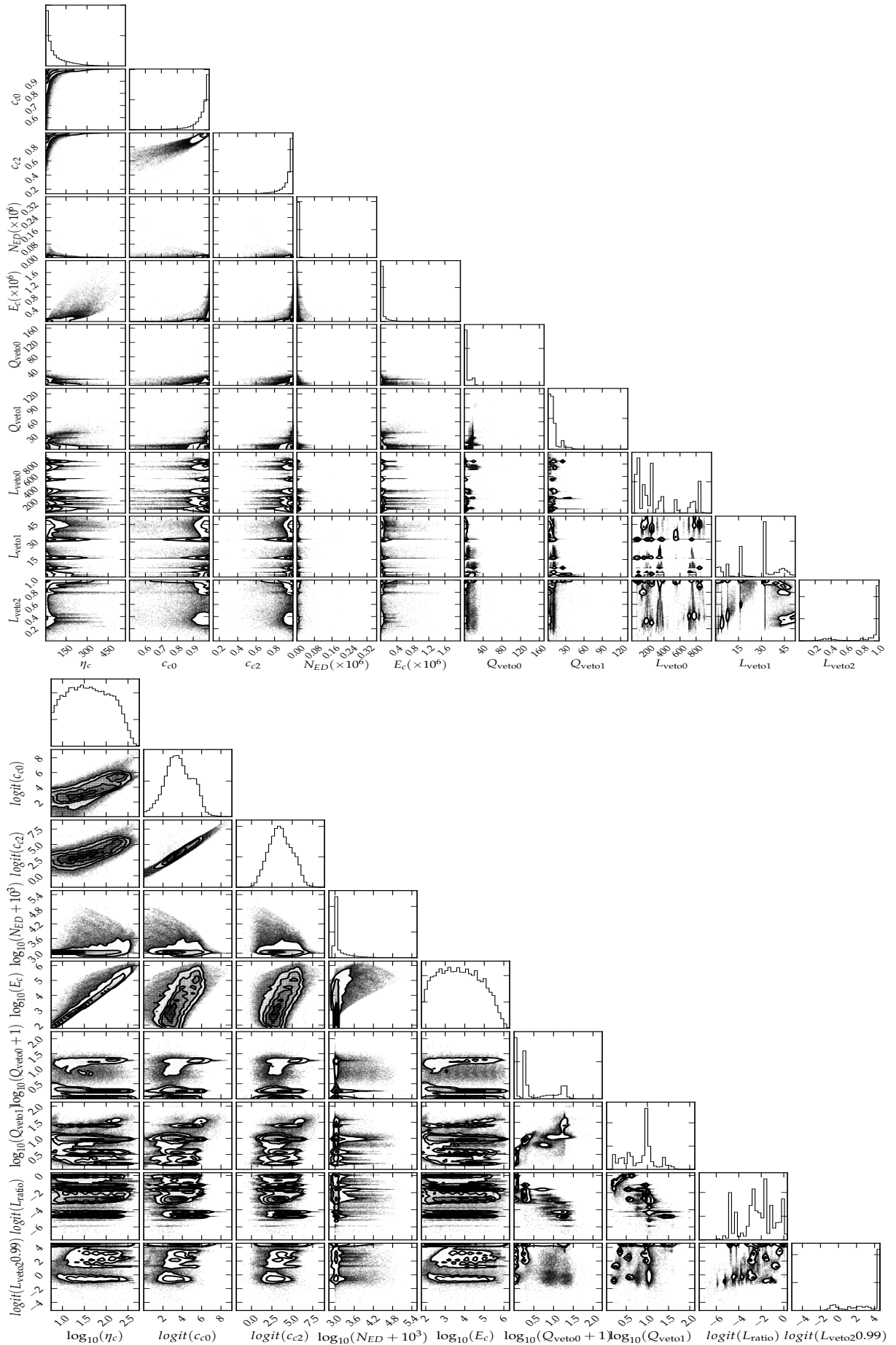


FIG. 5: The corner plot showing the one and two dimensional distribution of the reparametrized cWB trigger attribute for the signals before (top) and after (bottom) the reparameterization.

- [1] V. Gayathri, D. Lopez, P. R. S., I. S. Heng, A. Pai, and C. Messenger, *Phys. Rev. D* **102**, 104023 (2020).
- [2] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. X* **11**, 021053 (2021).
- [3] R. Abbott *et al.* (LIGO Scientific, VIRGO), (2021), [arXiv:2108.01045 \[gr-qc\]](https://arxiv.org/abs/2108.01045).
- [4] J. Aasi *et al.* (LIGO Scientific), *Class. Quant. Grav.* **32**, 074001 (2015), [arXiv:1411.4547 \[gr-qc\]](https://arxiv.org/abs/1411.4547).
- [5] F. Acernese *et al.* (Virgo Collaboration), *Class. Quant. Grav.* **32**, 024001 (2015).
- [6] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. Lett.* **125**, 101102 (2020).
- [7] R. Abbott *et al.* (LIGO Scientific, Virgo), *Astrophys. J. Lett.* **896**, L44 (2020), [arXiv:2006.12611 \[astro-ph.HE\]](https://arxiv.org/abs/2006.12611).
- [8] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Astrophys. J. Lett.* **892**, L3 (2020), [arXiv:2001.01761 \[astro-ph.HE\]](https://arxiv.org/abs/2001.01761).
- [9] R. Abbott *et al.* (LIGO Scientific, KAGRA, VIRGO), *Astrophys. J. Lett.* **915**, L5 (2021), [arXiv:2106.15163 \[astro-ph.HE\]](https://arxiv.org/abs/2106.15163).
- [10] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), [arXiv:2111.03606 \[gr-qc\]](https://arxiv.org/abs/2111.03606).
- [11] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. D* **101**, 084002 (2020), [arXiv:1908.03584 \[astro-ph.HE\]](https://arxiv.org/abs/1908.03584).
- [12] R. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA), *Phys. Rev. Lett.* **126**, 241102 (2021), [arXiv:2101.12248 \[gr-qc\]](https://arxiv.org/abs/2101.12248).
- [13] M. Ebersold and S. Tiwari, *Phys. Rev. D* **101**, 104041 (2020), [arXiv:2005.03306 \[gr-qc\]](https://arxiv.org/abs/2005.03306).
- [14] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Astrophys. J.* **874**, 163 (2019), [arXiv:1902.01557 \[astro-ph.HE\]](https://arxiv.org/abs/1902.01557).
- [15] J. Abadie *et al.* (LIGO Scientific), *Phys. Rev. D* **83**, 042001 (2011), [arXiv:1011.1357 \[gr-qc\]](https://arxiv.org/abs/1011.1357).
- [16] J. Abadie *et al.* (The LIGO Scientific Collaboration and The Virgo Collaboration), *Phys. Rev. D* **81**, 102001 (2010).
- [17] J. Abadie *et al.* (The LIGO Scientific Collaboration and The Virgo Collaboration), *Phys. Rev. D* **85**, 122007 (2012).
- [18] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. D* **95**, 042003 (2017).
- [19] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. D* **100**, 024017 (2019).
- [20] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), *Phys. Rev. D* **104**, 122004 (2021), [arXiv:2107.03701 \[gr-qc\]](https://arxiv.org/abs/2107.03701).
- [21] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, *Class. Quant. Grav.* **25**, 114029 (2008), [arXiv:0802.3232 \[gr-qc\]](https://arxiv.org/abs/0802.3232).
- [22] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. D* **93**, 122004 (2016).
- [23] M. Szczepańczyk, S. Klimenko, B. O'Brien, I. Bartos, V. Gayathri, G. Mitselmakher, G. Prodi, G. Vedovato, C. Lazzaro, E. Milotti, F. Salemi, M. Drago, and S. Tiwari, *Phys. Rev. D* **103**, 082002 (2021).
- [24] J. Powell, D. Trifirò, E. Cuoco, I. S. Heng, and M. Cavaglia, *Class. Quant. Grav.* **32**, 215012 (2015), [arXiv:1505.01299 \[astro-ph.IM\]](https://arxiv.org/abs/1505.01299).
- [25] J. Powell, A. Torres-Forné, R. Lynch, D. Trifirò, E. Cuoco, M. Cavaglia, I. S. Heng, and J. A. Font, *Class. Quant. Grav.* **34**, 034002 (2017), [arXiv:1609.06262 \[astro-ph.IM\]](https://arxiv.org/abs/1609.06262).
- [26] N. Mukund, S. Abraham, S. Kandhasamy, S. Mitra, and N. S. Philip, *Phys. Rev. D* **95**, 104059 (2017).
- [27] D. George, H. Shen, and E. A. Huerta, *Phys. Rev. D* **97**, 101501 (2018).
- [28] R. Biswas *et al.*, *Phys. Rev. D* **88**, 062003 (2013).
- [29] S. Rampone, V. Pierro, L. Troiano, and I. M. Pinto, *Int. J. Mod. Phys. C* **24**, 1350084 (2013), [arXiv:1401.5941 \[astro-ph.IM\]](https://arxiv.org/abs/1401.5941).
- [30] M. Lightman, J. Thurakal, J. Dwyer, R. Grossman, P. Kalmus, L. Matone, J. Rollins, S. Zairis, and S. Marka, *J. Phys. Conf. Ser.* **32**, 58 (2006).
- [31] M. Razzano and E. Cuoco, *Class. Quant. Grav.* **35**, 095016 (2018), [arXiv:1803.09933 \[gr-qc\]](https://arxiv.org/abs/1803.09933).
- [32] E. A. Huerta *et al.*, *Nature Astron.* **5**, 1062 (2021), [arXiv:2012.08545 \[gr-qc\]](https://arxiv.org/abs/2012.08545).
- [33] M. Cavaglia, S. Gaudio, T. Hansen, K. Staats, M. Szczepańczyk, and M. Zanolin, *Mach. Learn. Sci. Tech.* **1**, 015005 (2020), [arXiv:2002.04591 \[astro-ph.IM\]](https://arxiv.org/abs/2002.04591).
- [34] S. Vinciguerra, M. Drago, G. A. Prodi, S. Klimenko, C. Lazzaro, V. Necula, F. Salemi, V. Tiwari, M. C. Tringali, and G. Vedovato, *Class. Quant. Grav.* **34**, 094003 (2017), [arXiv:1702.03208 \[astro-ph.IM\]](https://arxiv.org/abs/1702.03208).
- [35] M. Cavaglia, S. Gaudio, T. Hansen, K. Staats, M. Szczepańczyk, and M. Zanolin, *Mach. Learn. Sci. Tech.* **1**, 015005 (2020), [arXiv:2002.04591 \[astro-ph.IM\]](https://arxiv.org/abs/2002.04591).
- [36] T. Mishra, B. O'Brien, V. Gayathri, M. Szczepańczyk, S. Bhaumik, I. Bartos, and S. Klimenko, *Phys. Rev. D* **104**, 023014 (2021), [arXiv:2105.04739 \[gr-qc\]](https://arxiv.org/abs/2105.04739).
- [37] E. Cuoco *et al.*, *Mach. Learn. Sci. Tech.* **2**, 011002 (2021), [arXiv:2005.03745 \[astro-ph.HE\]](https://arxiv.org/abs/2005.03745).
- [38] A. Iess, E. Cuoco, F. Morawski, and J. Powell, (2020), [arXiv:2001.00279 \[gr-qc\]](https://arxiv.org/abs/2001.00279).
- [39] V. Necula, S. Klimenko, and G. Mitselmakher, *J. Phys. Conf. Ser.* **363**, 012032 (2012).
- [40] S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. F. Da Silva, and G. Mitselmakher, *Phys. Rev. D* **93**, 042004 (2016).
- [41] M. D. *et al.*, *SoftwareX* **14**, 100678 (2021).
- [42] <https://gwburst.gitlab.io>.
- [43] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
- [44] G. Schwarz, *The Annals of Statistics* **6**, 461 (1978).
- [45] H. Bhat and N. Kumar, (2010).
- [46] LIGO Scientific Collaboration And Virgo Collaboration, "LIGO Virgo strain data from observing run O3a," <https://www.gw-openscience.org/03/03a> (2021).
- [47] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *SoftwareX* **13**, 100658 (2021).
- [48] M. Cabero *et al.*, *Class. Quant. Grav.* **36**, 15 (2019), [arXiv:1901.05093 \[physics.ins-det\]](https://arxiv.org/abs/1901.05093).
- [49] P. J. Sutton, (2013), [arXiv:1304.0210 \[gr-qc\]](https://arxiv.org/abs/1304.0210).
- [50] J. Powell and B. Müller, *Mon. Not. Roy. Astron. Soc.* **487**, 1178 (2019), [arXiv:1812.05738 \[astro-ph.HE\]](https://arxiv.org/abs/1812.05738).
- [51] E. P. O'Connor and S. M. Couch, *Astrophys. J.* **865**, 81 (2018), [arXiv:1807.07579 \[astro-ph.HE\]](https://arxiv.org/abs/1807.07579).
- [52] D. Radice, V. Morozova, A. Burrows, D. Vartanyan, and H. Nagakura, *Astrophys. J. Lett.* **876**, L9 (2019), [arXiv:1812.07703 \[astro-ph.HE\]](https://arxiv.org/abs/1812.07703).
- [53] J. Powell and B. Müller, *Mon. Not. Roy. Astron. Soc.* **494**, 4665 (2020), [arXiv:2002.10115 \[astro-ph.HE\]](https://arxiv.org/abs/2002.10115).

- [54] M. Obergaulinger and M. A. Aloy, *Mon. Not. Roy. Astron. Soc.* **492**, 4613 (2020), [arXiv:1909.01105 \[astro-ph.HE\]](#).
- [55] A. H. Nitz, C. D. Capano, S. Kumar, Y.-F. Wang, S. Kasta, M. Schäfer, R. Dhurkunde, and M. Cabero, (2021), [arXiv:2105.09151 \[astro-ph.HE\]](#).
- [56] R. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. D* **102**, 043015 (2020), [arXiv:2004.08342 \[astro-ph.HE\]](#).
- [57] O. M. M. Mohamed and M. Jaïdane-Saïdane, in *2009 17th European Signal Processing Conference* (2009) pp. 2273–2277.
- [58] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, Virgo, VIRGO), *Living Rev. Rel.* **21**, 3 (2018), [arXiv:1304.0670 \[gr-qc\]](#).
- [59] <https://www.gw-openscience.org/>.