Classical and Quantum Gravity

PURPOSE-LED PUBLISHING™

PAPER • **OPEN ACCESS**

Optimizing searches for gravitational wave bursts using coherent WaveBurst-2G

To cite this article: A Martini *et al* 2026 *Class. Quantum Grav.* **43** 055016

View the article online for updates and enhancements.

You may also like

- Beyond GWTC-3: analyzing and verifying new gravitational-wave events from community catalogues
Daniel Williams

- The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run
F Aubin, F Brighenti, R Chierici et al.

- *SWIFT* FOLLOW-UP OBSERVATIONS OF CANDIDATE GRAVITATIONAL-WAVE TRANSIENT EVENTS
P. A. Evans, J. K. Fridriksson, N. Gehrels et al.

This content was downloaded from IP address 193.205.210.73 on 27/05/2026 at 14:49

Classical and Quantum Gravity

Check for updates

OPEN ACCESS

RECEIVED
11 November 2025

REVISED
30 January 2026

ACCEPTED FOR PUBLICATION
17 February 2026

PUBLISHED
16 March 2026

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

Optimizing searches for gravitational wave bursts using coherent WaveBurst-2G

A Martini[1,2,*], A Miani[1,2], M Drago[3,4], C Lazzaro[5,6], F Salemi[3,4], S Bini[7], O Freitas[8,9], E Milotti[10,11], G Principe[10,11], S Tiwari[12], A Trovato[10,11], G Vedovato[13], Y Xu[14] and G A Prodi[1,2]

[1] Department of Physics, University of Trento, Trento, Italy
[2] INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, Trento, Italy
[3] Department of Physics, Sapienza University of Rome, Rome, Italy
[4] INFN, Sezione di Roma, Rome, Italy
[5] Department of Physics, University of Cagliari, Cagliari, Italy
[6] INFN, Sezione di Cagliari, Cagliari, Italy
[7] LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, United States of America
[8] Department of Astronomy and Astrophysics, University of Valencia, Valencia, Spain
[9] CF-UM-UP, University of Minho, Braga, Portugal
[10] Department of Physics, University of Trieste, Trieste, Italy
[11] INFN, Sezione di Trieste, Trieste, Italy
[12] Department of Physics, University of Zurich, Zurich, Switzerland
[13] INFN, Sezione di Padova, Padova, Italy
[14] IAC3, Universitat de les Illes Balears, Palma, Spain
[*] Author to whom any correspondence should be addressed.

E-mail: alessandro.martini-1@unitn.it

Keywords: gravitational waves astronomy, gravitational wave transients, minimally modeled detection methods, coherent data analysis methods, statistical significance assessment

Abstract

The most general searches for gravitational wave transients (GWTs) rely on data analysis methods that do not assume prior knowledge of the signal's waveform, direction, and arrival time on Earth. These searches provide data-driven signal reconstructions that are crucial both for testing available GW emission models and for discovering yet-to-be-unveiled sources. Here, we discuss the progresses in detection performances of the coherent WaveBurst second generation pipeline (cWB-2G), which is highly adaptable to minimally–modeled as well as model–informed searches for GWTs. Several search configurations for GWTs are examined using approximately 14.8 d of observation time from the third observing run by LIGO-Virgo-KAGRA (LVK). Recent enhancements include a ranking statistic fully based on multivariate classification with eXtreme gradient boosting, a thorough validation of the accuracy of the statistical significance of GWT candidates, and the measurement of the correlations of false alarms and simulated detections between different concurrent searches. For the first time, we provide a comprehensive comparison of cWB-2G performances on data from networks made of two and three detectors, and we demonstrate the advantage of combining concurrent searches for GWTs of generic morphology in a global observatory. This work offers essential insights for assessing our data analysis strategies in the ongoing and future LVK searches for generic GWTs.

1. Introduction

The observatories of the LIGO-Virgo-KAGRA (LVK) collaboration [1–3] completed the fourth observing run (O4) on 18 November 2025, with a catalog of more than two hundred gravitational wave transient (GWT) detections already published [4, 5]. All detected sources so far belong to the compact binary coalescence (CBC) class, for which the theory of general relativity provides detailed models which are suitable for matched-filter searches [6] and for testing fundamental physics [7–9]. In the future, we expect to observe other sources of GWTs for which there are no models as precise as those for

© 2026 The Author(s). Published by IOP Publishing Ltd

CBCs, such as core–collapse supernovae (CCSN) [10], neutron star (NS) excitations [11, 12], nonlinear memory effects [13], and accretion disk instabilities [14]. In addition, in this pioneering phase of gravitational wave astronomy, we need to perform searches for GWTs from unexpected sources. For all these GWT sources, LVK employs approaches which are either fully or partially agnostic to signal morphologies. The broadest searches for GWTs—called 'all-sky burst searches' — are designed to detect generic signals from any direction in the sky and at all times. This is a highly challenging task, and the ability to distinguish genuine GWTs from noise outliers is paramount. These all-sky searches are then specialized to look for short–duration burst GWTs (up to a few seconds in the sensitive frequency band [15–18]), and long–duration burst GWTs (up to $\sim 10^3$ s [18, 19–21]).

This work reports recent developments and tests carried out on one flavor of the data analysis pipeline coherent WaveBurst (cWB) [22–24], that contributed to all LIGO-Virgo and LVK minimally–modeled all–sky searches for GWTs since the first GW detection GW150914 [25], and played a crucial role in the detection of GW190521 [26] and GW231123 [27]. cWB carries out both low–latency searches and more thorough analyses on archived data, and it is capable of covering the entire sensitive frequency band. Several versions of cWB are actively being developed, known as cWB second generation (cWB-2G [24, 28, 29]), cWB cross power (cWB-XP [30, 31]) and cWB with Gaussian Mixture Model post-processing (cWB-GMM [32–34]).

Here, we present the latest methodological advancements of cWB-2G [35], which are exploited in the analysis of the ongoing fourth LVK observing run. Several search configurations for GWTs are tested on data from the LVK observing run O3, the most recent publicly released data set including both LIGO [1] and Virgo [2] observations. In particular, we select 14.8 days of concurrent observation time by LIGO–Hanford (H), LIGO–Livingston (L) and Virgo (V). Recent enhancements with respect to what was previously reported in [29] include the deployment of a ranking statistic fully based on machine learning (ML) methods, a thorough validation of the accuracy of the statistical significance of GWT candidates, and the measurement of the correlations of false alarms and simulated detections between different concurrent searches. For the first time, thanks to our comprehensive comparison of search performances when using two and three detectors' data, we are able to gauge the advantages and drawbacks of combining concurrent searches of GWTs of generic morphology in a global observatory of non-aligned detectors.

The paper is organized as follows: section 2 provides an overview of the cWB-2G workflow and the key improvements introduced in preparation for O4, leading to the latest release of cWB-2G [35]. Section 3 describes the dataset used in this study and the configurations of the analyses, in particular the supervised training and testing procedures of the eXtreme gradient boosting (XGBoost) multivariate classifier [36, 37]. Section 4 presents a method for assessing the consistency of the empirical background model with a Poisson point process. This is crucial to check the statistics of false alarms provided by the machine-learning ranking. In section 5 we discuss the correlation between false alarms from the different searches investigated in this paper. Finally, in section 6 we compare and combine searches for generic GWTs exploiting two and three detectors and we contrast an example of model–informed search for CBC coalescences with respect to model–agnostic searches.

2. General overview of cWB-2G analysis scheme

cWB-2G [24, 28, 35, 38] is an analysis pipeline employed in searches of generic GWT signals, designed to operate with little or no prior knowledge of the waveform models. Figure 1 shows a flowchart of cWB-2G which highlights the main logic blocks. The analysis proceeds across four main stages:

- **pre-production**, where the configuration of the analysis is defined, as for, e.g. the selection of the detector network, frequency range, time–frequency (TF) resolution levels, analyzed times, optional sky direction constraints and for the management of the submission of the computing jobs. In addition, the analysis scheme can search for candidate GWTs in the original data streams (*on–source* analysis), or alternatively estimate accidental false alarms in time–shifted data streams (*background* analysis, see 3.1), or estimate the detection efficiency for simulated GWT signals injected into the data streams (*simulation* analysis, see 3.2).
- **production of candidate triggers** and their summary statistics. The first step is a data pre-conditioning procedure, which includes the mitigation of spectral lines by regression and the data whitening filter. This is followed by a preselection of trigger candidates in the time-frequency plane, primarily based on the search for coincident excess power. The identified triggers are then passed to the more computationally demanding stage of constrained likelihood maximization: in this step, the coherent response of

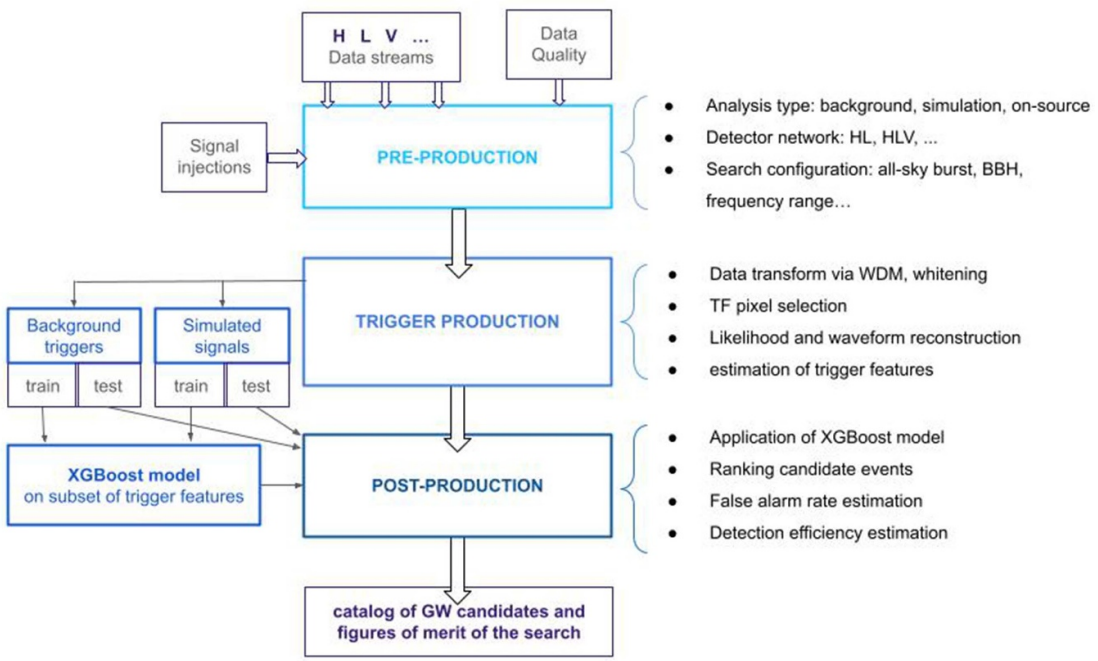


Figure 1. Workflow and data flow within cWB–2G.

two or more observatories is evaluated under the GWT signal hypothesis [22, 23]. The procedure further estimates the extrinsic and intrinsic properties of the candidate GWT, including its sky direction [39] and signal waveform [40], along with a set of features characterizing the selected trigger;

- **ranking of candidate triggers** by means of the XGBoost classifier, which has been previously trained on the trigger features provided by the background analysis and the simulation analysis, representing the accidental false alarms and the signal class, respectively.
- **reporting results**, that delivers the catalog of candidate GWTs, their reconstructed features, and several figures of merit of the analysis.

In cWB-2G, the input strain data are whitened and analyzed in the TF plane by the invertible orthonormal wavelet transform Wilson–Daubechies–Meyer (WDM) [41]. To preserve the sensitivity to diverse signal morphologies, cWB-2G utilizes concurrent WDM transforms at several different TF resolutions [41]. Once a trigger is identified as a potential GWT, cWB-2G reports several ad hoc statistics; we summarize here the most important ones. The network correlation coefficient, c_c, evaluates the coherence of a trigger in the network of detectors, and is defined as [24]

$$c_c = \frac{E_c}{E_c + E_n}, \tag{1}$$

where E_n is the residual noise energy and E_c is the coherent energy that quantifies the energy content of the GWT candidate. Both E_n and E_c are estimated in whitened data as square of the corresponding signal-to-noise ratio (SNR) over the network of detectors. The GWT detection statistic used through O3 LVK analyses is empirically defined as [42]

$$\eta_0 = \sqrt{\frac{E_c}{1 + \chi^2 \cdot (\max(1, \chi^2) - 1)}} \tag{2}$$

where $\chi^2 \equiv E_n/N_{df}$ is a proxy of the Chi-square statistics [38] and N_{df} is the number of independent wavelet amplitudes used to characterize the trigger.

In the current LVK observing run O4, cWB-2G exploits the XGBoost ML algorithm [36] to classify triggers according to their estimated authenticity, as explained in the following. The XGBoost algorithm is trained by learning false alarm and signal properties as described by a set of features, including η_0, c_c and several others described in appendix A. The specific novelty of this cWB-2G version [35] is that it ranks candidate triggers by relying solely on the output score provided by XGBoost, $W_{XGB} \in [0, 1]$.

For convenience, the newly adopted ranking statistic, η_0', is a monotone function obtained by stretching $W_{\rm XGB}$ in its higher score range to ease numerical resolution requirements at high candidate significance. Moreover, in all our analyses, the most significant false alarms show η_0' scores well below the upper bound. This new automatic ranking eliminates the complex manual tuning that has been intrinsic in previous cWB-2G versions enhanced with ML [29]. In fact, the XGBoost scores were previously used as a penalty factor multiplying the traditional test statistic η_0. In addition, to mitigate short–duration false alarm glitches, known as *blips* [43], that result was then fed in nonlinear penalization functions of a few morphological parameters of the candidate triggers. Such a complexity is not needed anymore with the current detection statistic η_0'.

This methodological advancement is calling for more extended tests and thorough validation procedures to exclude systematics, as they might emerge from e.g. ML overfitting or correlations in the data set representing false alarms. including the choices on the benchmarks of performaFurther details of the tuning of the XGBoost architecture, including the choices on the benchmarks of performances, are discussed in appendix A.1, while the training procedures used for different search types are discussed in section 3 with more details in appendix A.2. Section 4 and appendix B discuss new validation methods implemented concerning the accuracy of the assessment of the statistical significance of candidate GWTs.

3. Data set and configuration of the analyses

In this work, we compare different cWB-2G analyses of publicly available data from the second half of the LVK third observing run (O3b [44]) of advanced LIGO [45] and advanced Virgo detectors [46], more specifically data from 2020, January 6 (GPS 1262304000s) to February 14 (GPS 1265760000). Within this time range, there is a good coverage of coincident observation by the three detectors, corresponding to about 14.8 d of science quality data. We restrict our analyses to three detectors data, considering separately the two detector network made by LIGO-Hanford (H) and LIGO-Livingston (L) and the three detector network including Virgo (V). The duration of this data set is well-suited to the methodological goals of this work, and the associated computational load is not excessive.

Three different all-sky analyses are then developed by training different models of XGBoost post-processing:

- search of short duration GWTs on HL data (HL-burst);
- search of short duration GWTs on HLV data (HLV-burst);
- search of GWTs from binary black hole coalescences on HL data (HL-BBH);

All these cWB-2G analyses are performed in the frequency range $[16-2048]$ Hz, which is the same choice adopted in off-line searches for GWTs on O4 data, motivated by general considerations on the noise power spectral densities of the detectors and on the predicted spectral content of interesting sources such as CCSN, which may extend up to 2 kHz. The analyses employ the WDM transform [41] with seven TF resolutions with $\Delta t \times \Delta f$ in the range from $[256\text{Hz} \times 1.95\,\text{ms}]$ to $[4\text{Hz} \times 125\,\text{ms}]$. The sky directions are analyzed using segmentation Healpix level 5, corresponding to a resolution of solid angle $\sim$3.4 deg^2, which is fully sufficient for detection purposes.

The set of trigger features which are fed into XGBoost is search-dependent and are all estimated in whitened data (more details in appendix A). In the case of short burst searches, the input features include generic characteristics of triggers such as η_0, c_c as well as morphological information inspired to the characteristics of the dominant glitch populations, such as the similarity of a trigger to a single pulse and the effective number of signal cycles (see table A2). The latter features are especially useful to mitigate the effects of the so-called blip glitches [43, 46]. Information like estimated central frequency, duration and frequency bandwidth are not passed as features, with the aim to preserve the widest possible prior on the signal parameter space. The list of features is the same for the HL-burst and HLV-burst, but each feature is the result of the coherent reconstruction of the signal candidate in the different network of detectors.

In the search for BBH signals, the list of features of the short burst search is extended to include morphological characteristics that are model-specific to the target signal population, including estimates of chirp mass, similarity to chirping signal morphology, central frequency, frequency bandwidth, and duration (see table A3). The HL-BBH trigger list is identical to that of HL-burst; its specialization arises solely from the dedicated ranking process conducted using XGBoost.

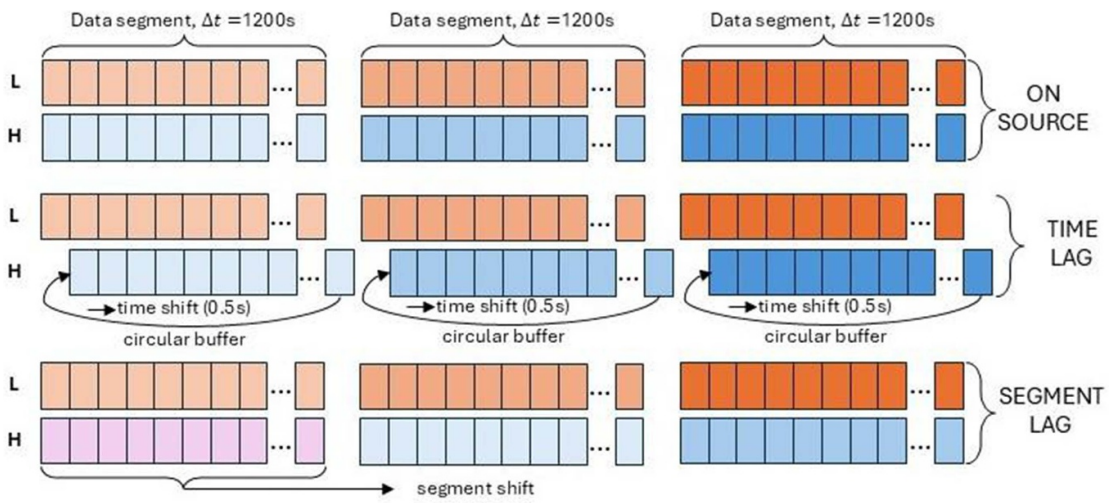


Figure 2. Implementation of time-slides in cWB-2G for two detectors: schematic view of three adjacent data segments. Top: actual data streams. Middle: time-lag treating each data segment as a circular buffer. Bottom: time shift by one data segment.

3.1. Background triggers for training and testing the XGBoost classifier

The background triggers used to train and test the XGBoost classifier are provided by running cWB-2G over time-slides. Each time-slide is implemented by shifting in time the detectors' data streams with a set of *unique* time shift values, meaning that they are selected avoiding any repetition of the same time difference between detector pairs. The live-times of different time-slides are typically varying by a few %, because the overlap of science quality data from different detectors changes with the applied time shifts.

For computational efficiency, cWB-2G analyzes the detectors' data streams in non-overlapping segments of 1200 s duration. Time-slides are implemented by combining two procedures: local time shifts within each data segment and segment–lags that shift data by multiples of the segments' duration, see figure 2. Local time shifts are implemented as multiples of 0.5 s and treat each data segment separately as a circular buffer. The 0.5 s step is enough to cancel the coherence of GWTs and noise outliers in the network. These local time-slides do not change the average time shift between detector pairs and provide up to 2400 instances of the noise[15]. Differently, the segment-lags procedure changes which data segments are overlapped and contributes a net average time shift between detector pairs which is a multiple of 1200 s.

In this work we use up to 29 segment-lags, corresponding to average time–shifts from -16800 s to $+16800$ s, for a total of up to 69 599 time-lags. This is suitable to investigate false alarm probabilities at $\sim$ a few tens per million and to check for possible systematics related to day-time activities at the detectors' sites[16]. By randomly choosing half of these time-lags, we obtain the set of background triggers used as noise model for training XGBoost. The complementary set is used for testing: it provides the reference distribution of the detection statistic for the null hypothesis. The statistical significance of a candidate event in terms of inverse false alarm rate (IFAR) is then estimated by its p–value multiplied by the total live-time of the testing set of time-lags.

Figure 3 shows the distribution of the background triggers reconstructed by cWB-2G on time-lags of HL whitened data in the duration vs frequency bandwidth plane. The minimum amplitude signal–to–noise ratio, SNR, which enables the reconstruction of a trigger is $\sim$5. Duration and bandwidth are estimated as standard deviations of the SNR^2 of each trigger in time and frequency. This set of a few millions background triggers is then fed into the XGBoost classifier, half for training and half for testing, to implement the HL–burst and HL–BBH searches. The distribution is dominated by short duration background triggers, close to the Heisenberg–Gabor bound on the time-frequency representation of triggers, i.e. product of standard deviations of time and frequency coordinates $\sigma_t \sigma_f \geqslant \frac{1}{4\pi}$. The prevalence of such background triggers is not surprising. In fact, the detection efficiency is typically higher for triggers with a compact time-frequency representation while the rejection of accidental coincidences is typically better for triggers showing more complex morphologies, thanks to the coherent analysis over the network of detectors. In addition, we know that short-duration noise outliers are very common in each

[15] Up to 2399 instances of noise triggers in case the segment-lag is zero.

[16] In previous cWB analyses of the LIGO observatory, the number of implemented time-lags has been beyond 1 million, see e.g. [47].

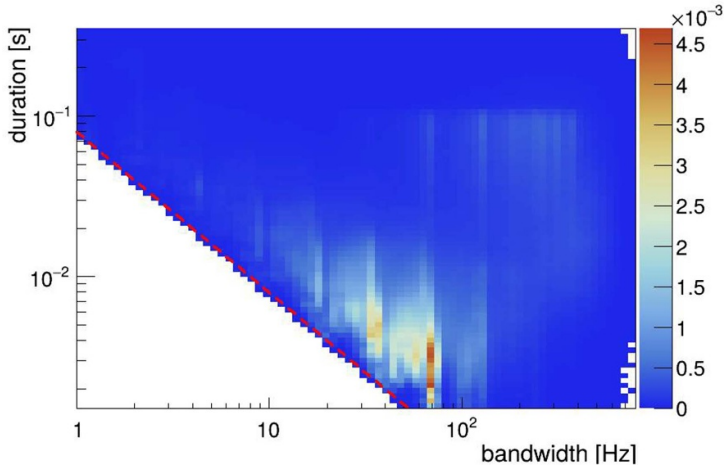


Figure 3. Distribution of duration and bandwidth of background triggers reconstructed by cWB-2G on time–lags of HL whitened data. Color scale shows fraction of counts per bin, that are uniform in log scale. The red dashed line shows the Heisenberg–Gabor limit for the time-frequency representation of triggers.

detector. A very similar distribution of background triggers is obtained by the analysis of HLV time-lags, and it is used for tuning the HLV–burst search. The distributions of the $\sim 10^5$ most significant background triggers according to the three different searches, as established by the respective XGBoost ranks, are discussed in section 4.

3.2. Gravitational wave signals for training and testing the XGBoost classifier

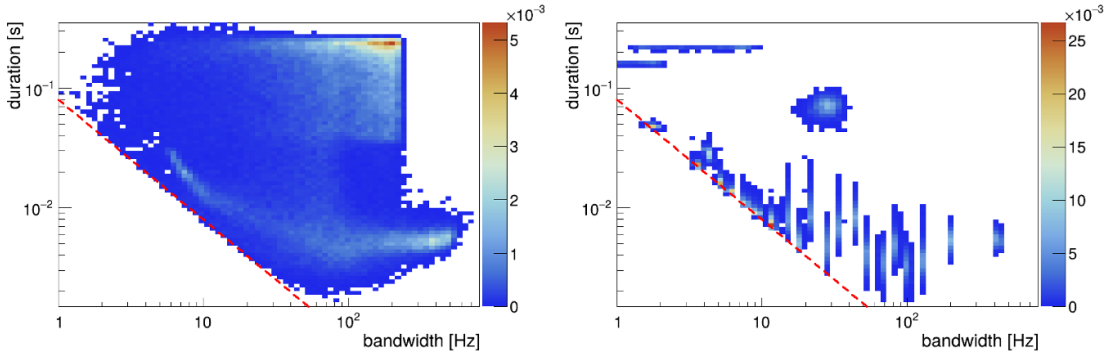
The simulated GWTs used for training and testing procedures are injected in GW strain data and analyzed by cWB-2G. The populations of simulated signals are chosen according to the goals of each search. All-sky searches for short duration GWTs of generic morphology lack a well-defined signal model. Therefore, the optimization of signal sets for training and testing HL-burst and HLV-burst is ill-defined, and their selection is based on qualitative criteria. It is particularly important that the training and testing populations differ sufficiently, with the former lacking distinctive morphological features, in order to assess the algorithm's ability to generalize to previously unseen signals. Instead, model-informed searches, such as HL-BBH, can exploit the same GWT population model for training and testing.

The qualitative criteria adopted for the selection of simulated GWTs to train the HL-burst and HLV-burst searches include:

- lower amplitude GWTs are likely more abundant than higher amplitude ones;
- all sky directions and arrival times during the observation should be uniformly distributed;
- preserve sensitivity to diverse GWTs, without assuming patterns in the signals' waveforms.

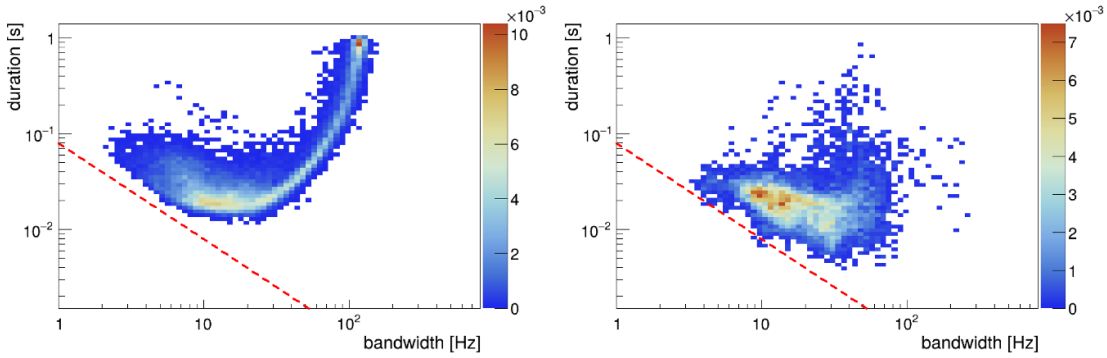
HL-burst and HLV-burst are trained using white noise burst (WNB) injections, which are GWTs with random polarization amplitudes as described in [29]. WNBs provide a challenging test case due to their noise–like properties. Their central frequency, bandwidth and duration parameters are randomly sampled across the searched ranges, in particular the duration and bandwidth span approximately from 2 ms to 0.3 s and from 2 Hz to 0.5 kHz respectively. For the signal model, duration and bandwidth are defined as standard deviations in time and frequency of the SNR ^{2} of the injected waveform after whitening, i.e. in zero–noise condition. The resulting distribution of reconstructed bandwidth and duration of these WNBs is shown in figure 4(a). The chosen population emphasizes training signals with larger TF spread, since their lower SNR density makes them harder to detect. We also enhance the population of signals near the Heisenberg–Gabor limit to match time-frequency characteristics of the dominant background triggers (see figure 3). The amplitudes of WNBs are simulated from a distribution $\propto \mathrm{SNR}^{-2}$, down to a minimum SNR in the network of detectors that corresponds to a negligible detection efficiency. This amplitude distribution enhances the training to signals along directions of weaker sensitivity of the network of detectors, consistent with the aim to make the sky coverage as isotropic as possible. This choice achieves very good results for diverse populations of sources, e.g. either uniformly distributed in space or else uniformly distributed in distance.

The signal test set comprises diverse GWTs morphologies traditionally used by LVK collaborations for burst searches [15–17]. These include:



(a) GWT set for training, WNBs (b) GWT set for testing, mixed morphologies

Figure 4. Distribution of reconstructed duration and bandwidth for the injected signals, zero–noise and whitened, used as training and testing sets in HL-burst and HLV-burst. The red dashed line shows the Heisenberg–Gabor limit. Color scale indicates fraction of counts per bin. Panel (a): training set of WNBs. Panel (b): testing set of ad–hoc waveform morphologies, including sine-Gaussians near the Heisenberg–Gabor limit, wide–band Gaussian pulses with durations ∼ a few ms, WNBs with larger time-frequency volumes and durations ∼0.1 s.



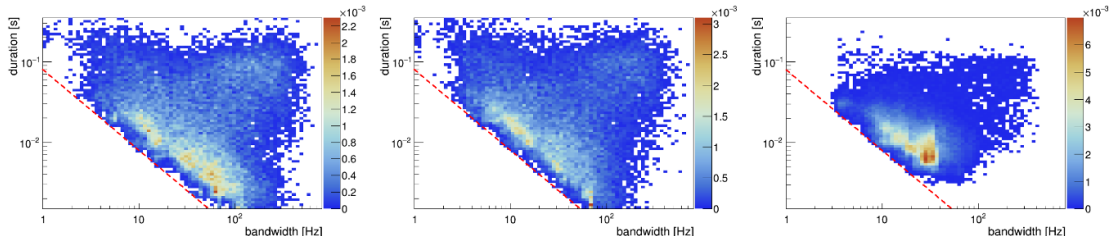
(a) injected CBC signals (b) recovered CBC signals

Figure 5. Distribution of CBC signals used for training and testing the HL–BBH search, in the duration-bandwidth plane. Panel (a) shows the simulated signals, zero–noise and whitened, with BNS and NS-BH signals appearing at longer durations, ∼1 s. Panel (b) shows the whitened signals found by cWB 2G, which are then passed to the XGBoost classifier. Color scale shows the fraction of counts per bin. The red dashed line indicates the Heisenberg–Gabor limit.

- Sine-Gaussian waveforms with elliptical (SGE) and linear sine-Gaussian (SG) polarization, at various central frequencies and Q-factors (3, 9, and 100), providing signals with compact TF representations, close to the Heisenberg–Gabor limit;
- WNBs at a few central frequencies, with random polarization amplitudes and spread over larger TF volumes;
- Linearly polarized Gaussian pulses (GA), short duration (from 0.1 to 4 ms) and wideband, whose reconstructed bandwidth is mainly set by the noise power spectral density of detectors.

The distribution of reconstructed duration and bandwidth of these test signals is shown in figure 4(b): clearly, they are just sparsely sampling the space of detectable short duration GWTs. More signal classes are usually added to extend the testing and the interpretation of the search; in fact, we also evaluate the sensitivity of HL-burst and HLV-burst using simulated CCSN signals and the model of CBC sources described below and shown in figure 5(a). Other aspects related to the distribution of simulated sources are discussed in section 6.

For the model–informed search HL-BBH, we use distinct sets of $\simeq 10^5$ injections each for training and testing, sampled from the Power Law + Peak mass distribution derived from LVK observing runs O1–O3 [48]. Figure 5 shows the distribution of signal duration and bandwidth of whitened signals, with panel (a) showing the injected ones, zero-noise, and panel (b) the found ones, as recovered by cWB-2G from the noisy data streams. The comparison between panels and further investigations reveal that HL-BBH has a lower sensitivity to BNS and NS-BH coalescences, which have longer duration and larger bandwidth, with standard deviations ∼1 s and ∼100 Hz in figure 5. This limitation could be mitigated



(a) HL-burst (b) HLV-burst (c) HL-BBH

Figure 6. Distributions of duration and bandwidth for the background triggers of the three searches passing an IFAR threshold of 1 week: (a) HL-burst, (b) HLV-burst, (c) HL-BBH. Color scales report the false alarm fraction per bin. The total count of false alarms ($\sim 10^5$) and the binning are the same in the three panels. We recall that the different false alarm sets of HL-burst and HL-BBH are resulting from the implementation of different XGBoost ranking procedures on the same set of background triggers found by cWB-2G from HL time-slides. The red dashed line indicates the Heisenberg–Gabor limit.

with a CBC-specific configuration of cWB-2G, in particular by tuning the trigger preselection and coherent analysis algorithms, but here we chose to focus on the effects of the tuning of the XGBoost post-processing and we kept the same cWB-2G configuration of the HL-burst. Section 6.3 discusses the resulting sensitivity to BBH and IMBH mergers.

4. Properties of false alarms and statistical significance of GWT candidates

The background triggers found by the three searches (HL-burst, HLV-burst and HL-BBH) show a variety of morphological features. Figure 6 compares their distributions as a function of duration and bandwidth, i.e. standard deviations of SNR ^{2} in time and frequency, considering all accidental triggers passing the IFAR threshold of 1 week from time slides. In all three searches, accidental triggers with a compact time-frequency morphology are still dominating the distributions, even though the ranking provided by XGBoost is mitigating the impact of the shorter duration triggers compared to the cWB-2G background shown in figure 3. HL-burst and HLV–burst show similar distributions of accidental triggers, see e.g. figures 6(a) and (b). In addition to the population which is peaked close to the Heisenberg–Gabor limit, a secondary component is visible with standard deviations in duration and bandwidth $\sim$0.1 s and a few hundred Hz respectively. We recall that, in HL-burst and HLV-burst searches, the duration and bandwidth features are not passed to the XGBoost classifier to preserve their agnostic character. Background triggers from HL-BBH are more compact in the duration-bandwidth plane, see figure 6(c). This is expected, given the properties of targeted signals shown in figure 5(a) and the exploitation by XGBoost of more features of reconstructed triggers, including here also the duration and bandwidth. However, the resulting mitigation of the dominating population of background triggers close to the Heisenberg-Gabor limit is still partial, mainly because of the residual overlap between distributions of estimated features of accidental triggers and of detected GWT injections.

The statistical significance of GWT candidates is estimated by comparison with the empirical distribution of false alarms from time-slides. This procedure can be affected by systematic errors on the point estimate and its uncertainty, that may be caused by e.g. correlations in the set of background triggers, non–stationary noise, residual effects due to the presence of genuine GWTs in the data, and finite sample size [49]. Given the novelty of the detection statistic adopted here, it is important to provide a thorough check of the accuracy of statistical significance estimates as well as to test if the fluctuations are consistent with a homogeneous Poisson point model. The expected counts of background triggers is numerically simulated by assuming a Poisson distribution of the counts per each time-lag $\mathcal{P}(k;\mu) = \mathrm{e}^{-\mu}\mu^k/k!$, see appendix B. Here, the expectation value μ is set equal to $\mathrm{FAR}(\eta_0')\ T$, where T is the live-time of each time-lag and $\mathrm{FAR}(\eta_0')$ is the measured value of the false alarm rate at threshold η_0' on the detection statistic.

The distribution of false alarm counts from time slides is in good agreement with the Poisson model for the three searches over a wide range of IFAR thresholds, from $\sim$1 d to $\sim$1 century. Figure 7 gives an example of the good agreement for an IFAR threshold of 1 week in HL-burst: the frequencies of occurrence of counts of false alarms above threshold are consistent with the Poisson model tuned to the empirical mean rate of events at that threshold. This agreement is confirmed for all searches in the entire range of tested IFAR thresholds, down to cumulative probabilities as low as a few 10^{-5}, limited by the total number of time slides performed in this work. Table 1 reports a summary of the agreement in probability of non-zero counts over the wide range of tested IFAR thresholds. From these checks, we can

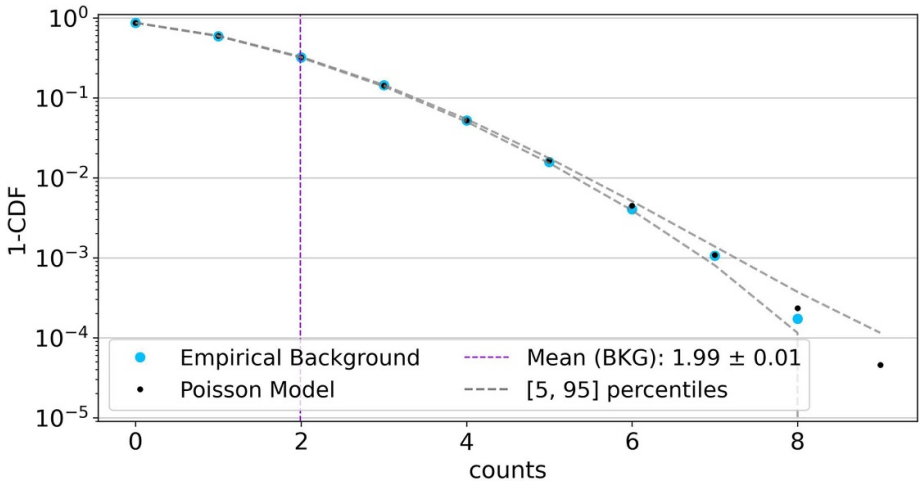


Figure 7. Cumulative distribution function of empirical false alarms counts from time slides for HL–burst at an IFAR thresholds of $\simeq 1$ week, compared to expectations from Poisson model. The y–axis reports the probability of occurrence of a false alarm count larger than the integer k in abscissa in one time–lag. Blue markers report results from time slides, black markers and dashed curves represent the expectations (mean plus 5% and 95% percentiles). The vertical dashed line indicates the mean counts per time-lag; no time–lags from the testing set show a false alarm count greater than 8. There is full agreement over all the investigated range of k in all the three searches at IFAR thresholds from $\sim$1 d to $\sim$1 century.

Table 1. Probability for a non-zero count of false alarms in one time–lag at selected FAR thresholds, labeled from A to F and listed in column 2 as with statistical uncertainties. The empirical estimates from time–slides ($\mathcal{P}_{\rm emp}(\text{counts} > 0)$, 3rd column) are consistent with expectations from the Poisson model ($\mathcal{P}_{\rm mod}(\text{counts} > 0)$, 4th column). The last column reports the standard deviation from the Poisson model. From top to bottom: results of the HL–burst, HLV–burst and HL–BBH.

HL–burst search				
Threshold	$\text{FAR}_{\rm emp} \pm \sigma^{\rm FAR}_{\rm emp}$	$\mathcal{P}_{\rm emp}(\text{counts} > 0)$	$\mathcal{P}_{\rm mod}(\text{counts} > 0)$	$\sigma_{\rm mod}$
A	(1.001 ± 0.001)/day	1.0	1.0	$5 \cdot 10^{-6}$
B	(1.000 ± 0.004)/week	0.864	0.863	0.002
C	(1.01 ± 0.02)/6month	0.074	0.074	0.001
D	(0.98 ± 0.03)/year	0.037	0.037	0.001
E	(0.99 ± 0.09)/10year	0.0038	0.0038	0.0003
F	(0.9 ± 0.3)/100year	0.00 034	0.00 035	0.0001
HLV–burst search				
Threshold	$\text{FAR}_{\rm emp} \pm \sigma^{\rm FAR}_{\rm emp}$	$\mathcal{P}_{\rm emp}(\text{counts} > 0)$	$\mathcal{P}_{\rm mod}(\text{counts} > 0)$	$\sigma_{\rm mod}$
A	(1.000 ± 0.002)/day	1.0	0.999 999	$7 \cdot 10^{-6}$
B	(0.999 ± 0.006)/week	0.871	0.868	0.003
C	(1.00 ± 0.03)/6month	0.075	0.075	0.002
D	(1.00 ± 0.04)/year	0.038	0.038	0.002
E	(1.0 ± 0.1)/10year	0.0039	0.0039	0.0005
F	(0.9 ± 0.4)/100year	0.0003	0.0003	0.0002
HL–BBH search				
Threshold	$\text{FAR}_{\rm emp} \pm \sigma^{\rm FAR}_{\rm emp}$	$\mathcal{P}_{\rm emp}(\text{counts} > 0)$	$\mathcal{P}_{\rm mod}(\text{counts} > 0)$	$\sigma_{\rm mod}$
B	(1.000 ± 0.004)/week	0.865	0.863	0.002
C	(1.00 ± 0.02)/6month	0.074	0.074	0.001
D	(0.99 ± 0.03)/year	0.038	0.037	0.001
E	(1.00 ± 0.09)/10year	0.0038	0.0038	0.0003
F	(0.9 ± 0.3)/100year	0.0003	0.0003	0.0001

conclude that there is no evidence for systematic biases of our point estimates of false alarm probabilities, nor for systematic deviations from the Poisson model.

In non-stationary noise, as e.g. related to human activities, one may expect a correlation of the rate of false alarms with the time shift between detectors. We analyzed separately the FAR estimates per each segment–shift at fixed values of the ranking statistics η_0', covering time shifts of LIGO Hanford data in

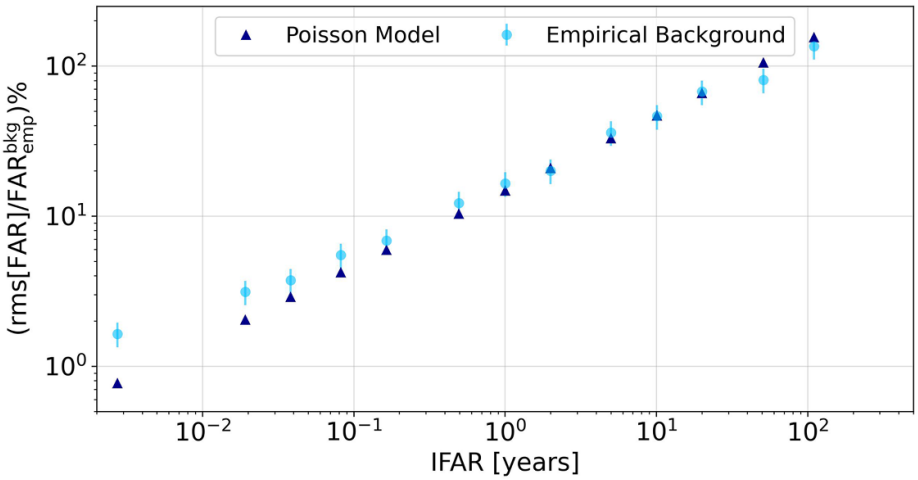


Figure 8. Relative fluctuations of empirical FAR estimates as a function of the average IFAR threshold for the HL–burst search. Blue markers report the RMS fluctuations in % on estimates of FAR normalized to their average value from 29 separate segment-shifts of LIGO Hanford detector, with error bars indicating the $\pm 1\sigma$ spread on these uncertainties. The relative fluctuations predicted by the numerical Poisson model at the same average IFAR thresholds are plotted as black triangles.

the range [−4 h 40 min, +4 h 40 min] in steps of 20 min, see section 3.1 and appendix B[17]. The resulting point estimates of FAR do not show a correlation with the time–shift applied between the LIGO detectors[18]. Moreover, we also investigated the root-mean-square (RMS) fluctuations of FAR point estimates from different segment-shifts as a function of the average FAR, see figure 8. These RMS fluctuations are explained by the Poisson model at IFAR thresholds of 2 months or larger. They exceed the poissonian statistical uncertainties only at lower IFAR thresholds and are anyway limited to <5% when using one segment-shift; therefore, we can neglect deviations from the Poisson model in the assessment of the statistical significance of GWT candidates.

Summarizing, these investigations brought no evidence for the presence of relevant systematic effects in our estimates of the statistical significance of GTW candidates in our searches on O3 data.

5. Correlations of background triggers in different searches

A key question in combining results from different searches is to take into account the joint significance of candidate GWTs. This section focuses on the correlation between background triggers across different searches, more specifically on the probability that a background trigger is in common to more searches. We compare false alarm results from pairs of searches using the same subset of time slides, more specifically, analyzing only common time–shifted live–times. In this condition, the probability that a background trigger in search A also appears in search B, $\mathcal{P}(B|A)$, is estimated by the count of common false alarms divided by the count of false alarms from search A. In case searches A and B are analyzing the same observing time and are set at the same IFAR threshold, $\mathcal{P}(B|A)$ is equal to $\mathcal{P}(A|B)$ by construction, since the two sets of background triggers will have the same expected counts, i.e. the same cardinality, and $\mathcal{P}(A) = \mathcal{P}(B)$. Therefore, the effective IFAR for the combined search made by the logical OR of two concurrent searches run at the same IFAR$'$ threshold is

$$\mathrm{IFAR} = \frac{\mathrm{IFAR}'}{2 - \mathcal{P}(A|B)} \tag{3}$$

where the denominator is the effective trials factor. This result is based on the addition rule in set theory, applied to the mean rate of false alarms of the combined search. Equation (3) follows from our assumptions of equal FAR thresholds of A and B searches and taking into account that FAR(AB)/FAR(A) $= \mathcal{P}(AB)/\mathcal{P}(A) = \mathcal{P}(B|A)$, where AB is the set of common false alarms.

HL-burst and HLV-bursts show a fraction of common false alarms; see table 2. Moreover, the expected symmetry $\mathcal{P}(A|B) = \mathcal{P}(B|A)$ is empirically obeyed, and the resulting probability of a common false

[17] Here, a time shift of −2 h synchronizes data streams to the same local time at Hanford and Livingston sites.

[18] We remark that the time shift applied to the Virgo detector plays a less relevant role on the statistic of false alarm for HLV analyses, given that for most sky directions and GW polarizations the network SNR is dominated by HL.

Table 2. Counts of background triggers at different search thresholds (from A to E, see table 1) on the same time-lags between H and L detectors: common and exclusive false alarms in HL-burst vs. HLV-burst (columns 2–4) and in HL-burst vs. HL-BBH (columns 5–7).

	False alarm counts					
	HL-burst vs HLV-burst			HL-burst vs HL-BBH		
Treshold	common	exclusive HL	exclusive HLV	common	exclusive burst	exclusive BBH
A	3092	14 533	14 412	—	—	—
B	470	2042	2034	126	2386	2479
C	59	229	234	9	279	283
D	7	41	46	0	48	50
E	2	4	2	0	4	5

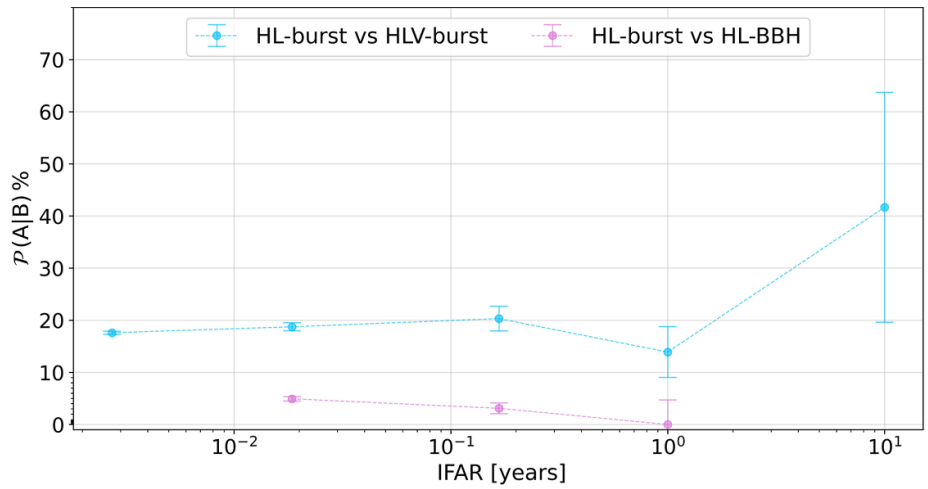


Figure 9. Measured probabilities $\pm 1\sigma$ error bars of common background triggers between concurrent searches HL-burst and HLV-burst (pale blue) and between HL-burst and HL-BBH (violet) as a function of the IFAR threshold.

alarm is $\mathcal{P}(B|A) \simeq 20\%$ for IFAR thresholds $\leqslant$ 1 year, see figure 9. While the statistical uncertainties limit the power of this investigation at higher IFAR thresholds, we can conservatively use a trials factor value equal to 2 for the combined search made by the logical OR of HL–burst and HLV–burst concurrently running at the same IFAR threshold. In fact, the measured trials factor $\simeq 1.8$ would only provide $a \simeq 10\%$ gain on the reported statistical significance of GWT candidates at IFAR $\leqslant$ 1 year.

The comparison between HL-burst and HL-BBH leads to the same main conclusion about the trials factor. Here, the probability of common false alarms is even lower, limited to a few %, see figure 9. This decorrelation is entirely due to the different ranking provided by XGBoost on the same set of transient candidates identified by cWB-2G on HL data. Here too, there is no evidence for a dependence of the probability of common false alarms on the statistical significance of the search, though the statistical uncertainties hinder this conclusion as our statistic vanishes at IFAR thresholds $>$ 1 year, see table 2.

6. Comparing sensitivities of different searches for GWTs

In this section, we discuss the sensitivities of the searches on the same observing time from LVK O3 public data (see section 3). Their detection efficiency are compared at equal statistical significance, i.e. selecting equal IFAR thresholds, by analyzing the same set of software signal injections. This procedure enables the investigation of common and exclusive detections in different searches. In section 6.1, we benchmark the sensitivity of HL-burst and HLV-burst, using the visible volume as a proxy for the probability of detection of GWTs. Section 6.2 discusses the performances of combined searches built from HL-burst and HLV-bursts. Finally, section 6.3 compares results of the model–informed HL-BBH and the agnostic HL-burst.

6.1. Comparing searches for generic GWTs on HL and HLV data

The performances of all–sky searches for short–duration bursts are compared by means of common signal injections from a set of ad-hoc waveforms, similar to the set used for testing the XGBoost classifier and described in section 3.2. The signals include linearly polarized GA, linearly polarized SG waveforms,

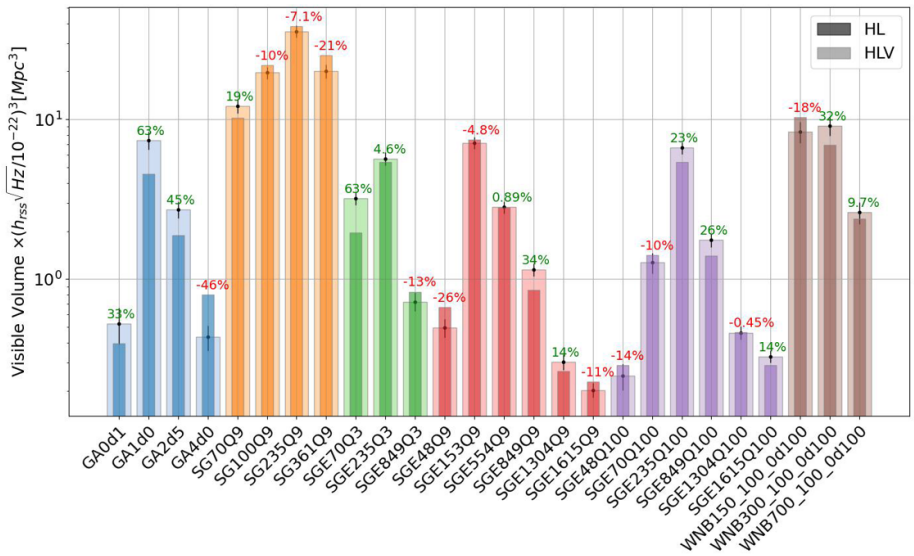


Figure 10. Visible volume of HLV-burst (light color) and HL-burst (full color) at threshold IFAR = 100 years on the same observing time from LVK O3. Sources are modeled as uniformly distributed in volume and emit at a fixed h_{rss} different GW waveform models: linearly polarized Gaussian pulses (GA, followed by xdy which stands for duration x.y ms), linearly polarized sine–Gaussian (SG, followed by the central frequency in Hz and Q value), elliptically polarized sine–Gaussian (SGE, followed by the central frequency in Hz and Q value) and white noise burst (WNB, followed by the central frequency and bandwidth in Hz, duration in (s).

elliptically polarized SGE and WNBs with random polarization amplitudes WNB. The polarization angle and, for SGEs, the cosine of the inclination angle are uniformly sampled.

Here, differently from what previously done in [29], we simulate a population of sources uniformly distributed in volume and radiating the same GW strength, i.e. standard GW sirens with the same absolute GW luminosity. In this simulation, we fixed the root–squared–sum strain amplitude at some reference distance, namely $h_{rss} = 10^{-22}/\sqrt{\mathrm{Hz}}$ at 1 Mpc. We consider this simple model as a suitable proxy for a wide class of plausible astrophysical populations of sources, respecting the generic criteria introduced in section 3.2[19].

In this context, the simplest benchmark which tracks the expected rate of detections is the visible volume, measured as $V_{vis}(\mathrm{IFAR}) = N_{detections}(\mathrm{IFAR})/\rho_{sources}$, where $N_{detections}(\mathrm{IFAR})$ is the count of sources detected by the search above the IFAR threshold and $\rho_{sources}$ is the count of simulated sources per unit volume, set to the same value for each waveform model. We remark that this procedure measures directly the product $V_{vis}(\mathrm{IFAR})h_{rss}^3$ for uniformly distributed standard sirens. Our visible volume results can be scaled easily when assuming a reference h_{rss} amplitude, as $V_{vis-hrss}(\mathrm{IFAR}) = V_{vis}(\mathrm{IFAR})\ [h_{rss}\sqrt{\mathrm{Hz}}/10^{-22}]^3$.

Figure 10 reports the visible volumes achieved by HLV-burst and HL-burst at IFAR threshold equal to 100 years, $V_{vis}(100\mathrm{y})$. At this high significance threshold, the two searches appear overall comparable: in half of the tested signal waveforms their visible volumes are within $\pm 15\%$ of each other, and HLV-burst performs better than HL-burst in 14 cases out of 26. This is the first demonstration that achieved visible volumes using latest public HLV data reach comparable values to those obtained using only the HL subset. Moreover, the overall performances of HL-burst reported in figure 10 show a mean relative improvement in V_{vis} of $\simeq +57\%$ with respect to what previously reported for cWB-2G on O3 HL data [29], with a median value of $+45\%$ and a variability between -30% and $+270\%$ on the 20 common waveform models. These progresses are indicative of the better performance of the new ranking procedure of candidates, fully based on the XGBoost classifier.

The variability of the visible volumes plotted in figure 10 is evidently correlated to the frequency bandwidth content of each waveform and to its polarization model. The variability in frequency is explained by the spectral properties of the noise of the detectors. The effect of the waveform polarization is mediated by the directional sensitivity of the detectors and by the following convention, which impacts the signal strength on Earth. For GA, SG and WNB signals, the emission strength is modeled as isotropic and the conventional h_{rss} is the actual root–squared–sum of the strain amplitude of the GWT

[19] Given the general aims and wide unknowns of our GWT searches, we did not consider effects from the cosmological red-shift, inhomogeneities in the nearby Universe or variable emission strength in the source model.

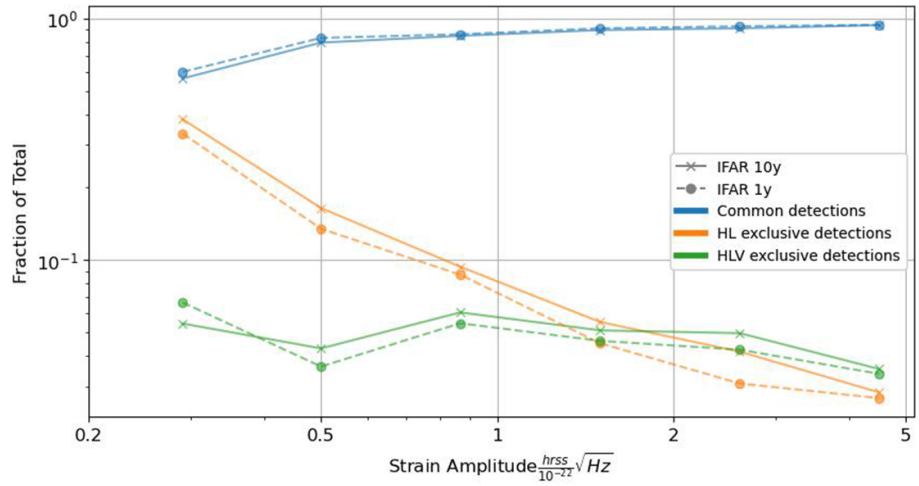


Figure 11. Fraction of software signal injections recovered exclusively by HL-burst (orange), by HLV-burst (green) and by both analyses (blue) as a function of signal strain amplitudes h_{rss}. Two significance thresholds are considered: IFAR $>$ 10 yr (continuous curves) and $>$ 1 yr (dashed curves). The injections equally weight the signal models plotted in figure 10, but here the source is imposing a grid of h_{rss} amplitudes.

at earth. Instead, SGE sources are modeled by the quadrupole emission formula and their conventional h_{rss} is referred to the optimal direction of emission; the signal is then projected on Earth by a random inclination angle with the line of sight. This is the main reason for the larger visible volumes in case of SG signals with respect to SGE ones.

A different set of signal injections is used to investigate the complementarity and correlation of detectable signals in HL–burst and HLV–burst as a function of signal strength. Here, the same set of waveform models reported in figure 10 are injected at a grid of selected h_{rss} amplitude values. Figure 11 shows the fractions of common and exclusive detections as a function of the signal strength for IFAR thresholds of 1 year and 10 years, summing together the detections from an equal number of software injections per waveform model. As expected, the strongest signals tested, with $h_{rss} \geqslant 2 \times 10^{-22}/\sqrt{\mathrm{Hz}}$, are typically found by both HL-burst and HLV-burst, with only a few percent of the detections exclusively identified by only one search. In this regime, the detection efficiency is close to unitary for most waveform models, and HL-burst and HLV-burst appear of comparable power. At not–so–strong amplitudes, the fraction of common detections decreases and the exclusive contribution of HL-burst gets more important. This asymmetry is consistent with previous reports of detection performances on O3 data [17, 29]. In fact, cWB-2G is constraining the search on HL data to the most favorable GW polarization component per each sky direction, while the exploitation of a detector of different directional sensitivity such as Virgo calls for considering both GW polarization components. The HL analyses then allow deploying stronger mitigation procedures of noise outliers with respect to HLV analyses. Overall, these results show only minor variations as a function of the search threshold in the IFAR range from 1 to 10 years[20].

To investigate the dependence of the visible volume on the distance to the source, we increased the statistics of the population of sources uniformly distributed in volume and selected three waveform models: one GA, one SGE and one WNB. Figure 12 reports the increments of the visible volume as a function of distance to the source for the HL-burst search. The counts per waveform model are here increased by a factor $\sim$20 with respect to the simulation of figure 10. The increments of visible volume show an obvious general behavior common to the three signal models, i.e. the probability of detection is high at the shortest distances and vanishes farther away. The behavior at larger distances depends on the signal model: in the case of randomly polarized WNBs the contributions to the visible volume steeply drop with distance, while for GA and SGE models the tails are more extended. We benchmark these effects by counting the fraction of detected signals beyond an effective radius defined as $[3/(4\pi)\ V_{vis}]^{1/3}$, where V_{vis} is the visible volume for the waveform model considered. These fractions are $\simeq$ 0.63, 0.53 and 0.38 in case of GA, SGE and WNB respectively, with a minor dependence on the search threshold

[20] Investigating the regime of lower signal amplitudes or higher IFAR thresholds is not technically feasible by this Monte Carlo. In fact, since the detection efficiency drops very rapidly with decreasing signal strength, it becomes computationally inefficient. The workaround which skips the analyses of too faint injections, which is deployed for the visible volume simulations, is not yet available here.

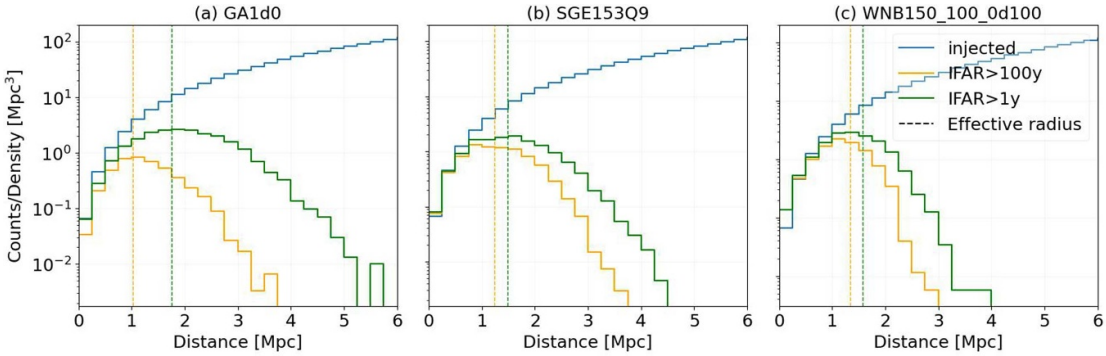


Figure 12. Visible volume of HL-burst as a function of source distance for three signal models, assuming a nominal strength $h_{rss} = 10^{-22}\sqrt{\text{Hz}}$ at a reference distance of 1 Mpc: (a) Gaussian pulse of duration 1 ms (GA1d0); (b) sine-Gaussian elliptically polarized at 153 Hz with $Q = 9$ (SGE153Q9); (c) white noise burst in the frequency band [150, 250] Hz with duration 0.1 s (WNB150_100_0d1). The orange and green histograms report the increments of visible volume per bin in distance, at thresholds IFAR = 100 y and 1 y respectively. The increments are estimated by the counts of detected sources belonging to the distance bin, divided by the density of simulated sources.

between IFAR = 1 yr and 100 yrs and statistical uncertainties of a few 10^{-3}. In other words, detectable sources belong to a more compact volume for WNBs, while for GAs detectable sources are more dispersed. This effect is expected, since the HL observatory is typically sensitive to only one GW polarization per each sky direction. WNBs provide GW polarization components that are on average equally weighted, so that along any incident direction they can trigger the HL observatory and are delineating a sharper horizon. At the other extreme, GA have a random linear polarization in the wavefront plane, which causes a larger variability of the signal projection on the HL observatory along any incident direction, resulting in a more blurred horizon.

Comparing results at IFAR thresholds of 1 yr and 100 yrs, the visible volume shrinks by a factor $\simeq$ 1.7 for the SGE and WNB models. Instead, for the GA the shrinking factor is much stronger, $\simeq$ 5.1. This is related to the similarity of GAs to the dominating family of stronger glitches in the LIGO detectors' data. In fact, the XGBoost training is tuned to the features of the background triggers, see 3.1, and glitch–like candidates are more frequently prevented from reaching a high rank than other waveform models.

6.2. Comparing combined searches for generic GWTs on HL and HLV data

We address here the question of what is the best performing short–duration burst search over a given observing time, when data are available from three detectors. We focus in particular on concurrent HL-burst and HLV-burst analyses set at the same IFAR threshold, and on the combined searches made by the logical OR ($\vee$) and logical AND ($\wedge$) of their candidates, respectively HL$\vee$HLV-burst and HL$\wedge$HLV-burst. We exploit here the same set of uniformly distributed sources used in the investigation of the visible volume as a function of distance, figure 12. With this signal set, whose amplitude distribution is $\propto h_{rss}^{-2}$, HL-burst is detecting more signals than HLV-burst for IFAR thresholds $<$ 100 years, in agreement with the fact that HL-burst performs better than HLV-burst for signal amplitudes in the lower amplitude range plotted in Figure 11. Taking HL-burst as reference, the fraction of signals that are common detections between HL-burst and HLV-burst is rather flat around $\sim$92% over most IFAR thresholds, for all the three waveform models tested. On the other side, the union of HL-burst and HLV-burst detections surpasses the HL-burst detections by an almost constant relative increment, $\sim$10%.

The comparison of visible volumes as a function of the IFAR threshold is shown in figure 13. The visible volume of HL$\vee$HLV-burst is estimated by counting the union of the detected signal sets from HL-burst and HLV-burst analyses and by halving the IFAR of HL-burst and HLV-burst. In fact, these analyses show mostly independent false alarms on O3 data, and the use of a trials factor of 2 brings a small $\simeq$ 10% bias on the conservative side (see section 5). For the first time, we are demonstrating a systematic gain in visible volume of HL$\vee$HLV-burst with respect to HL-burst for IFAR thresholds in a wide range, up to $\sim$10 years. The relative gain in probability of detection on O3 data is up to $\sim$8% in the three signal models investigated, and decreases as the IFAR threshold increases. The threshold range where HL$\vee$HLV-burst surpasses HL-burst is narrower in the case of the waveform model GA. This

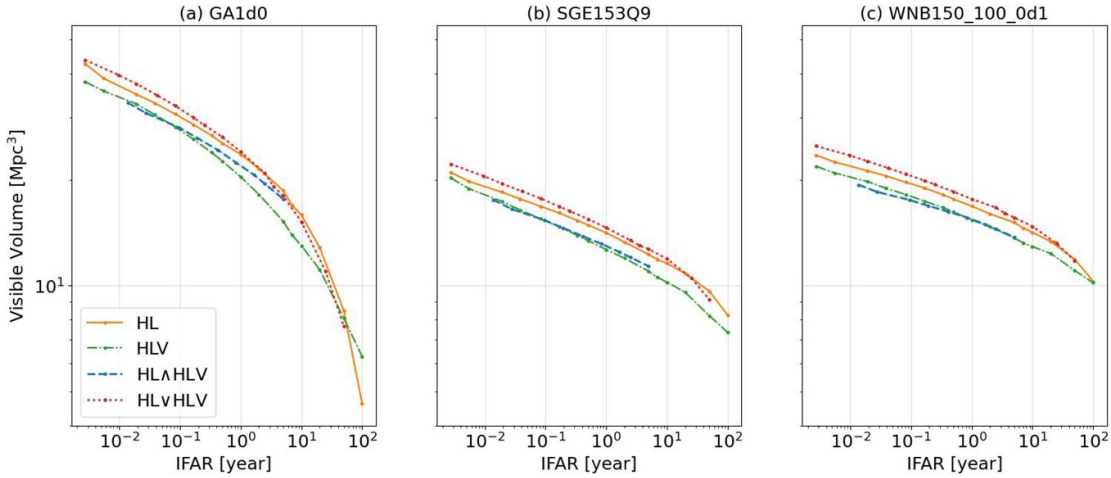


Figure 13. Comparison of visible volumes achieved by different searches as a function of the IFAR threshold, for the waveform models GA1d0 (a), SGE153Q9 (b) and WNB150_100_0*d*1 (c). The colored curves and belts show visible volumes and one σ statistical uncertainties for HL-burst (orange), HLV-burst (green), HL∨HLV-burst (red) and HL∧HLV-burst (blue).

reflects the observation that the visible volumes for GA signals show a cutoff knee at lower IFAR values than for other waveforms.

The visible volume of HL∧HLV-burst is estimated by counting the common detections of HL-burst and HLV-burst from the same Monte Carlo. The IFAR of HL∧HLV-burst is instead higher by a factor of $\simeq 5$ with respect to the original HL-burst and HLV-burst, up to a significance of $\sim$5 years[21]. As shown in figure 13, the resulting visible volumes of HL∧HLV-burst are not competitive with respect to the performances of HL-burst in the plotted range of thresholds.

6.3. Comparing sensitivities of agnostic and BBH–model–informed searches

The model-informed search HL-BBH is midway between generic searches, that do not assume a signal model, and template-based searches, which rely on accurate and extensive template banks. As described in section 3.2, our HL-BBH utilizes the same cWB-2G data processing and same candidate triggers of HL-burst, while it exploits a specific training of the XGBoost classifier that targets compact binary coalescences using a more extended set of morphological features of triggers. In this case, HL-BBH can be deployed on top of the HL-burst with a negligible addition of computational resources, while delivering better detection performances for BBH and IMBH sources with respect to HL-burst. However, the performances of this HL-BBH are suffering from the model agnostic preselection of triggers, which favors candidates with compact TF representations and is far from optimal in the case of lower chirp mass signals. There are recent examples of more advanced model–informed searches based on cWB targeting CBCs [42, 50] or hyperbolic encounters of compact objects [51].

Figure 14 compares the receiver operating curves of the model–informed and model–agnostic searches over the stellar-mass CBC and IMBH population models described at the end of section 3.2: the advantage in detection efficiency of HL-BBH is evident in the entire false alarm rate range. In particular, the probability of detection of stellar-mass CBC sources is approximately doubled with respect to HL-burst for candidate significances IFAR between 10 and 500 years, while the systematic improvement for IMBH signals is relevant but less striking, since HL-burst is already efficient for short–duration GWTs. However, this is accomplished at the cost of highly constraining the detectable signal parameter's space of HL-BBH: the visible volumes achieved by HL-BBH for all the signal morphologies shown in figure 10 become negligible with respect to what HL-burst achieves. This is expected, since estimated signal features such as central frequency, frequency bandwidth, duration, chirp mass and similarity to chirping signal morphology are used by HL-BBH for the multivariate classification of events.

[21] The false alarm statistics available is insufficient to estimate a useful lower limit to the resulting IFAR of HL∧HLV-burst for higher search thresholds, see figure 9.

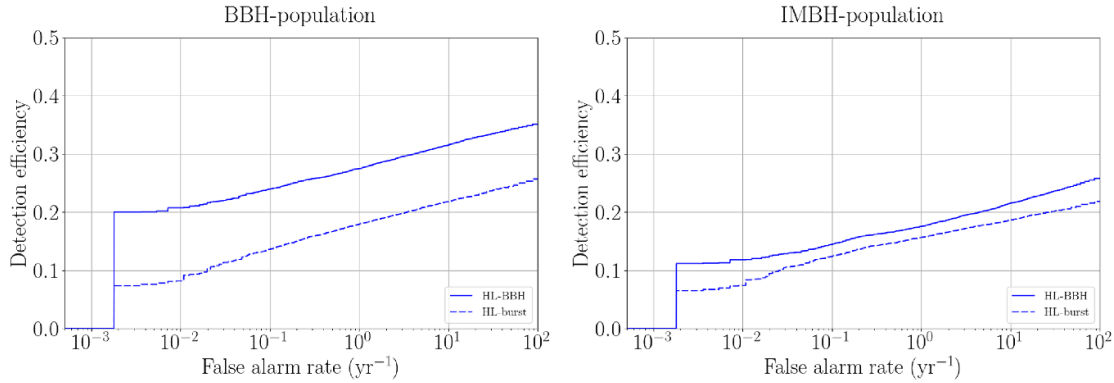


Figure 14. Detection efficiency versus IFAR for an astrophysical population of CBC sources of stellar mass (left) and for the subset of IMBH sources (right) using the unmodeled search HL-burst (dotted line) and the model-informed search HL-BBH (continuous line). For the population model used see end of section 3.2.

7. Summary and conclusions

This work is a status update of the 2G version of the cWB pipeline, one of the pipelines used by the LVK Collaboration to search for generic GWTs in the current observing run (O4). All the results discussed here are based on the analyses of 14.8 days of Hanford-Livingston-Virgo data released by LVK within the Observing Run 3, the most recent publicly available three–detector data. The main innovations are implemented in the post-processing stage of cWB-2G, where the ranking of candidate triggers is now entirely performed by the XGBoost classifier. This enhances the flexibility of the scope of cWB-2G, since its training more easily adapts to the characteristics of the detectors' network, of the noise triggers and of the target signal population. In particular, we focus on three searches: two targeting generic GWTs and using HL and HLV data, one targeting BBH and IMBH mergers using HL data.

We implement enhanced tools for validating the statistical significance of GWT candidates assessed by means of time-slides over a broad range of thresholds, at IFAR from 1 d to 1 century. Moreover, we measure the correlation and complementarity of background triggers and simulated GWTs across different searches. These methods enable us to provide for the first time a comprehensive comparison of searches for generic GWTs on Hanford-Livingston and Hanford-Livingston-Virgo data.

The most directly applicable results are reported in section 6, including benchmarks of the sensitivity to populations of sources and insights on the probability of detection as a function of distance to the source. For what concerns searches for short duration GWTs which run concurrently on the same observing time, the current cWB-2G version is capable to reach comparable performances when using HL or HLV data from O3 at high significance thresholds. Moreover, we demonstrate for the first time that the logical OR combination of the two concurrent searches provides a gain in probability of detection of short duration GWTs over a wide range of IFAR thresholds, improving visible volumes up to $\sim$8% with respect to that exploiting only HL data at the same IFAR, which was the search previously setting the best performances [17]. In addition, we demonstrate how the search can be model–informed by acting on the XGBoost classifier, and we compare a search for binary black holes of stellar and intermediate mass with a search for generic GWTs.

Based on this work, we shall design advantageous combinations of all-sky searches for GWTs provided by concurrent analyses, either unmodeled or weakly-modeled, on different combinations of GW detectors. In particular, our results provide important insights for assessing data analysis strategies in searches for short–duration GWTs in the LVK observing run O4, and in future observations. The methods discussed here have been demonstrated on the open source cWB-2G pipeline with XGBoost post-processing, but in principle they can be generalized to compare and combine all-sky searches for generic GWTs based on different pipelines.

We plan to extend this work on all-sky all-times GWT surveys to searches for targeted sources which leverage on additional information on their sky positions and GWT arrival times. This development will call for investigating different weakly-modeled approaches, as well as comparing and combining results from concurrent analyses on different detectors' configurations. For what concerns programming code developments, we are developing a Python-based implementation of the cWB-2G pipeline, PyCWB [52, 53], which promises to easy further methodological progresses and usability by a broader research community.

Acknowledgments

The authors are grateful to Dr Leigh Smith for helpful discussions and feedback on this manuscript.

This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

The authors are grateful for computational resources provided by the LIGO Laboratory, supported by National Science Foundation Grants PHY-0757058 and PHY-0823459.

This work has been funded by the European Union-Next Generation EU, Mission 4 Component 1 CUP J53D23001550006 with the PRIN Project No. 202275HT58. This work is partially supported by ICSC—Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union—NextGenerationEU.

MD acknowledges the support of the Sapienza Grants No. RM123188F3F2172A and RG1241905E47F60C. OGF acknowledges support by the Spanish Agencia Estatal de Investigaci'on (Grants PID2021-125485NB-C21 and PID2024-159689NB-C21) funded by MCIN/AEI/10.13039/501100011033 and ERDF A way of making Europe, as well as by the Portuguese Foundation for Science and Technology (FCT) through doctoral scholarship UI/BD/154358/2022, and CF-UM-UP through Strategic Funding UIDB/04650/2020.

Y X was supported by the Universitat de les Illes Balears (UIB); the Spanish Agencia Estatal de Investigación grants PID2022-138626NB-I00, RED2024-153978-E, RED2024-153735-E, funded by MICIU/AEI/10.13039/501100011033 and the ERDF/EU; and the Comunitat Autónoma de les Illes Balears through the Conselleria d'Educació i Universitats with funds from the European Union—NextGenerationEU/PRTR-C17.I1 (SINCO2022/6719) and from the European Union—European Regional Development Fund (ERDF) (SINCO2022/18146).

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://gwosc.org/ [55].

Appendix A. Configuration of analyses

A.1. Tuning of XGBoost hyperparameters

In this work, we utilize the XGBoost classifier [36], whose structure is defined by hyperparameters. The tuning of the hyperparameters is performed on a validation dataset using a K-fold cross validation: the original dataset is randomly shuffled and split into k different subsets, of which $K-1$ are used for training the model, and the remaining one is used to validate the performance. The procedure is repeated K times, looping over the subsets, so that at every iteration, a different subset is treated as an independent validation set. The evaluation results are computed and stored at every iteration, and they are combined at the end to provide a final, quantitative assessment of the model's performance. Hyperparameters' tuning involves evaluating performance metrics such as the receiver operating characteristic (ROC) curve and the precision-recall (PR) curve. In particular, we considered the areas under the PR curve (PR-AUC) and ROC curve (ROC-AUC). The former is particularly relevant for imbalanced datasets, since it benchmarks the trade-off between false positives and the rate of false negatives. In addition, we considered a particular version of the ROC-AUC metric built from the logarithmic values of detection efficiency and false alarm probability. In fact, this enhances the benchmark's sensitivity to the region of confident detections, i.e. in the lower range of investigated false alarm probabilities.

Table A1. List of XGBoost hyperparameters (first column), their allowed range (second column), the tuned values for HL-burst and HLV burst (third column), and the tuned values for HL-BBH (fourth column).

Parameter	range	HL and HLV-burst	HL-BBH
Learning_rate	(0,1)	0.003	0.03
Max_depth	$(0,\infty)$	8	13
Min_child_weight	$(0,\infty)$	3	2
Colsample_bytree	(0,1]	0.5	0.7
Subsample	(0,1]	0.8	0.6
Gamma	$[0,\infty]$	0.4	1.0

Table A2. List of the trigger's features reconstructed by cWB-2G and passed to XGBoost for HL-burst and HLV-burst.

Feature	Description
norm	sum of energies appearing in all TF resolution layers divided by energy of the event. Lower values mean the signal energy is concentrated in very few TF layers, a characteristic which is very common for background triggers.
noise	$\sqrt{\frac{1}{\sum_{\rm det}\frac{1}{N_{\rm det}}}}$ where $N_{\rm det}$ is the noise energy in each detector.
c_c	network correlation coefficient.
χ^2	noise energy per degree of freedom.
mSNR/likelihood	max value of SNR^2 in single detectors divided by the total likelihood in the network.
ecor/likelihood	sum of the off-diagonal terms of likelihood matrix normalized to the total likelihood.
rho0_100d0	effective correlated SNR, i.e. η_0 from equation (2), capped at 100.
Qa	similarity of waveform to one dominant pulse, inspired by blip glitches.
Qp	effective number of oscillations in the waveform.

The tuning of XGBoost's hyperparameters has been performed manually and separately for the short duration burst searches and for the HL-BBH, which use different sets of trigger features (see next subsection). The selected values of hyperparameters are reported in in table A1. We checked that the performances of the searches were not crucially depending on plausible changes with respect to the selected values.

Here, **learning_rate** regulates how much each generated decision tree affects the successive one by shrinking the weights of the features (the lower the less prone to overfitting). The **max_depth** parameter determines the deepness of each decision tree, deeper trees can capture more complex patterns in the data, but may also lead to overfitting. The other parameters **min_child_weight**, **colsample_bytree**, **subsample**, and **gamma** act as different forms of regularization to preserve the conservativeness of the algorithm and prevent overfitting [36, 38].

A.2. XGBoost training procedure

Training the XGBoost ML algorithm provides a decision model capable of discriminating between GWT triggers and background triggers. The background triggers are sampled by analyzing time-slides of the original data, that provide an off-source data set which is typically more than 10^5 times longer than the on-source observing time. A Monte Carlo of GWT injections provides the set representing genuine signals.

These data sets are split in two separate halves, one half is used for training XGBoost, while the second half is used to test the trained model by measuring the statistical significance of candidates and the detection efficiency for the simulated population of GWTs. Further alternative population models of GWTs are simulated to interpret the search under different conditions.

The 9 trigger's features that cWB-2G passes to XGBoost for the HL-burst and HLV-burst are described in table A2. HL-BBH instead uses the 15 features listed in table A3.

Table A3. List of the trigger's features reconstructed by cWB-2G and passed to XGBoost for HL-BBH.

Feature	Description
norm	same as in burst searches.
c_c	same as in burst searches.
χ^2	same as in burst searches.
sSNR/likelihood	SNR^2 of L detector divided by the total likelihood in the network.
ecor/likelihood	same as in burst searches.
rho0_25d0	effective correlated SNR, i.e. η_0 from equation (2), capped at 25.
Qa	same as in burst searches.
Qp	same as burst searches.
frequency	mean frequency of the recovered SNR^2.
duration	standard deviation of the time distribution of the recovered SNR^2.
bandwidth	standard deviation of the frequency distribution of the recovered SNR^2.
mchirp	estimate of chirpmass [54].
mchirp3-5	three different metrics of the goodness-of-fit of the chirpmass estimate

Appendix B. Poisson checks and background stability

Here we report the details of the study discussed in section 4 of this paper. The following tables show the empirical estimate of the FAR (FAR^{slag}_{emp}) for each segment-lag and η_0' thresholds ($\{A, B, C, D, E, F\}$) for the tree different searches: HL-burst (table B1), HLV-burst (table B3), and HL-BBH (table B2) respectively. Each FAR^{slag}_{emp} is provided with its relative 95% credible interval in square brackets, computed from the homogeneous Poisson point model described in section 4. The FAR^{slag}_{emp} values which fall off this credible interval are highlighted in bold. The $\sigma[FAR^{slag}_{emp}]$ and $\sigma[FAR^{slag}_{th}]$ report the variances of the $FAR^{slag}_{emp\,/\,th}$ set for each η_0' threshold value. The last two rows of each table list the empirical and theoretical FAR (FAR^{bkg}_{emp} and FAR^{bkg}_{th} respectively) estimates for the entire background with their uncertainties.

Figure B1 shows, for HL-burst, that the empirical FAR estimate for each segment-lag, FAR^{slag}_{emp}, does not display any trend as a function of the time shift. This is a visual representation of the information encapsulated in table B1.

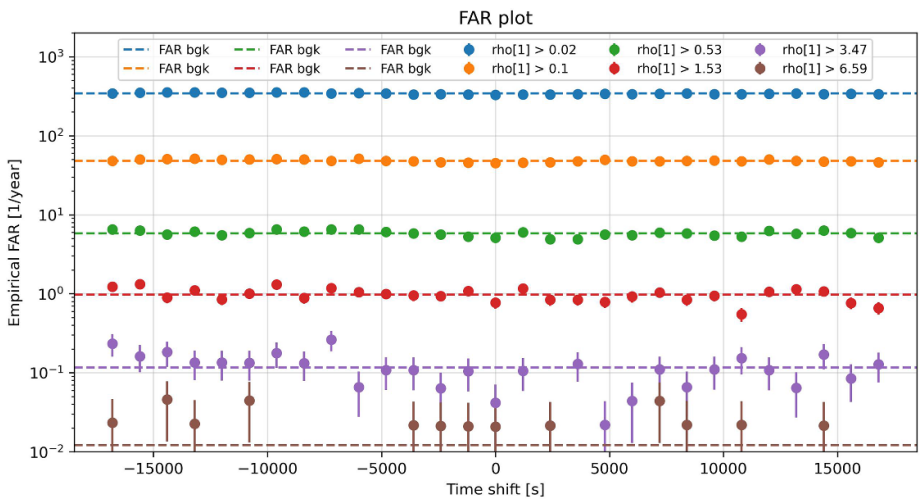


Figure B1. FAR values at fixed η_0' as a function of the average shift in time of disjoint sets of time slides. The time shift is applied to the Hanford data stream.

Table B1. Comparison of FAR values at fixed η_0' thresholds as a function of the average shift in time of disjoint sets of time slides for the HL−burst search. The time shift is applied to the Hanford data stream. The FAR estimates that fluctuate more than 2 standard deviations from the mean FAR (last row) are highlighted in bold.

	A_{HL}	B_{HL}	C_{HL}	D_{HL}	E_{HL}	F_{HL}
	FAR^{slag}_{emp} $[FAR^{5\%}_{slag,\,th}, FAR^{95\%}_{slag,\,th}]$					
Time shift [s]	[1/day]	[1/week]	[1/6 month]	[1/year]	[1/10years]	[1/100years]
−16 800	1.0 [0.99, 1.01]	0.98 [0.97, 1.04]	1.1 [0.8, 1.2]	1.1 [0.7, 1.2]	1.4 [0.2, 1.9]	2.3 [0.0, 4.7]
−15 600	1.01 [0.99, 1.01]	1.02 [0.97, 1.04]	1.1 [0.8, 1.2]	**1.3** [0.8, 1.2]	1.4 [0.2, 1.9]	0.0 [0.0, 4.6]
−14 400	**1.02** [0.99, 1.01]	1.02 [0.97, 1.04]	0.9 [0.8, 1.2]	0.8 [0.7, 1.2]	1.4 [0.2, 1.8]	2.3 [0.0, 4.6]
−13 200	**1.03** [0.99, 1.01]	**1.04** [0.97, 1.03]	1.1 [0.8, 1.2]	1.0 [0.8, 1.2]	1.1 [0.2, 1.8]	2.3 [0.0, 4.5]
−12 000	**1.02** [0.99, 1.01]	1.02 [0.97, 1.03]	**0.8** [0.8, 1.2]	0.8 [0.7, 1.2]	0.7 [0.2, 1.8]	0.0 [0.0, 4.5]
−10 800	**1.02** [0.99, 1.01]	1.02 [0.97, 1.03]	1.0 [0.8, 1.2]	0.9 [0.7, 1.2]	0.9 [0.2, 1.8]	2.2 [0.0, 4.5]
−9600	**1.02** [0.99, 1.01]	**1.04** [0.97, 1.03]	1.2 [0.8, 1.2]	1.2 [0.7, 1.2]	1.1 [0.2, 1.8]	0.0 [0.0, 4.5]
−8400	**1.02** [0.99, 1.01]	1.03 [0.97, 1.04]	0.9 [0.8, 1.2]	0.8 [0.7, 1.2]	0.9 [0.2, 1.8]	0.0 [0.0, 4.4]
−7200	0.99 [0.99, 1.01]	0.97 [0.97, 1.03]	1.2 [0.8, 1.2]	1.1 [0.8, 1.2]	**2.0** [0.2, 1.7]	0.0 [0.0, 4.4]
−6000	1.0 [0.99, 1.01]	**1.04** [0.97, 1.03]	1.0 [0.8, 1.2]	0.9 [0.8, 1.2]	0.7 [0.2, 1.7]	0.0 [0.0, 4.4]
−4800	0.99 [0.99, 1.01]	0.98 [0.97, 1.03]	1.1 [0.8, 1.2]	0.9 [0.7, 1.2]	0.7 [0.2, 1.7]	0.0 [0.0, 4.3]
−3600	**0.97** [0.99, 1.01]	**0.96** [0.97, 1.03]	1.0 [0.8, 1.2]	0.9 [0.7, 1.2]	0.6 [0.2, 1.7]	2.2 [0.0, 4.3]
−2400	**0.97** [0.99, 1.01]	**0.94** [0.97, 1.03]	**0.8** [0.8, 1.2]	0.9 [0.8, 1.2]	**0.2** [0.2, 1.9]	0.0 [0.0, 4.2]
−1200	**0.96** [0.99, 1.01]	**0.94** [0.97, 1.03]	1.0 [0.8, 1.2]	1.0 [0.8, 1.2]	0.8 [0.2, 1.9]	2.1 [0.0, 4.2]
0	**0.98** [0.99, 1.01]	**0.96** [0.97, 1.03]	0.9 [0.8, 1.2]	0.8 [0.8, 1.2]	0.4 [0.4, 1.9]	2.1 [0.0, 4.2]
1200	0.99 [0.99, 1.01]	**0.96** [0.97, 1.03]	1.2 [0.8, 1.2]	**1.3** [0.8, 1.2]	1.1 [0.2, 1.7]	0.0 [0.0, 4.2]
2400	**0.99** [0.99, 1.01]	**0.96** [0.97, 1.03]	0.9 [0.8, 1.2]	0.9 [0.8, 1.2]	0.4 [0.2, 1.7]	0.0 [0.0, 4.3]
3600	0.99 [0.99, 1.01]	1.01 [0.97, 1.03]	0.9 [0.8, 1.2]	0.9 [0.7, 1.2]	1.5 [0.2, 1.7]	0.0 [0.0, 4.3]
4800	1.01 [0.99, 1.01]	**1.04** [0.97, 1.03]	0.9 [0.8, 1.2]	0.9 [0.8, 1.2]	**0.2** [0.2, 1.7]	0.0 [0.0, 4.4]
6000	1.0 [0.99, 1.01]	1.0 [0.97, 1.03]	1.2 [0.8, 1.2]	1.0 [0.8, 1.2]	0.4 [0.2, 1.8]	0.0 [0.0, 4.4]
7200	1.01 [0.99, 1.01]	1.0 [0.97, 1.03]	1.2 [0.8, 1.2]	1.1 [0.7, 1.2]	1.1 [0.2, 1.8]	**4.4** [0.0, 4.4]
8400	1.01 [0.99, 1.01]	1.02 [0.97, 1.03]	1.0 [0.8, 1.2]	0.9 [0.8, 1.2]	0.7 [0.2, 1.8]	2.2 [0.0, 4.4]
9600	1.0 [0.99, 1.01]	1.03 [0.97, 1.03]	1.0 [0.8, 1.2]	1.0 [0.8, 1.2]	1.1 [0.2, 1.8]	0.0 [0.0, 4.4]
10 800	1.0 [0.99, 1.01]	1.0 [0.97, 1.03]	**0.8** [0.8, 1.2]	**0.7** [0.8, 1.2]	1.8 [0.2, 1.8]	2.2 [0.0, 4.4]
12 000	1.01 [0.99, 1.01]	**1.05** [0.97, 1.03]	1.1 [0.8, 1.2]	1.2 [0.7, 1.2]	1.1 [0.2, 1.7]	0.0 [0.0, 4.3]
13 200	1.01 [0.99, 1.01]	1.02 [0.97, 1.03]	1.1 [0.8, 1.2]	1.2 [0.8, 1.2]	0.9 [0.2, 1.7]	0.0 [0.0, 4.3]
14 400	1.01 [0.99, 1.01]	0.99 [0.97, 1.03]	1.1 [0.8, 1.2]	1.2 [0.8, 1.2]	**1.9** [0.2, 1.7]	2.1 [0.0, 4.3]
15 600	1.0 [0.99, 1.01]	1.01 [0.97, 1.03]	1.0 [0.8, 1.2]	0.9 [0.8, 1.2]	1.1 [0.2, 1.9]	0.0 [0.0, 4.2]
16 800	1.0 [0.99, 1.01]	0.97 [0.97, 1.03]	0.9 [0.8, 1.2]	0.8 [0.8, 1.2]	1.3 [0.2, 1.7]	0.0 [0.0, 4.3]
$\sigma\left[FAR^{slag}_{emp}\right]$	$\pm 0.01(6)$	$\pm 0.03(1)$	$\pm 0.1(2)$	$\pm 0.1(6)$	$\pm 0.4(6)$	$\pm 1.(2)$
$\sigma\left[FAR^{slag}_{th}\right]$	± 0.007	$\pm 0.02(0)$	$\pm 0.1(0)$	$\pm 0.1(5)$	$\pm 0.4(6)$	$\pm 1.(4)$
$FAR^{bkg}_{emp} \pm \sigma^{bkg}_{emp}$	(1.001 ± 0.001)	(1.000 ± 0.004)	(1.01 ± 0.02)	(0.98 ± 0.03)	(0.99 ± 0.09)	(0.9 ± 0.3)
$FAR^{bkg}_{th} \pm \sigma^{bkg}_{th}$	(1.00144 ± 0.00001)	(1.00041 ± 0.00004)	(1.0083 ± 0.0002)	(0.9835 ± 0.0003)	(0.9897 ± 0.0009)	(0.910 ± 0.003)

Table B2. Comparison of η_0' values at fixed rho thresholds as a function of the average shift in time of disjoint sets of time slides for the HL−BBH search. The time shift is applied to the Hanford data stream. The FAR estimates that fluctuate more than 2 standard deviations from the mean FAR (last row) are highlighted in bold.

	B_{BBH}		C_{BBH}		D_{BBH}		E_{BBH}		F_{BBH}	
	FAR^{slag}_{emp} $[FAR^{5\%}_{slag,\,th}, FAR^{95\%}_{slag,\,th}]$									
Time shift [s]	[1/week]		[1/6 month]		[1/year]		[1/10years]		[1/100years]	
−16 800	0.997	[0.966, 1.035]	**1.211**	[0.827, 1.187]	1.22	[0.73, 1.24]	1.4	[0.23, 1.86]	0.0	[0.0, 4.7]
−15 600	0.984	[0.966, 1.034]	0.857	[0.822, 1.181]	1.01	[0.75, 1.23]	1.39	[0.23, 1.85]	2.3	[0.0, 4.6]
−14 400	1.013	[0.966, 1.035]	1.086	[0.823, 1.177]	1.06	[0.74, 1.24]	0.46	[0.23, 1.83]	0.0	[0.0, 4.6]
−13 200	1.003	[0.966, 1.034]	0.879	[0.834, 1.172]	0.8	[0.76, 1.24]	1.13	[0.23, 1.8]	2.3	[0.0, 4.5]
−12 000	0.994	[0.967, 1.035]	**1.197**	[0.827, 1.174]	1.19	[0.75, 1.24]	0.67	[0.22, 1.79]	0.0	[0.0, 4.5]
−10 800	0.998	[0.966, 1.034]	1.038	[0.826, 1.183]	1.06	[0.75, 1.23]	0.45	[0.22, 1.79]	0.0	[0.0, 4.5]
−9600	1.009	[0.966, 1.035]	1.001	[0.834, 1.179]	1.08	[0.75, 1.23]	1.34	[0.22, 1.78]	0.0	[0.0, 4.5]
−8400	1.024	[0.966, 1.034]	0.981	[0.826, 1.179]	0.85	[0.76, 1.24]	0.88	[0.22, 1.76]	0.0	[0.0, 4.4]
−7200	0.999	[0.967, 1.034]	**0.819**	[0.83, 1.179]	0.86	[0.75, 1.23]	1.09	[0.22, 1.75]	2.2	[0.0, 4.4]
−6000	1.016	[0.966, 1.034]	**1.242**	[0.839, 1.177]	1.03	[0.75, 1.23]	0.44	[0.22, 1.74]	0.0	[0.0, 4.4]
−4800	1.02	[0.967, 1.034]	0.933	[0.835, 1.172]	0.96	[0.75, 1.24]	1.52	[0.22, 1.74]	0.0	[0.0, 4.3]
−3600	0.987	[0.967, 1.033]	1.025	[0.831, 1.176]	1.0	[0.77, 1.23]	1.29	[0.42, 1.73]	4.3	[0.0, 4.3]
−2400	1.019	[0.968, 1.034]	0.91	[0.836, 1.174]	0.83	[0.75, 1.23]	0.85	[0.42, 1.69]	0.0	[0.0, 4.2]
−1200	**0.939**	[0.968, 1.033]	**0.836**	[0.836, 1.17]	0.89	[0.76, 1.24]	1.04	[0.42, 1.88]	0.0	[0.0, 4.2]
0	0.986	[0.967, 1.033]	**0.831**	[0.831, 1.173]	0.82	[0.76, 1.23]	0.83	[0.21, 1.87]	0.0	[0.0, 4.2]
1200	1.006	[0.967, 1.033]	1.027	[0.836, 1.175]	1.02	[0.75, 1.23]	1.48	[0.21, 1.91]	2.1	[0.0, 4.2]
2400	1.012	[0.967, 1.033]	1.038	[0.835, 1.177]	1.14	[0.76, 1.22]	1.28	[0.21, 1.71]	2.1	[0.0, 4.3]
3600	1.01	[0.967, 1.034]	1.09	[0.831, 1.176]	1.0	[0.77, 1.23]	0.65	[0.22, 1.73]	0.0	[0.0, 4.3]
4800	1.002	[0.967, 1.035]	1.024	[0.839, 1.177]	1.12	[0.75, 1.23]	0.65	[0.22, 1.74]	0.0	[0.0, 4.4]
6000	0.969	[0.967, 1.034]	0.844	[0.833, 1.173]	0.8	[0.76, 1.23]	1.1	[0.22, 1.75]	0.0	[0.0, 4.4]
7200	0.994	[0.966, 1.034]	0.991	[0.826, 1.178]	0.87	[0.76, 1.24]	0.66	[0.22, 1.76]	0.0	[0.0, 4.4]
8400	0.995	[0.965, 1.034]	1.067	[0.836, 1.177]	0.76	[0.76, 1.24]	0.88	[0.22, 1.76]	0.0	[0.0, 4.4]
9600	1.007	[0.966, 1.034]	0.946	[0.836, 1.177]	0.93	[0.74, 1.24]	1.76	[0.22, 1.76]	**6.6**	[0.0, 4.4]
10 800	0.972	[0.966, 1.034]	0.909	[0.833, 1.172]	1.04	[0.76, 1.23]	0.66	[0.22, 1.75]	0.0	[0.0, 4.4]
12 000	0.984	[0.967, 1.034]	0.964	[0.834, 1.181]	0.9	[0.75, 1.24]	0.43	[0.22, 1.73]	0.0	[0.0, 4.3]
13 200	1.019	[0.968, 1.034]	**1.189**	[0.835, 1.178]	0.99	[0.76, 1.23]	1.07	[0.43, 1.93]	0.0	[0.0, 4.3]
14 400	1.011	[0.967, 1.033]	1.025	[0.833, 1.174]	1.01	[0.76, 1.22]	1.71	[0.21, 1.92]	0.0	[0.0, 4.3]
15 600	1.011	[0.968, 1.033]	0.965	[0.828, 1.178]	1.13	[0.75, 1.21]	0.85	[0.21, 1.91]	2.1	[0.0, 4.2]
16 800	1.022	[0.967, 1.033]	1.098	[0.831, 1.172]	**1.24**	[0.76, 1.22]	1.07	[0.21, 1.72]	2.1	[0.0, 4.3]
$\sigma\left[FAR^{slag}_{emp}\right]$	$\pm 0.01(8)$		$\pm 0.1(2)$		$\pm 0.1(3)$		$\pm 0.3(8)$		$\pm 1.(6)$	
$\sigma\left[FAR^{slag}_{th}\right]$	± 0.02		$\pm 0.1(0)$		$\pm 0.1(5)$		$\pm 0.4(7)$		$\pm 1.(4)$	
$FAR^{bkg}_{emp} \pm \sigma^{bkg}_{emp}$	(1.000 ± 0.004)		(1.00 ± 0.02)		(0.99 ± 0.03)		(1.00 ± 0.09)		(0.9 ± 0.3)	
$FAR^{bkg}_{th} \pm \sigma^{bkg}_{th}$	(1.00007 ± 0.00004)		(1.0003 ± 0.0002)		(0.9862 ± 0.0003)		(1.0008 ± 0.0009)		(0.898 ± 0.003)	

Table B3. Comparison of FAR values at fixed η_0' threshold as a function of the average shift in time of disjoint sets of time slides for the HLV−burst search. The time shift is applied to the Hanford data stream. The FAR estimates that fluctuate more than 2 standard deviations from the mean FAR (last row) are highlighted in bold.

	A_{HLV}	B_{HLV}	C_{HLV}	D_{HLV}	E_{HLV}	F_{HLV}
	FAR^{slag}_{emp} $[FAR^{5\%}_{slag,\,th}, FAR^{95\%}_{slag,\,th}]$					
Time shift [s]	[1/day]	[1/week]	[1/6 month]	[1/year]	[1/10years]	[1/100years]
0	0.996 [0.987, 1.012]	0.98 [0.97, 1.03]	0.99 [0.84, 1.18]	0.92 [0.76, 1.24]	1.26 [0.42, 1.89]	2.1 [0.0, 4.2]
1200	**0.977** [0.987, 1.012]	0.97 [0.97, 1.03]	1.05 [0.84, 1.17]	1.08 [0.76, 1.26]	**0.22** [0.22, 1.74]	0.0 [0.0, 4.3]
3600	**0.967** [0.987, 1.012]	**0.95** [0.97, 1.03]	0.9 [0.84, 1.17]	1.05 [0.75, 1.25]	1.72 [0.22, 1.72]	0.0 [0.0, 4.3]
6000	0.99 [0.987, 1.013]	0.99 [0.97, 1.03]	0.94 [0.84, 1.17]	0.89 [0.76, 1.24]	0.65 [0.22, 1.74]	0.0 [0.0, 4.3]
7200	**1.018** [0.987, 1.012]	**1.05** [0.97, 1.03]	1.04 [0.83, 1.17]	0.96 [0.77, 1.25]	1.53 [0.22, 1.75]	2.2 [0.0, 4.4]
8400	0.996 [0.987, 1.012]	0.99 [0.97, 1.03]	0.95 [0.83, 1.17]	0.95 [0.76, 1.24]	1.08 [0.22, 1.73]	0.0 [0.0, 4.3]
9600	1.008 [0.987, 1.012]	1.02 [0.97, 1.03]	0.97 [0.83, 1.18]	1.01 [0.76, 1.25]	0.43 [0.22, 1.73]	0.0 [0.0, 4.3]
10 800	**1.026** [0.987, 1.012]	1.03 [0.97, 1.03]	1.07 [0.83, 1.18]	1.11 [0.77, 1.24]	1.07 [0.21, 1.71]	4.3 [0.0, 4.3]
12 000	**1.015** [0.987, 1.013]	1.03 [0.97, 1.03]	0.91 [0.83, 1.17]	0.9 [0.77, 1.24]	0.64 [0.21, 1.71]	2.1 [0.0, 4.3]
13 200	**1.016** [0.987, 1.012]	1.01 [0.97, 1.03]	1.15 [0.83, 1.17]	1.02 [0.77, 1.26]	0.64 [0.21, 1.92]	0.0 [0.0, 4.3]
14 400	0.993 [0.987, 1.012]	0.98 [0.97, 1.03]	0.98 [0.84, 1.17]	0.97 [0.76, 1.25]	1.27 [0.21, 1.9]	0.0 [0.0, 4.2]
16 800	0.996 [0.987, 1.012]	0.99 [0.97, 1.03]	1.03 [0.83, 1.18]	1.11 [0.77, 1.25]	1.46 [0.21, 1.88]	0.0 [0.0, 4.2]
$\sigma\left[FAR^{slag}_{emp}\right]$	$\pm 0.01(7)$	$\pm 0.02(7)$	$\pm 0.06(9)$	$\pm 0.07(5)$	$\pm 0.4(6)$	$\pm 1.(4)$
$\sigma\left[FAR^{slag}_{th}\right]$	$\pm 0.007(7)$	$\pm 0.02(0)$	$\pm 0.1(0)$	$\pm 0.1(5)$	$\pm 0.4(6)$	$\pm 1.(4)$
$FAR^{bkg}_{emp} \pm \sigma^{bkg}_{emp}$	(1.000 ± 0.002)	(0.999 ± 0.006)	(1.00 ± 0.03)	(1.00 ± 0.04)	(1.0 ± 0.1)	(0.9 ± 0.4)
$FAR^{bkg}_{th} \pm \sigma^{bkg}_{th}$	(0.99973 ± 0.00002)	(0.99945 ± 0.00006)	(1.0001 ± 0.0003)	(0.9996 ± 0.0004)	(0.9996 ± 0.0004)	(0.895 ± 0.004)

ORCID iDs

A Martini 0009-0004-9683-0961
A Miani 0000-0001-7737-3129
M Drago 0000-0002-3738-2431
C Lazzaro 0000-0001-5993-3372
F Salemi 0000-0002-9511-3846
S Bini 0000-0002-0267-3562
O Freitas 0000-0002-2898-1256
E Milotti 0000-0001-7348-9765
G Principe 0000-0003-0406-7387
S Tiwari 0000-0003-1611-6625
A Trovato 0000-0002-9714-1904
G Vedovato 0000-0001-7226-1320
Y Xu 0000-0001-8697-3505
G A Prodi 0000-0001-5256-915X

References

[1] Aasi J *et al* 2015 Advanced LIGO *Class. Quantum Grav.* **32** 074001
[2] Acernese F *et al* 2015 Advanced Virgo: a second-generation interferometric gravitational wave detector *Class. Quantum Grav.* **32** 024001
[3] Aso Y, Michimura Y, Somiya K, Ando M, Miyakawa O, Sekiguchi T, Tatsumi D and Yamamoto H 2013 Interferometer design of the KAGRA gravitational wave detector *Phys. Rev. D* **88** 043007
[4] The LIGO Scientific Collaboration, The Virgo Collaboration, and the KAGRA Collaboration 2025 GWTC-4.0: updating the gravitational-wave transient catalog with observations from the first part of the fourth LIGO-Virgo-KAGRA observing run (https://doi.org/10.48550/arXiv.2508.18082)
[5] Gravitational wave open science center (available at: https://gwosc.org/)
[6] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration 2025 Gwtc-4.0: Methods for identifying and characterizing gravitational-wave transients (https://doi.org/10.48550/arXiv.2508.18081)
[7] Abbott B P *et al* 2019 Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1 *Phys. Rev. D* **100** 104036
[8] Abbott R *et al* 2021 Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog *Phys. Rev. D* **103** 122002
[9] Abbott R *et al* 2025 Tests of General Relativity with GWTC-3 *Phys. Rev. D* **112** 084080
[10] Abdikamalov E *et al* 2022 *Gravitational Waves From Core-Collapse Supernovae* (Springer) pp 909–45
[11] Kouvaris C 2025 Gravitational waves from magnetars *Int. J. Mod. Phys. D* **34** 2550037
[12] Giliberti E and Cambiotti G 2022 Starquakes in millisecond pulsars and gravitational waves emission *Mon. Not. R. Astron. Soc.* **511** 3365–76
[13] Ebersold M and Tiwari S 2020 Search for nonlinear memory from subsolar mass compact binary mergers *Phys. Rev. D* **101** 104041
[14] Gottlieb O, Levinson A and Levin Y 2024 In ligo's sight? vigorous coherent gravitational waves from cooled collapsar disks *Astrophys. J. Lett.* **972** L4
[15] Abbott B P *et al* 2017 All-sky search for short gravitational-wave bursts in the first advanced LIGO run *Phys. Rev. D* **95** 042003
[16] Abbott B P *et al* 2019 All-sky search for short gravitational-wave bursts in the second advanced LIGO and advanced Virgo Run *Phys. Rev. D* **100** 024017
[17] Abbott R *et al* 2021 All-sky search for short gravitational-wave bursts in the third advanced LIGO and advanced Virgo run *Phys. Rev. D* **104** 122004
[18] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration 2025 All-sky search for short gravitational-wave bursts in the first part of the fourth LIGO-Virgo-KAGRA observing run *Phys. Rev. D* **112** 102005
[19] Abbott B P *et al* 2018 All-sky search for long-duration gravitational wave transients in the first advanced LIGO observing run *Class. Quantum Grav.* **35** 065009
[20] Abbott B P *et al* 2019 All-sky search for long-duration gravitational-wave transients in the second advanced LIGO observing run *Phys. Rev. D* **99** 104033
[21] Abbott R *et al* 2021 All-sky search for long-duration gravitational-wave bursts in the third advanced LIGO and advanced Virgo run *Phys. Rev. D* **104** 102001
[22] Klimenko S, Mohanty S, Rakhmanov M and Mitselmakher G 2005 Constraint likelihood analysis for a network of gravitational wave detectors *Phys. Rev. D* **72** 122002
[23] Klimenko S , Yakushin I, Mercer A and Mitselmakher G 2008 A coherent method for detection of gravitational wave bursts *Class. Quantum Grav.* **25** 114029
[24] Klimenko S *et al* 2016 Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors *Phys. Rev. D* **93** 042004
[25] Abbott B P *et al* 2016 Observation of gravitational waves from a binary black hole merger *Phys. Rev. Lett.* **116** 061102
[26] Abbott R *et al* 2020 Gw190521: A binary black hole merger with a total mass of $150 m_{\odot}$ *Phys. Rev. Lett.* **125** 101102
[27] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration 2025 Gw231123: a binary black hole merger with total mass 190-265 $m_{\odot}$ *ApJL* **993** L25
[28] Drago M *et al* 2021 coherent waveburst, a pipeline for unmodeled gravitational-wave data analysis *SoftwareX* **14** 100678
[29] Szczepańczyk M J *et al* 2023 Search for gravitational-wave bursts in the third advanced LIGO-Virgo run with coherent waveburst enhanced by machine learning *Phys. Rev. D* **107** 062002
[30] Klimenko S 2022 Wavescan: multiresolution regression of gravitational-wave data (https://doi.org/10.48550/arXiv.2201.01096)

[31] Mishra T, Bhaumik S, Gayathri V, Szczepańczyk M J, Bartos I and Klimenko S 2025 Gravitational waves detected by a burst search in LIGO/Virgo's third observing run *Phys. Rev. D* **111** 023054
[32] 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 *Phys. Rev. D* **105** 063024
[33] Smith L, Ghosh S, Sun J, Gayathri V, Heng I S and Pai A 2024 Enhancing search pipelines for short gravitational-wave transients with Gaussian mixture modeling *Phys. Rev. D* **110** 083032
[34] Ghosh S, Smith L, Sun J, Pai A, Heng I S and Gayathri V 2025 Leveraging cross-detector parameter consistency measures to enhance sensitivities of gravitational-wave searches *Phys. Rev. D* **112** 063026
[35] Klimenko S *et al* 2024 cwb pipeline library: 6.4.6
[36] extreme gradient boosting (available at: https://github.com/dmlc/xgboost) (Accessed 10 March 2025)
[37] Chen T and Guestrin C 2016 Xgboost: A scalable tree boosting system *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (ACM) pp 785–94
[38] coherent waveburst web-site (available at: https://gwburst.gitlab.io) (Accessed 10 March 2025)
[39] Klimenko S, Vedovato G, Drago M, Mazzolo G, Mitselmakher G, Pankow C, Prodi G, Re V, Salemi F and Yakushin I 2011 Localization of gravitational wave sources with networks of advanced detectors *Phys. Rev. D* **83** 102001
[40] Salemi F, Milotti E, Prodi G,, Vedovato G, Lazzaro C, Tiwari S, Vinciguerra S, Drago M and Klimenko S 2019 Wider look at the gravitational-wave transients from GWTC-1 using an unmodeled reconstruction method *Phys. Rev. D* **100** 042003
[41] Necula V, Klimenko S and Mitselmakher G 2012 Transient analysis with fast Wilson-Daubechies time-frequency transform *J. Phys.: Conf. Ser.* **363** 012032
[42] Mishra T, O'Brien B, Gayathri V, Szczepańczyk M, Bhaumik S, Bartos I and Klimenko S 2021 Optimization of model independent gravitational wave search for binary black hole mergers using machine learning *Phys. Rev. D* **104** 023014
[43] Cabero M *et al* 2019 Blip glitches in advanced LIGO data *Class. Quantum Grav.* **36** 155010
[44] Abbott R *et al* 2023 Open data from the third observing run of LIGO, Virgo, KAGRA and GEO *Astrophys. J. Suppl. Ser.* **267** 29
[45] Buikema A *et al* 2020 Sensitivity and performance of the advanced LIGO detectors in the third observing run *Phys. Rev. D* **102** 062003
[46] Acernese F *et al* 2023 Virgo detector characterization and data quality: results from the o3 run *Class. Quantum Grav.* **40** 185006
[47] Abbott B P *et al* 2016 Observing gravitational-wave transient gw150914 with minimal assumptions *Phys. Rev. D* **93** 122004
[48] Abbott R *et al* 2023 Population of merging compact binaries inferred using gravitational waves through GWTC-3 *Phys. Rev. X* **13** 011048
[49] Was M, Bizouard M-A, Brisson V, Cavalier F, Davier M, Hello P, Leroy N, Robinet F and Vavoulidis M 2010 Limitations of the time slide method of background estimation *Class. Quantum Grav.* **27** 194014
[50] Mishra T *et al* 2025 Gravitational Waves Detected by a Burst Search in LIGO/Virgo's Third Observing Run *Phys. Rev. D* **111** 023054
[51] Bini S, Tiwari S, Xu Y, Smith L, Ebersold M, Principe G, Haney M, Jetzer P and Prodi G A 2024 Search for hyperbolic encounters of compact objects in the third LIGO-Virgo-KAGRA observing run *Phys. Rev. D* **109** 042009
[52] Yumeng X *et al* 2023 PycWB: a user-friendly, modular, and python-based framework for gravitational wave unmodelled search (available at: https://github.com/PycWB/pycwb)
[53] (Available at: https://github.com/PycWB/pycwb)
[54] Tiwari V, Klimenko S, Necula V and Mitselmakher G 2015 Reconstruction of chirp mass in searches for gravitational wave transients *Class. Quantum Grav.* **33** 01LT01
[55] Available at: https://gwosc.org/