Understanding Combined Results From Multiple GW Searches Using Information Theory

Oleksandra Lukina
University of South Dakota

Mentor: Derek Davis
LIGO, California Institute of Technology
(Dated: August 7, 2023)

Interim Report 2
LIGO SURF 2023

Determining whether a gravitational wave (GW) signal is of astrophysical origin or is caused by terrestrial noise still presents a challenge to the GW community. Current searches estimate the significance of events by calculating the false alarm rate (FAR) and p_{astro}, but these results are limited to a single search pipeline. In this work, we suggest a method of combining GW information by calculating harmonic mean FARs for different groups of GW searches and using them as the basis for calculating a joint p_{astro}. Using this approach, we compare the effectiveness of different combinations of searches, revisit the significance of previously detected events, and investigate the correlations between pipelines using the language of information theory.

I. INTRODUCTION

Since the first detection of gravitational waves (GWs) [1], the total number of GW candidates reported by LIGO-Virgo-KAGRA (LVK) Collaboration reached 90 [2] and continues to grow [3]. At the same time, determining whether a certain signal has astrophysical origin or is caused by terrestrial noise still remains a challenge, which leads to the uncertainty in the number of detected compact mergers [4].

Noisy local environments that are difficult to model, observations with multiple detectors, and inability to shield the instruments from GW signals result in the dependence of the estimated significance on a search analysis [5]. Two main approaches that are used to search for events include matching signals to the compact binary coalescence waveform templates and searching for transient signals across the network of detectors with minimal modeling. Variations of the first method are used in PyCBC [6], GstLAL [7], MBTA [8], IAS [9], and OGC [10] pipelines, and the second method is used in the cWB pipeline [2]. There are several technical differences between search pipelines, so they result in different estimations of the probability of an astrophysical origin. However, all pipelines are designed to search for the same signals, so their results are not fully independent and should correlate.

In order to calculate significance of events, we introduce two quantities, false alarm probability and p_{astro}. False alarm probability is a probability of observing a coincidence or a "false alarm" with a signal-to-noise ratio (SNR) equal or higher than a certain value. As a result, to confirm the presence of a signal, one must show that the probability to obtain the observed event in a dataset that only contains noise is smaller than a given threshold [5]. It is also possible to convert false alarm probability into a related quantity called the false alarm rate (FAR), which is measured in yr^{-1} and is usually included in GW catalogs, e.g., [2].

On the other hand, p_{astro} is defined as a probability that a GW candidate has astrophysical origin and is not caused by terrestrial noise. It is calculated by combining the rates at which triggers – outputs of a search pipeline – are generated by both astrophysical and noise sources, i.e., both false and true alarm rates [11].

II. METHODS

The goal of this project is to develop a method of combining information from multiple GW search pipelines and analyze what new knowledge these combinations provide about observed compact mergers. In addition, we are interested in analysing the contributions of the IAS and OGC pipelines, developed outside of the LVK collaboration, to the results of the internal pipelines.

In order to do this, we calculate the combined FAR using FARs from individual pipelines that correspond to the same events as a harmonic mean, which has a statistical advantage for combining results from dependent tests [12]. This corresponds to the equation below:

$$\frac{1}{FAR_{1...N}} = \frac{1}{N}\left(\frac{1}{FAR_1} + \frac{1}{FAR_2} + ... + \frac{1}{FAR_N}\right) \quad (1)$$

We can rewrite this equation using inverse false alarm rates (IFARs), measured in units of time, as follows:

$$IFAR_{1...N} = \frac{1}{N}\sum_{i=1}^{N} IFAR_i \quad (2)$$

Combined IFAR is then used as a new statistic that characterizes GW events which allows us to revisit their significance and observe how it changes devending on a specific combination of pipelines and whether we use real data or an injection set.

In addition, applying the approach used for a toy model in [13], we use IFAR distributions of the injection set to investigate the correlations between pipelines.

III. RESULTS

Using the equation (2) described in the Methods section, we calculated combined IFARs for all combinations of all available search pipelines for three different datasets:

1. The set of IFARs of all triggers that were detected by at least one pipeline among PyCBC (highmass and all-sky treated as two separate pipelines), GstLAL, or MBTA during the O3a obsering run [14], accesed via [15].

2. Same set as in dataset 1 but with addition of two pipelines external to the LVK collaboration, IAS [16] and OGC [10], data for which was accessed via [17] and [18] respectively.

3. The set of IFARs for an injection set that includes five LVK searches, namely PyCBC (BBH and hyperbank treated as two separate pipelines), GstLAL, MBTA, and CWB that contains 512431 injections.

After calculating combined IFARs for all events that were detected by at least one of the pipelines included in a dataset, we calculated the number of events that passes the threshold of $FAR > 1$ yr for each combination of different number of pipelines. Depending on a combination, joint searches detected between 25 and 36 events in dataset 1, 25-41 events in dataset 2, and 71763-96173 events in dataset 3 as compared to 23-36 events detected by each individual pipeline in datasets 1 and 2 and 18260-77289 events in the injection dataset.

Based on these calculations, we identified the combinations of pipelines that detect the most events for each number of combined pipelines and presented these results in a form of decision trees shown in Fig. 1. The combinations of pipelines are labeled according to the first letters of the searches that the combination includes, namely GstLAL – "g", PyCBC – "p", PyCBC_highmass – "h", MBTA – "m", CWB – "c", IAS – "i", and OGC – "o". For example, among the combinations of two pipelines used in dataset 1, the highest number of events (41) was detected by the combination of GstLAL ("g") and IAS ("i") searches, which is labeled as "gi".

Interestingly, the maximum number of detected events decreases as we add more searches including real data (datasets 1 and 2) and increases for the injection set (dataset 3). The addition of events that comes from adding more pipelines to the combined IFAR calculation for the injection set is illustrated in Fig. 2, which shows that GstLAL detects about 80% of the number of events that the combination of all LVK pipelines detects, and the combination of GstLAL and PyCBC detects about 90% of that number. In addition, we calculated that the harmonic mean combination of all pipelines recovers about 1.1% more events than the combination that uses the Bonferroni correction (taking the maximum IFAR among the pipelines and applying a trials factor).

Next, we compared the lists of candidates passing the threshold of combined $IFAR > 1\ yr$ between the combination of LVK pipelines (GstLAL, two PyCBC searches, and MBTA) and the combination of internal and external pipelines (GstLAL, two PyCBC searches, MBTA, IAS, and OGC). The list of events in given in Table I, which shows that the combination of LVK pipelines detect 33 events, and the combination of all searches detects 35 events. It is important to not that, although GW190514_065416 and GW190725_174728 only passed the IFAR threshold after the inclusion of two external pipelines, both events were officially reported in the O3a LVK publication [19].

Finally, we report the preliminary results of the qualitative estimation of correlation between different LVK pipelines that are illustrated using the corner plot in Fig. 3. The plot shows the pairwise correlations between IFAR distributions for CWB, GstLAL, MBTA, PyCBC_BBH, and PyCBC_hyperbank searches.

IV. NEXT STEPS

The final part of my SURF project will be dedicated to the calculation of joint p_{astro} using the results from multiple pipelines. In order to do this, we will need to fit FAR distributions to the signal and noise model and analyze how the probability of a signal being of an astrophysical origin changes depending on which pipelines we take into the account. We are particularly interested in seeing if the combined results obtained for O3a candidates suggest change in the classification of any of the events.

In addition, we are planning to continue the analysis of the correlations between search pipelines by finding a quantitative representation of the results shown in Table I. Moreover, we intend to find an explanation of why we observe a decrease in detected events with addition of pipelines for real data, but an increase in the same case for the injected data. We are interested in supporting our conclusions using the language of information theory and in terms of concepts like mutual information, interaction information, and total correlation.

GW	GPS Time	LVK IFAR (yr)	LVK+ IFAR (yr)	LVK Detected	LVK+ Detected
GW190408_181802	1238782700	1.17×10^{14}	7.82×10^{13}	+	+
GW190412_053044	1239082262	1.30×10^{26}	8.66×10^{25}	+	+
GW190413_134308	1239198206	2.14	1.67×10^{3}	+	+
GW190421_213856	1239917954	1.06×10^{2}	1.99×10^{3}	+	+
GW190425_081805	1240215503	7.39	4.93	+	+
GW190503_185404	1240944862	1.07×10^{5}	7.33×10^{4}	+	+
GW190512_180714	1241719652	3.24×10^{10}	2.16×10^{10}	+	+
GW190513_205428	1241816086	1.88×10^{4}	1.45×10^{4}	+	+
GW190514_065416	1241852074	9.14×10^{-2}	8.16	–	+
GW190517_055101	1242107479	8.07×10^{2}	2.22×10^{3}	+	+
GW190519_153544	1242315362	1.22×10^{5}	8.32×10^{4}	+	+
GW190521_030229	1242442967	1.94×10^{2}	2.64×10^{2}	+	+
GW190521_074359	1242459857	4.97×10^{31}	3.31×10^{31}	+	+
GW190527_092055	1242984073	1.11	1.10	+	+
GW190602_175927	1243533585	2.29×10^{6}	1.53×10^{6}	+	+
GW190620_030421	1245035079	2.24×10^{1}	1.49×10^{1}	+	+
GW190630_185205	1245955943	1.80×10^{9}	1.20×10^{9}	+	+
GW190701_203306	1246048404	4.82×10^{1}	3.21×10^{1}	+	+
GW190706_222641	1246487219	5.84×10^{3}	5.93×10^{3}	+	+
GW190707_093326	1246527224	9.32×10^{13}	6.22×10^{13}	+	+
GW190708_232457	1246663515	8.10×10^{2}	5.40×10^{2}	+	+
GW190720_000836	1247616534	5.66×10^{6}	3.77×10^{6}	+	+
GW190725_174728	1248112066	7.14×10^{-1}	1.49	–	+
GW190727_060333	1248242631	9.15×10^{8}	6.10×10^{8}	+	+
GW190728_064510	1248331528	4.61×10^{14}	3.07×10^{14}	+	+
GW190803_022701	1248834439	4.06	3.54	+	+
GW190814_211039	1249852257	4.62×10^{10}	3.08×10^{10}	+	+
GW190828_063405	1251009263	4.98×10^{25}	3.32×10^{25}	+	+
GW190828_065509	1251010527	1.03×10^{4}	8.82×10^{3}	+	+
GW190910_112807	1252150105	8.72×10^{1}	5.81×10^{1}	+	+
GW190915_235702	1252627040	3.61×10^{4}	2.62×10^{4}	+	+
GW190924_021846	1253326744	4.97×10^{8}	3.31×10^{8}	+	+
GW190925_232845	1253489343	3.47×10^{1}	7.72×10^{1}	+	+
GW190929_012149	1253755327	1.72	1.67×10^{3}	+	+
GW190930_133541	1253885759	3.56×10^{1}	1.74×10^{3}	+	+

TABLE I: The list of O3a events passing the threshold of $IFAR > 1\ yr$ for the combination of LVK pipelines (PyCBC, GstLAL, MBTA) and with the addition of the IAS and the OGC pipelines to the combination (labeled LVK+).

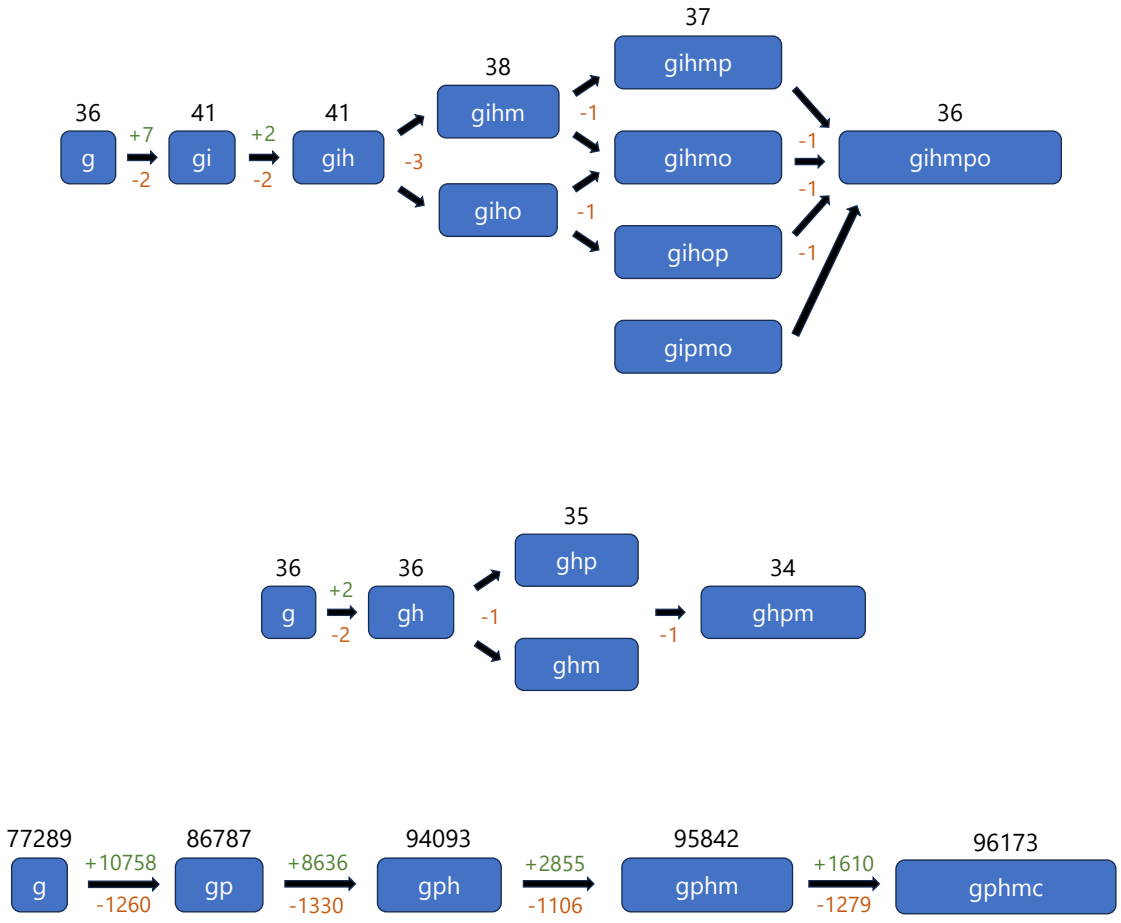

FIG. 1: Combinations of pipelines that detect the most events for each number of searches. GstLAL – "g", PyCBC – "p", PyCBC_highmass – "h", MBTA – "m", CWB – "c", IAS – "i", and OGC – "o". Top: Combinations of LVK, IAS, and OGC pipelines based on O3a data. Middle: Combinations of LVK pipelines based on O3a data. Bottom: LVK pipelines based on the injection set. Numbers above the boxes show the number of candidates passing the threshold of $IFAR > 1\ yr$, and numbers above and below the arrows indicate the number of events lost and gained due to the addition of an extra pipeline to the calculation of the combined IFARs.

[1] LIGO Scientific Collaboration and Virgo Collaboration. GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev. D*, 93:122003, Jun 2016.

[2] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *arXiv:2111.03606*, Nov 2021.

[3] https://observing.docs.ligo.org/plan.

[4] Floor S. Broekgaarden. ChatGPT scores a bad birdie in counting gravitational-wave chirps. *arXiv:2303.17628*, Apr 2023.

[5] C. Capano, T. Dent, C. Hanna, et al. Systematic errors in estimation of gravitational-wave candidate significance. *Phys. Rev. D*, 96:082002, Oct 2017.

[6] S. A. Usman, A. H. Nitz, I. W. Harry, et al. The PyCBC search for gravitational waves from compact binary coalescence. *Classical and Quantum Gravity*, 33(21):215004, Oct 2016.

[7] S. Sachdev, S. Caudill, H. Fong, et al. The GstLAL Search Analysis Methods for Compact Binary Mergers

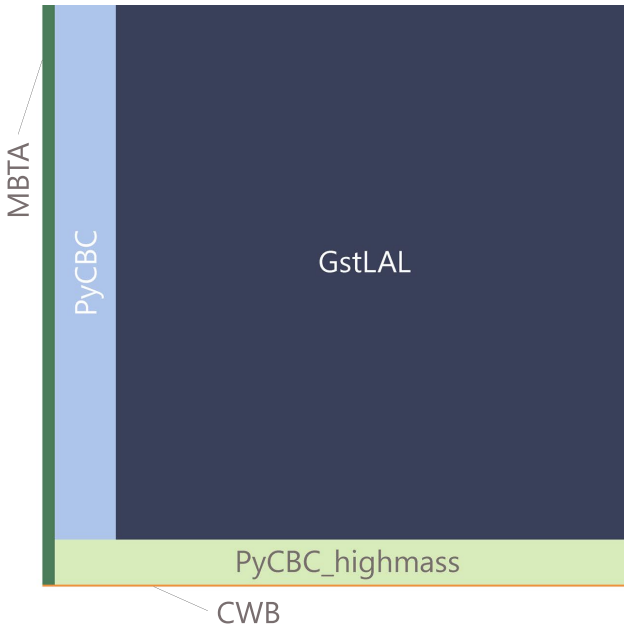

FIG. 2: Increase in the number of events passing the threshold upon the addition of results from more pipelines to the calculation of the combined IFAR for the injection set. An illustration of the numerical information given in the bottom part of Fig. 1. For example, the dark blue box shows the number of events detected by GstLAL, and the combination of a dark blue and light blue sections gives the number of events detected by the combination of two pipelines.

in Advanced LIGO's Second and Advanced Virgo's First Observing Runs. *arXiv:1901.08580*, Jan 2019.

[8] F Aubin, F Brighenti, R Chierici, D Estevez, G Greco, G M Guidi, V Juste, F Marion, B Mours, E Nitoglia, and et al. The mbta pipeline for detecting compact binary coalescences in the third ligo–virgo observing run. *Classical and Quantum Gravity*, 38(9):095004, 2021.

[9] T. Venumadhav, B. Zackay, J. Roulet, et al. New search pipeline for compact binary mergers: Results for binary black holes in the first observing run of Advanced LIGO. *Phys. Rev. D*, 100:023011, Jul 2019.

[10] Alexander H. Nitz, Sumit Kumar, Yi-Fan Wang, Shilpa Kastha, Shichao Wu, Marlin Schäfer, Rahul Dhurkunde, and Collin D. Capano. 4-OGC: Catalog of gravitational waves from compact-binary mergers. Dec 2021.

[11] LIGO Scientific Collaboration and Virgo Collaboration. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Phys. Rev. X*, 9:031040, Sep 2019.

[12] Daniel J. Wilson. The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences*, 116(4):1195–1200, 2019.

[13] S. Banagiri, C. P. L. Berry, G. S. Cabourn Davies, et al. A Unified $p_{\rm astro}$ for Gravitational Waves: Consistently Combining Information from Multiple Search Pipelines. *arXiv:2305.00071*, May 2023.

[14] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Candidate data release, Nov 2021.

[15] https://zenodo.org/record/5546665.

[16] Seth Olsen, Tejaswi Venumadhav, Jonathan Mushkin, Javier Roulet, Barak Zackay, and Matias Zaldarriaga. New binary black hole mergers in the ligo-virgo o3a data. *Phys. Rev. D*, 106:043009, Aug 2022.

[17] https://github.com/seth-olsen/new_BBH_mergers_O3a_IAS_pipeline.

[18] https://github.com/gwastro/4-ogc.

[19] The LIGO Scientific Collaboration and the Virgo Collaboration. Gwtc-2.1: Deep extended catalog of compact binary coalescences observed by ligo and virgo during the first half of the third observing run, 2022.

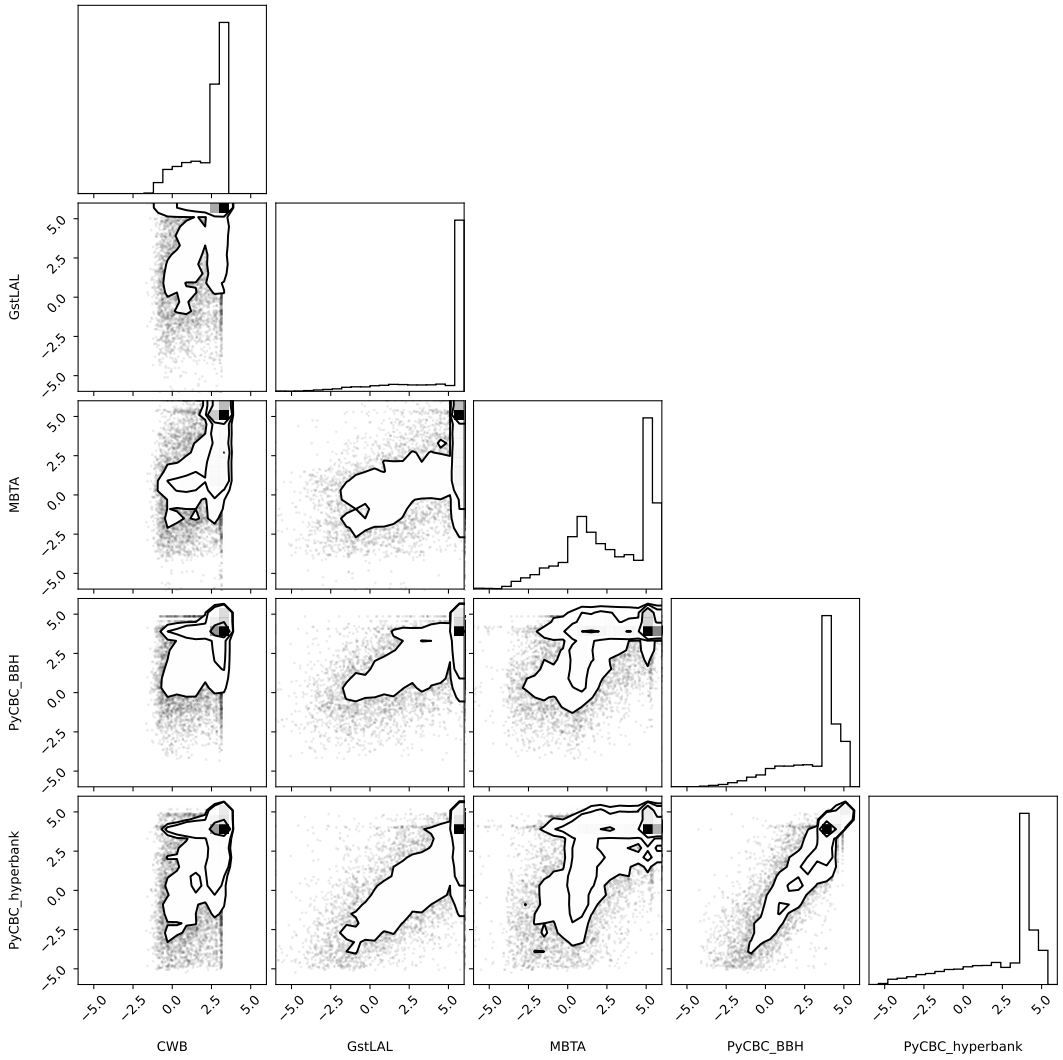

FIG. 3: Qualitative analysis of correlation between IFAR (yr) distributions of different LVK pipelines.