

Systematic bias in dark siren statistical methods and its impact on Hubble constant measurement

Viviane Alfradique¹*

Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud 150, 22290-180 Rio de Janeiro, RJ, Brazil

Clécio R. Bom

Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud 150, 22290-180 Rio de Janeiro, RJ, Brazil

and Centro Federal de Educação Tecnológica Celso Suckow da Fonseca,

Rodovia Márcio Covas, lote J2, quadra J—Itaguaí (Brazil)

Tiago Castro²

INAF—Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131 Trieste, Italy;

INFN—Sezione di Trieste, I-34100 Trieste, Italy; IFPU—Institute for Fundamental Physics of the Universe, via Beirut 2, 34151, Trieste, Italy; and ICSC—Centro Nazionale di Ricerca in High Performance Computing,

Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy



(Received 25 March 2025; accepted 12 September 2025; published 29 September 2025)

The advent of multimessenger cosmology, marked by the detection of GW170817, demonstrated that standard sirens are a valuable cosmological probe. In the absence of an electromagnetic counterpart identification, gravitational waves (GWs) carry valuable information through the dark siren (DS) approach, where the source redshift is estimated using galaxy catalogs of potential hosts within the localization volume. However, the DS analysis can be affected by galaxy catalog incompleteness at the limits of gravitational-wave detectability, potentially introducing biases in the constraints on cosmological parameters. Focusing on GWs from binary black holes (BBH) detected by the LIGO, Virgo, and KAGRA (LVK) collaboration, we explore the possible systematic biases in the measurement of the Hubble constant (H_0). These biases may arise from (1) the incompleteness of catalogs due to the apparent magnitude thresholds of optical telescope sensitivity, and (2) the use of incorrect weighting schemes (e.g., using star formation or stellar mass as tracers of the host galaxy) for each potential host. We found that an unbiased estimate of H_0 can be obtained when the corrected weighting scheme is applied to a complete or volume-limited catalog. We use a complete galaxy catalog covering 90% of the localization probability ($A_{90\%}$) for each GW detection, employing stellar mass as a tracer. Our results show that a sample of 100 binary black hole events with $A_{90\%} < 10 \text{ deg}^2$ and measured luminosity distances below 1600(2500) Mpc, detected by the LVK at O4(O5) sensitivity, can provide a percent-level measurement of H_0 , with a precision of approximately 3%(1%). This number of detections is expected to be accumulated after approximately 8 and 3 years of observations with the LVK at O4 and O5 sensitivity, respectively. The O5 run provides a reduction in the H_0 uncertainty by 1.34 km/s/Mpc compared to the O4-like configuration. The H_0 precision increases to approximately 6% when it is assumed that every galaxy has an equal probability of being the host.

DOI: 10.1103/vd36-3mys

I. INTRODUCTION

The detection of the first gravitational wave, GW150914 [1], opened the field of gravitational-wave astronomy. Since that milestone, gravitational waves have emerged as a powerful tool for investigating the expansion history of the Universe and offer insights into the mass and spin distributions of gravitational wave sources [2]. The

expansion history can be analyzed using gravitational waves as *standard sirens* initially proposed by Schutz in [3]. In the standard sirens context, the GWs alone offer a direct measurement of the luminosity distance. However, the redshift cannot be directly obtained from the GWs, as it is degenerate with the source masses and must therefore be inferred from an associated electromagnetic counterpart. Until now, the only bright standard siren detected is the GW170817 [4], where the gravitational waves emerge from a merger of two neutron stars at a distance of 40^{+8}_{-14} Mpc.

*Contact author: vivianeapa@cbpf.br

Approximately two seconds after the merger, a short gamma-ray burst was detected by two distinct space telescopes: the Fermi Gamma-ray Burst Monitor [5] and the International Gamma-ray Astrophysics Laboratory [6]. Additionally, an optical kilonova (e.g., Arcavi *et al.* [7], Coulter *et al.* [8], Cowperthwaite *et al.* [9], Soares-Santos *et al.* [10], Chornock *et al.* [11], Kasliwal *et al.* [12], Nicholl *et al.* [13], Evans *et al.* [14], Pian *et al.* [15], Smartt *et al.* [16], Tanvir *et al.* [17], Valenti *et al.* [18]) was identified within the region of galaxy NGC 4993, confirming it as the host galaxy. This opens up the possibility of exploring cosmology through gravitational waves, providing a new independent measurement of the local expansion rate of the Universe, known as the Hubble constant, with an uncertainty of 15% [19].

Despite the success of GW170817 to probe cosmology, the absence of other standard sirens, combined with the low number of GW detections from binary neutron stars mergers [20], discourages their potential to dominate the field of gravitational-wave cosmology. As an alternative, cosmological information can be explored using the “dark siren” technique. As the name suggests, in this method the electromagnetic counterpart is assumed to be undetectable, and the source redshift is inferred from a statistical technique. The DS approach is further categorized into several distinct methods: (1) the spectral mass that relies on mapping the observed mass spectrum, the redshifted mass, i.e., $m_z = (1+z)m$, in luminosity distance to infer the redshift source [21–25], (2) the cross-correlation ones that explored the correlation between the GWs and galaxies distributions to estimate the host galaxy [26–30], and (3) the use of all flares in AGNs that might be associated to a BBH merger [31,32]. Apart from this, a promising way to identify the host galaxy is to focus on the galaxy catalog [3,33–35], where all the galaxies within the GW localization volume are identified as potential hosts. A probabilistic weight is assigned to each potential host based on its position, which is statistically marginalized over. This approach, known as the *statistical* method, has already been fully explored with real events of gravitational waves [35–41], the latest measurement [41] showed that from a catalog with 15 dark sirens it is possible to find a measure of H_0 with precision of 18% and when combined with the bright siren GW170817 is reduced to 6%.

The expectation is that the next generation of interferometers leads to a H_0 precision of $<1\%$ [42–44], which will require more rigorous control over potential systematic effects in DS cosmology. The study by [45] provides an extensive analysis, demonstrating that the statistical DS method, when applied to consistent simulated data, yields cosmological measurements without bias. In a mock data analysis, [46] also showed that the method is capable of providing unbiased inference for H_0 . The study further explored the effects of galaxy incompleteness and the weighting of host galaxies on cosmological inferences.

Recently studies [47,48] investigated the influence on H_0 when different galaxy properties are used as hosting probability, both works showed that a systematic bias is introduced into the H_0 estimation if the host probability is mismatched. Another potential source of systematic errors in the DS analysis is the mismodeling of the galaxy redshift uncertainty. In their study, [49] addressed this issue by modeling the redshift uncertainty using a Gaussian and a modified Lorentzian function. They showed that the systematic errors introduced by this approach are much smaller than the statistical ones. On the other hand, the probability distribution function of the photometric redshift for galaxy catalogs from wide-field surveys can be estimated using deep learning techniques [50–52]. The resulting distribution may exhibit more complexity than a simple Gaussian or Lorentzian form, highlighting the need to explore these systems beyond traditional models.

In this work, we investigate the systematic biases in the measurement of H_0 that may arise from statistical DS analyses. We focus on the effects of magnitude-limited catalogs and incorrect weighting of GW host galaxy probabilities, particularly those based on stellar mass (SM) and star formation rate (SFR). Although previous studies have explored these issues, our analysis extends the investigation in several key aspects. We used simulated sky maps from BAYESTAR [53], allowing a more realistic representation of measurements derived from gravitational wave detections. Furthermore, we go beyond traditional magnitude-limited samples by incorporating a volume-limited sample and investigating the bias that could arise from galaxy catalog incompleteness. In contrast to the analytical approximations often used in prior works, we quantify the selection function through a simulation-based approach, enhancing the robustness of our results. Finally, we analyze the impact on the precision of H_0 due to uncertainties in the GW luminosity distance, the 90% credible sky area, and the GW redshift range. This paper is organized as follows: In Sec. II, we present the statistical DS methodology. In Sec. III, we describe the simulation prescription for gravitational wave detections with their corresponding sky maps, and the mock galaxy catalog used. Our results are presented in Sec. IV, and we summarize and conclude in Sec. V.

II. DARK SIREN BAYESIAN ANALYSIS

Using the Bayes’ theorem, the H_0 posterior for a given i th GW event, with observed data d_{GW}^i , can be written as [45]

$$p(H_0|d_{\text{GW}}^i) \propto p(H_0)\mathcal{L}(d_{\text{GW}}^i|H_0), \quad (1)$$

$$\mathcal{L}(d_{\text{GW}}^i|H_0) \propto \frac{\int dz \mathcal{L}_{\text{GW}}(\hat{d}_L^i|d_L(z, H_0)) p_{\text{CBC}}(z)}{\int dz P_{\text{det}}^{\text{GW}}(z, H_0) p_{\text{CBC}}(z)}, \quad (2)$$

where $P_{\text{det}}^{\text{GW}}$ is the GW detection probability and it describes the probability of a GW source located at a redshift z to be detected based on specific detection conditions, $\mathcal{L}_{\text{GW}}$ is the GW likelihood of measuring a luminosity distance $\hat{d}_L$ given a H_0 value and a galaxy at a redshift z , and $p(H_0)$ is the prior distribution. p_{CBC} is the compact binary coalescence probability.

As defined in [45], the detection probability $P_{\text{det}}^{\text{GW}}(z, H_0)$ accounts for the fact that the analysis is limited to GW events detected above a certain threshold, d_{thr} . It is expressed as the integral of the GW likelihood over the data d_{GW} , conditioned on d_{GW} exceeding the detection threshold d_{thr} . The calculation of the term $P_{\text{det}}^{\text{GW}}$ requires integrating over all the realizations of GW events based on the binary population model parameters, considering different values of z and H_0 , and applying the detection condition. To achieve this, we follow the approach outlined in [46] and employ a Monte Carlo integration defined as

$$P_{\text{det}}^{\text{GW}}(z, H_0) = \frac{1}{N_{\text{GW}}} \sum_{i=1}^{N_{\text{GW}}} P_{\text{det}}^{\text{GW}}(d_{\text{GW}}^i, z, H_0), \quad (3)$$

where d_{GW} represents the GW data corresponding to an event (whether detected or not), which is associated with source parameters (such as inclination, right ascension, declination, component masses, and polarization) that may influence the detectability of the event. In the equation above, we have excluded the dependence on the sky position, assuming that the probability is approximately uniform across the sky. The term $P_{\text{det}}^{\text{GW}}(d_{\text{GW}}^i, z, H_0)$ is assumed to be equal to one if the GW event d_{GW}^i satisfies the detection condition, and zero otherwise. The detection criterion is that the network signal-to-noise ratio (SNR) must be greater than 12 and 45 for O4 and O5 runs, respectively; the measured luminosity distance must be below 1600(2500) Mpc for the O4(O5) observing run; and an additional condition is applied to the sky localization area to ensure the selection of well-localized events. For this, we consider two configurations: $A_{90\%} < 10 \text{ deg}^2$ or $A_{90\%} < 100 \text{ deg}^2$. Thus, $P_{\text{det}}^{\text{GW}}(z, H_0)$ is a smooth function whose values range from 1 to 0. The Monte Carlo equation, as defined above, was computed using simulations of 100,000 GW events from BBH mergers, following the procedure and assumptions for the population properties described in Sec. III B. These GW simulations were conducted for 17 distinct values of H_0 within the prior range of [40, 120] km/s/Mpc. The resulting function was then interpolated to ensure it was continuously defined across the entire prior range. We computed the $P_{\text{det}}^{\text{GW}}(z, H_0)$ function for each adopted detection condition configuration. Figure 1 presents $P_{\text{det}}^{\text{GW}}(z, H_0)$ as a function of z , computed for different values of H_0 values within the redshift range of interest for this analysis, assuming O4 and O5 scenarios.

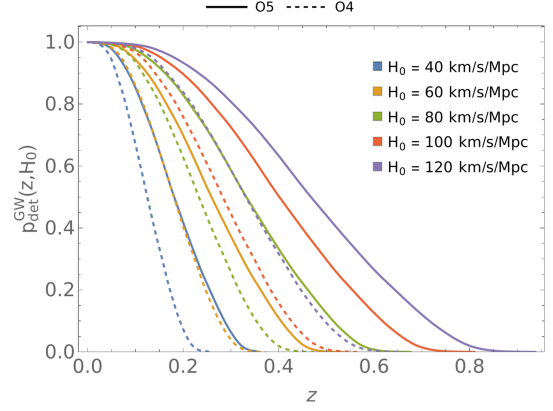


FIG. 1. GW detection probability as a function of z , with each color representing a different value of H_0 . These results were generated using the design sensitivity curves for O4 (dashed lines) and O5 (solid lines), a BBH population, and assuming that the detection criteria are a network SNR above 12, a 90% confidence localization area ($A_{90\%}$) smaller than 100 deg^2 , and measured luminosity distance below 1600 Mpc for O4 (2500 Mpc for O5).

The term p_{CBC} can be modeled as described in Borghi *et al.* [25], Perna *et al.* [48], Gray *et al.* [54], Mastrogiovanni *et al.* [55],

$$p_{\text{CBC}}(z) = \frac{\int dx_{\text{gal}} p_{\text{rate}}(z, x_{\text{gal}}) p_{\text{cat}}(z, x_{\text{gal}})}{\int dz dx_{\text{gal}} p_{\text{rate}}(z, x_{\text{gal}}) p_{\text{cat}}(z, x_{\text{gal}})} \quad (4)$$

with p_{rate} represent the probability that a galaxy with redshift z and specific properties x_{gal} hosts a GW event and p_{cat} is the probability of a galaxy with these properties being located at redshift z , essentially corresponding to the galaxy catalog probability. Here, we model the GW host probability p_{rate} based on the galaxy properties, specifically the stellar mass M_* or the star formation rate ρ_{SFR} , i.e., $x_{\text{gal}} = \{M_*, \rho_{\text{SFR}}\}$. This probability is defined in the detector frame; therefore, it must be multiplied by $(1+z)^{-1}$ to account for the time dilation caused by the expansion of the Universe between the source and the detector frames. We assume that the galaxy catalog is complete and provides the precise values of $\{M_*, \rho_{\text{SFR}}, z\}$, such that p_{cat} can be written as a delta function

$$p_{\text{cat}}(z, x_{\text{gal}}) \approx \frac{1}{N_{\text{gal}}} \sum_{i=1}^{N_{\text{gal}}} \delta(z - \hat{z}_{\text{gal}}^i) \delta(x_{\text{gal}} - \hat{x}_{\text{gal}}^i), \quad (5)$$

where N_{gal} is the number of galaxies in the catalog. The marginal GW likelihood was computed assuming a Gaussian function as described in [53]

$$\mathcal{L}_{\text{GW}}(\hat{d}_L^j | d_L(z, H_0), \hat{\Omega}_i) \approx \frac{p(\hat{\Omega}_i)}{\sqrt{2\pi}\sigma(\hat{\Omega}_i)} \times \exp \left[-\frac{(d_L(z, H_0) - \hat{d}_L^j(\hat{\Omega}_i))^2}{2\sigma^2(\hat{\Omega}_i)} \right], \quad (6)$$

where the $p(\hat{\Omega}_i)$ is the position probability, while $\hat{d}_L^j(\hat{\Omega}_i)$ and $\sigma(\hat{\Omega}_i)$ denote the mean and standard deviation of the luminosity distance, respectively. The solid angle of the i th HEALPIX pixel is given by $\hat{\Omega}_i$.

By substituting Eq. (5) into Eq. (4), we obtain the final form of $p_{\text{CBC}}(z)$ as

$$p_{\text{CBC}}(z) \propto \frac{1}{N_{\text{gal}}} \sum_{i=1}^{N_{\text{gal}}} p_{\text{rate}}(z, \hat{x}_{\text{gal}}^i) \delta(z - \hat{z}_{\text{gal}}^i), \quad (7)$$

that simplifies the likelihood Eq. (2) for the measured luminosity distance corresponding to the j th gravitational wave event to

$$\mathcal{L}(d_{\text{GW}}^j | H_0) \propto \sum_{i=1}^{N_{\text{gal}}} \mathcal{L}_{\text{GW}}(\hat{d}_L^j | d_L(\hat{z}_{\text{gal}}^i, H_0), \hat{\Omega}_i) p_{\text{rate}}(\hat{z}_{\text{gal}}^i, \hat{x}_{\text{gal}}^i) / \sum_{i=1}^{N_{\text{gal}}} P_{\text{det}}^{\text{GW}}(\hat{z}_{\text{gal}}^i, H_0) p_{\text{rate}}(\hat{z}_{\text{gal}}^i, \hat{x}_{\text{gal}}^i), \quad (8)$$

where $\hat{\Omega}_i$ represents the solid angle of the HEALPIX pixel in which the i th galaxy is located.

For a set of N_{GW} luminosity distances observed from gravitational waves, $\{\hat{d}_L^j\} (\equiv \{d_{\text{GW}}^j\})$, and assuming the independence of each data, the above equation is extended to

$$\mathcal{L}(\{\hat{d}_L^j\} | H_0) \propto \prod_{j=1}^{N_{\text{GW}}} \mathcal{L}(\hat{d}_L^j | H_0). \quad (9)$$

III. MOCK DATA

A. Mock galaxy catalog

This study employs the Magneticum galaxy catalog [56,57] as our mock galaxy dataset. The Magneticum simulation suite contains several cosmological hydrodynamic simulations that account for the evolution of dark and luminous matter under the influence of galaxy and baryonic processes. The galaxy catalogs used in this work were extracted from the boxes 2 and 2b [58]. These simulations are performed within a comoving periodic volume of $(500 \text{ Mpc})^3$ for box 2 and a 6 times larger volume for box 2b. The light cone covers an area of 5143 deg^2 (a 1/8th of a sky-wedge) and extends to a redshift of $z = 1.3$. The simulations assumes a

standard Λ CDM cosmological model, with the following parameters: $\Omega_m = 0.272$, $\Omega_b = 0.0456$, $\Omega_\Lambda = 0.728$, $n = 0.963$, $H_0 = 70.4 \text{ km/s/Mpc}$, and $\sigma_8 = 0.809$. The light cone was generated by extracting galaxies from snapshots of the simulation at various redshifts, replicating the box when needed to ensure a consistent representation of the cosmic volume across the observational limits. The galaxy properties, including stellar mass, star formation rate, metallicity, and absolute magnitude, were derived directly from the simulation output. These quantities were computed based on detailed models of gas dynamics, stellar evolution, and feedback processes, providing a comprehensive view of the galaxy population at different epochs. The Magneticum catalog offers a diverse range of galaxy properties, which are crucial for our analysis to consistently describe the large-scale structure, galaxy evolution, and cosmological parameter estimation.

B. GW simulations

In order to compute the H_0 posterior under different observational conditions, the Bayesian framework described in Sec. II was applied to a catalog of simulated BBH gravitational wave detections, along with galaxy information from the Magneticum catalog. The gravitational wave detections were simulated using the Python package LIGO.SKYMAP [59], which includes the BAYESTAR code, an efficient tool for reconstructing gravitational wave sky localization. We perform an end-to-end simulation of 100,000 BBH mergers for the LIGO and Virgo network detector with sensitivity comparable to that attained in the O4 and O5 runs. The amplitude spectral densities used for the O4 run are available at [60]. For the O5 run, we used the public sensitivity curves presented in [61] and available at [62]. For the LIGO detectors, we used the curve labeled as “AplusDesign”; for Virgo, “avirgo_O5low_NEW”; and for KAGRA, “kagra_80 Mpc.” The BBH population follows a mass distribution defined by the power law plus peak model with parameters fixed to the best values found in [2].

To simulate an injected GW event, a galaxy from the Magneticum catalog was first selected as the host, with the GW host probability p_{rate} weighted by stellar mass, star formation rate or equally, i.e., $p_{\text{rate}} \propto \{M_\star, \rho_{\text{SFR}}, 1\}$. In this way, the 3D position (right ascension, declination, and redshift) of each injected GW event was set to match the true sky position of its host galaxy. We assume the IMRPhenomD frequency-domain waveform model to produce the GW signals using the BAYESTAR software. After generating the injections, a matched-filter analysis was performed to compute the SNR for each injected event. An event was considered detected if the network SNR is above 12 and at least 2 detectors have a single-detector SNR above 4. Finally, the 3-dimensional sky map was reconstructed for all the detections GW event with the BAYESTAR. To select the best events for the DS analysis in each observing run, we adopted the following criteria:

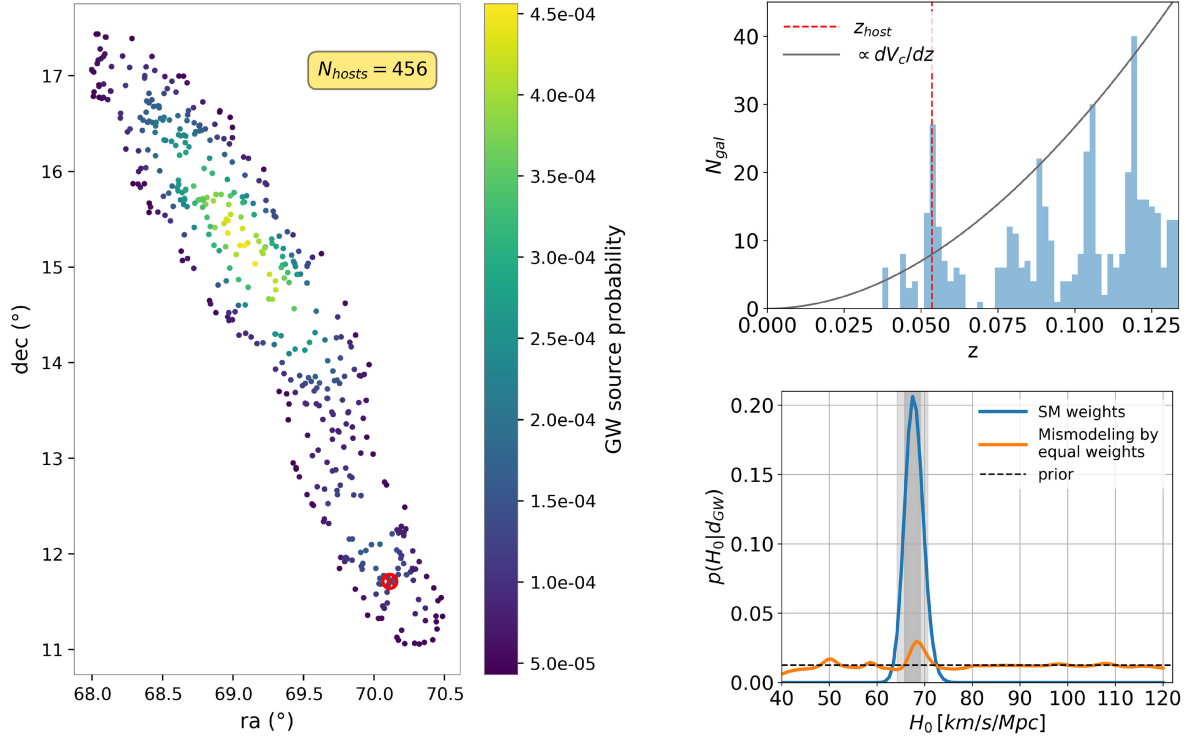


FIG. 2. Left: the sky localization of the potential host galaxies of a GW example case from our mock catalog. Each point represents a galaxy, colored by the positional probability of the GW source, computed with BAYESTAR. The red circle with a cross indicates the true position of the host galaxy. Right: the redshift distribution of the potential host galaxies, extracted from Magneticum for this particular GW event, is shown in the top panel. The dashed red line represents the z value of the true host, and the gray solid line represents a uniform distribution in comoving volume. The bottom panel shows the H_0 posterior for the DS method, assuming, incorrectly, equal weights (in orange) and weighting the galaxies by their stellar mass (in blue).

(1) for the O4 run, GW events with an measured luminosity distance $d_{L,\text{mea}} < 1600$ Mpc and an SNR above 12, and (2) for the O5 run, GW events with an measured luminosity distance $d_{L,\text{mea}} < 2500$ Mpc and an SNR above 45. The choice to use the measured, rather than the true, luminosity distance ensures consistency with the hierarchical Bayesian likelihood and mitigates potential biases in the analysis, as discussed in [63]. In this way, despite the O5 configuration selecting more distant GW events, possibly with larger uncertainties in luminosity distance, the more restrictive SNR cut ensures that the distribution of luminosity distance uncertainties is equivalent in both observing run configurations. The GW events used in the analysis were also selected based on their 90% confidence level (CL) sky localization area. The detection configurations were divided into: (1) events with $A_{90\%} < 100$ deg², and a more restrictive set, (2) events with $A_{90\%} < 10$ deg².

After constructing the three-dimensional sky map for the gravitational wave detection sample, we perform a cross-match between the resulting localizations and the Magneticum galaxy catalog. Although the sky-localizations provide the luminosity distance posterior along each line-of-sight, most wide survey catalogs that probe the universe

at extragalactic distances lack precise distance measurements, primarily due to their dependence on an intrinsic calibration process (as seen in the *Standard Candles*). Therefore, to obtain results closer to the observed universe, we constructed a three-dimensional subcatalog containing all galaxies within the 90% credible area. This subcatalog includes galaxies whose redshift values correspond to the 90% CL bounds of the luminosity distance, with the minimum and maximum redshifts determined by assuming the lower and upper values of H_0 from our prior, respectively. In our 100 GW subcatalogs, we find that, on average, approximately four GW events have their true host galaxies located outside the 90% CL. Although these true hosts are not included in the 90% CL, the posterior obtained using the 90% CL remains statistically consistent with that obtained utilizing the broader 99% CL. As an example, the left panel of Fig. 2 presents all the galaxies within the 90% CL of the sky localization area for a specific simulated GW event. The top panel of Fig. 2 shows the redshift distribution of these potential host galaxies, with the true redshift of the host galaxy indicated by a dashed line. Finally, we applied this subcatalog to each simulated GW detection using the statistical method outlined in

Sec. II to estimate the Hubble constant. To explore the effect of galaxy incompleteness on the H_0 measurement, no corrections for completeness were applied in the calculations. The H_0 posterior for the example GW event illustrated in Fig. 2 is shown in the bottom panel. In these results, we see that, as the position of the true host galaxy coincides with a region of overdensity, the H_0 posterior distribution is consistent with the fiducial value of 70.4 km/s/Mpc.

IV. RESULT AND DISCUSSION

In this section, we present our estimation of the Hubble constant for the LVK network during the O4 and O5 runs, using the simulated mock data in the statistical framework

described in Sec. II. We performed an analysis considering four distinct configurations of apparent magnitude completeness, characterized by magnitude cuts of $m < \{22, 21, 20, 19\}$, two detector sensitivities mimicking the O4 and O5 runs, and three different weighting schemes for the host galaxy. These schemes assume: (1) equal probability for all galaxies in the catalog to host the GW source, (2) stellar mass, and (3) star formation as a tracer for the host galaxy. In practice, since stellar mass and star formation are not directly observed, it is common to use a galaxy's K-band and B-band luminosities as proxies for stellar mass and star formation, respectively. For each configuration, we performed 30 independent realizations

TABLE I. H_0 measurements for different GWs detections configurations (a weighting host galaxies by the stellar mass or equal weighting, different apparent magnitude thresholds, and two cuts over the measured luminosity distance, $d_{L,\text{mea}} < \{1600, 2500\}$ Mpc) across 30 realizations of a sample of 100 GWs with $A_{90\%} < 10 \text{ deg}^2$, except where explicitly stated otherwise. The galaxy completeness is defined as written in Eq. (11). The Hubble parameter H_0 , its uncertainties, and the bias (ΔH_0) are given in units of km/s/Mpc. The H_0 bias is defined in Eq. (10).

	Mth	Skymap coverage	Completeness	H_0	σ_{H_0}	$\frac{\sigma_{H_0}}{H_0}$	ΔH_0	$\frac{\Delta H_0}{\sigma_{H_0}}$
$d_{L,\text{mea}} < 1600 \text{ Mpc}$ and SNR > 12-O4								
No weighting	...	100%	100%	$70.60^{+5.56}_{-3.41}$	4.49	0.06	1.84 ± 3.43	0.41 ± 0.76
No weighting ($A_{90\%} < 100 \text{ deg}^2$)	...	100%	100%	$71.00^{+10.40}_{-7.72}$	9.06	0.13	0.68 ± 5.78	0.08 ± 0.64
No weighting (magnitude-limited)	22.0	99.9%	98%	$70.00^{+4.42}_{-2.86}$	3.64	0.05	0.98 ± 1.94	0.27 ± 0.53
No weighting (magnitude-limited)	21.0	99.4%	86%	$70.00^{+4.07}_{-3.16}$	3.61	0.05	0.55 ± 2.07	0.15 ± 0.57
No weighting (magnitude-limited)	20.0	91.7%	56%	$50.50^{+13.81}_{-1.19}$	7.50	0.15	-13.84 ± 7.58	-1.85 ± 1.01
No weighting (magnitude-limited)	19.0	58.9%	24%	$40.0^{+10.62}_{-0.00}$	5.31	0.13	-22.24 ± 5.24	-4.19 ± 0.99
No weighting (volume-limited)	22.0	99.9%	89%	$71.52^{+4.82}_{-3.11}$	3.96	0.06	0.18 ± 1.71	0.05 ± 0.43
No weighting (volume-limited)	21.0	95.7%	63%	$71.52^{+2.99}_{-2.90}$	2.94	0.04	0.21 ± 1.77	0.07 ± 0.60
No weighting (volume-limited)	20.0	61.5%	29%	$71.52^{+2.37}_{-2.79}$	2.58	0.04	0.24 ± 1.74	0.09 ± 0.67
No weighting (volume-limited)	19.0	25.3%	10%	$71.52^{+3.15}_{-1.96}$	2.38	0.03	0.32 ± 1.78	0.13 ± 0.75
SM weighting	...	100%	100%	$70.50^{+2.40}_{-2.17}$	2.28	0.03	0.39 ± 1.36	0.17 ± 0.60
SM mismodeling by equal	...	100%	100%	$70.50^{+4.07}_{-2.65}$	3.36	0.05	0.75 ± 2.00	0.22 ± 0.07
SM mismodeling by SFR	...	100%	100%	$70.50^{+3.63}_{-2.47}$	3.05	0.04	0.71 ± 1.90	0.23 ± 0.62
Equal mismodeling by SM	...	100%	100%	$71.00^{+5.21}_{-4.23}$	4.72	0.07	3.62 ± 8.68	0.77 ± 1.84
$d_{L,\text{mea}} < 2500 \text{ Mpc}$ and SNR > 45-O5								
No weighting	...	100%	100%	$70.71^{+1.47}_{-2.47}$	1.97	0.03	0.38 ± 1.51	0.19 ± 0.77
No weighting ($A_{90\%} < 100 \text{ deg}^2$)	...	100%	100%	$69.90^{+6.14}_{-2.07}$	4.11	0.06	0.89 ± 3.86	0.22 ± 0.94
No weighting (magnitude-limited)	22.0	99.9%	89%	$70.71^{+0.84}_{-2.86}$	1.85	0.03	-0.14 ± 1.41	-0.08 ± 0.76
No weighting (magnitude-limited)	21.0	95.7%	63%	$68.28^{+2.38}_{-1.77}$	2.08	0.03	-3.64 ± 7.46	-1.75 ± 3.59
No weighting (magnitude-limited)	20.0	61.5%	29%	$67.47^{+0.72}_{-14.36}$	7.54	0.11	-6.97 ± 7.64	-0.92 ± 1.01
No weighting (magnitude-limited)	19.0	25.3%	10%	$44.85^{+18.77}_{-0.16}$	9.46	0.21	-16.03 ± 8.33	-1.69 ± 0.88
No weighting (volume-limited)	22.0	99.9%	89%	$70.71^{+1.19}_{-2.62}$	1.91	0.03	0.22 ± 1.64	0.12 ± 0.86
No weighting (volume-limited)	21.0	95.7%	63%	$70.71^{+1.26}_{-2.45}$	1.86	0.03	0.22 ± 1.64	0.12 ± 0.88
No weighting (volume-limited)	20.0	61.5%	29%	$70.71^{+1.29}_{-2.34}$	1.81	0.03	0.22 ± 1.64	0.12 ± 0.91
No weighting (volume-limited)	19.0	25.3%	10%	$70.71^{+1.23}_{-2.07}$	1.65	0.02	0.22 ± 1.37	0.13 ± 0.83
SM weighting	...	100%	100%	$70.71^{+0.68}_{-1.41}$	1.04	0.01	0.30 ± 0.57	0.29 ± 0.55
SM mismodeling by equal	...	100%	100%	$70.71^{+2.95}_{-1.95}$	2.45	0.03	0.79 ± 2.45	0.32 ± 1.00
SM mismodeling by SFR	...	100%	100%	$70.71^{+0.51}_{-2.03}$	1.27	0.02	-0.18 ± 0.42	-0.14 ± 0.33
Equal mismodeling by SM	...	100%	100%	$69.09^{+7.06}_{-1.07}$	4.07	0.06	5.16 ± 16.13	1.27 ± 3.96

of the combined posteriors on H_0 using a sample of 100 DSs, analyzing the precision and quantifying the biases relative to the injected H_0 value. This approach ensures that our results are not dependent on a single realization of the population, allows us to assess the statistical fluctuations due to population realizations, and leads to more robust conclusions. We adopt a uniform prior on H_0 in the range $[40, 120]$ km/s/Mpc.

In Table I, we present the final constraint on the Hubble constant for DSs selected based on the 90% CL of the localization probability being below 10 deg^2 , except where explicitly stated otherwise. We define the bias of the H_0 estimator as the mean difference between the peak of the combined posterior, obtained from a sample of 100 GWs events, and the fiducial value from the *Magneticum* simulation ($H_0^{\text{fid}} = 70.4 \text{ km/s/Mpc}$), computed over 30 independent realizations. Specifically, let $\hat{H}_0^{(i)}$ denote the posterior peak from the combined posterior in the i th realization. The bias is then given by

$$\Delta H_0 = \frac{1}{30} \sum_{i=1}^{30} (\hat{H}_0^{(i)} - H_0^{\text{fid}}). \quad (10)$$

The final H_0 measurement was obtained by averaging the posterior distributions from all 30 realization of a sample of 100 GWs. For the computation of the completeness fraction [see its definition in Eq. (11)], we used the reference distance corresponding to the detection threshold: O4 $d_{L,\text{ref}} = 1600 \text{ Mpc}$ for O4 and $d_{L,\text{ref}} = 2500 \text{ Mpc}$ for O5. Our results indicate that is possible to obtain a unbiased final H_0 constraint for a galaxy catalogue volume-limited, even under a restrictive apparent magnitude threshold,

$m < 19$, where many distant galaxies have already been removed. We found that the H_0 bias increases with the apparent magnitude threshold adopted for the O4 run, reaching a maximum of $\Delta H_0 = 0.32 \pm 1.78 \text{ km/s/Mpc}$ for $m < 19$. In contrast, for the O5 run, the H_0 bias remains constant at approximately 0.22 km/s/Mpc across all apparent magnitude thresholds considered. This demonstrates that the volume-limited catalog is capable of effectively representing the true distribution of galaxy overdensities and underdensities along the line of sight. Furthermore, the results for different galaxy weighting schemes indicate that even if the wrong model is assumed, the final H_0 constraint remains unbiased. However, both the uncertainty and bias in the H_0 measurement increase in this mismodeling scenario, suggesting that a biased measurement may occur in some cases. This will be explored in detail in Sec. IV B. The maximum bias magnitude of H_0 occurs when it is assumed that all galaxies have an equal probability of hosting the GW event, but the SM is incorrectly applied as the weighting scheme. We found a bias of $3.67 \pm 8.68 \text{ km/s/Mpc}$ for O4, which increases to $5.16 \pm 16.13 \text{ km/s/Mpc}$ for O5. The minimum bias occurs when the true BBH merger distribution follows SM but is recovered using SFR, resulting in a bias of $0.71 \pm 1.90 \text{ km/s/Mpc}$ ($-0.18 \pm 0.42 \text{ km/s/Mpc}$) for O4 (O5). The improvement achieved with SFR, compared to equal weighting, can be attributed to the positive correlation between SM and SFR (see Fig. 8), which facilitates the correct assignment of probability for the majority of possible host galaxies.

Figure 3 shows the combined H_0 posterior for 100 simulated DSs detected in an O4- and O5-like configuration. The addition of the SM weighting reduces the uncertainty in the H_0 measurement (compare the green

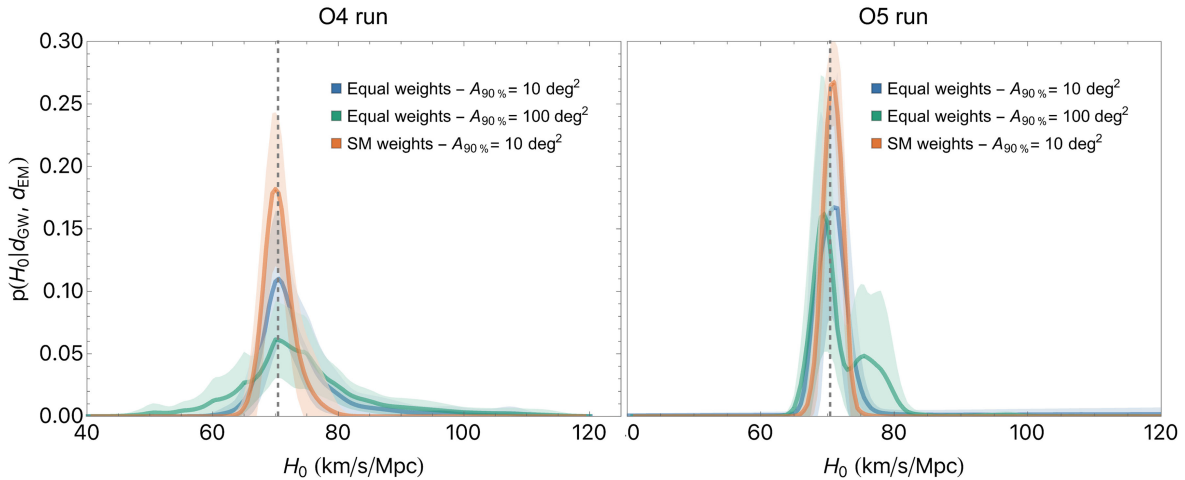


FIG. 3. Comparison of H_0 posteriors between three different configurations for O4-like (left panel) and O5-like (right panel) configuration: (1) in blue, using the fact that all galaxy in the catalog are equally likely to be the host and selected only the best-localized DS events (localization area is less than 10 deg^2), (2) in green, also assuming the same weigh scheme of (1) but selected only events with $A_{90\%} < 100 \text{ deg}^2$, and (3) in orange, using the best-localized DS events weighting by the stellar mass. The shaded area and the thick line represent the 68% CL and the mean of the 30 realization. The gray dashed line is the fiducial H_0 value from the *Magneticum* simulation.

and blue curves), an improvement of a factor of 1.97 for O4 and 1.89 for O5 run. All the estimations are in agreement with the fiducial H_0 value. In the best-case scenario, assuming stellar mass weighting, H_0 can be constrained to 3% (1%) for O4-like (O5-like) configurations. The comparison between the O4 and O5 results shows that the next terrestrial detector sensitivity can cause a reduction of 2.52 km/s/Mpc and 1.34 km/s/Mpc in the H_0 uncertainty for the equal and SM weighting schemes, respectively. This shows that the benefit of enhanced detector sensitivity is comparable to the improvements achieved by using SM as the GW host tracer.

A. Incompleteness in the volume-limited galaxy catalog

The statistical DS analysis is based on the idea of using galaxy catalogs to infer the redshift information of GW sources. However, galaxy catalogs are magnitude-limited due to the sensitivity of telescopes. As a result, GW signals may originate from distances beyond the completeness of these catalogs, meaning the true host galaxy might not be included, or the redshift distribution of galaxies along the line of sight may not accurately represent the true distribution. This can lead to a systematic bias in the measurement of the Hubble constant (see the Appendix). In this subsection, we instead explore volume-limited galaxy catalogs and analyze the impact of catalog incompleteness on the measurement of the Hubble constant.

Figure 4 presented the Magneticum galaxy catalog completeness for different apparent magnitude limits. As defined in [46], the galaxy catalog completeness is

$$f_{\text{completeness}}(d_L, m_{\text{th}}) = \frac{\sum_{i=1}^{N_{\text{gal}}^{\text{th}}(m_{\text{th}})} \Theta(d_L - d_{L,i})}{\sum_{i=1}^{N_{\text{gal}}} \Theta(d_L - d_{L,i})}, \quad (11)$$

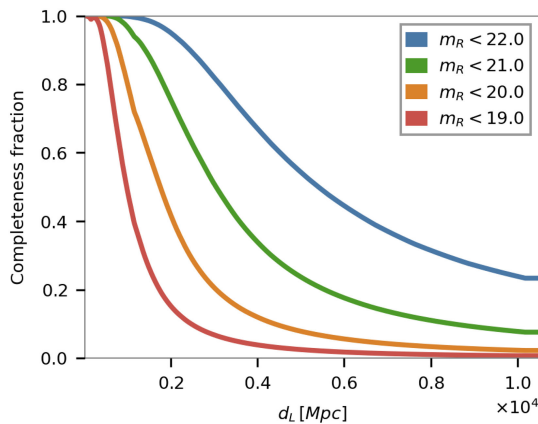


FIG. 4. Magneticum catalog galaxy completeness, defined in Eq. (11), as a function of the luminosity distance. The colors represent the galaxy catalog assuming an apparent magnitude limit of 22 (blue), 21 (orange), 20 (green), and 19 (red).

where N_{gal} is the number of galaxies inside the Magneticum catalog, $N_{\text{gal}}^{\text{th}}$ is the number of galaxies after the apparent magnitude cut m_{th} , Θ is the Heaviside function and d_L is a reference luminosity distance. This completeness is defined as the fraction of galaxies inside the magnitude-limited catalog (with threshold m_{th}) relative to a complete catalog (i.e., without any magnitude cut), as a function of d_L . As expected, the results in Fig. 4 show that the completeness fraction decreases with the reduction of the apparent magnitude limit, highlighting the impact of this effect on the intrinsic distribution of galaxies in the catalog. The abrupt change in the distribution of galaxy distances has implications for the H_0 posterior computed in the statistical DS analysis, as the overdensities in the galaxy distribution lead to peaks in the H_0 posterior.

The incompleteness of the galaxy catalog has been extensively studied in previous works [25,34,35,46,55,64,65], where a correction term was introduced to account for the unobserved galaxies in the catalog distribution. As pointed out in [65] and confirmed in [46], assuming a homogeneous completion for a nonuniform patch of sky can either overestimate or underestimate the galaxy contributions in certain regions. This introduces bias in the H_0 estimation, particularly when there is no electromagnetic data support and the contribution from the out-of-catalog term dominates. [64,65] have extended the completeness methodology to the non-uniform case in an effort to mitigate these limitations. However, this approach remains an approximation of the real data and can still be a source of bias and uncertainties. In the absence of completeness corrections, the inherent bias of a magnitude-limited galaxy catalog, which favors the brightest and/or nearest objects and excludes many distant galaxies, prevents consistency with the true redshift distribution of galaxies along the line of sight. In the context of the statistical DS methodology, this lack of distant galaxies introduces a strong bias toward lower H_0 values (see the Appendix), as the observed galaxy population becomes excessively concentrated at low redshifts.

One potential solution to this issue is to make the galaxy catalog volume limited. To achieve this, we follow the same steps described in previous works [37,38]: (1) the maximum redshift for each GW event was determined by converting the higher 90% CL bound of the luminosity distance measurement into the corresponding maximum redshift, assuming the largest value of H_0 in the prior; (2) we computed the absolute magnitude for the assumed apparent magnitude threshold at that maximum redshift, and then remove all galaxies within catalog that is above this absolute magnitude threshold. In all the steps, we assumed the fiducial Λ CDM cosmology. This limiting absolute magnitude represents the minimum value that a galaxy, located at the maximum redshift, must have to achieve an apparent magnitude above the limiting apparent magnitude of the galaxy sample. The limiting absolute magnitude is adjusted

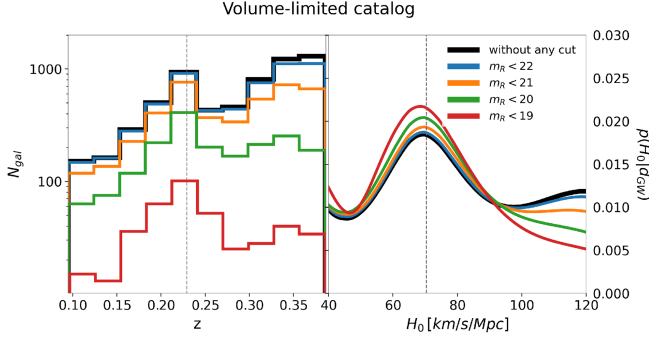


FIG. 5. Redshift distribution and H_0 posterior for a simulated DS event using a volume-limited galaxy catalog with different magnitude thresholds. The black line shows the results using the full Magneticum galaxy catalog, while the other colors represent specific apparent magnitude cuts. The dashed gray line in the redshift distribution indicates the true redshift of the DS event, and in the H_0 posterior, the dashed gray line represents the fiducial value.

for each GW source, meaning that the galaxies used have different magnitude ranges (distant GW sources will have a more restrictive limiting absolute magnitude than closer GW sources in a galaxy catalog limited by a fixed apparent magnitude value) in each case. This approach relies on the assumption that the final redshift distribution of the possible host galaxies remains relatively independent of magnitude.

In Fig. 5 we present the results with different volume-limited catalogs for a simulated O4-like GW source, localized at $z \approx 0.23$ where the Magneticum catalog is already incomplete. Compared to the magnitude-limited catalog results (see Fig. 11 in the Appendix), the volume-limited case produces an H_0 posterior distribution more consistent with the one obtained with a complete catalog. The peak of the posterior does not shift for any absolute magnitude threshold. Nonetheless, a slight change in the H_0 posterior is noticeable at the boundaries. The H_0 posterior decreases at high H_0 values and slightly increases at low H_0 values, corresponding to the change in the redshift distribution. According to the results reported in Table I, the volume-limited catalog provides an unbiased H_0 measurement across all magnitude thresholds adopted in this study. In the O4 scenario, the bias reaches a minimum value of 0.18 ± 1.71 km/s/Mpc for $m_R < 22$ and a maximum of 0.32 ± 1.78 km/s/Mpc for $m_R < 19$. For the O5 run, the bias remains constant at 0.22 km/s/Mpc across all apparent magnitude thresholds, with a slight decrease in the bias uncertainty for $m_R < 19$.

To understand the limit to which the effect of galaxy incompleteness can be controlled, Fig. 6 shows the cumulative sum of the ratio between the offset of the H_0 posterior peak, δH_0 , and the corresponding standard deviation, as a function of the GW source redshift. The offset δH_0 is defined as the difference between the H_0 posterior peak

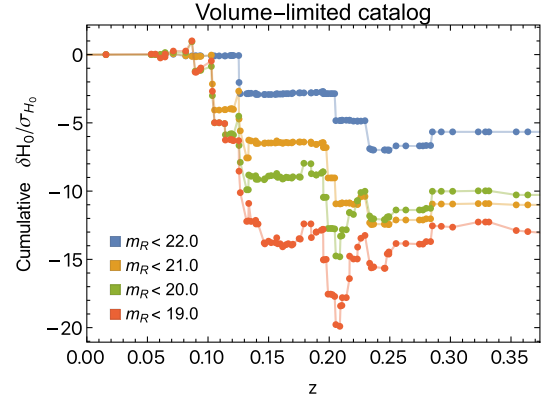


FIG. 6. Cumulative sum of the ratio between the H_0 posterior peak offset and the standard deviation, $\delta H_0 / \sigma_{H_0}$, as a function of GW source redshift for various apparent magnitude cuts (different colors). The H_0 posterior peak offset is defined as the difference between the H_0 posterior peak with both apparent and absolute magnitude cuts and that obtained without any cuts. In the text, we explain how the absolute magnitude was determined.

obtained using a catalog with an apparent (and absolute) magnitude cut and that obtained using a complete catalog, i.e., without any cut. This is computed for a volume-limited catalog using one realization of a sample of 100 GWs. The cumulative sum value represents the total ratio of H_0 posterior peak offset to σ_{H_0} accumulated from each GW source in the simulated sample. Each point in the figure corresponds to the sum of these ratios up to the redshift of that GW source. Figure 6 shows that the H_0 offset remains less than the standard deviation ($\delta H_0 / \sigma_{H_0} < 1$) for GW sources detected at low redshifts ($z < 0.1$). In this regime, the catalog is nearly complete, and the offset can be considered negligible. On the other hand, for GW events beyond $z = 0.1$, the Magneticum catalog becomes less complete, resulting in a cumulative sum of the ratio of the H_0 posterior peak offset to the standard deviation of approximately -5.7 , -11.0 , -10.3 , and -13.0 for apparent magnitude thresholds of 22, 21, 20, and 19, respectively. This represents a reduction in the ratio $\delta H_0 / \sigma_{H_0}$ of approximately 59%, 62%, 79%, and 85% compared to the results obtained with the magnitude-limited catalog (see Fig. 12 in the Appendix) for $m_R < 22$, 21, 20, and 19, respectively. As can be seen in Fig. 6, the H_0 posterior peak offset increases with the source redshift until it reaches a plateau at high z , where the effect of the absolute magnitude threshold is minimum compared to the apparent magnitude threshold. We demonstrate that a volume-limited catalog results in a lower cumulative sum of $\delta H_0 / \sigma_{H_0}$ compared to a magnitude-limited catalog for all adopted apparent magnitude thresholds. Additionally, Fig. 7 illustrates the transition between the volume-limited and magnitude-limited catalogs (represented by the label “No M_R cut” in the figure).

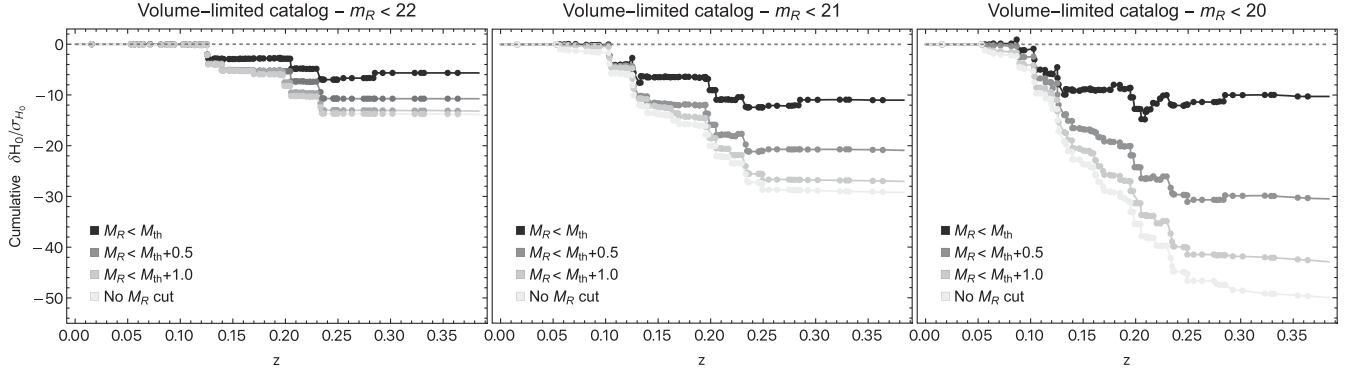


FIG. 7. Cumulative sum of the ratio between the H_0 posterior peak offset and the standard deviation, $\delta H_0/\sigma_{H_0}$, as a function of GW source redshift for volume-limited catalogs. For each curve with a gray color scale, a distinct absolute magnitude threshold is adopted. The darkest gray represents the conservative value computed, as described in the text, while the lighter gray corresponds to a more flexible cut (higher values for the magnitude threshold), which includes fainter galaxies.

B. Mismodeling the GW galaxy host

We now turn our attention to exploring the potential systematic effects that may arise in the estimation of H_0 when incorrect weighting schemes are applied to the GW host probability. We investigate this effect using the O4-like sample of DS events, which was constructed assuming stellar mass as the GW host tracer, and investigate mismodeling by assuming equal weighting and star formation rate weighting. Figure 8 shows the distributions of galaxy properties (stellar mass, star formation rate, and redshift) in the Magneticum galaxy catalog for three injection sets. The 2D corner plot of SM-SFR exhibits a positive correlation

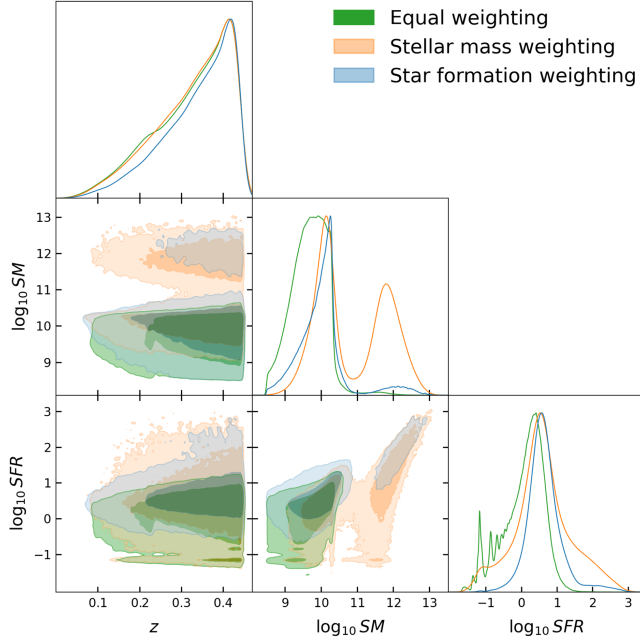


FIG. 8. Distributions of stellar mass, star formation rate, and redshift from the Magneticum galaxy catalog, with each color representing results obtained using different weighting schemes.

between the two quantities when SFR weighting is applied, indicating that an unbiased H_0 estimation can still be recovered even if the SM weighting scheme is incorrectly used. However, when the true GW merger rates follow the SM, a bimodal distribution emerges in SM, breaking the positive correlation with SFR. This reveals that a systematic bias can be introduced in the H_0 measurement if SFR is incorrectly assumed as the GW host tracer. A similar result was found in the universemachine galaxy catalog discussed in [47], where the bimodal distribution appeared in SFR instead of SM.

The results presented in Table I indicate that an unbiased H_0 estimation can still be recovered (note that in the last column, the H_0 bias measurements are consistent with each other for the mismodeling cases) even when BBH mergers follow the SM and are mismodeled using SFR. This may seem counterintuitive based on the discussion in the previous paragraph and could lead to the incorrect conclusion that an improper weighting scheme cannot introduce bias. However, a careful analysis of the individual H_0 posteriors reveals that the presence of well-localized GW events (i.e., those with a low number of galaxies, $\lesssim 200$, along the line of sight) is responsible for producing an unbiased estimation in the combined result. Figure 9 shows how the mean of the H_0 posterior peak offset varies with redshift, the number of possible host galaxies (N_{gal}), and the uncertainty in the luminosity distance of the GW source. The H_0 offset is defined as the difference between the posterior peak obtained using an incorrect weighting scheme and that obtained using the correct one. We used bin widths of 0.005 for z , 100 for N_{gal} , and 20 Mpc for σ_{d_L} , respectively. Assuming an incorrect GW host weighting scheme, we observe that the individual H_0 posterior peak can be offset by up to $\delta H_0 \approx 40$ km/s/Mpc, or even higher, in the case of imprecise luminosity distance and sky localization measurements (which are related to N_{gal}). Our results are in

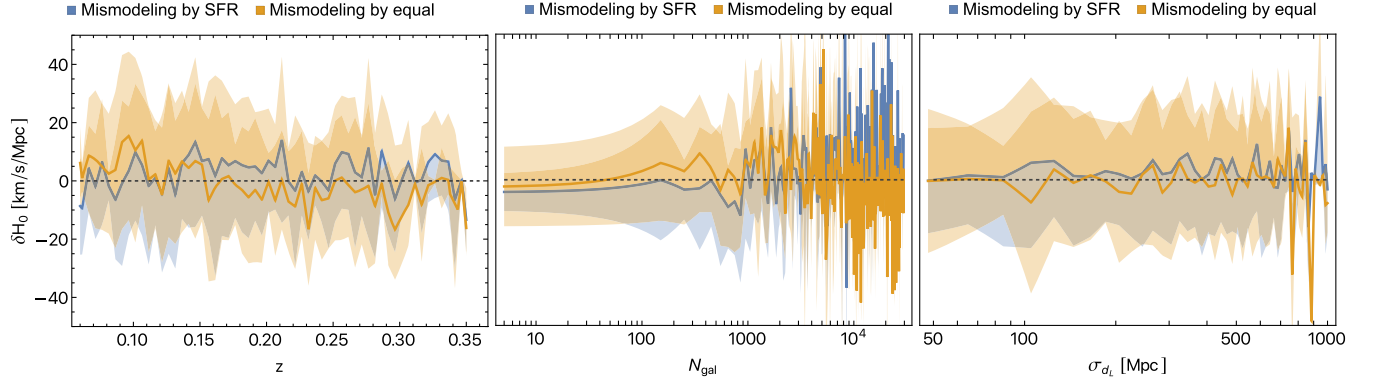


FIG. 9. Offset of the Hubble constant posterior peak resulting from an incorrect weighting scheme, as a function of redshift z , number of galaxies N_{gal} , and luminosity distance uncertainty σ_{d_L} . The solid line represents the mean bias computed for each event posterior over the 30 distinct realizations, while the shaded region corresponds to the statistical uncertainty in the bias estimation. The blue color represents the results obtained under the incorrect assumption that the hosts follow the SFR, while the orange curve corresponds to the assumption of equal weights for all galaxies.

agreement with those of [47], which demonstrated that an incorrect weighting scheme can lead to a significant bias in the “transitional regime” when $N_{\text{gal}} = 10^2\text{--}10^4$.

In addition, we explored the results discussed above using probability-probability (PP) plots [66] for the H_0 posterior, assuming the correct weighting scheme (stellar mass as a GW host tracer) and an incorrect weighting scheme (weighting by SFR or equal probability). The PP plot presents the cumulative distributions of two different datasets; if they follow the identity line in the range $[0, 1]$, the two distributions are approximately equal, indicating no

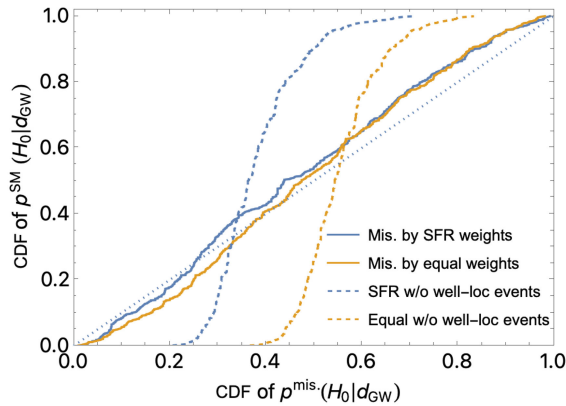


FIG. 10. PP plot comparing the posterior distributions of H_0 under different weighting schemes: the corrected scheme (SM) versus the incorrect ones (SFR or equal weights). The y-axis shows the cumulative distribution function (CDF) of the posterior found with the corrected weighting scheme, and the x-axis shows the CDF of the posterior assuming the incorrect weighting scheme. The solid lines represent the results considering all 100 GW events in the sample, while dashed lines correspond to the case where well-localized GW events are excluded. Deviations from the identity line (dotted) indicate discrepancies between the posteriors caused by incorrect modeling of the weighting scheme.

bias. Using the mean of H_0 posterior from 30 independent realizations, each considering 100 detected GW events with $A_{90\%} < 10 \text{ deg}^2$, we constructed the PP plot (see solid lines in Fig. 10 for the two mismodel case). In both cases, we observe a slight deviation of the curves from the identity line. Additionally, we repeated the analysis presented in Table I, but excluding all well-localized GW events. These results are represented as a dashed line in Fig. 10. The interpretation of this PP plot shows that the absence of well-localized events leads to an offset of the H_0 posterior peak toward higher values and leads to a broader posterior distribution. In particular, when the star formation rate (or an equal-weighting scheme) is incorrectly applied, the estimated values of the Hubble constant are $H_0 = 74.5^{+12.2}_{-12.5} \text{ km/s/Mpc}$ ($H_0 = 71.0^{+12.3}_{-16.9} \text{ km/s/Mpc}$).

C. Understand the H_0 precision from the dark sirens: The impact of selection cuts

The quality of cosmological constraints derived from DSs is inherently tied to the coverage and completeness of the galaxy catalog, as well as the precision of the gravitational-wave luminosity distance measurement. Therefore, identifying and selecting the most informative DSs is essential for optimizing their contribution to constraining cosmological parameters. In the following paragraph, we investigate how the H_0 forecasts change by different selection criteria, including (1) the uncertainty in the luminosity distance, (2) varying redshift ranges, and (3) the sky localization area.

First, we report constraints on H_0 constraints derived from a DS mock catalog, assuming an equal-weight scheme with different selection criteria based on the luminosity distance measurement of GW sources. Specifically, we consider three distinct cuts on the fractional error in the GW luminosity distance measurement, defined as the ratio of the uncertainty to the mean value ($\sigma_{d_L}/\bar{d}_L$): < 0.2 , 0.15 , and 0.1 . All selected DSs satisfied the sky localization

TABLE II. H_0 measurements for different selection cuts (redshift range, luminosity distance uncertainty, and sky localization area) applied to GW detections with $A_{90\%} < 10 \text{ deg}^2$ (except where explicitly stated otherwise) and a luminosity distance thresholds of 1600 Mpc.

Selection cuts	N_{GW}	H_0 [km/s/Mpc]	σ_{H_0} [km/s/Mpc]
$z < 0.1$	506	$70.71^{+0.88}_{-0.31}$	0.6
$z < 0.2$	1951	$71.52^{+0.35}_{-0.84}$	0.6
$z < 0.3$	2834	$70.71^{+0.48}_{-0.52}$	0.5
$\sigma_{d_L}/\bar{d}_L < 0.1$	482	$71.52^{+0.62}_{-0.96}$	0.79
$\sigma_{d_L}/\bar{d}_L < 0.15$	897	$71.51^{+0.53}_{-0.88}$	0.71
$\sigma_{d_L}/\bar{d}_L < 0.2$	1339	$70.71^{+0.92}_{-0.39}$	0.65
$A_{90\%} < 10 \text{ deg}^2$	314	$69.90^{+3.39}_{-3.20}$	3.29
$A_{90\%} < 20 \text{ deg}^2$	660	$69.90^{+1.23}_{-2.42}$	1.83
$A_{90\%} < 50 \text{ deg}^2$	1160	$69.90^{+0.60}_{-0.86}$	0.73
$A_{90\%} < 100 \text{ deg}^2$	1630	$69.90^{+0.82}_{-0.49}$	0.66

criterion $A_{90\%} < 10 \text{ deg}^2$. In general, we find that events with a fractional error of 20% can still contribute to improving constraints on H_0 , leading to a reduction of 18% and 8% in the uncertainty compared to the cases with $\sigma_{d_L}/\bar{d}_L = 0.1$ and 0.15, respectively. In addition, we find that with a sample of approximately 1,440 DSs detected by LVK O4, it is possible to achieve an H_0 uncertainty of 0.65 km/s/Mpc. These results are summarized in Table II.

Furthermore, we explore how our forecasts change for deeper redshift ranges and investigate the limits of statistical DS analysis as a cosmological probe, achievable with current GW detections. We found that the most informative DSs are localized at low redshift ($z < 0.1$), and the H_0 uncertainty decreases only by 0.1 km/s/Mpc when including sources up to $z < 0.3$. This result arises from the significant degradation in the 3D localization accuracy of events at high redshift, which makes it difficult to identify their host galaxies, resulting in uninformative H_0 posteriors. In the next-generation GW detectors, the sky localization volume will be measured with greater precision, enabling the extraction of more cosmological information from high-redshift GW sources. This scenario was explored by [43] in the era of third-generation gravitational wave detectors, such as the Einstein Telescope and Cosmic Explorer, with a complete galaxy catalog extending to $z < 1$. The authors demonstrated that with a sample of well-localized DSs ($\text{SNR} > 300$), it is possible to achieve a precision of 0.8% for the Hubble constant and 10% for the matter energy density parameter.

Finally, we also investigate how the precision of H_0 degrades as the GW sky localization area increases. This analysis is particularly important for understanding the selection criteria that ensure the identification of the most cosmologically informative DS events. Specifically for this analysis, we selected DS events that satisfy the sky

localization criterion $A_{90\%} < 100 \text{ deg}^2$. In the previous paragraph, we discussed how the precision of H_0 varies as a function of the luminosity distance uncertainty. We observed that as the luminosity distance uncertainty increases, we approach the non-informative limit, where DS events are unable to improve the precision of H_0 . This same limit can also be investigated in terms of sky localization area, as larger areas contain a greater number of galaxies, making it more challenging to identify the host galaxy. Consequently, we can assess the impact on H_0 measurements in terms of catalog coverage over the localization area or the redshift depth. We found that it is possible to identify informative DSs with sky localization areas in the range $50 \text{ deg}^2 < A_{90\%} < 100 \text{ deg}^2$ (see the results reported in Table II). These events provide a reduction of approximately $\sim 10\%$ (80%) in the H_0 uncertainty compared to the results for $A_{90\%} < 50 \text{ deg}^2$ ($A_{90\%} < 10 \text{ deg}^2$). The most significant improvement, a reduction of approximately $\sim 60\%$ in the H_0 precision, occurs with the inclusion of events in the range $20 \text{ deg}^2 < A_{90\%} < 50 \text{ deg}^2$. This improvement degrades to approximately 10% when including events in the range $50 \text{ deg}^2 < A_{90\%} < 100 \text{ deg}^2$.

V. CONCLUSIONS

The advent of dedicated deep-field surveys, combined with the rapidly increasing number of gravitational wave detections, is set to establish DSs as a powerful tool for cosmological probing. Therefore, given the approximation of the percent-level precision, it is essential to carefully consider the potential systematic biases in H_0 inference that could arise in such analyses. Here we investigate the systematic of the DSs analysis when (1) incorrectly weighted the GW host, and (2) an incomplete galaxy catalog is used. We also investigated the potential of DSs for cosmological parameter inference.

First, we performed a series of end-to-end simulations of binary black hole mergers from the LVK detector network with a sensitivity similar to O4 and O5 runs, combined with a galaxy mock catalog generated from the Magneticum simulations. We found it possible to obtain unbiased measurements when recovering with the same injected model. The average results over the 30 realization provided an initial indication that it is possible to recover unbiased H_0 measurements even when injected by SM and incorrectly weighting the host galaxy scheme assuming SFR or equal probability. However, a detailed analysis of the individual posteriors revealed that the H_0 posterior peak can be offset by approximately $\delta H_0 \gtrsim 40 \text{ km/s/Mpc}$ for GW detections with $\sigma_{d_L} > 10^2 \text{ Mpc}$ and $N_{\text{gal}} > 10^3$. This result was masked by the presence of well-localized events, which did not exhibit significant bias in the H_0 measurements. These results highlight a potential solution to reduce such biases, as the next generation of GW

facilities will improve the localization volume of gravitational waves, leading to an increase in the number of well-localized events observed.

The absence of incompleteness correction in a volume-limited galaxy catalog does not introduce any bias in the H_0 measurement. We obtained H_0 measurements that are statistically consistent within 1σ for all adopted apparent magnitude thresholds. These results demonstrate the capability of the volume-limited catalog to accurately recover the true distribution of galaxy overdensities and underdensities along the line of sight.

Finally, we characterized the constraining power of DSs. We find a H_0 precision of 6%(3%) for 100 O4-like (O5-like) detected GW events with a 90% CL sky localization area less than 10 deg^2 , assuming no host-galaxy weighting. This number of detections is expected to be achieved after approximately 8 years of observation at O4 sensitivity and 3 years at O5 sensitivity with the LVK. The precision decreased to 13% (6%) for the O4 (O5) run when the sky localization area increased to $<100 \text{ deg}^2$. We show that weighting the potential host galaxies by stellar mass improves the constraints, reducing the uncertainty by approximately half compared to the case without weighting for both O4-like and O5-like configurations. Specifically, the uncertainty decreases by 2.28 km/s/Mpc in the O4-like configuration and by 1.04 km/s/Mpc in the O5-like configuration, resulting in an H_0 precision of 3%(1%) for O4(O5) run. A detailed analysis on how the selection criteria affect the H_0 forecast reveals that events with $0.15 < \sigma_{d_L}/\bar{d}_L < 0.2$ and $50 \text{ deg}^2 < A_{90\%} < 100 \text{ deg}^2$ can modestly improve the H_0 measurement. The most informative events are localized at low redshift, but DSs at $z > 0.2$ lead to a decrease of 0.1 km/s/Mpc in the H_0 uncertainty.

The results presented here are based on the simplifying assumption that galaxy redshifts are measured perfectly, i.e., redshift uncertainties are ignored. This is an optimistic scenario, as galaxies are likely to have significant photometric redshift uncertainties due to the lack of spectroscopic data at the redshift limits considered in this study. As demonstrated in [45], photometric redshift uncertainties broaden the H_0 posterior and, if an inconsistent Bayesian analysis is adopted, can introduce bias. These results were derived under the assumption of a Gaussian approximation for the redshift likelihood. However, in real-world applications, machine learning or template fitting algorithms are typically used to compute the photo- z probability distribution function. It is therefore crucial to investigate how non-Gaussian features in the photo- z PDF impact the results of this analysis. We leave this investigation to a future study.

ACKNOWLEDGMENTS

We acknowledge Alex Saro and Klaus Dolag for the Magneticum [56]. The authors made use of Sci-Mind servers

machines developed by the CBPF AI LAB team and would like to thank Paulo Russano and Marcelo Portes de Albuquerque for all the support in infrastructure matters. V. A. thanks Antonella Palmese for her hospitality and useful discussions during a visit to McWilliams Center for Cosmology & Astrophysics. C. R. B. acknowledges the financial support from CNPq (316072/2021-4) and from FAPERJ (Grants No. 201.456/2022 and No. 210.330/2022) and the FINEP Contract No. 01.22.0505.00 (ref. 1891/22).

DATA AVAILABILITY

The data are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX: INCOMPLETENESS GALAXY CATALOG: MAGNITUDE-LIMITED CASE

In this appendix, we present a complementary discussion about galaxy incompleteness in Hubble constant measurements, as introduced in Sec. IV A, within the context of magnitude-limited galaxy surveys.

To exemplify the effect of magnitude-limited galaxy incompleteness, in the left panel of Fig. 11 we applied various apparent magnitude limits on the r -band for a simulated O4-like GW source, localized at $z \approx 0.23$ where the Magneticum catalog is already incomplete. As the catalogs become less complete, fewer galaxies are visible at high redshift ($z \approx 0.23$), which reduces the peak of the redshift distribution, with the majority of galaxies concentrated at lower z . This change leads to a systematic bias in the H_0 posterior: the maximum of the posterior moves to lower H_0 values.

The results presented in Table I for the magnitude-limited surveys indicate that the final H_0 constraint remains unbiased when the galaxy catalog is incomplete, with a

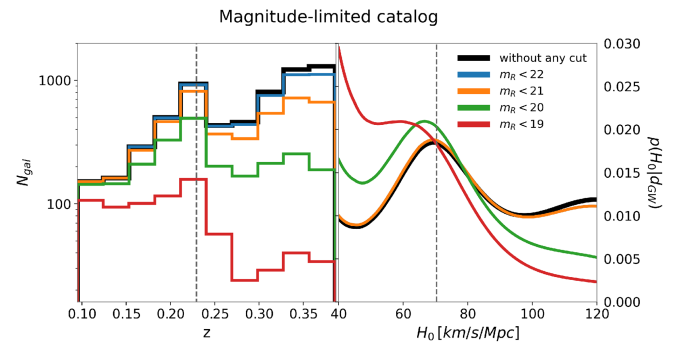


FIG. 11. Effect of the apparent magnitude cut on the redshift distribution and the H_0 posterior for a simulated DS event. The black curve corresponds to results derived from the complete Magneticum galaxy catalog, while the colored curves denote subsets defined by specific apparent magnitude limits. In the redshift distribution panel, the dashed gray line marks the true redshift of the DS event; in the H_0 posterior panel, it indicates the fiducial value.

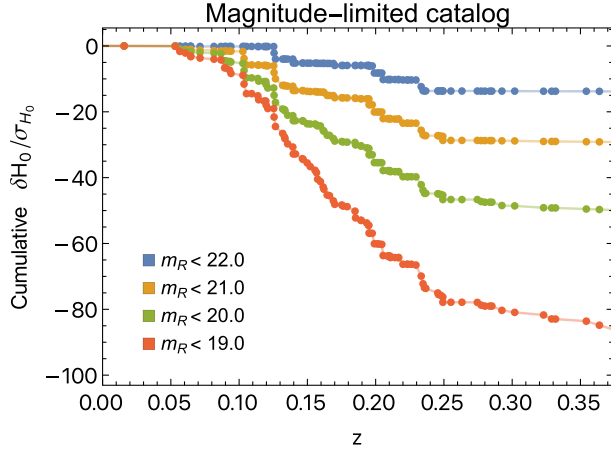


FIG. 12. Cumulative distribution of $\delta H_0 / \sigma_{H_0}$, the offset of the H_0 posterior peak normalized by its standard deviation, as a function of GW source redshift, for different apparent magnitude limits (shown in color). This corresponds to the magnitude-limited survey version of Fig. 12.

minimum apparent magnitude threshold of 21.0, corresponding to 86% completeness for the O4 run configuration, and a threshold of 20.0 (29% completeness) for the

O5-like scenario. A more restrictive apparent magnitude threshold, $m_R < 19.0$, will result in a biased H_0 measurement, with maximum bias of -22.24 km/s/Mpc and -16.03 km/s/Mpc for the O4 and O5 observing runs, respectively. The difference between the results is mainly due to the fact that we are considering a higher SNR threshold for O5, resulting in events with better sky localization and, consequently, less susceptibility to the effect of catalog incompleteness.

Figure 12 demonstrates that the H_0 offset is negligible ($\delta H_0 / \sigma_{H_0} < 1$) for nearby GW sources ($z \lesssim 0.06$), where the galaxy catalog is nearly complete. Beyond this limit, the cumulative value of the normalized H_0 offset begins to increase, a regime increasingly affected by catalog incompleteness. For an apparent magnitude limit of 19, we find that the cumulative sum of $\delta H_0 / \sigma_{H_0}$ reaches approximately -84.8 , highlighting the potential for significant bias under strong incompleteness. As expected, an apparent magnitude threshold of 22 yields the minimum cumulative sum value, -13.8 . We can also observe an abrupt increase in the offset for $m_R < 20$, which is a direct consequence of the decrease in the galaxy completeness fraction within the distance range of interest (see Fig. 4).

-
- [1] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [2] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Population of merging compact binaries inferred using gravitational waves through GWTC-3, *Phys. Rev. X* **13**, 011048 (2023).
 - [3] B. F. Schutz, Determining the Hubble constant from gravitational wave observations, *Nature (London)* **323**, 310 (1986).
 - [4] B. P. Abbott, R. Abbott, T. D. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GW170817: Observation of gravitational waves from a binary neutron star inspiral, *Phys. Rev. Lett.* **119**, 161101 (2017).
 - [5] A. Goldstein *et al.*, An ordinary short gamma-ray burst with extraordinary implications: Fermi-GBM detection of GRB 170817A, *Astrophys. J.* **848**, 14 (2017).
 - [6] V. Savchenko, C. Ferrigno, E. Kuulkers *et al.*, Integral detection of the first prompt gamma-ray signal coincident with the gravitational-wave event GW170817, *Astrophys. J. Lett.* **848**, L15 (2017).
 - [7] I. Arcavi, G. Hosseinzadeh, D. A. Howell *et al.*, Optical emission from a kilonova following a gravitational-wave-detected neutron-star merger, *Nature (London)* **551**, 64 (2017).
 - [8] D. A. Coulter, R. J. Foley, C. D. Kilpatrick *et al.*, Swope supernova survey 2017a (SSS17a), The optical counterpart to a gravitational wave source, *Science* **358**, 1556 (2017).
 - [9] P. S. Cowperthwaite, E. Berger, V. A. Villar *et al.*, The electromagnetic counterpart of the binary neutron star merger LIGO/Virgo GW170817. II. UV, Optical, and near-infrared light curves and comparison to kilonova models, *Astrophys. J. Lett.* **848**, L17 (2017).
 - [10] M. Soares-Santos *et al.*, The electromagnetic counterpart of the binary neutron star merger LIGO/Virgo GW170817. I. Discovery of the optical counterpart using the dark energy camera, *Astrophys. J. Lett.* **848**, 7 (2017).
 - [11] R. Chornock, E. Berger, D. Kasen *et al.*, The electromagnetic counterpart of the binary neutron star merger LIGO/Virgo GW170817. IV. Detection of near-infrared signatures of r-process nucleosynthesis with Gemini-South, *Astrophys. J. Lett.* **848**, L19 (2017).
 - [12] M. M. Kasliwal, E. Nakar, L. P. Singer *et al.*, Illuminating gravitational waves: A concordant picture of photons from a neutron star merger, *Science* **358**, 1559 (2017).
 - [13] M. Nicholl, E. Berger, D. Kasen *et al.*, The electromagnetic counterpart of the binary neutron star merger LIGO/Virgo GW170817. III. Optical and UV spectra of a blue kilonova from fast polar ejecta, *Astrophys. J. Lett.* **848**, L18 (2017).

- [14] P. A. Evans, S. B. Cenko, J. A. Kennea *et al.*, Swift and NuSTAR observations of GW170817: Detection of a blue kilonova, *Science* **358**, 1565 (2017).
- [15] E. Pian, P. D’Avanzo, S. Benetti *et al.*, Spectroscopic identification of r-process nucleosynthesis in a double neutron-star merger, *Nature (London)* **551**, 67 (2017).
- [16] S. J. Smartt, T.-W. Chen, A. Jerkstrand *et al.*, A kilonova as the electromagnetic counterpart to a gravitational-wave source, *Nature (London)* **551**, 75 (2017).
- [17] N. R. Tanvir, A. J. Levan, C. Gonzalez-Fernandez *et al.*, The emergence of a Lanthanide-Rich kilonova following the merger of two neutron stars, *Astrophys. J. Lett.* **848**, L27 (2017).
- [18] S. Valenti, D. J. Sand, S. Yang, E. Cappellaro, L. Tartaglia, A. Corsi, S. W. Jha, D. E. Reichart, J. Haislip, and V. Kouprianov, The discovery of the electromagnetic counterpart of GW170817: Kilonova AT 2017gfo/DLT17ck, *Astrophys. J. Lett.* **848**, L24 (2017).
- [19] B. Abbott, R. Abbott, T. D. Abbott *et al.*, A gravitational-wave standard siren measurement of the Hubble constant, *Nature (London)* **551**, 85 (2017).
- [20] R. Abbott, T. D. Abbott, F. Acernese *et al.* (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, *Phys. Rev. X* **13**, 041039 (2023).
- [21] J. Ezquiaga, Hearing gravity from the cosmos: GWTC-2 probes general relativity at cosmological scales, *Phys. Lett. B* **822**, 136665 (2021).
- [22] J. Ezquiaga and D. Holz, Spectral sirens: Cosmology from the full mass distribution of compact binaries, *Phys. Rev. Lett.* **129**, 061102 (2022).
- [23] A. Farah, T. Callister, J. Ezquiaga, M. Zevin, and D. E. Holz, No need to know: Astrophysics-free gravitational-wave cosmology, *Astrophys. J.* **978**, 153 (2025).
- [24] I. Magaña Hernandez and A. Ray, Beyond gaps and bumps: Spectral siren cosmology with non-parametric population models, *arXiv:2404.02522*.
- [25] N. Borghi, M. Mancarella, M. Moresco, M. Tagliazucchi, F. Iacovelli, A. Cimatti, and M. Maggiore, Cosmology and astrophysics with standard sirens and galaxy catalogs in view of future gravitational wave observations, *Astrophys. J.* **964**, 191 (2024).
- [26] S. Mukherjee, B. D. Wandelt, S. M., B. D. Nissanke, and A. Silvestri, Accurate precision cosmology with redshift unknown gravitational wave sources, *Phys. Rev. D* **103**, 043520 (2021).
- [27] S. Mukherjee, B. D. Wandelt, and J. Silk, Testing the general theory of relativity using gravitational wave propagation from dark standard sirens, *Mon. Not. R. Astron. Soc.* **502**, 1136 (2021).
- [28] S. Mukherjee, A. Krolewski, B. D. Wandelt, and J. Silk, Cross-correlating dark sirens and galaxies: Constraints on H_0 from GWTC-3 of LIGO–Virgo–KAGRA, *Astrophys. J.* **975**, 189 (2024).
- [29] T. Ghosh, S. More, S. Bera, and S. Bose, Bayesian framework to infer the hubble constant from the cross-correlation of individual gravitational wave events with galaxies, *Phys. Rev. D* **111**, 063513 (2025).
- [30] J. Ferri, I. L. Tashiro, L. R. Abramo, I. Matos, M. Quartin, and R. Sturani, A robust cosmic standard ruler from the cross-correlations of galaxies and dark sirens, *J. Cosmol. Astropart. Phys.* **04** (2025) 008.
- [31] C. R. Bom and A. Palmese, Standard siren cosmology with gravitational waves from binary black hole mergers in active galactic nuclei, *Phys. Rev. D* **110**, 083005 (2024).
- [32] A. Palmese, M. Fishbach, C. J. Burke, J. Annis, and X. Liu, Do LIGO/Virgo black hole mergers produce AGN flares? The case of GW190521 and prospects for reaching a confident association, *Astrophys. J. Lett.* **914**, L34 (2021).
- [33] W. Del Pozzo, Inference of cosmological parameters from gravitational waves: Applications to second generation interferometers, *Phys. Rev. D* **86**, 043011 (2012).
- [34] H.-Y. Chen, M. Fishbach, and D. E. Holz, A two per cent Hubble constant measurement from standard sirens within five years, *Nature (London)* **562**, 545 (2018).
- [35] M. Fishbach, R. Gray, I. Magaña Hernandez *et al.*, A standard siren measurement of the Hubble constant from GW170817 without the electromagnetic counterpart, *Astrophys. J. Lett.* **871**, L13 (2019).
- [36] M. Soares-Santos, A. Palmese, W. Hartley *et al.*, First measurement of the Hubble constant from a dark standard siren using the dark energy survey galaxies and the LIGO/Virgo binary-black-hole merger GW170814, *Astrophys. J. Lett.* **876**, L7 (2019).
- [37] A. Palmese, J. deVicente, M. E. S. Pereira *et al.*, A statistical standard siren measurement of the Hubble constant from the LIGO/Virgo gravitational wave compact object merger GW190814 and dark energy survey galaxies, *Astrophys. J. Lett.* **900**, L33 (2020).
- [38] A. Palmese, C. R. Bom, S. Mucesh, and W. G. Hartley, A standard siren measurement of the Hubble constant using gravitational-wave events from the first three LIGO/Virgo observing runs and the DESI legacy survey, *Astrophys. J.* **943**, 56 (2023).
- [39] W. Ballard, A. Palmese, I. Magaña Hernandez *et al.*, A dark siren measurement of the Hubble constant with the LIGO/Virgo gravitational wave event GW190412 and DESI galaxies, *Res. Notes AAS* **11**, 250 (2023).
- [40] V. Alfradique, C. R. Bom, A. Palmese, G. Teixeira *et al.*, A dark siren measurement of the Hubble constant using gravitational wave events from the first three LIGO/Virgo observing runs and DELVE, *Mon. Not. R. Astron. Soc.* **528**, 3249 (2024).
- [41] C. R. Bom, V. Alfradique, A. Palmese, G. Teixeira, L. Santana-Silva, A. Santos, and P. Darc, A dark standard siren measurement of the Hubble constant following LIGO/Virgo/KAGRA O4a and previous runs, *Mon. Not. R. Astron. Soc.* **535**, 961 (2024).
- [42] S. Borhanian, A. Dhani, A. Gupta, K. G. Arun, and B. S. Sathyaprakash, Dark sirens to resolve the Hubble–Lemaître tension, *Astrophys. J. Lett.* **905**, L28 (2020).
- [43] N. Muttoni, D. Laghi, N. Tamanini, S. Marsat, and D. Izquierdo-Villalba, Dark siren cosmology with binary black holes in the era of third-generation gravitational wave detectors, *Phys. Rev. D* **108**, 043543 (2023).
- [44] L.-G. Zhu and X. Chen, The dark side of using dark sirens to constrain the Hubble–Lemaître constant, *Astrophys. J.* **948**, 26 (2023).

- [45] J. R. Gair, A. Ghosh, R. Gray *et al.*, The Hitchhiker’s guide to the galaxy catalog approach for dark siren gravitational-wave cosmology, *Astron. J.* **166**, 22 (2023).
- [46] R. Gray, I. Magaña Hernandez, H. Qi *et al.*, Cosmological inference using gravitational wave standard sirens: A mock data analysis, *Phys. Rev. D* **101**, 122001 (2020).
- [47] A. G. Hanselman, A. Vijaykumar, M. Fishbach, and D. Holz, Gravitational-wave dark siren cosmology systematics from galaxy weighting, *Astrophys. J.* **979**, 9 (2025).
- [48] G. Perna, S. Mastrogiovanni, and A. Ricciardone, Investigating the impact of galaxies’ compact binary hosting probability for gravitational-wave cosmology, *Astron. Astrophys.* **698**, A128 (2025).
- [49] C. Turski, M. Bilicki, G. Dálya, R. Gray, and A. Ghosh, Impact of modelling galaxy redshift uncertainties on the gravitational-wave dark standard siren measurement of the Hubble constant, *Mon. Not. R. Astron. Soc.* **526**, 6224 (2023).
- [50] E. V. R. Lima, L. Sodr  Jr., C. R. Bom *et al.*, Photometric redshifts for the S-PLUS survey: Is machine learning up to the task?, *Astron. Comput.* **38**, 100510 (2022).
- [51] G. Teixeira, C. Bom, L. Santana-Silva *et al.*, Photometric redshifts probability density estimation from recurrent neural networks in the DECam local volume exploration survey data release 2, *Astron. Comput.* **49**, 100886 (2024).
- [52] C. Li, Y. Zhang, C. Cui *et al.*, A photometric redshift catalogue of galaxies from the DESI legacy imaging surveys DR10, *Astrophys. J.* **168**, 233 (2024).
- [53] L. P. Singer and L. R. Price, Rapid Bayesian position reconstruction for gravitational-wave transients, *Phys. Rev. D* **93**, 024013 (2016).
- [54] R. Gray, F. Beirnaert, C. Karathanasis *et al.*, Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues, *J. Cosmol. Astropart. Phys.* **12** (2023) 023.
- [55] S. Mastrogiovanni, D. Laghi, R. Gray, G. C. Santoro, A. Ghosh, C. Karathanasis, K. Leyde, D. A. Steer, S. Perri s, and G. Pierra, Joint population and cosmological properties inference with gravitational waves standard sirens and galaxy surveys, *Phys. Rev. D* **108**, 042002 (2023).
- [56] M. Hirschmann, K. Dolag, A. Saro, L. Bachmann, S. Borgani, and A. Burkert, Cosmological simulations of black hole growth: AGN luminosities and downsizing, *Mon. Not. R. Astron. Soc.* **442**, 2304 (2014).
- [57] A. Ragagnin, K. Dolag, V. Biffi *et al.*, A web portal for hydrodynamical, cosmological simulations, *Astron. Comput.* **20**, 52 (2017).
- [58] <http://www.magneticum.org/simulations.html>.
- [59] <https://lscsoft.docs.ligo.org/ligo.skymap/>.
- [60] https://git.ligo.org/lscsoft/gwistat/-/tree/master/psd?ref_type=heads.
- [61] KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration, Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Rev. Relativity* **21**, 3 (2018).
- [62] <https://dcc.ligo.org/LIGO-T2000012/public>.
- [63] R. Essick and M. Fishbach, Ensuring consistency between noise and detection in hierarchical Bayesian inference, *Astrophys. J.* **962**, 169 (2024).
- [64] A. Finke, S. Foffa, F. Iacovelli, M. Maggiore, and M. Mancarella, Cosmology with LIGO/Virgo dark sirens: Hubble parameter and modified gravitational wave propagation, *J. Cosmol. Astropart. Phys.* **08** (2021) 026.
- [65] R. Gray, C. Messenger, and J. Veitch, A pixelated approach to galaxy catalogue incompleteness: Improving the dark siren measurement of the Hubble constant, *Mon. Not. R. Astron. Soc.* **512**, 1127 (2022).
- [66] M. B. Wilk and R. Gnanadesikan, Probability plotting methods for the analysis for the analysis of data, *Biometrika* **55**, 1 (1968).