

Echoes from the dark: Galaxy catalog incompleteness in standard siren cosmology

Nicola Borghi^{1,2,3,*}, Michele Moresco^{1,2}, Matteo Tagliacruzchi^{1,2,3}, and Giulia Cuomo^{4,5}

¹ Dipartimento di Fisica e Astronomia “Augusto Righi”–Università di Bologna, Via Gobetti 93/2, I-40129 Bologna, Italy

² INAF-OAS, Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, I-40129 Bologna, Italy

³ INFN-Sezione di Bologna, Viale Bertini Pichat 6/2, 40127 Bologna, Italy

⁴ Dipartimento di Fisica “G. Occhialini”, Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

⁵ INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

Received 22 September 2025 / Accepted 14 December 2025

ABSTRACT

Gravitational wave observations can be combined with galaxy catalogs to constrain cosmology and test modified gravity theories using the standard siren method. However, galaxy catalogs are intrinsically incomplete due to observational limitations, potentially leaving host galaxies undetected, thereby weakening constraints and potentially introducing systematic errors. In this work, we present a self-consistent framework to study catalog incompleteness and host weighting effects, implemented in the publicly available CHIMERA pipeline. We obtained joint cosmological and astrophysical population constraints from 100 binary black hole (BBH) events in a LIGO-Virgo-KAGRA O5-like configuration using spectroscopic galaxy catalogs with varying completeness levels and stellar-mass host weighting schemes. We find percent-level constraints on H_0 with complete catalogs, reaching precisions of 1.6%, 1.3%, and 0.9% for constant, linear, and quadratic mass weighting, respectively. As completeness decreases, the precision degrades following a sigmoid trend, with a threshold and steepness that increase for stronger weightings. Simultaneously, the correlation between H_0 and the BBH population mass scale increases, making results more sensitive to assumptions about the astrophysical population. Remarkably, 2% precision remains achievable even when catalogs contain only 50% of the potential host galaxies within the gravitational wave detection horizon, while 1% precision requires host probabilities scaling with stellar mass squared. The results are robust against host weighting mismodeling, even at moderate completeness levels. This work further highlights the importance of spectroscopic galaxy surveys in standard siren cosmology and provides a pathway for developing the science case of future facilities.

Key words. gravitational waves – catalogs – galaxies: evolution – cosmological parameters

1. Introduction

Gravitational waves (GWs) from compact binary coalescences (CBCs) are maturing as a robust observational probe to constrain cosmology and fundamental physics, offering a new perspective on existing cosmological tensions (Verde et al. 2019; Moresco et al. 2022; Valentino et al. 2025). Because GWs carry direct information on the CBC luminosity distance, they can be used as standard distance indicators or “standard sirens” (Schutz 1986; Holz & Hughes 2005). With the additional source redshift information, it is then possible to constrain cosmology via the luminosity distance-redshift relation. However, the source redshift cannot be obtained directly from current GW data, as it is degenerate with the source masses.

When an electromagnetic counterpart is identified and the host galaxy redshift is measured, a “bright siren” measurement becomes possible. To date, the only confirmed case is GW170817 (Abbott et al. 2017b), constraining the Hubble constant to a precision of $\sim 10\%$ (Abbott et al. 2017a; Hotokezaka et al. 2019; Bulla et al. 2022; Palmese et al. 2024). The latest release of the Gravitational-Wave Transient Catalog 4 (GWTC-4.0) by the LIGO-Virgo-KAGRA (LVK) Collaboration contains 219 confident CBC detections, the majority of which are binary black holes (BBHs) without electromagnetic counterparts (Abbott et al. 2019, 2021, 2023a; Abac et al. 2025d). It is

therefore essential to develop methods to extract the cosmological information without relying on counterpart detection.

These “dark siren” methods include two main complementary approaches. The first consists of statistically inferring the redshift from catalogs of potential host galaxies within the gravitational wave localization volume (Schutz 1986; Del Pozzo 2012; Chen et al. 2018; Fishbach et al. 2019; Gray et al. 2020; Palmese et al. 2020; Finke et al. 2021; Gair et al. 2023). The second consists of breaking the mass-redshift degeneracy by modeling intrinsic astrophysical properties. This is possible with “spectral sirens”, sources whose source-frame mass distribution contains features such as peaks, gaps, or changes in slope (Chernoff & Finn 1993; Taylor et al. 2012; Farr et al. 2019; Mastrogiovanni et al. 2021; Mancarella et al. 2022; Ezquiaga & Holz 2022; Mali & Essick 2025). Additional approaches rely on the cross-correlation between the spatial distribution of GWs and galaxies (e.g., Oguri 2016; Mukherjee et al. 2021; Ferri et al. 2025; Pedrotti et al. 2025).

Recent standard siren analyses (Mastrogiovanni et al. 2023; Gray et al. 2023), including the latest GWTC-4.0 cosmology analysis (Abac et al. 2025b), have advanced beyond treating catalog and spectral siren methods as independent. Instead, they are jointly combined within a hierarchical Bayesian framework, making it possible to robustly marginalize over the uncertainties of the astrophysical and cosmological parameters simultaneously. Several pipelines are

* Corresponding author: nicola.borghi6@unibo.it

publicly available, including *gwcsmo* (Gray et al. 2020, 2023), *icargw* (Mastrogiovanni et al. 2023, 2024), and CHIMERA (Borghi et al. 2024; Tagliacruzchi et al. 2025). For further details on the application of GWs to cosmology, we refer to Palmese & Mastrogiovanni (2025) and Pierra & Mastrogiovanni (2025).

While galaxy catalogs provide valuable information, they are intrinsically limited by observational effects that can cause host galaxies to be missed, potentially weakening cosmological constraints (e.g., Gray et al. 2020; Cross-Parkin et al. 2025) and introducing systematic errors (Perna et al. 2025; Hanselman et al. 2025; Alfradique et al. 2025). This ‘‘incompleteness’’ directly relates to assumptions about the CBC host galaxy population, as GW sources may preferentially reside in certain types of galaxies that are more easily detected and thus better represented in surveys. The link between CBCs and their host galaxies has been increasingly investigated through population synthesis studies (e.g., Artale et al. 2020; Santoliquido et al. 2022) and observational analyses (Vijaykumar et al. 2024), although more BBH events are needed to establish robust observational constraints. As GW data quality improves, accounting for galaxy catalog incompleteness and weighting will become increasingly important. Extending current mock data analyses to model these effects and test for potential systematic errors in future standard siren measurements is a necessary step forward.

In our previous work (Borghi et al. 2024, hereafter B24), we forecasted constraints on H_0 from 100 BBH mergers detected in an LVK O5-like scenario. We obtained percent-level constraints on H_0 , while assuming a complete galaxy catalog with spectroscopic redshift measurements and considering only galaxies with stellar masses of $\log_{10}(M_*/M_\odot) > 10.5$ as potential hosts. While this approximation was physically motivated by the expectation that BBHs form in more massive galaxies (e.g., Artale et al. 2020; Santoliquido et al. 2022), a more realistic approach requires including the full galaxy population in the analysis. Naively extending to all masses without additional weighting, however, would implicitly assume that low-mass $\log_{10}(M_*/M_\odot) \approx 6$ galaxies have the same probability of hosting BBH mergers as massive $\log_{10}(M_*/M_\odot) \approx 12$ ones. This strong assumption, which extends the BBH hosting probability over several orders of magnitude, requires accounting for BBH host weighting and observational effects in galaxy surveys.

In this paper, we present a self-consistent framework to study the effects of galaxy catalog incompleteness and host weighting in standard siren analyses, we incorporate it into a state-of-the-art analysis pipeline, and we provide updated and more realistic forecasts. This framework can be readily applied to ongoing and upcoming surveys such as *Euclid* (Euclid Collaboration; Mellier et al. 2025) and the Dark Energy Spectroscopic Instrument survey (DESI; DESI Collaboration 2024), used to inform the science case of future facilities such as the Wide-Field Spectroscopic Telescope (WST; Mainieri et al. 2024), and implemented in systematic studies for current cosmological analyses (Agarwal et al. 2025). The paper is organized as follows. In Sect. 2, we introduce the statistical framework and its implementation in the CHIMERA code. In Sect. 3, we describe the generation of mock galaxy and GW catalogs. Sect. 4 provides the results of the incompleteness effects in standard siren constraints. Finally, in Sect. 5 we conclude the paper and discuss future applications of this method.

2. Method

We considered a population of GW sources, each described by the source-frame parameters, θ (such as redshift z , sky position

$\hat{\Omega}$, and binary masses m_1, m_2), which globally follow a population distribution, $p_{\text{pop}}(\theta | \lambda)$, described by hyperparameters, λ . Given a dataset of N independent GW detections, $\{\mathbf{d}_i\}$, the likelihood of λ can be obtained with the hierarchical Bayesian formalism (Loredo 2004; Mandel et al. 2019; Vitale et al. 2022)

$$p(\{\mathbf{d}_i\} | \lambda) \propto \prod_{i=1}^N \frac{\int p(\mathbf{d}_i | \theta, \lambda) p_{\text{pop}}(\theta | \lambda) d\theta}{\xi(\lambda)}, \quad (1)$$

where $p(\mathbf{d}_i | \theta, \lambda)$ is the likelihood of individual GW events, and the selection function is defined as

$$\xi(\lambda) = \int P_{\text{det}}(\theta, \lambda) p_{\text{pop}}(\theta | \lambda) d\theta, \quad (2)$$

which corrects for the fraction of detectable events given a target population, with $P_{\text{det}}(\theta, \lambda) \in [0, 1]$ being their detection probability. Equation (1) assumes a scale-invariant rate prior so that the likelihood depends only on the shape of the population distribution, not on the overall rate normalization (Mandel et al. 2019).

It is crucial that $P_{\text{det}}(\theta, \lambda)$ is consistent with the detection pipeline. To ensure this, Eq. (2) is typically estimated by Monte Carlo integration, using a suite of simulated signals injected into the detection pipelines. This involves generating N_{gen} events with parameters θ_j , then computing

$$\hat{\xi}(\lambda) = \frac{1}{N_{\text{gen}}} \sum_{i=1}^{N_{\text{det}}} \frac{p_{\text{pop}}(\theta_i | \lambda)}{p_{\text{draw}}(\theta_i)}, \quad (3)$$

where the summation runs over the N_{det} detected injections, reweighted by the ratio of the target population to the injection distribution.

2.1. Population

To model the BBH population, we separated the hyperparameters, λ , into three categories: those that describe the underlying cosmology, λ_c ; the mass distribution, λ_m ; and the redshift distribution, λ_z . This allowed us to write the population model as

$$p_{\text{pop}}(\theta | \lambda) = p(m_1, m_2 | \lambda_m) p(z, \hat{\Omega} | \lambda_c, \lambda_z). \quad (4)$$

The first term represents the intrinsic mass distribution, typically described using parametric models, although nonparametric approaches are also being studied (e.g. Edelman et al. 2022; Abbott et al. 2023b; Farah et al. 2025). The second term acts as a ‘‘redshift prior’’, which can be informed by external electromagnetic data, such as galaxy catalogs. Equation (4) assumes that the intrinsic BBH mass distribution does not evolve significantly on cosmological timescales, as supported by current data (see Callister 2024; Abac et al. 2025c). More generally, we can incorporate dependencies on host galaxy properties, \mathcal{G} (such as stellar mass, metallicity, and star formation rate), that may influence BBH formation, evolution, and ultimately their merger rate. In this case, the population prior is marginalized over galaxy properties,

$$p_{\text{pop}}(\theta | \lambda) = \int p_{\text{pop}}(\theta, \mathcal{G} | \lambda) d\mathcal{G}. \quad (5)$$

Here we do not consider observational uncertainties in galaxy properties to focus exclusively on incompleteness effects. We will address the inclusion of survey-dependent galaxy weight uncertainties in a subsequent work.

2.2. Redshift prior

The core of the methodology presented in this paper is the construction of a redshift prior informed by a galaxy catalog that includes host galaxy probability and completeness. The underlying assumption is that CBC events only occur inside galaxies. The correlation between BBHs and their host galaxy properties is an active area of research, and various theoretical models predict different dependencies based on formation channels, metallicity effects, and dynamical processes. In this work, we adopt a two-step approach to model these correlations: (i) we treat stellar mass as a property that varies on a galaxy-by-galaxy basis; and (ii) we model the BBH merger rate density as a parametric function in z . So we can rewrite the redshift prior of Eq. (5) as

$$p(z, \hat{\Omega}, \mathcal{G} | \lambda_c, \lambda_z) \propto p_{\text{gal}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) \frac{\psi(z | \lambda_z)}{(1+z)}, \quad (6)$$

where $p_{\text{gal}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c)$ is the probability that a galaxy with properties \mathcal{G} at redshift z and sky location $\hat{\Omega}$ hosts a GW event, $\psi(z | \lambda_z)$ is the BBH merger rate density, and the factor $(1+z)^{-1}$ converts the source-frame time to the detector-frame time.

As a first step, we introduce the case for a complete galaxy catalog that contains all potential host galaxies. In this case, we can define $p_{\text{gal}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) = p_{\text{cat}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c)$ and compute this probability as a weighted mean over the contribution of each galaxy, g , in the catalog

$$p_{\text{cat}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) = \frac{\sum_g w_g p(z, \mathcal{G} | \tilde{z}_g, \lambda_c) \delta(\hat{\Omega} - \hat{\Omega}_g)}{\sum_g w_g}, \quad (7)$$

where w_g is the weight of the g -th galaxy, $p(z, \mathcal{G} | \tilde{z}_g, \lambda_c)$ the galaxy redshift posterior, and δ is a Dirac delta distribution of the galaxy sky position which can be treated as errorless. Equation (7) assumes that the uncertainties in w are negligible. The cosmological dependence enters because we do not know the true redshift, since the galaxy catalog contains observed redshift measurements \tilde{z}_g and associated uncertainties $\tilde{\sigma}_{z,g}$. From these quantities, which we assumed to be Gaussian, we constructed the likelihoods,

$$p(z | \tilde{z}_g, \lambda_c) = \frac{\mathcal{N}(z; \tilde{z}_g, \tilde{\sigma}_{z,g}^2) p_{\text{bkg}}(z, \mathcal{G} | \lambda_c)}{\int \mathcal{N}(z; \tilde{z}_g, \tilde{\sigma}_{z,g}^2) p_{\text{bkg}}(z, \mathcal{G} | \lambda_c) dz}. \quad (8)$$

where p_{bkg} is our prior knowledge of the host background distribution. For instance, a typical assumption for a complete galaxy catalog is a uniform distribution in comoving volume, $p_{\text{bkg}} \propto dV_c/dz$. Alternatively, one could adopt a flat prior (e.g., Gray et al. 2023). We further discuss this quantity in the context of an incomplete catalog in Sect. 2.3.

2.3. Incompleteness

In GW cosmology, a catalog is considered ‘‘complete’’ when it contains all potential BBH host galaxies. In practice, galaxy observations are inevitably affected by selection effects that can depend on galaxy properties \mathcal{G} , limiting the completeness of the catalog. Therefore, Eq. (7) has to be extended as follows (Chen et al. 2018; Finke et al. 2021):

$$p_{\text{gal}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) = f_{\mathcal{R}} p_{\text{cat}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) + (1 - f_{\mathcal{R}}) p_{\text{miss}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c), \quad (9)$$

with

$$f_{\mathcal{R}} \equiv \int_0^{z_{\text{max}}} P_{\text{comp}}(z, \hat{\Omega}, \mathcal{G}) p_{\text{bkg}}(z, \mathcal{G} | \lambda_c) dz, \quad (10)$$

where P_{comp} is the completeness function, p_{bkg} is the background distribution of the potential hosts, and $f_{\mathcal{R}}$ is an overall completeness fraction that weights the contribution of the catalog term p_{cat} (Eq. 7) over the term representing the missing hosts p_{miss} . Importantly, $f_{\mathcal{R}}$ must be calculated over a sufficiently large region up to z_{max} that includes all GW events that can be detected.

The completeness function is defined in the domain $[0, 1]$ and can be computed as

$$P_{\text{comp}}(z, \hat{\Omega}, \mathcal{G}) = \min\left(\frac{n_{\text{cat}}(z, \mathcal{G})}{n_{\text{theo}}(z, \mathcal{G})}, 1\right), \quad (11)$$

where $n_{\text{cat}}(z)$ is the catalog number density and $n_{\text{theo}}(z)$ is the theoretical target number density, both depending on the properties of the host galaxies, \mathcal{G} . Compared to the galaxy catalog term in Eq. (7), $n_{\text{cat}}(z)$ has to be computed over larger volumes to ensure adequate statistics and smooth out local inhomogeneities. We note that this approach, also used in Finke et al. (2021), differs from the one used in `gwcosmo` and `icarogw`, where the observed catalog distribution is instead modeled as the integral of a theoretical distribution down a limiting magnitude (Gray et al. 2023; Mastrogiovanni et al. 2024). The key advantage of measuring the catalog density with a more physically motivated method is that it enables modeling completeness effects beyond simpler magnitude cuts, such as color-color selection or stellar mass cuts, more flexibly representing the selection functions of current galaxy surveys.

The distribution of galaxies can be described empirically using the parametric formulation introduced by Schechter (1976). It is convenient to express the theoretical number density of potential hosts as

$$n_{\text{theo}}(z, \mathcal{G}) = \int_{M_{\star, \text{min}}(z)}^{M_{\star, \text{max}}(z)} \Phi(M_{\star}, z) w(M_{\star}, z) dM_{\star}, \quad (12)$$

where $\Phi(M_{\star}, z)$ is the Schechter mass function, $M_{\star, \text{min}}(z)$ and $M_{\star, \text{max}}(z)$ are the minimum and maximum masses of galaxies hosting BBH mergers, and $w(M_{\star}, z)$ accounts for potential host weighting assumptions (see Sect. 3.1 and Sect. 3.2). The use of Schechter functions is widespread in recent standard siren analyses, where galaxy luminosity or absolute magnitude is typically used as a weighting factor (e.g., Abbott et al. 2023c; Mastrogiovanni et al. 2023; Turски et al. 2025). Here, we instead weight by stellar mass directly to build a physically consistent framework tied directly to host galaxy properties. Moreover, value-added catalogs for stellar masses are expected to be released for current wide-field surveys such as DESI (DESI Collaboration 2024) and *Euclid* (Euclid Collaboration: Mellier et al. 2025). The background probability distribution of potential host galaxies, p_{bkg} can then be obtained by multiplying the theoretical galaxy density by the comoving volume element,

$$p_{\text{bkg}}(z | \lambda_c, \mathcal{G}) \propto n_{\text{theo}}(z, \mathcal{G}) \frac{dV_c(z)}{dz}. \quad (13)$$

In general, this quantity depends on cosmological parameters, λ_c , but not on H_0 , as n_{theo} scales as H_0^3 and dV_c/dz as H_0^{-3} . Finally, under the assumption that missing galaxies are homogeneously distributed with respect to the background, we can write

$$p_{\text{miss}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) = \frac{1 - P_{\text{comp}}(z, \hat{\Omega}, \mathcal{G})}{1 - f_{\mathcal{R}}} p_{\text{bkg}}(z, \mathcal{G} | \lambda_c). \quad (14)$$

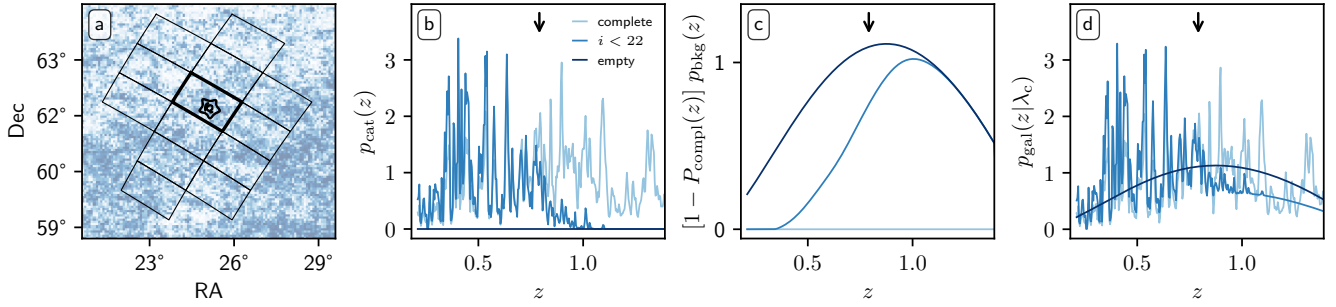


Fig. 1. Effects of galaxy catalog incompleteness within the CHIMERA framework. (a) Sky localization and pixelization (represented with the grid) for a GW event from a stellar mass-weighted MICECATv2 catalog (the underlying blue points). The marker shows the host true position and pixel, from which the distributions in the subsequent panel are computed. (b) Catalog redshift distribution for complete (light blue), $i < 22$ cut (blue), and empty catalogs (dark blue flat line). The arrow shows the true host redshift. (c) Completion terms accounting for missing galaxies; dark blue shows the background distribution used in standard siren constraints. (d) Resulting galaxy distribution terms, individually normalized for better comparison.

This approach, known as “homogeneous completion,” differs from “multiplicative completion,” which assumes that missing galaxies trace the distribution of those present in the catalog (Finke et al. 2021), and “variance completion,” which incorporates galaxy correlation lengths (Dalang & Baker 2024; Leyde et al. 2024, 2026). In this work, we adopt the homogeneous completion method, which is self-consistent with the incompleteness effects simulated in our analysis (see Sect. 3.5). With all the necessary components in place, the galaxy term corrected for galaxy catalog incompleteness and host weighting becomes

$$p_{\text{gal}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) = \int P_{\text{comp}}(z, \hat{\Omega}, \mathcal{G}) p_{\text{bkg}}(z, \mathcal{G} | \lambda_c) dz \left[p_{\text{cat}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) + [1 - P_{\text{comp}}(z, \hat{\Omega}, \mathcal{G})] p_{\text{bkg}}(z, \hat{\Omega}, \mathcal{G} | \lambda_c) \right] \quad (15)$$

2.4. Implementation in CHIMERA

This section summarizes the implementation of the method in CHIMERA (Borghi et al. 2024; Tagliacruzchi et al. 2025). The latest release (v2) presented in Tagliacruzchi et al. (2025) is fully implemented in the JAX framework (Bradbury et al. 2018), enabling hyperparameter vectorization and GPU acceleration, and includes more efficient population functions and likelihood estimation algorithms. Together with this paper, we release an updated version of CHIMERA (v2.1) that enables proper treatment of galaxy catalog incompleteness and host weighting effects¹.

Likelihood. The likelihood computation (Eq. 1) uses kernel density estimation (KDE) to model the GW data posterior. The KDEs are then integrated in equal-area Healpix pixels (Gorski et al. 2005) that divide the sky area of the GW event (see Fig. 1a). CHIMERA employs three algorithms of increasing computational efficiency: (i) a full 3D kernel in (RA, Dec, z); (ii) a “many-1d” approach that marginalizes the 3D posterior using a 1D redshift KDE for each sky pixel; and (iii) a “single-1d” algorithm in which the GW information is collapsed onto a single sky pixel. The GW probability is then evaluated on a redshift grid where the KDE has support, considering all the hyperparameter space. In this paper, the (ii) approach is used. Although both approximate methods (ii) and (iii) show negligible differ-

ences from (i) for the datasets analyzed here (Tagliacruzchi et al. 2025), we adopt the more conservative “many-1d” method.

Catalog term. To handle an efficient computation, the catalog term (Eq. 7) of a given pixelized GW dataset can be pre-computed and saved to be loaded separately. This is crucial as there might be more than 1 million galaxies in a pixel. This quantity is H_0 -independent as the numerator and denominator of the galaxy redshift posterior (Eq. 8), scales in the same way as H_0 , but must be re-computed for each host weighting scheme and galaxy catalog. Figure 1b shows an example for three catalogs: complete, i -band magnitude limited at $i < 22$, and empty. The arrow indicates the redshift of the true host galaxy.

Completeness. The completeness function is computed in multiple N_{masks} masks following the approach of Finke et al. (2021). Healpix pixels are grouped together using a k-means algorithm based on the number of galaxies inside each pixel. For each mask, the completeness function (Eq. 11) is evaluated by binning the observed galaxies into redshift bins and comparing their density to a theoretical background distribution. As for the catalog term, the completeness function can be saved to disk and reloaded to be later interpolated onto the specific redshift grid where the likelihood is evaluated. Note that the pixelization scheme used for the completeness can be coarser than that of the GW data. In this work, we adopt the default settings ($N_{\text{masks}} = 5$, $n_{\text{side}} = 32$, and 100 redshift bins) that have been chosen to ensure adequate statistics within each mask.

Background distribution. The background galaxy distribution is obtained by integrating the Schechter mass function over the relevant host mass range with any applied host weighting (Eq. 12). This term is used in both the (pre)computation of the completeness function, P_{comp} , and, crucially, is evaluated at each likelihood call and therefore must be defined efficiently, as it directly impacts the overall computational performance of the pipeline. In CHIMERA, this is optimized using just-in-time (JIT) compilation and vectorization across all redshift grids. An example illustrating the background galaxy distributions is shown in Figure 1c.

Galaxy term. The catalog and completeness correction terms are combined as in Eq. (15). Figure 1d illustrates this interplay

¹ Available at <https://github.com/CosmoStatGW/CHIMERA>

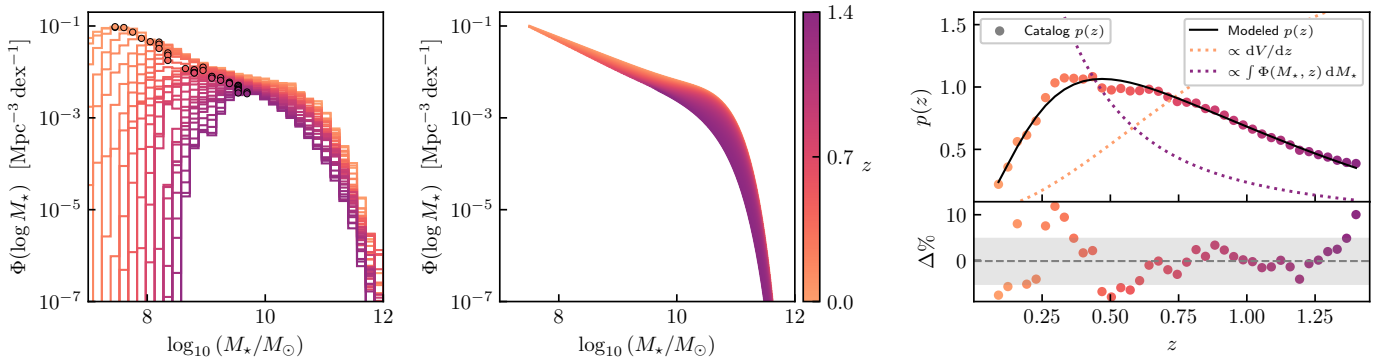


Fig. 2. Stellar mass and redshift distribution of the parent galaxy catalog. *Left panel:* Stellar mass function in different redshift bins measured from the original MICECATv2 galaxy catalog. The points indicate the peak number density at each redshift bin. *Central panel:* Theoretical stellar mass function modeled as a double Schechter function evolving in redshift. *Right panel:* Redshift distribution of the parent galaxy catalog. The black line shows the theoretical distribution obtained from the product of the cosmological volume element dominating at low- z (dotted orange line) and the integral of the Schechter damping the distribution at high- z (dotted purple line). The bottom panel shows the percentage difference, highlighting the median (gray line) and the 16–84 percentile range (gray band).

at different levels of completeness. For a complete catalog, all information comes from the catalog term, where galaxies are individually weighted according to the assumed hosting probabilities. On the other hand, in the absence of a galaxy catalog, all information comes from the completion term, and host weighting enters through the background distribution. In the intermediate case of a catalog complete down to a given magnitude, the relative contribution varies with redshift, showing a transition near $z \approx 1$ where the completeness begins to decline in the galaxy catalog adopted in this work (see Sect. 3.1).

3. Catalogs

This section describes the generation of mock galaxy and GW catalogs. First, we built our parent galaxy catalog, which represents all galaxies in our synthetic universe and is described by a parametric, redshift-evolving stellar mass function (Sect. 3.1). From this parent catalog, we derive three sub-catalogs of potential BBH hosts by applying different host weighting schemes (Sect. 3.2). For each host catalog, we generate mock GW events following a prescribed population model (Sect. 3.3) and simulate their detection with a Fisher matrix approach (Sect. 3.4). Finally, we apply observational completeness cuts to the parent galaxy catalog to simulate realistic survey conditions (Sect. 3.5). The parent galaxy catalog, either complete or with cuts applied, serves as the input catalog for the parameter estimation runs. Further details on this framework are provided in Appendix A.

3.1. Parent galaxy catalog

The mock galaxy catalog used in this work is obtained from the MICE Grand Challenge light-cone simulation (MICECATv2), which populates one-eighth of the sky (5157 deg^2) and reproduces a Dark Energy Survey-type catalog completeness down to an observed magnitude of $i < 24$ out to redshift $z = 1.4$ (Fosalba et al. 2015a,b; Crocce et al. 2015; Carretero et al. 2015; Hoffmann et al. 2015). The simulation assumes a flat Λ CDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_{m,0} = 0.25$, and $\Omega_{\Lambda,0} = 0.75$.

We selected a subsample of MICECATv2 so that the redshift distribution can be described with a parametric function and efficiently computed at each likelihood call of CHIMERA, while pre-

serving realistic galaxy clustering. We first computed the stellar mass distribution of the original catalog, defining the minimum mass as the mass corresponding to the peak density at each redshift bin (Fig. 2, left panel). We then modeled this distribution with a redshift-evolving double Schechter function (Fig. 2, central panel), where the evolution is driven by the minimum mass. Lastly, we subsampled the original catalog so that our model can describe the background galaxy distribution of the catalog with a typical percentage difference below 5% (Fig. 2, right panel). To compute the theoretical number density (Eq. 12) the Schechter function is integrated from a redshift-evolving minimum mass to $\log_{10}(M_*/M_\odot) \approx 12$. Further details are given in Appendix A. The resulting catalog, identified as the parent galaxy catalog throughout this paper, contains approximately 335 million galaxies (about 67% of MICECATv2) and is more than 200 times larger than the high-mass subsample of MICECATv2 galaxies used in B24.

3.2. Potential host galaxies

The connection between BBH and host galaxy properties is a subject of ongoing research. On theoretical grounds, population synthesis studies suggest that low- z BBH mergers preferentially occur in galaxies with higher stellar mass and more metal-rich (e.g., Artale et al. 2020; Santoliquido et al. 2022). On the observational side, GW population studies are only beginning to explore these dependencies (e.g., Vijaykumar et al. 2024).

In this study, we assign the probability that each galaxy hosts a BBH merger as a function of its stellar mass without applying mass cuts. We considered three weighting schemes:

$$w \propto M_\star^{\alpha_M}, \quad \text{where } \alpha_M = 0, 1, 2. \quad (16)$$

These cases are compared in Fig. 3 alongside the weighting adopted in B24 and the low- z trend from the population synthesis study of Artale et al. (2020). Despite the uniform ($\alpha_M = 0$) weighting being maximally agnostic, it unrealistically assigns equal hosting probability to all galaxies, regardless of their stellar mass, across a range spanning several orders of magnitude. Applying a stellar-mass cut is a practical attempt to mitigate this issue. The $\alpha_M = 1$ weighting is more in line with theoretical predictions in the local Universe, where most of the BBH events are observed (see Sect. 3.4). Finally, the $\alpha_M = 2$ weighting allows

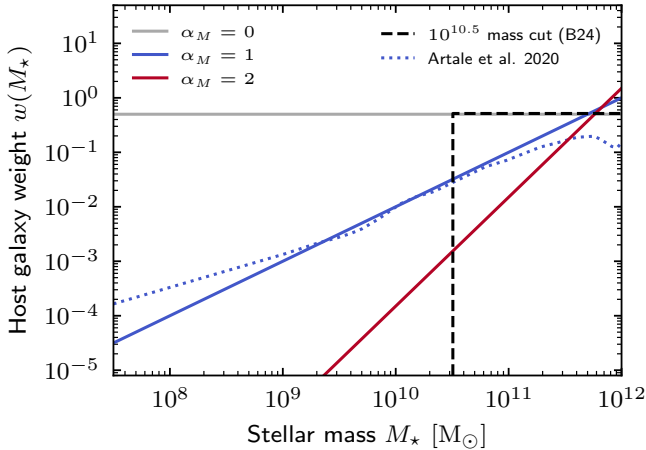


Fig. 3. Host galaxy weighting schemes considered in this work: constant (gray), linear (blue), and quadratic (red) mass weighting. For comparison, we show the $z = 0.1$ trend from Artale et al. (2020) (dotted blue line) and the cut used in previous work (solid line, as in Borghi et al. 2024).

us to explore the case of an even stronger dependence on stellar mass.

We remark that the parent catalog is complete above a given minimum stellar mass that evolves with redshift (Fig. 2). In the likelihood, we integrate the Schechter function over this mass range to ensure statistical consistency. This allows us to probe the regime in which the galaxy density is highest and therefore least informative for the dark siren method. In practice, this occurs for $\alpha_M = 0$ at low z , where most of the events are located. However, since low-mass galaxies are increasingly missed at higher redshift, caution would be required if extending this analysis to constrain $H(z)$ at higher z with the $\alpha_M = 0$ catalog (or even more so with $\alpha_M < 0$), as the galaxy density of this parent sample may not necessarily reflect the true underlying population. This issue does not affect mocks generated with $\alpha_M > 0$, since the weights preferentially select high-mass galaxies where the fidelity of the parent sample is highest.

3.3. Mock GW catalogs

For each weighting scheme, we populated the corresponding host galaxy catalog with BBH mergers, generating three mock GW catalogs hereafter labeled as mock0, mock1, mock2 for $\alpha_M = 0, 1, 2$ respectively. We assumed that the overall BBH merger rate follows the cosmic star formation rate density from Madau & Dickinson (2014):

$$\psi(z; \lambda_z) \propto \frac{(1+z)^\gamma}{1 + \left(\frac{1+z}{1+z_p}\right)^{\gamma+\kappa}}, \quad (17)$$

where $\psi(z) \propto (1+z)^\gamma$ at low z , peaks at z_p , and then declines as $\psi(z) \propto (1+z)^{-\kappa}$. The redshifts are then converted to luminosity distances by assuming the cosmology of MICECATv2. For the mass distribution, we adopt the phenomenological model ‘‘PowerLaw+Peak’’ (PLP, Talbot & Thrane 2018), which is the preferred model up to LVK GWTC-3 (Abbott et al. 2023b). Equation (4) is then factorized as

$$p(m_1, m_2 | \lambda_m) = p(m_1 | \lambda_m) p(m_2 | m_1, \lambda_m), \quad (18)$$

where the primary mass $p(m_1 | \lambda_m)$ is modeled as a superposition of a power-law with slope α in the domain

$m_1 \in [m_{\text{low}}, m_{\text{high}}]$ and a Gaussian centered at μ_g with width σ_g and mixing fraction λ_g . The low-mass boundary is smoothed by a factor set by δ_m . The secondary mass is modeled as a power-law with index β in the domain $m_2 \in [m_{\text{low}}, m_1]$.

In summary, our GW population is described by these twelve cosmological and astrophysical hyperparameters:

$$\begin{aligned} \lambda_c &= \{H_0\} \\ \lambda_z &= \{\gamma, k, z_p\} \\ \lambda_m &= \{\alpha, \beta, \delta_m, m_{\text{low}}, m_{\text{high}}, \mu_g, \sigma_g, \lambda_g\}. \end{aligned} \quad (19)$$

The fiducial values and prior ranges chosen for this work are the same as those in B24 (see their Table 1), ensuring that the forecasts can be directly compared in terms of the GW population. In particular, we fix $\Omega_{m,0}$ to the value adopted in MICEcatv2.

3.4. Detected GW events

To simulate detections, we use GWFAST (Iacovelli et al. 2022a,b). We assumed quasi-circular non-precessing BBH systems so that the following detector frame parameters characterize each waveform:

$$\theta^{\text{det}} = \{\mathcal{M}, \eta, d_L, \theta, \phi, \iota, \chi_1^z, \chi_2^z, \psi, t_c, \Phi_c\}, \quad (20)$$

where \mathcal{M} is the detector-frame chirp mass, η is the symmetric mass ratio, d_L is the luminosity distance, $\theta = \pi/2 - \text{Dec}$ and $\phi = \text{RA}$ are the sky position angles, ι refers to the inclination angle of the binary’s orbital angular momentum with respect to the line of sight, $\chi_{1/2,z}$ are the dimensionless spin parameters along the direction of the orbital angular momentum, ψ is the polarization angle, t_c is the coalescence time, and Φ_c is the phase at coalescence. The first five parameters are drawn from mock GW catalogs presented in Sect. 3.3, while the remaining parameters follow standard priors: inclination angle $\cos \iota \sim \mathcal{U}(0, \pi)$; spin components $\chi_{1,2}^z \sim \mathcal{U}(-1, 1)$; polarization angle $\psi \sim \mathcal{U}(0, \pi)$, coalescence time $t_c \sim \mathcal{U}(0, 1)$ (in units of fraction of a day), and coalescence phase $\Phi_c \sim \mathcal{U}(0, 2\pi)$. For each source, we simulated GW emission and computed the network signal-to-noise ratio (S/N) using the IMRPhenomHM (London et al. 2018; Kalaghatgi et al. 2020) waveform approximant. In this work, we consider an O5-like network configuration², which includes the two LIGO (Ligo & Aasi 2015), Virgo (Acernese et al. 2015), and KAGRA (Aso et al. 2013) instruments, as well as a LIGO detector located in India (Abbott et al. 2016). We assume a minimum frequency of 10 Hz and a duty cycle of 100%.

We selected a subsample of 100 BBHs with a network signal-to-noise ratio, $S/N > 25$. This cut, chosen to be consistent with B24, is designed to yield the 100 best BBH events over approximately one year of observation. In general, higher S/N events have tighter sky localization and fewer potential host galaxies, resulting in stronger H_0 constraints in the dark siren regime. In contrast, pure spectral siren analyses (empty catalog limit), included here for comparison, would benefit from including lower S/N events to better resolve features in the mass distribution (see e.g., Abbott et al. 2023b).

Figure 4 presents the main properties of the 100 BBH events generated for each of the three host weighting schemes. The top

² The amplitude spectral densities can be found at <https://dcc.ligo.org/LIGO-T2000012/public>. In particular, we use AplusDesign for the three LIGO detectors, avirgo_05low_NEW for Virgo, and kagra_80Mpc for KAGRA.

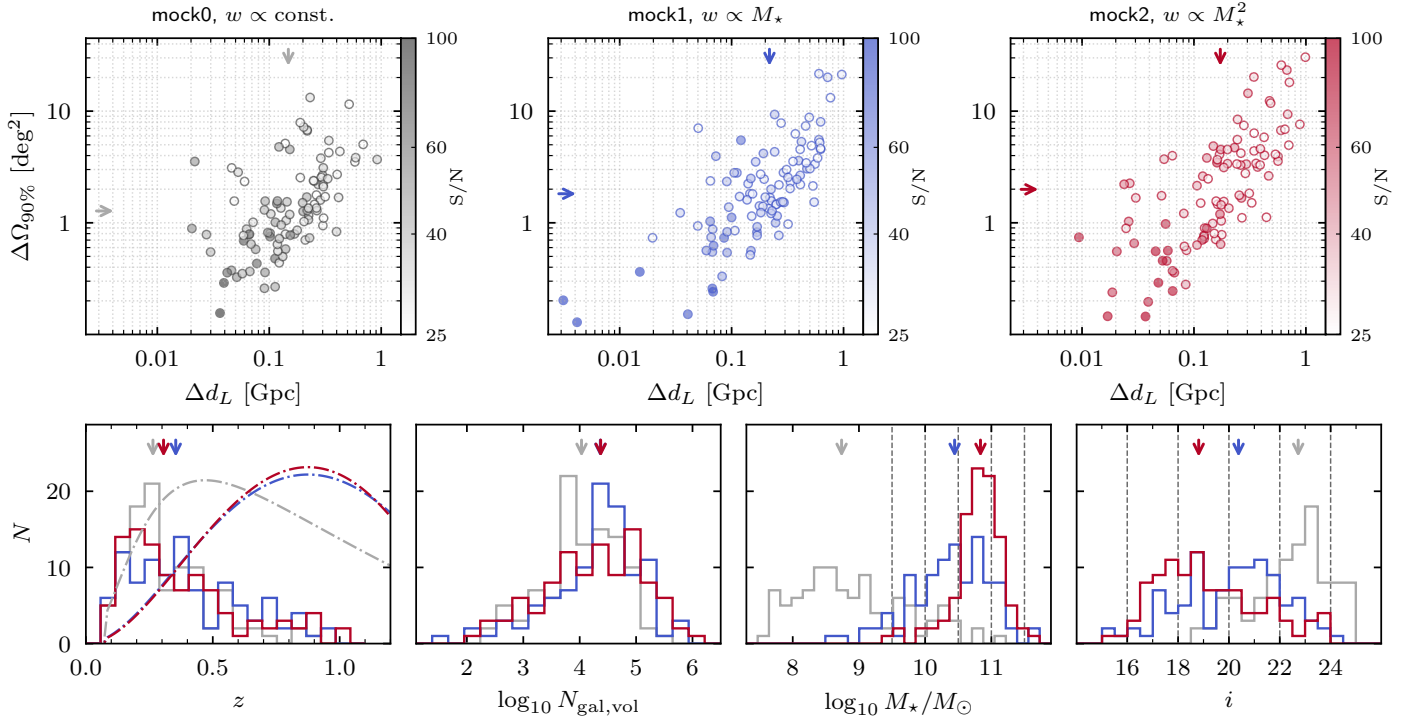


Fig. 4. Main properties of 100 O5-like BBH events with $S/N > 25$ drawn from MICECATv2 with three host weighting schemes: unweighted (mock0, gray), weighted by stellar mass (mock1, blue), and weighted by stellar mass squared (mock2, red). *Top panels:* Distribution in the sky localization (90% C.L.) versus luminosity-distance uncertainty plane color-coded by the network S/N . *Bottom panels:* Distributions of source redshift, number of galaxies within the localization volume ($N_{\text{gal,vol}}$), host galaxy stellar mass, and i -band magnitude for the various samples; the arrows indicate the median values. The dot-dashed lines show the background redshift distribution of potential BBH host galaxies used to generate GW events. The vertical dashed line shows the mass and magnitude cuts that are applied in this work to study incompleteness effects.

panels show the 90% C.L. sky localization area against the luminosity distance uncertainty, color-coded by S/N . The bottom panels show the redshift distribution, galaxy counts within the localization volume ($N_{\text{gal,vol}}$), and host galaxy properties (stellar mass M_* and i -band magnitude). Note that $N_{\text{gal,vol}}$ assumes a flat prior on $H_0 \in [60, 80] \text{ km s}^{-1} \text{ Mpc}^{-1}$. The observed GW events are located at $z \lesssim 1$, with sky localization areas ranging from 0.1 to a few tens of square degrees. The values of $N_{\text{gal,vol}}$, a key indicator of how strongly each GW event can constrain H_0 in the dark siren regime, range from $O(10^2)$ to $O(10^5)$, with the maximum set by the finite redshift coverage of the current catalog. These values are significantly larger than those in B24, where the best-localized events contained only a few galaxies and could therefore be treated as “pseudo-bright” sirens. Differences among the three mocks reflect both the background host galaxy distribution (see Fig. 4, bottom left panel) and the specific realization of the sampled GW events. Due to the lower redshift distribution of their hosts, mock0 events are observed with slightly higher S/N and better sky localization, with median values of 34.1 and 1.3 deg^2 , respectively, compared to mock1 (32.5 and 1.8 deg^2) and mock2 (31.6 and 2.0 deg^2). As a result, the median $N_{\text{gal,vol}}$ differs between the mocks: ≈ 6500 for mock0 and ≈ 12000 for both mock1 and mock2.

We compute selection effects by generating a large set of injections with GWFAST using the same $S/N > 25$ threshold applied to the construction of the GW catalog. We verify that these events span the full range of detectable luminosity distances and binary masses. The full population includes 2×10^7 generated sources, resulting in 1×10^6 detections at $S/N > 25$ in the O5-like scenario considered in this work. This dataset is then used to estimate the selection effects following Eq. (3).

3.5. Observed galaxy catalogs

Galaxy catalogs are generally constructed to achieve a well-defined completeness threshold, typically set by apparent magnitude or stellar mass limits. Magnitude-complete surveys maintain uniform flux thresholds across the survey footprint. The Galactic plane, which covers approximately one third of the full sky, is typically surveyed at shallower magnitudes or excluded from deeper surveys because of severe dust extinction, which limits the detection of fainter galaxies. Some well-known examples include SDSS (York et al. 2000), GAMA (Driver et al. 2011), DESI (DESI Collaboration 2024), zCOSMOS (Lilly et al. 2007), and VIPERS (Guzzo et al. 2014). Mass-complete catalogs are built from subsamples of galaxies that are selected to ensure homogeneity in terms of intrinsic physical properties. As a result, these catalogs are better suited to study the evolution of galaxies throughout cosmic time (e.g., Pozzetti et al. 2010; Weaver et al. 2023).

In this work, we considered the case of a spectroscopic galaxy catalog with uncertainties on the galaxy redshifts of $\sigma_z = 0.001(1+z)$. We then generated catalogs including incompleteness effects by removing galaxies according to the following mass and magnitude cuts:

$$\begin{aligned} \log_{10} M_* > \log_{10} M_{*,\text{cut}} \in \{9.5, 10, 10.5, 11, 11.5\} \\ i < i_{\text{cut}} \in \{24, 22, 20, 18, 16\} \end{aligned} \quad (21)$$

We also examine two limiting cases: a complete catalog (no selection cuts) representing ideal survey conditions, and an empty catalog (spectral siren analysis only) representing the worst-case scenario where no potential host galaxies are observed.

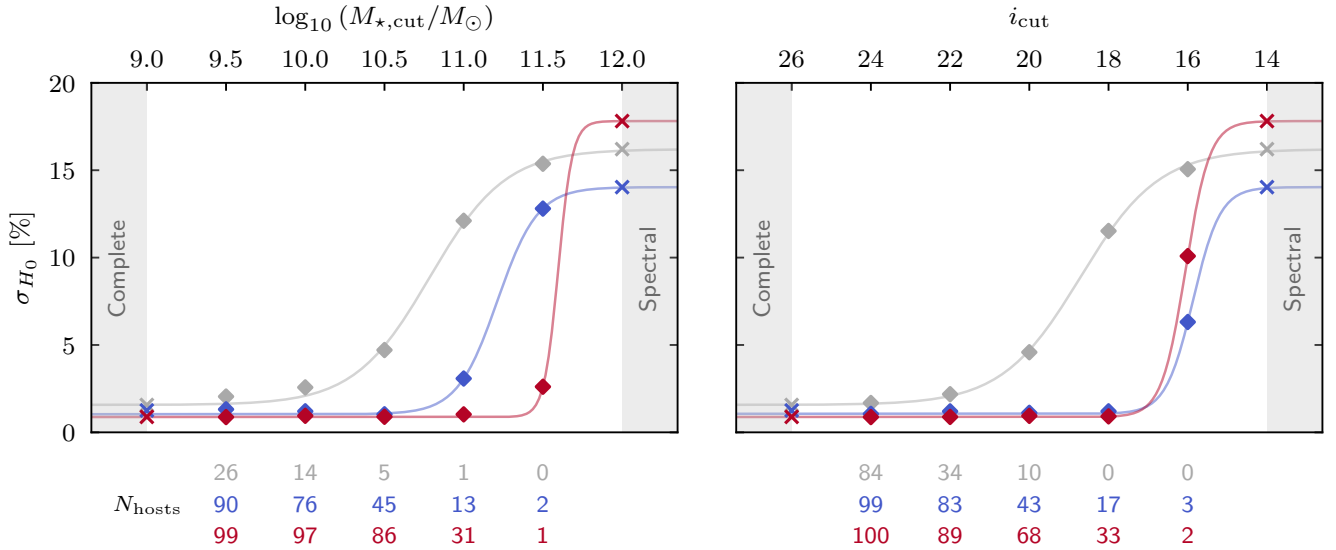


Fig. 5. Effects of galaxy survey incompleteness on H_0 constraints for different stellar mass (left panel) and i -band magnitude (right panel) completeness cuts across three host weighting schemes: unweighted (mock0, gray), weighted by stellar mass (mock1, blue), and weighted by stellar mass squared (mock2, red). The values annotated below each panel indicate the number of host galaxies retained in the catalog after applying each cut.

4. Results

This section presents the results of the full standard siren analysis for all 36 configurations, combining three host galaxy weighting schemes (Eq. 16) with twelve catalog completeness cuts (Eq. 21, including the complete and empty cases).

The posterior distribution is sampled with the affine-invariant MCMC sampler emcee (Foreman-Mackey et al. 2013), adopting wide uniform priors. We ensure convergence by requiring that the number of samples is at least $50\times$ the integrated autocorrelation time for each hyperparameter. The entire analysis, enabled by the enhanced version of CHIMERA (Tagliacruzchi et al. 2025), took about 50 k CPU hours on the LEONARDO supercomputer hosted by CINECA (Turisini et al. 2024). The results are quoted as median values with symmetric 68.3% credible intervals.

Figure 5 shows the marginalized constraints on H_0 in all the completeness and weighting configurations. The full table of results can be found in Appendix B. We obtained unbiased estimates of H_0 within 1σ across all configurations, even at high levels of incompleteness. The remaining hyperparameters are all recovered within 2σ . This demonstrates the robustness of both the statistical methodology and the framework used to generate mock catalogs. In the following sections, we present the results for the complete catalog, empty catalog, and intermediate completeness configurations. We then examine the correlations between H_0 and the hyperparameters describing the astrophysical population. Finally, we assess potential biases that may arise from wrong assumptions on the host galaxy weighting.

4.1. Complete galaxy catalog

When the galaxy catalog contains all potential hosts, we obtain percent-level constraints on H_0 . The precision improves when host weighting favors more massive galaxies, with $\sigma_{H_0} = 1.6\%$, 1.3% , 0.9% for mock0, mock1, mock2, respectively. The main reason for this improvement is the decrease in the effective number of galaxies within the GW localization volume, $N_{\text{gal,vol}}$, that significantly contribute to the catalog term once the weighting is applied. We verify that the trend is independent of the spe-

cific realization of GW events by performing one-dimensional analyses on the H_0 posterior, keeping all the other hyperparameters fixed to their fiducial values. This is further supported by the fact that mock0, despite having lower $N_{\text{gal,vol}}$ values (see Fig. 4, bottom panels), provides the weakest constraints.

These results are remarkable given that typical $N_{\text{gal,vol}}$ values are $200\times$ larger than in B24. Even with this increased galaxy density, constraints degrade by only a factor of two in the least informative case. We attribute this to the fact that, even when the analysis is extended to more realistic and denser catalogs, the redshift uncertainties provided by a spectroscopic galaxy catalog remain sufficiently small to preserve the information content of the catalog term (Eq. 7). Consequently, through the combined analysis of multiple events, the true value of H_0 can still be statistically recovered with high precision.

4.2. Empty galaxy catalog

When no galaxies are included, cosmological constraints are obtained with a pure spectral siren approach, which relies on assumptions on the source-frame BBH mass distribution to break the mass-redshift degeneracy. We find $\sigma_{H_0} = 16\%$, 14% , 18% for mock0, mock1, mock2, respectively. These results are consistent with forecasts based on similar methodologies and BBH population assumptions (Mancarella et al. 2022; Leyde et al. 2022; Borghi et al. 2024). Unlike the complete catalog case, the trend with the weighting exponent α_M is no longer present here, as the constraints depend more strongly on the specific realization of the GW events, in particular how well the relevant features of the mass distribution are mapped. These spectral siren results establish the limiting case for studying how cosmological constraints degrade as galaxy catalog completeness decreases.

4.3. Modeling incompleteness effects on H_0

To describe how galaxy catalog incompleteness affects the precision of H_0 constraints, we modeled the evolution of the

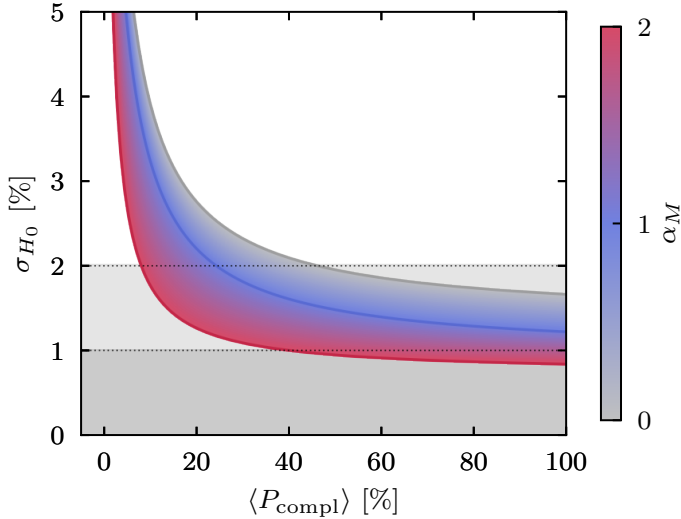


Fig. 6. Percentage of uncertainty on marginalized H_0 constraints as a function of the average completeness and host mass weighting exponent α_M obtained from a full standard siren analysis of 100 O5-like BBH events.

percentage uncertainty, σ_{H_0} , using a sigmoid function that smoothly connects the case of a complete catalog ($\sigma_{H_0}^{\text{comp}}$) to the empty catalog/spectral siren limit ($\sigma_{H_0}^{\text{spec}}$):

$$\sigma_{H_0}(x) = \frac{\sigma_{H_0}^{\text{spec}} - \sigma_{H_0}^{\text{comp}}}{1 + \exp[-k(x - x_{\text{th}})]} + \sigma_{H_0}^{\text{comp}}, \quad (22)$$

where x is the chosen completeness parameter (such as stellar mass or magnitude), x_{th} is the threshold at which the uncertainty transitions most rapidly between the two regimes, and k sets the steepness of the transition.

For mass-completeness cuts, we find $x_{\text{th}} \equiv \log_{10} M_{\text{th}} = 10.8, 11.2, 11.6$ and $k = 4.2, 11.1, 22.4$ for *mock0*, *mock1*, and *mock2*, respectively. As α_M increases, the threshold shifts toward higher stellar masses, indicating that missing less massive (less probable) hosts ultimately has a negligible impact on the H_0 constraints. Notably, with mass and mass-squared weighting, percent-level precision is still achievable as long as the catalog is complete at $\log_{10}(M_{\star}/M_{\odot}) > 11$. The annotations in Fig. 5 show the number of true host galaxies remaining after each completeness cut. Percent-level constraints are obtained when at least about 20 events retain their hosts after completeness cuts or, equivalently, when the catalog completeness is high enough to keep a sufficient fraction of true host galaxies in the catalog. Interestingly, there is one configuration (*mock2* with $\log_{10}(M_{\text{cut}}/M_{\odot}) = 11.5$) for which this is possible thanks to a single lucky event. We underscore the importance of spectroscopic follow-ups to target the localization region of these best-constraining ‘‘golden’’ events with future galaxy surveys.

For magnitude-completeness cuts, we find $x_{\text{th}} \equiv i_{\text{th}} = 18.7, 15.9, 16.0$ and $k = 1.0, 3.0, 3.3$ for *mock0*, *mock1*, and *mock2*, respectively. We do not observe a clear trend of the threshold with α_M , as the difference between the *mock1* and *mock2* results is mainly driven by the specific realization of the GW events analyzed. This is also expected given that the distributions of i -band magnitudes of the true hosts are quite similar in the two mocks (see Fig. 4 lower right). In both *mock1* and *mock2*, percent-level constraints can be obtained with catalogs containing all potential BBH hosts down to magnitudes $i < 17$. However, we emphasize that realistic catalogs are charac-

terized by specific and more complex selection functions; therefore, caution should be taken when extrapolating these results to the observed surveys.

To summarize our results, we computed the comoving volume-averaged completeness within the GW detector horizon,

$$\langle P_{\text{compl}} \rangle \equiv \frac{1}{V_c(z_{\text{hor}})} \int_0^{z_{\text{hor}}} P_{\text{compl}}(z) dV_c \quad \text{with } z_{\text{hor}} = 1.4, \quad (23)$$

where we recall that P_{compl} refers to potential BBH host galaxies rather than the full galaxy population and, being integrated over the GW horizon, is non-local. While a more local definition would better capture the completeness of the most informative events and correlate more directly with the H_0 constraints, we adopt this integrated metric to ensure a consistent comparison across all the different configurations.

Figure 6 shows the trends of σ_{H_0} at varying $\langle P_{\text{compl}} \rangle$, smoothed with a Gaussian kernel and interpolated as a function of α_M . In the worst-case scenario of uniform weighting, achieving a 2% constraint on H_0 requires a spectroscopic catalog with at least 50% completeness within the GW detector horizon. In the more realistic case of mass weighting, the same completeness level would provide a 1.5% constraint, while achieving 1% precision is only possible if BBH hosting probability scales with the stellar mass squared.

4.4. Correlation with astrophysical parameters

The study of correlations between H_0 and astrophysical hyperparameters across the different configurations provides insights into how strongly cosmological constraints depend on assumptions about the BBH mass distribution. When the catalog term is not informative, a strong anticorrelation typically appears between H_0 and the mass scale parameter μ_g . This happens because higher H_0 values shift sources to higher z at fixed luminosity distance, requiring smaller source-frame masses to match the observations. This behavior has been observed in real data (e.g., Abbott et al. 2023c; Mali & Essick 2025; Abac et al. 2025b), in mock analyses when using photometric galaxy catalogs (Borghi et al. 2024), and is clearly evident in this work as the completeness decreases.

We generally find a strong anti-correlation between H_0 and μ_g at average completeness levels below 10%, with Spearman correlation coefficients computed from MCMC chains reaching values of $\rho \approx -0.7$. When the catalog completeness is greater than 50%, these correlations weaken substantially. Percent-level constraints on H_0 are obtained only when $|\rho| < 0.15$, at which point the anti-correlation is broken for all three mocks. Figure 7 illustrates the case of *mock0*, where the anti-correlation is broken at completeness levels above 25%, which correspond, in our mock catalogs, to a stellar mass completeness cut of $\log_{10}(M_{\star}/M_{\odot}) > 10$. We also observed weaker, less significant anti-correlations between H_0 and the edges of the mass distribution m_{high} and m_{low} .

4.5. Mismodeling hosting weights

We investigate potential biases introduced by incorrect host galaxy weighting assumptions by re-analyzing the GW events from *mock1* using incorrect weighting schemes during parameter estimation (i.e., $\alpha_M = 0$ and $\alpha_M = 2$), while keeping the parent catalog and all other assumptions fixed. These weighting choices propagate through the likelihood by affecting individual galaxy weights (Eq. 7), the completeness function (Eq. 11), and

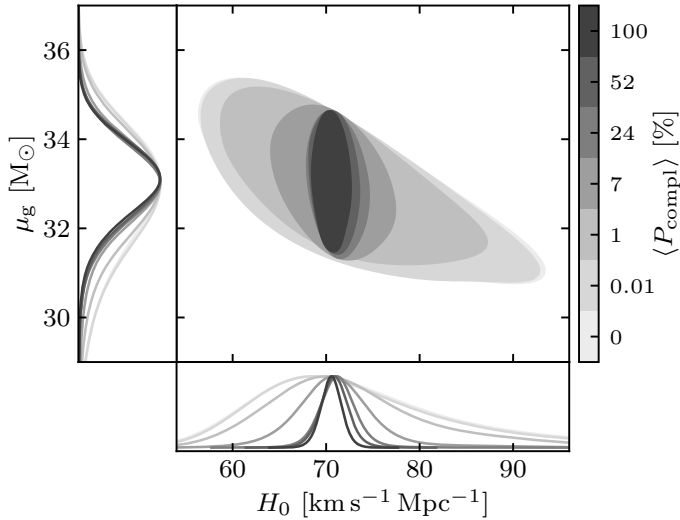


Fig. 7. Constraints on the $H_0 - \mu_g$ plane obtained with 100 O5-like BBH events from the unweighted (mock 0) dataset at varying galaxy catalog average completeness.

the background distribution (Eq. 13). Further details and supporting figures are provided in Appendix C.

For complete spectroscopic catalogs, incorrect weighting introduces negligible systematic bias ($|\Delta H_0| < 1 \text{ km s}^{-1} \text{ Mpc}^{-1}$), with mass-squared weighting surprisingly producing tighter constraints. At moderate average completeness levels ($\approx 20\%$), a mass weighting stronger than the true one ($\alpha_M = 2$) still yields unbiased results, while uniform weighting ($\alpha_M = 0$) introduces biases of $\approx 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$. When catalogs become highly incomplete or empty (spectral siren regime), both incorrect weighting assumptions produce biases, exceeding 1σ for $\alpha_M = 0$ and 2σ for $\alpha_M = 2$. This demonstrates the importance of accurately modeling the host population when galaxy catalog information becomes sparse.

Our results are in agreement with previous studies of host galaxy weighting systematics. In particular, [Hanselman et al. \(2025\)](#) and [Alfradique et al. \(2025\)](#) also find unbiased H_0 constraints in the case of a complete catalog with stellar-mass mismodeled by equal weighting. In this work, we confirm the existence of a regime where host weight mismodeling does not affect H_0 constraints also at higher level of incompleteness. The robustness of these constraints is likely due to the assumption of very informative catalog terms with spectroscopic redshift uncertainties. However, assumptions on the GW data can also play a role. Dedicated mock data challenges will be crucial to compare analysis frameworks and better assess these effects.

5. Conclusions

In this work, we have studied the cosmological potential of standard sirens, focusing on the impact of galaxy catalog incompleteness and host galaxy weighting on H_0 constraints. Below, we summarize the main results of the paper.

1. We presented a self-consistent framework to study galaxy catalog incompleteness and host weighting effects in standard siren analyses and implement it in the publicly available CHIMERA pipeline (v2.1)³. The updated version of the code is designed to be more modular and efficient for large galaxy catalogs.

2. We built our parent sample from the MICECATv2 mock galaxy catalog and described it with a parametric stellar mass distribution. We then generated three realizations of 100 BBH events at $S/N > 25$ detectable by a LIGO-Virgo-KAGRA O5-like network over about one year of observation. Each realization is based on a different host galaxy weighting scheme: unweighted, weighted by stellar mass, and weighted by stellar mass squared. We explore standard siren constraints achievable with spectroscopic catalogs, simulating incompleteness through stellar mass or magnitude cuts.
3. We obtained percent-level constraints on H_0 even at moderate incompleteness. In the best-case scenario of a complete spectroscopic catalog, precision reaches 1.6%, 1.3%, and 0.9% for no weighting, mass weighting, and mass-squared weighting, respectively. Remarkably, 2% precision is achievable even when the completeness of potential hosts, averaged within the GW horizon, reaches 50%. Achieving 1% precision in the assumed configurations is only possible if BBH hosting probability scales with the stellar mass squared.
4. We characterized how H_0 constraints degrade with decreasing catalog completeness using a sigmoid function. We find that the completeness threshold and the steepness of the sigmoid transition increase for higher mass weighting schemes, demonstrating quantitatively that missing progressively more low-mass host galaxies has negligible impact on the cosmological constraints.
5. We studied the correlations between H_0 and the astrophysical population parameters. A strong anti-correlation between H_0 and the BBH mass scale parameter μ_g (with Spearman coefficients $\rho \approx -0.7$) is present at average completeness levels below $\approx 10\%$. This correlation weakens substantially above 50% completeness. Percent-level constraints on H_0 are achieved when $|\rho| < 0.15$, demonstrating the transition from the spectral to the dark siren regime.
6. We assessed systematic effects from mismodeling host galaxy weighting. With complete spectroscopic catalogs, incorrect weighting does not introduce significant H_0 biases. At moderate average completeness ($\approx 20\%$), stronger-than-true mass weighting yields unbiased results, while weaker weighting introduces $10 \text{ km s}^{-1} \text{ Mpc}^{-1}$ biases. For highly incomplete or empty catalogs (spectral siren regime), mismodeling the weighting in either direction introduces significant biases, highlighting the importance of accurately modeling the host population when galaxy catalog information becomes sparse.

These results are based on some simplifying assumptions. First, we assume that galaxy properties are measured with negligible uncertainty. Second, we assume that we know the true underlying galaxy distribution. Third, we simulate catalog incompleteness with simple stellar mass and magnitude cuts, while actual surveys follow more complex selection functions as they might target different galaxy populations and be subject to different selection effects. In future work, we plan to test the impact of these assumptions by applying realistic survey conditions tailored to ongoing and upcoming large-scale surveys. We also plan to assess the impact of photometric redshift measurements on our results, with particular focus on the trade-off between catalog completeness and the precision of redshift measurements. Together, these steps will be crucial to push the cosmological constraints to higher z and study the expansion history $H(z)$. Finally, our study assumes a duty cycle of 100% at nominal detector sensitivity for all instruments in LVK O5-like configuration, providing a simple benchmark for expected yearly O5

³ Available at <https://github.com/CosmoStatGW/CHIMERA>

performance. Under more realistic observing conditions, achieving the same precision on H_0 would require proportionally longer observation times. Despite these limitations, our results highlight the crucial importance of spectroscopic galaxy surveys for standard siren cosmology. Moreover, our framework demonstrates that even moderately incomplete spectroscopic catalogs can yield competitive cosmological constraints, even when hosting probabilities are not properly modeled.

To put our results in context, current constraints from GWTC-4.0 provide $\approx 20\%$ uncertainties on H_0 from 141 CBC dark sirens, using a galaxy catalog where only about one quarter of the galaxies have spectroscopic redshifts (Abac et al. 2025b). Our forecast of 1–2% precision with 100 O5-like BBH events and a spectroscopic catalog confirms and extends the findings of our previous work (B24), which achieved percent-level constraints using a complete sample of massive hosts with $\log_{10}(M_{\star}/M_{\odot}) > 10.5$. Here we validate those results across the full galaxy-mass range and present a comprehensive end-to-end analysis of catalog incompleteness and host weighting effects on dark-siren constraints. Our approach employs realistic mock catalogs, marginalization on the astrophysical population parameters, and hierarchical inference consistent with current LVK cosmology analyses, extending earlier explorations (Gray et al. 2020) and going beyond more simplified analytical models used in recent studies (Perna et al. 2025; Hanselman et al. 2025; Alfradique et al. 2025; Cross-Parkin et al. 2025). Our results further demonstrate that competitive H_0 measurements will be achievable with an O5-like configuration with moderately complete spectroscopic surveys, and that host weight mismodeling does not necessarily introduce significant systematics on H_0 even at moderate completeness levels.

These findings are particularly relevant for ongoing galaxy surveys such as *Euclid* (Euclid Collaboration: Mellier et al. 2025), DESI (DESI Collaboration 2024), and the proposed WST mission (Mainieri et al. 2024). With the increasing number of GW detections, it will be crucial to understand the interplay between the GW sources and host galaxy properties. Looking ahead, third-generation GW detectors such as the Einstein Telescope (Punturo et al. 2010; Abac et al. 2025a) and Cosmic Explorer (Reitze et al. 2019) will be transformative in terms of detections, potentially revealing more detailed structures in the GW population and correlation with host galaxy properties (Chen et al. 2024). At the same time, space-based detectors such as LISA (Amaro-Seoane et al. 2023) will allow us to test these correlations at completely different black hole scales. Together, these advancements will be crucial to unlock the full potential of GWs as cosmological probes.

Acknowledgements. We thank Michele Mancarella and Lucia Pozzetti for helpful suggestions and comments. NB and MM acknowledge support from the grant ASI n. 2024-10-HH.0 “Attività scientifiche per la missione *Euclid* – fase E”. MM acknowledges support from MIUR, PRIN 2022 (grant 2022NY2ZRS 001). MM acknowledges the financial contribution from the grant PRIN-MUR 2022 2022NY2ZRS 001 “Optimizing the extraction of cosmological information from Large Scale Structure analysis in view of the next large spectroscopic surveys” supported by Next Generation EU. MT acknowledges the funding from the European Union – NextGenerationEU, in the framework of the HPC project – “National Center for HPC, Big Data and Quantum Computing” (PNRR – M4C2 – I1.4 – CN00000013 – CUP J33C22001170001). GC is supported by ERC Starting Grant No. 945155–GWmining, Cariplo Foundation Grant No. 2021-0555, MUR PRIN Grant No. 2022-Z9X4XS, Italian-French University (UIF/UF) Grant No. 2025-C3-386, MUR Grant “Progetto Dipartimenti di Eccellenza 2023-2027” (BiCoQ), and the ICSC National Research Centre funded by NextGenerationEU. This material is based upon work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation. We also acknowledge the CINECA award under the ISCRA initiative for the availability of high perfor-

mance computing resources through ISCRA-C project LIGEA HP10CXW8OV and HP10CD9GBT (PI N. Borghi). *Software:* CHIMERA (Borghi et al. 2024; Tagliazucchi et al. 2025), GWFAST (Iacovelli et al. 2022b), numpy (Harris et al. 2020), JAX (Bradbury et al. 2018), emcee (Foreman-Mackey et al. 2013), matplotlib (Hunter 2007), datashader (Bednar et al. 2025), pastamarkers (PASTA Collaboration 2024).

References

- Abac, A., Abramo, R., Albanesi, S., et al. 2025a, ArXiv e-prints [arXiv: 2503.12263]
- Abac, A. G., Abouelfettouh, I., Acernese, F., et al. 2025b, ArXiv e-prints [arXiv: 2509.04348]
- Abac, A. G., Abouelfettouh, I., Acernese, F., et al. 2025c, ArXiv e-prints [arXiv: 2508.18083]
- Abac, A. G., Abouelfettouh, I., Acernese, F., et al. 2025d, ArXiv e-prints [arXiv: 2508.18082]
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, *Liv. Rev. Relativ.*, 19, 1
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, *Nature*, 551, 85
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, *Phys. Rev. Lett.*, 119, 161101
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, *Phys. Rev. X*, 9, 31040
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021, *Phys. Rev. X*, 11, 21053
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2023a, *Phys. Rev. X*, 13, 41039
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2023b, *Phys. Rev. X*, 13, 011048
- Abbott, R., Abe, H., Acernese, F., et al. 2023c, *ApJ*, 949, 76
- Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, *Classical Quantum Gravity*, 32, 24001
- Agarwal, A., Dupletsa, U., Leyde, K., et al. 2025, *ApJ*, 987, 47
- Alfradique, V., Bom, C. R., & Castro, T. 2025, *Phys. Rev. D*, 112, 063561
- Amaro-Seoane, P., Andrews, J., Arca Sedda, M., et al. 2023, *Liv. Rev. Relativity*, 26, 2
- Artale, M. C., Bouffanais, Y., Mapelli, M., et al. 2020, *MNRAS*, 495, 1841
- Aso, Y., Michimura, Y., Somiya, K., et al. 2013, *Phys. Rev. D*, 88, 43007
- Bednar, J. A., Crail, J., Thomas, I., et al. 2025, <https://doi.org/10.5281/Zenodo.3844614>
- Borghi, N., Mancarella, M., Moresco, M., et al. 2024, *ApJ*, 964, 191
- Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, <http://github.com/jax-ml/jax>
- Bulla, M., Coughlin, M. W., Dhawan, S., & Dietrich, T. 2022, *Universe*, 8, 289
- Callister, T. A. 2024, ArXiv e-prints [arXiv: 2410.19145]
- Carretero, J., Castander, F. J., Gaztañaga, E., Croce, M., & Fosalba, P. 2015, *MNRAS*, 447, 646
- Chen, H.-Y., Fishbach, M., & Holz, D. E. 2018, *Nature*, 562, 545
- Chen, H.-Y., Ezquiaga, J. M., & Gupta, I. 2024, *Classical Quantum Gravity*, 41, 125004
- Chernoff, D. F., & Finn, L. S. 1993, *ApJ*, 411, L5
- Croce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, *MNRAS*, 453, 1513
- Cross-Parkin, M. L., Howlett, C., Davis, T. M., & Khetan, N. 2025, *PASA*, 42, e149
- Dalang, C., & Baker, T. 2024, *JCAP*, 2024, 24
- Del Pozzo, W. 2012, *Phys. Rev. D*, 86, 43011
- DESI Collaboration (Adame, A. G., et al.) 2024, *AJ*, 167, 62
- Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, *MNRAS*, 413, 971
- Edelman, B., Doctor, Z., Godfrey, J., & Farr, B. 2022, *ApJ*, 924, 101
- Euclid Collaboration (Mellier, Y., et al.) 2025, *A&A*, 697, A1
- Ezquiaga, J. M., & Holz, D. E. 2022, *Phys. Rev. Lett.*, 129, 61102
- Farah, A. M., Callister, T. A., Ezquiaga, J. M., Zevin, M., & Holz, D. E. 2025, *ApJ*, 978, 153
- Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, *ApJ*, 883, L42
- Ferri, J., Tashiro, I. L., Abramo, L. R., et al. 2025, *JCAP*, 2025, 008
- Finke, A., Foffa, S., Iacovelli, F., Maggiore, M., & Mancarella, M. 2021, *JCAP*, 2021, 26
- Fishbach, M., Gray, R., Hernandez, I. M., et al. 2019, *ApJ*, 871, L13
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Fosalba, P., Croce, M., Gaztañaga, E., & Castander, F. J. 2015a, *MNRAS*, 448, 2987
- Fosalba, P., Gaztañaga, E., Castander, F. J., & Croce, M. 2015b, *MNRAS*, 447, 1319
- Gair, J. R., Ghosh, A., Gray, R., et al. 2023, *AJ*, 166, 22
- Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, 622, 759
- Gray, R., Hernandez, I. M., Qi, H., et al. 2020, *Phys. Rev. D*, 101, 122001
- Gray, R., Beirnaert, F., Karathanasis, C., et al. 2023, *JCAP*, 2023, 23
- Guzzo, L., Scodreggio, M., Garilli, B., et al. 2014, *A&A*, 566, A108

- Hanselman, A. G., Vijaykumar, A., Fishbach, M., & Holz, D. E. 2025, *ApJ*, **979**, 9
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, *MNRAS*, **447**, 1724
- Holz, D. E., & Hughes, S. A. 2005, *ApJ*, **629**, 15
- Hotokezaka, K., Nakar, E., Gottlieb, O., et al. 2019, *Nat. Astron.*, **3**, 940
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Iacovelli, F., Mancarella, M., Foffa, S., & Maggiore, M. 2022a, *ApJ*, **941**, 208
- Iacovelli, F., Mancarella, M., Foffa, S., & Maggiore, M. 2022b, *ApJS*, **263**, 2
- Kalaghatgi, C., Hannam, M., & Raymond, V. 2020, *Phys. Rev. D*, **101**, 103004
- Leja, J., Speagle, J. S., Johnson, B. D., et al. 2020, *ApJ*, **893**, 111
- Leyde, K., Mastrogiovanni, S., Steer, D., Chassande-Mottin, E., & Karathanasis, C. 2022, *JCAP*, **2022**, 12
- Leyde, K., Baker, T., & Enzi, W. 2024, *JCAP*, **2024**, 13
- Leyde, K., Baker, T., & Enzi, W. 2026, *JCAP01*, **2026**, 13
- Ligo, S. C., Aasi, J., et al. 2015, *Classical Quantum Gravity*, **32**, 74001
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, **172**, 70
- London, L., Khan, S., Fauchon-Jones, E., et al. 2018, *Phys. Rev. Lett.*, **120**, 161102
- Loredo, T. J. 2004, *AIP Conf. Proc.*, **735**, 195
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Mainieri, V., Anderson, R.I., Brinchmann, J., et al. 2024, ArXiv e-prints [arXiv: **2403.05398**]
- Mali, U., & Essick, R. 2025, *ApJ*, **980**, 85
- Mancarella, M., Genoud-Prachex, E., & Maggiore, M. 2022, *Phys. Rev. D*, **105**, 64030
- Mandel, I., Farr, W. M., & Gair, J. R. 2019, *MNRAS*, **486**, 1086
- Mastrogiovanni, S., Leyde, K., Karathanasis, C., et al. 2021, *Phys. Rev. D*, **104**, 062009
- Mastrogiovanni, S., Laghi, D., Gray, R., et al. 2023, *Phys. Rev. D*, **108**, 42002
- Mastrogiovanni, S., Pierra, G., Perriès, S., et al. 2024, *A&A*, **682**, A167
- Moresco, M., Amati, L., Amendola, L., et al. 2022, *Liv. Rev. Relativity*, **25**, 6
- Mukherjee, S., Wandelt, B. D., Nissanke, S. M., & Silvestri, A. 2021, *Phys. Rev. D*, **103**, 043520
- Oguri, M. 2016, *Phys. Rev. D*, **93**, 83511
- Palmese, A., & Mastrogiovanni, S. 2025, ArXiv e-prints [arXiv: **2502.00239**]
- Palmese, A., deVicente, J., Pereira, M. E. S., et al. 2020, *ApJ*, **900**, L33
- Palmese, A., Kaur, R., Hajela, A., et al. 2024, *Phys. Rev. D*, **109**, 063508
- PASTA Collaboration (Borghi, N., et al.) 2024, ArXiv e-prints [arXiv: **2403.20314**]
- Pedrotti, A., Mancarella, M., Bel, J., & Gerosa, D. 2025, ArXiv e-prints [arXiv: **2504.10482**]
- Perna, G., Mastrogiovanni, S., & Ricciardone, A. 2025, *A&A*, **698**, A128
- Pierra, G., & Mastrogiovanni, S. 2025, ArXiv e-prints [arXiv: **2507.10597**]
- Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, *A&A*, **523**, A13
- Punturo, M., Abernathy, M., Acernese, F., et al. 2010, *Class. Quantum Gravity*, **27**, 194002
- Reitze, D., Adhikari, R. X., Ballmer, S., et al. 2019, *BAAS*, **51**, 35
- Santoliquido, F., Mapelli, M., Artale, M. C., & Boco, L. 2022, *MNRAS*, **516**, 3297
- Schechter, P. 1976, *ApJ*, **203**, 297
- Schutz, B. F. 1986, *Nature*, **323**, 310
- Tagliazucchi, M., Moresco, M., Borghi, N., & Fiebig, M. 2025, *A&A*, **702**, A244
- Talbot, C., & Thrane, E. 2018, *ApJ*, **856**, 173
- Taylor, S. R., Gair, J. R., & Mandel, I. 2012, *Phys. Rev. D*, **85**, 023535
- Turisini, M., Cestari, M., & Amati, G. 2024, *J. Large-scale Research Facilities* **JLSRF**, **9**
- Turski, C., Brozzetti, M. L., Dálya, G., Punturo, M., & Ghosh, A. 2025, ArXiv e-prints [arXiv: **2505.13568**]
- Valentino, E. D., Said, J. L., Riess, A., et al. 2025, *Phys. Dark Universe*, **49**, 101965
- Verde, L., Treu, T., & Riess, A. G. 2019, *Nat. Astron.*, **3**, 891
- Vijaykumar, A., Fishbach, M., Adhikari, S., & Holz, D. E. 2024, *ApJ*, **972**, 157
- Vitale, S., Gerosa, D., Farr, W.M., & Taylor, S.R. 2022, ArXiv e-prints [arXiv: **2007.05579**]
- Weaver, J. R., Davidzon, I., Toft, S., et al. 2023, *A&A*, **677**, A184
- York, D. G., Adelman, J., Anderson, J. E., Jr, et al. 2000, *AJ*, **120**, 1579

Appendix A: Mock catalogs generation

Figure A.1 summarizes the mock data generation framework presented in this work. The parent galaxy catalog used in this work is obtained from MICECATv2 (Fosalba et al. 2015a,b; Crocce et al. 2015; Carretero et al. 2015; Hoffmann et al. 2015). To ensure an efficient computation of the galaxy term (Eq. 15) while preserving realistic density and clustering properties, the full catalog was subsampled and modeled assuming a theoretical stellar mass function (SMF). From this catalog, we generated three sub-catalogs of potential host galaxies varying the hosting probability exponent α_M as follows: unweighted ($\alpha_M = 0$), mass weighted ($\alpha_M = 1$), squared mass weighted ($\alpha_M = 2$). For each of these catalogs, we generated a GW population based on the prescriptions of Sect. 3.3 and computed the detected events with a Fisher Matrix approach as described in Sect. 3.4. These GW catalogs were analyzed in CHIMERA together with the galaxy catalog, taking into account catalog incompleteness cuts.

To describe the parent galaxy catalog, we adopted a double Schechter function (e.g., Pozzetti et al. 2010; Leja et al. 2020):

$$\Phi(M, z) dM = \begin{cases} \left[\Phi_1^*(z) \left(\frac{M}{M^*} \right)^{\alpha_1} + \Phi_2^*(z) \left(\frac{M}{M^*} \right)^{\alpha_2} \right] e^{-\frac{M}{M^*} \frac{dM}{M^*}} & \text{if } M \geq M_{\text{lim}}(z), \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A.1})$$

where $\Phi_{1,2}^*$ are the characteristic number densities of the two components, $\alpha_{1,2}$ are the low-mass slope parameters, and M^* is the characteristic mass scale separating the power-law regime ($M \ll M^*$) from the exponential cutoff regime ($M \gg M^*$). M_{lim} is the stellar mass completeness limit, representing the minimum stellar mass above which the galaxy catalog is complete. We incorporate redshift evolution by allowing Φ_1^* , Φ_2^* , and $\log_{10} M_{\text{lim}}$ to vary according to a quadratic function:

$$p(z) = a_0 + a_1(1+z) + a_2(1+z)^2, \quad (\text{A.2})$$

while keeping $\log_{10} M^*$, α_1 , and α_2 fixed. This parameterization ultimately provides a model consistent with recent observational constraints (see Leja et al. 2020; Weaver et al. 2023). The parent galaxy catalog was then built as follows.

1. We binned the original MICECATv2 catalog into 40 redshift bins spanning $0.07 < z < 1.4$ and 40 logarithmic stellar mass bins ranging from $6.5 < \log_{10}(M_*/M_\odot) < 12.2$. Within each redshift bin, we computed the catalog SMF and define M_{lim} as the stellar mass corresponding to the peak number density (see Fig. 2, left panel).
2. We fit a quadratic function (Eq. A.2) to $\log_{10} M_{\text{lim}}(z)$ across redshift bins and remove all galaxies with a stellar mass below M_{lim} in each redshift bin. This is a first step to ensure that the resulting catalog is actually mass-complete above $M_{\text{lim}}(z)$ and can be accurately described by the theoretical SMF.
3. In each redshift bin, the catalog SMF was smoothed with a 1σ Gaussian filter to reduce the statistical fluctuations (Fig. 2, central panel) and fitted with a double Schechter function following Eq. (A.1), with redshift evolution modeled via the quadratic parameterization in Eq. (A.2). To improve the fit convergence, we fix M^* and slopes $\alpha_{1,2}$ to values consistent with the continuity model from Leja et al. (2020) applied to the COSMOS2020 catalog (Weaver et al. 2023). The resulting coefficients are provided in Table A.1.
4. At this point, we had a smooth parametric model of the MICECATv2 SMF across its full mass and redshift range. We validated it by comparing the catalog number density

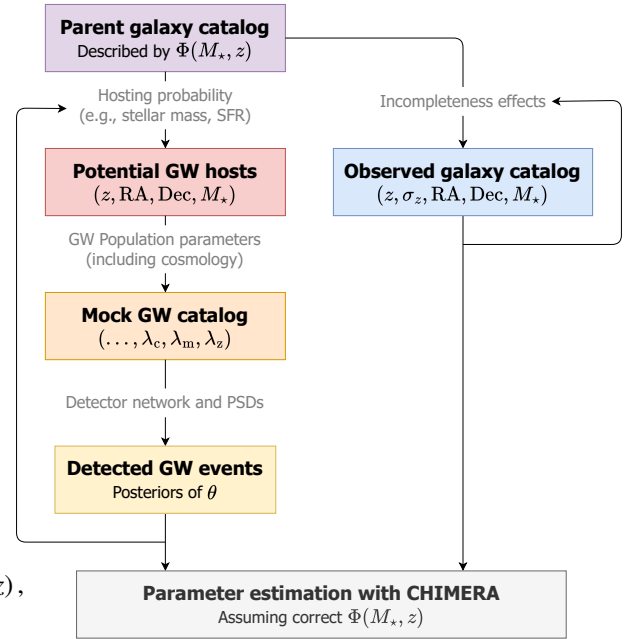


Fig. A.1. Overview of our end-to-end simulation framework for testing the impact of different galaxy catalog completeness cuts and host weighting schemes on standard siren constraints.

Table A.1. Coefficients for the redshift-evolving double Schechter function describing the parent galaxy catalog.

Parameter	a_0	a_1	a_2
$\log_{10} M_{\text{lim}}$	2.90	5.32	-1.04
$\log_{10} \Phi_1$	1.48×10^{-3}	3.44×10^{-3}	-1.68×10^{-3}
$\log_{10} \Phi_2$	3.23×10^{-3}	-2.40×10^{-3}	0.57×10^{-3}
$\log_{10} M^*$	10.5	0	0
α_1	-0.3	0	0
α_2	-1.5	0	0

Notes. Schechter parameters $\{\log_{10} M^*, \alpha_1, \alpha_2\}$ are kept fixed in the fit.

with the theoretical number density obtained by integrating $\Phi(M, z)$ (Fig. 2, central panel). The average difference remains below 5%, with oscillations likely due to synthetic effects inherent in the simulation. This very good agreement demonstrates that our parametric SMF accurately represents the underlying galaxy population.

The resulting catalog, referred to as the parent catalog throughout this paper, contains approximately 335 million galaxies, representing about 67% of the full MICECATv2 catalog. This subsample is more than 200 times larger than the high-mass subsample of MICECATv2 galaxies used in Borghi et al. (2024).

The framework presented here has been implemented in CHIMERA (v2.1) with a vectorized approach to quickly evaluate the Schechter function (Eqs. A.1 and A.2) and its integral across multiple sky pixels. The catalog term and completeness functions can be pre-computed and stored for different sky masks (see Sect. 2). This means that with the above parameterization, it is possible to efficiently compute the background term (Eq. 13) and thus the final galaxy term (Eq. 15).

Table A.2. Constraints on H_0 from 100 BBHs in an LVK O5-like scenario for different galaxy catalog completeness levels and host weightings.

Completeness	mock 0 - $w \propto \text{const.}$			mock 1 - $w \propto M_\star$			mock 2 - $w \propto M_\star^2$		
	$\langle P_{\text{compl}} \rangle$ (%)	N_{host}	H_0 ($\text{km s}^{-1} \text{Mpc}^{-1}$)	$\langle P_{\text{compl}} \rangle$ (%)	N_{host}	H_0 ($\text{km s}^{-1} \text{Mpc}^{-1}$)	$\langle P_{\text{compl}} \rangle$ (%)	N_{host}	H_0 ($\text{km s}^{-1} \text{Mpc}^{-1}$)
Complete	89.8	100	$70.6^{+1.1}_{-1.1}$ (1.6%)	89.6	100	$70.3^{+0.9}_{-0.9}$ (1.3%)	93.8	100	$69.9^{+0.6}_{-0.6}$ (0.9%)
$\log_{10}(M_\star/M_\odot) > 9.5$	51.5	26	$70.9^{+1.5}_{-1.4}$ (2.0%)	85.8	90	$70.4^{+0.9}_{-0.9}$ (1.3%)	93.8	99	$70.1^{+0.6}_{-0.6}$ (0.9%)
$\log_{10}(M_\star/M_\odot) > 10.0$	23.6	14	$71.2^{+1.9}_{-1.8}$ (2.6%)	71.9	76	$70.6^{+0.9}_{-0.8}$ (1.2%)	93.5	97	$70.2^{+0.7}_{-0.6}$ (0.9%)
$\log_{10}(M_\star/M_\odot) > 10.5$	6.94	5	$70.9^{+3.4}_{-3.3}$ (4.7%)	44.5	45	$70.2^{+0.7}_{-0.7}$ (1.0%)	89.5	86	$70.2^{+0.6}_{-0.6}$ (0.9%)
$\log_{10}(M_\star/M_\odot) > 11.0$	0.83	1	$71.2^{+10.5}_{-6.7}$ (12%)	12.6	13	$70.7^{+2.6}_{-1.7}$ (3.1%)	52.3	31	$70.0^{+0.7}_{-0.6}$ (1.0%)
$\log_{10}(M_\star/M_\odot) > 11.5$	0.01	0	$71.2^{+13.9}_{-8.0}$ (15%)	0.53	2	$72.4^{+12.8}_{-5.7}$ (13%)	5.56	1	$68.9^{+1.8}_{-1.8}$ (2.6%)
$i < 24$	65.6	84	$70.5^{+1.1}_{-1.2}$ (1.7%)	76.8	99	$70.4^{+0.7}_{-0.7}$ (1.1%)	85.2	100	$70.2^{+0.6}_{-0.6}$ (0.9%)
$i < 22$	14.4	34	$70.5^{+1.5}_{-1.6}$ (2.2%)	42.6	83	$70.3^{+0.9}_{-0.8}$ (1.2%)	53.5	89	$70.0^{+0.6}_{-0.6}$ (0.9%)
$i < 20$	2.63	10	$69.9^{+3.0}_{-3.4}$ (4.6%)	20.9	43	$70.4^{+0.8}_{-0.8}$ (1.1%)	30.8	68	$70.0^{+0.7}_{-0.6}$ (0.9%)
$i < 18$	0.40	0	$69.9^{+9.2}_{-6.9}$ (12%)	7.22	17	$69.7^{+0.9}_{-0.8}$ (1.2%)	15.2	33	$70.0^{+0.6}_{-0.6}$ (0.9%)
$i < 16$	0.03	0	$70.9^{+13.4}_{-7.9}$ (15%)	1.14	3	$71.4^{+7.7}_{-1.3}$ (6.3%)	4.17	2	$77.0^{+6.9}_{-8.7}$ (10%)
Empty/Spectral	0	0	$71.5^{+14.7}_{-8.5}$ (16%)	0	0	$77.0^{+13.8}_{-7.8}$ (14%)	0	0	$72.2^{+14.9}_{-10.9}$ (18%)

Notes. $\langle P_{\text{compl}} \rangle$ is a comoving volume-weighted average within the GW detector horizon, $0 < z < 1.4$.

Appendix B: Results table

Table A.2 reports the marginalized H_0 constraints at different completeness configurations for the three host galaxy weighting schemes studied in this work. For each configuration, we provide the average completeness, $\langle P_{\text{compl}} \rangle$, computed following Eq. (23). We stress that P_{compl} refers to potential BBH host galaxies not to the full galaxy population (see its definition in Eq. 11). We also include the number of actual host galaxies N_{host} remaining in the catalog after applying observational selection cuts.

Appendix C: Impact of wrong host galaxy weighting

Figure C.1 shows the bias in marginalized H_0 constraints when incorrect host weighting assumptions are applied during parameter estimation. We tested the mass-weighted mock1 as our fiducial case across various completeness cuts. When correctly setting $\alpha_M = 1$, all configurations yield unbiased H_0 constraints within 1σ . The shift in the median H_0 in the spectral siren regime is likely due to the specific realization of GW events. Therefore, wrong configurations are also expected to be skewed toward this value.

For complete spectroscopic catalogs, incorrect weighting introduces systematic errors that are negligible compared to statistical uncertainties (with $|\Delta H_0| < 1 \text{ km s}^{-1} \text{Mpc}^{-1}$). Interestingly, the mass-squared weighting ($\alpha_M = 2$) yields slightly tighter constraints compared to the true weighting. This indicates that for the configurations studied in this work, a stronger down-weighting of low-mass galaxies reduces the effective number of galaxies contributing to the catalog term without introducing substantial bias. As average completeness decreases below $\approx 20\%$, using a stronger-than-true mass weighting still yields unbiased results. This is also expected as $\alpha_M = 1$ and $\alpha_M = 2$ configurations have similar background distributions (see Fig. 4, lower left panel). In contrast, using $\alpha_M = 0$ produces biases of about $10 \text{ km s}^{-1} \text{Mpc}^{-1}$. When catalogs become highly incomplete or empty (spectral siren regime), both incorrect weighting assumptions produce biases, exceeding 1σ for $\alpha_M = 0$ and 2σ for $\alpha_M = 2$.

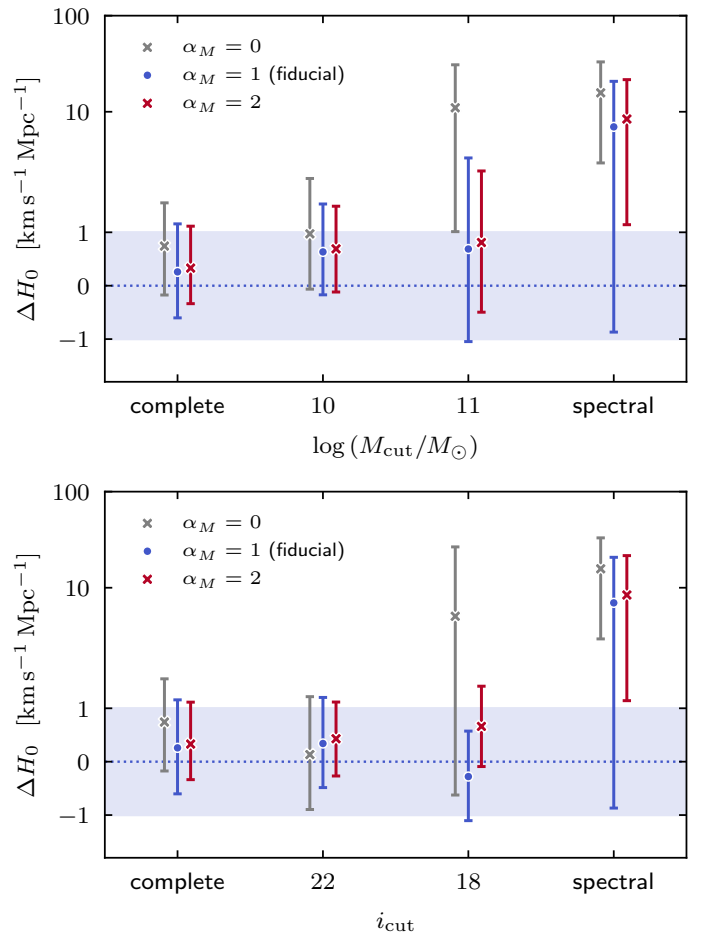


Fig. C.1. Bias on marginalized H_0 constraints at varying completeness levels for 100 BBH detections in an LVK O5-like scenario generated from a mass-weighted catalog (mock 1) assuming three host weighting schemes: constant (wrong), linear (fiducial), and squared (wrong) stellar mass weighting.