

# Study of systematics on the cosmological inference of the Hubble constant from gravitational wave standard sirens

Grégoire Pierra<sup>1,\*</sup>, Simone Mastrogiovanni<sup>2</sup>, Stéphane Perriés<sup>1</sup> and Michela Mapelli<sup>3</sup>

<sup>1</sup>*Université Lyon, Université Claude Bernard Lyon 1, CNRS, IP2I Lyon/IN2P3, UMR 5822, F-69622 Villeurbanne, France*

<sup>2</sup>*INFN, Sezione di Roma, I-00185 Roma, Italy*

<sup>3</sup>*Institut für Theoretische Astrophysik, ZAH, Universität Heidelberg, 69120 Heidelberg, Germany*



(Received 11 January 2024; accepted 13 March 2024; published 9 April 2024)

Gravitational waves (GWs) from compact binary coalescences can constrain the cosmic expansion of the Universe. In the absence of an associated electromagnetic counterpart, the spectral siren method exploits the relation between the detector frame and the source frame masses to jointly infer the parameters of the mass distribution of black holes (BHs) and the cosmic expansion parameter  $H_0$ . This technique relies on the choice of the parametrization for the source mass population of BHs observed in binary black hole mergers (BBHs). Using astrophysically motivated BBH populations, we study the possible systematic effects affecting the inferred value for  $H_0$  when using heuristic mass models like broken power law, power law plus peak, and multipeak distributions. We find that, with 2000 detected GW mergers, the resulting  $H_0$  obtained with a spectral siren analysis can be biased up to  $3\sigma$ . The main sources of this bias come from the failure of the heuristic mass models used so far to account for a possible redshift evolution of the mass distribution and from their inability to model unexpected mass features. We conclude that future dark siren GW cosmology analyses should make use of source mass models able to account for redshift evolution and capable to adjust to unforeseen mass features.

DOI: [10.1103/PhysRevD.109.083504](https://doi.org/10.1103/PhysRevD.109.083504)

## I. INTRODUCTION

September 2015 marks the date of the first detection of a gravitational wave (GW) [1] emitted by the merger of a binary black hole (BBH). Since then, the LIGO-Virgo-KAGRA scientific collaboration has detected more than 90 GW signals from compact binary coalescences (CBCs) [2–4]. These detected GW signals enable us to directly measure the luminosity distances  $d_L$  of the sources without making any assumption on the cosmological model. However, GW signals do not directly provide the source redshift unless an electromagnetic (EM) counterpart is observed. The redshift is required to measure the Universe’s expansion. This is an interesting prospect given the current tension on the measurements of the Universe’s local expansion rate, the Hubble constant  $H_0$  [5]. The multiple GW detections opened the field of GW cosmology. To exploit the tens of GW sources observed without an EM counterpart, additional methods have been developed to obtain the source redshift. The “spectral siren” [6] method relies on the intrinsic degeneracy between the redshift  $z$  of a source and its detector  $m^{\text{det}}$  and source frame  $m^{\text{s}}$  masses, through the relation  $m^{\text{det}} = (1+z)m^{\text{s}}$ . The redshift of the sources is implicitly estimated by measuring

the detector masses and jointly fitting the source mass distribution and cosmological parameters [7–9]. This method depends on the choice of the phenomenological models for the BBH source mass distribution and their capacity to describe the true underlying BBH population [6,7,10–18]. The spectral siren method is also tightly connected with the “galaxy catalog method” [15–17,19], which identifies the possible galaxy hosts of the GW source by using galaxy surveys. Even with the extra information from galaxy surveys, BBH source mass models still have to be used. In fact, in the case where the galaxy survey is highly incomplete, the source mass models are the ones that provide the implicit redshift of the GW sources [8].

In this study, we explore how the commonly used phenomenological source mass models can fit complex and astrophysically motivated distributions of BBH mergers. The rationale of this study is to understand how a mismatch between the source mass models and the population of BBHs would translate to an  $H_0$  bias. To make this study, we start with a BBH population generated synthetically by modeling several astrophysical processes. Then we perform a hierarchical Bayesian inference [9] to measure  $H_0$  alongside the phenomenological BBH mass models.

This paper is organized as follows. In Sec. II we describe the framework built to simulate the GW detections and introduce the basic principles of hierarchical Bayesian

\*g.pierra@ip2i.in2p3.fr

inference. We also present the phenomenological BBH population models for the mass spectrum. In Sec. III, we discuss several sanity checks to test the validity of the framework. We also investigate the robustness of the spectral siren method when a wrong mass model is used for the inference and when the source mass spectrum includes an unmodeled redshift evolution. In Sec. IV, we consider a spectral siren analysis using a realistic catalog of BBH mergers, finding a biased estimate of  $H_0$  if simple phenomenological mass models are employed. Section IV also argues about the possible different sources of the bias, specifically the importance of the redshift evolution of the mass spectrum when estimating the cosmological parameters. Finally, in Sec. V we draw the conclusions for this study.

## II. BAYESIAN INFERENCE, GW OBSERVATIONS, AND CBC POPULATION MODELING

In this section, we describe our simulation procedure. We first simulate BBH observations using a detection criteria based on the signal-to-noise ratio (SNR) that we describe in Sec. II A. Then we use the hierarchical Bayesian inference, summarized in Sec. II B, to reconstruct cosmological and population properties.

### A. A fast generator of GW observations

The workflow that we use to generate the GW observations is the following. From a population of BBH mergers described by the two source masses and the redshift of the merger, we compute an approximate value of the optimal SNR. Instead of calculating the SNR from the full waveform and the matched filter, we use the relation below as a proxy for the optimal SNR [7,20],

$$\rho = \delta \times 9 \left[ \frac{\mathcal{M}_c}{\mathcal{M}_{c,9}} \right]^{\frac{5}{6}} \left[ \frac{d_{L,9}}{d_L} \right], \quad (1)$$

where  $\mathcal{M}_c$  is the binary detector chirp mass,  $d_L$  is the luminosity distance,  $\rho$  is the optimal SNR, and  $\delta$  is a projection factor. Although Eq. (1) can underestimate the SNR of events with chirp masses above  $30M_\odot$ , it should suffice to provide us with a representative sample of the BBH population. The luminosity distance and the binary chirp mass are computed assuming a flat  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmology with  $H_0 = 67.8 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $\Omega_{m,0} = 0.308$ , consistent with cosmic microwave background measurements from Planck [21]. The parameters with a superscript of “9” are the detector chirp mass and luminosity distance values for which a signal with  $\delta = 1$  would have a SNR of 9. We set the reference chirp mass to  $\mathcal{M}_{c,9} = 25 M_\odot$  and the luminosity distance  $d_{L,9} = 1.5 \text{ Gpc}$  [22]. The projection factor  $\delta$  captures the variations of the optimal SNR when dealing with a three-detector network and encapsulates the combined impact of the polarization of the GWs and the sky localization. We

draw it according to a cumulative density function introduced in [23].

To mimic the presence of detector noise in the data, we also calculate a “detected” SNR  $\rho^{\text{det}}$  from the optimal SNR. The detected SNR is drawn from a  $\chi^2$  distribution with 6 degrees of freedom (two for each detector). A GW is detected if its detected SNR exceeds a SNR threshold of 12 and the GW frequency corresponding to the innermost stable circular orbit (ISCO) is higher than 15 Hz. The GW and ISCO frequencies are related by

$$f_{\text{GW}} = 2f_{\text{ISCO}}, \quad (2)$$

where  $f_{\text{ISCO}}$  is approximated as [24]

$$f_{\text{ISCO}} = \frac{1}{2\pi} \frac{1}{6^{3/2}} \frac{c^3}{G} \left( \frac{1M_\odot}{M_{\text{tot}}^d} \right) 10^3 \text{ Hz} \quad (3)$$

For this study, we chose to not generate errors on the detector frame masses and luminosity distances. We assume that we are perfectly able to measure them from the GW signal. We make this choice for two motivations. The first is that we want to ease the computational load of the analysis and explore multiple test cases with thousands of GW detections. The second is that we want to maximize the potential effect of the systematics arising from the phenomenological mass models, which could be hidden by the error budget on the masses.

### B. The hierarchical likelihood

The detection of GW events can be described by an inhomogeneous Poisson process in the presence of selection biases [7–9,25]. For a set of  $N_{\text{GW}}$  GW signals detected over an observation time  $T_{\text{obs}}$ , the probability of having a specific GW dataset  $\{x\}$  given population hyperparameters  $\Lambda$  can be written as

$$\mathcal{L}(\{x\}|\Lambda) \propto e^{-N_{\text{exp}}(\Lambda)} \prod_i^{N_{\text{GW}}} T_{\text{obs}} \int d\theta dz \times \mathcal{L}_{\text{GW}}(x_i|\theta, z, \Lambda) \frac{1}{1+z} \frac{dN_{\text{CBC}}}{d\theta dz dt_s}(\Lambda). \quad (4)$$

The hyperparameters  $\Lambda$  describe the BBH population and also include cosmological expansion parameters. For a flat  $\Lambda$ CDM cosmology, the expansion parameters are the Hubble constant  $H_0$  and the fraction of matter density today  $\Omega_{m,0}$ . Other population hyperparameters are the ones controlling the shape of the BBH mass spectrum. We describe the phenomenological models for the BBH mass spectrum in Sec. II C and in more detail Appendix A 1. In Eq. (4) the individual GW likelihood  $\mathcal{L}_{\text{GW}}(x_i|\theta, \Lambda)$  gauges the errors on the estimation of the intrinsic parameters  $\theta$  for a certain GW event  $x_i$ . For our study, the intrinsic parameters  $\theta = (m_1^s, m_2^s)$  denote the masses in the source frame. In Eq. (4), the factor  $\frac{1}{1+z}$  accounts for the time

dilatation between the detector and source frame times  $t_d$  and  $t_s$ .

$$\frac{dN_{\text{CBC}}}{d\theta dz dt_s}(\Lambda) \quad (5)$$

is the CBC merger rate per source time, comoving volume, and masses. Finally,  $N_{\text{exp}}$  is the expected number of GW detections during a time  $T_{\text{obs}}$ . This term encapsulates the GW selection biases and can be written as

$$N_{\text{exp}}(\Lambda) = T_{\text{obs}} \int d\theta dz P_{\text{det}}(z, \theta, \Lambda) \frac{dN_{\text{CBC}}}{d\theta dz dt_s}(\Lambda) \frac{1}{1+z}, \quad (6)$$

where  $P_{\text{det}}(z, \theta, \Lambda)$  represents the probability of detecting a certain CBC at a redshift  $z$ , characterized by some specific GW parameters  $\theta$  and given cosmological hyperparameters  $\Lambda$ . The detection probability is given by the integral of the GW likelihood over all the detectable data realizations, namely,

$$P_{\text{det}}(z, \theta, \Lambda) = \int_{x \in \text{det}} dx \mathcal{L}_{\text{GW}}(x|\theta, z, \Lambda). \quad (7)$$

The spectral siren CBC rate is then parametrized as

$$\begin{aligned} \frac{dN_{\text{CBC}}}{d\theta dz dt_s} &= \frac{dN_{\text{CBC}}}{d\vec{m}_s dV_c dt_s} \frac{dV_c}{dz}(\Lambda) \\ &= R_0 \psi(z; \Lambda) p_{\text{pop}}(\vec{m}_s|\Lambda) \times \frac{dV_c}{dz}. \end{aligned} \quad (8)$$

In Eq. (8),  $R_0$  is the local ( $z = 0$ ) CBC merger rate per time per comoving volume, the primary source frame masses are defined by  $\vec{m}_s = (m_1^s, m_2^s)$ , and  $\psi(z; \Lambda)$  is the phenomenological function for the CBC merger rate that evolves with redshift with  $\psi(z = 0; \Lambda) = 1$ . The  $p_{\text{pop}}$  term is a probability density function of the phenomenological models for the source frame masses  $p_{\text{pop}}(\vec{m}_s|\Lambda)$ . The last term in Eq. (8) is the differential of the comoving volume per redshift, which for a flat  $\Lambda$ CDM cosmology is

$$\frac{dV_c}{dz} = \left[ \frac{c}{H_0} \right]^3 \left[ \int_0^z \frac{dz'}{E(z')} \right]^2, \quad (9)$$

where the term  $E(z)$  in Eq. (9) is the dimensionless Hubble parameter defined as

$$E(z) = \frac{H(z)}{H_0} = \sqrt{\Omega_m(1+z)^3 + (1-\Omega_m)}. \quad (10)$$

An important assumption that we implicitly made in Eq. (8) when writing  $p_{\text{pop}}(\vec{m}_s|\Lambda)$  is that the CBC mass spectrum does not evolve in redshift. This assumption is typically made in most of the state-of-the-art literature [8,9,26,27]. In the following, we will discuss how the assumption for the source mass distribution to be independent on redshift could translate to an  $H_0$  bias.

### C. The phenomenological population models of BBHs

Phenomenological CBC merger rate models of masses and redshift are central to this study. Currently, spectral siren analysis as in [26] uses simple parametric models describing the distribution of source frame masses and CBC merger rate. For the masses, we employ three of these models: the broken power law (BPL), the power law plus peak (PLP), and the multipeak (MLTP). The three mass models are represented in Fig. 1. The CBC merger rate as a function of redshift is parametrized following the Madau and Dickinson [28] star formation rate. In Appendixes A 1 and A 2, we provide the detailed probability density functions for the models and the set of population parameters and prior ranges used in the rest of the paper. In the following, we provide a general description of these models.

The BPL model [29] is the simplest model that we employ. It is composed of a power law, with a smooth tapering at its low mass edge, plus a break in the power law at a mass  $m_{\text{break}}$ . The low mass smoothing is introduced to model the effects of the stellar progenitor metallicity [29], which could create a smooth transition for black hole production. The break in the power law mimics the left

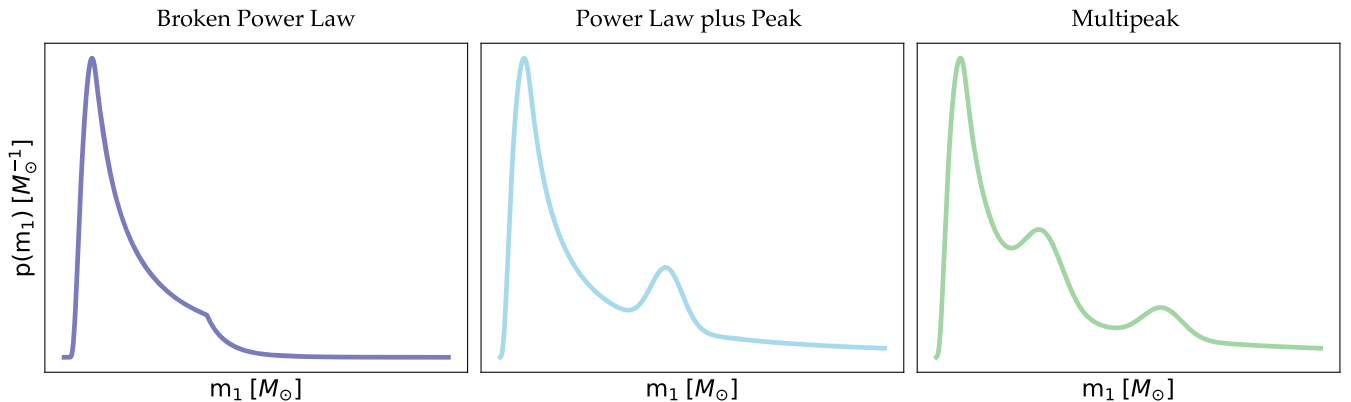


FIG. 1. Simple representation of the three phenomenological mass models, the broken power law (left), the power law plus peak (middle), and the multipeak (right).

edge of the pair-instability supernovae [30] gap. The second power law, defined after the break, could explain the existence of a second population of BBHs inside the pair-instability supernovae gap due to dynamically formed BBHs.

The PLP model [31] is based on a power law with the same low mass smoothing as the BPL model and incorporates a Gaussian peak introduced to account for a potential accumulation of BBH just before the pair-instability supernovae gap [31]. The PLP mass model is more flexible than the BPL model: the position and the width of the peak can vary and allow the model to potentially catch high mass events, without changing the lower part of the spectrum. The fraction of events in the Gaussian peak is given by the parameter  $\lambda_{\text{peak}}$  (see Appendix A 1).

The MLTP mass model is built similar to the PLP model, but with a second Gaussian peak available at higher masses. The motivations for the second Gaussian component are that BBH systems could arise from second-generation black holes produced by hierarchical mergers. The MLTP model is more flexible and can reduce into a PLP model, if the secondary component is not supported by the Bayesian inference.

In previous studies [7,8,26], a significant correlation between the Hubble constant and possible local overdensities in the BBH mass spectrum was found. In particular, the position of the Gaussian peak for the PLP model ( $\mu_g$ ) showed a strong correlation with  $H_0$ .

### III. APPLICATION TO ANALYTICAL BBH POPULATIONS

Here we consider diverse populations of BBHs generated from parametric models to study how the PLP, BPL, and MLTP models infer  $H_0$  for several scenarios. In Sec. III A, we use the same mass models (BPL, PLP, and MLTP) for the generation of the BBHs catalog for the inference. In Sec. III B, we still generate BBHs from one of our three models, but we perform the inference using the other two. In Sec. III C we simulate GW events with an extended version of the PLP model that includes a redshift evolution of the Gaussian peak in order to test the bias on  $H_0$  obtained by the redshift-independent models. We will quantify the  $H_0$  bias according to what confidence interval (C.I.) the true values of  $H_0$  are found from the hierarchical posterior. In principle, biases should be quantified using parameter-parameter (PP) plots that indicate what is the fraction of time that the true value of  $H_0$  is found in a given C.I. for the posterior. As an example, if no bias is present, the true value of  $H_0$  would be found in 50% of the C.I. for 50% of the cases and so on. However, here we do not calculate PP plots, as they are computationally demanding. We argue for the presence of possible biases when the true value of  $H_0$  is excluded at  $2\sigma$  or more C.I. The motivation to choose this criterion is that the probability of obtaining such discrepancy due to statistical fluctuations, and not due to biases, is only 5%.

In the rest of the paper, we will use the posterior predictive checks (PPCs) to assess how well the PLP, BPL, and MLTP models can fit the mass spectrum. The astrophysical posterior predictive distribution, given the observed GW events, for redshift and source masses is defined as

$$p_{\text{pop}}(z, m_1^s, m_2^s | \{x\}) = \int d\Lambda p_{\text{pop}}(z, m_1^s, m_2^s | \Lambda) p(\Lambda | \{x\}). \quad (11)$$

The posterior predictive distribution for the luminosity distance and detector frame masses is defined as

$$p_{\text{obs}}(d_L, m_1^d, m_2^d | \{x\}) = \int d\Lambda p_{\text{pop}}(d_L, m_1^d, m_2^d | \Lambda) \times p(\Lambda | \{x\}) P_{\text{det}}(d_L, m_1^d, m_2^d). \quad (12)$$

We will discuss how the mapping between source and detector frame, and the mismatch of the BBH mass spectrum, will translate to an  $H_0$  bias.

#### A. Using the same mass models for simulation and inference

We simulate three test populations of BBH mergers, with the three source mass distributions and the CBC merger rate evolution described before. The parameters of the population models used for the simulation of the catalogs are summarized in Appendix B, in Tables V–VIII. For the three populations, we generate 2000 detected GW events, which will be used in the spectral siren analysis. We choose to work with 2000 GW events since for this number of detections it is typically possible to have a data-informed posterior bounded in the prior range  $H_0 \in [20, 140] \text{ km s}^{-1} \text{ Mpc}^{-1}$ . We jointly estimate the cosmological parameter  $H_0$ , as well as all the mass and the CBC merger rate parameters using hierarchical Bayesian inference.

Figure 2 presents the marginal posterior distributions for  $H_0$  obtained for each source mass model. For all the simulations, the true value of the Hubble constant is estimated within the 90% C.I. We verified that the true values of all the other population parameters describing the CBC merger rate and mass spectrum are contained in the 90% C.I. of their marginal posterior. The inferred  $H_0$  is more precise for the PLP and the MLTP mass models rather than the BPL model. The more precise constraint on  $H_0$  is given by the fact that the PLP and MLTP models contain sharper features in the source mass spectrum. Similar results have been obtained using  $\sim 20$  independent population realizations for each of the three  $H_0$  posteriors shown in Fig. 2.

Figure 3 shows the PPCs for the BPL, PLP, and MLTP mass models. The figure demonstrates that all the structures in the mass and rate spectrum are well reconstructed by the analysis. In particular, the inference can correctly reconstruct all the features present in the mass spectra. In conclusion, when using the correct mass model for

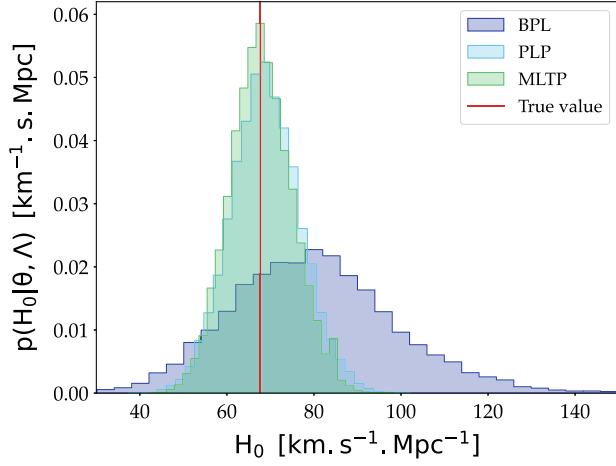


FIG. 2. Marginal posterior of the Hubble constant derived from the spectral siren inference, with the PLP, BPL, and MLTP mass models and a catalog of 2000 detected GW events. The vertical red line is the injected value in the simulation, fixed to the Planck15 value of  $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

inference, both the correct population and cosmological expansion parameters are recovered. Additionally, this test demonstrates the importance of having features in the BBH mass spectrum to improve the precision for cosmological expansion parameters.

### B. Using different mass models for simulation and inference

We generate three populations of 2000 detected CBCs from the BPL, PLP, and MLTP mass models and then we perform the estimation of populations and cosmological parameters with the other two mass models.

Figure 4 shows the marginal posterior distributions of the Hubble constant inferred. We find that, when the detections are simulated with the BPL model and the inference performed with PLP and MLTP models, the true value of  $H_0$  is contained in the 68% C.I. despite the wrong mass model being used for the inference. In contrast, when simulating with the PLP and MLTP models and inferred

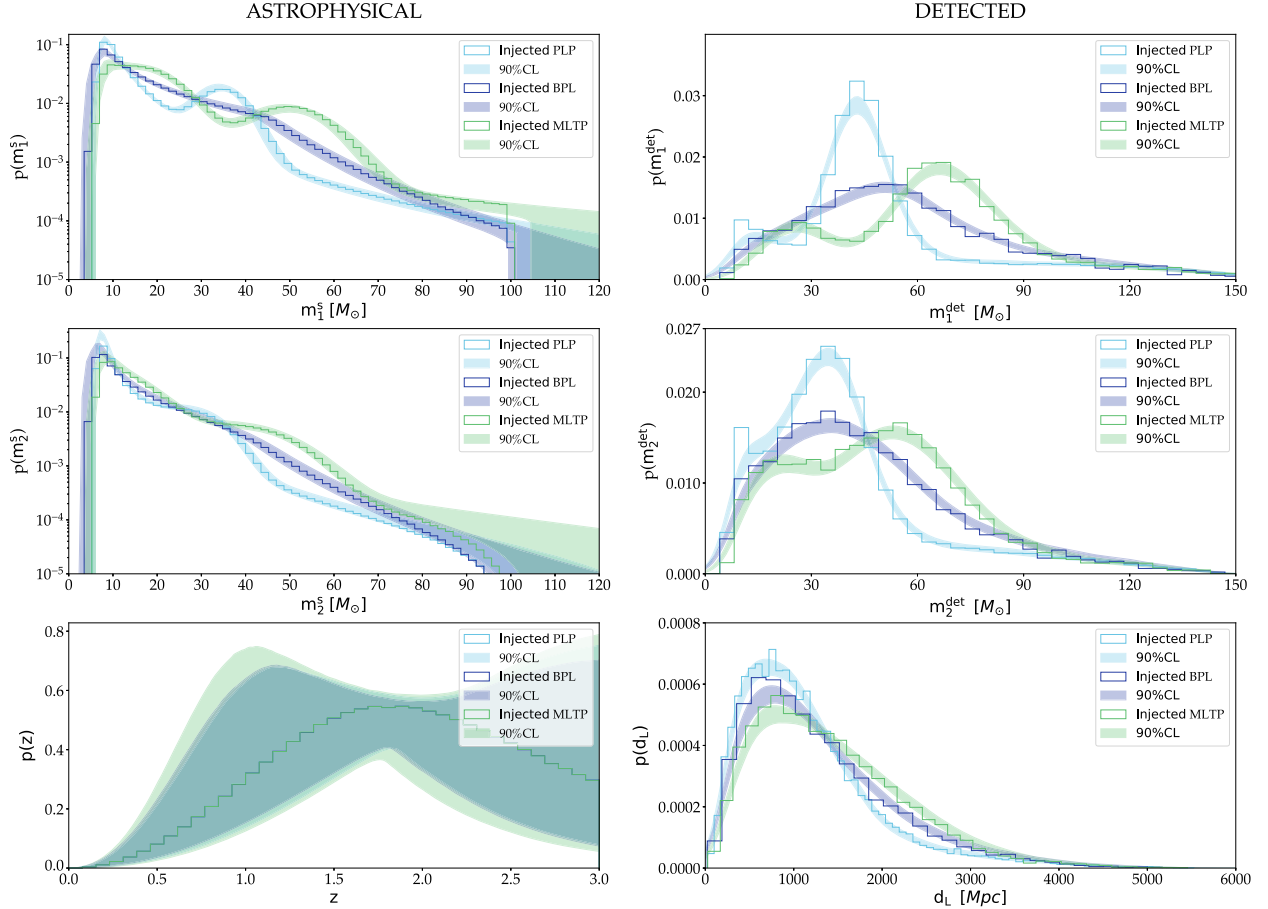


FIG. 3. Posterior predictive checks for the three catalogs of BBHs simulated and inferred with the BPL model (in blue), the PLP model (in light blue), and the MLTP mass model (in green) with 2000 GW events. The left column correspond to the source frame masses and redshift, from top to bottom  $m_1^s$ ,  $m_2^s$ , and  $z$ . The injected populations are under the form of a histogram, and the 90% CL reconstructed spectrum is shown as colored contours. The right column displays the detector frame masses and luminosity distances, with selection effects taken into account, from top to bottom  $m_1^{\text{det}}$ ,  $m_2^{\text{det}}$ ,  $d_L$ .

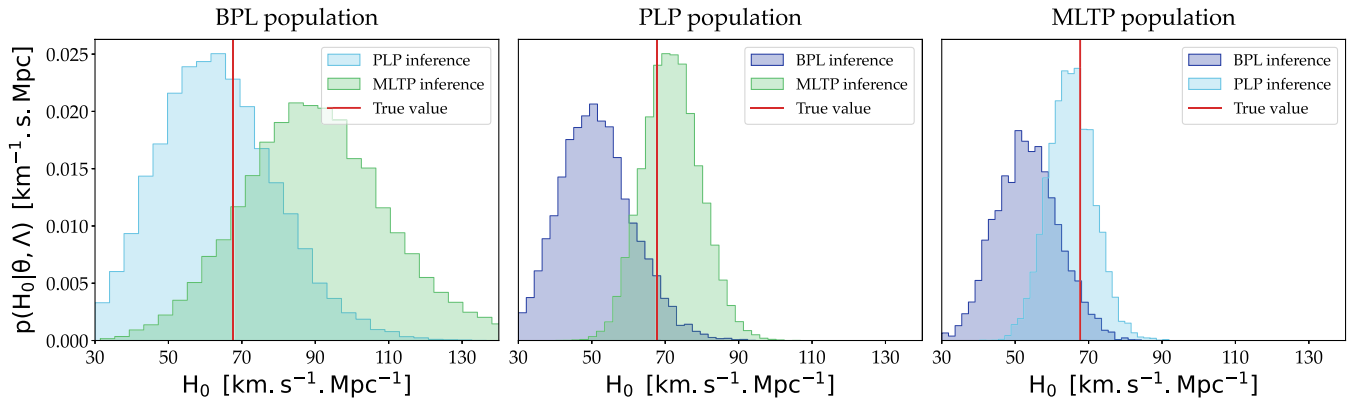


FIG. 4. Left: the marginal distribution of the Hubble constant inferred with the PLP and the MLTP mass models while generating the BBH catalog with the BPL model. Middle: produced simulating the BBH events with the PLP model and inferred with the BPL and the MLTP models. Right:  $H_0$  results obtained while generating the GW events with the MLTP model and inferred with the BPL and PLP mass models. On each of the plots, the true value of the Hubble constant in each population is represented by the vertical red line.

with the BPL model, we find the true value of  $H_0$  is excluded at a 98.7% C.I. level. This bias is linked to the difficulties of the BPL model in catching all the sharp mass features contained in the PLP and MLTP models mass spectra. The mismatching of the mass features result in a systematically mismatched redshift for the sources and then on a bias on  $H_0$ . We have repeated the study cases presented in this section with  $\sim 20$  independent population realizations, finding that the result in Fig. 4 is not a mere statistical fluctuation.

Similar to Sec. III A, we look at the PPC to understand the cosmological results. Figure 5 shows the PPCs for the case in which the simulated population was generated from a MLTP model and the Bayesian inference performed with the BPL and PLP models. From the source frame mass spectrum of  $m_1$  and  $m_2$ , we see that the PLP model inference can reconstruct a peak located around  $50M_\odot$ , while the reconstructed mass spectrum from the BPL model misses the high mass feature by underestimating it. The resulting bias in  $H_0$  can be understood as, for lower values of  $H_0$ , GW events are placed at lower redshifts, hence with higher source frame masses. By lowering the  $H_0$  value, the BPL model is trying to fit the overdensity of high mass events produced by the MLTP model. Concerning the simulations with the most complex of the models, the MLTP model, we find that both the PLP and BPL models are not able to correctly reconstruct the lowest part of the mass spectrum. However, the mismatch is not sufficient to introduce an  $H_0$  bias for the reconstruction with the PLP model. The CBC merger rate as a function of redshift, in all cases, is correctly reconstructed. In conclusion, the sharper mass features in the BBH mass spectrum can source a significant bias on  $H_0$ .

### C. Mass models response to a redshift evolution of the mass spectrum

In this section, we perform a test to examine the robustness of redshift-independent mass models to a

possible evolution of the mass spectrum in redshift. In doing so, we will assume a linear dependence between the features of the mass spectrum and the redshift. This simple model has already been explored in [6] for the evolution of the low and high edges of the mass spectrum, finding a resulting mild  $H_0$  bias when the low edge of the mass spectrum is not modeled correctly. In this paper, we modify the PLP mass model to include a peak position linearly evolving with redshift. The redshift evolution of the mass spectrum for the simulated BBH population is incorporated in the parameter that governs the position of the Gaussian component, while the edges are kept fixed in redshift. This type of evolution could be the result of an interplay of the pair-instability supernova (PISN) mass scale and time delays between binary formation and merger [14,32] that highly suppress the mass spectrum above the PISN scale (the Gaussian peak). The impact of such models has been explored in [33] on Gravitational-Wave Transient Catalog (GWTC) 3, finding no statistically significant preference and a consistent  $H_0$  posterior against the nonevolving PLP model.

Here, we do not suppress the mass spectrum above the PLP model peak, but just assume a linear dependency,

$$\mu_g(z) = \mu_g^0 + z(\mu_g^1 - \mu_g^0), \quad (13)$$

where  $\mu_g^0 = \mu_g(z=0)$  and  $\mu_g^1 = \mu_g(z=1)$  are the positions of the Gaussian peak at  $z=0$  and  $z=1$ . We fix  $\mu_g^0 = 30M_\odot$  and we considered seven values for  $\mu_g^1$  from  $25M_\odot$  to  $35M_\odot$ .

For each simulation, we run a full hierarchical Bayesian inference using the redshift-independent PLP model to reconstruct  $H_0$  as well as the other population parameters from 2000 GW detections. We report the  $H_0$  estimation with 68.3% C.I. for these analyses in Fig. 6, showing the variations of the inferred  $H_0$  value as a function of the position of  $\mu_g^1$ . For instance, if the Gaussian peak shifts of about five solar masses from redshift 0 to redshift 1,  $H_0$  is

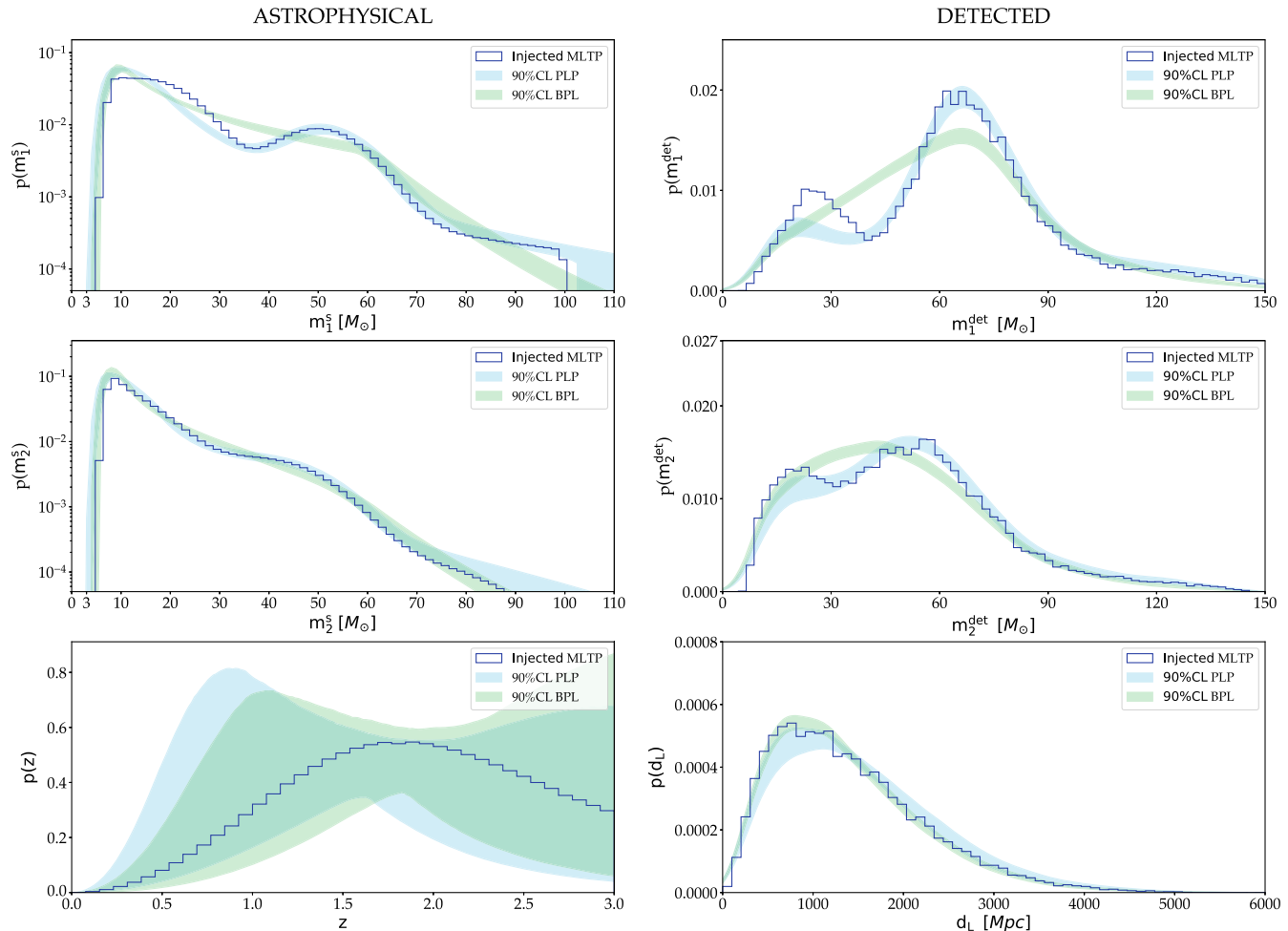


FIG. 5. Posterior predictive checks for the wrong model analysis. The blue histograms are the injected population of BBHs, generated with a MLTP mass model, with the same set of parameters as depicted in Sec. III A. The two colored contours are the 90% CL reconstructed populations, inferred, respectively, with the PLP model in light blue and the BPL mass model in green. This posterior predictive check corresponds to the result presented in Fig. 4 (right).

found outside the 99.7% C.I. Figure 6 shows that the redshift evolution of the structures in the mass spectrum can bias the inferred value of the Hubble constant. Even when considering a mild evolution of the mass spectrum in redshift, the bias on the inferred  $H_0$  is significant. We also notice that the systematic bias obtained on  $H_0$  is linearly proportional to the magnitude of the redshift evolution of the mass spectrum feature. As we will describe in the following paragraph, the bias is introduced by the fact that the source frame mass spectrum is not reconstructed correctly. Comparable results have been achieved utilizing ten independent simulation realizations for each data point, showing the consistency of the result presented in Fig. 6.

Figure 7 summarizes the PPC for this test case. The right column shows the detected mass distributions for three of the seven simulations. In the detector frame, the three simulations display similar distributions for the masses and luminosity distance of detected GW events. The distributions are the ones that are fit by the nonevolving

PLP models. However, in the source frame, the three simulations have a significantly different mass spectrum (due to redshift evolution). The PLP model is not able to catch this evolution and produces a wrong reconstruction of the mass spectrum. In particular, the nonevolving PLP model always reconstructs a peak located at  $30M_\odot$ . This is the consequence of the fact that about 80% of the GW detections are located at low redshift ( $z < 0.3$ ), where, even for redshift evolving models, the peak of the Gaussian component is located around  $\mu_g^0 = 30M_\odot$ .

The introduction of an  $H_0$  bias is due to the misplacing of the Gaussian component for events at higher redshifts. When  $\mu_g^1 > \mu_g^0$  and the PLP model is not able to fit it, we obtain an  $H_0$  biased to higher values. The motivation for this bias is that, for high  $H_0$  values, GW events are placed at higher redshifts and hence at lower source masses. By placing the events at lower source masses, the PLP model is able to accommodate them in the peak that fits around  $30M_\odot$  in the source frame. When  $\mu_g^1 < \mu_g^0$ , we observe a bias

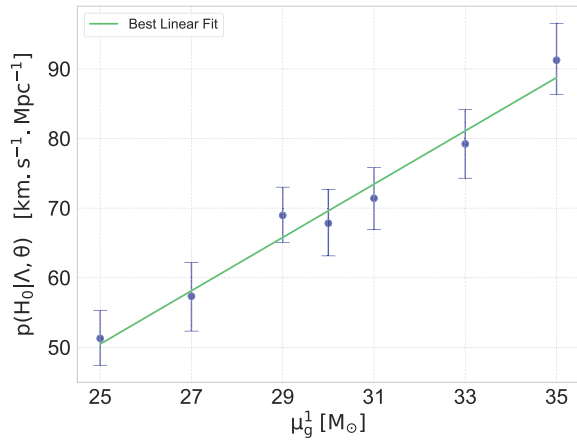


FIG. 6. Evolution of the  $H_0$  value inferred with respect to the final position of the PLP model peak at  $z = 1$ . The blue points correspond to the median of the inferred posterior of the Hubble constant, for each of the analysis. The point of reference (no redshift evolution) is for  $\mu_g = 30$  at  $z = 1$ . The error bars are the  $1\sigma$  intervals of the posteriors and the green line is the best linear fit passing by all points.

to lower values of  $H_0$ . Similar to the previous case, this is due to the fact that GW events are placed at lower redshifts; therefore, they will have a higher source mass that can be placed in the peak at  $30M_\odot$ . This simple test case shows that the evolution in redshift of a feature in the mass spectrum could potentially bias the GW-based  $H_0$  estimation.

#### D. Summary of the tests

To summarize, the  $H_0$  bias is closely linked to the ability of the mass model to catch and reconstruct the features of the CBC mass spectrum. A bias on the  $H_0$  value can be introduced from several factors.

We have shown that, in the case where the true underlying population of CBCs presents sharp features and the reconstruction model is not able to include them, a significant bias on  $H_0$  can be obtained. We have also shown that, if any of the mass features of the spectrum evolve in redshift, an  $H_0$  bias can be introduced even if the evolution is of the order of a few solar masses.

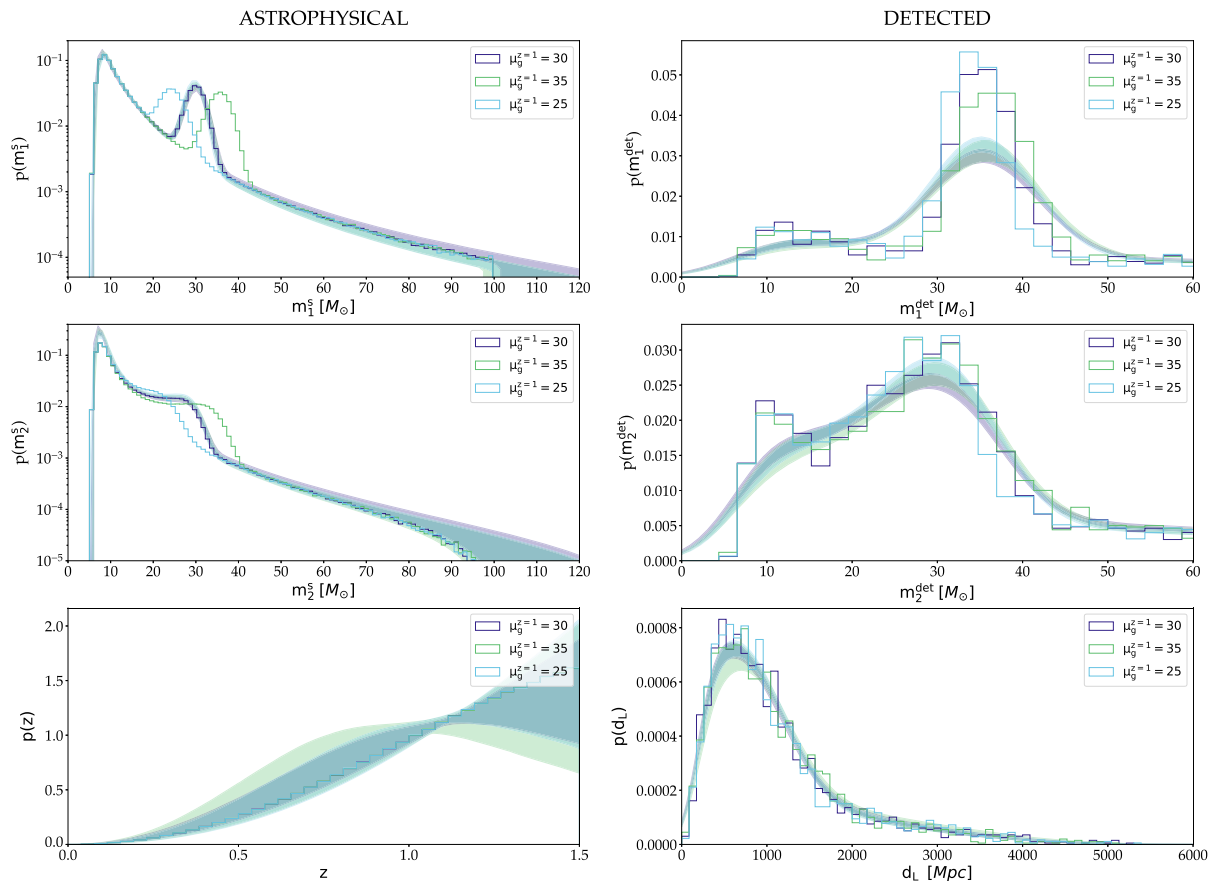


FIG. 7. Posterior predictive check for the simulated populations, incorporating an additional linear evolution of the Gaussian component  $\mu_g(z)$  with respect to the redshift, between  $z = 0$  and  $z = 1$ . The colored contours are the 90% CL estimated spectrum obtained with the spectral siren analysis. The colored histograms display the injected populations in each run, denoted by the position of their Gaussian peak at  $z = 1$ . Only three out of the seven populations are shown: the  $\mu_g = 30$  analysis in purple (reference run with no redshift evolution), the  $\mu_g = 35M_\odot$  in green, and the  $\mu_g = 25$  in blue.



#### IV. APPLICATION TO A COMPLEX BBH POPULATION

In this section, we study how the BPL, PLP, and MLTP models respond to a population of BBHs generated from synthetic astrophysical simulations. We use a catalog, referred to as “A03,” of BBH mergers formed by various formation channels from [34,35]. We then present the results of the hierarchical Bayesian analysis obtained with the BPL, PLP, and MLTP mass models in Sec. IV B. Finally, in Sec. IV C, we investigate the possible sources of bias for the Hubble constant.

##### A. General description of the A03 catalog

For this study, we employ the A03 catalog from [35]. This model consists of four different channels: isolated binary evolution and dynamical assembly of BBHs in young, globular, and nuclear star clusters. They are mixed based on their corresponding redshift-dependent rates as described in [35]. The dynamical formation channels assume an initial black hole mass function obtained with the MOBSE population synthesis code [36]. In model A03, we assume the delayed model of [37] for core-collapse supernovae, the pair-instability supernova formalism of [38], and we take into account the dependence of the black hole (BH) mass function on the metallicity of the progenitor stars, as discussed in [36]. The BHs then pair up dynamically with other BHs and harden because of dynamical interactions in their host star clusters [34]; we also allow for hierarchical mergers of BBHs, as described in [35]. The isolated binary evolution channel consists of systems evolved with the MOBSE binary population synthesis code accounting for binary star evolution [36]. For these runs, we include a treatment for mass transfer, common envelope (with efficiency parameter  $\alpha = 1$ ), tidal evolution, natal kicks, and gravitational wave decay as described in [36]. A further discussion of the astrophysical processes included in model A03 and the corresponding uncertainties are beyond the scope of this study. Here, we explore the possibility that nontrivial structures of the mass distribution for BBHs can introduce a bias on the estimation of  $H_0$ .

Figure 8 shows the distribution of the primary mass  $m_1^s$  for different redshift bins. The BBH mass spectrum presents several features: Moving from lower to higher masses, a feature at  $10M_\odot$ , a “valley” between  $13$  and  $15M_\odot$ , and a second large peak at  $16M_\odot$ . In the high mass region of the spectrum, a local overdensity of BBHs is visible around  $25M_\odot$ , plus several less significant peaks up to  $90M_\odot$ . These structures are also visible on the  $m_2^s$  distribution. Another property of the A03 catalog is the evolution in redshift of the primary and secondary source frame masses. From Fig. 8, we notice that the structures in the mass spectrum are evolving as the redshift increases. The feature at  $20M_\odot$  is nonexistent

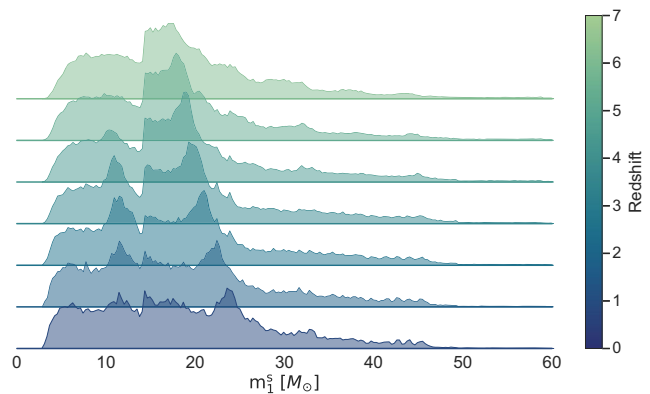


FIG. 8. Kernel density estimation of the primary masses  $m_1^s$  in source frame from the A03 BBH catalog. Each plot corresponds to a slice in redshift, going from  $z = 0$  up to  $z = 7$ . This figure highlights the overall mass spectrum evolution with respect to the redshift, especially the appearance of peaks and their shift.

for mergers below  $z = 2$  and emerges while shifting to the lower masses in the higher redshift bins.

Figure 9 shows the BBH mass spectrum in a redshift range of  $0 < z < 1.75$ . This range corresponds to the typical range in which our simulation is able to detect BBH mergers. The evolution of the BBH mass spectrum at low redshifts is fainter than the one at high redshift; nevertheless, peaks around  $11M_\odot$  appear while shifting to higher masses and the structure around  $25M_\odot$  is drifting to lower masses. Since real GW events at high redshift ( $z > 1$ ) are hardly detectable in the O4-like network, this effect is thought to be subdominant for population and cosmological inferences [39]. If these structures are evolving with respect to the redshift, as discussed in [40–44], the mass models could mismatch the features and produce a biased value of  $H_0$  as shown in Sec. II C.

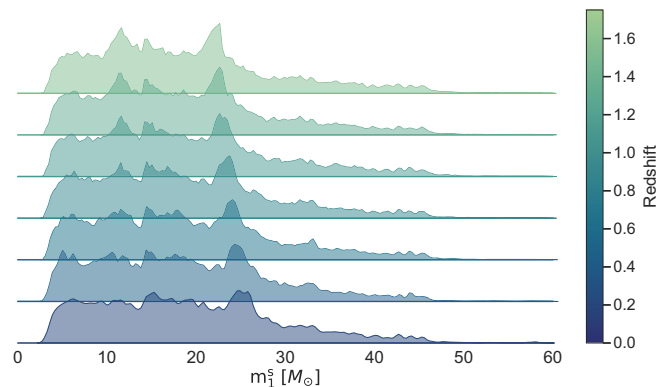


FIG. 9. Kernel density estimation of the primary masses  $m_1^s$  in source frame from the A03 BBH catalog. Each plot corresponds to a slice in redshift, going from  $z = 0$  up to  $z = 1.75$ . This figure highlights the overall mass spectrum evolution with respect to the redshift, especially the appearance of peaks and their shift.

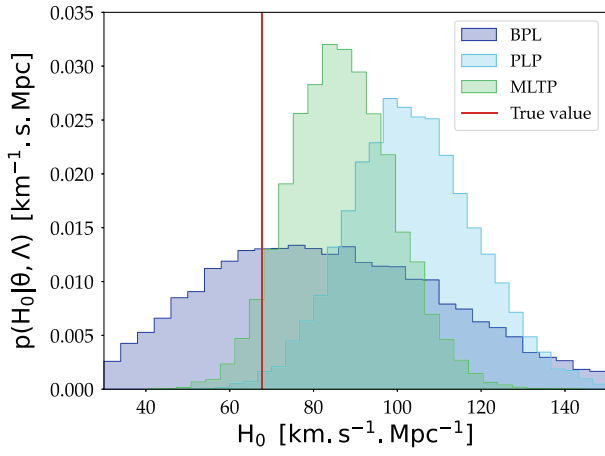


FIG. 10. Marginal posterior distributions of the Hubble constant  $H_0$  obtained with 2000 GW events, using the PLP model in light blue, the BPL model in blue, and the MLTP mass model in green. The vertical red line represents the true value of the Hubble constant intrinsic to the A03 catalog and equal to  $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## B. Vanilla analysis of the A03 catalog

Using the A03 BBH catalog, and with the framework described in Sec. II A, we generate GW catalogs made of 2000 detected GW events. For each catalog, we perform a full hierarchical Bayesian inference, reconstructing posteriors on  $H_0$ , as well as the other population parameters of the redshift rate and the three redshift-independent BBH mass models. The prior ranges used for the Bayesian inference are reported in Tables I–IV.

Figure 10 shows the marginal posterior distributions of the Hubble constant estimated. When using the PLP and MLTP mass models, we find an  $H_0$  bias toward higher values. In particular, we estimate for the PLP model  $H_0 = 103^{+11}_{-10}$  and  $H_0 = 87^{+9}_{-8} \text{ km s}^{-1} \text{ Mpc}^{-1}$  at 68.3% C.I. The true value of  $H_0$  is excluded from the 99.7% C.I. for the PLP model and the 95% C.I. for the MLTP model. From Fig. 10, it is possible to note that the BPL mass model obtains an uninformative  $H_0$  posterior. We obtain a value of  $H_0 = 82^{+20}_{-19} \text{ km s}^{-1} \text{ Mpc}^{-1}$ , which includes the injected  $H_0$

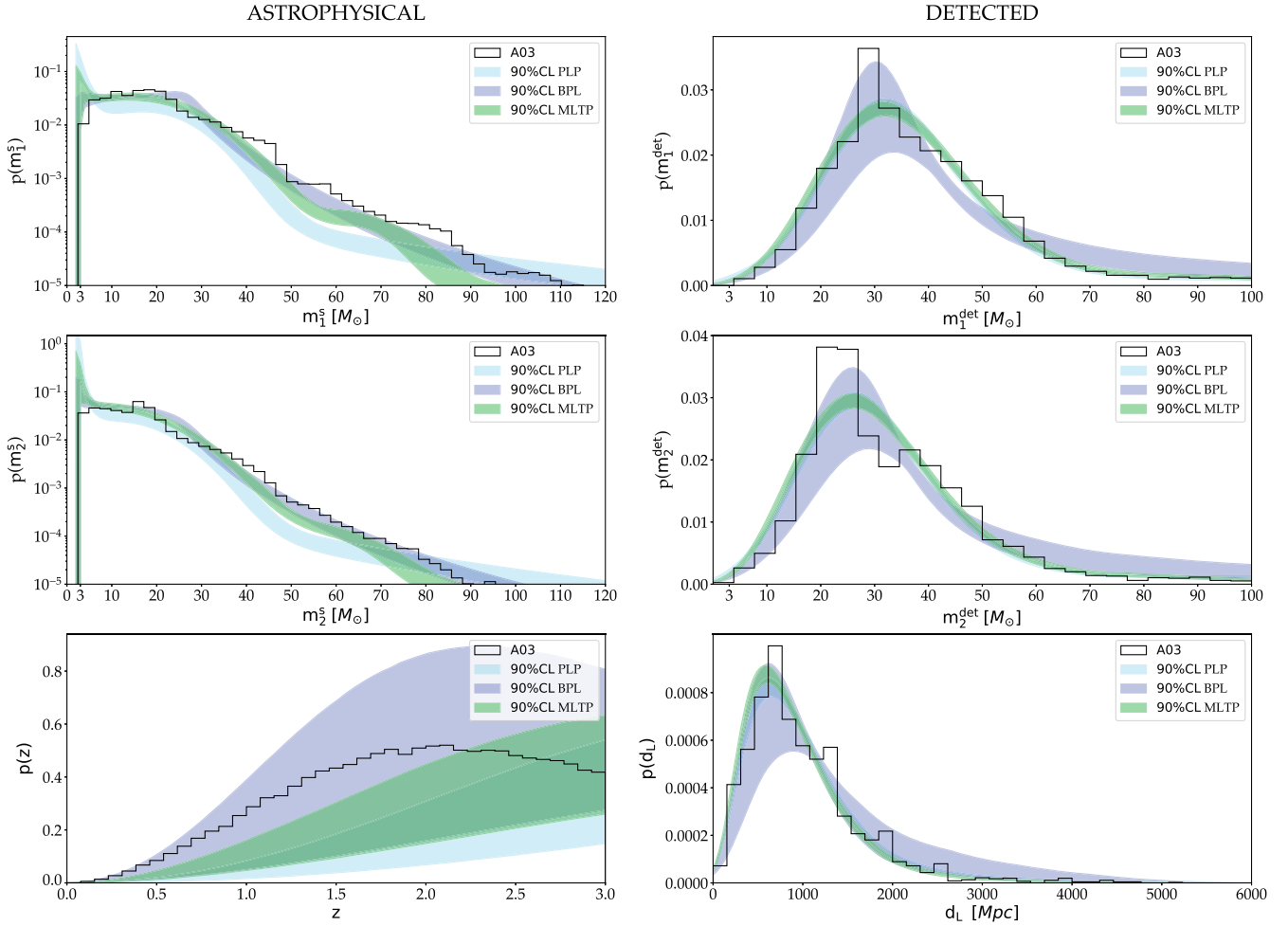


FIG. 11. Posterior predictive check for the spectral siren analysis with 2000 GW observations, generated from the A03 BBH catalog. The plain black histograms are the injected population, in the source frame (left column) and in the detector frame (right column). The three colored contours depict the 90% CL for the three inferences (BPL, PLP, and MLTP).

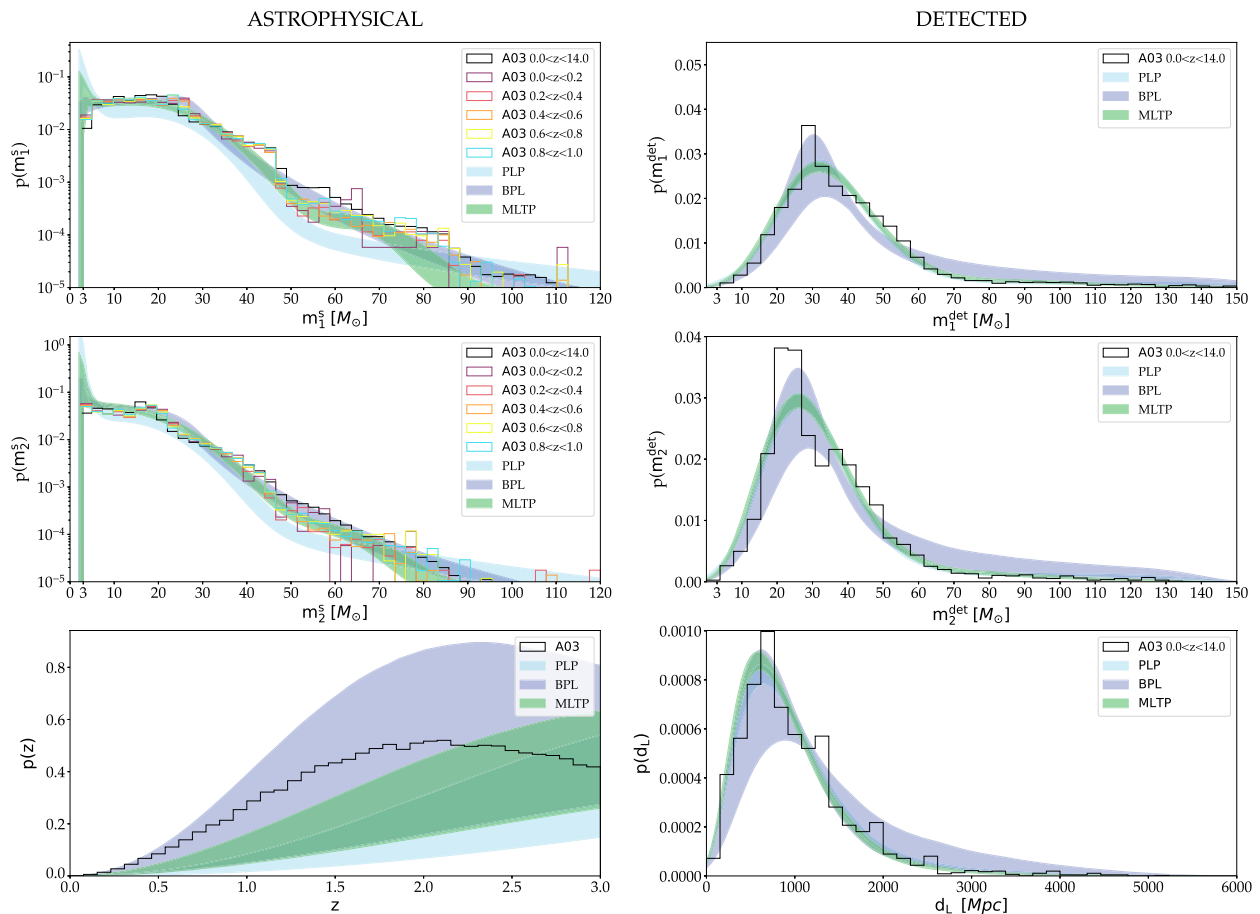


FIG. 12. Posterior predictive check of the spectral siren analysis of 2000 GW observations generated from the A03 BBH catalog. The plain black histograms are the injected population, in the source frame (left column) and in the detector frame (right column). Each colored histogram corresponds to a bin in redshift from  $z = 0$  to  $z = 2$ . The three colored contours depict the 90% CL computed with the BPL, PLP, and MLTP mass models.

within the 68% C.I. As we will argue later, the BPL model recovers a less informative posterior on  $H_0$ , as it does not contain any strong mass feature. We note that, even if the BPL model does not display an  $H_0$  bias for 2000 events, it might still display it as more and more GW detections are collected. The bias is not introduced by any of the other population parameters “railing” over the prior ranges. In fact, we verified that all the parameters are well constrained in their prior ranges. Equivalent findings have been produced through  $\sim 15$  distinct simulation realizations for each of the three posterior distributions displayed in Fig. 10.

The PPC plot in Fig. 11 gives insights into the origin of the  $H_0$  bias. While the reconstructed distributions of detector frame masses and luminosity distance do not show any particular deviation between the injected population and the recovered one, the reconstructed source population display larger differences. In the source frame, the CBC merger rate is correctly reconstructed by the BPL model, whereas the PLP and MLTP models significantly underestimate the presence of BBHs between  $40M_\odot$  and  $80M_\odot$ . Therefore, from observed data, the PLP and MLTP models

predict a significantly larger number of BBHs at lower source masses. This is achieved by pushing the value of the Hubble constant at higher values so that events are found at higher redshifts and have on average a lower value of source masses. Another indicator in support of the finding that the PLP and MLTP models place events at higher redshifts is the reconstruction of the CBC merger rate in redshift. In fact, from the bottom left plot of Fig. 11, we can see that the PLP and MLTP models reconstruct a biased merger rate in redshift that prefers higher values of redshifts for the BBH mergers. Instead, the BPL model inference manages to fit the overall trend of the A03 source frame population. The BPL model reconstructs better the source frame masses between  $40M_\odot$  and  $80M_\odot$ .

Even though the PLP and MLTP models have enough degrees of freedom to approximate a smooth distribution of masses (as the ones present in the A03 catalog), they still fail to reconstruct the mass spectrum. The motivation for which the PLP and MLTP models are not able to correctly reconstruct the mass spectrum of the A03 catalog

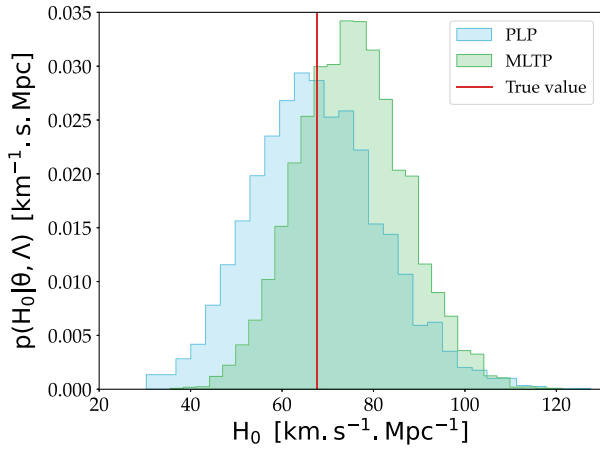


FIG. 13. Marginal posterior distributions of the Hubble constant  $H_0$  obtained with 2000 GW events, using the PLP and the MLTP mass models. The vertical purple line represents the true value of the Hubble constant intrinsic to the A03 catalog and equal to  $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . This distribution is obtained using the A03 catalog, where the redshift evolution of the mass spectrum has been removed.

and  $H_0$  is not trivial. A first hypothesis for this result was that the PLP and MLTP models were catching local peaks in the mass spectrum (at a fixed redshift) that were evolving in redshift. In fact, from the posterior distributions, we observed that the PLP and MLTP models were finding local peaks at low masses ( $\approx 17M_\odot$ ) with a large standard deviation and a higher peak at  $\approx 30M_\odot$ . To test this hypothesis, we performed a PPC in Fig. 12, dividing the true population in redshift bins. Unfortunately, the PPC does not clearly show a discrepancy between the population reconstruction and any of the mass spectra for the redshift bins. As this test is inconclusive on the origin of the  $H_0$  bias, we performed additional studies that we discuss below.

### C. Investigating the sources of the $H_0$ bias: Blinding the mass-redshift relation

In order to test further the possibility that the redshift evolution of the mass spectrum is introducing an  $H_0$  bias, we performed a simulation where we blind the A03's mass

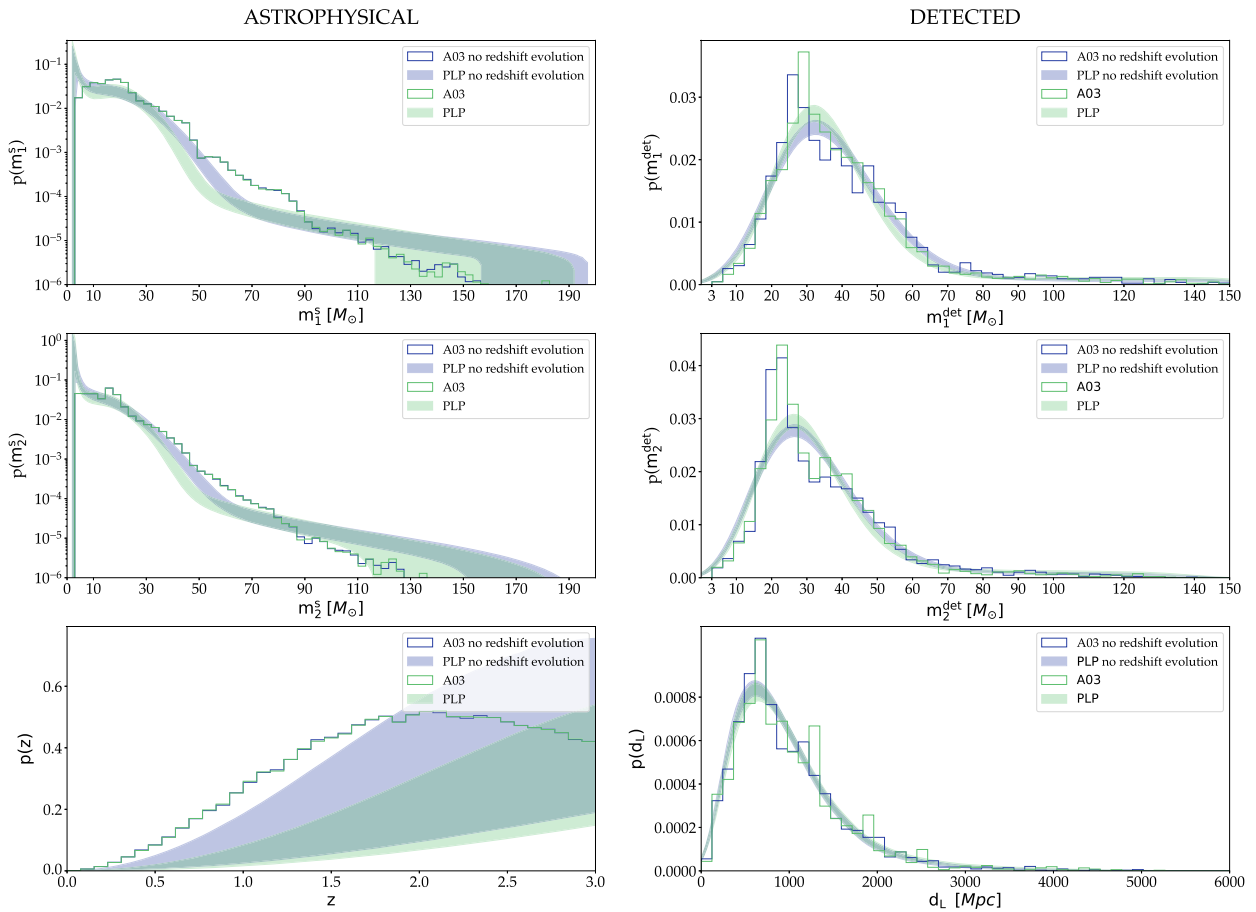


FIG. 14. Comparison of posterior predictive checks for the spectral siren analysis of the A03, with and without redshift evolution of the mass spectrum, using 2000 GW events. The blue contour and histograms are the vanilla population and are the same results shown in Fig. 11. The purple histograms are the new A03 population with no redshift evolution of the mass spectrum and the purple contours are the 90% CL inferred spectrum with the population, denoted “A03 no redshift evolution.”

spectra to the redshift evolution. To remove the redshift dependency of the mass spectrum, we randomly shuffle the pairs of BBH merger redshifts and masses. This procedure artificially removes the redshift dependence of the mass spectrum while conserving its nontrivial shape. For the Bayesian inference, we only use the PLP and MLTP mass models since they were displaying a non-negligible bias of  $H_0$ .

The result of the joint inference with the PLP mass model is presented in Fig. 13. For the PLP model, we obtain  $H_0 = 67.0_{-8.9}^{+9.7} \text{ km s}^{-1} \text{ Mpc}^{-1}$  and for the MLTP model we obtain  $H_0 = 75.13_{-7.9}^{+7.8} \text{ km s}^{-1} \text{ Mpc}^{-1}$ , the true value of  $H_0$  is contained within the 68.3% C.I. of the posterior. Figure 14 shows the PPC plot for this test. The figure overlaps the reconstruction for the original A03 catalog and the one after the redshift-blinding procedure. From the detector frame point of view, we see that the two reconstructed distributions do not significantly differ. However, from the source frame point of view, the PLP model, for the redshift-blinded A03 catalog, is able to reconstruct better the BBH masses present in the range  $30M_\odot$ – $60M_\odot$ . Although the reconstruction of the mass spectrum above  $60M_\odot$  is still poor, the PLP model for the nonredshift evolving A03 is able to include the true value of  $H_0$ . As a consequence, also the reconstruction of the BBH merger rate as a function of redshift slightly improves, resulting in an underestimation of about 20%–30% of the true CBC merger rate. We have also found similar results for the MLTP model. We verified by running the simulations with 20 different population realizations that the value of  $H_0$  obtainable by the PLP and MLTP models from the redshift-blinded A03 is always unbiased.

It is important to notice that, even after blinding the redshift evolution of the A03 mass spectrum, the PLP and MLTP models are not yet able to completely reconstruct the source frame distributions. As such, we still expect a systematic bias on  $H_0$  to be present and hidden by the large statistical uncertainties that can be seen in Fig. 13. Identifying the source of these systematic biases is not trivial, but from the test that we performed, we are sure that the redshift evolution of the mass spectrum is expected to be the dominant source of the  $H_0$  bias.

## V. CONCLUSIONS

In this work we have extensively studied the interplay between systematic biases on  $H_0$  and the reconstruction of the BBH mass spectrum for Spectral Siren Cosmology. We performed several simulations using BBH merger catalogs generated with three of the current mass models used in literature and we also considered a catalog of synthetically simulated BBHs.

In Sec. III, we showed that simple phenomenological models that do not contain sharp features could introduce a bias on  $H_0$  if the true population of BBHs includes a local overdensity of sources. Moreover, we showed that the redshift evolution of source frame mass features can induce an  $H_0$  bias if not accounted for in the mass models. In particular, the Hubble constant would act as a free parameter and shift to higher values if the true source mass spectrum is underestimated, hence placing more events at higher redshift and, equivalently, decreasing the global source mass spectrum. This result is, respectively, true if the source mass spectrum is overestimated, creating a shift of  $H_0$  to lower values.

In Sec. IV, using the A03 BBH catalog, we showed that, even if the detector frame is well reconstructed by the inference, we might still get a biased value of the Hubble constant. Deviations up to  $2\sigma$  and  $3\sigma$  from the injected value were observed for the MLTP and PLP model inferences, respectively. We observed that, while each model yielded to different populations in the source frame, they produce remarkably similar populations in the detector frame. This discrepancy in the source frame mapping appeared to be related to the response of the models to the GW events, as each model attempted to accommodate to the true population in its own unique manner. Consequently, the Hubble constant is used as a compensatory factor, when necessary, to align the inferred populations with the observed data. Finally, we find that the  $H_0$  bias obtained with the PLP and MLTP mass models disappears when the redshift evolution of the source mass spectrum is removed from the population.

In light of these findings, we demonstrate that the most commonly used mass models, namely PLP, BPL, and MLTP models, combined with the spectral siren analysis for GW cosmology, are not suitable choices if the BBH population exhibits even a slight redshift evolution of its mass spectrum. In addition, even if the mass spectrum is not evolving with redshift, the mismatching of sharp structures in the BBH mass distribution can also be responsible for a biased estimation of the Hubble constant. We showed that this kind of assumption can be dangerous for the spectral siren method and, more generally, for methods that make use of nonevolving parametric models of the mass spectrum, since the constraining power on the Hubble constant relies on the ability of the phenomenological mass models to fit the position of structures in the mass spectrum. Given that the true distribution of BBHs in our Universe remains unknown and may differ significantly from the current state-of-the-art models, future GW cosmology studies based on the spectral siren method should incorporate source mass models capable of absorbing effects such as the redshift evolution of the mass spectrum and possessing sufficient degrees of freedom to accommodate unforeseen mass features. Failure to do so may introduce biases in future Hubble constant measurements, especially when dealing with a large number of detected GW events.

## ACKNOWLEDGMENTS

The authors are grateful for computational resources provided by the LIGO Laboratory and supported by the National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. We thank Viola Sordini for the helpful comments and discussions on the manuscript. M. M. acknowledges financial support from the European Research Council for the ERC Consolidator Grant No. DEMOBLACK, under Contract No. 770017, and from the German Excellence Strategy via the Heidelberg Cluster of Excellence (EXC 2181–390900948) STRUCTURES.

## APPENDIX A: MASS AND REDSHIFT POPULATION MODELS

### 1. Source mass models

In this appendix, we present in more detail the three phenomenological source mass models used in this paper for the spectral siren inference. These models are the current mass models used by the LIGO-Virgo-KAGRA scientific collaboration for cosmology results [2–4]. We give the mathematical expressions of priors and probability density functions for the power law plus peak, broken power law, and the multipeak model. Note that they are the same models described in [9]. All three models are made of combinations of truncated power law and truncated Gaussian distributions.

The usual truncated power law distribution is given by

$$\mathcal{P}(x|x_{\min}, x_{\max}, \alpha) = \begin{cases} \frac{1}{N} x^\alpha, & (x_{\min} \leq x \leq x_{\max}), \\ 0, & \text{otherwise,} \end{cases} \quad (\text{A1})$$

where  $N$  is the normalization factor defined as

$$N = \begin{cases} \ln\left(\frac{x_{\max}}{x_{\min}}\right), & (\text{if } \alpha = -1), \\ \frac{1}{\alpha+1} (x_{\max}^{\alpha+1} - x_{\min}^{\alpha+1}), & \text{otherwise} \end{cases} \quad (\text{A2})$$

The truncated Gaussian distribution with mean  $\mu$  and standard deviation  $\sigma$ ,

$$\mathcal{G}_{[a,b]}(x|\mu, \sigma) = \begin{cases} \frac{1}{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], & a \leq x \leq b, \\ 0, & \text{otherwise,} \end{cases} \quad (\text{A3})$$

where  $N$  is also the normalization factor expressed as

$$N = \int_b^a \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (\text{A4})$$

*The broken power law.* This model, introduced in [29], is made of two truncated power laws attached at a breaking point  $b$ ,

$$b = m_{\min} + (m_{\max} - m_{\min})f, \quad (\text{A5})$$

where  $f$  belongs to  $[0, 1]$ . When  $f = 0$ , the breaking point  $b$  is equal to  $m_{\min}$ . The probability density functions for  $m_1^s$  and  $m_2^s$  are then defined as

$$\begin{aligned} \pi(m_1^s|m_{\min}, m_{\max}, \alpha) &= \frac{1}{N} \left[ \mathcal{P}(m_1^s|m_{\min}, b, -\alpha_1) \right. \\ &\quad + \frac{\mathcal{P}(b|m_{\min}, b, -\alpha_1)}{\mathcal{P}(b|b, m_{\max}, -\alpha_2)} \\ &\quad \left. \times \mathcal{P}(m_1^s|m_{\max}, b, -\alpha_2) \right], \end{aligned} \quad (\text{A6})$$

$$\pi(m_2^s|m_{\min}, m_1^s, \beta) = \mathcal{P}(m_2^s|m_{\min}, \beta, m_1^s). \quad (\text{A7})$$

$N$  is the normalization factor defined as

$$N = 1 + \frac{\mathcal{P}(b|m_{\min}, b, -\alpha_1)}{\mathcal{P}(b|b, m_{\max}, -\alpha_2)}. \quad (\text{A8})$$

Table I presents the parameters that control the BPL model distribution, as well as the prior distributions used to perform the joint inference of the cosmology and the BBH population presented in Sec. IV.

*The power law plus peak.* This model, composed of a truncated power law and a Gaussian component, was first introduced in [45]. The probability density functions of the primary masses are given by

$$\begin{aligned} \pi(m_1^s|m_{\min}, m_{\max}, \alpha) &= (1 - \lambda)\mathcal{P}(m_1^s|m_{\min}, m_{\max}, -\alpha) \\ &\quad + \lambda\mathcal{G}(m_1^s|\mu_g, \sigma_g), \quad (0 \leq \lambda \leq 1), \end{aligned} \quad (\text{A9})$$

$$\pi(m_2^s|m_{\min}, m_{\max}, \alpha) = \mathcal{P}(m_2^s|m_1^s, m_{\min}, \beta). \quad (\text{A10})$$

TABLE I. Summary of the prior ranges used with the BPL mass models for the spectral siren analysis.

Parameter	Description	Prior
$\alpha_1$	Power law index number 1 primary mass	$\mathcal{U}(-4, 10)$
$\alpha_2$	Power law index number 2 primary mass	$\mathcal{U}(-4, 10)$
$\beta$	Power law index secondary mass	$\mathcal{U}(-4, 10)$
$m_{\min}$	Minimum value of the source mass [ $M_\odot$ ]	$\mathcal{U}(1M_\odot, 10M_\odot)$
$m_{\max}$	Maximum value of the source mass [ $M_\odot$ ]	$\mathcal{U}(100M_\odot, 200M_\odot)$
$\delta_m$	Smoothing parameter [ $M_\odot$ ]	$\mathcal{U}(0M_\odot, 10M_\odot)$
$b$	Breaking point [ $M_\odot$ ]	$\mathcal{U}(0M_\odot, 1M_\odot)$

TABLE II. Summary of the prior ranges used with the PLP mass models for the spectral siren analysis.

Parameter	Description	Prior
$\alpha$	Spectral index for the power law of the primary mass distribution	$\mathcal{U}(1, 10)$
$\beta$	Spectral index for the power law of the mass ratio distribution	$\mathcal{U}(-4, 10)$
$m_{\min}$	Minimum mass of the primary mass distribution [ $M_{\odot}$ ]	$\mathcal{U}(1M_{\odot}, 12M_{\odot})$
$m_{\max}$	Maximum mass of the primary mass distribution [ $M_{\odot}$ ]	$\mathcal{U}(50M_{\odot}, 200M_{\odot})$
$\lambda_g$	Fraction of the model in the Gaussian component	$\mathcal{U}(0, 1)$
$\mu_g$	Mean of the Gaussian in the primary mass distribution [ $M_{\odot}$ ]	$\mathcal{U}(10M_{\odot}, 40M_{\odot})$
$\sigma_g$	Width of the Gaussian in the primary mass distribution [ $M_{\odot}$ ]	$\mathcal{U}(6M_{\odot}, 17M_{\odot})$
$\delta_m$	Range of mass tapering at the lower end of the mass distribution [ $M_{\odot}$ ]	$\mathcal{U}(0M_{\odot}, 12M_{\odot})$

TABLE III. Summary of the prior ranges used with the MLTP mass models for the spectral siren analysis.

Parameter	Description	Prior
$\alpha$	Power law index primary mass	$\mathcal{U}(1, 10)$
$\beta$	Power law index secondary mass	$\mathcal{U}(-4, 10)$
$m_{\min}$	Minimum value of the source mass [ $M_{\odot}$ ]	$\mathcal{U}(1M_{\odot}, 10M_{\odot})$
$m_{\max}$	Maximum value of the source mass [ $M_{\odot}$ ]	$\mathcal{U}(50M_{\odot}, 200M_{\odot})$
$\delta_m$	Smoothing parameter [ $M_{\odot}$ ]	$\mathcal{U}(0M_{\odot}, 10M_{\odot})$
$\mu_g^{\text{low}}$	Mean of the lower Gaussian peak [ $M_{\odot}$ ]	$\mathcal{U}(11M_{\odot}, 30M_{\odot})$
$\mu_g^{\text{high}}$	Mean of the higher Gaussian peak [ $M_{\odot}$ ]	$\mathcal{U}(40M_{\odot}, 80M_{\odot})$
$\sigma_g^{\text{low}}$	Standard deviation of the higher Gaussian peak [ $M_{\odot}$ ]	$\mathcal{U}(6M_{\odot}, 17M_{\odot})$
$\sigma_g^{\text{high}}$	Standard deviation of the higher Gaussian peak [ $M_{\odot}$ ]	$\mathcal{U}(6M_{\odot}, 17M_{\odot})$
$\lambda_g$	Proportion of events in the peaks	$\mathcal{U}(0, 1)$
$\lambda_g^{\text{low}}$	Proportion of events in the lower peak	$\mathcal{U}(0, 1)$

TABLE IV. Summary of the prior ranges used with the CBC merger rate for the spectral siren analysis.

Parameter	Description	Prior
$\gamma$	Slope of the power law regime for the rate evolution before the point $z_p$	$\mathcal{U}(0, 10)$
$k$	Slope of the power law regime for the rate evolution after the point $z_p$	$\mathcal{U}(0, 6)$
$z_p$	Redshift turning point between the power law regimes with $\gamma$ and $k$	$\mathcal{U}(0, 4)$

Table II presents the parameters that control the PLP model distribution, as well as the prior distributions used to perform the joint inference of the cosmology and the BBH population presented in Sec. IV.

*The multipeak.* This model is an extension of the PLP model; it consists of a truncated power law and two Gaussian components. This parametrization was first applied in [29] and the primary mass probability density functions (PDFs) are given by

$$\begin{aligned} \pi(m_1^s | m_{\min}, m_{\max}, \alpha) = & [(1 - \lambda)\mathcal{P}(m_1^s | m_{\min}, m_{\max}, -\alpha) \\ & + \lambda\lambda_{\text{low}}\mathcal{G}(m_1^s | \mu_g^{\text{low}}, \sigma_g^{\text{low}}) \\ & + \lambda(1 - \lambda_{\text{low}})\mathcal{G}(m_1^s | \mu_g^{\text{high}}, \sigma_g^{\text{high}})], \end{aligned} \quad (\text{A11})$$

$$\pi(m_2^s | m_{\min}, m_{\max}, \beta) = \mathcal{P}(m_2^s | m_1^s, m_{\min}, \beta). \quad (\text{A12})$$

Table III presents the parameters that control the MLTP model distribution, as well as the prior distributions used to perform the joint inference of the cosmology and the BBH population presented in Sec. IV.

All of the three mass models described above are normalized functions, since they are all made of normalized PDFs.

## 2. Merger rate evolution model

In this paper, we use a specific parametrization of the CBC merger rate  $\mathcal{R}(z)$ , also denoted  $p_{\text{pop}}(z|\Lambda)$  in Sec. II as a function of the redshift. This model is constructed after the Madau and Dickinson star formation rate [28]. Using the same notation as in Eq. (8), the CBC merger rate is

expressed as

$$p_{\text{pop}}(z|\Lambda) = [1 + (1 + z_p)^{-\gamma-k}] \frac{(1+z)^\gamma}{1 + \left(\frac{1+z}{1+z_p}\right)^{\gamma+k}} \quad (\text{A13})$$

Table I presents the parameters that control the CBC merger rate distribution, as well as the prior distributions used to perform the joint inference of the cosmology and the BBH population presented in Sec. IV.

## APPENDIX B: INJECTED BBHS POPULATIONS

This appendix presents the values for each of the mass and CBC merger rate parameters, chosen to perform the spectral siren analysis of Sec. III, for each of the following mass models: BPL, PLP, and MLTP.

### 1. Mass parameters simulated

#### a. Broken power law parameters

TABLE V. Summary of the prior ranges used with the BPL mass models for the spectral siren analysis.

Parameter	Description	Injected value
$\alpha_1$	Power law index number 1 primary mass	1.5
$\alpha_2$	Power law index number 2 primary mass	5.5
$\beta$	Power law index secondary mass	1.4
$m_{\text{min}}$	Minimum value of the source mass [ $M_\odot$ ]	$5M_\odot$
$m_{\text{max}}$	Maximum value of the source mass [ $M_\odot$ ]	$100M_\odot$
$\delta_m$	Smoothing parameter [ $M_\odot$ ]	$5M_\odot$
$b$	Breaking point [ $M_\odot$ ]	$0.4M_\odot$

#### b. Power law plus peak parameters

TABLE VI. Summary of the values injected to construct the PLP model mass distribution of the BBH populations analyzed in Sec. III.

Parameter	Description	Injected value
$\alpha$	Spectral index for the power law of the primary mass distribution	2
$\beta$	Spectral index for the power law of the mass ratio distribution	1
$m_{\text{min}}$	Minimum mass of the primary mass distribution [ $M_\odot$ ]	$5M_\odot$
$m_{\text{max}}$	Maximum mass of the primary mass distribution [ $M_\odot$ ]	$100M_\odot$
$\lambda_g$	Fraction of the model in the Gaussian component	0.1
$\mu_g$	Mean of the Gaussian in the primary mass distribution [ $M_\odot$ ]	$35M_\odot$
$\sigma_g$	Width of the Gaussian in the primary mass distribution [ $M_\odot$ ]	$5M_\odot$
$\delta_m$	Range of mass tapering at the lower end of the mass distribution [ $M_\odot$ ]	$5M_\odot$



### c. Multipeak parameters

TABLE VII. Summary of the values injected to construct the MLTP mass distribution of the BBH populations analyzed in Sec. III.

Parameter	Description	Injected value
$\alpha$	Power law index primary mass	2
$\beta$	Power law index secondary mass	1
$m_{\min}$	Minimum value of the source mass [ $M_{\odot}$ ]	$5M_{\odot}$
$m_{\max}$	Maximum value of the source mass [ $M_{\odot}$ ]	$100M_{\odot}$
$\delta_m$	Smoothing parameter [ $M_{\odot}$ ]	$5M_{\odot}$
$\mu_g^{\text{low}}$	Mean of the lower Gaussian peak [ $M_{\odot}$ ]	$16M_{\odot}$
$\mu_g^{\text{high}}$	Mean of the higher Gaussian peak [ $M_{\odot}$ ]	$50M_{\odot}$
$\sigma_g^{\text{low}}$	Standard deviation of the higher Gaussian peak [ $M_{\odot}$ ]	$8M_{\odot}$
$\sigma_g^{\text{high}}$	Standard deviation of the higher Gaussian peak [ $M_{\odot}$ ]	$8M_{\odot}$
$\lambda_g^{\text{high}}$	Proportion of events in the peaks	0.7
$\lambda_g^{\text{low}}$	Proportion of events in the lower peak	0.8

## 2. Rate parameters simulated

TABLE VIII. Summary of the values injected to construct the CBC merger rate of the BBH populations analyzed in Sec. III.

Parameter	Description	Injected value
$\gamma$	Slope of the power law regime for the rate evolution before the point $z_p$	3
$k$	Slope of the power law regime for the rate evolution after the point $z_p$	3
$z_p$	Redshift turning point between the power law regimes with $\gamma$ and $k$	2

- 
- [1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
- [2] R. Abbott *et al.* (KAGRA, VIRGO, and LIGO Scientific Collaborations), 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).
- [3] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, *Phys. Rev. X* **11**, 021053 (2021).
- [4] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, *Phys. Rev. D* **109**, 022001 (2024).
- [5] W. L. Freedman, Measurements of the Hubble constant: Tensions in perspective, *Astrophys. J.* **919**, 16 (2021).
- [6] J. M. Ezquiaga and D. E. Holz, Spectral sirens: Cosmology from the full mass distribution of compact binaries, *Phys. Rev. Lett.* **129**, 061102 (2022).
- [7] S. Mastroianni, K. Leyde, C. Karathanasis, E. Chassande-Mottin, D. A. Steer, J. Gair, A. Ghosh, R. Gray, S. Mukherjee, and S. Rinaldi, On the importance of source population models for gravitational-wave cosmology, *Phys. Rev. D* **104**, 062009 (2021).
- [8] S. Mastroianni, D. Laghi, R. Gray, G. C. Santoro, A. Ghosh, C. Karathanasis, K. Leyde, D. A. Steer, S. Perries, and G. Pierra, A novel approach to infer population and cosmological properties with gravitational waves standard sirens and galaxy surveys, *Phys. Rev. D* **108**, 042002 (2023).
- [9] S. Mastroianni, G. Pierra, S. Perriés, D. Laghi, G. C. Santoro, A. Ghosh, R. Gray, C. Karathanasis, and K. Leyde,

- ICAROGW: A PYTHON package for inference of astrophysical population properties of noisy, heterogeneous and incomplete observations, [arXiv:2305.17973](https://arxiv.org/abs/2305.17973).
- [10] S. R. Taylor, J. R. Gair, and I. Mandel, Hubble without the Hubble: Cosmology using advanced gravitational-wave detectors alone, *Phys. Rev. D* **85**, 023535 (2012).
- [11] D. Wysocki, J. Lange, and R. O’Shaughnessy, Reconstructing phenomenological distributions of compact binaries via gravitational wave observations, *Phys. Rev. D* **100**, 043012 (2019).
- [12] W. M. Farr, M. Fishbach, J. Ye, and D. Holz, A future percent-level measurement of the Hubble expansion at redshift 0.8 with Advanced LIGO, *Astrophys. J. Lett.* **883**, L42 (2019).
- [13] M. Mancarella, E. Genoud-Prachex, and M. Maggiore, Cosmology and modified gravitational wave propagation from binary black hole population models, *Phys. Rev. D* **105**, 064030 (2022).
- [14] S. Mukherjee, The redshift dependence of black hole mass distribution: Is it reliable for standard sirens cosmology?, *Mon. Not. R. Astron. Soc.* **515**, 5495 (2022).
- [15] R. Gray, Gravitational wave cosmology: Measuring the Hubble constant with dark standard sirens, Ph.D. thesis, University of Glasgow, 2021.
- [16] 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).
- [17] J. R. Gair *et al.*, The hitchhiker’s guide to the galaxy catalog approach for gravitational wave cosmology, *Astron. J.* **166**, 22 (2023).
- [18] K. Leyde, S. Mastrogiovanni, D. A. Steer, E. Chassande-Mottin, and C. Karathanasis, Current and future constraints on cosmology and modified gravitational wave friction from binary black holes, *J. Cosmol. Astropart. Phys.* **09** (2022) 012.
- [19] B. F. Schutz, Determining the Hubble constant from gravitational wave observations, *Nature (London)* **323**, 310 (1986).
- [20] M. Fishbach, D. E. Holz, and W. M. Farr, Does the black hole merger rate evolve with redshift?, *Astrophys. J. Lett.* **863**, L41 (2018).
- [21] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, *Astron. Astrophys.* **594**, A13 (2016).
- [22] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and VIRGO Collaborations), Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Rev. Relativity* **21**, 3 (2018).
- [23] M. Dominik, E. Berti, R. O’Shaughnessy, I. Mandel, K. Belczynski, C. Fryer, D. E. Holz, T. Bulik, and F. Pannarale, Double compact objects III: Gravitational wave detection rates, *Astrophys. J.* **806**, 263 (2015).
- [24] B. Moore, M. Favata, K. G. Arun, and C. K. Mishra, Gravitational-wave phasing for low-eccentricity inspiralling compact binaries to 3PN order, *Phys. Rev. D* **93**, 124061 (2016).
- [25] S. Vitale, D. Gerosa, W. M. Farr, and S. R. Taylor, Inferring the properties of a population of compact binaries in presence of selection effects, in *Handbook of Gravitational Wave Astronomy*, edited by C. Bambi, S. Katsanevas, and K. D. Kokkotas (Springer, Singapore, 2021), 10.1007/978-981-15-4702-7\_45-1.
- [26] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Constraints on the cosmic expansion history from GWTC–3, *Astrophys. J.* **949**, 76 (2023).
- [27] R. Gray *et al.*, Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues, *J. Cosmol. Astropart. Phys.* **12** (2023) 023.
- [28] P. Madau and M. Dickinson, Cosmic star formation history, *Annu. Rev. Astron. Astrophys.* **52**, 415 (2014).
- [29] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Population properties of compact objects from the second LIGO-Virgo gravitational-wave transient catalog, *Astrophys. J. Lett.* **913**, L7 (2021).
- [30] R. Farmer, M. Renzo, S. E. de Mink, P. Marchant, and S. Justham, Mind the gap: The location of the lower edge of the pair instability supernovae black hole mass gap, [arXiv:1910.12874](https://arxiv.org/abs/1910.12874).
- [31] C. Talbot and E. Thrane, Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization, *Astrophys. J.* **856**, 173 (2018).
- [32] C. Karathanasis, B. Revenu, S. Mukherjee, and F. Stachurski, GWSim: PYTHON package for creating mock GW samples for different astrophysical populations and cosmological models of binary black holes, *Astron. Astrophys.* **677**, A124 (2023).
- [33] C. Karathanasis, S. Mukherjee, and S. Mastrogiovanni, Binary black holes population and cosmology in new lights: Signature of PISN mass and formation channel in GWTC-3, *Mon. Not. R. Astron. Soc.* **523**, 4539 (2023).
- [34] M. Mapelli *et al.*, Hierarchical black hole mergers in young, globular and nuclear star clusters: The effect of metallicity, spin and cluster properties, *Mon. Not. R. Astron. Soc.* **505**, 339 (2021).
- [35] M. Mapelli, Y. Bouffanais, F. Santoliquido, M. A. Sedda, and M. C. Artale, The cosmic evolution of binary black holes in young, globular, and nuclear star clusters: Rates, masses, spins, and mixing fractions, *Mon. Not. R. Astron. Soc.* **511**, 5797 (2022).
- [36] N. Giacobbo, M. Mapelli, and M. Spera, Merging black hole binaries: The effects of progenitor’s metallicity, mass-loss rate and Eddington factor, *Mon. Not. R. Astron. Soc.* **474**, 2959 (2018).
- [37] C. L. Fryer, K. Belczynski, G. Wiktorowicz, M. Dominik, V. Kalogera, and D. E. Holz, Compact remnant mass function: Dependence on the explosion mechanism and metallicity, *Astrophys. J.* **749**, 91 (2012).
- [38] M. Mapelli, M. Spera, E. Montanari, M. Limongi, A. Chieffi, N. Giacobbo, A. Bressan, and Y. Bouffanais, Impact of the rotation and compactness of progenitors on the mass of black holes, *Astrophys. J.* **888**, 76 (2020).
- [39] R. Abbott *et al.* (KAGRA, VIRGO, and LIGO Scientific Collaborations), Population of merging compact binaries

- inferred using gravitational waves through GWTC-3, *Phys. Rev. X* **13**, 011048 (2023).
- [40] K. Belczynski, V. Kalogera, and T. Bulik, A comprehensive study of binary compact objects as gravitational wave sources: Evolutionary channels, rates, and physical properties, *Astrophys. J.* **572**, 407 (2001).
- [41] M. Spera and M. Mapelli, Very massive stars, pair-instability supernovae and intermediate-mass black holes with the SEVN code, *Mon. Not. R. Astron. Soc.* **470**, 4739 (2017).
- [42] M. Mapelli, N. Giacobbo, E. Ripamonti, and M. Spera, The cosmic merger rate of stellar black hole binaries from the Illustris simulation, *Mon. Not. R. Astron. Soc.* **472**, 2422 (2017).
- [43] S. Vitale, W.M. Farr, K. Ng, and C.L. Rodriguez, Measuring the star formation rate with gravitational waves from binary black holes, *Astrophys. J. Lett.* **886**, L1 (2019).
- [44] M. Renzo, R.J. Farmer, S. Justham, S.E. de Mink, Y. Göteborg, and P. Marchant, Sensitivity of the lower-edge of the pair instability black hole mass gap to the treatment of time dependent convection, *Mon. Not. R. Astron. Soc.* **493**, 4333 (2020).
- [45] C. Talbot, R. Smith, E. Thrane, and G. B. Poole, Parallelized inference for gravitational-wave astronomy, *Phys. Rev. D* **100**, 043030 (2019).