Cosmology with standard sirens

Bernardo Porto Veronese*

PPGCOSMO, UFES

(Dated: November 29, 2023)

Abstract

The detection of gravitational waves has marked a new era in multi-messenger astronomy. In particular, they can serve as cosmological probes to understand the history of cosmic expansion. This manuscript introduces the general idea behind the method, serving as a pedagogical introduction to gravitational-wave standard sirens. We highlight some of the challenges in the field, and explore current state-of-the-art methods to perform Bayesian inference on cosmological parameters using gravitational-wave data. We also present a simple proof-of-concept implementation[a] on mock data to discuss more technical aspects of the data analysis framework.

I. INTRODUCTION

The idea of using gravitational waves (GWs) from compact binary mergers to measure cosmological parameters was first introduced by Bernard Schutz in 1986 [1]. These signals directly provide a measurement of the luminosity distance measurement to the source, which is therefore independent of the cosmic distance ladder. With the addition of redshift information, measurements can therefore be made of those cosmological parameters which impact the expansion history of the Universe, such as the Hubble constant (H_0). This approach is independent of all other local measurements to date.

The standard siren method probes the expansion history of the universe with the distance-redshift relation, with which one can infer the cosmological parameters such as H_0 and the dark energy equation of state parameter w: [2]

$$
\begin{aligned}
D_l(z) &= (1+z)\frac{c}{H_0\sqrt{\Omega_K}}\sinh\left[\sqrt{\Omega_K}\int_0^z \frac{H_0}{H(z')dz'}\right] \\
\frac{H(z)}{H_0} &= \sqrt{\Omega_m(1+z)^3+\Omega_K(1+z)^2+\Omega_{de}(1+z)^{3(1+w)}}.
\end{aligned}
\tag{1}
$$

To lighten notation, we have omitted the 0-subscript next to the Ω_i's, although they correspond to the present day values in the above equation. Note that using Eq. (1) requires specifying a cosmological model.

[a] The code, as well as the source files of this manuscript, are available in `https://github.com/binado/standard-sirens/tree/main`.

* bernardo.veronese@edu.ufes.br

A. Bright sirens

From the GW data, it is possible to infer the luminosity distance to the binary source, but not the redshift, as the latter is degenerate with the chirp mass in the GW waveform modelling. It is therefore necessary to complement the data with another source of information that provides the redshift measurement. Multi-messenger observations, such as neutron star mergers with electromagnetic counterparts like short gamma-ray bursts or kilonovae, provide the most straight-forward measurement [3, 4]. An electromagnetic counterpart like a kilonova can typically be pinpointed to a specific galaxy, thereby identifying the host galaxy of the GW merger. The GW signal provides the distance to the host galaxy, while its electromagnetic spectrum provides the redshift. These sources are typically referred to as bright sirens. So far, the only confirmed such event has been the binary neutron star detection GW170817, which occurred so exceptionally close to our galaxy - at $d \sim 40\,\mathrm{Mpc}$ - that a direct, model-independent estimation of H_0 with Hubble's law,

$$v_H = H_0 d, \tag{2}$$

could be made by measuring the Hubble flow velocity v_H, resulting in $H_0 = 70.0^{+12.0}_{-8.0}$ km s^{-1} Mpc^{-1} [5].

There are systematic effects and uncertainties that need to be taken care of. The accuracy of the GW luminosity distance measurement is typically of the order of 10%. The main source of uncertainty comes from the degeneracy between the distance and inclination angle of the source. The latter is defined as the angle between the line-of-sight vector from the source to the detector and the orbital-angular momentum of the binary system. This degenrac may be broken by orbital black hole precession [6] or by binaries with asymmetric mass ratios that emit measurable higher- order GW harmonics [7]. In parallel, detector calibration uncertainty also plays a role in this uncertainty, in particular the error on the amplitude of the measured strain.

From the counterpart side, the redshift measurement is affected by peculiar velocity corrections, which are more important for closer sources, as was the case of GW170817 mentioned above.

B. Where are the optical counterparts?

As stated above, almost all GW events have been detected without an EM counterpart. Selection effects play a role in this: optical transients from the CBCs are mostly expected in binary neutron star (BNS) mergers, which are more difficult to detect due to the smaller masses of the sources. At the same time, there are more galaxies at higher redshifts, where optical counterparts are more unlikely to be detected due to fainter signals and more significant noise from objects in the line-of-sight direction of the source.

These *dark sirens* can be used to probe the expansion of the universe provided that they are complemented with an external redshift measurement. In his original paper, Schutz suggested that this information could be inferred from galaxy catalogs: each galaxy's redshift contributes to a hypothetical estimation of H_0, such that the galaxy structure within a GW event's localisation volume is reflected in the H_0 posterior it produces. How informative the individual events are will depend strongly on their localisation volumes. By combining the contributions of many events, the true value of H_0 will be measured as other values will statistically average out. Such analyses have been carried out in the literature, see [8–12]. As an example, Ref. 13 applied the galaxy catalog method with the two best localized dark sirens, GW170814 and GW190814, and the photo-z catalog from the Dark Energy Survey (DES) [14]. A joint analysis with the bright siren event GW170817 provided an $\sim 18\%$ improvement on the 68% confidence interval compared to inferring H_0 with GW170817 alone. Another study used 8 well-localised dark sirens alone to infer $H_0 = 79.8^{+19.1}_{-12.8}$ [15].

At higher redshifts, galaxy-wide surveys are incomplete, and the probability that the catalog contains the merger's host galaxy decreases. On the other hand, both the the gravitational wave sources and galaxies are tracers of the matter density, and therefore, they are spatially correlated through the underlying matter field. Therefore, if the events are well localized, angular correlations between galaxy distributions in redshift and merger distributions in luminosity distance may be used to infer cosmological parameters. Some authors have explored this idea with forecasts for 3G detectors [16]. Refs. 17 and 18 used simulated data to analyse the method's constraining power on H_0 for different numbers of events. They were able to measure the Hubble constant with 2.5% accuracy for $\mathcal{O}(100)$ events. Finally, this technique is not exclusive to standard sirens, and can be applied to any redshift-free distance tracer, such as type Ia supernovae [19].

Alternative methods have been explored in the literature where the inference was set up using GW data alone. One such method consists of using the prior knowledge of the star formation rate and time delay distribution of binary mergers for modelling the prior on the merger population distribution [20–22]. In order to avoid biases by using a fixed merger distribution, it is important to jointly fit its hyperparameters with the GW data. Another approach, called the *spectral siren* method, uses features in the black hole mass distribution to break the source-frame mass and redshift degeneracy with GW data alone [23].

The different methods presented above are not at all independent. Galaxy catalogs, for instance, are most effective with well-localised, low-redshift events. Such events need to be marginalized over a smaller number of galaxies than those that encompass large volumes so that they are more likely to provide better constraints. Better localized events also tend to be less distant, which means that a standard siren analysis would mostly be sensitive to the Hubble constant rather than to other cosmological parameters. At higher redshifts, the catalogs become increasingly uninformative due to their incompleteness. Indeed, it is possible to jointly use galaxy catalog data and black hole population models in a single likelihood formalism [24, 25].

Finally, some authors have investigated the potential of gravitational waves as velocity and density tracers, going beyond the traditional sense of a *standardsiren* as a distance indicator [26, 27].

This manuscript is organized as follows. In Sec. II, we will go over the statistical formalism for data analysis with standard sirens, and we will then specialize in the galaxy catalog method. In Sec. III, we outline the methodology of a proof-of-concept analysis of the discussed bayesian framework using a real galaxy catalog. We present and discuss our results in Sec. IV and we finish with some concluding remarks in Sec. V.

II. STATISTICAL FRAMEWORK

In gravitational-wave astronomy, one line of interest is extracting the distributional properties of a population of sources based on a set of observations which are drawn from that distribution. Any methodology that leads to unbiased estimates of the population parameters must simultaneously account for measurement uncertainties and selection effects. One way with which the latter affects the observed population is a Mamquist bias: the loudest or

brightest sources are more likely to be detected. The standard formalism for extracting the true source population parameters by incorporating these biases in the analysis is frequently labeled as Hierarchical Bayesian inference, see [28–30].

In the discussion below, we will follow the framework outlined in Ref. 31 (hereinafter HG23), which is a pedagogical resource on the galaxy catalog approach.

The GW population distribution is sampled with a set of $N_{\rm obs}$ *observed* events with true parameters $\{\vec{\theta}_i\}$, $i \in \{1, \cdots, N_{\rm obs}\}$. We do not have direct access to the true parameters because of noise; instead, we have a set of measured data $\{\vec{d}_i\}$. The $\vec{\theta}_i$ are the individual object parameters, although we are interested in the population hyperparameters, which we call $\vec{\Lambda}$. We cannot determine $\vec{\Lambda}$ directly, but we can compute the posterior probability given the observations. In the usual Bayesian formalism,

$$p(\vec{\Lambda}|\{\vec{d}_i\}) = \frac{p(\{\vec{d}_i\}|\vec{\Lambda})\pi(\vec{\Lambda})}{p(\{\vec{d}_i\})} \tag{3}$$

where $p(\{\vec{d}_i\}|\vec{\Lambda})$ is the likelihood of observing the dataset given the population properties, $\pi(\vec{\Lambda})$ is the prior on $\vec{\Lambda}$ and $p(\{\vec{d}_i\})$ is the evidence, which is the integral of the numerator over $\vec{\Lambda}$.

In the spirit of Ref. 29, we first start with the idealized scenario where the event parameters are perfectly measured. The total likelihood for the set of $N_{\rm obs}$ independent measurements is then

$$p(\{\vec{\theta}_i\}|\vec{\Lambda}) = \prod_{i=1}^{N_{\rm obs}} \frac{p_{\rm pop}(\vec{\theta}_i|\vec{\Lambda})}{\int p_{\rm pop}(\vec{\theta}_i|\vec{\Lambda}) d\vec{\Lambda}} \tag{4}$$

where $p_{\rm pop}(\vec{\theta}|\vec{\Lambda})$ is related to the number density dN of objects expected to be found in the region $[\vec{\theta}, \vec{\theta} + d\vec{\theta}]$:

$$dN = N p_{\rm pop}(\vec{\theta}|\vec{\Lambda}) d\vec{\theta} \tag{5}$$

We shall build an incrementally more robust model than Eq. (4). Let us first consider the presence of selection effects: not all events are equally likely to be detected. We can encode this with a detection probability $p_{\rm det}$. In the perfect measurement idealization, this detection probability becomes a function of the parameters $\vec{\theta}$ only. In the general case, where noise is present, the detection probability is a function of the data. Let $\mathcal{D}$ be the set of all data. To

determine whether an event is detectable, one can use a detection statistic $\rho_{\mathcal{D}}$, which can be calculated for each piece of data. In practice, this statistic can be the signal-to-noise ratio (SNR), the false-alarm rate, etc. We split $\mathcal{D}$ into two disjoints sets, $\mathcal{D}_<$ and $\mathcal{D}_\geq$, according to whether $\rho_{\mathcal{D}}$ is smaller than a threshold $\rho_{\rm tr}$ or not. Then

$$p_{\rm det}(\vec{\theta}) = \int_{\mathcal{D}_\geq} p(\vec{d}|\vec{\theta}) d\vec{d} \tag{6}$$

The probability of observing a particular dataset $\vec{d}$ given the assumed population distribution parameterised by $\vec{\Lambda}$ is

$$p(\vec{d}|\vec{\Lambda}) = \frac{\int p(\vec{d}|\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\Lambda}) d\vec{\theta}}{\alpha(\vec{\Lambda})} \tag{7}$$

where $\alpha(\vec{\Lambda})$ is a normalization factor integrated over the set of detectable data,

$$\begin{aligned}
\alpha(\vec{\Lambda}) &= \int_{\mathcal{D}_\geq} d\vec{d} \int p(\vec{d}|\vec{\theta}) p_{\rm pop}(\vec{d}|\vec{\Lambda}) d\vec{\theta} &(8)\\
&= \int \left[\int_{\mathcal{D}_\geq} p(\vec{d}|\vec{\theta}) d\vec{d} \right] p_{\rm pop}(\vec{d}|\vec{\Lambda}) d\vec{\theta} &(9)\\
&= \int p_{\rm det}(\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\Lambda}) d\vec{\theta} &(10)
\end{aligned}$$

Hence, in the presence of both measurement uncertainties and selection effects, Eq. (4) becomes

$$p(\{\vec{d_i}\}|\vec{\Lambda}) = \prod_{i=1}^{N_{\rm obs}} \frac{\int p(\vec{d_i}|\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\Lambda}) d\vec{\theta}}{\int p_{\rm det}(\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\Lambda}) d\vec{\theta}} \tag{11}$$

We can also include the population rate into the framework. The probability of observing k events with an expected number of detections $N_{\rm det}$ is given by a Poisson distribution as

$$p(k|N_{\rm det}) = e^{-N_{\rm det}} (N_{\rm det})^{N_{\rm obs}} \tag{12}$$

The usual $N_{\rm obs}!$ term is absent in Eq. (12) because the events are distinguishable from the observed data. When accouting for selection effects, the expected number of detections $N_{\rm det}$ becomes

$$N_{\rm det}(\vec{\Lambda}) = \int_{\mathcal{D}_{\geq}} d\vec{d} \int p(\vec{d}|\vec{\theta}) \frac{dN}{d\vec{\theta}} d\vec{\theta} \tag{13}$$

$$= \int \left[\int_{\mathcal{D}_{\geq}} p(\vec{d}|d)\vec{d} \right] \frac{dN}{d\vec{\theta}} d\vec{\theta} \tag{14}$$

$$= \int p_{\rm det}(\vec{\theta}) \frac{dN}{d\vec{\theta}} d\vec{\theta} \tag{15}$$

$$= \int p_{\rm det}(\vec{\theta}) N p_{\rm pop}(\vec{\theta}|\vec{\Lambda}) d\vec{\theta} \tag{16}$$

$$= N\alpha(\Lambda) \tag{17}$$

where the last two equalities are derived from Eq. (5) and Eq. (10) respectively. The full posterior with the population rate is then

$$\begin{aligned} p(\vec{\Lambda}, N|\vec{d}) &= p(N|\vec{\Lambda}, \vec{d}) p(\vec{\Lambda}|\vec{d}) \\ &= e^{-N_{\rm det}} (N_{\rm det})^{N_{\rm obs}} \pi(N) \pi(\vec{\Lambda}) \alpha(\vec{\Lambda})^{-N_{\rm obs}} \prod_{i=1}^{N_{\rm obs}} \int p(\vec{d_i}|\vec{\theta}) p_{\rm pop}(\vec{\theta}|\vec{\Lambda}) d\vec{\theta} \end{aligned} \tag{18}$$

If a prior $\pi(N) \propto 1/N$ is assumed on the population rate, then the posterior can be marginalized over N:

$$\int e^{-N_{\rm det}} (N_{\rm det})^{N_{\rm obs}} \frac{dN}{N} = \int e^{-N_{\rm det}} (N_{\rm det})^{N_{\rm obs}-1} dN_{\rm det} \tag{19}$$

$$= (N_{\rm obs} - 1)! \tag{20}$$

So far, the framework we developed has been general; we now specify to the gravitational wave case. The individual event parameters $\vec{\theta}$ describe the compact binary coalescence (CBC): the individual masses and spins, sky position, polarization, inclination angle, luminosity distance, and redshift. The population parameters $\vec{\Lambda}$ can be split into three groups: mass, rate and cosmological parameters. The mass parameters specify the GW mass model, such as minimum and maximum mass, slopes, the positions of any features in the mass distribution function, etc. These are used in the spectral siren method. The rate parameters are used in the model to describe how the CBC merger rate evolves with redshift. Finally, the cosmological parameters are the constants which appear in Eq. (1), namely $\{H_0, \Omega_m, \Omega_{de}, w\}$.

A. A simplified approach

In this section, we reproduce the formalism developed in HG23. In that paper, the authors perform a mock data analysis of the dark siren approach using the galaxy catalog method to demonstrate its capability to recover an unbiased posterior for H_0. They consider that the remaining cosmological parameters, such as Ω_m and $w_{\rm DE}$, are fixed to fiducial values. They make a series of simplifying assumptions on p_{det}, p_{pop} and $p(\vec{d}|\vec{\theta})$, which we discuss below.

The selection effects, encoded in the $\alpha(\Lambda)$ term from Eq. (10), not only depend on the underlying population modelling prior $p_{\rm pop}$, but they are also affected by the (sky-dependent) GW detector sensitivity and detector-frame mass. A complete analytical treatment is thus intractable. Hence, the detection efficiency must be estimated by drawing synthetic objects from a fiducial distribution, $p_{\rm draw}(\theta)$, drawing corresponding data from the likelihood function $p(\vec{d}|\vec{\theta})$, and "injecting" these data into the pipeline used to produce the GW event catalog, recording which observations are detected. The number of required injections to estimate the selection function precisely enough scales linearly with the size of the event catalog [32].

Instead, we make the following simplification: detection is assumed to happen if the *measured* luminosity distance is smaller than a threshold, $d_L^{\rm th}$. If the GW likelihood is taken to be gaussian, that is,

$$\mathcal{L}(\hat{d}_L^i|d_L(z,H_0)) = \frac{1}{\sqrt{2\pi}\sigma_{d_L}} \exp\left[-\frac{1}{2}\left(\frac{\hat{d}_L^i - d_L(z,H_0)}{\sigma_{d_L}}\right)^2\right], \tag{21}$$

then the detection probability can be expressed analytically with Eq. (6):

$$\begin{aligned} p_{\rm det}(z,H_0) &= \int_{-\infty}^{d_L^{\rm th}} \mathcal{L}(\hat{d}_L^i|d_L(z,H_0))d\hat{d}_L^i \\ &= \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\hat{d}_L^i - d_L(z,H_0)}{\sqrt{2}\sigma_{d_L}}\right)\right]. \end{aligned} \tag{22}$$

where erf is the unilateral error function of the standard normal distribution. The uncertainty σ_{d_L} is taken to be a constant fraction of d_L, such that $\sigma_{d_L}/d_L = C < 1$.

The galaxy catalog information is used to compute the population model prior $p_{\rm pop}(\vec{\theta}|\vec{\Lambda}) = p_{\rm pop}(z|H_0)$. Each galaxy in the catalog contributes with a term

$$\mathcal{L}_{\rm EM}(\hat{z}_i|z_i)p(z_i|H_0), \tag{23}$$

where z_i and $\hat{z}_i$ are the galaxy's true and measured values, respectively. The likelihood encodes the measurement uncertainty, while the redshift prior depends on our knowledge of the galaxy distribution on redshift. A simple choice is to pick $p(z_i|H_0)$ to be uniform in a comoving volume, $p(z_i|H_0) \propto dV_c/dz$. The posterior probability becomes

$$p(H_0, \{z_g\}|d_{\rm EM}, d_{\rm GW}) \propto \alpha^{-1}(H_0) \left[\sum_{i=1}^{N_{\rm gal}} \mathcal{L}_{\rm GW}(d_{\rm GW}|d_L(z_i, H_0))\right] \prod_{j=1}^{N_{\rm gal}} \mathcal{L}_{\rm EM}(\hat{z}_j|z_j)p(z_j|H_0). \tag{24}$$

The dependence on the true galaxy redshifts $\{z_g\}$ can then be marginalized over to recover the posterior distribution on H_0 only.

Note that in the above expression we are implicitly neglecting cross-correlations between galaxies, for instance due to clustering. On the approximation that the galaxy redshifts are measured perfectly, the likelihood $\mathcal{L}_{\rm EM}$ becomes a delta function, and the posterior for a single GW event reduces to a simple form:

$$p(H_0|d_{\rm EM}, d_{\rm GW}) \propto \frac{\sum_{i=1}^{N_{\rm gal}} \mathcal{L}_{\rm GW}(d_L^{\rm GW}|d_L(\hat{z}_i, H_0))}{\sum_{i=1}^{N_{\rm gal}} p_{\rm det}(\hat{z}_i, H_0)}. \tag{25}$$

Alternatively, we model the photo-z redshift likelihood as a gaussian,

$$\mathcal{L}_{\rm EM}(\hat{z}_j|z_j) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left[-\frac{1}{2}\left(\frac{\hat{z}_j - z_j}{\sigma_z}\right)^2\right], \tag{26}$$

with $\sigma_z \sim \min\{0.033(1+\mathrm{z}), 0.015\}$, following an empirical fit described in Ref. 12.

III. METHOD

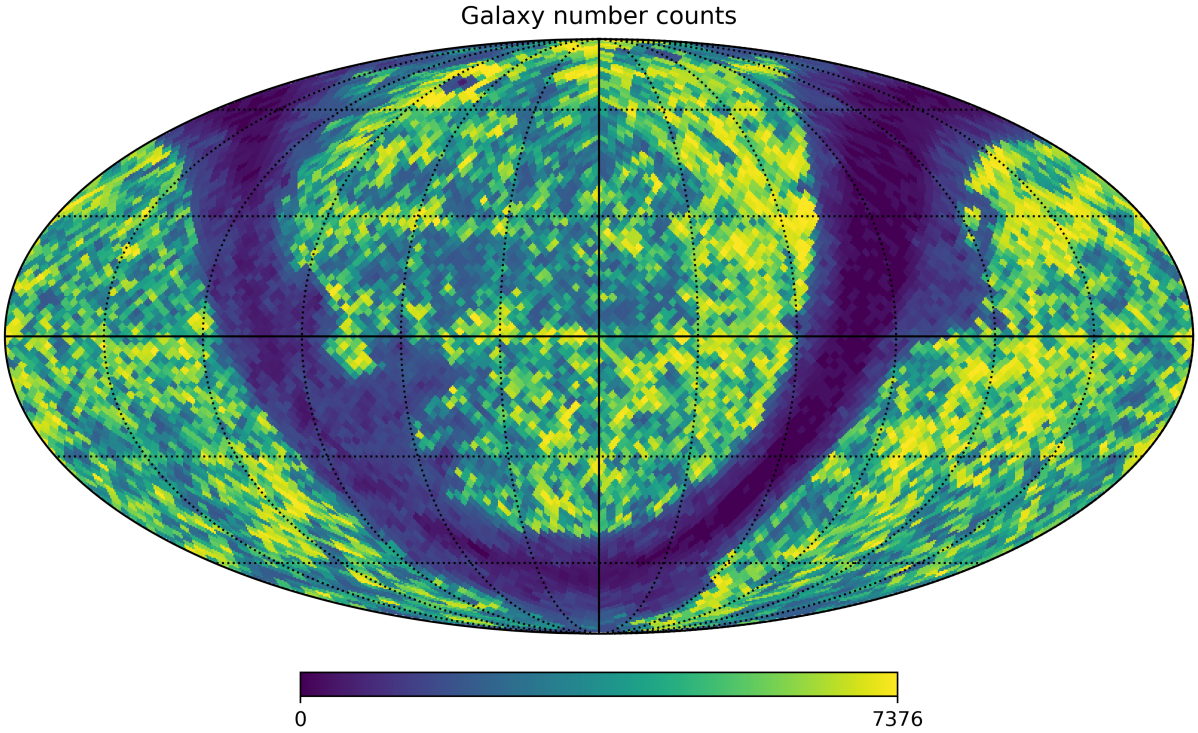
We follow the framework developed in Sec. II A as follows. Unlike HG23, we use real data from the `GLADE+` catalog [33, 34] [1], which is a compilation of six different astronomical surveys, containing about 22 million galaxies. It covers the full sky with a 20% completeness up to about 800 Mpc.

[1] The catalog is publicly available for download at `https://glade.elte.hu/`.

A. Preparing the galaxy catalog

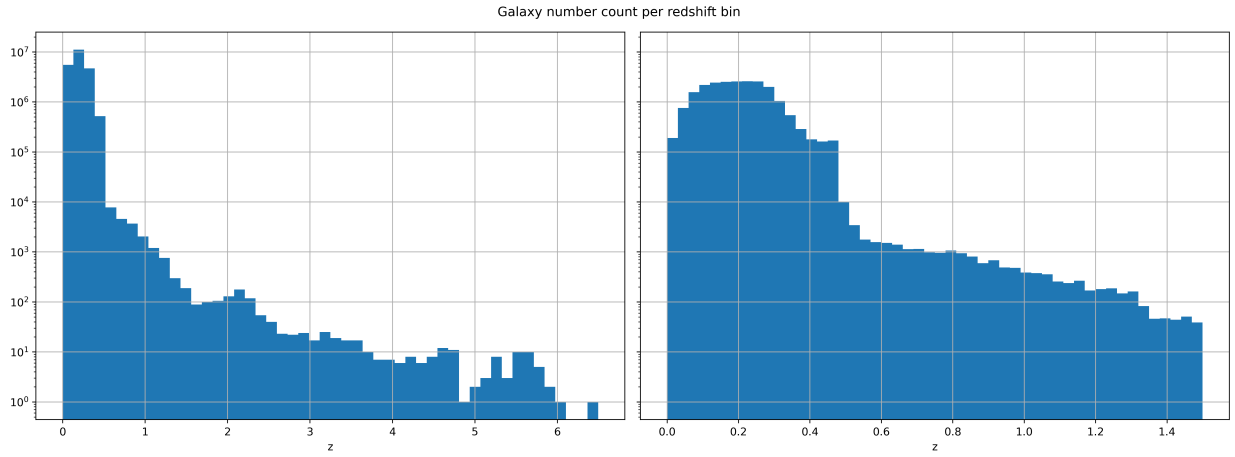
We implement a simple post-processing pipeline to the `GLADE+` catalog. We remove objects that are identified as quasars or galaxy clusters, objects with no measured redshift, and close-by galaxies ($z < 0.05$) with no peculiar velocity corrections. The sky locations of the galaxies are assigned to unique, equally-sized pixels using `healpy`, a Python implementation of the HEALPIX software [35, 36] [2]. The `nside` parameter, which controls the pixel resolution, is set to 32. The total number of pixels is $12\texttt{nside}^2 = 12888$, and each pixel covers approximately 1.8° of the sky. We do not analyse the impact of the choice of this parameter on our results, which has been shown not to be significant [37]. Fig. 1 illustrates the galaxy distribution in the catalog across the whole sky, and Fig. 2 shows a histogram of the galaxy distribution per redshift.

FIG. 1. A Mollweide projection of the galaxy number counts across the sky after our post-processing step of the `GLADE+` catalog.



[2] `http://healpix.sourceforge.net`

FIG. 2. *Left:* a histogram of the galaxy number count per redshift bin in the (whole-sky) catalog. *Right:* the same histogram for all galaxies up to $z = 1.5$.



B. The gravitational wave data

The gravitational wave data is generated from the catalog directly:

1. A direction (θ, φ) on the sky is picked, and all galaxies within an angular radius α of it are selected as candidate hosts for the coalescences [3].

2. A sample of $N_{\rm gw}$ redshifts are drawn uniformly from said galaxies in the range $0 \leq z \leq z_{\rm max}$, where $z_{\rm max}$ is a constant value fixed *a priori*.

3. The drawn redshifts are converted into true luminosity distances supposing a fiducial flat ΛCDM cosmology with $\Omega_m = 0.3$ and $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$. Each distance value is associated to a mock gravitational wave event.

4. The *measured* luminosity distances are drawn from their respective true values consistently with the GW gaussian likelihood of Eq. (21).

To mitigate biases from the structure distribution in a particular direction, we sample mergers in different directions and perform the inference on all of them.

Similar to HG23, we analyse two separate cases: the perfect-redshift limit and the full likelihood from Eq. (24). We set $\sigma_L/d_L \in \{0.1, 0.2, 0.3\}$, and we fix $z_{\rm max} = 1.4$ and $d_L^{\rm th} =$

[3] The direction is redrawn if there are less than $N_{\rm min} \sim N_{\rm GW}$ galaxies within it.

1550 Mpc. We choose a uniform prior $H_0 \in [20, 140]\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Since we leave the remaining cosmological parameters fixed, we implement a simple grid method to compute the marginalized likelihood.

Our methodology is evidently a simple proof of concept of the galaxy catalog approach to dark siren cosmography. Still, it is important to highlight the limitations to our method, as well as the approximations that are made. Firstly, the simplified data-generating process: we are generating mock events *uniformly* from the available galaxies in the catalog. The resulting redshift distribution of CBCs is inevitably inaccurate for two reasons: the galaxies may not equally be likely to host these mergers; this may also be a function of their masses, luminosities, metallicities, etc. Secondly, the galaxy distribution in the catalog does not reproduce the real distribution due to the the latter's incompleteness. The selected redshifts will be artificially small because of this. This should not spoil the analysis, so long as the mock events' generation is consistent with the likelihood model. In particular, we expect that the posterior distribution should peak around the fiducial value of H_0 used in the $z - d_L$ conversion. The galaxy distribution in the catalog, however, is not expected to be reproduce this background cosmology. Moreover, the (assumed true) redshifts of the coalescences are drawn from real measurements, and therefore do not correspond to the true redshits of their corresponding hosts. This distinction does not play a role when the spectroscopic$-z$ approximation is considered, namely the likelihood in Eq. (25), although it becomes important for the broad case where redshift uncertainty is incorporated into the analysis.

A number of approximations were also made in our likelihood model. The GW likelihood $p(\vec{d}|\vec{\theta})$ does not, in general, follow a normal distribution. In real analyses, while a similar ansatz can be used for the conditional probability of the distance to the event given a sky direction [38], a more accurate procedure is to apply Bayes's theorem and to discretely sample the posterior $p(\vec{\theta}|\vec{d})$ reconstructed in the single-event parameter estimation with assumed prior $\pi(\vec{\theta})$ [29]. Likewise, the photo$-z$ PDF may be more complicated than Eq. (26), and can be estimated with more sophisticated methods, such as neural networks [15, 39]. Moreover, as already mentioned, a more realistic computation of the selection effects involves injecting artificial signals and computing the number of detections at the selected sensitivity. Finally, when using real GW data, the set of candidate host galaxies should be extracted from the credible regions estimated from the events' skymaps - for instance, with the `BAYESTAR`

package [40].

We have also explicitly ignored completeness corrections to the magnitude-limited galaxy catalog. In our toy model, this is not as problematic since the CBCs are generated from the catalog itself. In the general case, however, not accounting for the incompleteness will make the inference biased as well as far less informative for events that occured at higher redshifts, which are commonplace for BBHs. See Ref. 41 for a review of different completeness corrections adopted in the literature.

The implementation is available in the Github link `https://github.com/binado/standard-sirens/tree/main`. The goal of the repository is to make it a playground for exploring different strategies for performing inference with standard sirens.

IV. RESULTS AND DISCUSSION

The results are presented in Figs. 3 - 4, for the spectroscopic-like and photometric-like redshift uncertainties, respectively. Fig 5 represents a side-by-side comparison of both cases for different levels of luminosity distance uncertainty. As expected, a GW likelihood with smaller uncertainty generates a more informative posterior. Nevertheless, since we are dealing with simulated data, it is imporatant to keep in mind that the position and the height of the posterior peaks fluctuate depending on the seed of the random generator used throughout the code. Since each curve is generated from a set of independent mock events, it is thus not meaningful to draw conclusions from their direct comparison out of single pipeline run. The features in line-of-sight redshift distribution are imprints of the structure along that direction, and the GW data, by construction, approximately reproduces them. To analyse the dependence of the reconstructed posterior on the data in a quantitative manner, we run the inference pipeline $N_{\mathrm{sim}} = 100$ times, each with a different fiducial value of H_0 drawn uniformly from the prior range $[20, 140]\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. We compute the minimal symmetric $\alpha-$credible interval (CI) in which these fiducial values are located in the corresponding posterior distributions. We then analyse the histogram distribution of α. If the inference is unbiased, we expect the fiducial values to lie within the symmetric CI with percentile α a number αN_{sim} of times. This is an example of a p-p plot. This procedure is computationally viable because our 1-dimensional posterior does not take longer than 1 minute to compute. In real analyses, however, where (multiple) cosmological parameters are jointly estimated

with, for instance, the black hole mass distribution, running several MCMC chains may be intractable, or require a clever parallelization scheme. The p-p plots are represented in Figs. 6. We observe that all curves approximately follow the straight 45° line which, as argued above, is the expected behaviour for an unbiased inference. There are no significant deviations from it at any interval.

FIG. 3. The posterior probability distribution on H_0 for the inference under the perfect redshift measurement assumption, and different values of $C = \sigma_L/d_L$. The dotted black line represents the fiducial value used to generate the GW data.

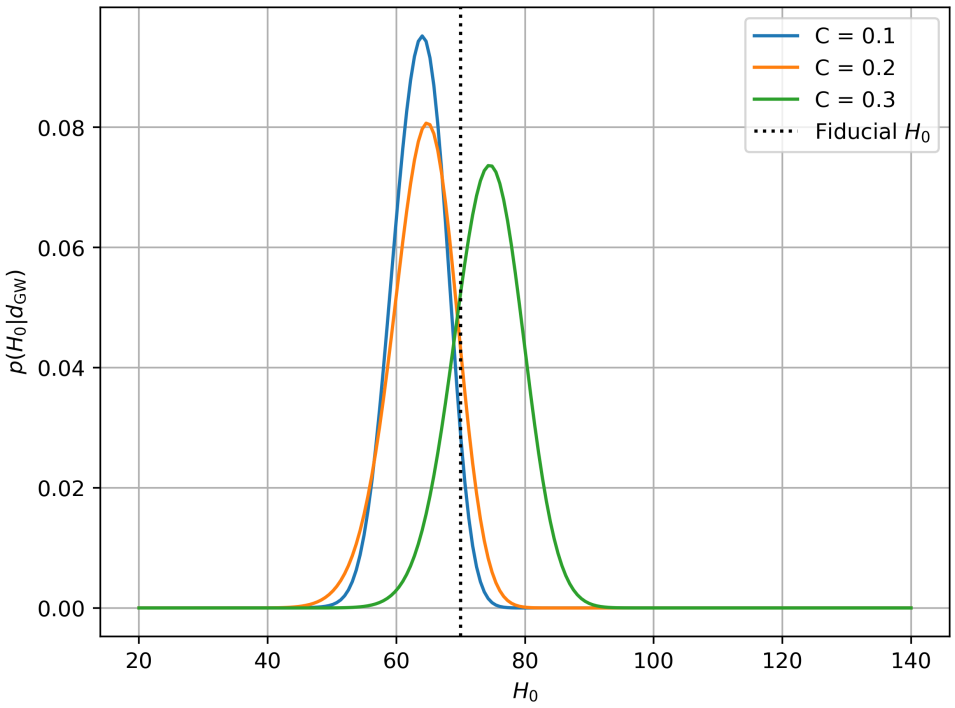
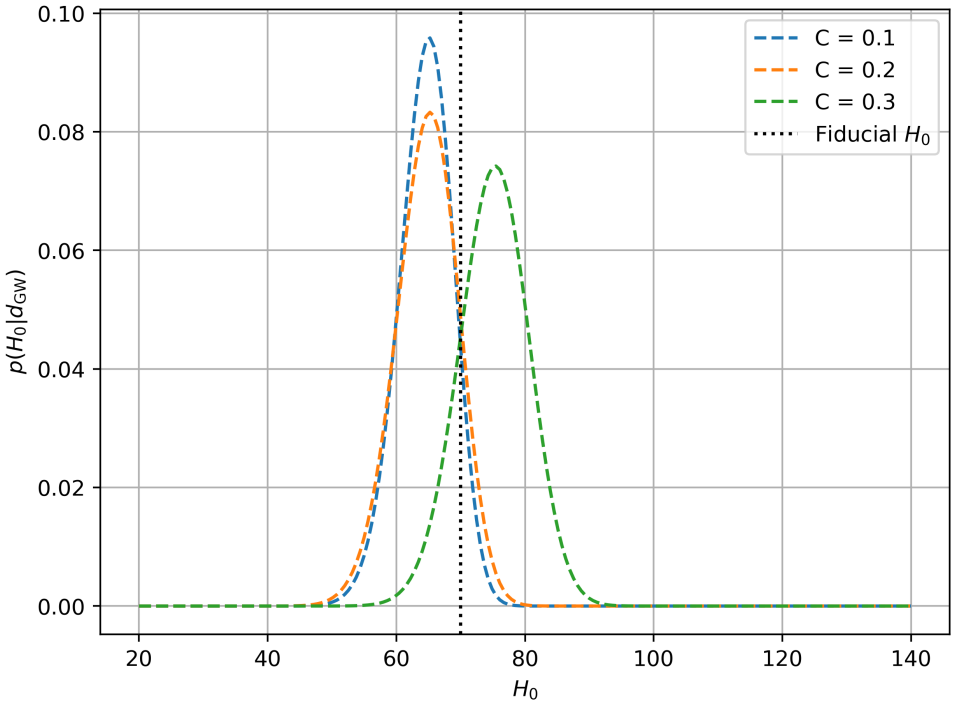


FIG. 4. The posterior probability distribution on H_0 for the inference under the normally distributed redshift measurement assumption, and different values of $C = \sigma_L/d_L$. The dotted black line represents the fiducial value used to generate the GW data.



Finally, the analysis could take other directions as well. In HG23, the authors analysed limiting cases in the object distributions, such as the few-galaxy, many-mergers configuration, where the redshift of different CBC's may originate from the same host galaxy, in which case they are correlated. We have also not explored the impact of the choice of fixed parameters, such as $z_{\max}$ or $d_L^{\rm th}$, on the robustness of the results. Likewise, we could also promote the remaining cosmological parameters to free hyperparameters of the model, and analyse the impact on the constraining power on H_0.

V. CONCLUSION

This manuscript is intended as a pedagogical introduction to gravitational-wave standard sirens. With more detections of mergers accumulated in the next LIGO-VIRGO-KAGRA runs, and especially with the advent of third generation interferometer networks in the next decade, gravitational waves may be a powerful and independent tool to measure the

FIG. 5. The posterior probability distribution on H_0 for the inference under both spec$-z$ and photo-z cases, and for $C = 0.1$ (top), $C = 0.2$ (middle), and $C = 0.3$ (bottom). The dotted black lines represent the fiducial value used to generate the GW data.

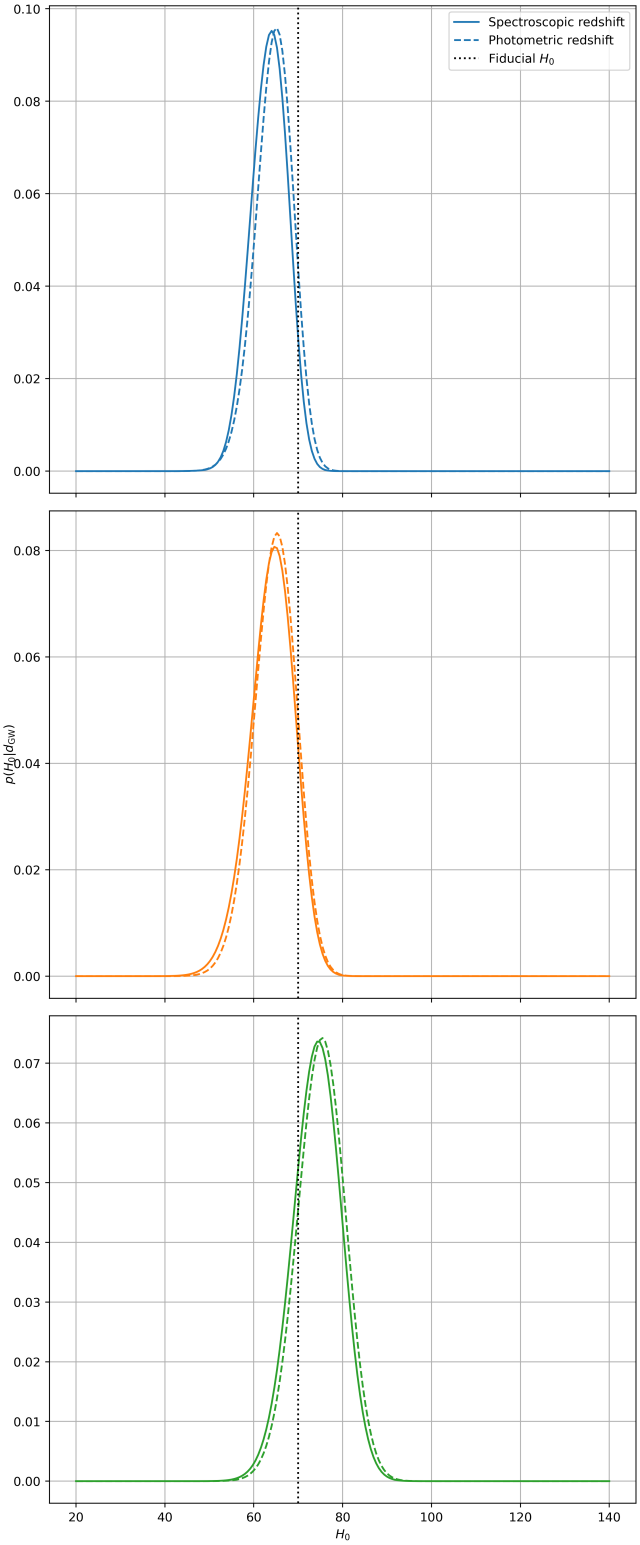
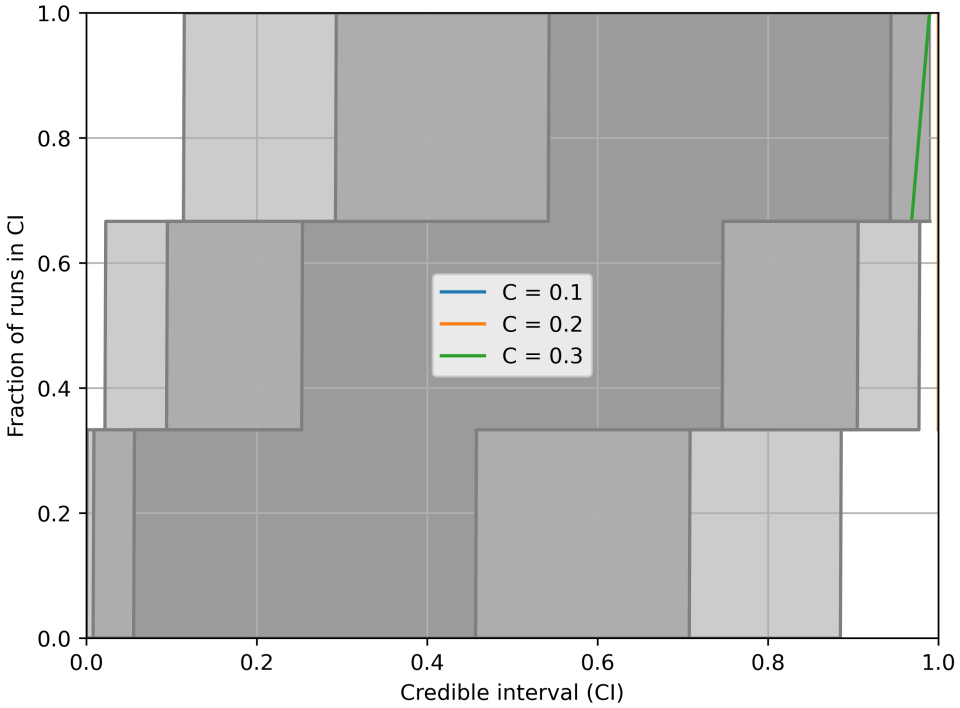


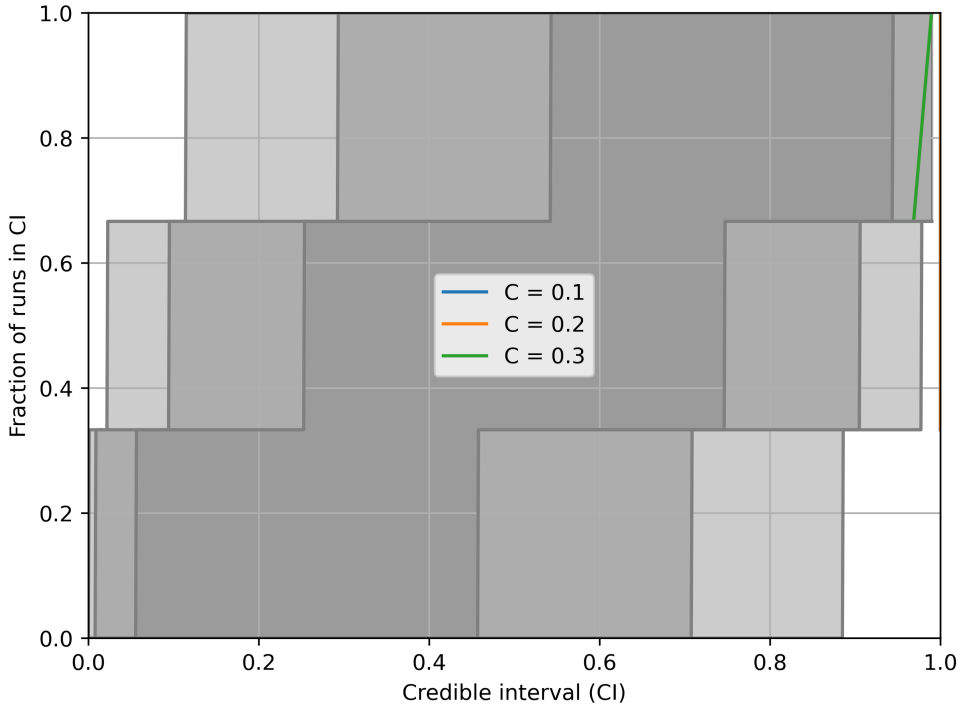
FIG. 6. p-p plot for the inference on perfect redshift measurement limit. The pipeline was run 100 times. The blue, orange and green curves correspond to $C = 0.1\,0.2,\,0.3$ respectively. The grey regions correspond to the 1σ, 2σ and 3σ confidence levels.



cosmic expansion history of the Universe. Nevertheless, there are still challenges to deal with, which are mostly related to mitigating the systematics in the process: the detector calibration uncertainty, the measurement uncertainty of the source's luminosity distance, and so on. On the dark siren case, where no optical counterpart is available, these systematic uncertainties are even more pronounced: while the informative power of galaxy catalogs is significantly diminished at higher redshifts, using fixed BBH population assumptions, which are more useful there, may dominate the outcome of the inference [42]. Therefore, while the standard-siren methodology is still limited by statistical uncertainties, the multiple years of data collection process will be plenty of time for scientists in the field to develop more sophisticated tools to model the systematics into statistical uncertainties that can be marginalised over.

In the spirit of presenting pedagogical resources for a hands-on introduction to the field, we have also presented a simple proof-of-concept implementation of an inference pipeline using simulated dark sirens and the galaxy catalog method. Although our approach sligthly

FIG. 7. p-p plot for the inference on perfect redshift measurement limit. The pipeline was run 100 times. The blue, orange and green curves correspond to $C = 0.1\,0.2, 0.3$ respectively. The grey regions correspond to the 1σ, 2σ and 3σ confidence levels.



differs from Ref. 31 in that the redshift data comes from a real catalog, `GLADE+`, we reach a similar conclusion, namely that that a data-generating process consistent with the likelihood model assumptions led to an unbiased inference on the Hubble constant. A natural extension of this work is to modify the implementation to work with real gravitational-wave data, and to include other statistical methods studied in the literature, for instance incorporating black hole population models.

[1] B. F. Schutz, Determining the Hubble Constant from Gravitational Wave Observations, Nature **323**, 310 (1986).

[2] D. W. Hogg, Distance measures in cosmology, (1999), arXiv:astro-ph/9905116.

[3] D. E. Holz and S. A. Hughes, Using gravitational-wave standard sirens, Astrophys. J. **629**, 15 (2005), arXiv:astro-ph/0504616.

[4] N. Dalal, D. E. Holz, S. A. Hughes, and B. Jain, Short grb and binary black hole standard

sirens as a probe of dark energy, Phys. Rev. D **74**, 063006 (2006), arXiv:astro-ph/0601275.

[5] B. P. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER), A gravitational-wave standard siren measurement of the Hubble constant, Nature **551**, 85 (2017), arXiv:1710.05835 [astro-ph.CO].

[6] Q. Yun, W.-B. Han, Q. Hu, and H. Xu, Precessing binary black holes as better dark sirens, Monthly Notices of the Royal Astronomical Society: Letters **527**, L60–L65 (2023).

[7] S. Vitale and H.-Y. Chen, Measuring the hubble constant with neutron star black hole mergers, Physical Review Letters **121**, 10.1103/physrevlett.121.021303 (2018).

[8] W. Del Pozzo, Inference of the cosmological parameters from gravitational waves: application to second generation interferometers, Phys. Rev. D **86**, 043011 (2012), arXiv:1108.1317 [astro-ph.CO].

[9] H.-Y. Chen, M. Fishbach, and D. E. Holz, A two per cent Hubble constant measurement from standard sirens within five years, Nature **562**, 545 (2018), arXiv:1712.06531 [astro-ph.CO].

[10] M. Fishbach *et al.* (LIGO Scientific, Virgo), A Standard Siren Measurement of the Hubble Constant from GW170817 without the Electromagnetic Counterpart, Astrophys. J. Lett. **871**, L13 (2019), arXiv:1807.05667 [astro-ph.CO].

[11] R. Gray *et al.*, Cosmological inference using gravitational wave standard sirens: A mock data analysis, Phys. Rev. D **101**, 122001 (2020), arXiv:1908.06050 [gr-qc].

[12] M. Soares-Santos *et al.* (DES, LIGO Scientific, Virgo), First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary–Black-hole Merger GW170814, Astrophys. J. Lett. **876**, L7 (2019), arXiv:1901.01540 [astro-ph.CO].

[13] A. Palmese *et al.* (DES), A statistical standard siren measurement of the Hubble constant from the LIGO/Virgo gravitational wave compact object merger GW190814 and Dark Energy Survey galaxies, Astrophys. J. Lett. **900**, L33 (2020), arXiv:2006.14961 [astro-ph.CO].

[14] T. D. E. S. Collaboration, The dark energy survey (2005), arXiv:astro-ph/0510346 [astro-ph].

[15] A. Palmese, C. R. Bom, S. Mucesh, and W. G. Hartley, A standard siren measurement of the hubble constant using gravitational-wave events from the first three ligo/virgo observing runs and the desi legacy survey, The Astrophysical Journal **943**, 56 (2023).

[16] M. Oguri, Measuring the distance-redshift relation with the cross-correlation of gravitational wave standard sirens and galaxies, Physical Review D **93**, 10.1103/physrevd.93.083511 (2016).

[17] S. Bera, D. Rana, S. More, and S. Bose, Incompleteness matters not: Inference of h 0 from binary black hole–galaxy cross-correlations, The Astrophysical Journal **902**, 79 (2020).

[18] S. Mukherjee, B. D. Wandelt, S. M. Nissanke, and A. Silvestri, Accurate precision cosmology with redshift unknown gravitational wave sources, Physical Review D **103**, 10.1103/physrevd.103.043520 (2021).

[19] S. Mukherjee and B. D. Wandelt, Beyond the classical distance-redshift test: cross-correlating redshift-free standard candles and sirens with redshift surveys (2018), arXiv:1808.06615 [astro-ph.CO].

[20] X. Ding, M. Biesiada, X. Zheng, K. Liao, Z. Li, and Z.-H. Zhu, Cosmological inference from standard sirens without redshift measurements, Journal of Cosmology and Astroparticle Physics **2019** (04), 033–033.

[21] C. Ye and M. Fishbach, Cosmology with standard sirens at cosmic noon, Physical Review D **104**, 10.1103/physrevd.104.043507 (2021).

[22] H. Leandro, V. Marra, and R. Sturani, Measuring the hubble constant with black sirens, Physical Review D **105**, 10.1103/physrevd.105.023523 (2022).

[23] J. M. Ezquiaga and D. E. Holz, Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries, Phys. Rev. Lett. **129**, 061102 (2022), arXiv:2202.08240 [astro-ph.CO].

[24] S. Mastrogiovanni, D. Laghi, R. Gray, G. C. Santoro, A. Ghosh, C. Karathanasis, K. Leyde, D. A. Steer, S. Perries, and G. Pierra, Joint population and cosmological properties inference with gravitational waves standard sirens and galaxy surveys, Phys. Rev. D **108**, 042002 (2023), arXiv:2305.10488 [astro-ph.CO].

[25] R. Gray, F. Beirnaert, C. Karathanasis, B. Revenu, C. Turski, A. Chen, T. Baker, S. Vallejo, A. E. Romano, T. Ghosh, A. Ghosh, K. Leyde, S. Mastrogiovanni, and S. More, Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues (2023), arXiv:2308.02281 [astro-ph.CO].

[26] A. Palmese and A. Kim, Probing gravity and growth of structure with gravitational waves and galaxies' peculiar velocity, Physical Review D **103**, 10.1103/physrevd.103.103507 (2021).

[27] V. Alfradique, M. Quartin, L. Amendola, T. Castro, and A. Toubiana, The lure of sirens: joint distance and velocity measurements with third-generation detectors, Mon. Not. Roy. Astron. Soc. **517**, 5449 (2022), arXiv:2205.14034 [astro-ph.CO].

[28] T. J. Loredo, Accounting for source uncertainties in analyses of astronomical survey data, AIP

Conf. Proc. **735**, 195 (2004), arXiv:astro-ph/0409387.

[29] I. Mandel, W. M. Farr, and J. R. Gair, Extracting distribution parameters from multiple uncertain observations with selection biases, Mon. Not. Roy. Astron. Soc. **486**, 1086 (2019), arXiv:1809.02063 [physics.data-an].

[30] 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* (Springer Singapore, 2021) pp. 1–60.

[31] J. R. Gair, A. Ghosh, R. Gray, D. E. Holz, S. Mastrogiovanni, S. Mukherjee, A. Palmese, N. Tamanini, T. Baker, F. Beirnaert, M. Bilicki, H.-Y. Chen, G. Dá lya, J. M. Ezquiaga, W. M. Farr, M. Fishbach, J. Garcia-Bellido, T. Ghosh, H.-Y. Huang, C. Karathanasis, K. Leyde, I. M. Hernandez, J. Noller, G. Pierra, P. Raffai, A. E. Romano, M. Seglar-Arroyo, D. A. Steer, C. Turski, M. P. Vaccaro, and S. A. Vallejo-Peña, The hitchhiker's guide to the galaxy catalog approach for dark siren gravitational-wave cosmology, The Astronomical Journal **166**, 22 (2023).

[32] R. Essick and W. Farr, Precision requirements for monte carlo sums within hierarchical bayesian inference (2022), arXiv:2204.00461 [astro-ph.IM].

[33] G. Dálya, G. Galgóczi, L. Dobos, Z. Frei, I. S. Heng, R. Macas, C. Messenger, P. Raffai, and R. S. de Souza, Glade: A galaxy catalogue for multimessenger searches in the advanced gravitational-wave detector era, Monthly Notices of the Royal Astronomical Society **479**, 2374–2381 (2018).

[34] G. Dálya, R. Díaz, F. R. Bouchet, Z. Frei, J. Jasche, G. Lavaux, R. Macas, S. Mukherjee, M. Pálfi, R. S. de Souza, B. D. Wandelt, M. Bilicki, and P. Raffai, Glade+ : an extended galaxy catalogue for multimessenger searches with advanced gravitational-wave detectors, Monthly Notices of the Royal Astronomical Society **514**, 1403–1411 (2022).

[35] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, Astrophys. J. **622**, 759 (2005), arXiv:astro-ph/0409513.

[36] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski, healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python, Journal of Open Source Software **4**, 1298 (2019).

[37] R. Gray, C. Messenger, and J. Veitch, A pixelated approach to galaxy catalogue incomplete-

ness: improving the dark siren measurement of the hubble constant, Monthly Notices of the Royal Astronomical Society **512**, 1127 (2022).

[38] L. P. Singer, H.-Y. Chen, D. E. Holz, W. M. Farr, L. R. Price, V. Raymond, S. B. Cenko, N. Gehrels, J. Cannizzo, M. M. Kasliwal, S. Nissanke, M. Coughlin, B. Farr, A. L. Urban, S. Vitale, J. Veitch, P. Graff, C. P. L. Berry, S. Mohapatra, and I. Mandel, Supplement: "going the distance: Mapping host galaxies of ligo and virgo sources in three dimensions using local cosmography and targeted follow-up" (2016, apjl, 829, l15), The Astrophysical Journal Supplement Series **226**, 10 (2016).

[39] V. Alfradique *et al.*, A dark siren measurement of the hubble constant using gravitational wave events from the first three ligo/virgo observing runs and delve (2023), arXiv:2310.13695 [astro-ph.CO].

[40] L. P. Singer and L. R. Price, Rapid bayesian position reconstruction for gravitational-wave transients, Phys. Rev. D **93**, 024013 (2016).

[41] C. Dalang and T. Baker, The clustering of dark siren host galaxies (2023), arXiv:2310.08991 [astro-ph.CO].

[42] R. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA), Constraints on the Cosmic Expansion History from GWTC–3, Astrophys. J. **949**, 76 (2023), arXiv:2111.03604 [astro-ph.CO].