The Hitchhiker's Guide to the Galaxy Catalog Approach for Dark Siren Gravitational-wave Cosmology

Jonathan R. Gair[1], Archisman Ghosh[2], Rachel Gray[3], Daniel E. Holz[4], Simone Mastrogiovanni[5],
Suvodip Mukherjee[6], Antonella Palmese[7,8,32], Nicola Tamanini[9], Tessa Baker[4], Freija Beirnaert[2], Maciej Bilicki[10],
Hsin-Yu Chen[11,12], Gergely Dálya[2], Jose Maria Ezquiaga[13], Will M. Farr[14,15], Maya Fishbach[16],
Juan Garcia-Bellido[17], Tathagata Ghosh[18], Hsiang-Yu Huang[19], Christos Karathanasis[20], Konstantin Leyde[21],
Ignacio Magaña Hernandez[22], Johannes Noller[23,24], Gregoire Pierra[25], Peter Raffai[26], Antonio Enea Romano[27,28],
Monica Seglar-Arroyo[20], Danièle A. Steer[29,30], Cezary Turski[2], Maria Paola Vaccaro[20], and
Sergio Andrés Vallejo-Peña[27]

[1] Max Planck Institute for Gravitationsphysik (Albert Einstein Institute), Am Mühlenberg 1, D-14476 Potsdam, Germany; jgair@aei.mpg.de
[2] Department of Physics and Astronomy, Ghent University, Proeftuinstraat 86, B-9000 Ghent, Belgium; archisman.ghosh@ugent.be
[3] Department of Physics and Astronomy, Queen Mary University of London, Mile End Road, London, E1 4NS, UK; r.gray@qmul.ac.uk
[4] Department of Physics, Department of Astronomy & Astrophysics, The University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[5] INFN, Sezione di Roma, I-00185 Roma, Italy
[6] Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai, 400005, India; suvodip@tifr.res.in
[7] McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA; palmese@cmu.edu
[8] Department of Physics, University of California Berkeley, 366 LeConte Hall MC 7300, Berkeley, CA 94720, USA
[9] Laboratoire des 2 Infinis—Toulouse (L2IT-IN2P3), Université de Toulouse, CNRS, UPS, F-31062 Toulouse Cedex 9, France; nicola.tamanini@l2it.in2p3.fr
[10] Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668 Warsaw, Poland
[11] Department of Physics, The University of Texas at Austin, Robert A Welch Hall, 105 E 24th st, Austin, TX 78712, USA
[12] Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[13] Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark
[14] Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA
[15] Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA
[16] Canadian Institute for Theoretical Astrophysics, David A. Dunlap Department of Astronomy and Astrophysics, and Department of Physics, 60 St George Street, University of Toronto, Toronto, ON M5S 3H8, Canada
[17] Instituto de Física Teórica UAM/CSIC, Universidad Autónoma Madrid, E-28049 Madrid, Spain
[18] Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India
[19] Department of Physics, National Central University, No. 300, Zhongda Rd., Zhongli District, Taoyuan City, 320317, Taiwan
[20] Institut de Física d'Altes Energies (IFAE), Universitat Autònoma de Barcelona, E-08193, Barcelona, Spain
[21] Université Paris Cité, CNRS, Astroparticule et Cosmologie, F-75006 Paris, France
[22] University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA
[23] Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK
[24] DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK
[25] Université Claude Bernard Lyon 1, CNRS, Institut de physique des deux infinis, 4 rue Enrico Fermi, F-69100 Villeurbanne, France
[26] Institute of Physics, Eötvös Loránd University, 1117 Budapest, Hungary
[27] Instituto de Física, Universidad de Antioquia, A.A.1226, Medellín, Colombia
[28] Icranet, Piazza della Repubblica 10, I-65122 Pescara, Italy
[29] Université Paris Cité, CNRS, Astroparticule et Cosmologie, F-75013 Paris, France
[30] Laboratoire de Physique de l'École Normale Suéperieure, ENS, CNRS, Université PSL, Sorbonne Université, Université Paris Cité, F-75005 Paris, France
mastrosi@roma1.infn.it

Abstract

We outline the "dark siren" galaxy catalog method for cosmological inference using gravitational wave (GW) standard sirens, clarifying some common misconceptions in the implementation of this method. When a confident transient electromagnetic counterpart to a GW event is unavailable, the identification of a unique host galaxy is in general challenging. Instead, as originally proposed by Schutz, one can consult a galaxy catalog and implement a dark siren statistical approach incorporating all potential host galaxies within the localization volume. Trott & Huterer recently claimed that this approach results in a biased estimate of the Hubble constant, H_0, when implemented on mock data, even if optimistic assumptions are made. We demonstrate explicitly that, as previously shown by multiple independent groups, the dark siren statistical method leads to an unbiased posterior when the method is applied to the data correctly. We highlight common sources of error possible to make in the generation of mock data and implementation of the statistical framework, including the mismodeling of selection effects and inconsistent implementations of the Bayesian framework, which can lead to a spurious bias.

---

[32] NASA Einstein Fellow.

1. Introduction

The use of gravitational waves (GWs) from compact binary mergers as standard sirens (Schutz 1986; Holz & Hughes 2005; Dalal et al. 2006) for cosmology is an idea which has finally come to fruition in recent years. 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 detection of the binary neutron star (BNS) GW170817 (Abbott et al. 2017b) and its electromagnetic (EM) counterpart (Abbott et al. 2017c) by the LIGO (Aasi et al. 2015) and Virgo (Acernese et al. 2015) GW detectors—which allowed the host galaxy, and hence the redshift of the merger, to be identified—led to the first GW measurement of H_0 (Abbott et al. 2017a). In the absence of an EM counterpart, redshift information from other sources can be used, such as (i) galaxy catalogs, using the statistical method (Del Pozzo 2012; Chen et al. 2018; Fishbach et al. 2019; Gray et al. 2020, 2022; Leandro et al. 2022) or the cross-correlation method exploring the spatial clustering between GW sources and galaxies (Oguri 2016; Bera et al. 2020; Mukherjee et al. 2020, 2021; Diaz & Mukherjee 2022); and (ii) "spectral siren" inference of the redshift from the astrophysical distributions of the GW sources themselves (Chernoff & Finn 1993; Taylor et al. 2012; Farr et al. 2019; Ezquiaga & Holz 2021; Mastrogiovanni et al. 2021; Ezquiaga & Holz 2022; Karathanasis et al. 2022; Leyde et al. 2022; Mukherjee 2022).

This paper will focus on the *dark siren + galaxy catalog* method (also informally known as the *statistical* or *dark siren* method),[33] in which galaxy catalogs are used to provide the redshift information for all potential host galaxies within a GW's localization volume, which are statistically averaged over. While less informative than the counterpart method on an event-by-event basis, the constraint strengthens as more events are included in the analysis. Given the current detection rates of bright and dark sirens (the latter have a factor >10 more detections), this method is expected to make a significant contribution to the GW constraint of H_0 (Chen et al. 2018). Indeed, this approach has already been implemented using dozens of available GW detections (Fishbach et al. 2019; Abbott et al. 2021a; Soares-Santos et al. 2019; Palmese et al. 2020, 2023; Abbott et al. 2023b; Finke et al. 2021).

This paper aims to act as a point of introduction for those new to the field of GW cosmology, and the dark siren + galaxy catalog method in particular, using mock data to build up a basic analysis following a Bayesian approach. This is motivated by the recent claim that GW dark sirens generally provide biased estimates of H_0 even under simplified assumptions (Trott & Huterer 2021, hereafter TH21). We show here

that this is not the case, and how it is incorrect modeling assumptions that lead to biased measurements. Given this, it is important to stress that a lot of the inconsistencies and biases explored in this work are not relevant for the case of realistic data, and instead primarily arise due to simplified assumptions and toy models for the GW sources and galaxy catalogs.

It is well understood that biases cannot arise in statistical analysis when the model and data-generating process are consistent. In simulation studies, it is possible to control the data-generation process, and so if biases appear these must be due to errors or inconsistencies when generating or analyzing the data. This should be used as a diagnostic tool to track down those errors and inconsistencies. This does not mean that analyses of real data are free of biases, since the true data-generating process for any observed process is unknowable. However, the true parameters are also not then known and so biases are hard to diagnose. It is certainly valuable, within a simulation study, to vary the assumptions about the data-generating process, while not changing the analysis, to understand what biases could appear in the analysis of real data. Indeed, this is the correct procedure for identifying potential sources of systematic error. However, this is only useful if a consistent analysis has first been shown to give unbiased results, and if the types of modifications to the data-generating process are controlled and physically well motivated. Any sources of bias identified in this way can then be mitigated by increasing the complexity of the analysis model.

This paper is organized as follows. Section 2 introduces the basic Bayesian framework, then describes the mathematics of how the mock GW and EM galaxy catalog data should be constructed. Section 3 gives details of the simulated data, then shows results for the scenario in which 200 GW detections are used to constrain H_0, and how the constraint changes in the limit of a large number of detections. Section 4 discusses some of the common mistakes made when applying the dark siren + galaxy catalog method, particularly when generating simplified mock data, which can lead to biased outcomes. It also highlights some of the real-world complications that need to be addressed in a true dark siren + galaxy catalog analysis, which are not in the scope of this paper. We note that this work is not intended to present a forecast of how well we will be able to constrain H_0, as we make some simplifying assumptions that are different from reality, but it is intended as a pedagogical work on the method.

2. Statistical Framework

Let us assume that we have observed a set of $N_{\rm obs}$ GW events, $\{x\}$, from which we are able to measure the luminosity distance of each source. In this basic example, H_0 can be determined by the fact that we measure the source luminosity distance from the GWs and can identify potential host galaxies from the catalog. According to Bayes' theorem, the posterior on H_0, given a set of detected GW events $\{x\}$, can be written as

$$p(H_0|\{x\}) \propto \mathcal{L}(\{x\}|H_0)p(H_0), \qquad (1)$$

where $p(H_0)$ is the prior on H_0. The likelihood $\mathcal{L}(\{x\}|H_0)$ can be described by an inhomogeneous Poisson process

---

[33] In general, any GW observed without an EM counterpart is a dark siren, and there is more than one method that can be used to produce cosmological measurements with dark sirens, so it is worth making the distinction. Similarly, all GW measurements of H_0 are statistical in that they require the information from many GW events to be combined in order to make a constraint.

(Mandel et al. 2019; Vitale et al. 2022):

$$\mathcal{L}(\{x\}|H_0) = \frac{e^{-N_{\exp}(H_0)}\,[N_{\exp}(H_0)]^{N_{\mathrm{obs}}}}{N_{\mathrm{obs}}!}$$
$$\times \prod_i^{N_{\mathrm{obs}}} \frac{\int dz \; \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z,H_0))\; p_{\mathrm{CBC}}(z)}{\int dz \; P_{\mathrm{det}}^{\mathrm{GW}}(z,H_0)\; p_{\mathrm{CBC}}(z)}. \quad (2)$$

Here, $\mathcal{L}_{\mathrm{GW}}(x_i|d_L(z,H_0))$ is the GW likelihood (the probability of measuring the data x_i given some luminosity distance d_L), $p_{\mathrm{CBC}}(z)$ is the probability that the source, a compact binary coalescence (CBC), is at redshift z for an observer on Earth, while $p_{\mathrm{det}}^{\mathrm{GW}}(z,H_0)$ is the GW detection probability used to account for selection biases. The term $N_{\exp}(H_0)$ is the number of expected GW detections for a given value of H_0. This term can be marginalized analytically by assuming a $1/N_{\exp}$ prior on the expected number of events (Fishbach et al. 2018), or equivalently the compact object binary merger rate. With this assumption, the pre-factor loses its dependence on H_0 and the likelihood for a single event reduces to

$$\mathcal{L}(x_i|H_0) = \frac{\int dz \, \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z,H_0))\, p_{\mathrm{CBC}}(z)}{\int dz \, P_{\mathrm{det}}^{\mathrm{GW}}(z,H_0)\, p_{\mathrm{CBC}}(z)}, \quad (3)$$

which is the one we use here. The quantity $p_{\mathrm{CBC}}(z)$ is being used here to represent the distribution of redshifts from which the GW source is drawn. If this is known, the likelihood, which is the probability distribution of observed data sets, can be found by marginalizing over the distribution of possible redshifts, which is what is done in Equation (3). In practice, we do not know $p_{\mathrm{CBC}}(z)$, but we construct it from observations of galaxies, as described in Section 2.1 below. The interpretation is then that we are using the *posterior* from EM observations as the *prior* for the GW data. This is perfectly legitimate but, to be rigorously correct, what the EM data provides is a distribution of possible GW source distributions, since the EM data is not perfect. In practice, this means that we should treat the unknown true redshifts of the galaxies as population-level parameters that we infer jointly from the EM and GW data, and marginalize over these after combining the individual GW measurements. In simpler terms, we should change the order of the product over events and the integrals over redshift in Equation (2) and the integral over redshift should be over the true redshifts of the galaxies that are possible hosts of the events. However, as we discuss in more detail in Section 2.3, it can be shown that in the limit in which the redshifts of the galaxies are perfectly known or the number density of CBCs is much lower than the number density of galaxies the hierarchical likelihood for H_0 can be reduced to Equation (3). In particular, the second hypothesis is perfectly reasonable considering that currently the rate of CBC mergers is estimated to be $\sim 10^{-6}$–$10^{-5}\,\mathrm{yr}^{-1}$ per galaxy (Abbott et al. 2021b). For this reason, Equation (3) is used in most current analyses, and this is perfectly legitimate. Where this distinction can matter, however, is in a mock-data scenario where the number of CBCs is artificially inflated (or the density of galaxies artificially reduced). For most of this paper, we will limit our discussion to the case in which the number density of

CBCs is much lower than the number density of galaxies. In Sections 2.3 and 3.3 only we will demonstrate when this assumption has an impact and how this can be mitigated.

As a final remark, it is important to note that the form of Equation (3) is based on the usual assumption that "detection" is a property of the observed data only, not the true parameters of the source. It can be convenient to simulate data by selecting events based on the true source parameters, but this is a modification to the *prior* rather than to the *likelihood* and would be inconsistent with the detection model assumed in Equation (3). A consistent analysis in this case would just retain the numerator of Equation (3), but with the prior $p_{\mathrm{CBC}}(z)$ renormalized by dividing by its integral over the range of events that are detectable (which will be dependent on the cosmological parameters) and the integral in the numerator truncated to the same range. While this correction should eliminate biases, we would not advocate this approach as it does not reflect the reality of an analysis of real data and might therefore give misleading results.

2.1. Galaxy Catalog Modeling

In this section, we describe how to build the probability, $p_{\mathrm{CBC}}(z)$, of a CBC occurring at redshift z, under the assumption that CBCs occur in galaxies and so will trace the distribution of galaxies in the universe in some fashion. As described above, we will also implicitly assume that the number density of CBCs (constrained to be 18–44 $\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}$ by ground-based observations to date; Abbott et al. 2023a) is much smaller than the number density of galaxies (bright galaxies have a local comoving density of $\sim 10^8\,h^{-3}\,\mathrm{Gpc}^{-3}$; Hahn et al. 2023) and so each CBC can be treated as an independent draw from the redshift distribution inferred from the galaxy catalog. Note that in this section (and for the rest of the paper) we will discuss the distribution of galaxies purely as a function of redshift, and thus neglect their spatial distribution in R.A. and decl. This closely follows the assumptions and approximations made in TH21. Neglecting this aspect does not automatically introduce problems, but it is an unrealistic setup that increases the risk mentioned above about relative number densities of CBCs and galaxies.

The probability that a merger will occur at a redshift z is the product of the probability that *there is a galaxy at z*, $p_{\mathrm{cat}}(z)$, and the probability of a galaxy at redshift z *hosting a GW merger*, $p_{\mathrm{rate}}(z)$:

$$p_{\mathrm{CBC}}(z) = \frac{p_{\mathrm{rate}}(z)p_{\mathrm{cat}}(z)}{\int_0^\infty p_{\mathrm{rate}}(z)p_{\mathrm{cat}}(z)dz}. \quad (4)$$

The rate—or signal emission—probability is given by

$$p_{\mathrm{rate}}(z) = \frac{\frac{R(z)}{1+z}}{\int_0^\infty \frac{R(z)}{1+z}dz}, \quad (5)$$

where $R(z)$ is the merger rate of GWs in their source frame, usually expressed in cubic gigaparsecs per year. The merger rate is usually parameterized as $R(z) \propto (1+z)^\gamma$ in the redshift region $0 < z < 2$ (Fishbach & Holz 2017) to account for a possible evolution of the merger rate. If GW mergers are uniform in comoving volume and source-frame time, this term is constant. The $1/1+z$ factor accounts for the effect of time dilation due to the expansion of the universe between the

source frame and the detector frame. In absence of any galaxy catalog observation, $p_{cat}(z)$ can be constructed with a uniform comoving volume distribution and

$$p_{CBC}(z) = \frac{\frac{R(z)}{1+z}\frac{dV_c}{dz}}{\int \frac{R(z)}{1+z}\frac{dV_c}{dz}dz}, \tag{6}$$

where

$$\frac{dV_c}{dz} = 4\pi \left[\frac{c}{H_0}\right]^3 \left[\int_0^z \frac{dz'}{E(z')}\right]^2. \tag{7}$$

Note that the above prior does not depend on H_0, but it depends on other cosmological parameters through the expansion history ($H(z)/H_0 \equiv E(z) = [\Omega_m(1+z)^3 + \Omega_\Lambda]^{1/2}$ for a flat Lambda cold dark matter, ΛCDM, cosmological model) and the parameterization of the rate term. This is the prior used for analyses making use of mass information (Chernoff & Finn 1993; Taylor et al. 2012; Farr et al. 2019; Mastrogiovanni et al. 2021; Ezquiaga & Holz 2022; Karathanasis et al. 2022; Leyde et al. 2022; Mukherjee 2022), where cosmology and $p_{CBC}(z)$ are fit jointly. Note also that if the CBC rate presents some features in redshift, such as peaks, it might help to measure cosmology even in the absence of counterparts or galaxy information (Ye & Fishbach 2021).

Physically, $p_{CBC}(z)$ is something like the redshift distribution of galaxies that have sourced a compact merger, the signal of which has passed through the Earth. It is interesting to note that Equation (6) can be also defined from a more "physical" argument starting from the rate of CBC mergers seen from an observer of Earth. This quantity can be expressed as

$$\frac{dN_{CBC}}{dzdt_d} = \frac{dN_{CBC}}{dV_cdt_s}\frac{dV_c}{dz}\frac{dt_s}{dt_d}$$
$$= R(z)\frac{dV_c}{dz}\frac{1}{1+z}, \tag{8}$$

where t_d and t_s are the times in the detector and source frame, and the number of CBCs per comoving volume per source-frame time is by definition the CBC merger rate.

For this basic mock data, when calculating Equation (4) we will neglect the rate term in Equation (5) in order to align more closely with the approach taken in TH21. As long as this rate assumption is treated self-consistently when generating the mock data and analyzing it, this will not introduce any bias to the results. Let us note that this is a simplified description where the CBC rates only depend on the universe epoch (redshift); in a more general case, we might even model that more luminous galaxies are more likely to host CBCs. Therefore, in the remaining of the paper, we will approximate

$$p_{CBC}(z) \approx p_{cat}(z). \tag{9}$$

Moreover, here we want to exploit the fact that we are provided with a galaxy survey. We want to build $p_{cat}(z)$ given the observation of N_{gal} galaxies with *measured redshifts* $\{\hat{z}_g\}$. In other words, now we are computing $p_{cat}(z|\{\hat{z}_g\})$. In this computation, we will assume that the galaxy catalog is complete.[34] For details on how galaxy catalog incompleteness can be incorporated into such an analysis, see, for example,

Chen et al. (2018), Fishbach et al. (2019), and Gray et al. (2020, 2022). While constructing $p_{cat}(z|\{\hat{z}_g\})$, we must consider that $\{\hat{z}_g\}$ are not the *true* redshifts, $\{z_g\}$. We can take into account this uncertainty using the laws of probabilities, namely

$$p_{cat}(z|\{\hat{z}_g\}) = \int d\{z_g\} p_{gal}(z|\{z_g\}) p_{red}(\{z_g\}|\{\hat{z}_g\}), \tag{10}$$

where $p_{gal}(z|\{z_g\})$ is the probability to have a galaxy at redshift z when we have a set of true redshifts for the galaxies. This is simply given by

$$p_{gal}(z|\{z_g\}) = \frac{1}{N_{gal}}\sum_i^{N_{gal}} \delta(z - z_g^i), \tag{11}$$

where δ is the Dirac delta distribution. The probability $p_{red}(\{z_g\}|\{\hat{z}_g\})$ encodes the fact that we are not perfectly able to measure the true redshift of the galaxies, but we have a *measured value*. We label this probability with "red" to indicate that it refers to the measurement uncertainties due to redshift. Since each galaxy measure is independent of the others,

$$p_{red}(\{z_g\}|\{\hat{z}_g\}) = \prod_j p_{red}(z_g^j|\hat{z}_g^j). \tag{12}$$

By plugging Equations (11)–(12) in Equation (10), we obtain that

$$p_{CBC}(z) \approx p_{cat}(z|\{\hat{z}_g\}) = \frac{1}{N_{gal}}\sum_i^{N_{gal}} p_{red}(z|\hat{z}_g^i). \tag{13}$$

At this point, we can immediately note that if the galaxy redshifts are perfectly measured, Equation (13) reduces to a sum of Dirac delta functions:

$$p_{CBC}(z) \approx p_{cat}(z|\{\hat{z}_g\}) = \frac{1}{N_{gal}}\sum_i^{N_{gal}} \delta(z - \hat{z}_g^i), \tag{14}$$

and the likelihood in Equation (3) can be expressed analytically as

$$\mathcal{L}(x_i|H_0) = \frac{\sum_i^{N_{gal}} \mathcal{L}_{GW}(x_i|d_L(\hat{z}_g^i, H_0))}{\sum_i^{N_{gal}} P_{det}^{GW}(\hat{z}_g^i, H_0)}. \tag{15}$$

In Section 2.3, we show that the hierarchical likelihood still reduces to this equation even if we relax the assumption that the CBC density is lower than the galaxy density.

However, we are usually in a situation where the redshift of the galaxies is measured with large uncertainties. The term $p_{CBC}(z)$ is our *prior* on the CBC redshift distribution. In order for it to be used as such, the sum in Equation (13) must be over the redshift posteriors of individual galaxies. In this case, we will construct the likelihood of the individual galaxies in the same manner as was used to construct the GW likelihood, but an additional prior must be applied to convert this to a posterior. That is,

$$p_{red}(z|\hat{z}_g^i) = \frac{\mathcal{L}_{red}(\hat{z}_g^i|z)p_{bg}(z)}{\int \mathcal{L}_{red}(\hat{z}_g^i|z)p_{bg}(z)dz}, \tag{16}$$

where $\mathcal{L}_{red}(\hat{z}_g^i|z)$ is the likelihood model used to generate *observed* redshifts from the true redshift of each galaxy, and $p_{bg}(z)$ is a chosen prior on redshift that reflects our best belief

---

[34] "Complete" is defined here to mean that the catalog contains all galaxies that can emit CBCs in the universe.

on the background distribution of galaxies. We adopt a Gaussian redshift likelihood model like in TH21:

$$\mathcal{L}(\hat{z}_g^i|z) = \frac{1}{\sqrt{2\pi}\,\sigma_z}e^{-\frac{(\hat{z}_g^i-z)^2}{2\sigma_z^2}}, \qquad (17)$$

with $\sigma_z = 0.013(1+z)^3 \leqslant 0.015$. Note that the redshift uncertainty explicitly depends on the *true* redshift, and this has to be taken into account in the modeling.

Regarding priors, in the absence of any special knowledge about the galaxy distribution, the simplest choice we can make is to set $p_{\rm bg}(z)$ to be uniform in comoving volume. This can appear as "double counting," but it is the more conservative choice in absence of any data-driven information since we would expect galaxies to be nearly uniformly distributed in comoving volume. In the limit that $\sigma_z \to \infty$, i.e., we detect galaxies but we do not measure their redshifts, $p_{\rm CBC}(z)$ will then represent galaxies uniform in comoving volume. While in the limit that $\sigma_z \to 0$, the prior choice $p_{\rm bg}(z)$ will not matter and $p_{\rm CBC}(z)$ will be given by the sum of δ-like peaks located at the measured redshift of galaxies.

2.2. Gravitational-wave Data Modeling

The problems of detection and source parameters estimation of GW signals are complex and an active field of current research; we refer the reader to Abbott et al. (2020) for a more in-depth discussion. Here, we just discuss the fundamental aspects of detection and parameter estimation for GW signals.

GW signals are detected using low-latency algorithms, able to calculate signal-to-noise ratios and false-alarm rates using either a template bank (Messick et al. 2017; Nitz et al. 2017; Aubin et al. 2021; Cannon et al. 2021; Dal Canton et al. 2014) or a superposition of wavelets (Klimenko & Mitselmakher 2004). Typically, GW signals with false-alarm rates lower than 1–10 per year are selected for population analyses (Abbott et al. 2021, 2023a). The choice of a detection threshold is important to correct for selection biases. Once a signal is classified as detected, the source's physical parameters are estimated using a Bayesian sampling of the GW likelihood:

$$\mathcal{L}(x_i|\boldsymbol{\theta}) \propto e^{-\frac{1}{2}(x_i-h(\boldsymbol{\theta})|x_i-h(\boldsymbol{\theta}))}, \qquad (18)$$

where $\boldsymbol{\theta}$ is the set of GW source parameters, h is the GW signal, and x_i indicates the GW strain data. The scalar product is given in Fourier frequency space by

$$(a|b) = 4\,\mathrm{Re}\left[\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S(f)}\right], \qquad (19)$$

where $S(f)$ is the one-sided power spectral density of the noise.

For this analysis, we do not use the full GW likelihood, and instead use a toy model similar to the one used in TH21. The GW likelihood in Equation (3) is a central quantity of the mock study. The GW likelihood provides a model to generate error budgets on the estimation of the luminosity distance and it is also used to define the detection probability:

$$P_{\rm det}^{\rm GW}(z, H_0) = \int_{\rm detectable} \mathcal{L}_{\rm GW}(x|d_L(z,H_0))dx, \qquad (20)$$

where the integral is done over all the possible detectable GW signals. In principle, the GW likelihood and detection probability should take into account all the GW parameters, but in this simplified mock study we will just model it as a function of d_L. We simplify the GW likelihood by defining an "observed" luminosity distance, $\hat{d}_L^i$, which replaces the "observed data" x_i. We use a likelihood model:

$$\mathcal{L}_{\rm GW}(\hat{d}_L^i|d_L(z, H_0)) = \frac{1}{\sqrt{2\pi}\,\sigma_{d_L}}e^{-\frac{(\hat{d}_L^i-d_L(z,H_0))^2}{2\sigma_{d_L}^2}}, \qquad (21)$$

where $\sigma_{d_L} = Ad_L(z, H_0)$, with A a constant fractional error. Note that Equation (21) is a probability density function of the "measured" luminosity distance, $\hat{d}_L^i$. In other words, given a value of the true luminosity distance d_L, the probability of obtaining a certain measured value $\hat{d}_L^i$ is distributed according to a normal distribution. However, we note that the reconstruction of the "true" luminosity distance d_L given an observed value $\hat{d}_L^i$ is not Gaussian.[35] We also note that, while a simple likelihood model for d_L is sometimes assumed in GW analyses (e.g., Palmese et al. 2021) using the `bayestar` (Singer & Price 2016) sky maps, in general, this likelihood will take more complicated forms.

Regarding the detection process, we assume that GWs are detectable if their *measured* luminosity distance is smaller than a threshold of $\hat{d}_L < \hat{d}_L^{\rm thr} = 1550$ Mpc. The GW detection probability in Equation (20) can be written using the likelihood model in Equation (21) as

$$\begin{aligned} P_{\rm det}^{\rm GW}(z, H_0) &= \int_{-\infty}^\infty \Theta(\hat{d}_L; \hat{d}_L^{\rm thr})\mathcal{L}_{\rm GW}(\hat{d}_L|d_L(z,H_0))d\hat{d}_L \\ &= \int_{-\infty}^{\hat{d}_L^{\rm thr}} \mathcal{L}_{\rm GW}(\hat{d}_L|d_L(z,H_0))d\hat{d}_L \\ &= \frac{1}{2}\left[1 + \mathrm{erf}\left[\frac{\hat{d}_L^{\rm thr} - d_L(z,H_0)}{\sqrt{2}Ad_L(z,H_0)}\right]\right]. \end{aligned} \qquad (22)$$

In the above equations, Θ is a Heaviside step function of $\hat{d}_L$, dropping to 0 for $\hat{d}_L > \hat{d}_L^{\rm thr}$, and "erf" is the unilateral error function of a standardized normal distribution. In TH21, it is argued that $P_{\rm det}^{\rm GW}(z, H_0)$ is a Heaviside step function dropping to 0 at $\hat{d}_L^{\rm thr}$. This is correct only in the limit that $\sigma_{d_L} \to 0$ (for small error budgets).

In Figure 1, we plot the detection probability as a function of the true luminosity distance of the GW. When the GW likelihood has a nonnegligible error, the GW detection probability does not drop sharply to zero. This is expected, as the scattering process of d_L makes us able to detect sources whose true luminosity distance is above the detection threshold. The detection probability reduces to a Heaviside step function only when the distance error goes to 0 (i.e., when we are perfectly able to measure luminosity distance).

---

[35] This is due to the fact that there is a d_L dependence in the denominator of Equation (21) and

$$p(d_L^i|\hat{d}_L^i) = \frac{\mathcal{L}(\hat{d}_L^i|d_L^i)\Pi(d_L^i)}{\int \mathcal{L}(\hat{d}_L^i|d_L^i)\Pi(d_L^i)dd_L^i},$$
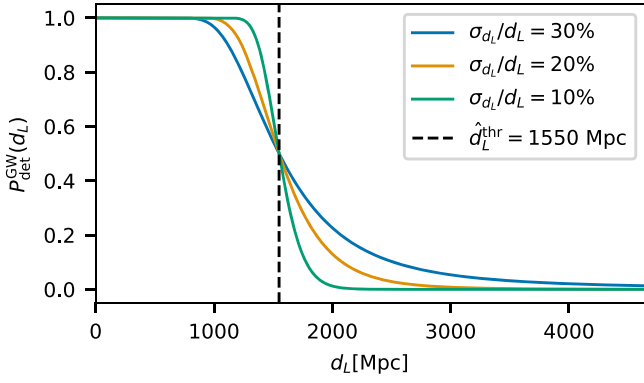
where Π is a prior term.

Figure 1. GW detection probability computed with the toy likelihood model in Equation (22) as a function of the *true* luminosity distance. The colors indicate various fractional errors on the luminosity distance. The vertical dashed line indicates the detection threshold for the *measured* luminosity distance, $\hat{d}_L$. We note that the detection probability only reduces to a Heaviside step function in the limit of $\sigma_{d_L} \to 0$. If $\sigma_{d_L} \neq 0$, a step function on *true* luminosity distance cannot be used to account for selection effects, else the H_0 posterior may be biased in a way that depends on σ_{d_L}.

2.3. Full Likelihood Derivation

As described above, when writing down Equation (3) we have effectively assumed that the GW likelihood depends on a distribution $p_{\mathrm{CBC}}(z)$, which is known. We then proceeded to construct $p_{\mathrm{CBC}}(z)$ from EM observations, $\{\hat{z}_g\}$, of galaxies in the catalog via Equation (13). Using the EM data as a prior for the GW data is perfectly legitimate, but the usual form of the GW likelihood, Equation (3) is then just an approximation. The GW likelihood actually depends on the true, and unknown, redshifts of the galaxies, $\{z_g\}$. In the absence of selection effects, this distinction does not matter for a single event, which we will now demonstrate. As written earlier in Equation (2), the GW likelihood for a single observation, x_i, depends only on the possible luminosity distances that the CBC could have. Previously, we wrote this down as a continuous distribution over redshifts, but if we instead assume that the CBC comes, with equal probability, from one of a set of N_{gal} galaxies with redshifts $\{z_g^j\}$, the GW likelihood can be written as

$$\mathcal{L}_{\mathrm{GW}}(x_i|\{z_g\}, H_0) = \frac{1}{N_{\mathrm{gal}}} \sum_{j=1}^{N_{\mathrm{gal}}} \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z_g^j, H_0)), \quad (23)$$

the likelihood for the EM observations is

$$\mathcal{L}_{\mathrm{EM}}(\{\hat{z}_g\}|\{z_g\}, H_0) = \prod_{j=1}^{N_{\mathrm{gal}}} \mathcal{L}_{\mathrm{red}}(\hat{z}_g^j|z_g^j), \quad (24)$$

and the prior on the redshifts of all the galaxies in the set of possible hosts (ignoring clustering) is

$$p(\{z_g\}) = \prod_{j=1}^{N_{\mathrm{gal}}} p_{\mathrm{bg}}(z_g^j), \quad (25)$$

which we assume to be independent of the prior on H_0, so the joint prior on the population-level parameters is $p(\{z_g\}$,

$H_0) = p(\{z_g\})p(H_0)$. The joint likelihood for the EM observations comprising the catalog and the GW data is

$$
\begin{aligned}
p(x_i, \{\hat{z}_g\}|H_0, \{z_g\}) &= \mathcal{L}_{\mathrm{GW}}(x_i|\{z_g\}, H_0) \\
&\times \mathcal{L}_{\mathrm{EM}}(\{\hat{z}_g\}|\{z_g\}, H_0), \quad (26)
\end{aligned}
$$

from which we obtain the posterior on the population parameters, $(H_0, \{z_g\})$, via Bayes' theorem:

$$
\begin{aligned}
p(H_0, \{z_g\}|x_i, \{\hat{z}_g\}) &\propto \frac{1}{N_{\mathrm{gal}}} \left[\sum_{j=1}^{N_{\mathrm{gal}}} \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z_g^j, H_0)) \right] \\
&\times \prod_{j=1}^{N_{\mathrm{gal}}} [\mathcal{L}_{\mathrm{red}}(\hat{z}_g^j|z_g^j) p_{\mathrm{bg}}(z_g^j)]. \quad (27)
\end{aligned}
$$

We can now integrate out the unknown true galaxy redshifts, $\{z_g\}$. For each term in the sum over galaxies entering the GW likelihood there is precisely one of these integrals that is over the corresponding galaxy redshift. The other integrals are independent of the GW data and reduce to the evidence for the EM observation of that galaxy:

$$\mathcal{Z}_j = \int \mathcal{L}_{\mathrm{red}}(\hat{z}_g^j|z_g^j) p_{\mathrm{bg}}(z_g^j) dz_g^j. \quad (28)$$

We deduce that the marginalized posterior takes the form

$$
\begin{aligned}
p(H_0|x_i, \{\hat{z}\}) &\propto \frac{1}{N_{\mathrm{gal}}} \left(\sum_{j=1}^{N_{\mathrm{gal}}} \left[\prod_{k \neq j} \mathcal{Z}_k \right] \right. \\
&\times \left. \left(\int \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z_g^j, H_0)) \mathcal{L}_{\mathrm{red}}(\hat{z}_g^j|z_g^j) p_{\mathrm{bg}}(z_g^j) \, dz_g^j \right) \right) \\
&= \frac{1}{N_{\mathrm{gal}}} \left[\prod_{k=1}^{N_{\mathrm{gal}}} \mathcal{Z}_k \right] \sum_{j=1}^{N_{\mathrm{gal}}} \\
&\times \left(\int \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z_g^j, H_0)) \frac{\mathcal{L}_{\mathrm{red}}(\hat{z}_g^j|z_g^j) p_{\mathrm{bg}}(z_g^j)}{\mathcal{Z}_j} \, dz_g^j \right) \\
&= \frac{1}{N_{\mathrm{gal}}} \left[\prod_{k=1}^{N_{\mathrm{gal}}} \mathcal{Z}_k \right] \sum_{j=1}^{N_{\mathrm{gal}}} \left(\int \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z_g^j, H_0)) p_{\mathrm{red}}(z_g^j|\hat{z}_g^j) \, dz_g^j \right) \\
&\propto \int \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z, H_0)) \left[\frac{1}{N_{\mathrm{gal}}} \sum_{j=1}^{N_{\mathrm{gal}}} p_{\mathrm{red}}(z|\hat{z}_j) \right] dz \\
&= \int \mathcal{L}_{\mathrm{GW}}(x_i|d_L(z, H_0)) p_{\mathrm{cat}}(z) \, dz,
\end{aligned}
$$
$$(29)$$

and so we recover the result derived earlier.[36] However, this alternative way of deriving the posterior reveals two important corrections that are in principle present but negligible in practice.

---

[36] This is not exactly the earlier result since it depends on $p_{\mathrm{cat}}(z)$ rather than $p_{\mathrm{CBC}}(z)$. This is because when writing down Equation (23) we assumed all galaxies were equally likely to host CBCs. In this limit $p_{\mathrm{cat}}(z) = p_{\mathrm{CBC}}(z)$. We can easily modify this assumption by including additional weights, $p_{\mathrm{rate}}(z_g^j)$, in each term in Equation (23) and replacing N_{gal} by $\sum p_{\mathrm{rate}}(z_g^j)$ in the pre-factor.

First, we have ignored GW selection effects in the above. These are straightforward to include by replacing $\mathcal{L}_{GW}$ by $\mathcal{L}_{GW}/p_{det}(H_0, \{z_g\})$. The detection probability, $p_{det}(H_0, \{z_g\})$, is the integral of the GW likelihood over data sets deemed above the threshold and hence included in the analysis, i.e.,

$$p_{det}(H_0, \{z_g\}) = \frac{1}{N_{gal}} \sum_{j=1}^{N_{gal}} \int_{-\infty}^{d_L^{thr}} \mathcal{L}_{GW}(x|d_L(z_g^j, H_0)) dx$$
$$= \frac{1}{N_{gal}} \sum_{j=1}^{N_{gal}} p_{det}^{GW}(d_L(z_g^j, H_0)). \quad (30)$$

However, this is a function of the *true* values of the galaxy redshifts. This dependence of the denominator of the GW likelihood on $\{z_g\}$ breaks the separability of the integrals that we exploited above, unless the true redshifts of the galaxies are perfectly known. In this case, it can be seen that the full hierarchical likelihood reduces to Equation (15). The same thing happens if we include rate weighting in Equation (23), in the way discussed in the earlier footnote. In practice, we will not know the true redshifts of the galaxies. The detection probability (and any rate weighting) are effectively an average of the galaxy redshift distribution over the whole volume within which GW sources can be observed. If that volume is sufficiently large, as is the case for current GW detectors, then the average galaxy redshift will be approximately uniform in comoving volume, and so the dependence of the detection probability (or rate weighting) on the actual galaxy redshifts will be relatively weak and so this term can be factored out, reducing the result to the simpler expression used in this paper.

The second correction arises when considering multiple GW observations. The likelihood is the product of likelihoods for each GW observation, but each one of those likelihoods includes the sum over galaxies. This will introduce cross terms which are not simply the product of the individually marginalized likelihoods but marginals of the product of the likelihoods over the true redshift of the galaxies. These terms represent corrections for the case when multiple GW events are observed from the same galaxy. In practice, these corrections are small, as there are typically $1/N_{gal}$ times fewer of them than the dominant independent-host terms. These terms will only become important once a significant number of sources in the catalog share a host. With typical merger rates of a few tens per cubic gigaparsec per year and approximately one galaxy per cubic megaparsec, the typical spacing of mergers in any given galaxy is millions of years, so we will essentially never be in a regime where these corrections matter.[37] Once again, we emphasize that these terms only arise when the galaxy redshift measurements are imperfect. When galaxy redshifts are known perfectly, Equation (15) is still exact when analyzing many GW sources.

To conclude this section, we note that the inconsistency described here does not represent a fundamental limitation of the galaxy catalog approach. The analysis can be done consistently by computing the posterior using the full

expression

$$p(H_0|\{x\}) \propto p(H_0) \int d\{z_g\} \left[\prod_{j=1}^{N_{obs}} \frac{\mathcal{L}_{GW}(x_j|\{z_g\}, H_0)}{p_{det}(H_0, \{z_g\})} \right]$$
$$\times \left[\prod_{k=1}^{N_{gal}} \mathcal{L}_{red}(\hat{z}_g^k|z_g^k) \right] p_{bg}(\{z_g\})$$

$$\mathcal{L}_{GW}(x_j|\{z_g\}, H_0) = \sum_{i=1}^{N_{gal}} p_{rate}(z_g^i) \mathcal{L}_{GW}(x_j|d_L(z_g^i, H_0)),$$
$$(31)$$

and $p_{det}(H_0, \{z_g\})$ is given by

$$p_{det}(H_0, \{z_g\}) = \sum_{j=1}^{N_{gal}} p_{rate}(z_g^j) \int_{-\infty}^{d_L^{thr}} \mathcal{L}_{GW}(x|d_L(z_g^j, H_0)) dx$$
$$= \sum_{j=1}^{N_{gal}} p_{rate}(z_g^j) p_{det}^{GW}(d_L(z_g^j, H_0)).$$
$$(32)$$

Note that we have omitted a factor of $1/\sum_{k=1}^{N_{gal}} p_{rate}(z_g^k)$ from both $\mathcal{L}_{GW}(x_j|\{z_g\}, H_0)$ in Equation (31) and $p_{det}(H_0, \{z_g\})$ in Equation (32) as this cancels. This expression involves multiplication of a sum over galaxies for each GW event, followed by an integration over all the unknown true galaxy redshifts. This is extremely computationally expensive, which is why it is better to use the standard approximate expression, especially since the latter is expected to be accurate in any application to real data. Nonetheless, to further illustrate this issue and its resolution, we will show a simplified example of the application of the full-likelihood framework in Section 3.3. As a final remark, we note that the rate dependence in the full likelihood can be handled by introducing a parameterized model, $p_{rate}(z|\lambda)$, and inferring the parameters λ simultaneously with H_0.

3. Results

In this section, we present results on H_0 using the statistical framework discussed in Section 2. We start by describing the simulation of mock data in Section 3.1, and in Section 3.2 we show results on H_0 using the statistical framework in the limit that the number density of galaxies is higher than the number density of GWs. In Section 3.3 we simulate a case where the number density of galaxies is lower than the number density of mergers.

3.1. Simulating the Mock Data

We now describe step-by-step how we create the mock data. For this mock study, we use galaxies taken from MICEcat v1.0,[38] the Grand Challenge (Fosalba et al. 2014, 2015; Carretero et al. 2015; Crocce et al. 2015; Hoffmann et al. 2015). The MICEcat simulation is a light-cone simulation covering one-eighth of the sky out to a redshift of 1.4, down to halo masses of $2.2 \times 10^{11} h^{-1} M_\odot$ with a total of about 205 million galaxies. The fiducial cosmological model in MICEcat v1.0 is a flat ΛCDM model with cosmological parameters $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.25$, $\Omega_b = 0.044$, $\Omega_\Lambda = 0.75$, n_s

---

[37] Space is big. Really big. You just will not believe how vastly hugely mind-bogglingly big it is.

[38] http://maia.ice.cat/mice/

= 0.95, and $\sigma_8 = 0.8$. We also choose this model for our simulations. We note that MICEcat was previously used in Fishbach et al. (2019), which showed explicitly that dark siren estimations of H_0 are unbiased. Below, we highlight how MICEcat is employed for this work.

When generating our mock data, we follow the method in TH21 as closely as possible. We take two sky directions, referred to as "Direction 1" and "Direction 2," and for each of these we select galaxies found in an opening angle from the line of sight (LOS) of $\theta = 1$ and 5 deg, leading to four different scenarios. To simplify the notation, we adopt the nomenclature D_{15} to indicate Direction 1 with an opening angle of 5 deg, and similarly for the other combinations of directions and opening angles. On top of this choice, to reduce the overall number of galaxies being considered, we also subsample the galaxies in Direction 2 by half.

To simulate GW events, given N_{gal} galaxies with a set of *true* redshifts $\{z_g\}$, we randomly choose N_{ev} of them in which to simulate a GW signal. We decide to draw GWs only from galaxies with true redshift below $z_{max} = 1.4$. This is reasonable because the luminosity distance at this redshift (~ 10 Gpc) is much larger than our threshold distance for the GW events. The probability of detecting a GW event from that distance, even when allowing for fluctuations due to the difference between measured and true luminosity distance, is negligible. For each GW signal, we compute the true luminosity distance, d_L^i, by converting the true redshift using the fiducial MICEcat cosmological parameters. Using the likelihood in Equation (21), we draw an observed value for the luminosity distance $\hat{d}_L^i$ following a Gaussian centered at d_L^i with $\sigma = A d_L^i$. If $\hat{d}_L^i$ is lower than a threshold of $\hat{d}_L^{thr} = 1550$ Mpc (chosen to approximately match the detectability threshold used in TH21), we label the signal as detected and it can be used for cosmological inference.

A crucial aspect of the simulation is the choice of z_{draw} in comparison to the choice of $\hat{d}_L^{thr}$. Following TH21, we might be tempted to simulate GW events only from galaxies with redshift below a z_{draw} value obtained converting $\hat{d}_L^{thr}$ to a redshift for given cosmological parameters (H_0 and Ω_m). This procedure would allow us to detect only GW sources with a *true* luminosity distance below the detection range, and this is inconsistent with the assumed framework to correct for selection biases. As we argued in Figure 1, we would expect to find a significant number of sources that are detected even if their *true* luminosity distance is higher than $\hat{d}_L^{thr}$ (the detection range). On the contrary, by selecting $z_{max} = 1.4$, we are able to simulate GWs at high redshift that become detectable due to fortuitous noise fluctuations. Note that we can always make the choice of injecting GW events in an arbitrarily reduced range, $z_1 < z < z_2$, which TH21 did for $z_1 = 0$ and $z_2 = 0.3$. However, we would then need to account for this choice, using a $p_{rate}(z) = 0$ if $z < z_1$ or $z > z_2$, that must be consistently implemented in the statistical framework. In our analysis, we account for the fact that $p_{rate}(z) = 0$ for $z > 1.4$, in order to avoid introducing any possible bias.

The final step of the simulation is to calculate $p_{CBC}(z)$ to use for the inference in the cases for which we assume that the galaxy redshifts are not perfectly known. To do so, we draw a set of observed redshifts $\{\hat{z}_g\}$ from the true ones $\{z_g\}$, according to the likelihood model in Equation (17). By using Equation (16), we build a posterior on redshift for every galaxy, which is later used to build the approximant for $p_{CBC}(z)$. To build the approximant, we use a uniform in comoving volume prior; note

that this prior does not depend on H_0, namely

$$p_{bg}(z) = \frac{\frac{dV_c}{dz}}{\int_0^\infty \frac{dV_c}{dz} dz}, \quad (33)$$

where the H_0 dependence cancels out, and it depends only on the value of $\Omega_m = 0.25$ considered in the MICEcat simulation.

In Figure 2, we show the reconstruction of the galaxy density profile as a function of redshift for the different LOSs. We can see that the constructed interpolant for p_{CBC} tracks the true galaxy density profile of the catalog.

3.2. H_0 Inference in the Low-galaxy Density Limit

Following TH21, we simulate 200 GW dark sirens for each of the four directions D_{11}, D_{15}, D_{21}, and D_{25} with three different scenarios for the error on the GW likelihood: $\sigma_{d_L}/d_L = 10\%$, 20%, and 30%. For this case, we will assume that we perfectly know the redshift of the galaxies. We will use the hierarchical likelihood in Equation (15), which we have shown to be formally correct even in the low-galaxy density limit, and which happens for D_{11} and D_{21}.

In Figure 3, we show the H_0 posteriors that we obtain for each LOS and for each error budget of the GW likelihood. From the plot, we can see that the LOSs with more galaxies correspond to less constraining power on H_0. This is expected, as the GW localization volume will include a larger number of galaxies, and the effect of structures in the galaxy distribution (which can help to provide useful redshift information) is reduced. Interestingly, we note that for the cases with $\sigma_{d_L}/d_L = 30\%$ it is possible to obtain a posterior which is slightly more informative than the case of $\sigma_{d_L}/d_L = 20\%$. This might be counterintuitive, but it is consistent with the fact that for the case of $\sigma_{d_L}/d_L = 30\%$ some GW events might be generated at the redshift edge of our simulation $z_{draw} = 1.4$. This effectively introduces another redshift scale (an edge) to the inference, which then can help in inferring H_0. This effect can be visualized from the GW likelihood. In Figure 4, we plot the GW likelihood as function of redshift for $H_0 = 140$ km s^{-1} Mpc^{-1} for a signal detected at d_{thr} with $\sigma_{d_L}/d_L = 20\%$ and $\sigma_{d_L}/d_L = 30\%$. We can see that the GW likelihood for $\sigma_{d_L}/d_L = 30\%$ has some nonnegligible support beyond $z_{draw} = 1.4$, which is excluded from our statistical analysis. This adds extra information and allows us quickly to exclude high values of H_0, as shown in Figure 3.

The posteriors in Figure 3 do not show obvious biases. The curve for Direction 2 with a 5° opening angle has a multimodal structure, but this is just the result of a statistical fluctuation in the sampled population. To show more quantitatively that the statistical method for dark sirens does not present any significant bias, we generate a probability–probability (PP) plot. PP plots are used to test that parameters subject to Bayesian inference (in our case H_0) are consistently recovered. To generate a PP plot, we repeat the H_0 inference for 200 GW signals 100 times, each time drawing an injected H_0 value between the explored prior range of [20, 140] km s^{-1} Mpc^{-1}. We then check in what credible intervals the true value for each simulation is found. If the analysis is performed properly, we expect to see that, for example, 40% of injections are found in the 40% credible intervals. Figure 5 shows the PP plot for our mock data with two LOSs and varying the error budget on d_L. As we can see from the plot, the Bayesian framework is suitable to perform the H_0 inference as in all cases we are diagonal within 3σ. In particular,
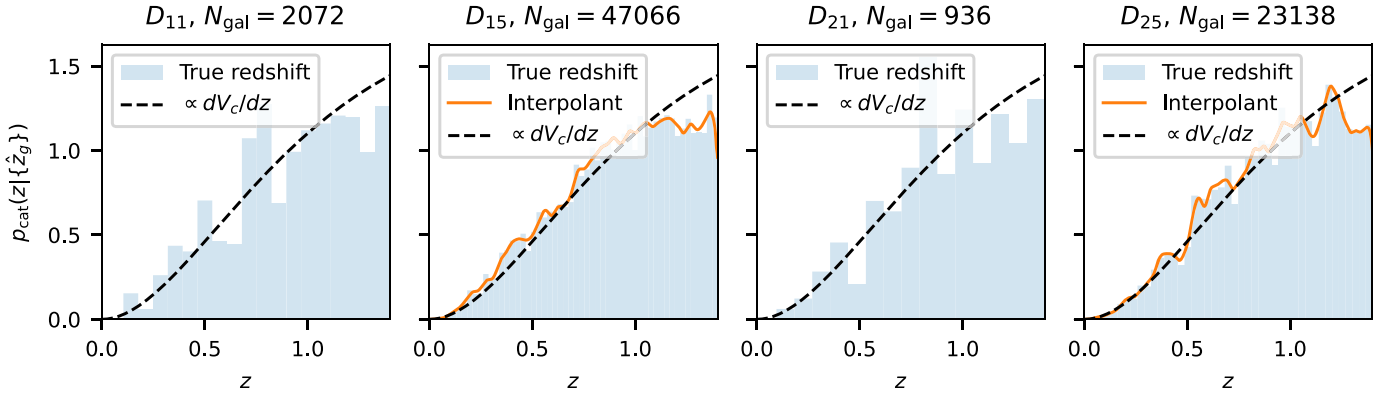
Figure 2. Reconstructions of the galaxy density profile $p_{cat}(z)$ for each of the populations considered in the mock data. The blue histograms indicate the true redshift of the galaxies, the orange lines are the reconstructions of the density profile using Equation (13), and the black dashed lines mark the uniform in comoving volume distributions. For the LOSs with small galaxy numbers, we do not fit the redshift interpolant as these are used only in the limit that we perfectly know the redshift (in order to not introduce a bias due to the breaking of the statistical framework).
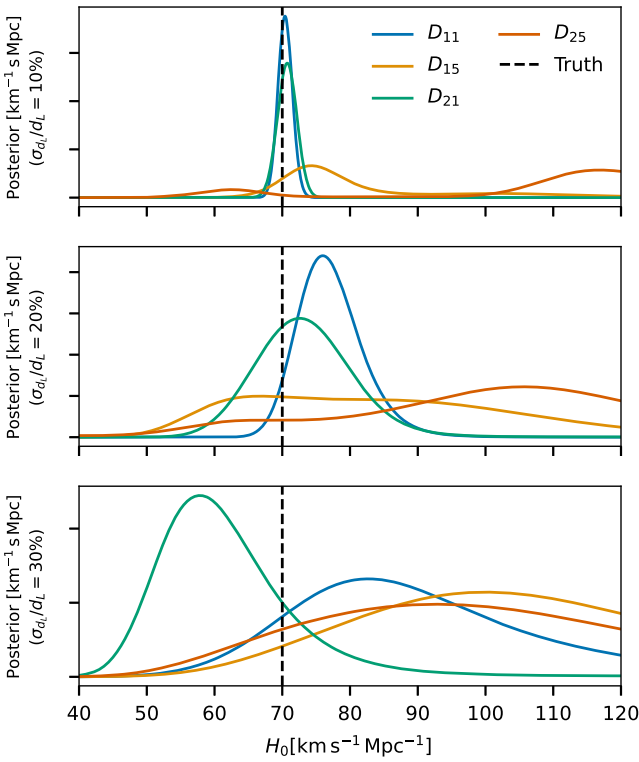
Figure 3. Posteriors of H_0 derived for four different LOSs (D_1 and D_2) in the MICEcat simulations, and for two different sky areas (1° and 5° opening angle for D_{1x} and D_{2x}, respectively) around those directions. The top, middle, and bottom panels assume a standard deviation on the luminosity distance that is 10%, 20%, and 30% of the luminosity distance, respectively. All the posteriors are the result of combining 200 GW events and assuming galaxy redshift is perfectly known. We do not find a bias. Note that any particular choice of the random seed used for the simulation will lead to fluctuations of the posteriors around the true value. However, by performing several simulations one should find that these fluctuations are compatible with the confidence levels at which the true value is found (see the discussion about PP plots in the text).

we recover a diagonal PP plot for D_{25}, which was the case that showed the second posterior mode in Figure 3.

Moreover, we have performed simulations also assuming errors on the galaxy redshifts. In the limit that we have $N_{GW} \geqslant N_{gal}$, the simplified framework cannot be used,[39] e.g., cases D_{11} and D_{21},

---

[39] This an extremely unrealistic limit, where we assume that multiple GW sources are hosted in the same galaxy.
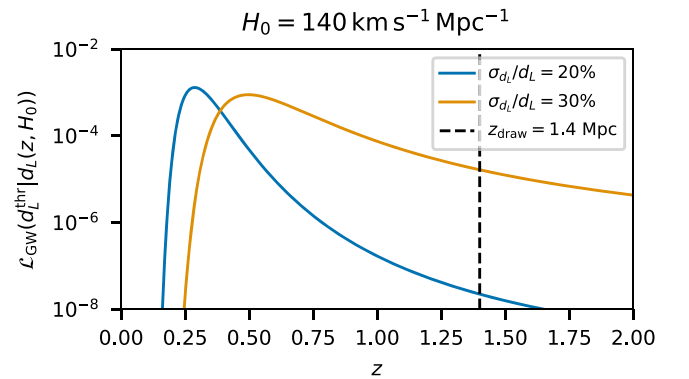
Figure 4. GW likelihood (vertical axis) as a function of the redshift (horizontal axis) for $H_0 = 140$ km s^{-1} Mpc^{-1}. The lines show the case of $\sigma_{d_L}/d_L = 20\%$ (blue line) and $\sigma_{d_L}/d_L = 30\%$ (orange line). The vertical dashed line indicates z_{draw}.
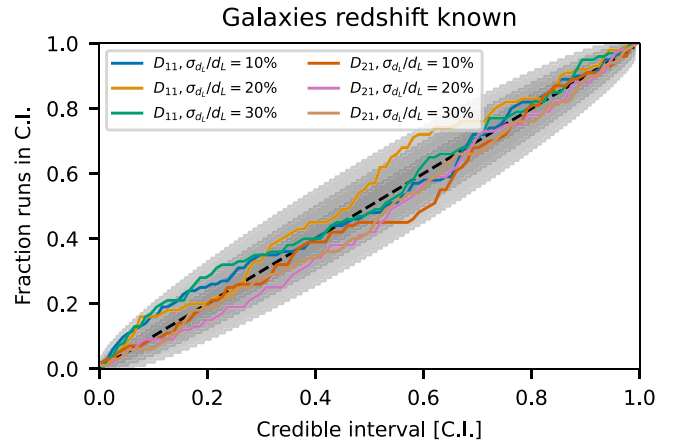
Figure 5. PP plot generated for the Direction 1 considered here and a nominal error on d_L of 20%–30%. The shaded area shows the 1, 2, and 3σ fluctuations of the PP plot around the ideal case, computed from a binomial distribution with n equal to the number of realizations and p equal to the credible interval level. Our simulations show that we are able to recover the input value of H_0 without incurring any significant bias.

otherwise there could be a bias; see Section 3.3 for more details. Therefore, we restrict the simulation to D_{15} and D_{25}. Figures 6 and 7 show a sample of H_0 posteriors and the PP plots generated from these cases. We can observe that the statistical dark sirens method is unbiased for this case.
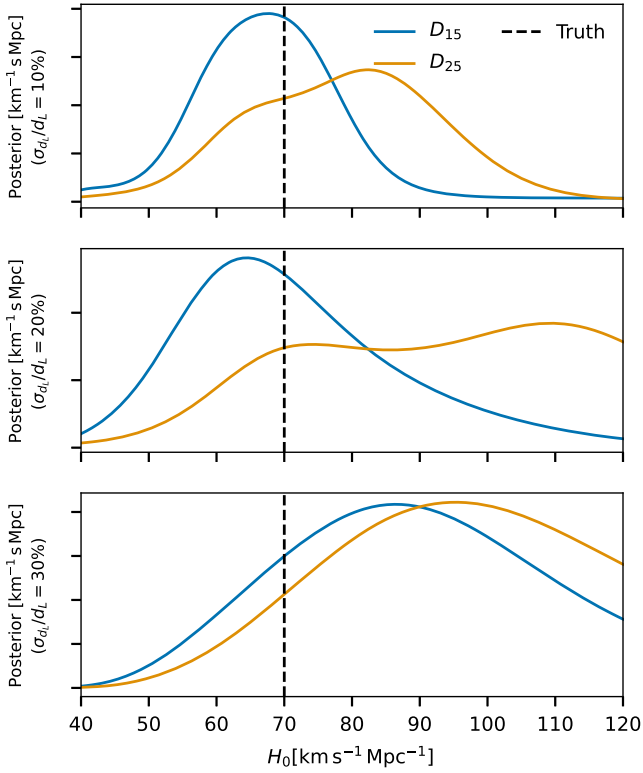
Figure 6. Posteriors of H_0 derived for four different LOSs (D_1 and D_2) in the MICEcat simulations, and for two different sky areas (1° and 5° opening angle for D_{1x} and D_{2x}, respectively) around those directions. The top, middle, and bottom panels assume a standard deviation on the luminosity distance that is 10%, 20%, and 30% of the luminosity distance, respectively. All the posteriors are the result of combining 200 GW events and assume errors on galaxy redshift. We do not notice any significant bias.

3.3. The One-galaxy Limit

As described in Section 2.3, the standard analysis makes some approximations that are only valid in the limit that the GW detection volume contains a sufficiently large number of galaxies. To illustrate this, we consider here an extreme and unrealistic example in which the galaxy catalog contains only one galaxy, the redshift of which is known imperfectly from EM observations. As in previous cases, we generate a catalog of 200 GW sources and construct the posterior on H_0 using the standard analysis that employs the approximate likelihood. In Figure 8, we show example posteriors for catalogs generated with $H_0 = 70$ km s^{-1} Mpc^{-1}, assuming the uncertainty in the EM redshift measurement is $\delta z/z = 3\%$ or 0.3%. We see that, with a 3% error, the posterior is shifted significantly to the right and shows a large bias. For the smaller redshift uncertainty the posterior appears to be unbiased, but in fact a small bias is still present on average. This is revealed by constructing a PP plot over 100 random realizations of the universe. The PP plot is shown in Figure 9, wherein there is a clear and significant departure from the diagonal, even when $\delta z/z = 0.3\%$.

These biases arise from the fact that in the standard analysis the approximation has been used that the true redshift of the galaxy can be marginalized out of the likelihood separately for each GW event. To carry out a consistent analysis, we must use the full likelihood given in Equation (31). In the case that there is only one galaxy, with unknown true redshift z and observed
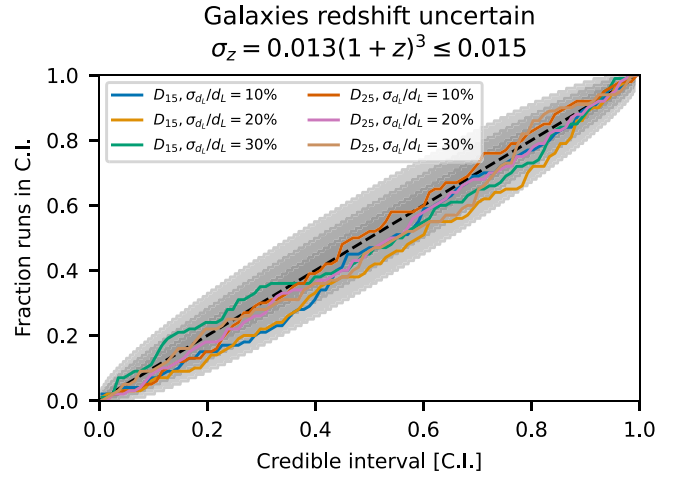
Figure 7. PP plot generated for directions 1 and 2 and a nominal error on d_L of 10%–30%. The shaded area shows the 1, 2, and 3σ fluctuations of the PP plot around the ideal case. Our simulations show that we are able to recover the input value of H_0 without incurring any significant bias.
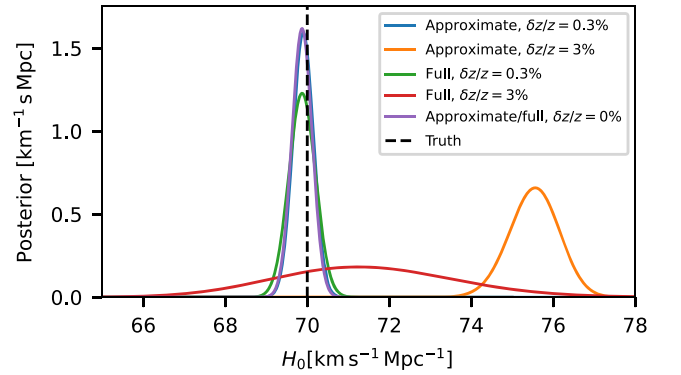
Figure 8. Comparison between posteriors obtained in the one-galaxy limit, with 200 observed GW sources, computed using the standard approximate likelihood given in Equation (3) or using the full likelihood given in Equation (34). We show results for fractional errors in the EM measurement of the galaxy redshift of $\delta z/z = 0.3\%$ and 3%, and also the perfect measurement limit $\delta z/z = 0\%$.

redshift $\hat{z}$, this simplifies to

$$p(H_0|\{x\}) \propto p(H_0) \int dz \left[\prod_{j=1}^{N_{obs}} \frac{\mathcal{L}_{GW}(x_j|d_L(z, H_0))}{p_{det}(H_0, z)} \right]$$
$$\times [\mathcal{L}_{red}(\hat{z}|z)] p(z). \quad (34)$$

The results from analyzing the same data sets with the full likelihood are also shown in Figures 8 and 9. We see that the full likelihood broadens the posterior, making it consistent with the true value even in the case of large redshift uncertainties, and it generates a PP plot that is perfectly consistent with the expected diagonal line. The figures also show results computed in the perfect-redshift measurement limit, $\delta z/z = 0\%$. In this limit the full and approximate likelihoods are exactly equivalent and so we recover unbiased results using either likelihood.

The reason for going through this example is that it illustrates that it is important to be careful when generating mock catalogs that they have a realistic galaxy density. Using galaxy catalogs that are too sparse can lead to biases because of the inconsistency in the standard analysis that becomes
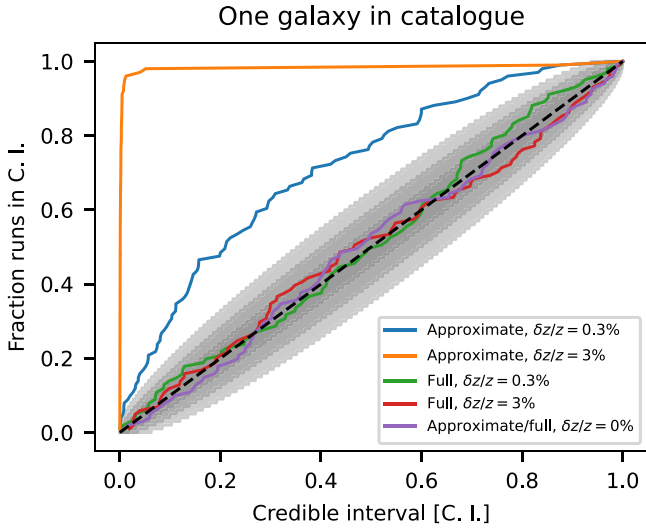
One galaxy in catalogue

Figure 9. PP plots for the one-galaxy analysis, with 200 observed GW sources, comparing the results using the approximate and full likelihoods. We consider the same values of the fractional error in the redshift measurement that were shown in Figure 8.
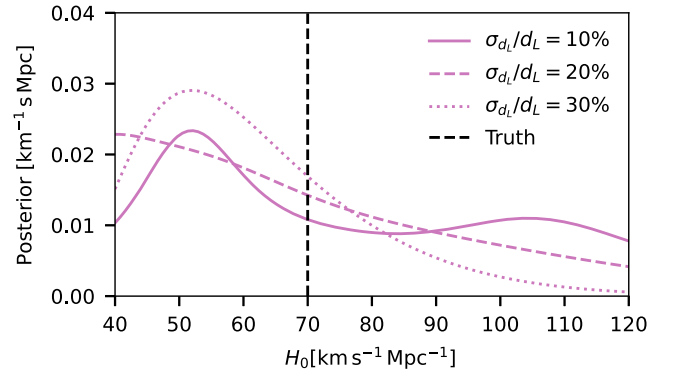

Figure 10. H_0 posteriors for 200 GW events drawn from a uniform in comoving volume distribution of 10^6 galaxies. The several lines indicate various fractional errors on d_L for the GW likelihood model. The absence of clustering significantly weakens the result, rendering it nearly uninformative in the limit of large numbers of galaxies, but does not incur in significant biases.

apparent in this limit. However, these are not indicative of a fundamental problem in the analysis, and can be avoided by using the full likelihood of Equation (31).

4. Discussion

We have shown that the dark siren statistical method, utilizing a galaxy catalog of potential hosts, is able to produce unbiased estimates of H_0. This is true, however, only if the statistical framework is consistent with the generative model of the population. An analysis model that is inconsistent with the simulations will lead to biases.

TH21 concludes that the bias on H_0 is introduced by the absence of galaxy clustering, which is exacerbated when the number of galaxies in a given direction is large (e.g., the cases D_{15} and D_{25}), or when the error budget on the GW luminosity distance is large (leading to the clustering effects being washed out in the GW localization volume). We argue that the conclusions in TH21 are likely due to an inconsistent procedure for the generation of their mock data and calculating the selection effects for their GW events. The following sections will demonstrate that an absence of structure in the redshift distribution of galaxies does not lead to a biased estimate of H_0 (indeed, the result merely becomes uninformative), and that similar biases to those seen in TH21 can be produced by introducing inconsistencies between the mock-data generation and analysis.

4.1. The Absence of Clustering

In Section 3, we have shown that in the limit that there are many galaxies along the LOS (e.g., D_{15} and D_{25}) we recover an unbiased H_0 result. As discussed above, in this limit we expect the constraining power on H_0 to significantly decrease. It can be demonstrated mathematically, as we do in the Appendix, that under some simplified assumptions the H_0 posterior is expected to be flat (assuming a flat prior) in this case.

To further demonstrate this, here we perform a full simulation of this measurement. We repeat the steps of simulating a Mock Data Challenge, explained in Section 3.1, with the only difference that the galaxy distribution is synthetically generated to be uniform in comoving volume

without any large-scale clustering. We draw 2×10^6 galaxies for this case in order to have a distribution that is as continuous as possible, and we use the same uncertainty model for the observed redshifts of the galaxies. In other words, $p_{gal}(z) \propto \frac{dV_c}{dz}$. In Figure 10, we show the H_0 posteriors of 200 GW signals combined for a relative error on d_L on the GW likelihood model of 10%, 20%, and 30%. We note that the recovered H_0 posterior is not as constraining as in the previous cases, as expected. The posterior does not display any noticeable bias in this case. Therefore, contrary to TH21, we conclude that the absence of clustering is not expected to introduce a bias on the H_0 estimation. In reality, the universe is *not* uniform in comoving volume on smaller scales, as we know that the universe's large-scale structure does exhibit clustering, but the effect of clustering is to enhance the H_0 constraint rather than being essential to it.

4.2. Possible Sources of Inconsistency in Implementing the Standard Siren Method

In this section, we speculate that possible inconsistencies between the statistical framework and the generation of mock data can easily introduce biases in the estimation of H_0. It is difficult to characterize the interplay of different issues, and how they might translate to a final H_0 bias; for this reason, we explore one possible inconsistency at a time. In what follows, we focus on a number of possible errors and demonstrate how they impact the H_0 posterior.

Inconsistency 1: double counting. A first issue is assigning GW events to the sample of MICEcat galaxies using a weight factor similar to Equation (13). However, since the sample of MICEcat galaxies is already distributed as Equation (13), this procedure effectively introduces a "double counting" problem. In other words, if the distribution of galaxies follows a z^2 profile, the distribution of GW events will follow a z^4 profile. This has a crucial impact on the H_0 inference when combined with errors in the luminosity distance generation and the presence of a detection threshold. Understanding the implication of the H_0 bias is not trivial for the dark siren approach, therefore we perform a simulation. In Figure 11 (top panel), we repeat the analysis as in Section 3 but generating GWs in galaxies with an extra weight factor given by Equation (13) (the analysis is then performed normally). As we see, this type of mismatch usually results in a bias toward lower values of H_0.
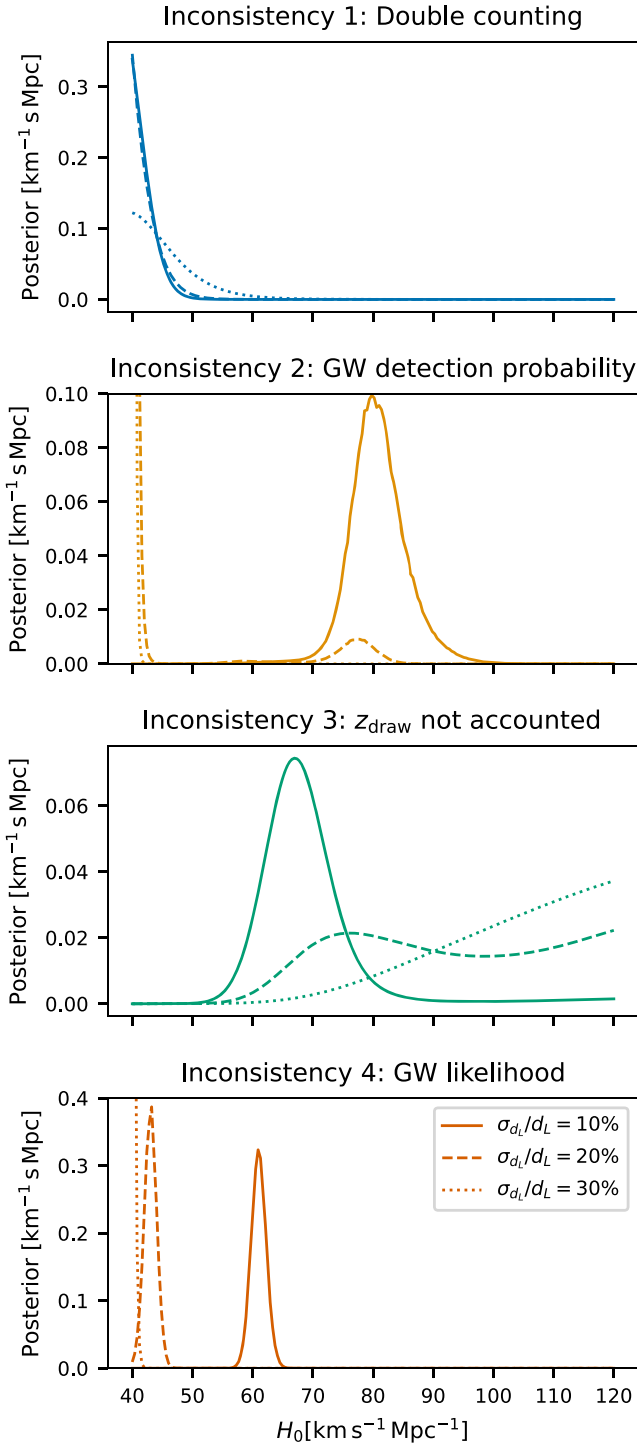
Figure 11. H_0 posteriors with 200 GW events for several cases with inconsistent choices in either the simulation of mock data or the statistical framework for the D_{15} LOS. The solid, dashed, and dotted lines are obtained with $\sigma_{d_L}/d_L = 10\%$, 20%, and 30%. First panel: GW events are generated by counting twice over the galaxy catalog. Second panel: the detection probability in the analysis is set as a Heaviside step function. Third panel: GW events are drawn from $z < 0.3$ and this correction is ignored in the analysis. Fourth panel: the luminosity distance dependence of the GW likelihood standard deviation is not properly taken into account in the formalism. Note that for this particular case alone we show D_{21}, which shows a more dramatic behavior than D_{15}, at least for the specific random seeds we tested.

Inconsistency 2: GW detection probability. A second issue is a general misinterpretation of observed quantities ("data") and latent variables ("true values") and their interaction with the

selection biases. As we detail in Section 2, the selection bias on the GW side should be corrected using the GW detection probability. The latter is defined by integrating the likelihood over all possible data sets that can be detected given a true value of the luminosity distance. TH21 assumes that the detection probability is a Heaviside step function of the true luminosity distance, with the probability dropping to 0 at $\hat{d}_L^{\text{thr}}$. This is correct only in the limit that there are no errors on the luminosity distance estimation. Figure 11 (middle panel) shows that, when the luminosity distance estimation comes with a nonnegligible uncertainty, assuming a Heaviside step function for the detection probability can bias the estimation of the Hubble constant to lower values.

Inconsistency 3: z_{draw} not accounted. A third potential misconception is the fact that GW events are drawn from galaxies with $z < 0.3$ but this aspect is not accounted for in the analysis. In order to understand the biases introduced by this type of error, we simulate GW events from galaxies at $z < 0.3$ and then build the interpolant $p_{\text{CBC}}(z)$ for the analysis using $z_{\text{draw}} = 1.4$. In other words, the analysis accounts also for selection effects due to GW hosts present between $0.3 < z < 1.4$, while the former simulation only considered GW hosts $z < 0.3$. In Figure 11 (third panel), we show that this type of error might introduce a high H_0 bias. We note that, in a real analysis, such as that done in Abbott et al. (2023b), we do not know if there is a maximum redshift at which GW events are generated. However, this is taken into account by fitting simultaneously with flexible merger-rate models alongside the value of H_0.

Inconsistency 4: GW likelihood mismodeling. Another possible source of error is treating the GW likelihood as if the standard deviation was not dependent on the true value of d_L, although the simulations are made assuming this dependence (see Section 2). By dropping the overall normalization factor in the GW likelihood, one is in practice ignoring a part of the likelihood that depends on the true luminosity distance. This causes a biased dependence of the H_0 posterior on the luminosity distance uncertainty, which we show in Figure 11 (bottom panel). We find that, in this case, the inconsistency has the effect of biasing H_0 toward lower values for increasing values of σ_{d_L}, similarly to what TH21 found. In the bottom panel of Figure 11, we show the results using a 1° light-cone simulation (instead of the 5° D_{15} simulation) as it shows a stronger bias change with σ_{d_L}. We also assume a Heaviside function for the selection effects as in TH21 so that the σ_{d_L} dependence on true luminosity distance is also not taken into account in this term. A general discussion about the misinterpretation of latent and observed variables, and the bias caused by assuming that observables depend on intrinsic properties in the mocks but not in the likelihood modeling in the context of GWs, is also provided[40].

Inconsistency 5: simulating dark siren events along the same line of sight. The statistical method used in TH21 is for a high-galaxy-density regime, but this is inconsistent with the generation of data from D_{11} and D_{21}, where it is possible for multiple GW events to be drawn from the same low-redshift host galaxy. Some of the H_0 biases for D_{11} and D_{21} might originate from this inconsistency, as we have shown in Section 3.3. However, in our own analyses, this produced a noticeable bias less frequently than the other potential sources

---

[40] in https://github.com/farr/MockPosteriors.

of bias we explored. Moreover, we find that a low-H_0 bias will arise if there are low-redshift overdensities along the LOS and the full-likelihood analysis is not used. When simulating multiple events along the same LOS, if there is an overdensity of galaxies in that direction, the H_0 posterior is expected to display a peak at an H_0 lower than the input value, which corresponds to the redshift of the overdensity and luminosity distance of the peak of the GW distribution (likely ∼1500 Mpc in this work, close to the detection horizon). Especially when using the light cones with fewer galaxies, there are very few galaxies at the redshift of interest, $z \lesssim 0.3$, so we are sensitive to the small number statistics of a peak in the redshift distribution that may shift from realization to realization (as one does not always get the same exact redshift peak, and thus corresponding H_0 peak, in all realizations). As there are significantly more galaxies and volume at larger redshifts than at lower redshifts, which results in the effect of the over-densities on the H_0 posterior being typically less pronounced in the former regime rather than in the latter, this effect is only more significant at lower H_0. Because in reality it is extremely unlikely to have multiple events from one narrow LOS, we advise against using a small-area light-cone simulation to make dark siren analyses. Alternatively, the presence of the same galaxies across multiple GW events needs to be taken into account with the full-likelihood analysis so as not to incur in a bias.

Inconsistency 6: aggressive truncation of the galaxy catalog. The sum over galaxies in Equation (13) extends over all galaxies that could possibly host GW sources. This is typically a large number, but for any given GW event the majority of these galaxies will lie in positions with little posterior support and can therefore be omitted. It is a common strategy to trim the catalog when doing an analysis by using the extent of the GW posterior in sky location and distance to eliminate galaxies from the catalog. This must be done carefully, to make sure that no galaxies are omitted that lie within the support of the distance posterior for any value of the cosmological parameters consistent with the prior. This is illustrated in Figure 2 of Soares-Santos et al. (2019), where posteriors on H_0 are compared for two different truncations of the galaxy catalog. It is always good practice to verify the (lack of an) impact of catalog truncation on any statistical GW measurement.

5. Conclusions

We have presented an introduction to the dark siren statistical method. We have explicitly demonstrated that, contrary to TH21, the dark siren statistical method can robustly recover an unbiased estimate of H_0. We expect the method to produce unbiased estimates of additional cosmological para-meters. We have also shown that, in the limit of the absence of galaxy clustering, the dark siren statistical method continues to provide an unbiased estimate of H_0. Notebooks showing how to reproduce this analysis can be found at https://github.com/simone-mastrogiovanni/hitchhiker_guide_dark_sirens.

We emphasize that this work is not a forecast, but instead is meant to provide a pedagogical introduction to the dark siren approach, highlighting common pitfalls. Because of the specific assumptions that we have made, this is not a realistic simulation of GW events and their detections, and we do not provide estimates for the future sensitivity of dark siren probes (see, e.g., Chen et al. 2018 and Gray et al. 2020 for forecasts).

This is why we do not specifically quote precision measure-ments of H_0 throughout the manuscript.

Moreover, we remind the reader that a data-analysis frame-work able to analyze real GW data is, in some aspects, more complicated than what is discussed in this paper. On the GW side, as has already been noted in the literature, selection effects based on GW sensitivity studies injecting signals in real data should also be carefully taken into account for standard sirens with counterparts (e.g., Farr 2019; Mortlock et al. 2019). Moreover, assumptions about the underlying population of compact object binaries can have a significant impact on estimating the Hubble constant not only for the galaxy catalog approach (Abbott et al. 2023b) but perhaps more importantly for a GW-only approach to standard siren cosmology (Mastrogio-vanni et al. 2021; Ezquiaga & Holz 2022; Karathanasis et al. 2022; Mukherjee 2022). On the galaxy catalog side, there are several challenges related to the completeness correction of the galaxy catalog and the presence of a possible correlation between galaxy intrinsic luminosities and merger rates (Gray et al. 2020, 2022). Also, techniques exploring spatial clustering using cross-correlation need to properly mitigate the effects from GW bias parameters and its redshift evolution, as demonstrated in Mukherjee et al. (2021, 2022). Clearly, the full Bayesian framework is more complicated for a statistical dark standard siren analysis as opposed to the case where a unique host galaxy is identified, so it is understandable that this method poses more challenges than the one simulated in this paper. However, it is important to note that *all* standard siren methods are dependent on the same assumptions and potential sources of mismodeling considered here.

Another caveat of this analysis, since it is built in part around the assumptions of TH21 to show how not to incur the biases they claim to find, is that of drawing a large number of events from the same LOS. In this analysis, we have used only two directions to simulate all the events of the GW sources. When drawing from the same LOS, the contribution of the large-scale structure is always the same, so your posterior may prefer the specific values of H_0 around the value corresponding to the overdensities along the LOS for distances close to the peak of the detected GW source distribution, thereby lacking the expected variation in the large-scale structure distribution. In reality, over different LOSs, the contributions from the different underdensities/overdensities will cancel out. We caution the reader from running the simulations we have made available using a number of GW events much larger than the number of galaxies.

To summarize, our analysis shows that, using the dark siren statistical method, we can measure the value of the Hubble constant H_0 without any bias, in contrast to the recent claim by TH21. Accurate measurements of the Hubble constant require correctly taking into account the selection effects in both the galaxy catalog side and GW side, and a proper modeling of the likelihoods in question. We encourage research studies focused on the impact of population assumptions and selection effects to advance the entire field of standard siren cosmology.

Appendix
Dark Sirens in the Absence of Galaxy Clustering: A Simplified Calculation

In this appendix, we show that under the assumptions that (i) the GW likelihood and detection criteria are only functions of the source luminosity distance and (ii) we are in the local universe and the distribution of galaxies is uniform, then we expect to obtain an uninformative H_0 likelihood. This is a mathematical proof that the statement in TH21, according to which in the absence of galaxy clustering one recovers a biased posterior, is incorrect. Let us remind that the hierarchical likelihood is given by

$$\mathcal{L}(\{x\}|H_0) \propto \prod_i^{N_{\text{obs}}} \frac{\int dz \mathcal{L}_{\text{GW}}(\hat{d}_L^i | d_L(z, H_0)) \, p_{\text{CBC}}(z)}{\int dz \, P_{\text{det}}^{\text{GW}}(z, H_0) \, p_{\text{CBC}}(z)}, \quad \text{(A1)}$$

and the posterior on H_0 can be calculated as

$$p(H_0|\{x\}) = \frac{\mathcal{L}(\{x\}|H_0)p(H_0)}{\int \mathcal{L}(\{x\}|H_0)p(H_0)dH_0}. \quad \text{(A2)}$$

If we are in the local universe, i.e., distance can be approximated as

$$d_L \approx \frac{cz}{H_0} \quad \text{(A3)}$$

and if the distribution of galaxies is uniform, we can write it as continuous and it is

$$p_{\text{CBC}}(z) \approx p_{\text{cat}}(z) = \frac{3}{z_{\text{max}}^3} z^2. \quad \text{(A4)}$$

In the above equation, z_{max} is a redshift limit up to which galaxies are generated. Moreover, if the single GW likelihood only depends on the source luminosity distance, and the detection criteria depend on a threshold luminosity distance $\hat{d}_L^{\text{thr}}$ such that $d_L(z_{\text{max}}, H_0) \gg \hat{d}_L^{\text{thr}}$ for all the H_0 values explored in the analysis, then the selection bias can be written as

$$\int P_{\text{det}}^{\text{GW}}(z, H_0) \, p_{\text{CBC}}(z) dz$$
$$= \int P_{\text{det}}^{\text{GW}}(d_L) \, p_{\text{CBC}}(z(d_L, H_0)) \left| \frac{\partial z}{\partial d_L} \right| \, dd_L = kH_0^3, \quad \text{(A5)}$$

where k is a constant. To solve the integral in the above equation, we have just performed a change of variable from z to d_L. We can perform a similar procedure for the numerator of Equation (A1) and show that this is given by

$$\int \mathcal{L}_{\text{GW}}(\hat{d}_L^i | d_L(z, H_0)) \, p_{\text{CBC}}(z) dz$$
$$= \int \mathcal{L}_{\text{GW}}(\hat{d}_L^i | d_L(z, H_0)) \, p_{\text{CBC}}(z(d_L, H_0)) \left| \frac{\partial z}{\partial d_L} \right| \, dd_L$$
$$= C_i H_0^3, \quad \text{(A6)}$$

where again we have performed a change of variable to solve the integral, and C_i is a normalization constant that depends on the signal detected from the data realization.

Note that, when calculating Equations (A5) and (A6), the hypothesis that $d_L(z_{\text{max}}, H_0) \gg \hat{d}_L^{\text{thr}}$ makes sure that C_i and k does not depend on H_0. If this condition is not satisfied, these constants will depend on H_0 and must be correctly included in the analysis. By using Equations (A5) and (A6) to calculate Equation (A1), we obtain that the GW likelihood is constant and does not depend on H_0, namely

$$\mathcal{L}(\{x\}|H_0) = \prod_i^{N_{\text{obs}}} \frac{C_i}{k}. \quad \text{(A7)}$$

Therefore, the posterior on H_0 equals the prior used in Equation (A2).

ORCID iDs

Jonathan R. Gair ⓘ https://orcid.org/0000-0002-1671-3668
Archisman Ghosh ⓘ https://orcid.org/0000-0003-0423-3533
Rachel Gray ⓘ https://orcid.org/0000-0002-5556-9873
Daniel E. Holz ⓘ https://orcid.org/0000-0002-0175-5064
Simone Mastrogiovanni ⓘ https://orcid.org/0000-0003-1606-4183
Suvodip Mukherjee ⓘ https://orcid.org/0000-0002-3373-5236
Antonella Palmese ⓘ https://orcid.org/0000-0002-6011-0530
Tessa Baker ⓘ https://orcid.org/0000-0001-5470-7616
Freija Beirnaert ⓘ https://orcid.org/0000-0002-4003-7233
Maciej Bilicki ⓘ https://orcid.org/0000-0002-3910-5809
Hsin-Yu Chen ⓘ https://orcid.org/0000-0001-5403-3762
Gergely Dálya ⓘ https://orcid.org/0000-0003-3258-5763
Jose Maria Ezquiaga ⓘ https://orcid.org/0000-0002-7213-3211

Will M. Farr ⬥ https://orcid.org/0000-0003-1540-8562
Maya Fishbach ⬥ https://orcid.org/0000-0002-1980-5293
Juan Garcia-Bellido ⬥ https://orcid.org/0000-0002-9370-8360
Tathagata Ghosh ⬥ https://orcid.org/0000-0001-9848-9905
Hsiang-Yu Huang ⬥ https://orcid.org/0000-0002-1665-2383
Christos Karathanasis ⬥ https://orcid.org/0000-0002-0642-5507
Konstantin Leyde ⬥ https://orcid.org/0000-0001-7661-2810
Ignacio Magaña Hernandez ⬥ https://orcid.org/0000-0003-2362-0459
Johannes Noller ⬥ https://orcid.org/0000-0003-2210-775X
Gregoire Pierra ⬥ https://orcid.org/0000-0003-3970-7970
Peter Raffai ⬥ https://orcid.org/0000-0001-7576-0141
Antonio Enea Romano ⬥ https://orcid.org/0000-0002-0314-8698
Monica Seglar-Arroyo ⬥ https://orcid.org/0000-0001-8654-409X
Danièle A. Steer ⬥ https://orcid.org/0000-0002-8781-1273
Cezary Turski ⬥ https://orcid.org/0000-0001-7122-6240
Maria Paola Vaccaro ⬥ https://orcid.org/0000-0003-3776-9246
Sergio Andrés Vallejo-Peña ⬥ https://orcid.org/0000-0002-6827-9509

References

Aasi, J., Abbott, B. P., Abbott, R., et al. 2015, CQGra, 32, 074001
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, Natur, 551, 85
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, PhRvL, 119, 161101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017c, ApJL, 848, L12
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020, CQGra, 37, 055002
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2021a, ApJ, 909, 218
Abbott, B. P., Abe, H., Acernese, F., et al. 2023b, ApJ, 949, 76
Abbott, R., Abbott, T. D., Abraham, S., et al. 2021b, ApJL, 913, L7
Abbott, Abe, H., Acernese, F., et al. 2023a, PhRvX, 13, 011048
Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, CQGra, 32, 024001
Aubin, F., Brighenti, F., Chierici, R., et al. 2021, CQGra, 38, 095004
Bera, S., Rana, D., More, S., & Bose, S. 2020, ApJ, 902, 79
Cannon, K., Caudill, S., Chan, C., et al. 2021, SoftX, 14, 100680
Carretero, J., Castander, F. J., Gaztanaga, E., Crocce, M., & Fosalba, P. 2015, MNRAS, 447, 650
Carretero, J., Tallada, P., Casals, J., et al. 2017, in Proc. European Physical Society Conf. on High Energy Physics, ed. P. Checchia (Trieste: SISSA), 488
Chen, H.-Y., Fishbach, M., & Holz, D. E. 2018, Natur, 562, 545
Chernoff, D. F., & Finn, L. S. 1993, ApJL, 411, L5
Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, MNRAS, 453, 1513
Dal Canton, T., et al. 2014, PhRvD, 90, 082004
Dalal, N., Holz, D. E., Hughes, S. A., & Jain, B. 2006, PhRvD, 74, 063006

Del Pozzo, W. 2012, PhRvD, 86, 043011
Diaz, C. C., & Mukherjee, S. 2022, MNRAS, 511, 2782
Ezquiaga, J. M., & Holz, D. E. 2021, ApJL, 909, L23
Ezquiaga, J. M., & Holz, D. E. 2022, PhRvL, 129, 061102
Farr, W. M. 2019, RNAAS, 3, 66
Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, ApJL, 883, L42
Finke, A., Foffa, S., Iacovelli, F., Maggiore, M., & Mancarella, M. 2021, JCAP, 2021, 026
Fishbach, M., Gray, R., Magaña Hernandez, I., et al. 2019, ApJL, 871, L13
Fishbach, M., & Holz, D. E. 2017, ApJL, 851, L25
Fishbach, M., Holz, D. E., & Farr, W. M. 2018, ApJL, 863, L41
Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015, MNRAS, 448, 2987
Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2014, MNRAS, 447, 1319
Gray, R., Hernandez, I. M., Qi, H., et al. 2020, PhRvD, 101, 122001
Gray, R., Messenger, C., & Veitch, J. 2022, MNRAS, 512, 1127
Hahn, C., Wilson, M. J., Ruiz-Macias, O., et al. 2023, AJ, 165, 253
Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, MNRAS, 447, 1724
Holz, D. E., & Hughes, S. A. 2005, ApJ, 629, 15
Karathanasis, C., Mukherjee, S., & Mastrogiovanni, S. 2022, arXiv:2204.13495
Klimenko, S., & Mitselmakher, G. 2004, CQGra, 21, S1819
Leandro, H., Marra, V., & Sturani, R. 2022, PhRvD, 105, 023523
Leyde, K., Mastrogiovanni, S., Steer, D. A., Chassande-Mottin, E., & Karathanasis, C. 2022, JCAP, 2022, 012
Mandel, I., Farr, W. M., & Gair, J. R. 2019, MNRAS, 486, 1086
Mastrogiovanni, S., Leyde, K., Karathanasis, C., et al. 2021, PhRvD, 104, 062009
Messick, C., Blackburn, K., Brady, P., et al. 2017, PhRvD, 95, 042001
Mortlock, D. J., Feeney, S. M., Peiris, H. V., Williamson, A. R., & Nissanke, S. M. 2019, PhRvD, 100, 103523
Mukherjee, S. 2022, MNRAS, 515, 5495
Mukherjee, S., Krolewski, A., Wandelt, B. D., & Silk, J. 2022, arXiv:2203.03643
Mukherjee, S., Wandelt, B. D., Nissanke, S. M., & Silvestri, A. 2021, PhRvD, 103, 043520
Mukherjee, S., Wandelt, B. D., & Silk, J. 2020, MNRAS, 494, 1956
Nitz, A. H., Dent, T., Dal Canton, T., Fairhurst, S., & Brown, D. A. 2017, ApJ, 849, 118
Oguri, M. 2016, PhRvD, 93, 083511
Palmese, A., Bom, C. R., Mucesh, S., & Hartley, W. G. 2023, ApJ, 943, 56
Palmese, A., deVicente, J., Pereira, M. E. S., et al. 2020, ApJL, 900, L33
Palmese, A., Fishbach, M., Burke, C. J., Annis, J., & Liu, X. 2021, ApJL, 914, L34
Schutz, B. F. 1986, Natur, 323, 310
Singer, L. P., & Price, L. R. 2016, PhRvD, 93, 024013
Soares-Santos, M., Palmese, A., Hartley, W., et al. 2019, ApJL, 876, L7
Tallada, P., Carretero, J., Casals, J., et al. 2020, A&C, 32, 100391
Taylor, S. R., Gair, J. R., & Mandel, I. 2012, PhRvD, 85, 023535
Trott, E., & Huterer, D. 2021, arXiv:2112.00241
Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2022, in Handbook of Gravitational Wave Astronomy, ed. C. Bambi (Berlin: Springer), 45
Ye, C., & Fishbach, M. 2021, PhRvD, 104, 043507