

How Low Can You Go: Constraining the Effects of Catalog Incompleteness on Dark Siren Cosmology

MADISON VANWYNGARDEN,^{1,2} MAYA FISHBACH,^{1,3,4} ADITYA VIJAYKUMAR,¹ ALEXANDRA G. GUERRERO,⁵ AND DANIEL E. HOLZ^{5,6,7}

¹*Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON. M5S 3H8, Canada*

²*Steward Observatory, University of Arizona, 933 N Cherry Ave, Tucson, AZ 85712, USA*

³*David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, 50 St. George St., Toronto ON. M5S 3H4, Canada*

⁴*Department of Physics, University of Toronto, 60 St. George St., Toronto, ON M5S 3H8, Canada*

⁵*Department of Physics, University of Chicago, Chicago, IL 60637, USA*

⁶*Department of Astronomy & Astrophysics, Enrico Fermi Institute, Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA.*

⁷*NSF-Simons AI Institute for the Sky (SkAI), 172 E. Chestnut St., Chicago, IL 60611, USA*

ABSTRACT

Gravitational waves (GWs) serve as standard sirens by directly encoding the luminosity distance to their source. When the host galaxy redshift is known, for example, through observation of an electromagnetic (EM) counterpart, GW detections can provide an independent measurement of the Hubble constant, H_0 . However, even in the absence of an EM counterpart, inferring H_0 is possible through the dark siren method. In this approach, every galaxy in the GW localization volume is considered a potential host that contributes to a measurement of H_0 , with redshift information supplied by galaxy catalogs. Using mock galaxy catalogs, we explore the effect of catalog incompleteness on dark siren measurements of H_0 . We find that in the case of well-localized GW events, if GW hosts are found in all galaxies with host halo masses $M_h > 2 \times 10^{11} M_\odot h^{-1}$, catalogs only need to be complete down to the 1% brightest magnitude $M_i < -22.43$ to draw an unbiased, informative posterior on H_0 . We demonstrate that this is a direct result of the clustering of fainter galaxies around brighter and more massive galaxies. For a mock galaxy catalog without clustering, or for GW localization volumes that are too large, using only the brightest galaxies results in a biased H_0 posterior. These results are important for informing future dark siren analyses with LIGO-Virgo-KAGRA as well as next-generation detectors.

1. INTRODUCTION

The $4\text{--}6\sigma$ discrepancy between low-redshift measurements of the Hubble constant, H_0 , (e.g. supernovae and cepheids; Scolnic et al. 2022; Riess et al. 2021; Uddin et al. 2023) and high-redshift measurements (e.g. cosmic microwave background; Planck Collaboration et al. 2020) has given rise to the current crisis in cosmology known as the Hubble tension (Freedman & Madore 2023). Whether these differences are a result of systematics in the cosmic distance ladder or a breakdown of Λ CDM cosmology is still under investigation. Alternative cosmological probes may be able to reduce the tension or provide insight into its cause (Moresco et al. 2022).

Decades before the first gravitational wave (GW) detection (Abbott et al. 2016), Schutz (1986) proposed using GW observations of compact binary mergers to measure the Hubble constant. GW signals are referred to as “standard sirens” because their waveform directly

encodes the luminosity distance to the source (Holz & Hughes 2005). However, their redshift is degenerate with the source-frame mass, making an independent measurement of the host galaxy’s redshift essential. When this redshift is available, a measurement of H_0 can be made without the need of a cosmic distance ladder (Schutz 1986; Holz & Hughes 2005). Bright sirens—GW sources with an electromagnetic (EM) counterpart—allow for identification of a unique host galaxy redshift. The identification of NGC 4993 as the host galaxy of the binary neutron star merger GW170817 (Abbott et al. 2017) allowed for the first measurement of H_0 from GWs (Abbott et al. 2017a,b; Coulter et al. 2017; Soares-Santos et al. 2017). However, no other GW signals with EM counterparts have been confidently confirmed.

There are different methods for estimating the redshifts of GW events without EM counterparts. These include the spectral siren, which uses the distribution of source-frame masses to infer cosmology (Taylor &

Gair 2012; Farr et al. 2019; Mastrogiovanni et al. 2021; Ezquiaga & Holz 2022); love sirens, which uses the neutron star equation of state (Messenger & Read 2012; Chatterjee et al. 2021); stochastic sirens, which uses the amplitude of the stochastic gravitational wave background (Cousins et al. 2025); and approaches that utilize cross-correlations between GW sources and large-scale structure (Namikawa et al. 2016; Bera et al. 2020; Mukherjee et al. 2021, 2024).

In this work, we focus on Schutz (1986)’s original proposal, which utilized GWs without an EM counterpart, known as dark sirens, and redshift information from galaxy catalogs. In the so-called “*dark siren* method,” every galaxy in the GW localization volume is considered a potential host that provides redshift information contributing to a measurement of H_0 (Del Pozzo 2012; Chen et al. 2018; Fishbach et al. 2019; Soares-Santos et al. 2019; Gray et al. 2020; Palmese et al. 2020; Abbott et al. 2023; Gair et al. 2023; Palmese et al. 2023; Mastrogiovanni et al. 2023; Gray et al. 2023). The H_0 measurement is obtained by combining results from each potential host galaxy. Though each dark siren measurement is less informative than a single bright siren measurement, the number of dark sirens far exceeds the number of bright sirens, allowing for measurements to be stacked over time to increase precision. The best localized events from observing runs O1-O4a from the LIGO-Virgo-KAGRA collaboration (LVK) has constrained H_0 to $\sim 20\%$ uncertainty (The LIGO Scientific Collaboration et al. 2025a; Bom et al. 2024). Dark siren cosmology will continue to play an important role with the inclusion of next-generation detectors, such as LISA (Auclair et al. 2023), Cosmic Explorer (Evans et al. 2021), and Einstein Telescope (Maggiore et al. 2020; Branchesi et al. 2023), which will greatly increase the number of dark sirens.

A persistent problem in dark siren cosmology is the incompleteness of galaxy catalogs. A GW event’s localization volume can contain hundreds of galaxies for relatively well-localized events to $\mathcal{O}(10^4 - 10^5)$ galaxies for a typical detection (Fishbach et al. 2019; Moresco et al. 2022; Chen & Holz 2016). Galaxy catalogs are inherently magnitude-limited and sky coverage and depth vary dramatically from survey to survey. For any dark siren analysis, it is possible that the true host of a GW event is not included in the galaxy catalog or catalogs overlapping with the localization volume. In this case, the host providing support for the correct value of H_0 would be lost. However, due to the clustering of galaxies, another in-catalog galaxy may still be present at the correct redshift (MacLeod & Hogan 2008). Previous studies have exploited this fact to develop completeness

corrections in which the missing galaxies in an incomplete catalog are assumed to be preferentially close to the in-catalog galaxies, leveraging large-scale structure to improve dark siren analyses (Finke et al. 2021; Dalang et al. 2024; Dalang & Baker 2024; Leyde et al. 2024, 2025; Borghi et al. 2025). Other studies have explored the effects of catalog incompleteness on constraining H_0 with real data from GWTC-3, including Naveed et al. (2025) and Beirnaert et al. (2025) which investigated the use of bright galaxies and galaxy cluster catalogs, respectively.

In this work, we further examine the effects of galaxy catalog incompleteness on the measurement of H_0 ; however, unlike previous studies, we do not apply any completeness correction and use a mock galaxy catalog to systematically investigate different limits of catalog completeness. We interpret and explain our results in the context of galaxy clustering. In Section 2, we describe the statistical framework and mock data used in this study. In Sections 3 and 4, we examine the results of using incomplete catalogs for both a simple test case and a more realistic scenario. We discuss the impact of galaxy clustering on our results in Section 5 before discussing caveats to our results in Section 6 and presenting our conclusions in Section 7.

2. METHODS

2.1. Statistical Framework

This section outlines the statistical framework used in our analysis, following the procedure and code provided by Gair et al. (2023). We begin by assuming that for a given set of GW signals, we observe a set of luminosity distances $\hat{d}_L$. We can then derive a posterior on H_0 using Bayes’ theorem,

$$p(H_0|\hat{d}_L) \propto \mathcal{L}(\hat{d}_L|H_0) p(H_0) \quad (1)$$

where $p(H_0)$ is the prior on H_0 and $\mathcal{L}(\hat{d}_L|H_0)$ is the likelihood of observing $\hat{d}_L$ given a value of H_0 . The likelihood for a single GW event is described as

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

where $\mathcal{L}_{\text{GW}}(\hat{d}_L^i|d_L(z, H_0))$ is the likelihood of measuring a GW source with observed luminosity distance $\hat{d}_L^i$ given its true luminosity distance $d_L(z, H_0)$, $p_{\text{CBC}}(z)$ is the probability that the compact binary coalescence (CBC) occurred at redshift z and $P_{\text{det}}^{\text{GW}}(z, H_0)$ is the GW detection probability. For our simulations, we make the simplifying assumption that the GW distance uncertainties are Gaussian, so that the GW likelihood can be written

as,

$$\mathcal{L}_{\text{GW}}(\hat{d}_L^i | d_L(z, H_0)) = \frac{1}{\sqrt{2\pi}\sigma_{d_L}} \exp \left[-\frac{1}{2} \frac{(\hat{d}_L^i - d_L(z, H_0))^2}{\sigma_{d_L}^2} \right], \quad (3)$$

where $\sigma_{d_L} = A d_L(z, H_0)$ and A is a constant fractional error on the luminosity distance. In the context of dark siren cosmology, $p_{\text{CBC}}(z)$ is the redshift distribution of potential host galaxies in our catalog. We assume that all galaxy redshifts are perfectly known, so $p_{\text{CBC}}(z)$ is given by a sum of Dirac delta functions, $\frac{1}{N_{\text{gal}}} \sum_i^{N_{\text{gal}}} \delta(z - z_{\text{gal}}^i)$. This simplifies the single-event likelihood to the following:

$$\mathcal{L}(\hat{d}_L^i | H_0) = \frac{\sum_i^{N_{\text{gal}}} \mathcal{L}_{\text{GW}}(\hat{d}_L^i | d_L(z, H_0))}{\sum_i^{N_{\text{gal}}} P_{\text{det}}^{\text{GW}}(z, H_0)}. \quad (4)$$

Our simulations also adopt a simplified GW detection probability of

$$\begin{aligned} P_{\text{det}}^{\text{GW}}(z, H_0) &= \int_{-\infty}^{\infty} \Theta(\hat{d}_L; d_L^{\text{thr}}) \mathcal{L}_{\text{GW}}(\hat{d}_L | d_L(z, H_0)) d\hat{d}_L \\ &= \frac{1}{2} \left(1 + \text{erf} \left[\frac{d_L(z, H_0) - d_L^{\text{thr}}}{\sqrt{2} A d_L(z, H_0)} \right] \right), \end{aligned} \quad (5)$$

where $\Theta(\hat{d}_L; d_L^{\text{thr}})$ is a Heaviside step function dropping to 0 for $\hat{d}_L > d_L^{\text{thr}}$. As in [Gair et al. \(2023\)](#), a GW source is detected if its measured luminosity distance is positive and less than our distance threshold, d_L^{thr} , of 1550 Mpc ($z \approx 0.3$ for a standard cosmology). The distance threshold must be applied to the observed luminosity distance rather than the true luminosity distance ([Essick & Fishbach 2024](#)). Due to fluctuations in detector noise, GW sources with true d_L above the threshold value may still be detected.

2.2. Mock Data

This work uses MICECAT, a mock galaxy catalog produced by the MICE collaboration from the MICE grand challenge light cone halo and galaxy catalog ([Fosalba et al. 2015a; Crocce et al. 2015; Fosalba et al. 2015b; Carretero et al. 2015; Hoffmann et al. 2015](#)). The catalog was generated using Halo Occupation Distribution and Halo Abundance Matching prescriptions to populate Friends of Friends dark matter halos with the following input cosmological parameters: $H_0 = 70$ km/s/Mpc, $\Omega_m = 0.25$, $\Omega_b = 0.044$, $\Omega_\Lambda = 0.75$. It is a lightcone covering one octant of the sky out to a redshift of 1.4, containing over 205 million galaxies, and is complete down to an absolute magnitude of $M_r < -18.9$ with host halo masses $M_h > 2.2 \times 10^{11} M_\odot h^{-1}$.

Table 1. Luminosities corresponding to the absolute magnitude cutoffs defining the brightest galaxy fractions in units of $L_\star$

	1% $[L_\star]$	10% $[L_\star]$	20% $[L_\star]$	$L_\star$
<i>i</i> -band	2.03	0.98	0.69	2.97×10^{10}
<i>g</i> -band	3.35	1.51	1.08	1.34×10^{10}

To simulate dark sirens, we first generate lines of sight (LOS) within the MICECAT catalog to populate with GW events. We begin by selecting 200 random RA and Dec pairs within the MICECAT range of 0 to 90 degrees. Lines of sight are then drawn by selecting all galaxies in the catalog within a 1-degree radius from a given RA and Dec pair ($\sim 3 \text{ deg}^2$ sky localizations). These lines of sight contain an average of $\sim 120,000$ galaxies out to $z = 1.4$.

We note that our 1-degree opening angle is smaller than typical GW sky localizations. For CBC candidates from observing runs O1 through O4a only $\sim 10\%$ had sky localizations smaller than 100 deg^2 ([The LIGO Scientific Collaboration et al. 2025a](#)). To explore this regime, we also test 5-degree opening angles ($\sim 75 \text{ deg}^2$ sky localizations) and recover qualitatively similar results to those reported below using the one-degree opening angles. However, events with significantly worse localizations are not informative. Although we keep the opening angle fixed throughout our analysis, we also explore the poorly-localized regime by increasing the distance measurement uncertainty σ_{d_L} which, analogously to larger sky areas, increases the localization volume and therefore the number of galaxies contributing to the analysis. Single-degree sky localizations are expected to be achievable with next-generation detectors ([Maggiore et al. 2024](#)). A more detailed discussion of our choice in localization volume is given in Section 6.

We draw GW events following the procedure of [Gair et al. \(2023\)](#) by first selecting galaxies with a true redshift $z < 1.4$ and using the input cosmological parameters to compute the true d_L . An observed GW distance ($\hat{d}_L$) is then drawn from a Gaussian centered on d_L with $\sigma_{d_L} = A d_L$ where A is 0.1, 0.2, or 0.3. This process is repeated until we have 200 GW events with $\hat{d}_L < d_L^{\text{thr}}$.

2.3. Catalog Incompleteness

Galaxy catalogs are inherently magnitude-limited, so the faintest galaxies in any region may not be observed. The incompleteness of realistic galaxy catalogs is often governed by the galaxies' apparent magnitudes, and the

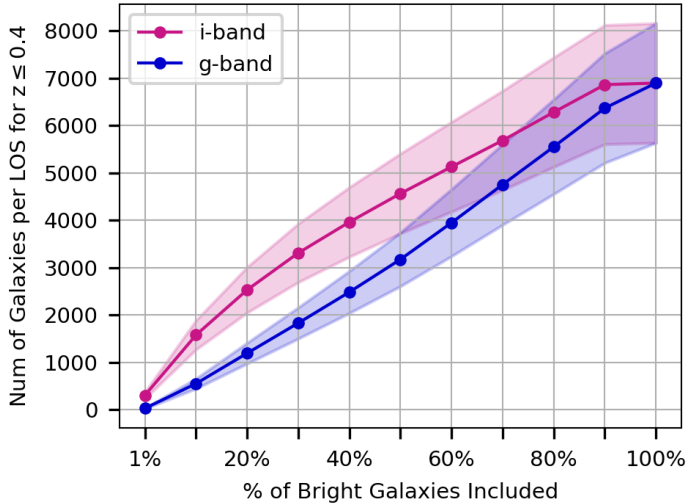


Figure 1. Average number of galaxies per 3 deg^2 LOS for $z \leq 0.4$, shown as a function of increasing brightness cutoffs in the i -band and g -band. The shaded region illustrates the standard deviation over the 200 LOS. There are greater numbers of bright i -band galaxies at this redshift range than bright g -band galaxies.

incompleteness fraction therefore varies with redshift. For example, the GLADE+ catalog used by the LVK collaboration for cosmological analysis from GWTC-3 and GWTC-4 is complete up to 47 Mpc, $\sim 55\%$ complete at 130 Mpc, and $\sim 20\%$ complete at 800 Mpc (Dálya et al. 2022; Abbott et al. 2023; The LIGO Scientific Collaboration et al. 2025b). They define completeness as the fraction of total expected B-band luminosity found in the catalog for different d_L limits. Here, we are interested in simulating a fixed completeness fraction at all redshifts, so we use absolute magnitude thresholds to remove galaxies from the catalog.

MICECAT reports galaxy magnitudes in the DES photometric bands (*grizY*). We convert these apparent magnitudes to absolute magnitudes using MICECAT’s input cosmological parameters. To test the extreme case of incompleteness, along each LOS we retain only the 1%, 10%, and 20% brightest galaxies (smallest absolute magnitudes) in the i - and g -band, which roughly trace stellar mass and recent star formation rate, respectively (Bell et al. 2003). We find that the 1% brightest galaxies in the i -band, for example, contain $\sim 5\%$ of the total luminosity in the catalog. Retaining the 1% brightest i -band galaxies is analogous to 10% completeness under the GLADE+ definition of completeness. Using the 10% and 20% brightest i -band galaxies is analogous to $\sim 28\%$ and $\sim 44\%$ completeness under their definition.

The luminosities corresponding to the magnitude cutoffs used to define the brightness fractions are shown in Table 1, given in units of characteristic luminosity

L_* . We calculate L_* for each photometric band from the characteristic magnitudes, M_* , reported in Blanton et al. (2003). The average number of galaxies at $z \leq 0.4$ per LOS for our 200 LOS is shown in Figure 1 for increasing brightness cutoffs. There are consistently more bright i -band galaxies at our redshifts of interest than bright g -band galaxies, which favor higher redshifts. This is consistent with the picture that most of the star formation in the Universe happens at $z \approx 2$ (Madau & Dickinson 2014).

3. INCOMPLETENESS ALONG A SINGLE LINE OF SIGHT

We begin with the test case of many GW events along a single line of sight. Though this is not realistic for GW detections, it provides a simple test case to examine the influence that the specific LOS has on the H_0 inference. We randomly draw 200 GW host galaxies from the full line of sight, assuming each galaxy is equally likely to host a GW event. However, we select only those galaxies with magnitudes less than the 10% cutoff shown in Table 1 to reconstruct the line of sight galaxy redshift distribution for H_0 inference. This means that only the 10% brightest galaxies are used for N_{gal} and $d_L(z, H_0)$ in the likelihood (see Equation 3).

The posteriors on H_0 for $\sigma_{d_L}/d_L = 0.1, 0.2$, and 0.3 are shown in Figure 2. Increasing the error on d_L is analogous to increasing the size of the GW localization volume. As expected, the posteriors widen and become less informative as the localization volume increases. The posteriors constructed with only the 10% brightest galaxies are remarkably similar to the posterior constructed with the full galaxy sample, especially in the best localized case. Using only a subset of bright galaxies also produces posteriors that are sharper than the full posterior. The i -band bright posterior better mimics the full posterior in the $\sigma_{d_L}/d_L = 0.2$ and 0.3 cases than the g -band bright posterior, indicating that the bright i -band galaxies are a better tracer of the full galaxy distribution at the redshift range of GW observations. Additionally, there are simply more bright i -band galaxies at these redshifts, as shown in Figure 1. This is further confirmed by the CDF of the galaxy redshift distributions for each galaxy population shown in the bottom panel of Figure 2, where the 10% brightest i -band galaxies are aligned with the full redshift distribution.

As discussed in Hanselman et al. (2024), for poorly localized GW events (large errors on d_L), the H_0 posterior becomes more sensitive to global trends in the redshift distribution (see also Perna et al. 2024), which differ between the bright i -band and bright g -band galaxies. For most LOS, bright g -band galaxies are preferentially

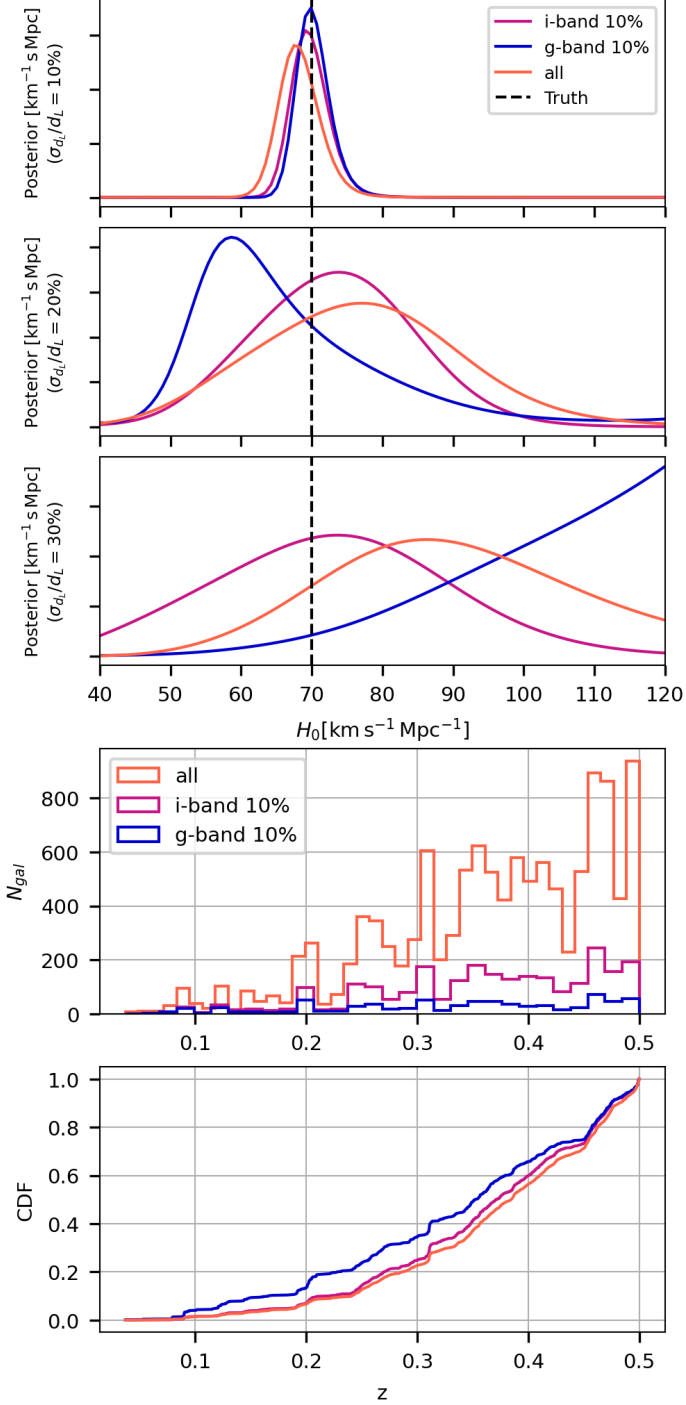


Figure 2. Posterior on H_0 for 200 GW events injected along one LOS and reconstructed with only the 10% brightest galaxies in the i - and g -band with 10%, 20%, and 30% error on GW luminosity distance. In the bottom panels are the normalized redshift distributions and the cumulative density function of the full LOS and the 10% brightest i - and g -band galaxies.

found at higher redshifts, following the star-formation rate which peaks at $z \sim 2$, and will therefore bias the H_0

posterior to higher values. Given small number statistics with this specific LOS, however, the distribution of the 10% brightest g -band galaxies skews to lower redshifts. The shapes of the posteriors are also sensitive to the galaxy clustering in this particular LOS, as over- or under-densities in the redshift distribution can cause the H_0 posteriors to fluctuate for individual events. The H_0 inference is unbiased only if the redshift distribution of the GW events matches the galaxy redshift distribution used in the reconstruction.

Overall, we find that using only the 10% brightest galaxies is sufficient for the dark siren measurement for well-localized GW events. In other words, there is little information lost from catalog incompleteness. Although we only show a single LOS, we repeat the inference with several LOS to verify that the qualitative trends hold. Averaging over three LOS, for the best localized case (10% distance error), the average difference between the 10% brightest i -band posterior and the ‘all’ posterior is 0.04σ , while the average difference between the 10% brightest g -band and ‘all’ is 0.28σ . Here, the difference is defined as the difference in mean between the bright and full posterior divided by the width of the full posterior. We have chosen to express the differences in terms of σ for clarity, but we note that the posteriors are not Gaussian. If we increase the catalog completeness and use the 20% brightest galaxies, the average differences in the full and bright posteriors become 0.01σ for bright i -band galaxies and 0.18σ for bright g -band galaxies in the best-localized case.

The similarities in posteriors between the bright sample of galaxies and the full sample, and the difference between i -band and g -band results can be explained by how the bright galaxies trace the redshift distribution of a given LOS. Galaxies tend to cluster around the most massive, and thus brightest, galaxies. This results in bright galaxies being a good tracer of the full galaxy distribution. To demonstrate this fact, we calculated the KS statistic between the redshift distribution of the brightest galaxies compared to the full distribution along our 200 LOS for $z \leq 0.4$. A histogram of the KS statistics is shown in Figure 3. The i -band galaxies at all brightness intervals have smaller KS statistics, and thus better match the full galaxy distribution than the bright g -band galaxies. Again, there are also more bright i -band galaxies at the redshifts of GW observations (see Figure 1). From Figure 3, we see the 1% brightest i -band galaxies trace the full line of sight distribution as well as a random selection of 1% of the galaxies along the LOS (the random sample is by definition drawn from the full galaxy distribution, and would thus have the smallest possible KS statistic for the given

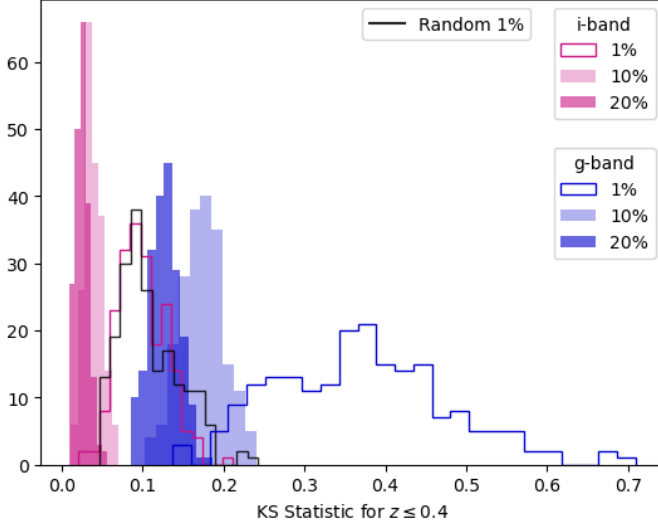


Figure 3. KS statistics between the redshift distributions of the 1%, 10%, and 20% brightest galaxies and the full galaxy sample along our 200 LOS for $z \leq 0.4$. We also include the KS statistics between the redshift distribution of a random 1% of galaxies with the full LOS for comparison.

sample size). The g -band galaxies are worse tracers, with the 1% brightest g -band galaxies differing from the full LOS by 20-60%. Increasing the number of g -band galaxies considered greatly reduces the KS statistic and LOS variation.

4. INCOMPLETENESS ALONG MANY LINES OF SIGHT

In a more realistic analysis, GW events will be isotropically distributed instead of being concentrated along single LOS. To simulate this case, we injected one GW source into each of the 200 LOS with luminosity distance uncertainties of 10, 20, and 30%. We reconstructed each LOS using only the galaxies with absolute magnitudes falling at or below the limits set in Table 1 for the 1, 10, and 20% brightest galaxies, respectively. The posterior generated for each GW event is then combined to form a final posterior as shown in Figure 4.

We begin our analysis of these results with the best localized case at $\sigma_{d_L}/d_L = 0.1$ (top row of Figure 4). The posteriors constructed from only the brightest galaxies are in strong agreement with the posteriors using all galaxies, even when retaining only the 1% brightest galaxies in the catalog. In this regime we also see little change in the posteriors as we increase the galaxy sample from the 1% brightest to the 20% brightest. With small errors on the luminosity distances of our GW events, the majority of the catalog is not necessary for building an informative posterior on H_0 .

In the case where $\sigma_{d_L}/d_L = 0.2$, larger differences in the posteriors are observed, with the bright i -band posterior most closely tracing the full catalog posterior. Using only the 1% brightest galaxies results in a g -band posterior that peaks sharply at the edge of the prior range. This is largely due to the fact that there are fewer g -band galaxies within this magnitude limit at our redshifts of interest from which to reconstruct the full galaxy distribution. The bright g -band galaxies that are included in the 1% brightest sample are biased towards higher redshifts, resulting in a posterior that peaks at larger values of H_0 . As the brightness fraction increases, the i -band posterior begins to converge with the full sample. The bright g -band posterior, on the other hand, has more support at the injected value of H_0 but differs significantly when compared to results obtained using all the galaxies.

Finally, in the $\sigma_{d_L}/d_L = 0.3$ case, all posteriors have strong support at high values of H_0 , similar to the results obtained in the single line of sight test. All three posteriors peak around the same value but the i and g -band bright posteriors most closely resemble each other rather than the full catalog line of sight, though all three begin to converge as the brightness fraction increases. As the error on d_L increases, structures in the redshift distribution begin to be washed out, so the posterior is informed by the overall redshift distribution rather than the specific large-scale structure of any LOS. These results are in good agreement with [Hanselman et al. \(2024\)](#). The posteriors in the worst localized $\sigma_{d_L}/d_L = 0.3$ regime are often less biased than the posteriors in the intermediate $\sigma_{d_L}/d_L = 0.2$ case. This is because at the intermediate localization, there is significant support from the non-host galaxies that bias the measurement. Although these galaxies also contribute in the worst localized case, their contribution is washed out as the overall posterior is broader and therefore less biased (while being less informative).

The influence of galaxy clustering is also diminished as the localization volume increases by increasing the sky localization. Widening our LOSs from a 1-degree opening radius to a 5-degree radius increases the widths of the H_0 posteriors. Using the same injected GW events and comparing at the 95% credible interval on H_0 for the smallest σ_{d_L} , the posterior constructed from the 10% brightest galaxies is a factor of 1.8 to 2.7 wider in the 5 degree LOS case for i -band and g -band, respectively. At the largest σ_{d_L} , the 10% brightest i -band and g -band posteriors are wider by a factor of 1.3 and 1.8.

4.1. Effects of additional galaxies

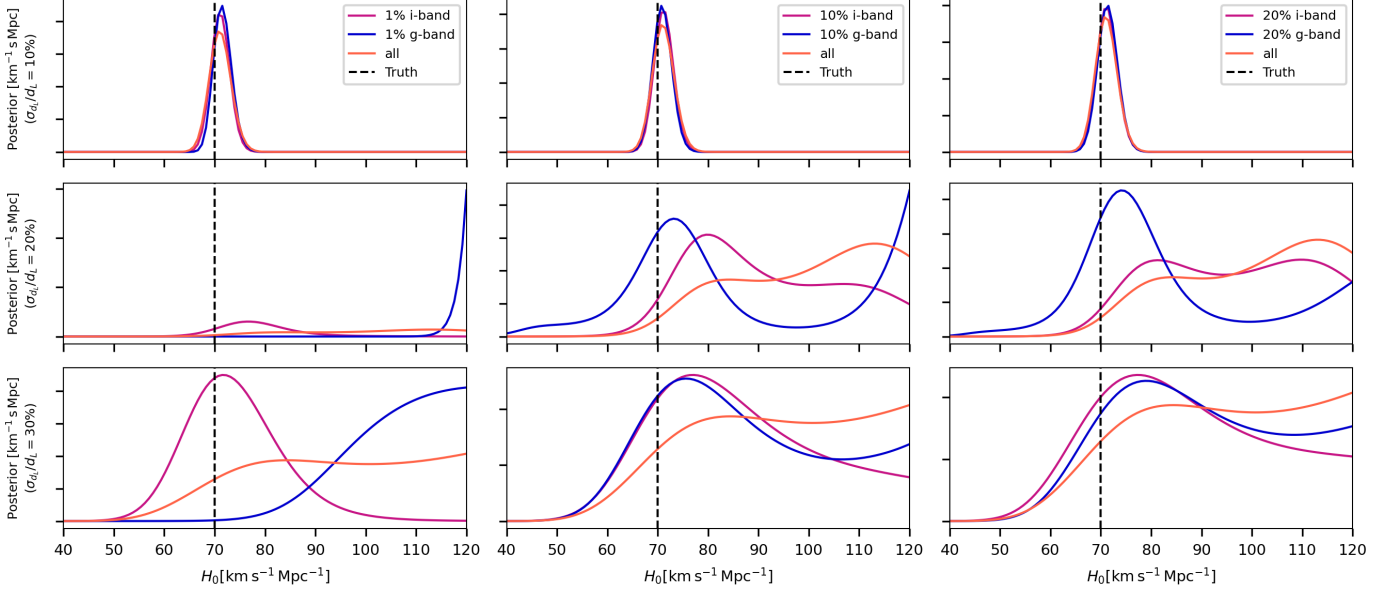


Figure 4. Resulting posteriors on H_0 from injecting 1 GW event into each of 200 LOS and reconstructing the LOS redshift distribution with the 1%, 10%, and 20% brightest galaxies in the i - and g -band.

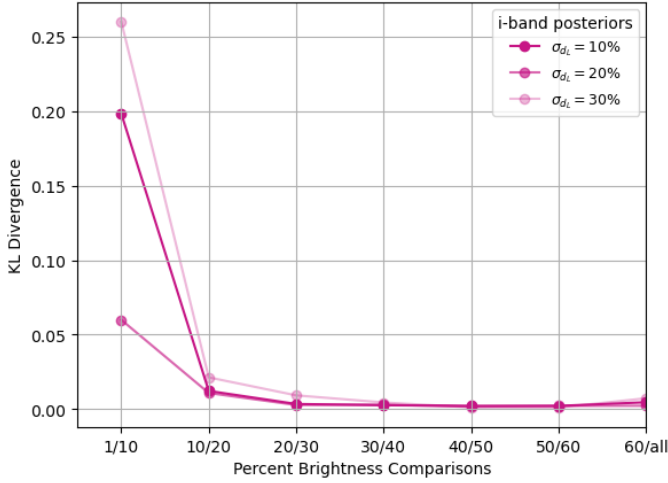


Figure 5. Median KL divergence between posteriors of increasing brightness fractions over 10 realizations of GW events. Here we show only results from posteriors constructed with the bright i -band galaxies at our three levels of d_L uncertainty.

The results of Figure 4 prompt the question: what fraction of galaxies do we actually need in order to avoid biasing our results? Figure 4 illustrates that as the brightness fraction increases, the bright galaxy posteriors begin to converge with the results from using all galaxies. To further investigate this effect, we calculate the Kullback-Leibler divergence (KL divergence) between the posteriors of increasing brightness fractions (e.g., the KL divergence between the H_0 posterior recovered with the 1% brightest versus the 10% brightest i -band galaxies) to quantify the information gained by

including additional bright galaxies. This statistic measures the difference, or relative entropy, between two probability distributions. The median KL divergences for the H_0 posteriors constructed from the bright i -band galaxies are shown in Figure 5 over ten realizations of the dark siren analysis described in Section 4.

From Figure 5 it is apparent that little information is gained after the inclusion of the 40% brightest i -band galaxies. Even adding galaxies beyond the 10% threshold leads to only a small gain in information. The limited effect of additional i -band galaxies can be explained by two factors. First, the bright i -band galaxies, even in small numbers, are good tracers of overall structure as seen in Figure 3. Second, the growth in the average number of bright i -band galaxies per LOS, shown in Figure 1, decreases as larger percentages of bright galaxies are considered, because most of the brightest i -band galaxies in the Universe are found at low redshifts.

Though not shown in Figure 5, the posteriors constructed from the bright g -band galaxies exhibit a similar behavior. For the well-localized events, there is little information gain after including the 40% brightest galaxies. However, the KL divergences between the 1% and 10% posteriors in the g -band case are much larger than the i -band galaxies. This likely occurs because there are far fewer g -band galaxies within the 1% brightness threshold at our redshifts of interest than i -band galaxies. As a result, there are realizations in which the posterior from the 1% brightest g -band galaxies is entirely uninformative. For $\sigma_{dL} > 10\%$, there are realizations in which the g -band posterior fails to converge with the full posterior even after including the 60% brightest

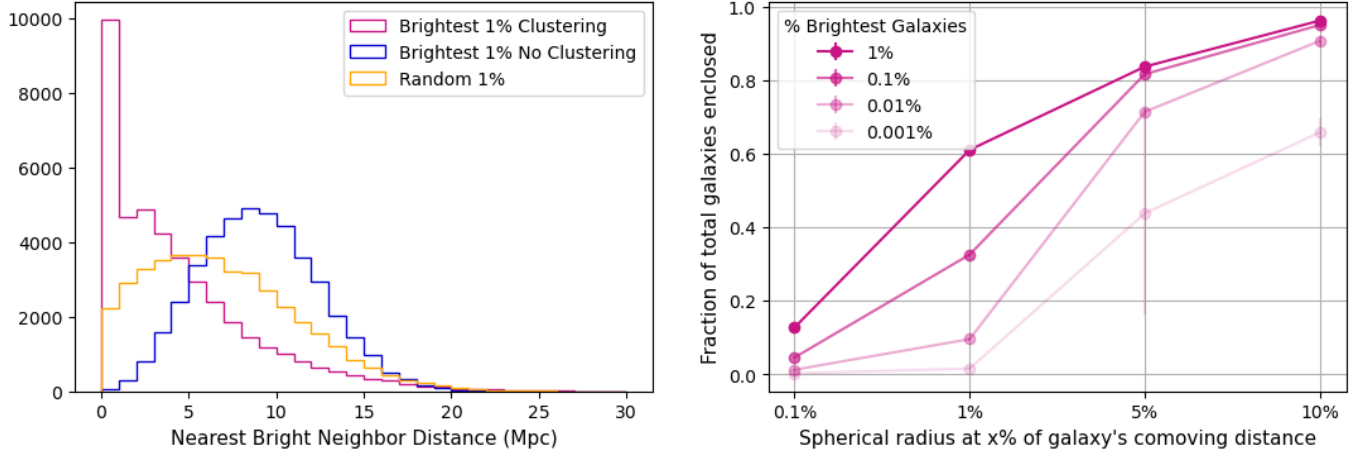


Figure 6. [Left] Distances in Mpc to the nearest bright neighbor galaxy for the 1% brightest *i*-band galaxies from MICECAT, the 1% brightest *i*-band galaxies without clustering, and 1% of galaxies drawn randomly from MICECAT, all for $z \leq 0.45$. [Right] For the 1%, 0.1%, 0.01%, and 0.001% brightest *i*-band galaxies, we draw a sphere with radius 0.1%, 1%, 5%, and 10% of that galaxy’s comoving distance and calculate the fraction of total galaxies with $z \leq 0.45$ enclosed by all the spheres. Results are averaged over three independent 20 square degree boxes in MICECAT.

galaxies. Again, even at 60% brightness, there remains an average 1000 galaxy difference per LOS between the number of bright *i*-band and *g*-band galaxies. Except for the well-localized regime, completeness levels greater than 60% are required for the H_0 posterior constructed with the bright *g*-band galaxies to converge to the posterior recovered from all galaxies.

5. IMPACT OF GALAXY CLUSTERING

The ability of an incomplete galaxy catalog to closely reproduce the posteriors constructed from a complete catalog is due to the role of galaxy clustering. Fainter galaxies tend to cluster around the most massive, brightest galaxies. In the case where the true GW host is missing from the catalog, the next closest bright galaxy is not far, as shown in the left panel of Figure 6, resulting in only a small redshift difference. For example, if the 90% faintest galaxies at $z \leq 0.4$ in the *i*-band are missing from our catalog, the fractional redshift difference between a faint galaxy and the closest (in comoving distance) bright galaxy is $\Delta z/z \leq 0.014$ for 99% of the faint galaxies. The 99th percentile of the fractional redshift difference between the 90% faintest *g*-band galaxies and bright galaxies is $\Delta z/z \leq 0.011$. These fractional redshift differences introduce such little error on H_0 that in the well-localized regime, an incomplete catalog produces a posterior just as informative as a complete one.

To better understand the role that galaxy clustering plays in the dark siren analysis, we construct a catalog that matches the properties of the MICECAT catalog but lacks galaxy clustering. We preserve the relations between redshift and *i*- and *g*-band magnitudes as in the MICECAT catalog, but randomly draw and re-assign

RA and Dec of each galaxy. This scrambles the sky positions, erasing the original large-scale structure while preserving the same trends between galaxy brightness and redshift.

The effects of clustering can be clearly seen in Figure 6. On the left, we show the nearest neighbor distances for the 1% brightest galaxies in the *i*-band, the same population but in our catalog without clustering, and a 1% of galaxies drawn randomly from MICECAT. The 1% brightest MICECAT galaxies are significantly more clustered than the other two galaxy populations, evident from the peak at nearest bright neighbor distances of a few Mpc. The plot on the right seeks to further probe just how clustered the brightest galaxies are. For a 20 square degree box in MICECAT, we begin by identifying the 1% brightest *i*-band galaxies at $z \leq 0.4$ and drawing a sphere around each bright galaxy with a radius corresponding to 10% of that galaxy’s comoving distance. We then calculate what fraction of all the galaxies in our catalog box fall within the bright-galaxy-centered spheres, considering galaxies up to $z \leq 0.45$. We begin with a radius of 10% of the galaxy’s distance, motivated by the smallest distance uncertainty considered in this study. We then considered smaller and smaller circles, at 5%, 1%, and 0.1% of the galaxy’s distance and find that large fractions of galaxies are still captured. We also consider even smaller subsets of bright galaxies, taking the 0.1%, 0.01%, and 0.001% brightest *i*-band galaxies, and find that at the largest radii tested (corresponding to typical GW localization errors), these small fractions of galaxies still capture all, or nearly all, of the galaxies in the catalog box. Figure 6 indicates that even if only a small fraction of bright galaxies are captured by a cata-

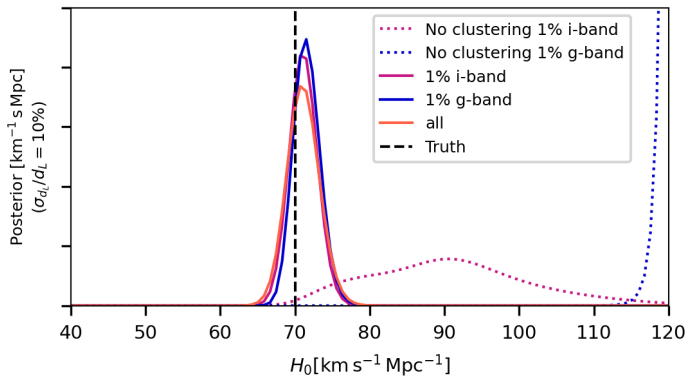


Figure 7. H_0 posterior constructed from the 1% brightest galaxies in the catalog without clustering as compared to the posteriors shown in Figure 4 for the best localized case.

log, the missing galaxies are likely found within a volume corresponding to a typical GW localization volume.

The clustering evident in Figure 6 is directly responsible for the negligible effects of catalog incompleteness seen in Figure 4 for well localized events. We further confirmed this by using our “no-clustering” catalog for a dark siren analysis. Following the same procedure as described in Section 4, we inject one GW event each into our 200 LOS and construct a posterior on H_0 using only the 1% brightest galaxies in the i and g -band to reconstruct the LOS redshift distribution. The results for the best-localized case are shown in Figure 7. Without the aid of galaxy clustering (dashed posteriors), catalog incompleteness becomes detrimental to the H_0 measurement. If the host is not present in the catalog, there are no longer other galaxies nearby providing support at the correct redshift. This leads to biased posteriors, and warrants methods to correct for incompleteness.

6. DISCUSSION

We have examined the effects of catalog incompleteness on constraining H_0 with the dark siren method. Our analysis relies on a number of assumptions to generate mock data. In this section, we will discuss the assumptions made and how they may impact the results presented in Sections 3 and 4.

The first assumption is that every galaxy is equally likely to host a GW source. When generating our mock data, we inject GW events randomly along each LOS. Though this choice allowed us to isolate the effects of catalog incompleteness, there is evidence that a galaxy’s stellar mass or star-formation rate is positively correlated with its probability of hosting a GW event (Vijaykumar et al. 2024; Adhikari et al. 2020). While our results indicate that a stellar-mass weighted galaxy sample (i.e. bright i -band galaxies) better traces the line of sight redshift distribution and is thus more useful for

a dark siren analysis, these results could change if GW sources preferentially came from star-forming (bright g -band) galaxies.

We also assume that redshifts are perfectly known. This is not the case in previous dark siren analyses, as photometric catalogs are often used due to their greater degree of completeness and sky coverage (e.g., Abbott et al. 2023; The LIGO Scientific Collaboration et al. 2025b). In this study, assuming photometric redshift errors would widen the H_0 posterior and introduce greater uncertainty to the measurement of H_0 . However, photometric catalogs will not qualitatively change our results about the effects of incompleteness. In fact, the limited impact of catalog incompleteness demonstrated in this work suggests that restricting galaxy catalogs to spectroscopic surveys for future dark siren analyses may be useful to reduce the possibility of an H_0 bias from photometric redshift errors (Turski et al. 2023).

Additionally, our choice of 1 deg radius LOSs is optimistic for current LVK capabilities. As discussed in Section 2.2, $\sim 10\%$ of candidates from O1-O4a had sky localizations less than 100 deg^2 . However, after the inclusion of LIGO-India in O5, the five-detector network (3 LIGOs plus Virgo and KAGRA) is expected to detect 38 BBHs with sky localizations $\leq 1 \text{ deg}^2$ at 90% credibility and ~ 1000 BBHs will have d_L errors of less than 10% per year (Maggiore et al. 2024). While the best-localized regime in this analysis may not be achievable with the most recent LVK observing run, it will be with the next observing run and even more so after the construction of next-generation detectors, such as Cosmic Explorer and Einstein Telescope.

We have demonstrated that only small fractions of the brightest galaxies in a GW localization volume are necessary for an unbiased dark siren analysis. How does this compare to the magnitude limits of real galaxy surveys? As an example, the DESI Luminous Red Galaxy Survey (Zhou et al. 2023) has an apparent magnitude limit of $z_{\text{fiber}} < 21.6$, where z_{fiber} is the z -band magnitude in a 1.5 arcsecond diameter centered on the galaxy. Using solar absolute magnitudes from Willmer (2018), at our threshold d_L , this corresponds to a luminosity cutoff of $3.47 \times 10^9 L_\odot$. This limit is far less conservative than the luminosity limits imposed in this study. Even surveys capturing only the brightest galaxies in the field could still be useful for a dark siren analysis. For example, the DESI bright galaxy survey’s bright sample institutes an apparent magnitude cutoff of $r < 19.5$ (Hahn et al. 2023), corresponding to a luminosity cutoff of $2.76 \times 10^{10} L_\odot$ at our threshold d_L . This is similar to the luminosity cutoff of our 10% brightest sample,

which provided more than enough galaxies to draw an informative posterior.

7. CONCLUSION

GWs are unique and useful cosmological probes. With the dark siren method it is possible to infer H_0 by treating every galaxy in the GW localization volume as a potential host. We have examined the effects of catalog incompleteness on our ability to measure H_0 using the dark siren method by injecting GW sources into a mock catalog, artificially removing galaxies, and completing a dark siren analysis using only the brightest galaxies along a given LOS.

We found that for well-localized events, there is little information loss due to catalog incompleteness. If GW hosts are found in all galaxies with host halo masses $M_h > 2 \times 10^{11} M_\odot h^{-1}$, using a catalog only complete down to the 1% brightest absolute magnitude $M_i < -22.43$ is sufficient to draw an unbiased, informative posterior on H_0 in this regime. Even smaller subsets of bright galaxies are sufficient tracers of a line of sight's redshift distribution. As measured by the KL divergence, including more than the $\sim 20\%$ brightest galaxies does not increase the information content of the H_0 posterior for well-localized events. These results are sensitive to our choice of bright i -band galaxies vs bright g -band galaxies. We found that the bright i -band galaxies, corresponding to greater stellar masses, were a better tracer of the full galaxy redshift distribution at the redshifts of GW observations, conservatively assuming that all galaxies are equally likely to host GW events. If GW events are preferentially found in stellar mass weighted galaxies, our conclusions would be amplified.

We showed that the efficacy of bright galaxies in a dark siren analysis is largely due to the influence of galaxy clustering. Faint galaxies cluster around the most massive and thus brightest galaxies, so if a faint GW host

galaxy is missing from the catalog, there is likely a bright galaxy nearby still providing support at the correct redshift. After artificially removing the galaxy clustering from our catalog, we found that using only the brightest galaxies was no longer sufficient for drawing an informative posterior.

We also conclude that it is only worth using the best-localized events in future dark siren analyses. As the localization volume increases, the effects of catalog incompleteness become noticeable as the H_0 posterior becomes more sensitive to trends in the global redshift distribution. Additionally, larger localization volumes are unable to constrain H_0 to precisions useful for cosmology. With the inclusion of LIGO-India to the current network of GW detectors in O5 and the potential of next-generation detectors, the number of well-localized events will greatly increase, providing a rich dataset for dark siren cosmology.

8. ACKNOWLEDGMENTS

This work was completed with support from the Fulbright U.S. Student Program, sponsored by the U.S. Department of State and administered by the Institute of International Education and the Foundation for Educational Exchange between Canada and the United States of America. MF acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant RGPIN-2023-05511, the University of Toronto Connaught Fund, and the Alfred P. Sloan Foundation. AV acknowledges support from the NSERC (funding reference number 568580). DEH and AGG were supported by NSF grants PHY-2110507 and PHY-2513312. DEH was also supported by the NSF-Simons AI-Institute for the Sky (SkAI) via grants NSF AST-2421845 and Simons Foundation MPS-AI-00010513, the Simons Collaboration on Black Holes and Strong Gravity, and the Kavli Institute for Cosmological Physics through an endowment from the Kavli Foundation.

REFERENCES

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, *PhRvL*, 116, 061102, doi: [10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102)
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, *Phys. Rev. Lett.*, 119, 161101, doi: [10.1103/PhysRevLett.119.161101](https://doi.org/10.1103/PhysRevLett.119.161101)
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, *ApJL*, 848, L12, doi: [10.3847/2041-8213/aa91c9](https://doi.org/10.3847/2041-8213/aa91c9)
- . 2017b, *Nature*, 551, 85, doi: [10.1038/nature24471](https://doi.org/10.1038/nature24471)
- Abbott, R., Abe, H., Acernese, F., et al. 2023, *ApJ*, 949, 76, doi: [10.3847/1538-4357/ac74bb](https://doi.org/10.3847/1538-4357/ac74bb)
- Adhikari, S., Fishbach, M., Holz, D. E., Wechsler, R. H., & Fang, Z. 2020, *ApJ*, 905, 21, doi: [10.3847/1538-4357/abbfb7](https://doi.org/10.3847/1538-4357/abbfb7)
- Auclair, P., Bacon, D., Baker, T., et al. 2023, *Living Reviews in Relativity*, 26, 5, doi: [10.1007/s41114-023-00045-2](https://doi.org/10.1007/s41114-023-00045-2)
- Beirnaert, F., Dálya, G., & Ghosh, A. 2025, *MNRAS*, doi: [10.1093/mnras/staf1432](https://doi.org/10.1093/mnras/staf1432)
- Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, *ApJS*, 149, 289, doi: [10.1086/378847](https://doi.org/10.1086/378847)

- Bera, S., Rana, D., More, S., & Bose, S. 2020, *ApJ*, 902, 79, doi: [10.3847/1538-4357/abb4e0](https://doi.org/10.3847/1538-4357/abb4e0)
- Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, *ApJ*, 592, 819, doi: [10.1086/375776](https://doi.org/10.1086/375776)
- Bom, C. R., Alfradique, V., Palmese, A., et al. 2024, *MNRAS*, 535, 961, doi: [10.1093/mnras/stae2390](https://doi.org/10.1093/mnras/stae2390)
- Borghi, N., Moresco, M., Tagliazucchi, M., & Cuomo, G. 2025, *Echoes from the Dark: Galaxy Catalog Incompleteness in Standard Siren Cosmology*. <https://arxiv.org/abs/2509.18243>
- Branchesi, M., Maggiore, M., Alonso, D., et al. 2023, *JCAP*, 2023, 068, doi: [10.1088/1475-7516/2023/07/068](https://doi.org/10.1088/1475-7516/2023/07/068)
- Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, *MNRAS*, 447, 646, doi: [10.1093/mnras/stu2402](https://doi.org/10.1093/mnras/stu2402)
- Chatterjee, D., Hegade K. R., A., Holder, G., et al. 2021, *PhRvD*, 104, 083528, doi: [10.1103/PhysRevD.104.083528](https://doi.org/10.1103/PhysRevD.104.083528)
- Chen, H.-Y., Fishbach, M., & Holz, D. E. 2018, *Nature*, 562, 545, doi: [10.1038/s41586-018-0606-0](https://doi.org/10.1038/s41586-018-0606-0)
- Chen, H.-Y., & Holz, D. E. 2016, *arXiv e-prints*, arXiv:1612.01471, doi: [10.48550/arXiv.1612.01471](https://doi.org/10.48550/arXiv.1612.01471)
- Coulter, D. A., Foley, R. J., Kilpatrick, C. D., et al. 2017, *Science*, 358, 1556, doi: [10.1126/science.aap9811](https://doi.org/10.1126/science.aap9811)
- Cousins, B., Schumacher, K., Ka-Wai Chung, A., et al. 2025, *arXiv e-prints*, arXiv:2503.01997, doi: [10.48550/arXiv.2503.01997](https://doi.org/10.48550/arXiv.2503.01997)
- Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, *MNRAS*, 453, 1513, doi: [10.1093/mnras/stv1708](https://doi.org/10.1093/mnras/stv1708)
- Dalang, C., & Baker, T. 2024, *Journal of Cosmology and Astroparticle Physics*, 2024, 024, doi: [10.1088/1475-7516/2024/02/024](https://doi.org/10.1088/1475-7516/2024/02/024)
- Dalang, C., Fiorini, B., & Baker, T. 2024, *Large scale structure prior knowledge in the dark siren method*. <https://arxiv.org/abs/2410.03275>
- Dálya, G., Díaz, R., Bouchet, F. R., et al. 2022, *MNRAS*, 514, 1403, doi: [10.1093/mnras/stac1443](https://doi.org/10.1093/mnras/stac1443)
- Del Pozzo, W. 2012, *PhRvD*, 86, 043011, doi: [10.1103/PhysRevD.86.043011](https://doi.org/10.1103/PhysRevD.86.043011)
- Essick, R., & Fishbach, M. 2024, *ApJ*, 962, 169, doi: [10.3847/1538-4357/ad1604](https://doi.org/10.3847/1538-4357/ad1604)
- Evans, M., Adhikari, R. X., Afle, C., et al. 2021, *arXiv e-prints*, arXiv:2109.09882, doi: [10.48550/arXiv.2109.09882](https://doi.org/10.48550/arXiv.2109.09882)
- Ezquiaga, J. M., & Holz, D. E. 2022, *PhRvL*, 129, 061102, doi: [10.1103/PhysRevLett.129.061102](https://doi.org/10.1103/PhysRevLett.129.061102)
- Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, *ApJL*, 883, L42, doi: [10.3847/2041-8213/ab4284](https://doi.org/10.3847/2041-8213/ab4284)
- Finke, A., Foffa, S., Iacovelli, F., Maggiore, M., & Mancarella, M. 2021, *JCAP*, 2021, 026, doi: [10.1088/1475-7516/2021/08/026](https://doi.org/10.1088/1475-7516/2021/08/026)
- Fishbach, M., Gray, R., Magaña Hernandez, I., et al. 2019, *ApJL*, 871, L13, doi: [10.3847/2041-8213/aaf96e](https://doi.org/10.3847/2041-8213/aaf96e)
- Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015a, *MNRAS*, 448, 2987, doi: [10.1093/mnras/stv138](https://doi.org/10.1093/mnras/stv138)
- Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015b, *MNRAS*, 447, 1319, doi: [10.1093/mnras/stu2464](https://doi.org/10.1093/mnras/stu2464)
- Freedman, W. L., & Madore, B. F. 2023, *JCAP*, 2023, 050, doi: [10.1088/1475-7516/2023/11/050](https://doi.org/10.1088/1475-7516/2023/11/050)
- Gair, J. R., Ghosh, A., Gray, R., et al. 2023, *AJ*, 166, 22, doi: [10.3847/1538-3881/acca78](https://doi.org/10.3847/1538-3881/acca78)
- Gray, R., Hernandez, I. M., Qi, H., et al. 2020, *PhRvD*, 101, 122001, doi: [10.1103/PhysRevD.101.122001](https://doi.org/10.1103/PhysRevD.101.122001)
- Gray, R., Beirnaert, F., Karathanasis, C., et al. 2023, *Journal of Cosmology and Astroparticle Physics*, 2023, 023, doi: [10.1088/1475-7516/2023/12/023](https://doi.org/10.1088/1475-7516/2023/12/023)
- Hahn, C., Wilson, M. J., Ruiz-Macias, O., et al. 2023, *AJ*, 165, 253, doi: [10.3847/1538-3881/accff8](https://doi.org/10.3847/1538-3881/accff8)
- Hanselman, A. G., Vijaykumar, A., Fishbach, M., & Holz, D. E. 2024, *arXiv e-prints*, arXiv:2405.14818, doi: [10.48550/arXiv.2405.14818](https://doi.org/10.48550/arXiv.2405.14818)
- Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, *MNRAS*, 447, 1724, doi: [10.1093/mnras/stu2492](https://doi.org/10.1093/mnras/stu2492)
- Holz, D. E., & Hughes, S. A. 2005, *ApJ*, 629, 15, doi: [10.1086/431341](https://doi.org/10.1086/431341)
- Leyde, K., Baker, T., & Enzi, W. 2024, *Journal of Cosmology and Astroparticle Physics*, 2024, 013, doi: [10.1088/1475-7516/2024/12/013](https://doi.org/10.1088/1475-7516/2024/12/013)
- . 2025, *Cosmic Cartography II: completing galaxy catalogs for gravitational-wave cosmology*. <https://arxiv.org/abs/2507.12171>
- MacLeod, C. L., & Hogan, C. J. 2008, *PhRvD*, 77, 043512, doi: [10.1103/PhysRevD.77.043512](https://doi.org/10.1103/PhysRevD.77.043512)
- Madau, P., & Dickinson, M. 2014, *ARA&A*, 52, 415, doi: [10.1146/annurev-astro-081811-125615](https://doi.org/10.1146/annurev-astro-081811-125615)
- Maggiore, M., Iacovelli, F., Belgacem, E., Mancarella, M., & Muttoni, N. 2024, *arXiv e-prints*, arXiv:2411.05754, doi: [10.48550/arXiv.2411.05754](https://doi.org/10.48550/arXiv.2411.05754)
- Maggiore, M., Van Den Broeck, C., Bartolo, N., et al. 2020, *JCAP*, 2020, 050, doi: [10.1088/1475-7516/2020/03/050](https://doi.org/10.1088/1475-7516/2020/03/050)
- Mastrogiovanni, S., Leyde, K., Karathanasis, C., et al. 2021, *Phys. Rev. D*, 104, 062009, doi: [10.1103/PhysRevD.104.062009](https://doi.org/10.1103/PhysRevD.104.062009)
- Mastrogiovanni, S., Laghi, D., Gray, R., et al. 2023, *Phys. Rev. D*, 108, 042002, doi: [10.1103/PhysRevD.108.042002](https://doi.org/10.1103/PhysRevD.108.042002)
- Messenger, C., & Read, J. 2012, *PhRvL*, 108, 091101, doi: [10.1103/PhysRevLett.108.091101](https://doi.org/10.1103/PhysRevLett.108.091101)

- Moresco, M., et al. 2022, *Living Rev. Rel.*, 25, 6, doi: [10.1007/s41114-022-00040-z](https://doi.org/10.1007/s41114-022-00040-z)
- Moresco, M., Amati, L., Amendola, L., et al. 2022, *Living Reviews in Relativity*, 25, 6, doi: [10.1007/s41114-022-00040-z](https://doi.org/10.1007/s41114-022-00040-z)
- Mukherjee, S., Krolewski, A., Wandelt, B. D., & Silk, J. 2024, *ApJ*, 975, 189, doi: [10.3847/1538-4357/ad7d90](https://doi.org/10.3847/1538-4357/ad7d90)
- Mukherjee, S., Wandelt, B. D., Nissanke, S. M., & Silvestri, A. 2021, *PhRvD*, 103, 043520, doi: [10.1103/PhysRevD.103.043520](https://doi.org/10.1103/PhysRevD.103.043520)
- Namikawa, T., Nishizawa, A., & Taruya, A. 2016, *PhRvL*, 116, 121302, doi: [10.1103/PhysRevLett.116.121302](https://doi.org/10.1103/PhysRevLett.116.121302)
- Naveed, K., Turski, C., & Ghosh, A. 2025, arXiv e-prints, arXiv:2505.11268, doi: [10.48550/arXiv.2505.11268](https://doi.org/10.48550/arXiv.2505.11268)
- Palmese, A., Bom, C. R., Mucesh, S., & Hartley, W. G. 2023, *ApJ*, 943, 56, doi: [10.3847/1538-4357/aca6e3](https://doi.org/10.3847/1538-4357/aca6e3)
- Palmese, A., deVicente, J., Pereira, M. E. S., et al. 2020, *ApJL*, 900, L33, doi: [10.3847/2041-8213/abaeff](https://doi.org/10.3847/2041-8213/abaeff)
- Perna, G., Mastrogiovanni, S., & Ricciardone, A. 2024, arXiv e-prints, arXiv:2405.07904, doi: [10.48550/arXiv.2405.07904](https://doi.org/10.48550/arXiv.2405.07904)
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6, doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910)
- Riess, A. G., Casertano, S., Yuan, W., et al. 2021, *ApJL*, 908, L6, doi: [10.3847/2041-8213/abdbaf](https://doi.org/10.3847/2041-8213/abdbaf)
- Schutz, B. F. 1986, *Nature*, 323, 310, doi: [10.1038/323310a0](https://doi.org/10.1038/323310a0)
- Scolnic, D., Brout, D., Carr, A., et al. 2022, *ApJ*, 938, 113, doi: [10.3847/1538-4357/ac8b7a](https://doi.org/10.3847/1538-4357/ac8b7a)
- Soares-Santos, M., Holz, D. E., Annis, J., et al. 2017, *ApJL*, 848, L16, doi: [10.3847/2041-8213/aa9059](https://doi.org/10.3847/2041-8213/aa9059)
- Soares-Santos, M., Palmese, A., Hartley, W., et al. 2019, *ApJL*, 876, L7, doi: [10.3847/2041-8213/ab14f1](https://doi.org/10.3847/2041-8213/ab14f1)
- Taylor, S. R., & Gair, J. R. 2012, *PhRvD*, 86, 023502, doi: [10.1103/PhysRevD.86.023502](https://doi.org/10.1103/PhysRevD.86.023502)
- The LIGO Scientific Collaboration, the Virgo Collaboration, & the KAGRA Collaboration. 2025a, arXiv e-prints, arXiv:2509.04348, doi: [10.48550/arXiv.2509.04348](https://doi.org/10.48550/arXiv.2509.04348)
- The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. 2025b, arXiv e-prints, arXiv:2509.04348, doi: [10.48550/arXiv.2509.04348](https://doi.org/10.48550/arXiv.2509.04348)
- Turski, C., Bilicki, M., Dálya, G., Gray, R., & Ghosh, A. 2023, *MNRAS*, 526, 6224, doi: [10.1093/mnras/stad3110](https://doi.org/10.1093/mnras/stad3110)
- Uddin, S. A., Burns, C. R., Phillips, M. M., et al. 2023, arXiv e-prints, arXiv:2308.01875, doi: [10.48550/arXiv.2308.01875](https://doi.org/10.48550/arXiv.2308.01875)
- Vijaykumar, A., Fishbach, M., Adhikari, S., & Holz, D. E. 2024, *ApJ*, 972, 157, doi: [10.3847/1538-4357/ad6140](https://doi.org/10.3847/1538-4357/ad6140)
- Willmer, C. N. A. 2018, *ApJS*, 236, 47, doi: [10.3847/1538-4365/aabfdf](https://doi.org/10.3847/1538-4365/aabfdf)
- Zhou, R., Dey, B., Newman, J. A., et al. 2023, *AJ*, 165, 58, doi: [10.3847/1538-3881/aca5fb](https://doi.org/10.3847/1538-3881/aca5fb)