

BRIGHT GALAXIES AS REDSHIFT TRACERS FOR DARK SIREN COSMOLOGY

Muhammad Khuzifa Naveed

Student number: 02004453

Supervisors: Prof. Dr. Archisman Ghosh, Cezary Turski

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of
Master of Science in Physics and Astronomy

Academic year: 2024 - 2025



Department of Physics and Astronomy
Ghent Gravity Group

Bright Galaxies as Redshift Tracers for Dark Siren Cosmology

Muhammad Khuzaifa Naveed

Promotor Prof. Dr. Archisman Ghosh
Day-to-day supervisor Cezary Turski

AY 2024-2025

Muhammad Khuzaifa Naveed

Bright Galaxies as Redshift Tracers for Dark Siren Cosmology

Master Thesis, AY 2024-2025

Promotor: Prof. Dr. Archisman Ghosh

Day-to-day supervisor: Cezary Turski

Ghent University

Ghent Gravity Group

Department of Physics and Astronomy

Proeftuinstraat 86

9000 Ghent

Declaration

I hereby declare that this thesis is the result of my own work and has not been submitted for any degree or qualification at any other academic institution. All sources of information used have been appropriately acknowledged and referenced.

I affirm that the research, analysis, and implementation described in this thesis were conducted independently. To improve the clarity, grammar, and structure of the text, AI-based language models (such as ChatGPT and Gemini) were used for language refinement. Furthermore, an AI tool was employed to generate code for one specific figure within this thesis, focusing on its visual representation rather than the underlying data analysis or scientific interpretation. The intellectual contributions, scientific insights, and conclusions presented in this work are entirely my own.

Ghent, May 31, 2025

Muhammad Khuzaifa Naveed

Acknowledgement

Muhammad Khuzaifa Naveed

Abstract

Gravitational-wave (GW) observations have introduced a powerful, independent method to measure the Hubble constant, H_0 , through the use of so-called *standard sirens*. This thesis investigates the use of *dark sirens*, GW events without electromagnetic counterparts, for cosmological inference. In particular, it explores whether selecting the brightest galaxies as redshift tracers can improve the redshift prior and enhance the precision of H_0 estimation. By focusing on the most luminous galaxies, we partially mitigate the effects of catalog incompleteness currently plaguing dark siren measurements, effectively extending the reach of the catalog for cosmological analysis. Using the `gwcosmo` inference pipeline and the GLADE+ galaxy catalog, a series of brightness-ranked percentiles were tested on a subset of the GWTC-3 catalog. The results show that moderate pruning improves H_0 constraints, while aggressive cuts lead to information loss and potential bias. Simulated mock data challenges using the BUZZARD catalog support this approach and reveal a lower pruning threshold near 30% of the brightest galaxies. This method enhances the utility of incomplete galaxy catalogs in dark siren cosmology and contributes to the broader effort to resolve the Hubble tension.

Samenvatting

Zwaartekrachtgolven (GW) vormen een krachtig en onafhankelijk middel om de Hubble-constante, H_0 , te meten via de zogenaamde *standaard-sirenes*. Deze scriptie onderzoekt het gebruik van *donkere sirenes*, GW-waarnemingen zonder elektromagnetische tegenhanger, voor kosmologische inferentie. In het bijzonder wordt onderzocht of het kiezen van de helderste sterrenstelsels als roodverschuivingstracers de z -prior kan verbeteren en zo de precisie van de H_0 -schatting kan verhogen. Door ons te richten op de lichtkrachtigste sterrenstelsels wordt de impact van de onvolledigheid van sterrenstelscatalogi, een bekend probleem in donkere siren-analyses, gedeeltelijk verminderd, waardoor de effectieve diepte van de catalogus toeneemt. Met behulp van de `gwcosmo` pipeline en de GLADE+ catalogus werd een reeks helderheidspercentielen getest op een subset van de GWTC-3 catalogus. De resultaten tonen aan dat gematigde afkappingen de H_0 meeting verbeteren, terwijl agressieve afkappingen leiden tot verlies van informatie en een mogelijke bias. Gesimuleerde Mock Data Challenges met de BUZZARD catalogus ondersteunen deze aanpak en onthullen een praktische ondergrens van ongeveer 30% van de helderste sterrenstelsels. Deze methode vergroot de bruikbaarheid van onvolledige sterrenstelscatalogi in donkere siren-kosmologie en draagt bij aan de bredere inspanning om de Hubble-spanning op te lossen.

Contents

List of Figures	xiii
List of Tables	xv
List of Abbreviations	xvii
1 Introduction	1
1.1 The Hubble Constant	1
1.1.1 Early H_0 Measurements	1
1.2 Modern Cosmology and the Hubble Tension	2
1.2.1 Local measurements: Cosmic Distance Ladder	2
1.2.2 Indirect measurements: Cosmic Microwave Background	3
1.2.3 The Hubble Tension and quest for independent measurements	4
1.3 Gravitational-Wave (GW) Cosmology	5
1.4 Thesis Objectives	6
2 Gravitational-Wave Cosmology	9
2.1 Standard Siren Cosmology	9
2.2 Bright vs. Dark Sirens	11
2.3 State of the Art Dark Siren Methods	12
2.3.1 Redshift Inference Methods	13
2.3.2 Galaxy Weighting Choices	13
2.4 Refining the Redshift Prior: Brightest Galaxies as Tracers	15
3 The gwcosmo Pipeline	17
3.1 Bayesian Framework	17
3.2 Key Features of gwcosmo	20
3.3 Implementation and Applications	20
4 Data	23
4.1 GW Event Data	23
4.1.1 The GWTC-3 Catalog	23
4.1.2 Event Selection Criteria	23
4.1.3 Distance Posteriors and Sky Localization	24
4.1.4 Assumptions About the Source Population	24
4.2 Electromagnetic (EM) Galaxy Catalogs	25
4.2.1 The GLADE+ Galaxy Catalog	25
4.2.2 Catalog Completeness	25
4.2.3 Redshift Uncertainty Modeling	26

4.2.4	<i>K</i> -band Luminosity as Host Probability Proxy	26
5	GLADE+ Bright Galaxies as Redshift Tracers	27
5.1	Bright Galaxy Subsets	27
5.1.1	Trade-Offs and Limitations	28
5.2	Line-of-sight (LOS) Redshift Prior	31
5.2.1	LOS Redshift Prior Construction	31
5.3	Results	33
5.3.1	LOS Redshift Prior	33
5.3.2	Hubble Constant Posterior	35
5.3.3	Luminosity Retention	36
5.3.4	Cost-Benefit Trade-Off	37
6	Mock Data Challenge	39
6.1	The BUZZARD Mock Catalog	39
6.2	GWSim	42
6.3	H_0 Inference with <code>gwcosmo</code>	43
6.4	Results	44
6.4.1	Luminosity Retention in the BUZZARD Catalog	44
6.4.2	Tracing Large-Scale Structure with Luminous Galaxies	44
7	Conclusion	49
7.1	Summary of Findings	49
7.2	Limitations	50
7.3	Outlook and Future Work	50
	Bibliography	51

List of Figures

1.1	Hubble tension: local vs. cosmic microwave background (CMB)-based H_0 measurements.	5
1.2	Hubble tension and standard sirens.	6
2.1	The B_J -band and K -band luminosity function for the GLADE+ catalog.	14
2.2	Illustration of the proposed method, using bright galaxies as redshift tracers.	15
5.1	Normalized redshift distributions $p(z)$ for the GLADE+ galaxy catalogue and the brightest percentiles.	28
5.2	Spatial distribution of galaxies from the GLADE+ galaxy catalog and its subsets.	30
5.3	LOS redshift prior for GW170809 for different brightness-ranked GLADE subsets	34
5.4	H_0 posterior distributions for GW170809 using full and brightness-cut galaxy catalogs.	34
5.5	Cumulative H_0 posterior from the selected dark siren events using full GLADE+ and the different subsets.	35
5.6	GLADE+ luminosity retention.	37
6.1	BUZZARD luminosity retention.	44
6.2	r_{halo} distribution and evolution for BUZZARDm18 and its percentiles.	45

List of Tables

4.1	Priors on merger rate shape parameters.	24
4.2	Priors on the mass model parameters.	25
5.1	H_0 MAP values with 68% confidence ranges, alongside the maximum magnitude limits for GLADE+ and the different subsets.	36
6.1	BUZZARD cosmology	40
6.2	Central galaxy fraction for different m_{thr} for BUZZARD.	46

List of Abbreviations

ADDGALS	Adding Density-Determined Galaxies to Lightcone Simulations	40
BAO	Baryon Acoustic Oscillation	4
BBH	binary black hole	10
BCG	Brightest Cluster Galaxy	16
BNS	binary neutron star	10
CBC	compact binary coalescence	9
CMB	cosmic microwave background	xiii
DES	Dark Energy Survey	29
EM	Electromagnetic	xi
GW	Gravitational-Wave	xi
GWTC	Gravitational-Wave Transient Catalog	6
HST	Hubble Space Telescope	2
JWST	James Webb Space Telescope	3
KDE	kernel density estimation	20
LOS	line-of-sight	xii
LWK	LIGO, Virgo, KAGRA	6
MDC	Mock Data Challenge	29
MNRAS	Monthly Notices of the Royal Astronomical Society	27
NSBH	neutron star black hole	10
SDSS	Sloan Digital Sky Survey	40
SHOES	Supernovae H0 for the Equation of State	3
SNe	supernovae	9
SNe Ia	Type Ia supernovae	2

SNR	signal-to-noise ratio	7
TRGB	Tip of the Red Giant Branch	4
WISE	Wide-Field Infrared Survey Explorer	25

Introduction

1.1 The Hubble Constant

The Hubble constant, denoted H_0 , quantifies the present-day expansion rate of the Universe. An empirical relationship between galaxy redshift and distance was published by Edwin Hubble in 1929 (Hubble, 1929), showing that galaxies recede from us at velocities proportional to their distances. A few years earlier, Georges Lemaître had independently derived a similar velocity-distance relation in 1927 (Lemaître, 1927), based on theoretical grounds using general relativity and available data. This linear relation, now known as Hubble's law, is expressed as:

$$v = H_0 d \quad (1.1)$$

where v is the recession velocity and d is the proper distance to the galaxy. This relationship forms the foundation of modern observational cosmology, and established the expanding Universe as a cornerstone of modern cosmology laying the foundation for the standard cosmological model.

In modern terms, the Hubble constant also governs the shape of the luminosity distance-redshift relation and determines the age and size of the Universe. In a flat Λ CDM cosmology, the Hubble constant enters the luminosity distance formula as:

$$d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda}}. \quad (1.2)$$

Accurate measurements of H_0 are essential for understanding the expansion history of the Universe and for anchoring other cosmological parameters. However, different observational methods have yielded inconsistent values, leading to a discrepancy, what is now known as the *Hubble tension*.

1.1.1 Early H_0 Measurements

The quest to measure the Hubble constant has a long history, beginning with Hubble's original work in the 1920s. The first estimates of H_0 were based on the apparent brightness of Cepheid variable stars in nearby galaxies, which were used as standard candles (Hubble, 1929). These early measurements were limited by the available technology and the uncertainties in distance measurements.

In the 1930s, Edwin Hubble and Milton Humason measured redshifts and estimated distances to additional galaxies at Mount Wilson Observatory. Their work led to an initial estimate of $H_0 \approx 500 \text{ km s}^{-1} \text{ Mpc}^{-1}$, significantly higher than current values (Hubble, 1936). The discrepancy was largely due to underestimating galaxy distances and misidentifying Cepheid populations.

Throughout the 1940s and 1950s, astronomers like Walter Baade and Fritz Zwicky introduced new distance indicators, such as the use of Population I and II stars, and proposed Type Ia supernovae as standard candles (Zwicky, 1942; Baade, 1979a; Baade, 1979b). Baade's recalibration of Cepheid distances effectively halved Hubble's original value, leading to estimates closer to $H_0 \approx 250 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Baade, 1979a; Longair, 2006).

Later on, the advent of techniques such as surface brightness fluctuations and the discovery of quasars enabled distance measurements over greater cosmological scales. However, large discrepancies persisted, with estimates of H_0 ranging from 50 to $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Sandage, 1958; Vaucouleurs, 1972; Sandage and Tammann, 1982; De Vaucouleurs, 1985).

With the advent of the 21st century and significant advancements in both space- and ground-based observational capabilities, including high-resolution radio telescopes and precision optical instruments, two independent and fundamentally different methods for measuring the Hubble constant emerged. One approach relies on the cosmic distance ladder, anchored by Cepheid variables and Type Ia supernovae, while the other infers H_0 from early-Universe observations of the CMB, interpreted through the standard Λ CDM cosmological model.

1.2 Modern Cosmology and the Hubble Tension

1.2.1 Local measurements: Cosmic Distance Ladder

The traditional approach to measuring H_0 is the *cosmic distance ladder*, which combines a series of interconnected distance indicators. The ladder begins with geometric parallax for nearby stars, continues with Cepheid variable stars in more distant galaxies, and culminates in the use of Type Ia supernovae (SNe Ia) as standard candles at cosmological distances. Each rung of the ladder calibrates the next, enabling a distance-redshift relation that extends to hundreds of megaparsecs (Freedman et al., 2001; Freedman and Madore, 2010; Riess et al., 2019; Riess et al., 2022).

The 1980s and 1990s saw significant advancements in distance measurement techniques, including the use of the *Hubble Space Telescope (HST)* and improved calibration methods. The introduction of HST in 1990 allowed for more precise measurements of Cepheid variables in nearby galaxies, leading to a more accurate determination of H_0 . The HST Key Project, led by Freedman et al., used Cepheids in host galaxies of Type Ia supernovae to derive a value

of $H_0 = 72 \pm 8 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Freedman et al., 2001). This result significantly improved both the precision and reliability of local distance measurements.

The Supernovae H0 for the Equation of State (SH0ES) project, which built upon the HST Key Project, further refined the H_0 measurement using a larger sample of Cepheid-calibrated supernovae. Recently, in 2022, they reported a value of $H_0 = 73.30 \pm 1.04 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Riess et al., 2022). More recently, results from the James Webb Space Telescope (JWST) have also emerged (Freedman et al., 2024). These results, while somewhat consistent with previous measurements using data from both HST and JWST, lean towards the CMB-based values using only data from JWST, suggesting possible systematic errors in the earlier Cepheid calibration or the underlying assumptions of the cosmic distance ladder. Around the same time, Riess et al. (2024) used early JWST data to rule out systematic errors in Cepheid photometry as the source of the Hubble tension, providing further confidence in the HST Cepheid-based measurements, while further complicating the situation.

While the cosmic distance ladder achieves high precision, it involves multiple sources of systematic uncertainty at each step including stellar evolution models, metallicity corrections, dust extinction, and supernova standardization, all of which must be carefully calibrated and validated. With the advent of new observational techniques and instruments, the cosmic distance ladder continues to evolve, but it remains a complex and intricate process that requires careful consideration of systematic uncertainties.

1.2.2 Indirect measurements: Cosmic Microwave Background

The second approach to measuring H_0 relies on the *cosmic microwave background (CMB)* radiation, which provides a snapshot of the Universe at a much earlier time. The CMB is a relic radiation from the hot, dense state of the early Universe, and its temperature fluctuations encode information about the density and expansion rate of the Universe.

By fitting the observed power spectrum to the standard ΛCDM model, one can infer a value of the Hubble constant indirectly. This method relies on model assumptions, especially the flatness of the Universe and the constancy of dark energy. While not a direct measurement, the CMB-based value of H_0 is internally consistent and extremely precise.

The introduction of the *Planck* satellite in 2009 provided a new avenue for measuring H_0 through observations of the CMB. The CMB measurements, combined with the standard ΛCDM model, most recently yielded a value of $H_0 = 67.4 \pm 0.5 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Aghanim et al., 2020). This value is significantly lower than the local measurements obtained from Cepheid-calibrated supernovae, leading to the emergence of the so-called *Hubble tension*.

1.2.3 The Hubble Tension and quest for independent measurements

The Hubble tension, now exceeding 5σ , is one of the most significant unresolved issues in cosmology and has prompted investigations into both systematic errors and possible extensions to the standard cosmological model (Kamionkowski and Riess, 2023). Independent methods are therefore crucial for arbitrating between these measurements. These include:

- **Tip of the Red Giant Branch (TRGB)**

The TRGB method uses the well-defined luminosity at which low-mass stars ignite helium as a standard candle. This luminosity, shows minimal sensitivity to stellar population effects and dust, making it a robust alternative to Cepheid variables. The Carnegie-Chicago Hubble Program has applied this method to derive an independent calibration of Type Ia supernovae and obtained H_0 values closer to those inferred from the CMB (Freedman et al., 2024). As such, TRGB offers a valuable cross-check on the Cepheid-based ladder.

- **Baryon Acoustic Oscillations (BAOs)**

BAOs are imprints of sound waves in the early Universe, visible as a characteristic scale in the clustering of galaxies. This scale acts as a standard ruler and allows geometric distance measurements across cosmic time. When combined with redshift data and a calibration from the sound horizon measured by the CMB, BAOs can be used to infer H_0 (Cuceu et al., 2019; Alam et al., 2021). Although BAOs are not purely local measurements, they offer a powerful, independent cosmological probe that is less susceptible to astrophysical systematics.

- **Gravitational-wave standard sirens**

The use of gravitational-wave events as *standard sirens* (Schutz, 1986) provides a unique opportunity to measure H_0 independently. The GW sources provide a direct measurement of the luminosity distance through their waveform. When combined with a redshift, obtained either via EM counterparts or galaxy catalogs, these events can trace the distance-redshift relation and yield an independent estimate of the Hubble constant.

Standard sirens can be categorized into two broad categories: bright sirens and dark sirens, depending on whether an EM counterpart is detected. This thesis explores one implementation of this approach, focusing on *dark sirens*: gravitational-wave events without identified electromagnetic counterparts. By statistically associating these events with potential host galaxies in wide-field catalogs, we aim to extract constraints on H_0 . In particular, we investigate how restricting the galaxy catalog to its brightest subsets can improve the resulting cosmological inference.

Together, these emerging methodologies, TRGB, BAO, and standard sirens, form a triangulation of H_0 that spans distinct epochs and assumptions. Whether the Hubble tension reflects unknown systematics or fundamentally new physics, its resolution will likely come from the convergence (or divergence) of these independent measurements.

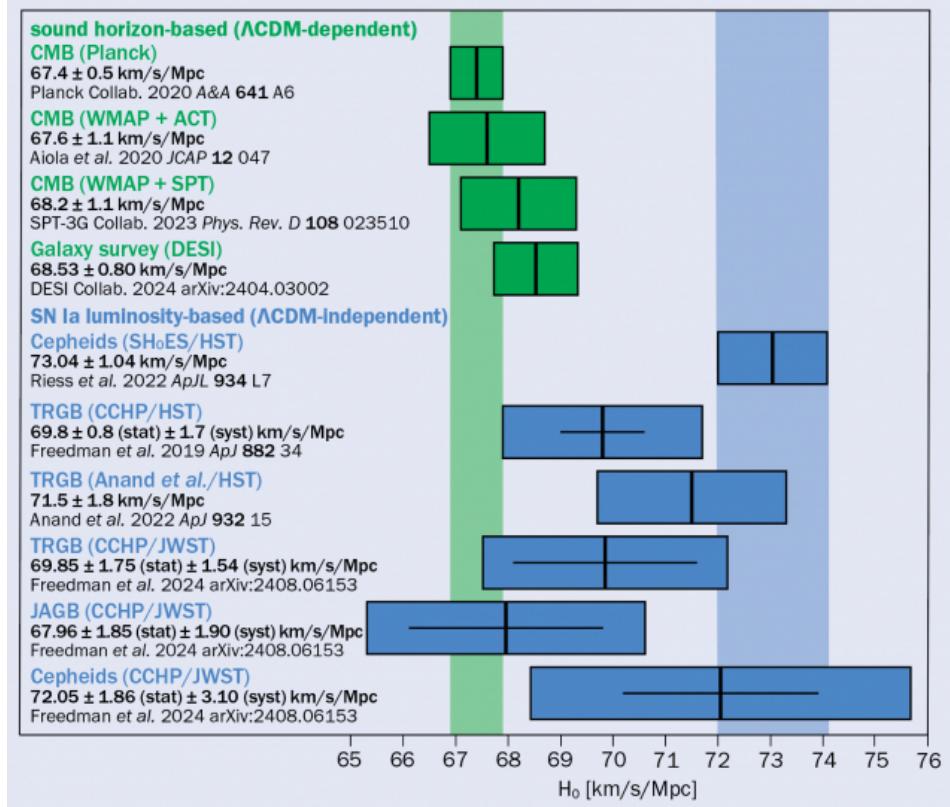


Fig. 1.1: Hubble tension: Comparison of the local and CMB-based measurements of H_0 over the years (CERN, 2025).

1.3 GW Cosmology

GW observations have introduced a novel, independent method for measuring cosmological distances: the so-called *standard sirens*. First proposed by Schutz (1986), standard sirens are the GW analogs of EM standard candles, such as Type Ia supernovae. Unlike EM methods, standard sirens enable a direct, calibration-free determination of the luminosity distance from the observed waveform of a compact binary coalescence.

The amplitude and frequency evolution of the GW signal encodes the luminosity distance, which can be inferred under the assumption of general relativity. However, GW detectors are not sensitive to the redshift of the source. Therefore, to place a GW event on the Hubble diagram and infer the Hubble constant H_0 , one must independently obtain the redshift.

In the case of bright sirens, events with an EM counterpart, such as GW170817 (Abbott et al., 2017), the redshift can be directly measured from the host galaxy spectrum. For the more common class of dark sirens, such as GW170814 (Soares-Santos et al., 2019), where no EM counterpart is detected, a statistical association is made between the GW localization volume and galaxy catalogs. This association yields a redshift probability distribution, which can be combined with the luminosity distance posterior to constrain H_0 .

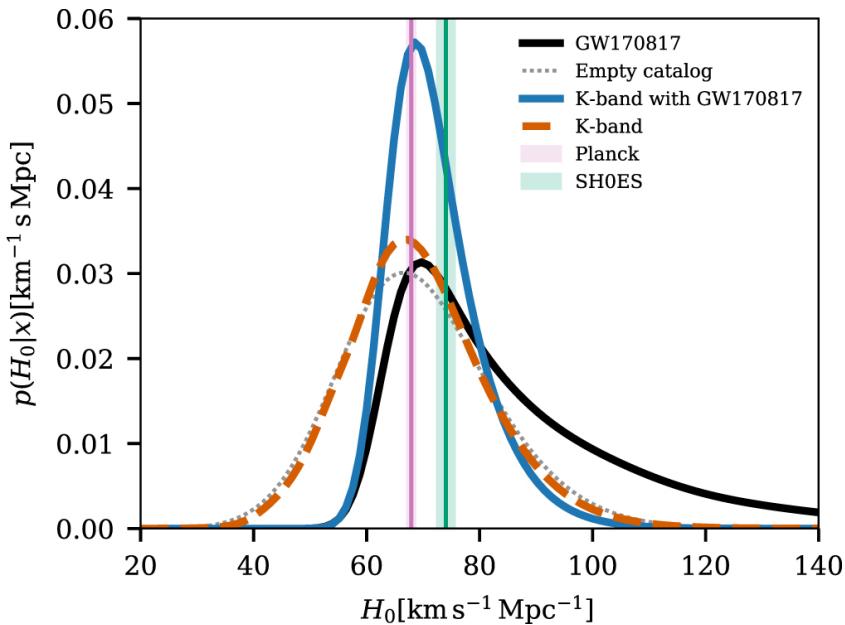


Fig. 1.2: Hubble constant posterior from Gravitational-Wave Transient Catalog (GWTC)-3 released by the LIGO, Virgo, KAGRA (LVK) collaboration, compared to the CMB and local measurements. The tension between these measurements is evident, with the GW posteriors (blue/orange), with/without the sole bright siren event, being consistent with both Planck (pink) and SHOES measurements, owing to the rather wide constraints. Additionally, the empty catalog (gray) and the bright siren (black) results are also plotted (Abbott et al., 2023a).

The promise of standard sirens lies in their independence from traditional distance ladders and CMB modeling. This makes them a crucial tool in addressing the Hubble tension. They are free from many astrophysical systematics affecting EM methods, such as dust extinction, metallicity effects, or empirical calibrations. Additionally, because GWs are less susceptible to environmental interactions, they can probe larger cosmological volumes with fewer intermediate assumptions. As GW detectors improve in sensitivity and the number of observed events increases, standard sirens are expected to play a progressively larger role in cosmological inference.

1.4 Thesis Objectives

The primary objective of this thesis is to investigate the impact of using brightness-ranked galaxy catalogs on the precision of H_0 measurements from dark sirens. By focusing on the brightest galaxies to trace the underlying matter distribution, we aim to improve the statistical constraints on H_0 while minimizing systematic uncertainties associated with galaxy catalog incompleteness.

This work involves H_0 inference using the `gwcosmo` pipeline, which combines GW distance posteriors with redshift priors obtained using galaxy catalogs. We will explore the effects of applying brightness cuts to the galaxy catalog, examining how these cuts influence the resulting H_0 posterior distributions. The analysis will be performed on a subset of

GWTC-3 events, focusing on those with high signal-to-noise ratios (SNRs) and no known electromagnetic counterparts. The results will be validated through the use of simulated catalogs and mock data challenges, allowing us to assess the robustness of our findings.

In Chapter 2, we will provide an overview of the theoretical framework for standard siren cosmology, including the principles of dark sirens. We will also discuss the challenges and limitations associated with this approach, particularly in the context of galaxy catalog incompleteness. Chapter 3 details the `gwcosmo` pipeline, which combines the GW distance posterior with the galaxy redshift prior, constructed using a galaxy catalog and corrected for catalog incompleteness, for H_0 inference. Chapter 4 details the data sources used in our analysis, including the GW event catalog and galaxy catalogs. In this chapter, we will also describe the selection criteria for the events and the modelling of missing EM data. In Chapter 5, we will detail the methodology used in our analysis, including the construction of brightness-ranked galaxy catalogs and the statistical techniques employed to derive redshift priors. In this chapter we also present our results, using the `GLADE+` galaxy catalog and GWTC event catalog, and discuss the impact of brightness cuts on the resulting H_0 posterior distributions. In Chapter 6, we will present the results of our analysis using simulated catalogs and mock data challenges, allowing us to assess the robustness of our findings. Finally, in Chapter 7, we will summarize our findings and discuss their implications for future research in cosmology.

Gravitational-Wave Cosmology

2.1 Standard Siren Cosmology

The detection of GW from compact binary coalescences (CBCs) have opened a transformative avenue in modern cosmology. These events acts as *standard sirens*, the gravitational analog of standard candles, as the luminosity distance (d_L) to a GW source is directly encoded in the strain amplitude and frequency evolution of the GW. The amplitude, corrected for antenna pattern and inclination, provides a direct, calibration-free distance measurement under the assumption of general relativity.

From Maggiore (2007), the strain measured by a GW detector can be expressed as:

$$h(t) = \frac{4}{d_L} \left(\frac{G\mathcal{M}_z}{c^2} \right)^{5/3} (\pi f(t))^{2/3} F(\iota, \psi, \theta, \phi) \quad (2.1)$$

where:

- $h(t)$ is the strain measured at time t ,
- d_L is the luminosity distance to the source,
- $\mathcal{M}_z = (1+z)\mathcal{M}$ is the redshifted chirp mass,
- $f(t)$ is the instantaneous GW frequency,
- $F(\iota, \psi, \theta, \phi)$ is a function that captures the detector response depending on inclination angle ι , polarization angle ψ , and sky position (θ, ϕ) .

The amplitude scaling with $1/d_L$ makes it possible to extract the luminosity distance directly from the signal, once the source's intrinsic properties and orientation are marginalized over. Furthermore, the $1/d_L$ scaling, in comparison to the $1/d_L^2$ scaling of EM signals, means that GW signals allow observations of sources at much larger distances, making them ideal for cosmological applications.

The direct self-calibrated measurement of the luminosity distance from the GW signal is a key advantage of standard sirens over traditional EM standard candles. This is in stark contrast to traditional EM standard candles, such as supernovae (SNe), where the distance is inferred from the observed flux and requires a calibration step to account for the intrinsic brightness of the source. The GW signal is also less affected by the intergalactic medium, as it is not subject to scattering or absorption like EM signals. This allows for a more direct measurement of the distance to the source not affected by dust extinction or other

astrophysical uncertainties that plague EM observations. This makes GW standard sirens a powerful tool for cosmology.

The concept of standard sirens was first introduced by Bernard F. Schutz, who noted that if the redshift of a GW source can be measured, one could use GW events to trace the expansion history of the Universe, in particular an independent measurement of the Hubble constant (H_0) (Schutz, 1986). This is due to the relation between the luminosity distance, redshift and the Hubble constant, in a flat Λ CDM cosmology, being:

$$d_L = \frac{c(1+z)}{H_0} \int_0^z \frac{dz'}{\sqrt{(1+z')^3\Omega_m + \Omega_\Lambda}} \quad (2.2)$$

Unlike traditional EM methods, which require cross-calibration across multiple rungs of the distance ladder (parallax, Cepheids, SNe Ia) with each step introducing uncertainties that compound, the standard siren approach provides a direct cosmological probe. This reduces systematic uncertainties and provides an independent check on other H_0 measurement techniques, possibly providing a solution to the Hubble tension, the current discrepancy between local and global measurements of the Hubble constant (Riess et al., 2019; Aghanim et al., 2020).

However, a key limitation is that while GW detectors provide a precise measurement of d_L , they do not directly measure the redshift. This necessitates an independent redshift measurement to place the source on the Hubble diagram. For *bright sirens*, this redshift is obtained from the host galaxy identified via an EM counterpart. In contrast, for *dark sirens*, the redshift is inferred statistically by cross-referencing the GW localization volume with galaxy catalogs or by leveraging population-based methods, as discussed in the following sections.

This forms the basis for *standard siren cosmology*, where GW sources are used as cosmic rulers. Over the past decade, this idea has transitioned from theoretical speculation to experimental reality, primarily through observations made by the LIGO, Virgo, KAGRA (LVK) collaboration. The first GW event, **GW150914**, was detected in 2015, marking the beginning of a new era in astrophysics and cosmology (Abbott et al., 2016). Since then, the LVK collaboration has detected numerous GW events, including binary black hole (BBH) mergers, binary neutron star (BNS) mergers, and neutron star black hole (NSBH) mergers. These observations have provided valuable insights into the nature of gravity, the formation of compact objects, and the expansion history of the Universe.

2.2 Bright vs. Dark Sirens

Standard sirens fall into two broad categories: *bright sirens* and *dark sirens*, depending on whether an electromagnetic counterpart is detected.

Bright sirens are rare but powerful. A prime example is **GW170817**, a BNS merger detected by LIGO-Virgo in 2017, which was accompanied by a short gamma-ray burst and subsequent kilonova (Abbott et al., 2017). This multi-messenger event enabled the unambiguous identification of its *host galaxy*, NGC 4993, providing both distance (from GW) and redshift (from optical spectroscopy) measurements. The resulting Hubble constant measurement, $H_0 = 70.0^{+12.0}_{-8.0} \text{ km s}^{-1} \text{Mpc}^{-1}$, was a major milestone demonstrating that GW observations could offer an independent and competitive cosmological probe (Abbott et al., 2017).

Such bright sirens provide a straightforward route to cosmological inference, using independent distance and redshift measurements. However, bright sirens require specific astrophysical conditions: the emission of EM signals strong enough to be detected, accurate sky localization, and timely follow-up by optical telescopes. Such conditions are only met for a small fraction of CBC events, especially those involving neutron stars. Majority of the observed mergers, especially binary black holes (BBHs) do not produce detectable EM counterparts, and are thus classified as dark sirens.

In the absence of a direct redshift measurement, dark sirens require a *statistical approach*. This involves cross-matching the sky localization and distance posterior from the GW event with a galaxy catalog covering the relevant region. The redshift information of a number of candidate galaxies is then used to statistically infer the likely redshift distribution of the source. This is done by constructing a *line-of-sight (LOS) redshift prior* for each GW event, which is then used to infer the Hubble constant. The LOS redshift prior is constructed by taking into account the galaxy number density and redshift distribution within the localization volume of the GW event. The GW localization volume is typically much larger than the volume of a galaxy, leading to a large number of galaxies that could potentially host the GW event. This results in a large number of potential redshift measurements, which can be used to construct a more accurate LOS redshift prior.

The statistical nature of dark sirens allows for the inclusion of a larger number of events, as it does not rely on the detection of an EM counterpart. This is particularly important for BBH events, which are more common and have a higher detection rate than BNS events. The LIGO, Virgo, KAGRA (LVK) collaboration has detected numerous dark siren events, which have been used to constrain the Hubble constant and test cosmological models (Abbott et al., 2023a).

As the number of BBH detections increases with each observing run, dark sirens will dominate the future of standard siren cosmology. However, this statistical method introduces new complexities and depends heavily on the quality and completeness of the galaxy catalog used, which plays a pivotal role in the reliability of the inferred Hubble constant. The incompleteness of the galaxy catalog can lead to biases in the inferred redshift distribution,

which can affect the accuracy of the Hubble constant measurement. This is particularly important for dark sirens, as they rely on the statistical association of GW events with galaxies in the catalog. State of the art dark siren methods have been developed to mitigate these issues, but they still rely on the quality and completeness of the galaxy catalog used. The next section discusses the current state of the art in dark siren cosmology, including the challenges and limitations of existing methods.

2.3 State of the Art Dark Siren Methods

The statistical framework for dark siren cosmology has evolved rapidly. Early implementations relied on basic overlap between GW localization volumes and precompiled galaxy catalogs. The first generation of tools, including `gwcosmo` (Gray et al., 2020; Gray et al., 2022; Gray et al., 2023) and `icarogw` (Mastrogiovanni et al., 2021; Mastrogiovanni et al., 2024), introduced a more rigorous Bayesian framework: for each GW event, a *LOS redshift prior* is constructed using galaxy number counts and redshifts, weighted by host probability, typically modeled using stellar mass proxies such as K -band luminosity.

One major challenge in this process is that current galaxy catalogs, such as GLADE+ (Dálya et al., 2022), are incomplete beyond low redshift (typically $z > 0.2 - 0.3$). This incompleteness leads to a nontrivial *out-of-catalog* term in the redshift prior, which must be modeled analytically using a *Schechter luminosity function* (Gray et al., 2023; Chen et al., 2024). This function estimates the number density of missing faint galaxies, often assuming fixed parameters (e.g., M_* , α) calibrated from deep surveys. If this modeling is inaccurate, the resulting posterior for H_0 can be biased or artificially broadened.

Despite these challenges, dark siren methods have been successfully applied to increasingly large datasets. The LVK collaboration has published joint analyses using CBC events from the **O1**, **O2**, and **O3** observing runs. For instance, Abbott et al. (2021a) and Abbott et al. (2023a) presented cumulative constraints on H_0 using numerous CBC events, demonstrating convergence toward a 10–15% precision regime, albeit still limited by catalog systematics.

These analyses highlight the potential of dark sirens to provide competitive cosmological constraints, especially as galaxy catalogs improve and detection rates increase. Future observing runs, such as **O4** and **O5**, are expected to significantly enhance the statistical power of dark siren cosmology, potentially reducing uncertainties in H_0 , providing better constraints on cosmological model, and testing the validity of general relativity on cosmological scales. The next generation of GW detectors, such as the ground-based **Einstein Telescope** (Abac et al., 2025) and **Cosmic Explorer** (Evans et al., 2021), and the space-based **LISA** observatory (Colpi et al., 2024), will further enhance the sensitivity and detection rates of GW events, opening up new avenues for cosmological exploration.

2.3.1 Redshift Inference Methods

In dark siren cosmology, two principal strategies exist for inferring the redshift of gravitational-wave sources: the *catalog-based method* (Gray et al., 2020; Gray et al., 2023; Chen et al., 2024) and the *population-based method* (Ezquiaga and Holz, 2022). The catalog method, briefly discussed earlier, statistically associates GW sources with galaxies from a survey by constructing a redshift prior along each line of sight, derived from galaxy positions and redshifts within the GW localization volume. This approach captures the clustering of galaxies and allows for redshift inference when direct EM counterparts are absent, but suffers from incompleteness at higher redshifts and spatial variation in survey depth (Gray et al., 2020; Gray et al., 2023; Chen et al., 2024). Conversely, the population method, often referred to as the *spectral siren approach*, leverages features in the observed distribution of source-frame binary parameters, especially masses, which are redshifted in the detector frame. If the intrinsic mass distribution contains recognizable structure (e.g., a cutoff or peak), its displacement due to cosmic redshift can break the mass-redshift degeneracy, enabling cosmological inference independent of galaxy surveys (Ezquiaga and Holz, 2022). However, this approach is sensitive to modeling assumptions about the underlying binary population.

Recent work has sought to improve robustness by combining dark siren methods with *spectral sirens*, extracting redshift information from the observed mass distribution of GW sources under population synthesis assumptions (Ezquiaga and Holz, 2022). These hybrid approaches aim to reduce dependence on incomplete catalogs and mitigate selection biases while extracting maximal cosmological information from current GW detections. Recently, both `gwcosmo` and `icarogw` have been updated to support such joint inference schemes (Gray et al., 2023; Mastrogianni et al., 2024).

2.3.2 Galaxy Weighting Choices

In the construction of the LOS redshift prior, each host galaxy is weighted by its probability of hosting a GW event. This probability is typically modeled as a function of the galaxy's luminosity, and relies on the assumption that the merger host probability of a galaxy is related to its luminosity in a specific band. The choice of band remains uncertain, with different bands correlating with different aspects. The blue band, for instance, correlates with star formation rate and is therefore appropriate for mergers with short time delays (coalescence shortly after formation), while the near-infrared band correlates with the total luminous mass and is thus appropriate for mergers with longer time delays (mergers in older, more evolved galaxies). It is not known how exactly galaxy color is correlated with the merger rate.

For modelling the out-of-catalog contribution, a Schechter luminosity function is assumed, which describes the distribution of galaxy luminosities in a given band. Wrong assumptions on the Schechter parameters, or incorrect description of selection biases can lead to biases in the inferred Hubble constant. The key assumption in constructing the out-of-catalog contribution is that the galaxy catalog is magnitude limited, i.e. galaxies are not detected

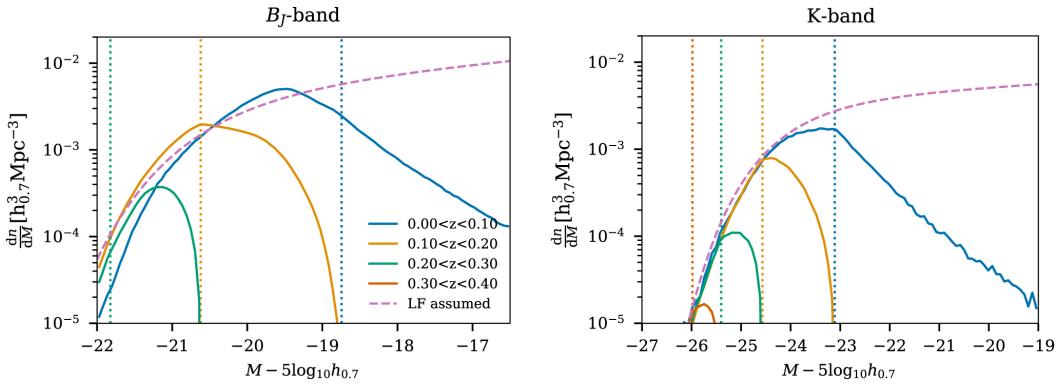


Fig. 2.1: The B_J -band (left) and K -band luminosity function (right) for the GLADE+ catalog. The solid lines show the luminosity function of the catalog, while the dashed lines shows the assumed Schechter function. The vertical dashed line indicates the median absolute magnitude threshold for galaxy detection. The bright end of the K -band luminosity seems to match the assumed Schechter function, while the B_J -band luminosity function does not. This indicates that the assumption of a magnitude-limited catalog is not valid for the B_J -band (Abbott et al., 2023a).

only because they are too faint. If other selection biases (based on e.g., colors or spectral features) were present, the out-of-catalog contribution would be biased, leading to an incorrect estimate of the redshift prior (Abbott et al., 2023a).

If the assumption of a magnitude-limited catalog is valid, then the luminosity function of the galaxy catalog would match the assumed Schechter function at its bright end, and start to decrease as it reaches the corresponding absolute magnitude threshold. As can be seen in Figure 2.1, this is indeed the case for the K -band luminosity function in GLADE+, which is well described by the assumed Schechter function. The B_J -band luminosity function, on the other hand, is not well described, and the assumption of a magnitude-limited catalog is not valid as there seem to be some additional missing galaxies at low redshift. This could lead to an incorrect estimate of the out-of-catalog contribution, and therefore an incorrect estimate of the redshift prior (Abbott et al., 2023a). For this reason, traditionally, the K -band luminosity has been used for galaxy weighting in GW cosmology inference pipelines. Another reason for using the K -band is, that the K -band luminosity is better correlated with the stellar mass of the galaxy, making it a more reliable tracer of the underlying matter distribution (Strazzullo et al., 2006; Sureshkumar et al., 2021; Abbott et al., 2023a).

However, this choice is not without its limitations. The K -band luminosity is not always available for all galaxies in the catalog, leading to potential biases in the selection of galaxies used to construct the redshift prior. For example, the K -band luminosity is only available for about 1 million galaxies in the GLADE+ catalog, which is a small fraction of the total number of galaxies in the catalog. This can lead to uncertainties in the redshift prior, which can in turn affect the inferred value of the Hubble constant. However, this effect should be relatively small, as we are interested in the overall matter distribution, and the K -band luminosity is a good tracer of the underlying matter distribution. This also forms the basis for this thesis, as we will be using the brightest galaxies in the K -band, in the GLADE+ galaxy catalog, to trace the underlying matter distribution for improved H_0 inference as discussed in the next section.

2.4 Refining the Redshift Prior: Brightest Galaxies as Tracers

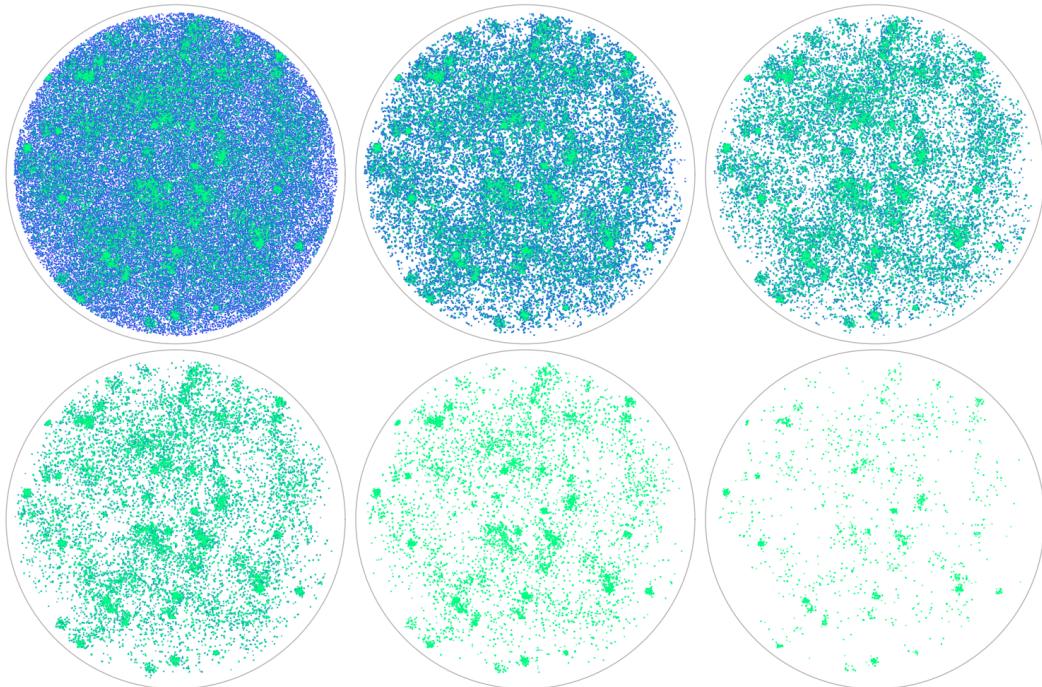


Fig. 2.2: Illustration of the proposed method for refining H_0 inference using the brightest galaxies as redshift tracers. The panels show different brightness cuts being applied to a galaxy catalog. The top left panel shows the case with a full galaxy catalog, while the subsequent panels (left-to-right) show the effect of applying increasingly strict brightness cuts. These depict how we can trace the matter distribution by using only the brightest galaxies. The last two panels (bottom center and right) show overly restrictive cuts, which would lead to a loss of information and a biased result. This is a conceptual visualization, generated with AI assistance, intended to represent the methodology rather than actual observational data.

The current state of the art in dark siren cosmology relies heavily on the quality and completeness of the galaxy catalog used to construct the LOS redshift prior. The incompleteness of the galaxy catalog can lead to biases in the inferred redshift distribution, which can affect the accuracy of the Hubble constant measurement. State of the art dark siren methods have been developed to account for the incompleteness, but they still rely on the quality and completeness of the galaxy catalog used.

One strategy to mitigate this limitation is by refining the construction of the redshift prior by focusing on the *brightest galaxies* to trace the underlying matter distribution. This would effectively make the catalog more complete and increase precision without introducing significant bias. The rationale is that the brightest galaxies are more likely to be associated with massive halos, which are in turn more likely to host detectable GW events. The brightest galaxies are also more likely to be catalogued, as they are easier to detect and measure. These galaxies would also crudely trace the underlying matter distribution, since they are more likely to be associated with massive halos. This would be particularly useful for dark

sirens, as they rely on the statistical association of GW events with galaxies in the catalog. By replacing the redshift of the true host galaxy with the redshift of a nearby bright galaxy in the catalog, the small error incurred would be insignificant compared to the overall uncertainty in the redshift prior and the luminosity distance measurement.

Figure 2.2 illustrates this approach, highlighting how the brightest galaxies can be used to trace the underlying matter distribution. The last two panels in the figure show an overly restrictive cut, which would lead to a loss of information and a biased redshift prior. This emphasizes the importance of selecting an optimal brightness cut that balances the need for completeness with the desire for precision. The goal would be to find the optimal brightness cut that maximizes the precision of the redshift prior while minimizing bias.

This approach, of using the brightest galaxies to trace the matter distribution, is similar to the *Brightest Cluster Galaxy (BCG)* method used in traditional cosmology, where the brightest galaxy in a cluster is used as a standard candle. The BCG method has been shown to be effective in reducing the scatter in distance measurements and improving the precision of cosmological constraints (Lauer et al., 2014). By applying a similar approach to dark sirens, one can potentially improve the precision of the inferred Hubble constant without introducing significant bias.

To implement this approach, we define a subset of the GLADE+ galaxy catalog, GLADEPXX, which includes only the top $XX\%$ of galaxies ranked by K -band luminosity, as the K -band luminosity is better associated with the mass of the galaxies (Strazzullo et al., 2006; Sureshkumar et al., 2021). The construction of the LOS redshift prior is then modified to account for this restricted catalog. Specifically, the Schechter luminosity function is adjusted to reflect the brighter subset, making the out-of-catalog contribution smaller, effectively making the catalog more complete. This modified prior is then used in the Bayesian framework to infer the Hubble constant.

The next chapters provide a detailed description of the methodology, including the `gwcosmo` inference pipeline, the selection criteria for GLADEPXX, the modeling of the out-of-catalog part and the validation of this approach using mock data.

The gwcosmo Pipeline

Based on the work presented in Gray et al. (2020), Gray et al. (2022), and Gray et al. (2023), gwcosmo is a Python package specifically designed for the joint inference of cosmological parameters using data from standard sirens and galaxy catalogs. The pipeline employs Bayesian inference techniques to simultaneously constrain cosmological parameters, such as the Hubble constant and dark energy equation of state parameters. It does this using various data inputs, including the gravitational wave strain data from gravitational wave observatories such as LIGO, Virgo, and KAGRA, entries from galaxy catalogs containing photometric or spectroscopic redshift information, as well as other relevant galaxy properties, and CBC source population models.

gwcosmo represents a significant advancement in extracting cosmological information especially from CBC events that lack unique EM counterparts. It addresses the challenge of redshift inference in dark siren cosmology by statistically marginalizing over potential host galaxies from a flux-limited catalog. This is achieved through hierarchical Bayesian modeling that incorporates GW strain data, the spatial and redshift distribution of galaxies, and astrophysical population models.

The pipeline is designed to be flexible and extensible, allowing users to customize the models and priors used in the analysis. It also provides tools for visualizing the results, including posterior distributions and confidence intervals for the inferred parameters. The package is built on top of established libraries such as numpy, scipy, and matplotlib, making it accessible to researchers familiar with these tools. The gwcosmo pipeline is a particularly useful tool in the intersection of gravitational wave astronomy and cosmology, as it enables the extraction of valuable cosmological information from gravitational wave events, which can complement traditional methods of measuring cosmological parameters. The package is publicly available and actively maintained and updated to incorporate new developments in both GW detection and cosmological modeling, ensuring that it remains a relevant tool for researchers in the field.

3.1 Bayesian Framework

The gwcosmo package implements a hierarchical Bayesian framework to infer cosmological and astrophysical parameters from GW observations, particularly when dealing with dark sirens, GW events lacking EM counterparts. This framework allows for the consistent incorporation of population models, redshift priors from galaxy catalogs, and detection selection effects (Chen et al., 2024).

Given a catalog of N_{det} detected GW events $\{x_{\text{GW}}\}$, the posterior on the set of hyperparameters Λ , which may include cosmological parameters (e.g., H_0), compact binary population parameters, and parameters of modified gravity, is given by:

$$p(\Lambda|\{x_{\text{GW}}\}, \{D_{\text{GW}}\}, I) \propto p(\Lambda|I) p(N_{\text{det}}|\Lambda, I) \prod_{i=1}^{N_{\text{det}}} p(x_{\text{GW},i}|D_{\text{GW},i}, \Lambda, I), \quad (3.1)$$

where $p(\Lambda|I)$ is the prior on the hyperparameters, and $p(N_{\text{det}}|\Lambda, I)$ encodes the expected number of detections, which depends on the underlying merger rate. The parameter D_{GW} represents whether a GW event is detected or not, and the parameter I represents any additional assumptions or information used in the analysis, not explicitly contained in Λ .

The likelihood $p(x_{\text{GW},i}|D_{\text{GW},i}, \Lambda, I)$ is the core of the analysis, as it encodes the information from the GW event. It is constructed using a combination of GW signal modeling and galaxy catalog data. It can be further decomposed, following Chen et al. (2024), into the following components, separating the GW signal likelihood from the detection probability:

$$\begin{aligned} \prod_{i=1}^{N_{\text{det}}} p(x_{\text{GW},i}|D_{\text{GW},i}, \Lambda, I) &\propto \left[\prod_i^{N_{\text{det}}} \int p(x_{\text{GW},i}|\theta, \Lambda, I) p(\theta|\Lambda, I) d\theta \right] \\ &\times \left[\int p(D_{\text{GW},i}|\theta, \Lambda, I) p(\theta|\Lambda, I) d\theta \right]^{-N_{\text{det}}} \end{aligned} \quad (3.2)$$

where θ represents the intrinsic parameters of the GW source (e.g., masses, spins), and $p(x_{\text{GW},i}|\theta, \Lambda, I)$ is the likelihood of the GW signal given the source parameters and cosmological model. The first term in the product represents the likelihood of the GW signal given the source parameters, while the second term accounts for the detection probability of the event.

The likelihood for each event incorporates both GW signal modeling and the line-of-sight (LOS) redshift prior derived from galaxy catalogs. For a dark siren event, where the host galaxy is not identified, the sky localization is divided into HEALPix pixels, and a redshift prior is computed per pixel based on the galaxy distribution (Gray et al., 2020; Chen et al., 2024). The likelihood then integrates over redshift and sky location:

$$p(x_{\text{GW},i}|\theta, \Lambda, I) = \int dz \sum_j^{N_{\text{pix}}} p(x_{\text{GW},i}|z, \theta, \Lambda, I) p(z|\Omega_j, \Lambda, I), \quad (3.3)$$

where $p(z|\Omega_j, \Lambda, I)$ is the LOS redshift prior for pixel Ω_j .

Similarly the detection probability $p(D_{\text{GW},i}|\theta, \Lambda, I)$ is computed by integrating over the redshift and sky localization:

$$p(D_{\text{GW},i}|\theta, \Lambda, I) = \int dz p(D_{\text{GW},i}|z, \theta, \Lambda, I) \sum_j^{N_{\text{pix}}} p(z|\Omega_j, \Lambda, I) \quad (3.4)$$

The GW signal provides a measurement of the *luminosity distance*, not redshift. Therefore, cosmology enters through the mapping $d_L(z; \Lambda)$, which can include modifications from

non-GR theories. For dark sirens, the redshift must be inferred statistically using host galaxy populations.

Detection probability and selection effects are corrected using a large set of injections. These injections refer to a large set of simulated CBC events, each defined by intrinsic parameters such as component masses (m_1, m_2), luminosity distance (d_L), sky location (RA, dec), and coalescence time (t_0). They are generated using Monte Carlo sampling from a known distribution over the N -dimensional parameter space relevant to CBCs. The purpose of these injections is to estimate the selection function, i.e., the probability that an event with given parameters would be detected by a GW detector network with specified sensitivities, duty cycles, and SNR thresholds. This estimation is performed via preference sampling, where the multidimensional integrals appearing in the likelihood are approximated by a weighted sum over injections:

$$p(D_{\text{GW}}|\Lambda, I) = \frac{1}{N_{\text{inj}}} \sum_{k=1}^{N_{\text{inj}}} w_k(\Lambda), \quad (3.5)$$

where the weights $w_k(\Lambda)$ encode how likely each injected event is under the assumed population and cosmological model. These selection corrections are crucial for unbiased inference, especially in the presence of detection thresholds and catalog incompleteness.

An important innovation in `gwcosmo` is the ability to simultaneously marginalize over population distributions, redshift priors, and selection functions. This modular structure allows consistent updates, for example, incorporating new BBH mass distribution models, or parameterized deviations from GR. Modified gravity effects, if modeled through the GW luminosity distance $d_L^{\text{GW}}(z)$, propagate into the likelihood via the distance-redshift conversion (Chen et al., 2024).

To ensure robustness, the package separates posterior contributions into numerators and denominators involving the event likelihood and selection function, respectively. Following the discussion earlier, the final likelihood for the entire catalog is given by:

$$p(\Lambda|\{x_{\text{GW}}\}) \propto \frac{\prod_i p(x_{\text{GW},i}|\Lambda)}{[p(D_{\text{GW}}|\Lambda)]^{N_{\text{det}}}}. \quad (3.6)$$

For bright sirens with identified hosts, the redshift prior is replaced by a Gaussian centered on the spectroscopic redshift. For dark sirens, the galaxy catalog, accounting for catalog incompleteness, provides a redshift distribution per sky pixel, accounting for photometric redshift uncertainties and catalog incompleteness.

This Bayesian framework enables self-consistent cosmological inference from GW data without EM counterparts and is central to the analyses carried out in this thesis.

3.2 Key Features of `gwcosmo`

Hereby we summarize the key features of the `gwcosmo` pipeline:

- **Joint Population and Cosmology Inference**

Simultaneously infers cosmological parameters (e.g., H_0 , Ω_m) and compact binary population hyperparameters (e.g., mass distribution, merger rate evolution), thereby avoiding biases that arise from fixing one while fitting the other. This fully Bayesian approach accounts for their mutual correlations and degeneracies.

- **Galaxy Catalog-Based Redshift Priors**

Constructs line-of-sight (LOS) redshift priors using spatial and photometric information from galaxy catalogs. These priors include weighting schemes based on the luminosities of the potential host galaxies, and account for catalog incompleteness using apparent magnitude thresholds to reflect observational selection effects.

- **Selection Effect Corrections**

Accurately models the detection efficiencies of GW detectors and observational biases in galaxy catalogs. The pipeline incorporates realistic detection probabilities and magnitude-limited selection functions, ensuring that cosmological inference is properly normalized and free from detection-induced biases.

- **Pixelated Sky Localization**

Implements a HEALPix-based discretization of the sky to handle dark sirens. The likelihood is evaluated by marginalizing over redshift and intrinsic parameters for each pixel, weighted by the redshift distribution of galaxies within that pixel.

- **Modular Cosmology Framework**

Supports inference under standard Λ CDM as well as extensions including w CDM, w_0-w_a CDM, and parameterized modifications of gravity. This enables testing of a broad class of cosmological and gravitational models using GW data.

3.3 Implementation and Applications

The `gwcosmo` pipeline is implemented in Python and is designed to enable robust Bayesian inference using GW data in conjunction with galaxy catalogs. It uses kernel density estimation (KDE) of posterior samples from GW parameter estimation to build smooth probability distributions in redshift and distance. For dark siren analyses, it performs efficient catalog summation over galaxies within each HEALPix pixel, accounting for photometric redshift uncertainties, magnitude limits, and catalog incompleteness. This approach allows the pipeline to statistically marginalize over possible host galaxies without requiring an electromagnetic counterpart.

The pipeline has been applied in recent analyses to test its cosmological inference capabilities. In particular, Gray et al. (2023) used `gwcosmo` to reanalyze 47 CBC events from the GWTC-3 catalog, combining these with the GLADE+ galaxy catalog to perform joint population and cosmology inference. Their analysis yielded the following constraint on the Hubble constant:

$$H_0 = 69^{+12}_{-7} \text{ km s}^{-1} \text{ Mpc}^{-1}$$

This value lies midway between the early-Universe measurement from *Planck* ($H_0 = 67.4 \pm 0.5 \text{ km s}^{-1} \text{ Mpc}^{-1}$) (Aghanim et al., 2020) and the local distance-ladder estimate from the SHOES project ($H_0 = 73.30 \pm 1.04 \text{ km s}^{-1} \text{ Mpc}^{-1}$) (Riess et al., 2022). While the uncertainty remains large due to the limited number of dark siren events and the incompleteness of the galaxy catalog, this result demonstrates the power of standard siren cosmology as an independent cosmology probe.

As the catalog of detected GW events grows in future observing runs, and galaxy surveys become deeper and more complete, the precision of dark siren measurements is expected to improve significantly. The modular design of `gwcosmo` makes it readily adaptable to incorporate new detector sensitivities, cosmological models, or improved photometric redshift techniques, positioning it as a key tool in the effort to resolve the Hubble tension through GW observations.

Data

The foundation of any standard siren cosmological analysis lies in the quality, completeness, and coverage of the data. In this thesis, we rely on two primary data sources: GW observations from the LIGO, Virgo, KAGRA (LVK) detectors, and galaxy catalogs providing EM redshift information. These datasets are used in tandem to statistically construct a redshift prior for each GW event and ultimately to constrain the Hubble constant H_0 .

4.1 GW Event Data

4.1.1 The GWTC-3 Catalog

The GW data used in this study are obtained from the *Gravitational-Wave Transient Catalog (GWTC)* (Abbott et al., 2019; Abbott et al., 2021b; Abbott et al., 2023b; Abbott et al., 2024). This catalog comprises all CBC events detected during the three observing runs (O1, O2, and O3) of the Advanced LIGO and Virgo detectors. GWTC includes over 93 confident detections, primarily binary black hole (BBH) mergers, along with a few neutron star black hole (NSBH) and binary neutron star (BNS) systems.

For each event, the catalog provides posterior samples for key source parameters inferred through Bayesian parameter estimation. These include the detector-frame component masses ($m_1^{\text{det}}, m_2^{\text{det}}$), the luminosity distance d_L , sky location (right ascension α and declination δ), orbital inclination angle ι , chirp mass \mathcal{M} , and various spin parameters (e.g., $\chi_{\text{eff}}, \chi_1, \chi_2$). These samples form the basis for constructing the luminosity distance posterior and are essential inputs to the standard siren inference pipeline.

4.1.2 Event Selection Criteria

To focus our analysis on dark siren cosmology, we restrict our sample to events that do not have any confirmed electromagnetic counterpart (i.e., no associated kilonova or GRB), have a network SNR exceeding 11, have somewhat good sky localization, and are accompanied by publicly available posterior samples for distance and sky position.

These criteria are designed to ensure both the statistical robustness of the inference and compatibility with existing galaxy catalogs. After applying these cuts, a subset of 47 events, with 42 BBHs, 3 NSBHs, and 2 BNSs, is retained for cosmological analysis.

4.1.3 Distance Posteriors and Sky Localization

Each event is characterized by a posterior distribution over the three-dimensional localization volume, typically encoded using the HEALPix format. The posterior provides the probability density $p(d_L, \alpha, \delta)$, where d_L is the luminosity distance and (α, δ) denote right ascension and declination.

These distance posteriors are crucial for constructing the LOS redshift prior when cross-matched with galaxy catalogs. Events with poor localization or multimodal distance distributions contribute more uncertainty to the inferred H_0 , emphasizing the importance of event quality.

4.1.4 Assumptions About the Source Population

Although this thesis primarily focuses on statistical redshift modeling, it is important to acknowledge that the cosmological inference also depends weakly on assumptions about the source population. For instance, `gwcosmo` allows one to specify priors on:

- Merger rate evolution with redshift, typically modeled as $R(z) \propto (1+z)^\kappa$
- Mass distributions of the binaries

For our analysis we assume that CBCs follow a Madau-Dickinson merger redshift evolution model (Madau and Dickinson, 2014):

$$R(z) = R_0(1+z)^\gamma \frac{1 + (1+z_p)^{-(\gamma+\kappa)}}{1 + \left(\frac{1+z}{1+z_p}\right)^{\gamma+\kappa}} \quad (4.1)$$

Here R_0 is the local merger rate, κ the high-z slope, γ the low-z slope and z_p the break-point. All the assumed priors, taken from Abbott et al. (2023c), are given in Table 4.1.

For BBH we use power-law with a Gaussian peak mass model, with powerlaw slope α , the mean of the Gaussian σ_g , the width of the peak μ_g , and the relative weight between power-law and the peak λ_g . We also assume the minimum and maximum masses for the black hole distribution. For neutron stars we assume that mass is uniformly distributed between $M_{min,NS}$ and $M_{max,NS}$. All the assumed priors, taken from Abbott et al. (2023c), are given in Table 4.2.

Tab. 4.1: Priors on merger rate shape parameters.

Parameter	Prior
R_0	$1/R_0$ (implicit)
κ	2.86
γ	4.59
z_p	2.47

Tab. 4.2: Priors on the mass model parameters.

Parameter	Prior
α	3.78
σ_g	3.88
μ_g	32.27
λ_g	0.03
$M_{min,BH}$	$4.98M_\odot$
$M_{max,BH}$	$112.5M_\odot$
$M_{min,NS}$	$1.0M_\odot$
$M_{max,NS}$	$3.0M_\odot$

4.2 EM Galaxy Catalogs

4.2.1 The GLADE+ Galaxy Catalog

For redshift information, we use the GLADE+ galaxy catalog (Dálya et al., 2022), an extended version of the original GLADE catalog (Dálya et al., 2018), designed for GW follow-up. GLADE+ is a composite catalog that combines several large surveys to achieve all-sky coverage:

- 2MASS XSC and 2MPZ: infrared-based all-sky surveys (Skrutskie et al., 2006; Bilicki et al., 2013)
- WISExSCOS: photometric redshifts from the Wide-Field Infrared Survey Explorer (WISE) and the SuperCOSMOS survey (Bilicki et al., 2016)
- HyperLEDA: spectroscopic redshift survey (Makarov et al., 2014)
- SDSS DR16Q: quasar catalog (Lyke et al., 2020)

As of the latest release, GLADE+ includes over 22 million objects with available positions, redshifts (spectroscopic or photometric), and photometry in multiple bands, most critically the K -band (Dálya et al., 2022).

4.2.2 Catalog Completeness

A significant limitation of GLADE+ is its incompleteness beyond redshift $z \sim 0.3$. While the catalog is approximately complete for bright galaxies at low redshift, its coverage of fainter galaxies or more distant regions is limited. This incompleteness introduces a bias when performing statistical redshift inference, as missing galaxies contribute to the out-of-catalog part of the redshift prior.

To address this, the `gwcosmo` pipeline models the galaxy population, for the out-of-catalog region, using a truncated Schechter luminosity function:

$$\phi(M) \propto 10^{0.4(\alpha+1)(M-M^*)} \exp[-10^{0.4(M-M^*)}] \quad (4.2)$$

where M is the absolute magnitude in the K -band, M^* is the characteristic magnitude, and α is the faint-end slope. For our analysis, we adopt the standard parameters: $\alpha = -1.09$, $M_K^* = -23.39 + 5 \log h$, consistent with Kochanek et al. (2001), and truncate at $M_{K,min}^* = -27.00 + 5 \log h$ and $M_{K,max}^* = -19.00 + 5 \log h$.

4.2.3 Redshift Uncertainty Modeling

Redshift measurements in GLADE+ are heterogeneous. Spectroscopic redshifts are generally precise, but photometric redshifts can have uncertainties on the order of $\sigma_z \sim 0.01 - 0.05$. These uncertainties are modeled by convolving the redshift of each galaxy with a Gaussian of fixed width, an assumption used in constructing the redshift prior.

Additionally, to reduce systematic error, galaxies with unphysical redshifts or unreliable photometry are filtered out during preprocessing. The net result is a cleaned, sky-localized, and redshift-tagged galaxy distribution that can be used for LOS prior construction.

Furthermore, K -corrections are applied in `gwcosmo` following Kochanek et al. (2001), and the apparent magnitude thresholds m_{thr} are computed per HEALPix pixel as a median apparent magnitude in a given pixel, giving a K -band threshold of 13.5 on average.

4.2.4 K -band Luminosity as Host Probability Proxy

In dark siren analysis, we must assign each galaxy a probability of being the true host. Following common practice, this probability is assumed to scale with stellar mass, for which K -band luminosity serves as a good proxy (Strazzullo et al., 2006; Sureshkumar et al., 2021). The host probability p_i for each galaxy is given by:

$$p_i \propto L_{K,i} \quad (4.3)$$

This choice is motivated by the assumption that more massive galaxies are more likely to host CBCs events. While simplistic, it provides a reasonable baseline.

GLADE+ Bright Galaxies as Redshift Tracers

The core idea investigated in this thesis is the use of **bright galaxies** as redshift tracers for GW events. The GLADE+ galaxy catalog is a comprehensive compilation of galaxies in the local Universe, designed primarily to support multi-messenger follow-up observations. However, it suffers from incompleteness at higher redshifts due to the limited sensitivity of current EM surveys.

To mitigate this limitation, we focus on the brightest galaxies in the catalog, which are more likely to be detected at higher redshifts and better trace the large-scale structure. By applying a brightness-based pruning strategy, we effectively enhance the redshift reach and completeness of the catalog, allowing GLADE+ to serve as a more accurate proxy for the underlying matter distribution relevant to dark siren cosmology.

The results and methodology developed in this thesis have been submitted for peer review to the *Monthly Notices of the Royal Astronomical Society (MNRAS)* and are available as a preprint on arXiv (Naveed et al., 2025), contributing to the growing body of research in GW cosmology.

5.1 Bright Galaxy Subsets

The analysis begins with the construction of **bright galaxy subsets**, denoted GLADEPXX, where only the top XX% of galaxies by cumulative K -band luminosity are included. The K -band luminosity is chosen as it is better associated with the mass of the galaxies (Strazzullo et al., 2006; Sureshkumar et al., 2021). Limiting the analysis to these bright subsets allows us to:

- Reduce the out-of-catalog correction by focusing on bright galaxies likely to be well-represented.
- Improve the signal-to-noise ratio of the statistical redshift prior.
- Minimize the inclusion of poorly characterized or faint galaxies.

The bright galaxy subsets are defined by selecting galaxies based on their absolute magnitude, which is inferred from their redshift and apparent magnitude. For a given subset, the dimmest galaxy sets the absolute magnitude limit, M_{max} , for that subset. This limit is then used in the Schechter luminosity function to adjust the out-of-catalog contribution, effectively

reducing its impact. By focusing on the brightest galaxies, we assume that they trace the mass distribution in the Universe more effectively, and are more likely to host CBC events.

This approach improves the completeness of the galaxy catalog in the high-redshift regime, as the out-of-catalog contribution becomes smaller. Figure 5.1 illustrates how the redshift distribution shift towards higher values making the catalog more complete, highlighting the effectiveness of this method in probing deeper z -ranges.

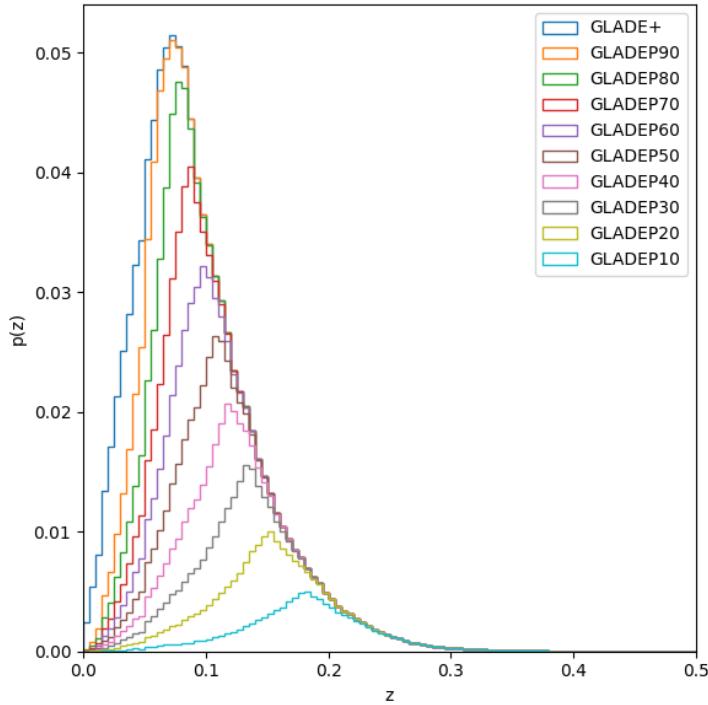


Fig. 5.1: Normalized redshift distributions $p(z)$ for the full GLADE+ galaxy catalogue (blue) and the brightest percentiles (other curves). Each subset, labeled GLADEPXX, includes only the top XX% brightest galaxies in the K -band. Focusing on increasingly brighter subsets shifts and sharpens the redshift distribution, effectively probing deeper ranges in z .

5.1.1 Trade-Offs and Limitations

While bright subsets reduce catalog incompleteness, they may inadvertently exclude genuine host galaxies, especially if those lie in less luminous systems or in underrepresented regions of the catalog. The effectiveness of these cuts depends on the intrinsic distribution of host galaxies, the depth and completeness of the original catalog, and the accuracy of K -band photometry.

One major limitation of this approach is that brightest galaxies may not be representative of the overall galaxy population, as they may be biased towards certain types of galaxies or regions of the sky. This effect is however mitigated by the fact that we are using the K -band luminosities, which are good tracer for the stellar mass (Strazzullo et al., 2006; Sureshkumar et al., 2021), and in turn the overall matter distribution. Furthermore, the bright galaxy

subsets are constructed from the GLADE+ catalog, which is designed to be a complete and representative sample of galaxies in the local Universe. This means that the bright galaxy subsets are likely to be representative of the overall galaxy population, and give a crude estimate of the matter distribution, even if they are not a complete sample.

Estimating the true redshift of a GW event by the nearest brightest galaxies is supposed to have a negligible effect on the results. This is because the bright galaxies are more likely to be located in regions of high density, such as galaxy clusters or groups, which are also likely to host GW events. This means that the bright galaxies are likely to be located in the same large-scale structure as the GW event, and thus provide a good estimate of the redshift of the event.

One important thing to note is that while the use of a bright subset enhances the completeness of the catalogue at high redshifts, it also introduces a trade-off. If the brightness cut is too restrictive, the exclusion of galaxies may lead to a loss of information and in turn an underrepresentation of the overall matter distribution, potentially biasing the results. One needs to carefully balance these considerations by comparing different brightness thresholds. A moderate brightness cut will maximize the benefit in terms of depth without incurring significant bias. The appropriate brightness cut can be established only via a set of astrophysics-motivated large-scale simulations. A *GW Mock Data Challenge (MDC)*, detailed in the next chapter, using the simulated BUZZARD galaxy catalog from the Dark Energy Survey (DES) Collaboration (DeRose et al., 2019; DeRose et al., 2022).

Another thing to note is that due to the brightness cut, we may lose the low-redshift galaxies, in turn losing information in the nearby Universe. This can be easily overcome in the future by combining the use of a complete galaxy catalog in the nearby Universe, with the bright galaxy subsets in the high-redshift regime. This will allow us to have a complete galaxy catalog in the nearby Universe, while still being able to probe deeper redshifts using the bright galaxy subsets. But this is not a trivial task, as it would require a smooth transition between the two cases. But nonetheless, this is a good first step towards improving the completeness of the galaxy catalog and future work could further refine this approach by using different redshift tracers in different redshift regimes, such as using the bright galaxy subsets in the high-redshift regime, and a complete galaxy catalog in the nearby Universe.

While the bright galaxy subsets may not be a perfect representation of the overall galaxy population, they are still a useful tool for improving the completeness of the galaxy catalog and reducing the out-of-catalog correction. The error incurred by using the bright galaxies as redshift tracers may be small compared to the current errors in luminosity distance measurements from the GW events, this may become a problem in the future as the GW measurements become more precise. In this case, one may need to use more sophisticated methods to estimate the redshift of the GW event, such as more complex models of the galaxy population, but for now this is a good first step towards improving the completeness of the galaxy catalog and reducing the out-of-catalog correction, by leveraging the currently available EM data.

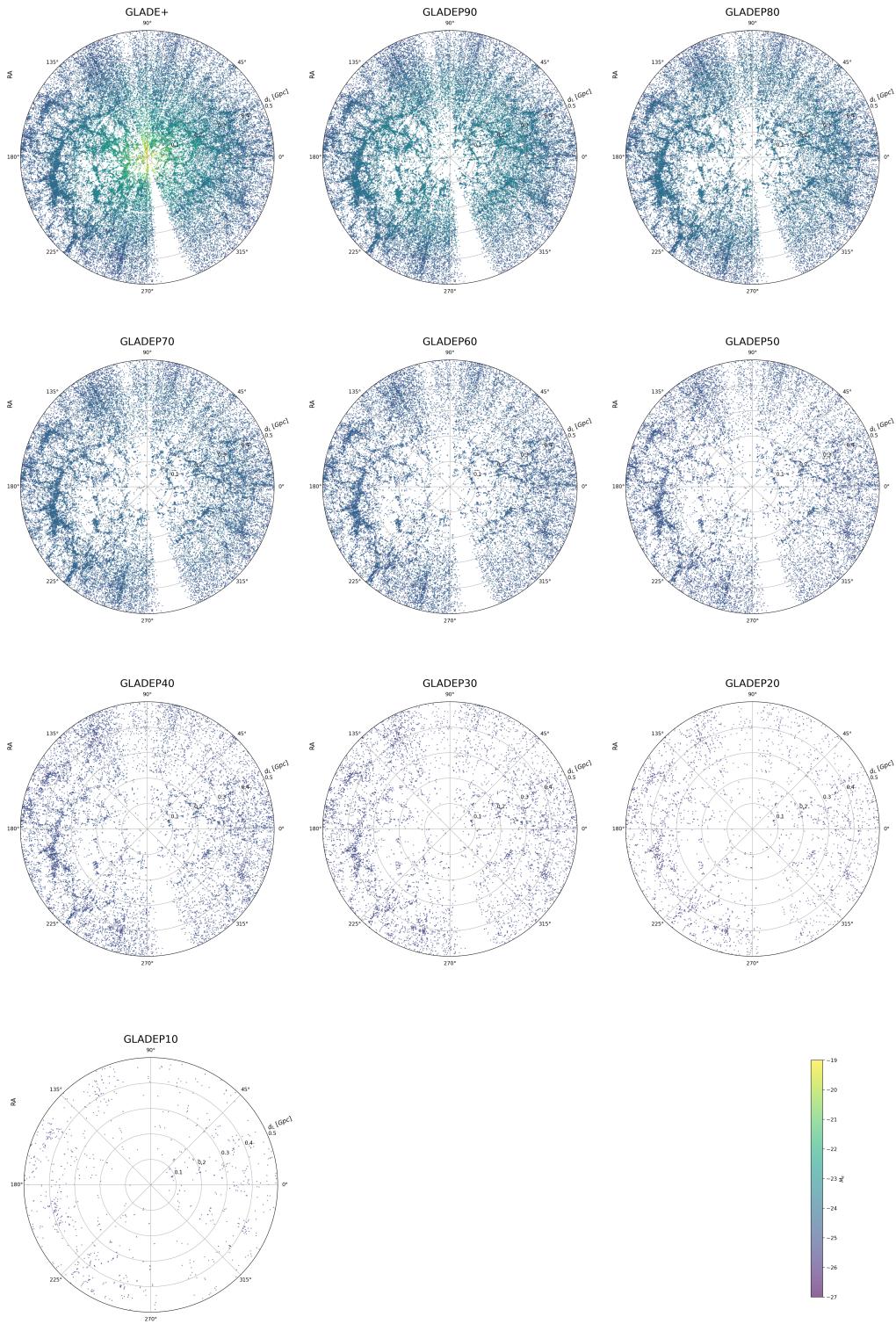


Fig. 5.2: Spatial distribution of galaxies from the GLADE+ catalog and its bright subsets, employed for dark standard siren cosmology. The plots show a 10-degree slice in declination, centered at 0° , with the radial coordinate representing luminosity distance d_L (in Gpc) and the angular coordinates being right ascension (RA). This shows how the bright galaxies trace the large-scale structure of the Universe.

5.2 LOS Redshift Prior

As `gwcosmo` requires a redshift prior for each GW event, we construct a LOS redshift prior from the GLADE+ galaxy catalog. The GLADE+ catalog is a comprehensive database of galaxies in the local Universe, which provides information about their positions, redshifts, and luminosities. This catalog is used to construct a redshift prior for each GW event, which is used in tandem with the luminosity distance posterior from the GW signal, to get a measurement for the Hubble constant H_0 .

5.2.1 LOS Redshift Prior Construction

The LOS redshift prior is constructed from the GLADE+ galaxy catalog, which contains a wealth of information about the galaxies in the local Universe. The redshift prior is constructed by taking into account the distribution of galaxies along the line of sight to the GW event, as well as their luminosity and redshift. This allows us to obtain a more accurate estimate for the redshift of the GW signal, which is crucial for cosmological measurements. The prior is constructed by dividing the sky into HEALPix pixels, and computing the redshift distribution of galaxies in each pixel. The redshift prior is then weighted using the luminosity of the potential host galaxies, allowing us to obtain a more accurate estimate for the Hubble constant.

Furthermore, the prior accounts for the incompleteness of the galaxy catalog, using source population models and the magnitude threshold calculated per pixel, which is particularly important for high-redshift events where the number of galaxies is significantly reduced. This allows us to separate the LOS redshift prior into an in-catalog and out-of-catalog contribution.

Taking into account, the fact that the host galaxy can be present, or not, inside the catalog, one can write the LOS redshift prior as:

$$p(z|\Omega_i, \Lambda, s, I) = \iint \sum_{g=G, \bar{G}} p(z, M, m, g|\Omega_i, \Lambda, s, I) dM dm \quad (5.1)$$

$$= p(G|\Omega_i, \Lambda, s, I) \iint p(z, M, m|G, \Omega_i, \Lambda, s, I) dM dm + p(\bar{G}|\Omega_i, \Lambda, s, I) \iint p(z, M, m|\bar{G}, \Omega_i, \Lambda, s, I) dM dm \quad (5.2)$$

The first term in the equation represents the contribution from galaxies that are present in the catalog, while the second term represents the contribution from galaxies outside the catalog. The two terms are weighted by their respective probabilities of being present or not in the catalog. The terms inside the integral are the priors on the redshift z , absolute magnitude M , and apparent magnitude of the galaxies m , informed by the galaxy catalog, within the sky area covered by pixel i . Here the parameters G/\bar{G} give the presence or absence of the galaxy in the catalog, Ω_i the sky location of the GW event, Λ the cosmological and

population hyperparameters of interest, s the presence of a GW source, and I the additional assumptions which are not explicitly expressed. One also needs to marginalize over the absolute magnitude M and the apparent magnitude m of the galaxy, as these determine, to the leading order, which galaxies are present in a flux-limited EM survey (Gray et al., 2023).

The integral in the in-catalog term can be expressed as the sum over the possible host galaxies in the catalog, weighted by their respective probabilities of being the host galaxy. These galaxies are weighted by their luminosity, which is a function of the absolute magnitude and redshift. The rationale being that the more luminous, and thus heavier galaxies are more likely to host CBC events, and therefore contribute more to the LOS redshift prior. This reduces the in-catalog part to a weighted sum over the galaxies in the catalog, where the galaxies are treated as point sources modeled by a Gaussian. This term can thus be expressed as:

$$\begin{aligned} \iint p(z, M, m | G, \Omega_i, \Lambda, s, I) dM dm &= \frac{1}{p(s|G, \Omega_i, \Lambda, I) N_{gal}(\Omega_i)} \\ &\times \sum_k^{N_{gal}(\Omega_i)} p(z|\hat{z}_k) p(s|z, M(z, \hat{m}_k, \Lambda), \Lambda, I) \end{aligned} \quad (5.3)$$

where the term $p(z|\hat{z}_k)$ represents the probability of a galaxy being at redshift z , given its observed redshift \hat{z}_k . This term is used to weight the contribution from each galaxy in the catalog, based on its observed redshift.

The integral in the out-of-catalog term marginalizes over the possible host galaxies not present in the catalog. This term is more complex, as it requires a model for the distribution of galaxies in the Universe, which is not directly available from the catalog. We use a Schechter luminosity function to model the distribution of galaxies in the Universe, which allows us to estimate the contribution from galaxies outside the catalog. The out-of-catalog term can be expressed as:

$$\begin{aligned} \iint p(z, M, m | \bar{G}, \Omega_i, \Lambda, s, I) dM dm &= \frac{1}{p(s|\bar{G}, \Omega_i, \Lambda, I) p(\bar{G}|\Omega_i, \Lambda, I)} \\ &\times \left[\Theta[z_{cut} - z] \int_{M(z, m_{th}(\Omega_i), \Lambda)}^{M_{max}(H_0)} p(z, M|\Lambda, I) p(s|z, M, \Lambda, I) dM \right. \\ &\quad \left. + \Theta[z - z_{cut}] \int_{M_{min}(H_0)}^{M_{max}(H_0)} p(z, M|\Lambda, I) p(s|z, M, \Lambda, I) dM \right] \end{aligned} \quad (5.4)$$

Here we also account for the EM selection effects of the catalog. Due to the flux limited nature of the galaxy catalog, the probability of a galaxy being present in the catalog depends on the galaxy's apparent magnitude, and whether it is greater or smaller than apparent magnitude threshold of the catalog along the same line of sight, $m_{th}(\Omega_i)$. The Heaviside function Θ is used to separate the two cases of the out-of-catalog contribution, depending on whether the redshift z is below or above a certain threshold z_{cut} . This is due to the exclusion of galaxies with redshift z greater than z_{cut} from the catalog, which is a result of unreliable redshift or color information at these higher redshifts.

The term $p(s|z, M, \Lambda, I)$ is the weighting factor for the contribution from each galaxy, based on its luminosity and redshift. The galaxies are weighted by their luminosity in the K -band, which is a good tracer of the mass of the galaxies (Strazzullo et al., 2006; Sureshkumar et al., 2021). Furthermore, the merger host probability is also taken into account, which is a function of the redshift. This is modeled by a Madau-Dickinson merger rate evolution model (Madau and Dickinson, 2014), which describes the evolution of the merger rate with redshift, discussed in Section 4.1.4. The term $p(s|z, M, \Lambda, I)$, also incorporates the source population models, used to pouplate the out-of-catalog contribution. These models are also discussed in Section 4.1.4.

The term $p(z, M|\Lambda, I)$ represents the luminosity funtion of the galaxies, taken to be the Schechter luminosity function, discussed in Section 4.2.2. The integration limits are set by the minimum and maximum absolute magnitudes of the galaxies, $M_{min}(H_0)$ and $M_{max}(H_0)$. These are H_0 -dependent, as the parameters of the Schechter luminosity function are also H_0 -dependent, but the final distribution remains insensitive to the exact values of H_0 (Gray et al., 2023).

5.3 Results

In this section we present the main outcomes of our analysis: how applying brightness cuts to the GLADE+ catalog alters LOS redshift prior for individual events, how these modified priors propagate into the inferred posterior on the Hubble constant H_0 . It should be noted that while `gwcosmo` is designed to perform joint population and cosmological inference, we focus here on the impact of the brightness cuts on the LOS redshift prior and the resulting H_0 posterior. For this reason, we do not perform a full joint inference, but rather use `gwcosmo` for H_0 inference, while keeping the population hyperparameters fixed to the values mentioned in Section 4.1.4.

To test the impact of catalog completeness, we apply successive cuts to the galaxy catalog by selecting only the brightest $XX\%$ of galaxies in K -band luminosity, forming subsets labeled GLADEPXX. This process shifts the effective redshift upwards, reducing the number of galaxies but improving the catalog's completeness.

5.3.1 LOS Redshift Prior

Figure 5.3 shows the LOS redshift prior for the event **GW170809** under different catalog cuts. Bright galaxy subsets (e.g., GLADEP20) result in sharper redshift distributions and reduce the weight of the out-of-catalog term. The low- z tail contributed by faint galaxies in the nearby Universe is also suppressed. Notably, these priors maintain consistency in shape, indicating that bright galaxies are reliable tracers of large-scale structure. The application of a brightness cut significantly modifies the LOS redshift distribution. Specifically, the bright galaxy subsets yield an amplified redshift prior at higher distances, effectively extending the reach of the catalogue, somewhat mitigating the incompleteness issues that arise at deeper redshifts. This behavior is qualitatively consistent across events.

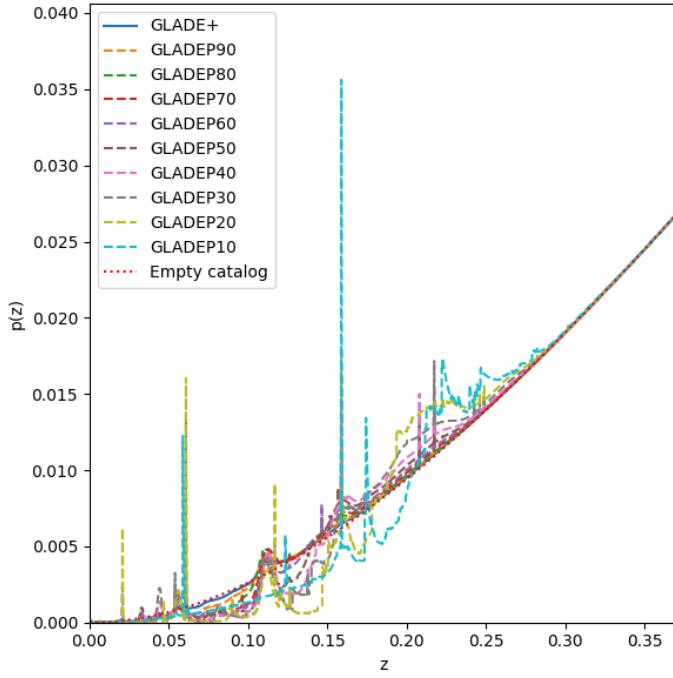


Fig. 5.3: LOS redshift prior for GW170809 for different brightness-ranked GLADE subsets. The full GLADE+ catalog (solid blue) is compared to the different subsets. Applying a brightness cut amplifies the prior at higher distances, extending the effective reach of the catalog and partially mitigating incompleteness at greater redshifts.

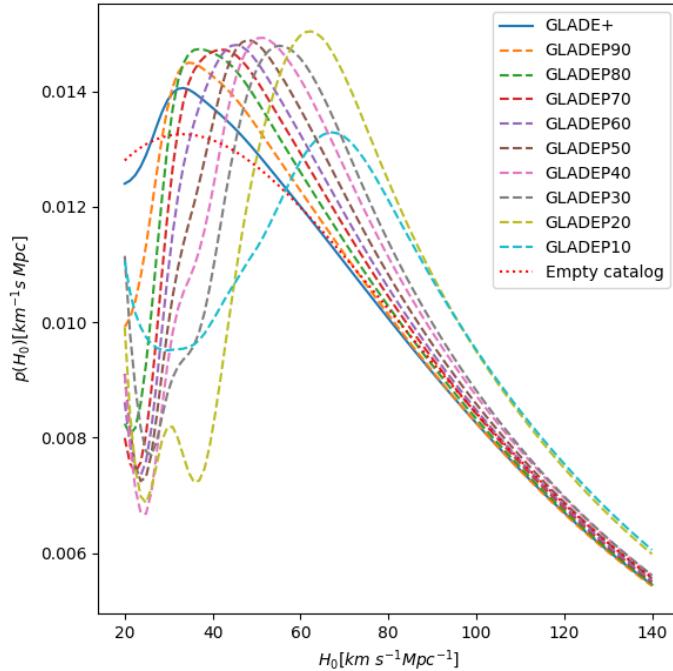


Fig. 5.4: H_0 posterior distributions for GW170809 using full and brightness-cut galaxy catalogs. GLADEP20 yields the tightest credible interval. GLADEP10 suffers from sparse sampling.

5.3.2 Hubble Constant Posterior

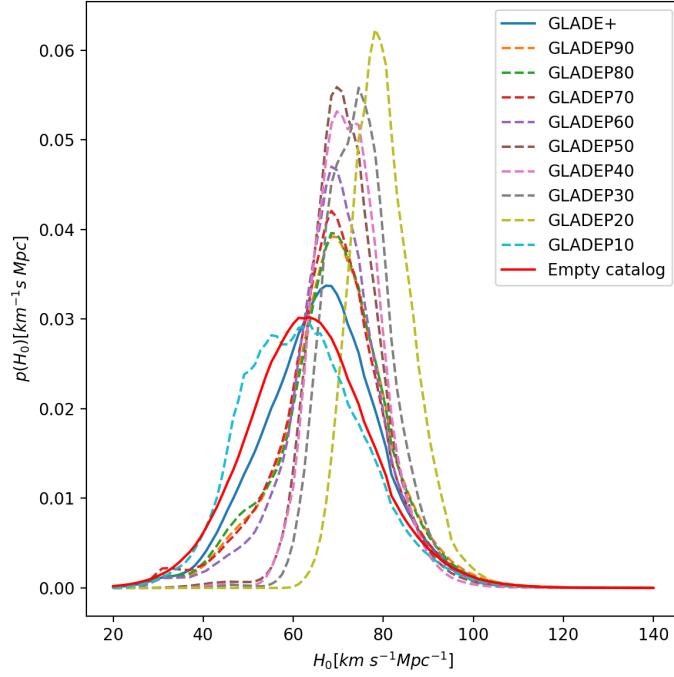


Fig. 5.5: Cumulative H_0 posterior from the selected dark siren events using full GLADE+ (blue) and the different subsets (dashed lines). The brightness-weighted catalog yields a tighter constraint without a significant shift.

Once one has the LOS redshift prior, one can use it to compute the posterior distribution of the Hubble constant H_0 given the GW event data, as the relation between the redshift z , luminosity distance d_L and the Hubble constant H_0 is given by:

$$d_L = \frac{c(1+z)}{H_0} \int_0^z \frac{dz'}{\sqrt{(1+z')^3 \Omega_m + \Omega_\Lambda}} \quad (5.5)$$

for a flat Λ CDM cosmology. At lower redshifts, this can be simplified to:

$$d_L \approx \frac{cz}{H_0} \quad (5.6)$$

Thus the LOS redshift prior, $p(z|\Omega_i, \Lambda, s, I)$, and the luminosity distance posterior, $p(d_L, \Omega | \text{GW})$, from the GW event data can be combined to obtain the posterior distribution of H_0 , which can then be used to obtain constraints on the value of H_0 .

Figure 5.4 shows the resulting H_0 posteriors for GW170809 under various catalog cuts. As the catalog is restricted to brighter galaxies, the H_0 posterior becomes increasingly narrow. For example, GLADEP20 yields a visibly tighter posterior compared to the full GLADE+, with a shift in the median. However, the most aggressive cut (GLADEP10) introduces broader tails, likely due to insufficient galaxy sampling in the localization volume.

We repeat this procedure for a subset of BBH events from GWTC-3 that meet our selection criteria. Figure 5.5 shows the combined H_0 posterior from the selected dark siren events

Tab. 5.1: Maximum *a posteriori* probabilities with 68% confidence ranges of the H_0 posterior distributions alongside the maximum magnitude limits for the different percentiles of the GLADE+ galaxy catalogue.

Catalogue	$M_{K,max}$	$H_0 [km s^{-1} Mpc^{-1}]$
GLADE+	-19.00	$67.87^{+8.97}_{-10.29}$
GLADEP90	-23.07	$68.94^{+9.24}_{-7.55}$
GLADEP80	-23.62	$68.93^{+9.25}_{-7.57}$
GLADEP70	-23.94	$68.63^{+8.61}_{-7.42}$
GLADEP60	-24.19	$68.85^{+7.72}_{-6.43}$
GLADEP50	-24.14	$70.05^{+6.12}_{-5.20}$
GLADEP40	-24.63	$69.94^{+7.46}_{-3.88}$
GLADEP30	-24.87	$74.70^{+4.69}_{-6.58}$
GLADEP20	-25.15	$78.28^{+5.53}_{-4.95}$
GLADEP10	-25.53	$63.46^{+6.39}_{-14.37}$

using the full GLADE+ catalog and the different subsets. Table 5.1 summarizes the H_0 posteriors for all cuts. The uncertainty is minimized around the GLADEP20 subset, showing 30-40% tighter constraints, while median values shift towards a higher value. While these tighter constraints are promising, the shift in median values could indicate a systematic bias introduced by the brightness cuts. This is particularly evident for the most extreme cut (GLADEP10), which yield a significantly lower median H_0 value, likely due to under-sampling and loss of information about the large-scale structure. For less extreme cuts, the median values doesn't show a huge shift. This suggests that moderate brightness cuts could improve precision without introducing much systematic bias.

5.3.3 Luminosity Retention

An important consideration when applying brightness-based cuts to galaxy catalogs is understanding how much of the total luminosity is retained as a function of the chosen percentile cut. Figure 5.6 shows the cumulative fraction of total K -band luminosity retained when selecting the brightest $X\%$ of galaxies in the GLADE+ catalog.

As the figure illustrates, the retained luminosity decreases rapidly when pruning the catalog based on brightness. At first glance, this might suggest that discarding the fainter galaxies results in the loss of a significant portion of cosmologically relevant information. However, this effect is largely driven by the overwhelming number of faint galaxies, each contributing only marginally to the total light output.

It is also important to note that the selection is made using the brightest percentiles of the catalog itself, rather than a parametric cut based on a fitted Schechter function. This empirical approach includes all galaxies below a certain magnitude threshold, which accentuates the cumulative effect of the numerous low-luminosity galaxies and contributes to the pronounced loss observed in total luminosity.

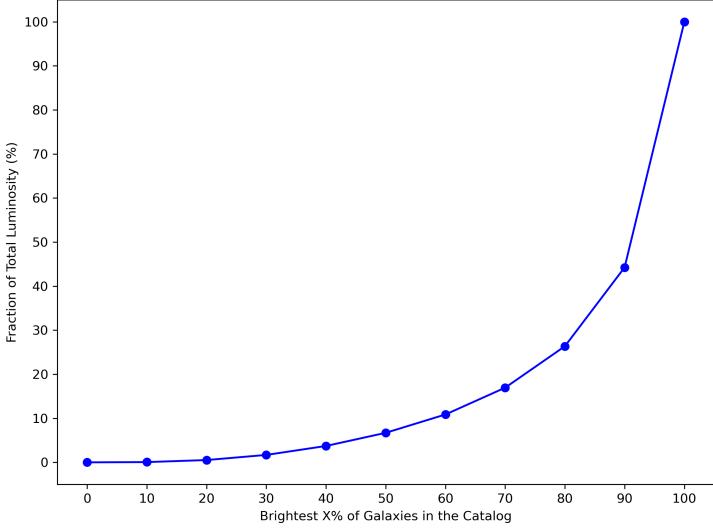


Fig. 5.6: Fraction of total K -band luminosity as a function of the brightest $X\%$ of galaxies in the catalog. The steep slope near 100% reflects the contribution of the numerous dim galaxies to the total luminosity budget.¹

The bright galaxies we retain are intrinsically more luminous and are typically associated with massive halos and central locations in galaxy groups or clusters. These bright, high-mass galaxies are more likely to trace the underlying matter distribution, which is essential for our purpose. By focusing on the top percentiles in brightness, we sacrifice luminosity but preserve the most structurally informative objects.

Therefore, while the pruning strategy leads to a steep loss in integrated light (as shown in the figure), this primarily reflects the exclusion of numerous dim galaxies that do not significantly enhance the galaxy density field used to construct the redshift prior. The retained bright galaxies should still provide a robust tracer of large-scale structure, particularly at higher redshifts where catalog completeness begins to decline.

5.3.4 Cost-Benefit Trade-Off

We observe that the precision of the H_0 posterior improves as the catalog is pruned to include only the brightest galaxies, up to an optimal point (around GLADEP20). Beyond this point, the constraints begin to degrade, with the posterior broadening and approaching the result obtained with an empty catalog. This degradation arises from under-sampling the galaxy distribution, highlighting a practical limit to brightness-based pruning.

While brightness cuts can enhance statistical precision by removing noisy or weakly informative galaxies, they also risk excluding key structure-tracing galaxies if pushed too far. This trade-off between precision and completeness must be carefully managed, as overly aggressive cuts can introduce bias and weaken the reliability of cosmological inference. The

¹This plot was generated using code partially adapted from scripts provided by Cezary Turski.

exact location of the optimal cut may depend on the specifics of the GW event, such as its localization volume, and the characteristics of the underlying galaxy catalog.

This balance is best understood within the framework of the Mock Data Challenge (MDC), which is discussed in the next chapter. The MDC allows us to systematically test how different cuts affect the inferred H_0 , providing a controlled environment to evaluate potential biases and performance.

Our results suggest that targeting the brightest galaxies in a well-characterized catalog can substantially enhance the precision of H_0 measurements from dark sirens. This method serves as a complementary approach to other strategies for mitigating catalog incompleteness, such as joint inference with galaxy clustering or population-based redshift estimation.

Mock Data Challenge

To validate the method developed in this thesis, we conduct a Mock Data Challenge (MDC) using simulated GW events and a mock galaxy catalog. The goal of this challenge is to assess the robustness of our inference pipelines and the impact of different brightness cuts on the measurements derived from GW observations. The MDC is designed to mimic real-world scenarios, allowing us to evaluate the performance of our methods under controlled conditions. This is done with the help of the BUZZARD mock catalogs, which provide a realistic simulation of the galaxy distribution and properties in the universe.

The MDC will involve generating synthetic GW events, simulating their detection by the LIGO, Virgo, KAGRA (LVK) detectors, and analyzing the resulting data using the `gwcosmo` inference pipelines. It is designed to determine the optimal brightness threshold which maximizes the measurement precision with bright subsets while keeping the biases in control. Tests performed in the process will help refine the methodology and ensure that the improved constraints on the Hubble constant are robust against additional systematic uncertainties.

6.1 The BUZZARD Mock Catalog

Cosmological analyses increasingly rely on simulated galaxy catalogs to test, validate, and calibrate inference pipelines. Among the most sophisticated of these are the BUZZARD mock catalogs presented by DeRose et al. (2019), a suite of synthetic sky simulations designed to closely emulate observations from the Dark Energy Survey (DES). The BUZZARD project aims to produce realistic mock universes that replicate the key statistical, spatial, and photometric properties of large-area surveys (DeRose et al., 2019; DeRose et al., 2022). These mocks are particularly useful for evaluating systematic uncertainties, testing survey strategies, and validating cosmological pipelines, including those involving standard sirens.

The BUZZARD mocks are constructed through a multi-step process that integrates large-scale dark matter simulations with empirical galaxy assignment and ray-traced gravitational lensing. The underlying dark matter distribution is generated using N -body simulations run using the 2LPTIC (Crocce et al., 2006) and LGADGET2 code (Springel, 2005). The simulations follow the evolution of particles under gravity across large comoving volumes, capturing the hierarchical growth of structure in a flat Λ CDM cosmology. The cosmological parameters used in the simulation are detailed in Table 6.1. Three lightcones covering different redshift regimes, denoted L1, L2, and L3, are used to build a deep mock catalog that spans from redshift $z \sim 0$ to $z \sim 2.35$, with increasing resolution and particle number at lower redshifts to mimic the increasing survey completeness at low z .

Tab. 6.1: The Λ CDM cosmology parameters used in the BUZZARD simulations.

Parameter	Description	Value
Ω_m	Matter density parameter	0.286
H_0	Hubble constant	$70 \text{ km s}^{-1} \text{ Mpc}^{-1}$
σ_8	Matter density fluctuation quantifying the clumpiness of matter distribution	0.815
n_s	Scalar spectral index: scale dependence of the primordial density perturbations	0.965
Ω_b	Baryon density parameter	0.046
N_{eff}	Effective number of neutrino species	3.046

Dark matter halos are identified in the simulation snapshots using the ROCKSTAR halo finder (Behroozi et al., 2012a), which groups particles into gravitationally bound objects based on adaptive phase-space density estimation. Temporal halo merger trees are constructed with the CONSISTENT-TREES algorithm (Behroozi et al., 2012b), enabling tracking of individual halos across cosmic time. These trees are crucial for assigning realistic galaxy properties and modeling assembly history.

Once the halo population is established, galaxies are inserted using the Adding Density-Determined Galaxies to Lightcone Simulations (ADDGALS) algorithm (Wechsler et al., 2022). ADDGALS is a semi-empirical prescription that associates galaxies with dark matter particles based on the local density field, using relationships inferred from observations such as the Sloan Digital Sky Survey (SDSS). Specifically, the algorithm calibrates the conditional luminosity function, the probability of a galaxy having a given luminosity in a particular local density environment, such that the final galaxy sample reproduces the observed luminosity function and galaxy clustering statistics (DeRose et al., 2019; Wechsler et al., 2022).

Each synthetic galaxy is assigned rest-frame and observed-frame photometry in multiple DES bands (e.g., g , r , i , z), drawn from a training set derived from real SDSS data (DeRose et al., 2019). This ensures that the color-magnitude distribution of BUZZARD galaxies matches the data, which is essential for generating realistic photometric redshifts. Photometric errors and selection effects are modeled to mimic DES observations, including depth variations and photometric scatter.

An important feature of BUZZARD is the inclusion of weak gravitational lensing through full-sky ray-tracing with the CALCLENS algorithm (Becker, 2013). CALCLENS computes the lensing convergence, shear, and magnification along each line of sight by integrating over the matter distribution in the simulation. These quantities are then assigned to each galaxy, enabling synthetic lensing measurements. This is particularly important for cosmological studies involving large-scale structure, cosmic shear, and galaxy bias calibration.

The final BUZZARD realizations contain over 800 million galaxies each and cover more than 10,000 square degrees of sky, comparable to the footprint of the full DES. The mocks include a variety of galaxy samples such as REDMAGIC (a luminous red galaxy sample with low photo-z scatter) and METACALIBRATION (used for weak lensing shear measurements). These samples are used in key DES analyses and their inclusion in BUZZARD makes the mock catalogs highly relevant for pipeline validation.

BUZZARD has been extensively validated against DES Year 1 (Y1)–DES Year 3 (Y3) data, demonstrating good agreement in angular correlation functions, redshift distributions, galaxy-galaxy lensing signals, and shear measurements. The photometric redshift performance in BUZZARD mimics that of DES pipelines, allowing realistic estimation of redshift uncertainty impacts (DeRose et al., 2019; DeRose et al., 2022). These qualities make BUZZARD an ideal environment for end-to-end tests of cosmological inference workflows.

The first BUZZARD data release, BUZZARD DR1, has been made publicly available. It consists of multiple simulations to generate a mock galaxy catalog covering the same sky area as the Dark Energy Survey (DES) surveys. It contains more than 3 billion galaxies with realistic properties, including redshift distributions, luminosity functions, and clustering statistics. In order to facilitate simulated analyses of existing and upcoming large scale structure surveys, photometry is provided in multiple bands, including the DECam (u, g, r, i, z, Y), VISTA (z, Y, J, H, K_s), WISE ($W1, W2$), Rubin (u, g, r, i, z, Y), and Roman (Y, J, H, K) passbands.

In this thesis, the BUZZARD mock catalog is used to evaluate the impact of galaxy catalog properties on the measurement of the Hubble constant H_0 from dark sirens. Specifically, they allow controlled tests of how catalog pruning, for instance, selecting only the brightest galaxies, affects the redshift prior and the resulting cosmological inference. Because BUZZARD includes realistic magnitude distributions, redshift evolution, and clustering properties, it provides a physically motivated and statistically robust testbed for these experiments.

For our analysis, we use the Roman K -band luminosities of the galaxies in the BUZZARD catalog, based on the rationale that K -band luminosities are good tracers for the stellar mass (Strazzullo et al., 2006; Sureshkumar et al., 2021). Furthermore, we apply an apparent magnitude threshold of 18 to the catalog to replicate realistic observational conditions, albeit with a somewhat optimistic depth. This apparent magnitude cut is chosen to ensure that the catalog contains a sufficient number of galaxies for statistical analysis. The resulting catalog, referred to as BUZZARD_{m18}, contains approximately 9.2 million galaxies with K -band apparent magnitudes brighter than 18, spanning redshifts from $z = 0$ to $z = 2.35$. The BUZZARD catalog is then cross-matched with the GW events generated by GWSim, detailed in the next section, to create a realistic dataset for our analysis. We then use this apparent magnitude limited catalog to select only the brightest galaxies, in order to test our results from GLADE+.

The use of BUZZARD thus supports one of the main goals of this thesis: to develop and validate a method for extracting cosmological information from gravitational-wave observations without electromagnetic counterparts, using only statistically constructed galaxy redshift priors. By using BUZZARD to test this methodology on a simulated sky, we can quantify biases, evaluate uncertainties, and refine our technique.

6.2 GWSim

To validate our methodology and assess the performance of standard siren inference using galaxy catalogs, we require a controlled and physically motivated sample of GW events with known properties. For this purpose, we use `GWSim` (Karathanasis et al., 2023), a publicly available Python package designed to generate mock catalogs of GW events from BBH mergers, incorporating both astrophysical source population models and cosmological assumptions.

`GWSim` enables simulation of GW events from a wide range of mass, spin, and merger rate distributions, within cosmological frameworks such as flat Λ CDM, w_0 CDM, and w_0-w_a CDM models. Each mock event is associated with a redshift and sky location, and can be assigned to a host galaxy either by sampling from an isotropic homogeneous distribution or by drawing directly from a user-provided galaxy catalog. In our analysis, we use `GWSim` in conjunction with the complete BUZZARD mock catalog to create realistic GW injections associated with galaxies that replicate the large-scale structure of the Universe.

The simulation pipeline consists of several modular steps:

- **Source Population Modeling**

The intrinsic BBH population is defined using a combination of merger rate models (e.g., constant, phenomenological, or delay-time models), mass distributions (e.g., power-law, power-law + Gaussian, or multi-peak), and spin distributions (e.g., uniform, Gaussian, or mass-correlated). We adopt a truncated power-law with Gaussian peak for the mass distribution and assume no spin dependence for simplicity. The same source population model discussed in Section 4.1.4 is used in order to isolate the effects of the galaxy catalog on H_0 inference.

- **Cosmological Framework**

For each GW source, the redshift and luminosity distance are related through cosmological parameters. In this study, we assume a flat Λ CDM cosmology with the same parameters used in the BUZZARD simulations (see Table 6.1).

- **Host Galaxy Assignment**

GW events are assigned to galaxies in the BUZZARD catalog using a merger-rate-weighted sampling. We optionally apply luminosity weighting in the K -band to preferentially select massive galaxies, as motivated by the correlation between K -band luminosity and stellar mass (Strazzullo et al., 2006; Sureshkumar et al., 2021). This is done by assigning the weight of each galaxy to be proportional to its Roman K -band luminosity. Assuming the same host galaxy weighting in the `gwcosmo` analysis and injection generation, detailed in Section 4, allows us to avoid the problem of incorrect host weighting in the inference step, which can lead to biased results.

- **Detection Modeling**

The GW strain and signal-to-noise ratio (SNR) are computed assuming the sensitivity

curves and duty cycles of the LIGO, Virgo, KAGRA (LVK) detector network for observation run O4 over a period of 1 year. Only events with network SNR exceeding a detection threshold ($\text{SNR} > 11$) are retained.

- **Parameter Estimation**

For each detected event, posterior samples of source parameters are generated using the Bilby package (Ashton et al., 2019; Ashton and Talbot, 2021). These include the luminosity distance, component masses, sky location, and binary inclination, forming the basis for cosmological inference.

6.3 H_0 Inference with `gwcosmo`

Once we have generated a set of mock GW events, we can cross-match them with the BUZZARD galaxy catalog and its brightest subsets to create realistic datasets for our analysis. For each catalog, this involves associating each event with its corresponding host galaxy, which is crucial for constructing the redshift prior and performing the cosmological inference. The cross-matching process takes into account the sky localization of the GW events and the spatial distribution of galaxies in the catalog, ensuring that we accurately capture the potential host galaxies for each event.

In order to model the out-of-catalog contribution to the redshift prior, we also need to account for the galaxies that are not included in the BUZZARD catalog. This is done by assuming a Schechter luminosity function for the galaxy population, as discussed in Section 4.2.2. Here, it has to be noted that the Roman K -band luminosity function in BUZZARD is not the same as the one used in earlier analysis using the GLADE+ catalog. For this reason, we estimate the Schechter function parameters for the BUZZARDm18 catalog, BUZZARD catalog with an apparent magnitude threshold of 18, using the $1/V_{max}$ estimator described in Schmidt (1968) and Takeuchi et al. (2000). This gives us a Schechter function with $\alpha = 1.26$, $M_K^* = -22.07 + 5 \log h$, $M_{K,min} = -26.00 + 5 \log h$, and $M_{K,max} = -18.00 + 5 \log h$.

Once we have the Schechter function parameters, we can use them to model the out-of-catalog contribution to the redshift prior. This allows us to construct the redshift prior for each GW event, and estimate the Hubble constant H_0 using `gwcosmo`, analogous to Chapter 5. Here it is to be noted that the source population model used in `gwcosmo` is the same as the one used in `GWSim`. Furthermore, the same cosmological parameters and host galaxy weighting are used in both `GWSim` injection generation and `gwcosmo` analysis to ensure consistency in the analysis. This is crucial for obtaining reliable results and minimizing biases in the inferred H_0 values, isolating the effects of the use of galaxy catalog subsets as redshift tracers.

The use of `GWSim` in conjunction with the BUZZARD catalog provides a self-consistent mock dataset that closely mirrors observational data. This allows us to assess how galaxy brightness cuts can affect the resulting H_0 posterior distributions and the overall performance of our inference pipelines.

6.4 Results

6.4.1 Luminosity Retention in the BUZZARD Catalog

As in the real-data analysis using GLADE+, we investigate the cumulative luminosity distribution of galaxies in the BUZZARD catalog to understand the effects of brightness pruning. Figure 6.1 shows the fraction of total luminosity retained as a function of the brightest percentile of galaxies.

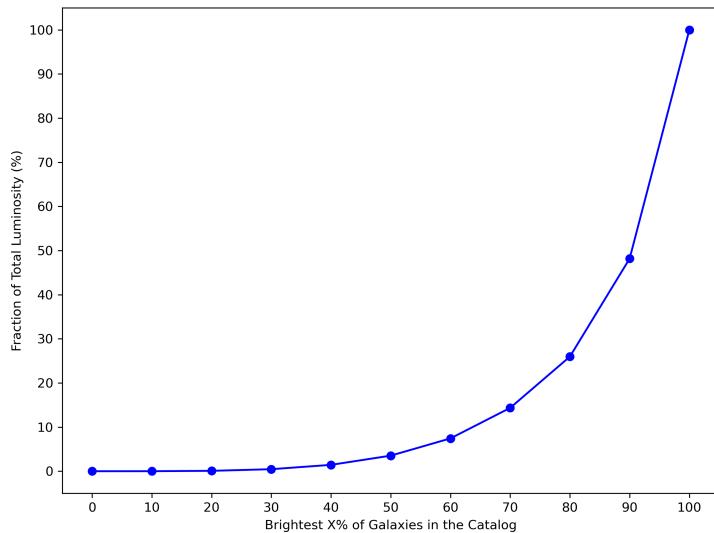


Fig. 6.1: Fraction of total K -band luminosity as a function of the brightest $X\%$ of galaxies in the BUZZARD catalog. As in the real catalog, a steep decline occurs, driven by the large population of faint galaxies.¹

This trend mirrors what was observed in the GLADE+ analysis, reinforcing the idea that although brightness pruning excludes a large number of faint galaxies, the retained bright galaxies remain effective tracers of large-scale structure. This is further discussed in the next section. The plot also highlights that brightness-based selection inherently sacrifices a large fraction of integrated luminosity, an effect amplified when using an empirical catalog rather than applying a parametric model like the Schechter function. Nevertheless, these retained galaxies are the most cosmologically informative for our purposes.

6.4.2 Tracing Large-Scale Structure with Luminous Galaxies

The BUZZARD mock catalog provides additional metadata for each galaxy, including its distance to the center of the nearest dark matter halo, denoted r_{halo} . This information is crucial for assessing the spatial relationship between galaxies and the underlying matter distribution, and in particular, for identifying potential host galaxies for GW events.

¹This plot was generated using code partially adapted from scripts provided by Cezary Turski.

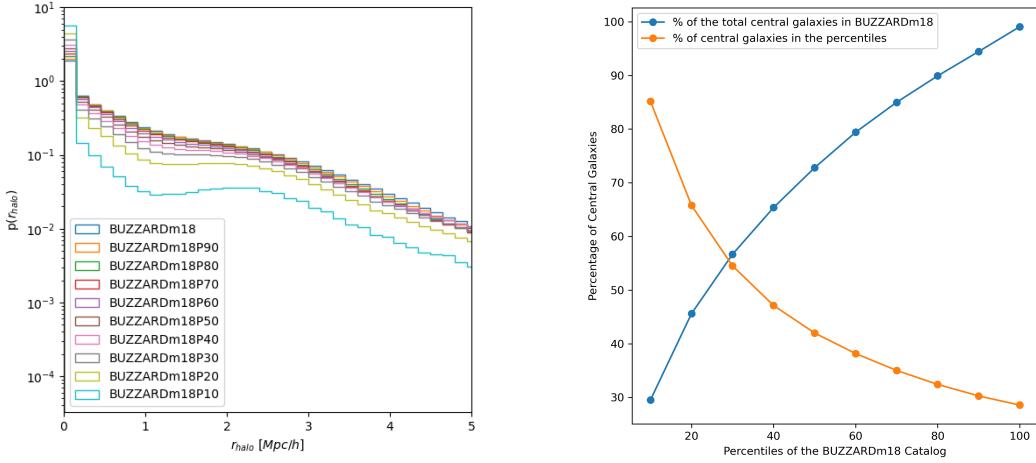


Fig. 6.2: **Left:** The distribution of the distance to the center of the nearest halo, r_{halo} , for the BUZZARDm18 catalog and its different percentiles, denoted BUZZARDm18PXX. **Right:** The evolution of the fraction of central galaxies in the BUZZARDm18 catalog as a function of the brightness cut. The central galaxy fraction is defined as the ratio of the number of central galaxies, $r_{\text{halo}} < 150 \text{ kpc}$, to the total number of galaxies in the catalog. The blue curve shows the remaining fraction of central galaxies when going from the BUZZARDm18 catalog to the BUZZARDm18PXX catalogs, while the orange curve shows the fraction of central galaxies in the BUZZARDm18PXX catalogs.

Figure 6.2 (left panel) shows the distribution of r_{halo} for the BUZZARDm18 catalog and for various brightness-ranked subsets, denoted BUZZARDm18PXX. The distribution peaks at small radii ($r_{\text{halo}} < 150 \text{ kpc}$), corresponding to galaxies located at or near halo centers, with a long tail extending to larger radii. This indicates that a significant portion of the catalog consists of galaxies embedded in halos.

To quantify the effect of brightness cuts on the galaxy-halo relationship, we define *central galaxies* as those with $r_{\text{halo}} < 150 \text{ kpc}$. The right panel of Figure 6.2 shows how the central galaxy fraction evolves across different brightness thresholds. The blue curve represents the retained fraction of central galaxies from the full BUZZARDm18 catalog, while the orange curve shows the fraction of central galaxies within each BUZZARDm18PXX subset.

These plots reveal a clear trend: as we move to brighter percentiles, a large number of outer (non-central) galaxies are removed, while a substantial fraction of central galaxies are retained. In fact, the relative proportion of central galaxies increases with stricter pruning. This suggests that brightness cuts selectively preserve galaxies that are more likely to trace the centers of halos, thereby maintaining sensitivity to the large-scale structure despite an overall reduction in the catalog size.

Table 6.2 further supports this observation. It shows the central galaxy fraction for different apparent magnitude thresholds. Across $m_{\text{thr}} = 14\text{--}18$, the central galaxy fraction remains roughly constant near 30%, but drops significantly for deeper thresholds (e.g., $m_{\text{thr}} = 20$) due to the increasing inclusion of dim, satellite, and field galaxies.

Tab. 6.2: Different apparent magnitude thresholds and their corresponding total number of galaxies and central galaxy fraction for the BUZZARD mock catalog.

m_{thr}	Total Number of Galaxies	Central Galaxy Fraction
14	41,018	0.28
16	588,841	0.29
18	9,232,370	0.29
20	90,635,665	0.13

This behavior highlights an important point: although pruning the catalog to its brightest percentiles results in the loss of a large fraction of the total luminosity, this loss is largely due to the removal of numerous faint galaxies. The retained luminous galaxies, while fewer in number, are more likely to reside at halo centers and act as effective tracers of the matter distribution. Thus, the information most relevant for dark siren cosmology is preserved.

However, caution is warranted at the extreme end of pruning. For example, in the BUZZARDm18P10 subset, while the majority of remaining galaxies are central, a significant number of central galaxies are also excluded. Given that only $\sim 30\%$ of galaxies in the full BUZZARDm18 catalog are central, the best-case scenario implies that pruning below 30% of the brightest galaxies inevitably leads to the loss of critical structure information and introduces potential bias.

This $\sim 30\%$ threshold appears robust across different magnitude cuts, as shown in Table 6.2. For very deep catalogs ($m_{thr} = 20$), the central galaxy fraction drops sharply, indicating a shift toward outer, dimmer galaxies that contribute less to the structure-tracing signal. Therefore, in our analysis with GLADE+, we should conservatively avoid cuts below this 30% threshold to prevent loss of critical information.

Furthermore, Figure 5.2 illustrates that pruning primarily removes galaxies in the nearby Universe, while preserving structure at higher redshifts. This opens the possibility of constructing hybrid redshift priors, using a complete catalog for low-redshift volumes and brightness-ranked subsets for deeper regions. However, such an approach would require carefully stitching together two different LOS redshift priors in a way that preserves smoothness and continuity, an effort that falls outside the scope of this thesis but represents an important direction for future work.

We should also be cautious about the fraction of central galaxies retained in the catalog, as excluding them disproportionately can bias the redshift distribution and thus affect cosmological inference. The BUZZARD catalog provides an ideal testbed to assess this impact in a controlled setting, where the true cosmology is known. Unfortunately, due to time constraints and technical challenges in processing and harmonizing the BUZZARD data, we were unable to conduct a full end-to-end bias quantification.

Nevertheless, the qualitative trends observed here and in partial mock runs strongly support the existence of a practical pruning floor around 30% brightness, below which inference becomes unreliable. The end-to-end MDC framework developed as part of this work is now

operational, requiring only minor refinements to support such systematic studies in the future.

Identifying the optimal brightness cut and studying the bias it introduces in inferred H_0 values will be a natural and impactful extension of this work. It would allow us to rigorously assess the trade-off between catalog completeness and cosmological precision, and ensure that our pruning strategies do not introduce systematic errors into dark siren cosmology.

Conclusion

The emergence of GW astronomy has opened a new frontier in observational cosmology. This thesis has explored the potential of using dark sirens, gravitational-wave events without electromagnetic counterparts, for the inference of the Hubble constant, H_0 . Specifically, we examined whether refining the galaxy catalog used in the redshift prior, by restricting it to its brightest subsets, can improve the cosmological constraints from these events.

7.1 Summary of Findings

We implemented a hierarchical Bayesian inference framework using the `gwcosmo` pipeline, which combines GW luminosity distance posteriors with redshift priors constructed from galaxy catalogs. The galaxy redshift prior was derived using the GLADE+ catalog and modified by applying brightness cuts to prioritize the most luminous galaxies in the K -band. These subsets, denoted GLADEPXX, were hypothesized to trace large-scale structure more efficiently due to their association with massive halos.

The analysis demonstrated that:

- Brightness-ranked catalogs improve the redshift prior, leading to tighter posteriors on H_0 by reducing low-likelihood host candidates in the GW localization volume.
- Moderate pruning yields improvement in H_0 precision as compared to using the full catalog, without introducing measurable bias. This improvement is attributed to the enhanced clustering signal from retaining mostly central, high-luminosity galaxies.
- Aggressive pruning (e.g., GLADEP10) results in degraded constraints and larger uncertainties, as the catalog no longer adequately traces the large-scale structure.
- A lower bound of approximately 30% of the brightest galaxies emerges as a practical pruning limit, in the best-case scenario where the brightest galaxies are strictly the central-most galaxies. Below this threshold, significant information about the underlying matter distribution is lost, particularly from central galaxies, increasing the risk of cosmological bias.
- Simulated mock challenges using the BUZZARD catalog support the robustness of the brightness-based pruning strategy but also emphasize the importance of catalog completeness, redshift depth, and optimal percentile thresholds. The 30% structural limit appears consistent across magnitude cuts, suggesting that future work should avoid overly aggressive pruning or instead develop hybrid approaches combining full and pruned catalogs.

7.2 Limitations

This work, while comprehensive, has several limitations:

- The number of GW events with high SNR and good localization remains small, limiting the statistical power of our results.
- Catalog incompleteness, particularly at high redshifts, introduces uncertainties that are only partially mitigated by brightness cuts.
- Due to time constraints and several technical challenges encountered with the BUZZARD mock catalog, we were unable to perform a full end-to-end bias quantification from pruning dim galaxies. Nevertheless, partial analyses and theoretical considerations support the proposed 30% threshold as a conservative lower bound.
- While a complete end-to-end Mock Data Challenge (MDC) framework has been developed, a few components still require minor refinements and consistency checks before full deployment in future analyses.

7.3 Outlook and Future Work

The outlook for dark siren cosmology is highly promising. With the advent of next-generation GW detectors such as Cosmic Explorer (Evans et al., 2021), Einstein Telescope (Abac et al., 2025) and LISA (Colpi et al., 2024), and deeper galaxy surveys (e.g., LSST (Ivezić et al., 2019), Euclid (Mellier et al., 2024)), both the number of detected events and the completeness of host catalogs are expected to improve significantly.

Future extensions of this work could include:

- Get a quantitative estimate of the bias introduced by the brightness cuts, with the devised end-to-end MDC framework.
- Testing for potential biases introduced by brightness cuts using larger and more realistic mock datasets.
- Developing adaptive redshift priors that vary with localization volume depth, combining full catalogs at low redshift with bright subsets at high redshift.
- Integrating clustering information or cross-correlations with large-scale structure to improve redshift inference (Afroz and Mukherjee, 2024).

Ultimately, the resolution of the Hubble tension will require a convergence of multiple independent probes. As shown in this thesis, dark sirens, when carefully analyzed, offer a robust and independent route to measuring H_0 . Furthermore, the use of bright galaxies as tracers of large-scale structure provides a promising avenue for improving the precision of cosmological constraints from GW observations. This methodology will play a growing role in the era of multi-messenger cosmology.

Bibliography

- Abac, Adrian, Raul Abramo, Simone Albanesi, et al. (2025). „The science of the Einstein telescope“. In: *arXiv e-print*, arXiv:2503.1226. arXiv: [2503.1226](#) (cit. on pp. 12, 50).
- Abbott, B. P. et al. (2017). „A gravitational-wave standard siren measurement of the Hubble constant“. In: *Nature* 551.7678, pp. 85–88. arXiv: [1710.05835 \[astro-ph.CO\]](#) (cit. on pp. 5, 11).
- Abbott, Benjamin P, R Abbott, TD Abbott, et al. (2016). „GW150914: First results from the search for binary black hole coalescence with Advanced LIGO“. In: *Physical Review D* 93.12, p. 122003 (cit. on p. 10).
- Abbott, Benjamin P et al. (2019). „GWTC-1: a gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs“. In: *Physical Review X* 9.3, p. 031040 (cit. on p. 23).
- Abbott, BP et al. (2021a). „A gravitational-wave measurement of the Hubble constant following the second observing run of advanced LIGO and Virgo“. In: *The Astrophysical Journal* 909.2, p. 218 (cit. on p. 12).
- Abbott, R et al. (2023a). „Constraints on the cosmic expansion history from GWTC-3“. In: *The Astrophysical Journal* (cit. on pp. 6, 11, 12, 14).
- Abbott, R. et al. (2024). „GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run“. In: *Physical Review D* 109.2, p. 022001 (cit. on p. 23).
- Abbott, Richard et al. (2021b). „GWTC-2: compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run“. In: *Physical Review X* 11.2, p. 021053 (cit. on p. 23).
- (2023b). „GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run“. In: *Physical Review X* 13.4, p. 041039 (cit. on p. 23).
 - (2023c). „GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run“. In: *Physical Review X* 13.4, p. 041039 (cit. on p. 24).
- Afroz, Samsuzzaman and Suvodip Mukherjee (2024). „Prospect of precision cosmology and testing general relativity using binary black holes–galaxies cross-correlation“. In: *Monthly Notices of the Royal Astronomical Society* 534.2, pp. 1283–1298 (cit. on p. 50).

- Aghanim, N. et al. (2020). „Planck 2018 results. VI. Cosmological parameters“. In: *Astron. Astrophys.* 641. [Erratum: *Astron. Astrophys.* 652, C4 (2021)], A6. arXiv: [1807.06209](https://arxiv.org/abs/1807.06209) [[astro-ph.CO](#)] (cit. on pp. 3, 10, 21).
- Alam, Shadab, Marie Aubert, Santiago Avila, et al. (2021). „Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory“. In: *Physical Review D* 103.8, p. 083533 (cit. on p. 4).
- Ashton, Gregory and Colm Talbot (2021). „Bilby-MCMC: an MCMC sampler for gravitational-wave inference“. In: *Mon. Not. Roy. Astron. Soc.* 507.2, pp. 2037–2051. arXiv: [2106.08730](https://arxiv.org/abs/2106.08730) [[gr-qc](#)] (cit. on p. 43).
- Ashton, Gregory et al. (2019). „BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy“. In: *Astrophys. J. Suppl.* 241.2, p. 27. arXiv: [1811.02042](https://arxiv.org/abs/1811.02042) [[astro-ph.IM](#)] (cit. on p. 43).
- Baade, Walter (1979a). „110. A Revision of the Extra-Galactic Distance Scale“. In: *A Source Book in Astronomy and Astrophysics, 1900–1975*. Ed. by Kenneth R. Lang and Owen Gingerich. Cambridge, MA and London, England: Harvard University Press, pp. 750–752 (cit. on p. 2).
- (1979b). „The resolution of Messier 32, NGC 205, and the central region of the Andromeda Nebula“. In: *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, pp. 744–749 (cit. on p. 2).
- Becker, Matthew R (2013). „CALCLENS: weak lensing simulations for large-area sky surveys and second-order effects in cosmic shear power spectra“. In: *Monthly Notices of the Royal Astronomical Society* 435.1, pp. 115–132 (cit. on p. 40).
- Behroozi, Peter S, Risa H Wechsler, and Hao-Yi Wu (2012a). „The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores“. In: *The Astrophysical Journal* 762.2, p. 109 (cit. on p. 40).
- Behroozi, Peter S, Risa H Wechsler, Hao-Yi Wu, et al. (2012b). „Gravitationally consistent halo catalogs and merger trees for precision cosmology“. In: *The Astrophysical Journal* 763.1, p. 18 (cit. on p. 40).
- Bilicki, Maciej, Thomas H Jarrett, John A Peacock, Michelle E Cluver, and Louise Steward (2013). „Two micron all sky survey photometric redshift catalog: A comprehensive three-dimensional census of the whole sky“. In: *The Astrophysical Journal Supplement Series* 210.1, p. 9 (cit. on p. 25).
- Bilicki, Maciej, John A Peacock, Thomas H Jarrett, et al. (2016). „WISE× SuperCOSMOS photometric redshift catalog: 20 million galaxies over 3π steradians“. In: *The Astrophysical Journal Supplement Series* 225.1, p. 5 (cit. on p. 25).
- CERN (Mar. 2025). *The Hubble tension; CERN Courier* (cit. on p. 5).
- Chen, Anson, Rachel Gray, and Tessa Baker (2024). „Testing the nature of gravitational wave propagation using dark sirens and galaxy catalogues“. In: *Journal of Cosmology and Astroparticle Physics* 2024.02, p. 035 (cit. on pp. 12, 13, 17–19).
- Colpi, Monica, Karsten Danzmann, Martin Hewitson, et al. (2024). „LISA definition study report“. In: *arXiv e-print*, arXiv:2402.07571. arXiv: [2402.07571](https://arxiv.org/abs/2402.07571) (cit. on pp. 12, 50).

- Crocce, Martín, Sebastián Pueblas, and Román Scoccimarro (2006). „Transients from initial conditions in cosmological simulations“. In: *Monthly Notices of the Royal Astronomical Society* 373.1, pp. 369–381 (cit. on p. 39).
- Cuceu, Andrei, James Farr, Pablo Lemos, and Andreu Font-Ribera (2019). „Baryon acoustic oscillations and the Hubble constant: past, present and future“. In: *Journal of Cosmology and Astroparticle Physics* 2019.10, p. 044 (cit. on p. 4).
- Dálya, Gergely, Gábor Galgócz, László Dobos, et al. (2018). „GLADE: A galaxy catalogue for multimessenger searches in the advanced gravitational-wave detector era“. In: *Monthly Notices of the Royal Astronomical Society* 479.2, pp. 2374–2381 (cit. on p. 25).
- Dálya, Gergely et al. (2022). „GLADE+: an extended galaxy catalogue for multimessenger searches with advanced gravitational-wave detectors“. In: *Monthly Notices of the Royal Astronomical Society* 514.1, pp. 1403–1411 (cit. on pp. 12, 25).
- De Vaucouleurs, G (1985). „Tycho’s supernova and the Hubble constant“. In: *Astrophysical Journal, Part 1 (ISSN 0004-637X)*, vol. 289, Feb. 1, 1985, p. 5-9. 289, pp. 5–9 (cit. on p. 2).
- DeRose, J. et al. (Jan. 2019). „The Buzzard Flock: Dark Energy Survey Synthetic Sky Catalogs“. In: arXiv: [1901.02401 \[astro-ph.CO\]](https://arxiv.org/abs/1901.02401) (cit. on pp. 29, 39–41).
- (2022). „Dark Energy Survey Year 3 results: Cosmology from combined galaxy clustering and lensing validation on cosmological simulations“. In: *Phys. Rev. D* 105.12, p. 123520. arXiv: [2105.13547 \[astro-ph.CO\]](https://arxiv.org/abs/2105.13547) (cit. on pp. 29, 39, 41).
- Evans, Matthew, Rana X Adhikari, Chaitanya Afle, et al. (2021). „A horizon study for cosmic explorer: science, observatories, and community“. In: *arXiv e-print, arXiv:2109.09882*. arXiv: [2109.09882](https://arxiv.org/abs/2109.09882) (cit. on pp. 12, 50).
- Ezquiaga, Jose María and Daniel E Holz (2022). „Spectral sirens: Cosmology from the full mass distribution of compact binaries“. In: *Physical Review Letters* 129.6, p. 061102 (cit. on p. 13).
- Freedman, Wendy L. and Barry F. Madore (2010). „The Hubble Constant“. In: *Annual Review of Astronomy and Astrophysics* 48. Volume 48, 2010, pp. 673–710 (cit. on p. 2).
- Freedman, Wendy L., Barry F. Madore, Brad K. Gibson, et al. (2001). „Final Results from the Hubble Space TelescopeKey Project to Measure the HubbleConstant“. In: *The Astrophysical Journal* 553.1, p. 47 (cit. on pp. 2, 3).
- Freedman, Wendy L., Barry F. Madore, In Sung Jang, et al. (2024). „Status report on the Chicago-Carnegie Hubble Program (CCHP): Three independent astrophysical determinations of the Hubble constant using the James Webb Space Telescope“. In: *arXiv e-prints*, arXiv-2408 (cit. on pp. 3, 4).
- Gray, Rachel et al. (2020). „Cosmological inference using gravitational wave standard sirens: A mock data analysis“. In: *Physical Review D* 101.12, p. 122001 (cit. on pp. 12, 13, 17, 18).
- (2022). „A pixelated approach to galaxy catalogue incompleteness: improving the dark siren measurement of the Hubble constant“. In: *Monthly Notices of the Royal Astronomical Society* 512.1, pp. 1127–1140 (cit. on pp. 12, 17).
 - (2023). „Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues“. In: *Journal of Cosmology and Astroparticle Physics* 2023.12, p. 023 (cit. on pp. 12, 13, 17, 21, 32, 33).

- Hubble, E. P. (1936). *Realm of the Nebulae* (cit. on p. 2).
- Hubble, Edwin (1929). „A relation between distance and radial velocity among extra-galactic nebulae“. In: *Proceedings of the National Academy of Sciences* 15.3, pp. 168–173. eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.15.3.168> (cit. on p. 1).
- Ivezić, Željko, Steven M Kahn, J Anthony Tyson, et al. (2019). „LSST: from science drivers to reference design and anticipated data products“. In: *The Astrophysical Journal* 873.2, p. 111 (cit. on p. 50).
- Kamionkowski, Marc and Adam G. Riess (2023). „The Hubble Tension and Early Dark Energy“. In: *Annual Review of Nuclear and Particle Science* 73. Volume 73, 2023, pp. 153–180 (cit. on p. 4).
- Karathanasis, Christos, Benoît Revenu, Suvodip Mukherjee, and Federico Stachurski (2023). „GWSim: Python package for creating mock GW samples for different astrophysical populations and cosmological models of binary black holes“. In: *Astronomy & Astrophysics* 677, A124 (cit. on p. 42).
- Kochanek, CS et al. (2001). „The k-band galaxy luminosity function“. In: *The Astrophysical Journal* 560.2, p. 566 (cit. on p. 26).
- Lauer, Tod R, Marc Postman, Michael A Strauss, Genevieve J Graves, and Nora E Chisari (2014). „Brightest cluster galaxies at the present epoch“. In: *The Astrophysical Journal* 797.2, p. 82 (cit. on p. 16).
- Lemaître, Georges (1927). „Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques“. In: *Annales de la Société Scientifique de Bruxelles*, A47, p. 49-59 47, pp. 49–59 (cit. on p. 1).
- Longair, Malcolm S. (2006). „The determination of cosmological parameters“. In: *The Cosmic Century: A History of Astrophysics and Cosmology*. Cambridge University Press, pp. 340–364 (cit. on p. 2).
- Lyke, Brad W, Alexandra N Higley, JN McLane, et al. (2020). „The Sloan Digital Sky Survey quasar catalog: Sixteenth data release“. In: *The Astrophysical Journal Supplement Series* 250.1, p. 8 (cit. on p. 25).
- Madau, Piero and Mark Dickinson (2014). „Cosmic star-formation history“. In: *Annual Review of Astronomy and Astrophysics* 52.1, pp. 415–486 (cit. on pp. 24, 33).
- Maggiore, Michele (Oct. 2007). *Gravitational Waves: Volume 1: Theory and Experiments*. Oxford University Press (cit. on p. 9).
- Makarov, Dmitry, Philippe Prugniel, Nataliya Terekhova, Hélène Courtois, and Isabelle Vauglin (2014). „HyperLEDA. III. The catalogue of extragalactic distances“. In: *Astronomy & Astrophysics* 570, A13 (cit. on p. 25).
- Mastrogiovanni, S et al. (2021). „On the importance of source population models for gravitational-wave cosmology“. In: *Physical Review D* 104.6, p. 062009 (cit. on p. 12).
- Mastrogiovanni, Simone et al. (2024). „ICAROGW: A python package for inference of astrophysical population properties of noisy, heterogeneous, and incomplete observations“. In: *Astronomy & Astrophysics* 682, A167 (cit. on pp. 12, 13).
- Mellier, Y, Abdurroúf Abdurroúf, JA Acevedo Barroso, et al. (2024). „Euclid. I. Overview of the Euclid mission“. In: *Astronomy & Astrophysics* (cit. on p. 50).

- Naveed, Khuzaifa, Cezary Turski, and Archisman Ghosh (2025). „Dark standard siren cosmology with bright galaxy subsets“. In: *arXiv preprint arXiv:2505.11268* (cit. on p. 27).
- Riess, Adam G, Gagandeep S Anand, Wenlong Yuan, et al. (2024). „JWST observations reject unrecognized crowding of cepheid photometry as an explanation for the hubble tension at 8σ confidence“. In: *The Astrophysical Journal Letters* 962.1, p. L17 (cit. on p. 3).
- Riess, Adam G., Stefano Casertano, Wenlong Yuan, Lucas M. Macri, and Dan Scolnic (2019). „Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond Λ CDM“. In: *Astrophys. J.* 876.1, p. 85. arXiv: [1903.07603 \[astro-ph.CO\]](https://arxiv.org/abs/1903.07603) (cit. on pp. 2, 10).
- Riess, Adam G, Wenlong Yuan, Lucas M Macri, et al. (2022). „A comprehensive measurement of the local value of the Hubble constant with 1 km s⁻¹ Mpc⁻¹ uncertainty from the Hubble Space Telescope and the SH0ES team“. In: *The Astrophysical journal letters* 934.1, p. L7 (cit. on pp. 2, 3, 21).
- Sandage, Allan (1958). „Current Problems in the Extragalactic Distance Scale.“ In: *Astrophysical Journal*, vol. 127, p. 513 127, p. 513 (cit. on p. 2).
- Sandage, Allan and GA Tammann (1982). „Steps toward the Hubble constant. VIII-The global value“. In: *Astrophysical Journal, Part 1*, vol. 256, May 15, 1982, p. 339-345. *Research supported by the Swiss National Science Foundation*; 256, pp. 339–345 (cit. on p. 2).
- Schmidt, Maarten (1968). „Space distribution and luminosity functions of quasi-stellar radio sources“. In: *Astrophysical Journal*, vol. 151, p. 393 151, p. 393 (cit. on p. 43).
- Schutz, Bernard F (1986). „Determining the Hubble constant from gravitational wave observations“. In: *Nature* 323.6086, pp. 310–311 (cit. on pp. 4, 5, 10).
- Skrutskie, Michael F, RM Cutri, R Stiening, et al. (2006). „The two micron all sky survey (2MASS)“. In: *The Astronomical Journal* 131.2, p. 1163 (cit. on p. 25).
- Soares-Santos, M. et al. (2019). „First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary–Black-hole Merger GW170814“. In: *Astrophys. J. Lett.* 876.1, p. L7. arXiv: [1901.01540 \[astro-ph.CO\]](https://arxiv.org/abs/1901.01540) (cit. on p. 5).
- Springel, Volker (2005). „The cosmological simulation code GADGET-2“. In: *Monthly notices of the royal astronomical society* 364.4, pp. 1105–1134 (cit. on p. 39).
- Strazzullo, Veronica, Piero Rosati, SA Stanford, et al. (2006). „The near-infrared luminosity function of cluster galaxies beyond redshift one“. In: *Astronomy & Astrophysics* 450.3, pp. 909–923 (cit. on pp. 14, 16, 26–28, 33, 41, 42).
- Sureshkumar, Unnikrishnan, A Durkalec, Agnieszka Pollo, et al. (2021). „Galaxy and Mass Assembly (GAMA)-Tracing galaxy environment using the marked correlation function“. In: *Astronomy & Astrophysics* 653, A35 (cit. on pp. 14, 16, 26–28, 33, 41, 42).
- Takeuchi, Tsutomu T, Kohji Yoshikawa, and Takako T Ishii (2000). „Tests of Statistical Methods for Estimating Galaxy LuminosityFunction and Applications to the Hubble Deep Field“. In: *The Astrophysical Journal Supplement Series* 129.1, p. 1 (cit. on p. 43).
- Vaucoleurs, G de (1972). „The velocity-distance relation and the Hubble constant for nearby groups of galaxies“. In: *Symposium-International Astronomical Union*. Vol. 44. Cambridge University Press, pp. 353–366 (cit. on p. 2).

Wechsler, Risa H, Joseph DeRose, Michael T Busha, et al. (2022). „ADDGALS: Simulated sky catalogs for wide field galaxy surveys“. In: *The Astrophysical Journal* 931.2, p. 145 (cit. on p. 40).

Zwicky, F (1942). „On the Frequency of Supernovae. II.“ In: *Astrophysical Journal, vol. 96*, p. 28 96, p. 28 (cit. on p. 2).