ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

School of Science
Department of Physics and Astronomy "A. Righi"
Master Degree Programme in Astrophysics and Cosmology

Who Speaks Louder?
Optimizing the Extraction of Information from Gravitational Waves for future Cosmological Analyses

Graduation Thesis

Presented by:
Martina Mori

Supervisor:
Prof. Michele Ennio Maria Moresco

Co-supervisors:
Nicola Borghi
Matteo Tagliazucchi

Academic Year 2024 - 2025

ἰοίην

Saffo, Fr. 182 V.

Abstract

Gravitational waves have emerged as a robust cosmological probe in an era where precision measurements increasingly challenge the completeness of the standard ΛCDM model. Future detectors upgrades will lead to a significant growth in the number of detections, with catalogs expected to reach orders of 10^5 events. These advances will enable tighter constraints on both cosmological parameters (e.g., the Hubble constant) and on the population properties of compact binaries. While such large data sets promise unprecedented statistical power, they also raise practical challenges: analyzing entire catalogs may become computationally prohibitive, and the information content of individual events may not be uniformly distributed. This thesis investigates how event selection strategies can optimize the extraction of information from large GW catalogs. Using simulated dark siren datasets derived from the MICE galaxy simulation with spectroscopic redshift uncertainties, cosmological and astrophysical parameters are inferred within a hierarchical Bayesian framework implemented in the `CHIMERA` pipeline. The analysis focuses on the impact of progressively stricter SNR thresholds, examining how constraints on cosmology and binary population parameters evolve as the event sample is reduced. Correlation studies and principal component analysis are employed to identify the dominant factors governing the informativeness of GW events. The analysis highlights how catalog subsampling can balance accuracy and computational demands, outlining strategies that will be increasingly important in the era of third–generation detectors.

Contents

Introduction

Cosmology has entered an era of precision, where the standard ΛCDM framework continues to provide an excellent description of most observations. The foundations of this model rest on the discovery that the Universe is expanding, first established through Hubble's observation of the linear relation between distance and recessional velocities, and, later, on measurements of distant Type Ia supernovae, that revealed that the expansion is accelerating. Together with measurements of the cosmic microwave background anisotropies and the large–scale distribution of galaxies, these observations established the ΛCDM model as the standard cosmological framework. The model assumes a spatially flat Universe dominated by cold dark matter and a cosmological constant (Λ), with baryons and radiation contributing only a minor fraction of the total energy density. This framework has proven remarkably successful, providing a consistent description of a wide range of cosmological probes, from the early Universe to late–time structure formation. At the same time, as measurements become increasingly precise, some tensions have emerged. The most debated is the discrepancy between estimates of the Hubble constant, H_0, inferred from the early Universe and from the local Universe. This so-called Hubble tension may point to new physics or to unaccounted systematics, but in either case it underlines the importance of developing independent probes of the cosmic expansion history (Verde, Treu and Riess (2019); Abdalla et al. (2022); Moresco et al. (2022)).

Gravitational waves represent one of the most promising probes in this context. Sources of GWs such as compact binary coalescences act as standard sirens, providing direct measurements of the luminosity distance without the need for external calibration. When combined with redshift information, they can be used to constrain H_0 and other

cosmological parameters (Schutz (1986)). Since the first detection of gravitational waves in 2015, the continuous improvement of the global network of detectors, with Advanced LIGO and Virgo upgrades, the addition of KAGRA, and future prospects like LIGO–India, has led to steadily increasing sensitivities and growing catalogs of events, with the most recent release (GWTC–4.0, Collaboration, Virgo Collaboration and KAGRA Collaboration (2025)) already including more than 200 candidates. Looking ahead, the advent of third–generation observatories such as the Einstein Telescope and Cosmic Explorer will bring an unprecedented number of observations, enhancing the role of GWs as precision tool for cosmology. This wealth of data promises extraordinary opportunities, but also raises new challenges related to data analysis, computational feasibility, and the identification of the most informative events. In particular, it will become crucial to assess whether every event contributes equally to parameter inference or whether a smaller subset carries most of the constraining power.

The central question of this work is therefore to explore the information content in GW; the goal is to derive strategies for optimizing the standard sirens analysis from a computation point of view, maximizing, at the same time, the scientific return in terms of robustness and accuracy of the constraints. This work is framed within this perspective, aiming to explore how gravitational waves detections can be used most effectively to extract cosmological and astrophysical information. Particular attention is given to the role of event selection and to the balance between scientific accuracy and computational feasibility in the era of large catalogs. We will base our analysis on specific GW simulations aimed at forecasting the behavior for the next O5 LVK observing run, to provide suggestions to optimize the analysis, also in view of the next generation third-generation GW detectors.

The thesis is organized as follows.

- Chapter 1 provides the theoretical background, introducing the cosmological framework, the fundamentals of general relativity and gravitational waves, and their use as standard sirens.
- Chapter 2 describes the methodology, presenting the hierarchical Bayesian framework, the models for the binary black hole population, and the `CHIMERA` pipeline.
- Chapter 3 introduces the simulated GW and galaxy catalogs and includes a detailed statistical analysis of their properties, with particular attention to correlation studies and dimensionality reduction.

- Chapter 4 presents the main results of the inference, examining both individual and combined constraints, the effect of SNR-based event selection, and the trade-off between accuracy and computational efficiency.
- Chapter 5 summarize the main results of our analysis, presenting also potential future developments.

1

Introduction to GWs

For nearly a century, Gravitational Waves (GW) were a theoretical prediction of General Relativity: subtle ripples of space-time produced by accelerating masses with a time-varying quadrupole moment. Their existence was firmly established only in 2015, with the first direct detection by the LIGO interferometers. This marked the beginning of gravitational wave astronomy and opened a new observational window onto the Universe.

Beyond confirming a key aspect of Einstein's theory, GW have quickly emerged as a promising tool for cosmology. In particular, signals from Compact Binary Coalescence (CBC) carry direct information on their distance, and can therefor be used as standard sirens.

This chapter provides the theoretical background: we first outline the cosmological framework and the observational tension surrounding the Hubble constant, then review the physics of gravitational waves and their detection, and finally introduce their role as cosmological probes.

1.1 The Cosmological Framework

Over the last century, numerous models have been developed to describe the large-scale properties and evolution of the Universe. The current adopted as Standard Cosmological Model is the Λ Cold Dark Matter (ΛCDM).

This model is built upon two primary assumptions:

General Relativity governs gravitational interactions. The evolution of space-time is described by Einstein's field equations, which relate the curvature of the Universe to its matter-energy content.

Cosmological Principle assumes that, on sufficiently large scales, the Universe looks the same in all directions and at all locations, so it is homogeneous and isotropic.

This involves a combination of energy components: baryonic matter, radiation, Cold Dark Matter (CDM), and a cosmological constant (Λ), interpreted as dark energy and introduced in the model to explain the accelerated expansion of the Universe which is assessed by observations.

1.1.1 The *FLRW* Metric and the Expanding Universe

A central consequence of general relativity applied to a homogeneous and isotropic Universe is the possibility for a non static universe. This was confirmed in 1929, when Edwin Hubble demonstrated through observations of distant galaxies that their spectral lines were systematically redshifted, indicating that galaxies are receding from us, as effect of the expansion of the Universe.

The velocity of recession was found to be proportional to their distance, a relationship now known as the *Hubble-Lemaître Law*:

$$v = H_0 \cdot d \tag{1.1}$$

where v is the recession velocity, d the proper distance to the galaxy, and H_0 the Hubble constant, representing the present-day expansion rate of the Universe.
This discovery provided the first direct observational evidence of cosmic expansion, replacing the notion of a static universe with a dynamic one.

Under the assumption of cosmological principle, the space-time metric can be written as *Friedmann Lemaître Robertson Walker (FLRW)* metric:

$$ds^2 = c^2dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right] \tag{1.2}$$

where ds^2 is the spacetime interval, c is the speed of light, $a(t)$ is the *scale factor*, describing how distances between cosmic objects evolve with cosmic time as effect of the expansion, ϕ and θ are the spherical coordinates and k is the *cosmological parameter*, describing the curvature of the geometry of space: $k = 0$ for a flat Universe, $k > 0$ for a closed Universe, and $k < 0$ for an open Universe.
The coordinates (r, θ, ϕ), called *comoving coordinates*, are fixed with respect to the expansion. The relation between physical and comoving distances is given by:

$$d_p = a(t)d_c \tag{1.3}$$

1.1.2 Elements of General Relativity

General Relativity (GR), proposed by Albert Einstein in 1915, revolutionized our understanding of gravity. Rather than interpreting gravity as a force acting at a distance as in Newtonian physics, GR interprets it as the effect of spacetime curvature induced by energy and momentum. Massive bodies cause spacetime to bend, and particles and light rays follow paths dictated by this curvature. In GR, spacetime is described as a four-dimensional differentiable manifold with a metric tensor $g_{\mu\nu}$, which defines the infinitesimal interval between two nearby events in this spacetime:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \tag{1.4}$$

Using natural units where the speed of light $c = 1$, the metric can be written in mixed form:

$$ds^2 = g_{00}dt^2 + 2g_{0i}dt\, dx^i + g_{ij}dx^i dx^j \tag{1.5}$$

where g_{00} corresponds to the time-time component, g_{ij} to space-space component, and g_{0i} to the mixed space-time component. The metric fully determines the causal and geometric structure of spacetime. In General Relativity, its form is constrained by Einstein's Field Equations, which relate spacetime curvature to the stress–energy tensor

$T_{\mu\nu}$ and to the cosmological constant Λ:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.6}$$

where $R_{\mu\nu}$ is the *Ricci tensor* and together with the *Ricci scalar* R provide the description of the curvature of spacetime.

1.1.3 Friedmann Equations

Applying Einstein's field equations (1.6) to the FLRW metric under the perfect fluid approximation we get the *Friedmann equations*, which provide a description of the evolution in time of the scale factor $a(t)$. These equations are:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2} \tag{1.7a}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{1.7b}$$

where ρ is the total energy density, p is the pressure of the cosmic fluid and we indicate with the dot the time derivatives.

Two crucial cosmological parameters emerge from these equations:

The **Hubble constant** H_0, defined as $H_0 = \dfrac{\dot{a}}{a}$ at the present time, quantifies the current expansion rate of the Universe.

The **density parameter** Ω, which describes the contribution of different components to the total energy density of the Universe. From the definition of the critical density derived by equating to zero Eq. (1.7a), while neglecting the cosmological component,

$$\rho_{crit} = \frac{3H_0^2}{8\pi G} \tag{1.8}$$

we can write the density parameters for different components as:

$$\Omega_i = \frac{\rho_i}{\rho_c} \tag{1.9}$$

Taking the sum of the densities of various components we can write the total

energy density of the universe:

$$\Omega_{TOT} = \sum \Omega_i = \Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1 \tag{1.10}$$

where Ω_m matter density, Ω_r is the radiation density, dark Ω_Λ is the cosmological component density and Ω_k represents the contribution of spatial curvature.

1.1.4 Equation of State and Cosmological Components

Friedmann equations are complemented by a third equation which is the *Equation of State*, determining the balance between density and pressure in a closed system with different fluid components. The standard form for Equation of State is:

$$p(t) = w(t)\rho(t) \tag{1.11}$$

In general $w_i(t)$ is taken as function of time, but in a given cosmological epoch it can be approximated with a constant value, w_i, depending on the dominant component of the epoch. With this assumption of a constant w_i, we can write the evolution of the energy density as

$$\rho_i(z) = \rho_{i,0}(1+z)^{3(1+w_i)} \tag{1.12}$$

in which each component results with a different ρ_i determined by the different w value:

Radiation (photons and relativistic particles): $\rho_r \propto a^{-4}$, $w = 1/3$

Matter (baryonic and cold dark matter): $\rho_m \propto a^{-3}$, $w = 0$

Dark Energy (cosmological constant): $\rho_\Lambda = constant$, $w = -1$

By combining Friedmann equations with 1.10 and 1.12 we can express the expansion rate in terms of redshift evolution of its main components:

$$H(z) = H_0 \left[\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_\Lambda\right]^{1/2} \tag{1.13}$$

1.1.5 Luminosity Distance

With the definition of the Hubble parameter, we can introduce the definition of *comoving distance* for the ΛCDM model:

$$d_c(z) = c\int_0^z \frac{dz'}{H(z')} = \frac{c}{H_0}\int_0^z \frac{dz'}{\sqrt{\Omega_R(1+z')^4 + \Omega_M(1+z)^3 + \Omega_\Lambda}} \tag{1.14}$$

which is related to the physical distance through the scale factor: $d_{phys} = a(z)d_c(z)$. From the definition of comoving distance, two fundamental quantities can be defined: the *angular diameter distance* and the *luminosity distance*. Recalling the expression of the scale factor a as a function of redshift, $a = 1/(1+z)$, we can insert it in the expression relating d_{phys} and d_{com} to get:

$$d_A(z) = \frac{c}{1+z}\int_0^z \frac{dz'}{H(z')} \tag{1.15}$$

which is the *angular diameter distance*.

We can also derive another distance definition, which is the one for *luminosity distance*, which is the quantity derived from the measured flux of a source while knowing its intrinsic luminosity:

$$F = \frac{L}{4\pi d_L^2}. \tag{1.16}$$

In terms of Hubble parameter we can express the luminosity distance as:

$$d_L(z) = c(1+z)\int_0^z \frac{dz'}{H(z')}. \tag{1.17}$$

In the local universe (low redshift) this simplifies to:

$$d_L(z) \approx \frac{cz}{H_0} \tag{1.18}$$

which tells us that with a measure of luminosity distance combined with redshift information we get a measurement of H_0. The role of GW in this framework enters because the amplitude of a GW signal from a CBC encodes a direct information on the luminosity distance of the source. For this reason, GW are referred as *standard sirens* (Sec. 1.3).

1.1.6 The Hubble tension

The expansion rate of the Universe, quantified by the Hubble constant H_0, can be measured in different ways depending on the cosmic epoch and on the distance tracers used. Over the past two decades, the rapid progress of observational cosmology has led to percent-level measurements of H_0, opening the era of so-called *precision cosmology*. But, instead of converging to a single value, different approaches have revealed a persistent discrepancy known as the *Hubble tension*.

A first family of methods relies on *standard candles*, astrophysical objects whose absolute luminosity is known. At short distances, Cepheid variables and RR Lyrae stars provide accurate distance estimates, while detached eclipsing binaries have fixed the distance to the Large Magellanic Cloud with high precision. At larger scales, stars at the Tip of the Red Giant Branch (TRGB) offer an additional standard candle, and Type Ia supernovae extend the ladder to cosmological distances (Moresco et al. (2022)). These distance-ladder techniques yield values of H_0 in the range $\sim$ 70–73 $\mathrm{km\,s^{-1}Mpc^{-1}}$, with the most precise Cepheid-calibrated ladder reporting $H_0 = 73.0 \pm 1.0$ $\mathrm{km\,s^{-1}Mpc^{-1}}$ (Riess et al. (2022)).

Complementary to candles are *standard rulers*, physical scales from early-universe physics. The most important are Baryon Acoustic Oscillations (BAO), coherent density fluctuations visible in the large-scale galaxy distribution, and the Cosmic Microwave Background (CMB), the relic radiation from recombination. Within ΛCDM model, these probes give consistent results, with CMB reporting $H_0 = 67.4 \pm 0.5$ $\mathrm{km\,s^{-1}}Mpc^{-1}$ (Aghanim et al. (2020)) and BAO producing similar values around 67-68 $\mathrm{km\,s^{-1}}Mpc^{-1}$ (Schöneberg et al. (2022)).

The discrepancy between early-universe (CMB, BAO) and late-universe probes has now reached the level of $\sim 5\sigma$. While it could in principle reflect statistical fluctuations or systematics uncertainties arising from different methods, independent probes consistently converge to lower values in the early universe and higher ones locally. This clear separation between the two families of measurements makes it unlikely that the tension arises from a single methodological bias, and instead suggests the possibility of new physics beyond ΛCDM (Verde, Treu and Riess (2019); Abdalla et al. (2022)). In this context, GW observations provide a crucial, entirely independent way to measure cosmological distances (see Sec. 1.3).

1.2 Fundamentals of Gravitational Waves

In Einstein's theory of GR, accelerated masses with a time-varying quadrupole moment generate ripples in the curvature of spacetime, known as Gravitational Waves. These waves propagate at the speed of light and carry energy, angular momentum, and information about their sources. In the weak-field, far-zone limit, the metric perturbation can be written as:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1 \tag{1.19}$$

The perturbation $h_{\mu\nu}$ satisfies the linearized Einstein equations. By imposing the Lorentz gauge condition

$$\partial^\mu \bar{h}_{\mu\nu} = 0, \qquad \bar{h}_{\mu\nu} = h_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}h, \tag{1.20}$$

the equations reduce to simple wave equations in vacuum,

$$\Box \bar{h}_{\mu\nu} = 0, \tag{1.21}$$

whose solutions are plane waves traveling at the speed of light.

An additional gauge freedom can be exploited by performing the transformation

$$h'_{\mu\nu} = h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu, \qquad \Box \xi_\mu = 0, \tag{1.22}$$

which allows us to choose the so-called *transverse-traceless (TT) gauge*. In this gauge, non-physical degrees of freedom are eliminated and only two independent components remain, corresponding to the two physical GW polarizations.

For a wave propagating along the z-axis, the metric perturbation takes the form

$$A^{TT}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{1.23}$$

$$A^{TT}_{\mu\nu} = h_+ e^{\mu\nu}_+ + h_\times e^{\mu\nu}_\times \tag{1.24}$$

with $e^{\mu\nu}_+$ and $e^{\mu\nu}_\times$ defined as:

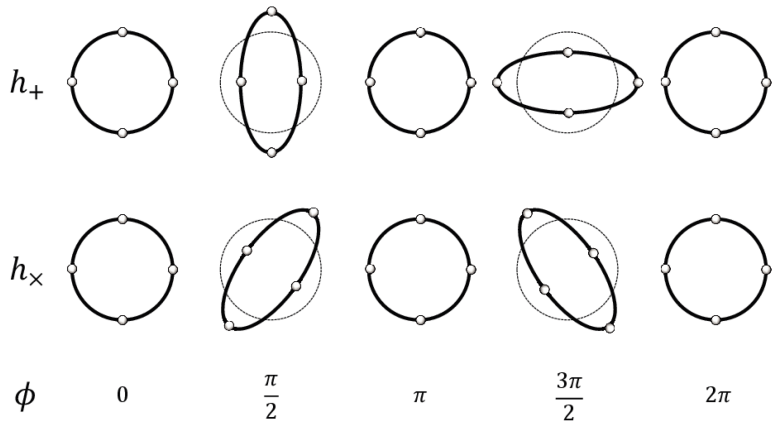


Figure 1.1: Deformation of a ring of test masses as effect of the h_+ and $h_\times$ polarization.(Image Credits: Antelis and Moreno (2017))

$$e_+^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad e_\times^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.25}$$

Here h_+ and $h_\times$ are the *plus* and *cross* polarizations. As the name suggests, the perturbation is transverse (no components along the propagation direction) and traceless ($h^\mu{}_\mu = 0$). These two polarizations represent the physical degrees of freedom of gravitational radiation. When passing through a ring of free test particles, they produce characteristic quadrupolar distortions: the h_+ polarization stretches space in one direction while compressing it in the orthogonal one, whereas $h_\times$ produces a similar effect rotated by 45° (Fig. 1.1). This provides the theoretical foundation for the interferometric detection of GWs.

For a binary system with component masses m_1 and m_2 in a quasi-circular orbit, the two polarizations observed at distance r are

$$h_+(t) = \frac{4}{r}\left(\frac{G\mathcal{M}}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm GW}}{c}\right)^{2/3}\frac{1+\cos^2\iota}{2}\cos(2\pi f_{\rm GW}t)\,, \tag{1.26a}$$

$$h_\times(t) = \frac{4}{r}\left(\frac{G\mathcal{M}}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm GW}}{c}\right)^{2/3}\cos\iota\,\sin(2\pi f_{\rm GW}t)\,, \tag{1.26b}$$

where ι is the inclination of the orbital plane relative to the line of sight, and the GW frequency is related to the orbital frequency by $f_{\mathrm{GW}} = 2f_{\mathrm{orb}}$. The *chirp mass* is defined as

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}, \tag{1.27}$$

and governs the amplitude and phase evolution of the signal.

Since the orbit shrinks due to GW emission, both the amplitude and frequency increase with time. Introducing the phase

$$\Phi(t) = 2\pi \int_{t_0}^{t} f_{\mathrm{GW}}(t')\, dt', \tag{1.28}$$

the polarizations can be rewritten as

$$h_+(t) = \frac{4}{r}\left(\frac{G\mathcal{M}}{c^2}\right)^{5/3}\left(\frac{\pi f_{\mathrm{GW}}}{c}\right)^{2/3}\frac{1+\cos^2\iota}{2}\cos\Phi(t), \tag{1.29a}$$

$$h_\times(t) = \frac{4}{r}\left(\frac{G\mathcal{M}}{c^2}\right)^{5/3}\left(\frac{\pi f_{\mathrm{GW}}}{c}\right)^{2/3}\cos\iota\,\sin\Phi(t). \tag{1.29b}$$

Close to merger, it is convenient to express the waveform in terms of the time to coalescence $\tau = t_{\mathrm{coal}} - t$, which makes explicit that both frequency and amplitude increase as $\tau \to 0$.

1.2.1 Sources of GW

GW can be generated by a wide range of astrophysical and cosmological processes. Depending on their morphology in the detector, signals are usually classified into four categories (Maggiore (2008); Tamanini et al. (to be published)):

- **Burst signals** are short-lived transients produced by violent, poorly modeled phenomena such as core-collapse supernovae or non-linear instabilities in neutron stars. Their unpredictable waveform makes them difficult to detect and identify, but they remain a promising probe of extreme astrophysics.
- **Continuous waves (CWs)** are nearly monochromatic and long-lasting signals. They are expected mainly from rapidly rotating, non-axisymmetric neutron stars with small surface deformations, or from compact binaries in very wide orbits. Their phase stability requires long integration times to reach detectable signal-to-noise ratios.

- **Stochastic backgrounds** consist of an incoherent superposition of many unresolved sources or random processes. An astrophysical background may arise from compact binaries too faint or too numerous to be distinguished individually, such as the Galactic population of double white dwarfs (important for the low-frequency band of LISA) or supermassive black-hole binaries driving a nanohertz background, recently hinted at by Pulsar Timing Arrays (PTA). Cosmological stochastic backgrounds could originate in the early Universe, for example from first-order phase transitions, cosmic strings, or inflationary quantum fluctuations.
- **Inspiral signals**, or CBC, are transient chirping signals produced during the final inspiral, merger, and ringdown of compact objects. They are the primary sources for current ground-based interferometers, operating in the ∼10–1000 Hz band, and so far the only class of sources directly detected. Depending on the nature of the components, CBCs are divided into:
  - *Binary Neutron Star (BNS)* mergers, such as GW170817, which produce both gravitational and electromagnetic counterparts (short GRBs, kilonovae, afterglows) and are therefore of particular importance for multi-messenger astronomy and cosmology.
  - *Neutron Star-Black Hole (NSBH)* coalescences, potentially accompanied by electromagnetic emission.
  - *Binary Black Holes (BBH)* mergers, which dominate the population of observed GW events. They do not produce electromagnetic signals, but provide crucial information on the mass and spin distribution of the black holes population.

1.2.2 GW detection

Ground-based interferometric GW detectors share the same basic layout: a power–recycled Michelson interferometer with Fabry–Perot arm cavities. A stabilized laser is split at a beam splitter and sent along two orthogonal, kilometer-scale arms, at the ends of which suspended mirrors act as test masses. The beams are reflected back and recombined at the splitter, with the interference pattern read out at the output photodetector (Maggiore (2008)). In the absence of a signal the output is kept at destructive interference, while the passage of a GW produces a differential change in the optical path ($\delta L \simeq hL$), leading to a measurable power variation. For a typical strain $h \sim 10^{-21}$ and detector's arms lenght $L \sim 4$ km, this corresponds to $\delta L \sim 4 \times 10^{-18}$ m.

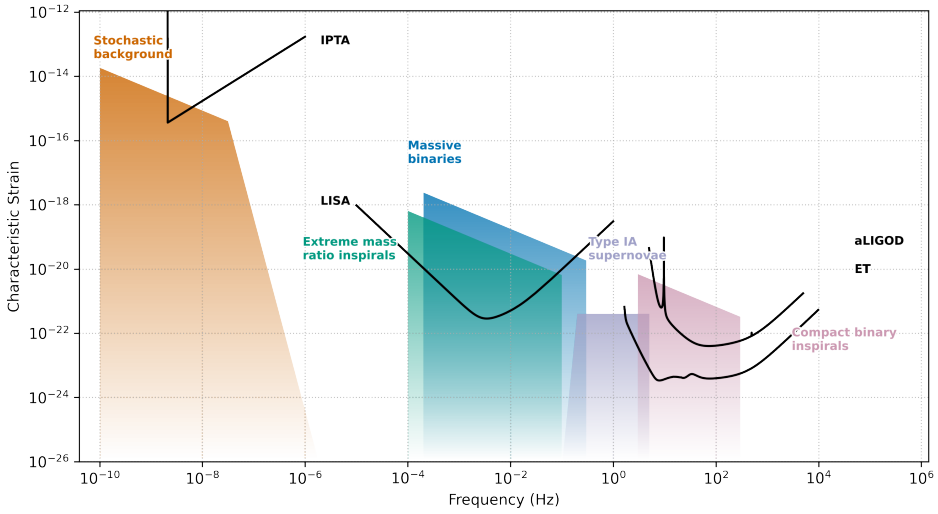


Figure 1.2: Characteristic strain as a function of frequency for selected GW detectors and source types. (Credits: Moore, Cole and Berry (2015))

To enhance their response, interferometers employ Fabry–Perot arm cavities, which increase the effective optical path by multiple reflections, and power recycling, which boosts the circulating laser power and reduces photon shot noise. Modern instruments also feature elaborate seismic isolation and operate in high vacuum to suppress environmental disturbances.

The sensitivity of detectors is limited by various *noise sources*, each dominant in a different frequency band. At low frequency ($\lesssim 10$ Hz) *displacement noise* from seismic motion, Newtonian gravity gradients, and suspension thermal fluctuations is the main limitation. Those produce a physical motion of the detector's components, resulting as a false GW strain. At higher frequencies, sensitivity is limited by *quantum optical noise*. Photon *shot noise*, due to random photon arrivals, dominates above a few hundred hertz, while *radiation-pressure noise*, caused by power fluctuations pushing on the mirrors, is most relevant at tens of hertz. These two effects define the *standard quantum limit*, and are mitigated with squeezed vacuum injection, power and signal recycling, and heavier test masses. Technical noise from the laser, electronics, or scattered light is also present but kept below the fundamental noise terms. As a result, current detectors achieve broadband strain sensitivities of order 10^{-23} $\mathrm{Hz}^{-1/2}$, with optimal performance between ~ 30 Hz and a few kHz, where stellar-mass compact binaries emitt most strongly.

The current global network of ground-based interferometric detectors is composed of the two Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors in the United States (Hanford and Livingston), Virgo in Italy, and Kamioka Gravitational Wave Detector (KAGRA) in Japan; a fifth site in India (LIGO–India) is under construction and will join in the coming years.

Since 2015 the LIGO-Virgo-KAGRA (LVK) collaboration has organized its observing campaigns in runs of increasing sensitivity and duration. The first direct detection, GW150914, occurred during O1 (2015–2016), marking the discovery of binary black hole mergers. In O2 (2016–2017) Virgo joined the network, and the joint observation of GW170817, a binary neutron star inspiral with rich electromagnetic follow-up, opened the era of multi-messenger astronomy. The subsequent O3 run (2019–2020) greatly expanded the catalog, with nearly ninety confirmed events including the first neutron star–black hole mergers and unusual massive BBH systems such as GW190521. The ongoing O4 run (2023–2025) operates with improved sensitivities and a growing number of candidate events, with Virgo and KAGRA intermittently joining the network.

The fifth observing run (O5) is planned to start around the late 2027 and will be longer and more sensitive. With upgraded instruments (LIGO A+, Advanced Virgo+, and KAGRA improvements) the network will extend its reach for compact binary mergers, increasing both the number of detections and the accuracy of source localization. In this phase the construction of LIGO–India will add a new site to the network, further enhancing sky coverage and opening the way for even more precise multi-messenger observations.

Looking ahead, third-generation ground detectors aim for about one order of magnitude improvement in broadband strain sensitivity and to access the frequency region below 10Hz. The core strategy is to build underground (as planned for the Einstein Telescope (ET)) which strongly reduces micro-seismic, which mainly propagate on the surface, and external disturbances (wind, temperature, scattered light). Underground siting also enables stable cryogenic setting, allowing for a mitigation of thermal noise in the $\sim$10–100 Hz band. In addition, improved optics and readout (optimized recycling and frequency-dependent squeezing with heavier test masses) keep quantum noise under control while maintaining stable interferometer operation (Punturo et al. (2010)).

The overall evolution of the detector network and the corresponding growth of the GW

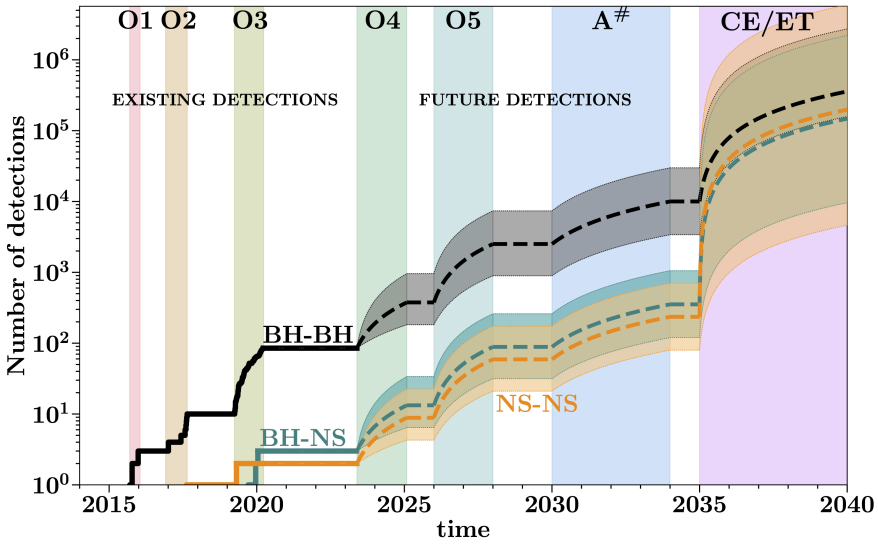


Figure 1.3: Number of detected mergers expected from binary black holes (*black*), black hole-neutron stars (*teal*), and binary neutron stars (*orange*) systems for different Observing runs. (Credits: Broekgaarden, Banagiri and Payne (2024))

catalogs expected from current and future observing runs, including third–generation observatories such as ET and CE, is summarized in Fig. 1.3.

1.3 Gravitational Waves as Cosmological Tools

The possibility of using GW as distance indicators was first proposed by Schutz (1986), who realized that Compact Binary Coalescence provide a direct measurement of their luminosity distance. Unlike electromagnetic probes, which rely on complex astrophysical modeling and calibration steps, the amplitude of GW signal directly encode d_L. This property makes GW sources *standard sirens*.

In the cosmological context, this property is of particular relevance to the current discrepancy between early and late Universe determinations of H_0, known as Hubble Tension (see Sec. 1.1.6). Gravitational waves provide an entirely independent probe: by combining the luminosity distance extracted from a GW signal with an external measurement of the source redshift, it is possible to directly test the distance–redshift relation and constrain cosmological parameters. In this way, GW standard sirens represent a powerful probe.

For Compact Binary Coalescence at cosmological distances, the strain (Eq. (1.29)) is modified by the expansion of the Universe: the emitted frequency is redshifted, the source-frame chirp mass is replaced by its redshifted value, and the proper distance r is replaced by the luminosity distance d_L:

$$f_{\rm obs} = \frac{f_{\rm src}}{1+z}, \qquad \mathcal{M}_{\rm obs} = (1+z)\,\mathcal{M}_{\rm src} \tag{1.30}$$

With these modifications, the observed polarizations become

$$h_+(t) = \frac{4}{d_L}\left(\frac{G\mathcal{M}_{\rm obs}}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm obs}(t)}{c}\right)^{2/3}\frac{1+\cos^2\iota}{2}\cos\Phi(t), \tag{1.31}$$

$$h_\times(t) = \frac{4}{d_L}\left(\frac{G\mathcal{M}_{\rm obs}}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm obs}(t)}{c}\right)^{2/3}\cos\iota\,\sin\Phi(t). \tag{1.32}$$

The interferometric detectors measure a linear combination of h_+ and $h_\times$, weighted by their antenna response functions F_+ and $F_\times$:

$$h(t) = h_+(t)F_+(\theta,\phi) \;+\; h_\times(t)F_\times(\theta,\phi) \tag{1.33}$$

where θ and ϕ are the angles defining the direction of propagation of the GW. From the observed waveform, different pieces of information can be extracted. The amplitudes of h_+ and $h_\times$ encode the inclination angle ι of the binary. The frequency evolution of the signal determines the redshifted chirp mass $\mathcal{M}_{\rm obs}$. Finally, the overall amplitude of the strain provides a direct measurement of the luminosity distance d_L. Putting these together, one obtains a self-consistent measurement of d_L. If an independent determination of the source redshift z is available, this yields a point on the distance–redshift relation, making compact binaries *standard sirens*.

1.3.1 Standard Sirens

One of the key challenges in gravitational wave cosmology is the degeneracy between mass and redshift. Without an independent redshift measurement, it is not possible to recover the source-frame chirp mass, and thus the true component masses that generated the GW signal. Separate approaches have been developed for this purpose, leading to the distinction between bright, dark, and spectral sirens.

Bright Sirens: In this case, the redshift information is provided by the direct obser-

vation of an electromagnetic counterpart to the GW event. Radiation produced by the merger can be observed across the electromagnetic spectrum, enabling a spectroscopic measurement of the host galaxy redshift from spectral line shifts. This method is feasible when at least one of the binary components is a neutron star, as in BNS or NSBH systems, where electromagnetic emission is expected. For BBH mergers, no light escapes and thus no direct counterpart can be observed. To date, only one confirmed bright siren has been detected: GW170817.

Dark Sirens: When no electromagnetic counterpart is observed, redshift information can be obtained from statistical association of the GW event to its host galaxy. The idea is to combine the GW localization volume with galaxy catalogs: by overlapping the 3D sky localization region of the event with cataloged galaxies, one can infer a probabilistic association between the GW source and potential hosts galaxies enclosed in the localization volume. This approach has been widely applied to BBH systems, where no EM signal is expected.

Spectral Sirens: This method relies on the statistical properties of the GW source population itself. For example, compact objects are expected to follow specific mass distributions. Since the observed chirp mass is redshifted, $\mathcal{M}_{\rm obs} = (1+z)\,\mathcal{M}_{\rm src}$, population features in the intrinsic mass distribution appear shifted in the observed distribution. By comparing the observed distribution of redshifted masses with astrophysical models, it is possible to statistically infer the redshift of events, and hence extract cosmological information even without direct host identification.

In this context, the present work is focuses to the use of dark sirens as cosmological probes. The analysis is performed within the hierarchical Bayesian inference formalism, which provides a consistent way to connect event-level parameters measured from GW signals with population properties and cosmological parameters. Using simulated catalogs, the work evaluates how sample selection influences the precision of cosmological inference and examines the impact of Signal to Noise Ratio (SNR) cuts on GW parameters. A central question is whether the cosmological information is evenly distributed across the full set of detected events, or whether specific subsets of events carry most of the constraining power. This becomes critical in view of the expected growth in the number of detections. So far, with the limited number of detections in O3 and O4, all events above threshold could be analyzed. This will no longer be possible with O5

and, especially, with third generation detectors, like the ET, which will deliver up to 10^5 events (Branchesi et al. (2023)). In such a large data volume regime, it becomes essential to determine which events carry the greatest cosmological and astrophysical information. By addressing this, the thesis contributes to defining strategies that will allow gravitational-wave cosmology to fully exploit their crucial role in cosmology.

2

Statistical Framework

Extracting cosmological information from GW observations requires a statistical framework that links the properties of individual events to the characteristics of the overall population. This is achieved using *Hierarchical Bayesian Inference*, a powerful method that enables the joint estimation of both event-level parameters and population-level hyperparameters describing astrophysical and cosmological distributions. The framework naturally incorporates measurement uncertainties, selection effects, and observational biases.

In this work, we perform hierarchical inference using `CHIMERA`, a dedicated pipeline that combines GW data with galaxy catalogs, enabling population and cosmological studies in a fully self-consistent way.

2.1 Hierarchical Bayesian Inference

Bayesian inference has become a fundamental tool in modern astrophysics, particularly in GW astronomy (Thrane and Talbot (2019)). Its power lies in the ability to incorporate prior knowledge, account for observational uncertainties, and update model probabilities as new data become available.

In the context of dark sirens, Bayesian methods provide a rigorous framework for inferring both the physical properties of individual sources and the underlying cosmological parameters. This is achieved by combining GW data with external astrophysical informations, such as galaxy catalogs. The approach becomes especially powerful when multiple events are analyzed jointly, enabling population-level inference through a hierarchical modeling structure.

This hierarchical approach is particularly powerful when dealing with complex models involving multiple layers of parameters. This is the case of GW cosmology, where we are not only interested in the properties of a single source, but also in the population properties that describe the entire ensemble of detected events. In this context, we distinguish between two levels of parameters:

- **Event-level parameters $\boldsymbol{\theta}_i$**: these describe the astrophysical and geometrical properties of the i-th source. For each detected event they are not observed, but inferred probabilistically from the noisy detector data (see Sec. 2.1.1).
- **Hyperparameters $\boldsymbol{\lambda}$**: These describe the statistical properties of the entire source population and are organized into three groups: $\boldsymbol{\lambda}_c$, the *cosmological parameters*, which enter through the distance–redshift relation; $\boldsymbol{\lambda}_m$, the *mass distribution parameters*, which control the intrinsic binary mass spectrum; and $\boldsymbol{\lambda}_z$, the *redshift evolution parameters*, which model the merger rate density as a function of cosmic time. The entire set of hyperparameters is summarized in Tab. 2.1, while a detailed description of population models for mass and merger rate distributions and related hyperparameters can be found in Sec. 2.2.

It is important to note that the event-level parameters $\boldsymbol{\theta}_i$ are not directly accessible through observations. Gravitational wave detectors instead yield data products $\boldsymbol{d}_i$, such as strain time series, from which information on $\boldsymbol{\theta}_i$ must be inferred. The statistical connection between the hyperparameters $\boldsymbol{\lambda}$ and the measured data $\boldsymbol{d}_i$ is therefore mediated by the parameters $\boldsymbol{\theta}_i$, following the hierarchical structure

$$\boldsymbol{d}_i \longrightarrow \boldsymbol{\theta}_i \longrightarrow \boldsymbol{\lambda}$$

This structure is known as *hierarchical* because it connects observed data ($\boldsymbol{d}_i$) to population-level parameters ($\boldsymbol{\lambda}$) through intermediate variables ($\boldsymbol{\theta}_i$).

	Parameter	*Description*	*Fiducial value*
$\boldsymbol{\lambda}_c$	H_0	Hubble constant $[Mpc\ s^{-1}\ km^{-1}]$	70.0
	Ω_m	Matter Energy density	0.25
$\boldsymbol{\lambda}_z$	γ	Slope at $z < z_p$	2.7
	z_p	z at the peak	2
	κ	Slope at $z > z_p$	3
$\boldsymbol{\lambda}_m$	α	Spectral index for the primary power law	3.4
	β	Spectral index for the secondary power law	1.1
	δ_m	Smoothing parameter $[M_\odot]$	4.8
	m_{low}	Minimum mass value $[M_\odot]$	5.1
	m_{high}	Maximum mass value $[M_\odot]$	87.0
	μ_g	Mean of the Gaussian peak $[M_\odot]$	34.0
	σ_g	Standard deviation of the Gaussian peak $[M_\odot]$	3.6
	λ_g	Strength of the Gaussian peak	0.039

Table 2.1: Hyperparameters adopted, grouped according to their role in describing the *cosmology* ($\boldsymbol{\lambda_c}$), the *merger–rate redshift evolution* ($\boldsymbol{\lambda_z}$), and the *mass distribution* ($\boldsymbol{\lambda_m}$). The fiducial value for each parameter is reported as reference.

2.1.1 Parameter Estimation

The data collected by a GW detector are a time series of strain $s(t)$ produced by a combination of the true GW signal $h(t)$ and detector noise:

$$s(t) = h(t) + n(t). \tag{2.1}$$

The role of parameter estimation is to disentangle the two and determine which sets of parameters $\boldsymbol{\theta}$ are most consistent with the observed strain.

The parameters can be grouped into two categories: *Intrinsic parameters* describe the source in its own rest frame, such as the component masses (m_1, m_2), spins $(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2)$, and, for neutron stars tidal deformabilities (Λ_1, Λ_2). They determine the dynamics of the binary and thus the detailed evolution of the gravitational waveform. *Extrinsic parameters* specify how the source is seen by the detector, including the luminosity distance d_L, sky position (RA, dec), orbital inclination ι, polarization angle ψ and coalescing phase and time ϕ_c and t_c. They affect the amplitude, orientation, and modulation of the observed signal.

From the detector point of view, none of these parameters are observed directly. Para-

meter estimation proceeds by comparing waveform templates with the observed strain $s(t)$, and by inferring which values of $\boldsymbol{\theta}$ allow the model to reproduce the data most consistently. In practice, one considers a family of possible waveforms templates, denoted as $h(t;\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \theta_1, \ldots, \theta_N$ represents the collection of both intrinsic and extrinsic parameters to be inferred, and evaluates how well each template matches the data. Combining this likelihood with prior information yields a *posterior distribution* for each parameter: a probability distribution describing which values of $\boldsymbol{\theta}$ are supported by the data (Christensen and Meyer (2022); Maggiore (2008)).

The result of parameter estimation for one event is not a single best-fit value, but a set of correlated probability distributions for all parameters. These posteriors encapsulate the information that will be combined across many events in hierarchical Bayesian inference, allowing us to connect the noisy strain data to population properties and cosmological parameters.

2.1.2 Bayes Theorem

Within this framework, *Bayes Theorem* relates the posterior distribution of the hyper-parameters to the likelihood of the data and prior information. Considering a population of $N_{\rm ev}$ GW events, each characterized by source-frame parameters $\boldsymbol{\theta}_i$ drawn from a population distribution $p_{\rm pop}(\boldsymbol{\theta}|\boldsymbol{\lambda})$, the posterior on $\boldsymbol{\lambda}$ reads:

$$p(\boldsymbol{\lambda} \,|\, \{\boldsymbol{d}_i\}) = \frac{\mathcal{L}(\{\boldsymbol{d}_i\} \,|\, \boldsymbol{\lambda})\, \pi(\boldsymbol{\lambda})}{p(\{\boldsymbol{d}_i\})}\,, \tag{2.2}$$

where $\pi(\boldsymbol{\lambda})$ is the *prior* distribution reflecting our knowledge or assumptions about $\boldsymbol{\lambda}$ before observing the data. $\mathcal{L}(\{\boldsymbol{d}_i\} \,|\, \boldsymbol{\lambda})$ is the *likelihood*, which quantifies the probability of the data given the model parameters. The denominator $p(\{\boldsymbol{d}_i\})$, known as the *evidence*, acts as a normalization factor obtained by integrating the likelihood over all possible values of $\boldsymbol{\theta}$.
In practice, the posterior is the main quantity of interest, as it encapsulates what can be inferred about the parameters after incorporating the data. The likelihood measures how well a given model explains the observations, while the prior allows us to include external information or theoretical expectations.

2.1.3 Likelihood and Selection Effects

The likelihood function is the core element of Bayesian inference, quantifying how probable it is to observe the data given a specific set of parameters.

Considering our set of $N_{\rm ev}$ observed events $\{\boldsymbol{d}_i\}$, for each event, we can define a likelihood that depends on the event-level parameters $\boldsymbol{\theta}_i$ and the population-level hyperparameters $\boldsymbol{\lambda}$. Assuming statistical independence between events, the hyper-likelihood can be written as the product of the likelihoods for each individual event:

$$\mathcal{L}(\{\mathbf{d}_i\}|\boldsymbol{\lambda}) = \prod_{i=1}^{N_{\rm ev}} \int p(\mathbf{d}_i|\boldsymbol{\theta}_i)\, p_{\rm pop}(\boldsymbol{\theta}_i|\boldsymbol{\lambda})\, d\boldsymbol{\theta}_i \tag{2.3}$$

Here, $p(\boldsymbol{d}_i|\boldsymbol{\theta}_i)$ is the event-level likelihood, while $p_{\rm pop}(\boldsymbol{\theta}_i|\boldsymbol{\lambda})$ is the population model defined by the hyperparameters $\boldsymbol{\lambda}$ (see Sec. 2.2).

However, not all sources in the astrophysical population are equally likely to be detected. The detectability of an event depends on source properties such as luminosity distance, binary orientation, and sky position, as well as on the sensitivity of the detector network. For binary black holes, the sensitive volume of a detector is strongly influenced by the component masses (Talbot and Thrane (2018)). More massive binaries generate higher-amplitude signals and are thus detectable at larger distances, but they also merge at lower frequencies and spend less time in the detector band. Moreover, cosmological redshift shifts the signal to lower frequencies and makes distant systems appear more massive in the detector frame.

This interplay between source properties and detector sensitivity results in an observed population that is biased with respect to the true astrophysical distribution. Such effects can be described through the *detection probability*, $P_{\rm det}(\boldsymbol{\theta})$, which gives the probability that a source with parameters $\boldsymbol{\theta}$ is observed:

$$P_{det}(\boldsymbol{\theta}) = \int_{\boldsymbol{d}\in\rm det} p(\mathbf{d}'|\boldsymbol{\theta})\; d\mathbf{d}'. \tag{2.4}$$

Taking this into account, the hyper-likelihood takes the form (Mandel, Farr and

Gair (2019); Vitale et al. (2021)):

$$\mathcal{L}(\{\mathbf{d}_i\}|\boldsymbol{\lambda}) = \prod_{i=1}^{N_{\rm ev}} \frac{\int p(\mathbf{d}_i|\boldsymbol{\theta}')\, p_{\rm pop}(\boldsymbol{\theta}'|\boldsymbol{\lambda}_i)}{\int_{d\in det} d\mathbf{d}' \int p(\mathbf{d}'|\boldsymbol{\theta}') p_{pop}(\boldsymbol{\theta}'|\boldsymbol{\lambda}) d\boldsymbol{\theta}'} \tag{2.5}$$

It is useful to factor out the detection probability from the likelihood by introducing the *selection function*:

$$\xi(\boldsymbol{\lambda}) = \int P_{\rm det}(\boldsymbol{\theta}')\, p_{\rm pop}(\boldsymbol{\theta}'|\boldsymbol{\lambda})\, d\boldsymbol{\theta}', \tag{2.6}$$

which represents the fraction of sources from the underlying population that would be detected for a given $\boldsymbol{\lambda}$. The corrected hyper-likelihood, properly accounting for selection effects, is then:

$$\mathcal{L}(\{\mathbf{d}_i\}|\boldsymbol{\lambda}) \propto \frac{1}{\xi(\boldsymbol{\lambda})^{N_{\rm ev}}} \prod_{i=1}^{N_{\rm ev}} \int p(\mathbf{d}_i|\boldsymbol{\theta}_i)\, p_{\rm pop}(\boldsymbol{\theta}_i|\boldsymbol{\lambda})\, d\boldsymbol{\theta}_i. \tag{2.7}$$

This normalization ensures that the inferred population parameters are not biased by the preferential detection of intrinsically louder or closer events. Neglecting the selection function $\xi(\boldsymbol{\lambda})$ would lead to biased inferences on the underlying population and cosmological parameters, as more easily detectable events would be overrepresented in the posterior.

If, in addition, we wish to include information on the total number of detections, $N_{\rm ev}$ can be modeled as a Poisson process with expected value $N_{\rm exp}(\lambda) = N_{\rm ev}\xi(\lambda)$, where N_{exp} is the number of expected events during the observation period, including also undetectable signals. In this case, the likelihood accounts for the additional Poisson term:

$$e^{-N_{exp}}(N_{exp})^{N_{\rm ev}} \tag{2.8}$$

Without the Poisson term the likelihood is *scale–free*: it allows inference on the shape of the population but not on the number of events.

2.1.4 Full Posterior

The full posterior over the hyperparameters, given the data from all events $\{\mathbf{d}_i\}$, is proportional to the product of the likelihood and the prior:

$$p(\boldsymbol{\lambda} \mid \{\mathbf{d}_i\}) \propto \mathcal{L}(\{\mathbf{d}_i\} \mid \boldsymbol{\lambda}) \times \pi(\boldsymbol{\lambda}) \tag{2.9}$$

Including Eq. (2.7) and 2.8, the full expression is written as:

$$p(\boldsymbol{\lambda}|\{\boldsymbol{d}_i\}) \propto e^{-N_{exp}}(N_{exp})^{N_{\rm ev}} \frac{\pi(\boldsymbol{\lambda})}{\xi(\boldsymbol{\lambda})^{N_{\rm ev}}} \prod_{i=1}^{N_{\rm ev}} \int p(\boldsymbol{d}_i|\boldsymbol{\theta}_i, \boldsymbol{\lambda})\, p_{\rm pop}(\boldsymbol{\theta}_i|\boldsymbol{\lambda})\, d\boldsymbol{\theta}_i \tag{2.10}$$

which includes, the population model $p_{\rm pop}(\theta_i|\boldsymbol{\lambda})$, which defines how the event-level parameters are distributed across the population; the selection function $\xi(\boldsymbol{\lambda})$, which accounts for detection biases and is essential to ensure unbiased inference. $N_{\rm exp}(\boldsymbol{\lambda})$ is the expected number of detections under $\boldsymbol{\lambda}$, and $\pi(\boldsymbol{\lambda})$ is the prior over the hyperparameters

To perform this analysis, we use the `CHIMERA` code, which is designed to handle large populations of GW events and their connection to galaxy catalogs. *Combined Hierarchical Inference Model for Electromagnetic and gRavitational Wave Analysis* `CHIMERA` computes the full posterior distribution over cosmological and population parameters by sampling the GW event posteriors, modeling the redshift prior informed by catalogs, incorporating the selection function, and efficiently evaluating the joint likelihood. A more detailed explanation can be found in Sec. 2.4.

2.2 Population Models

The population probability distribution $p_{\rm pop}(\boldsymbol{\theta}|\boldsymbol{\lambda})$ describes how the event-level parameters $\boldsymbol{\theta}$ of gravitational wave sources are distributed within the astrophysical population described by a set of hyperparameters $\boldsymbol{\lambda}$. We adopt a factorized form:

$$p_{\rm pop}(\boldsymbol{\theta}|\boldsymbol{\lambda}) = p(m_1, m_2|\boldsymbol{\lambda}_m)\, p(z, \hat{\Omega}|\boldsymbol{\lambda}_z, \boldsymbol{\lambda}_c), \tag{2.11}$$

where $p(m_1, m_2|\boldsymbol{\lambda}_m)$ models the *mass distribution* of the binary components, and $p(z, \hat{\Omega}|\boldsymbol{\lambda}_z, \boldsymbol{\lambda}_c)$ represents the *redshift and localization distribution*, which incorporates cosmological parameters $\boldsymbol{\lambda}_c$ and redshift-evolution parameters $\boldsymbol{\lambda}_z$.

This factorization assumes that the binary mass distribution is independent of redshift, which is a common approximation known as a *stationary* mass function. Although several studies have reported possible mass–redshift trends suggesting an evolution toward higher primary masses at larger redshifts (Gennari et al. (2025)), these effects are expected to become significant at redshifts higher than those reached by our current detector network. Within the redshift range probed by our catalogs, the stationary

approximation is therefore correct.

2.2.1 Mass distribution

The mass term can be separated into the distribution of the *primary* and *secondary* component masses:

$$p(m_1, m_2|\boldsymbol{\lambda}_m) = p(m_1|\boldsymbol{\lambda}_m)\ p(m_2|m_1, \boldsymbol{\lambda}_m), \tag{2.12}$$

with the condition that $m_1 \geq m_2$. This decomposition allows for flexible modeling of the primary mass spectrum and of the mass ratio m_1/m_2 distribution.

A variety of models have been proposed for the primary mass distribution $p(m_1|\boldsymbol{\lambda}_m)$, motivated by both population-synthesis predictions and observational constraints (Talbot and Thrane (2018); Roy, Son and Farr (2025)). The most widely used is the *Power Law + Peak (PLP)* model, described as

$$p(m_1|\alpha, \delta_m, m_{low}, m_{high}, \mu_g, \sigma_g, \lambda_g) \propto [(1-\lambda_g)\mathcal{P}(m_1) + \lambda_g \mathcal{G}(m_1)]\mathcal{S}(m_1) \tag{2.13}$$

which combines:

- a *truncated power-law* $\mathcal{P}(m_1; \alpha)$, with slope α and bounds $(m_{\rm low}, m_{\rm high})$, describing the bulk of the distribution;
- a *Gaussian peak* $\mathcal{G}(m_1; \mu_g, \sigma_g)$ representing an excess of sources at $m_1 \sim 30{-}40\,M_\odot$, potentially linked to pair-instability supernova physics (Talbot and Thrane (2018));
- a *smoothing function* $\mathcal{S}(m_1; \delta_m)$ to avoid a sharp cut-off at the lower edge of the mass distribution.

The relative contribution of the power-law and Gaussian components is set by a weighting expressed by λ_g. In Fig. 2.1a is reported the $p(m_1)$ mass distribution, which follows the PLP.

The secondary mass distribution $p(m_2|m_1, \boldsymbol{\lambda}_m)$ follows a power law with spectral index β inside the mass range $[m_{low}, m_1]$.

For BNS systems, the current number of detections is too small to robustly constrain the mass function, and a uniform distribution in a plausible mass range is often assumed as a practical choice.

2.2.2 Merger Rate distribution

The term $p(z, \hat{\Omega}|\boldsymbol{\lambda}_z, \boldsymbol{\lambda}_c)$ in the population function describes the redshift and sky localization distribution of the sources. It can be factorized into two components:

$$p(z, \hat{\Omega}|\boldsymbol{\lambda}_z, \boldsymbol{\lambda}_c) = \frac{p_{gal}(z, \hat{\Omega}|\boldsymbol{\lambda}_c) p_{rate}(z|\boldsymbol{\lambda}_z)}{\int p_{gal}(z', \hat{\Omega}'|\boldsymbol{\lambda}_c) p_{rate}(z'|\boldsymbol{\lambda}_z) dz' d\hat{\Omega}'}. \tag{2.14}$$

Here, $p_{gal}(z, \hat{\Omega}|\boldsymbol{\lambda}_c)$ represents the probability of finding a host galaxy at a given redshift and sky location, given the cosmological model $\boldsymbol{\lambda}_c$, while $p_{rate}(z|\boldsymbol{\lambda}_z)$ models the intrinsic merger rate density as a function of redshift, parametrized by $\boldsymbol{\lambda}_z$.

The galaxy term p_{gal} encodes the distribution of potential host galaxies and describes the probability of having an host galaxy at a given redshift z and sky localization Ω. The merger rate term $p_{\rm rate}$ describes how the rate of compact binary coalescences evolves with cosmic time. In the source frame, this evolution is often expressed as $\psi(z|\boldsymbol{\lambda}_z)$, while in the detector frame the observed rate is reduced by a factor $(1+z)$ due to cosmological redshift:

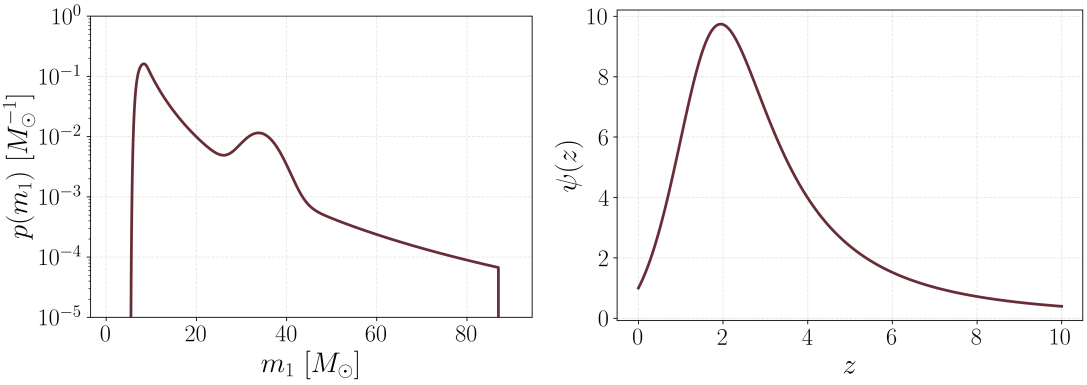
$$p_{rate}(z|\boldsymbol{\lambda}_z) \propto \frac{\psi(z|\boldsymbol{\lambda}_z)}{1+z}. \tag{2.15}$$

Several parametrizations can be adopted for $\psi(z)$, ranging from simple power-laws to more physically motivated models. In this work, we employ the *Madau–Dickinson* form (Madau and Dickinson (2014)), widely used to model star formation histories:

$$\psi(z|\boldsymbol{\lambda}_z) \propto \frac{(1+z)^{\gamma}}{1 + \left[\frac{1+z}{1+z_p}\right]^{\gamma+\kappa}}, \tag{2.16}$$

where γ is the slope which controls the rise of the rate at low redshift, κ sets slope of the decline at high redshift and z_p is the redshift at which the rate peaks.

This parametrization captures the general behavior expected for compact binary populations formed from stellar progenitors: an initial rise tracing the cosmic star formation rate, followed by a decline at high redshift. Fig. 2.1b shows the Madau–Dickinson merger rate model over the redshift range $0 \leq z \leq 10$.



(a) **(b)**

Figure 2.1: Assumed population models used in the analysis. *Left* (2.1a): Power-Law plus Peak mass function for the primary mass. *Right* (2.1b): Madau–Dickinson CBC rate model shown as a function of redshift in the range 0 to 10.

2.3 Catalog Completeness

In dark sirens cosmology, the redshift prior used for the association between GW events and potential host galaxies is derived from an external *galaxy catalog*. The underlying assumption is that compact binary mergers occur within galaxies, and that the distribution of these galaxies traces the distribution of possible GW hosts. In the idealised case of a *complete* galaxy catalog (one containing all galaxies that could plausibly host a GW event) the probability of finding a host with properties G (stellar mass, luminosity) at redshift z and sky position $\hat{\Omega}$ is given directly by the catalog:

$$p_{\rm gal}(z, \hat{\Omega}, G|\boldsymbol{\lambda}_c) = p_{\rm cat}(z, \hat{\Omega}, G|\boldsymbol{\lambda}_c). \tag{2.17}$$

This quantity can be computed as a weighted sum over all galaxies g in the catalog:

$$p_{\rm cat}(z, \hat{\Omega}, G|\boldsymbol{\lambda}_c) = \frac{\sum_g w_g\, p(z|\tilde{z}_g, \boldsymbol{\lambda}_c)\, \delta(\hat{\Omega} - \hat{\Omega}_g)}{\sum_g w_g}, \tag{2.18}$$

where the *weights*, w_g, encode the relative probability that galaxy g hosts the event, $\tilde{z}_g$ is its measured redshift, and δ is a Dirac delta function enforcing its sky localization. The term $p(z|\tilde{z}_g, \boldsymbol{\lambda}_c)$ represents the galaxy's redshift posterior, obtained from the measured value $\tilde{z}_g$ and its uncertainty $\sigma_{\tilde{z},g}$, combined with a prior $p_{\rm bkg}(z|\boldsymbol{\lambda}_c)$ describing the

background galaxy distribution. Assuming Gaussian measurement errors, we have:

$$p(z|\tilde{z}_g, \boldsymbol{\lambda}_c) = \frac{N(z; \tilde{z}_g, \sigma^2_{\tilde{z},g})\, p_{\rm bkg}(z|\boldsymbol{\lambda}_c)}{\int N(z'; \tilde{z}_g, \sigma^2_{\tilde{z},g})\, p_{\rm bkg}(z'|\boldsymbol{\lambda}_c)\, dz'}, \tag{2.19}$$

where $N(z; \mu, \sigma^2)$ is the Gaussian probability density function.

However, no realistic galaxy catalog is perfectly complete. Observational limitations, such as finite magnitude limits, incomplete sky coverage, and spectroscopic targeting biases, imply that some fraction of true host galaxies are missing from the catalog. This incompleteness grows with redshift due to the Malmquist bias, whereby intrinsically faint galaxies become undetectable. Ignoring this effect would bias cosmological inference, as the redshift prior would under-represent hosts at higher z.

We quantify *completeness* in a region S, defined by a range in redshift and sky position, for galaxies with properties G as:

$$P_{\rm comp}(S, G) = \frac{N_{\rm cat}(S, G)}{N_{\rm gal}(S, G)}, \tag{2.20}$$

where $N_{\rm cat}$ is the number of catalogued galaxies in S and $N_{\rm gal}$ the total number physically present. The latter can be modeled from a reference comoving number density $\bar{n}_{\rm gal}(G)$, assumed constant over S, such that:

$$P_{\rm comp}(S, G) = \frac{N_{\rm cat}(S, G)}{\bar{n}_{\rm gal}(G)\, V_c(S)}, \tag{2.21}$$

with $V_c(S)$ the comoving volume of S for the assumed cosmology.

When completeness is below unity, the galaxy probability is decomposed into a contribution from catalogued galaxies and a term p_{miss} describing the *missing* ones:

$$p_{\rm gal}(z, \hat{\Omega}, G|\boldsymbol{\lambda}_c) = f_R\, p_{\rm cat}(z, \hat{\Omega}, G|\boldsymbol{\lambda}_c) + (1 - f_R)\, p_{\rm miss}(z, \hat{\Omega}, G|\boldsymbol{\lambda}_c), \tag{2.22}$$

where

$$f_R = \int P_{\rm comp}(z, \hat{\Omega}, G)\, p_{\rm bkg}(z|\boldsymbol{\lambda}_c, G)\, dz \tag{2.23}$$

is the fraction of hosts expected to be present in the catalog, which depend on the completeness function P_{comp} and the background dirstibution p_{bkg}. f_R acts, in the

definition of p_{gal} as a weight for the two terms, p_{cat} and p_{miss}.

Several prescriptions exist for modeling $p_{\rm miss}$: in the *homogeneous completion* model, the missing galaxies are distributed uniformly within each region S, with number density $n_{\rm miss} = \bar{n}_{\rm gal}(G)\,[1 - P_{\rm comp}(S, G)]$. This yields an unbiased mean density but preserves the variance present in the catalog. In the *multiplicative completion* model, the missing galaxies trace the overdensity field, with an enhancement factor $b(S, G) = [1 - P_{\rm comp}(S, G)]/P_{\rm comp}(S, G)$, which increases the variance by $1/P^2_{\rm comp}$. In this work, following most current standard siren analyses, we adopt the homogeneous completion scheme, and we therefore define p_{miss} as:

$$p_{miss}(z, \hat{\Omega}, \mathcal{G}|\boldsymbol{\lambda}) = \frac{1 - P_{comp}(z, \hat{\Omega}, \mathcal{G})}{1 - f_R} p_{bkg}(z|\boldsymbol{\lambda}, \mathcal{G}). \tag{2.24}$$

The background distribution of galaxies which enter in Eqs.2.19, 2.23 and 2.24, is defined as follows:

$$p_{bkg}(z|\boldsymbol{\lambda}_c, \mathcal{G}) \propto \bar{n}_{gal}(z, \mathcal{G}) \frac{dV}{dz}(z|\boldsymbol{\lambda}_c) \tag{2.25}$$

where the term dV/dz include the cosmological dependence on H_0 and Ω_m, and $\bar{n}_{gal}(z, \mathcal{G})$ is the theoretical galaxy density distribution in redshift:

$$\bar{n}_{gal}(z, \mathcal{G}) = \int_{M_{min}(z)}^{M_{max}(z)} \Phi(M, z) w(M, z) dM. \tag{2.26}$$

Here, $\Phi(M, z)$ is the Shechter mass function, $M_{min}(z)$ and $M_{max}(z)$ are the lower and upper edges of the mass range for BBH hosts, and $w(M, z)$ is the weight associated to each galaxy which we take to be proportional to stellar mass. This choice assigns each galaxy a probability to host a BBH proportional to its stellar mass, and is motivated by theoretical studies indicating that, in the local Universe, BBH mergers preferentially occur in high mass and metal rich galaxies (Artale et al. (2020)).

Overall, completeness modeling enters directly into the construction of the redshift prior for GW–galaxy association:

$$p(z, \hat{\Omega}, G|\boldsymbol{\lambda}_c, \boldsymbol{\lambda}_z) \propto p_{\rm gal}(z, \hat{\Omega}, G|\boldsymbol{\lambda}_c)\, \frac{\psi(z|\boldsymbol{\lambda}_z)}{1 + z}, \tag{2.27}$$

where $\psi(z|\boldsymbol{\lambda}_z)$ is the source-frame merger rate evolution, and the factor $(1+z)^{-1}$ converts to the detector frame. By correctly accounting for catalog incompleteness, this prior

remains an unbiased tracer of the true host distribution, ensuring reliable inference on H_0 and other cosmological parameters.

2.4 CHIMERA

The *Combined Hierarchical Inference Model for Electromagnetic and gRavitational Wave Analysis* (`CHIMERA`) is a Python code that implements the hierarchical Bayesian formalism described in the previous sections for joint population and cosmological inference using GW events and galaxy catalogs (Borghi, Mancarella et al. (2024); Tagliazucchi et al. (2025); Borghi, Moresco et al. (2025)). The pipeline is fully built in `JAX` framework (Bradbury et al. (2018)), which enables accelerated array computations for high performance numerical computing. These feature is essential to evaluate high-dimensional hyperparameter spaces efficiently and to sustain the increased data volumes expected from next generation GW detectors and wide-field galaxy surveys.

Different pipelines have been released to perform astrophysical and cosmological inference with GW, like `MGCosmoPop` (Mancarella, Genoud-Prachex and Maggiore (2022)) and `GWPopulation` (Talbot, Smith et al. (2019)) especially designed for spectral sirens method, and `DarkSirensStat` (Finke et al. (2021)) and `cosmolisa` (Laghi et al. (2021)) for the dark sirens method. Recently, unified codes able to apply Bayesian framework to different methods have been developed, like `icarogw` (Mastrogiovanni et al. (2023)), `GWCOSMO` (Gray et al. (2023)) and `CHIMERA` (first introduced in Borghi, Mancarella et al. (2024) and further optimized in Tagliazucchi et al. (2025)).

Kernel Density Estimate (KDE): For each GW event i, Combined Hierarchical Inference Model for Electromagnetic and gRavitational Wave Analysis (`CHIMERA`) replaces the analytic likelihood in the full posterior (Eq. (2.10)) with a Kernel Density Estimate (KDE):

$$p(\boldsymbol{d}^{GW}|\lambda) \propto \frac{1}{\xi(\boldsymbol{\lambda})^{N_{\rm ev}}} \prod_{i=1}^{N_{\rm ev}} \int\int \mathcal{K}_i^{GW}(z,\hat{\Omega}|\boldsymbol{\lambda}_c,\boldsymbol{\lambda}_m) \frac{p_{gal}(z,\hat{\Omega}|\boldsymbol{\lambda}_c)\; p_{rate}(z|\boldsymbol{\lambda}_c)}{\int p_{gal}(z,\hat{\Omega}|\boldsymbol{\lambda}_c)\; p_{rate}(z|\boldsymbol{\lambda}_c) dz d\hat{\Omega}}\; dz d\hat{\Omega} \tag{2.28}$$

The key object is the GW *kernel*, $K_{\mathrm{GW},i}(z,\hat{\Omega}\,|\,\boldsymbol{\lambda}_c,\boldsymbol{\lambda}_m)$, which represents the marginaliz-

ation over the mass probability of the GW posterior.

$$\mathcal{K}_i^{GW}(z,\hat{\Omega}|\boldsymbol{\lambda}_c,\boldsymbol{\lambda}_m) = \int\int \frac{p(\boldsymbol{\theta}_i|\boldsymbol{d}_i^{GW},\boldsymbol{\lambda}_c)}{\pi(\boldsymbol{\theta}^{det})} \left|\frac{d\boldsymbol{\theta}_i}{d\boldsymbol{\theta}^{det}}\right| p(m_1,m_2|\boldsymbol{\lambda}_m)\, dm_1\, dm_2 \tag{2.29}$$

The KDEs are integrated over Healpix pixels of equal size used to divide the localization area of the GW event. `CHIMERA` provides three evaluation modes of increasing efficiency:

(1) *Full 3D KDE* in $(\mathrm{RA}, \mathrm{Dec}, z)$ space, by dividing the localization area of each event in N_{pix} pixels and evaluating in each pixel the KDE across the whole discretized redshift range;
(2) *Many–1D* method, in which the 3D posterior is marginalized into a 1D KDE in redshift for each pixel;
(3) *Single–1D* approximation, collapsing the GW posterior into a single redshift distribution.

The second approach is adopted in this work, as it provides an optimal balance between accuracy and computational cost.

Pixelization: Each GW event is pixelized with an adaptive procedure that chooses the best n_{side} value, defining the number of pixels per side, on the basis of the localization area of the event. This procedure ensures that the number of pixels is mantained homogeneous across events with different localization areas. With this pixelization, the angular integral over $d\Omega$ (Eq. (2.28)) is evaluated as a discrete sum of pixel contributions.

Selection Bias Term: The selection bias term, accounting for the correction to selection effects, is implemented in the code using the *injections* method: a set of $N_{\rm inj}$ simulated events describing an entire population of GW events is generated to test the effect of selection and to reflect the same bias on our data. `CHIMERA` solves selection term ξ integral in Eq. (2.6) with Monte Carlo, summing over GW injections:

$$\xi(\boldsymbol{\lambda}) \;\approx\; \frac{1}{N_{\rm inj}} \sum_{j=1}^{N_{\rm det}} \frac{p_{pop}(\theta_j|\lambda)}{p_{draw}(\theta_j)} = \frac{1}{N_{\rm inj}} \sum_{j=1}^{N_{\rm det}} s_j \tag{2.30}$$

where $N_{\rm inj}$ is the number of generated injections, $N_{\rm det}$ the detected subset, and p_{draw} is the probability distribution function from which the injections are extracted. To ensure numerical stability we introduce the *effective* number of injections (Farr et al. (2019)),

and require it to be larger than $5N_{\rm ev}$

$$N_{\rm eff}^{\rm inj} = \left[\sum_{j=0}^{N_{\rm det}} s_j\right]^2 \times \left[\sum_{j=0}^{N_{\rm det}} s_j^2 - \frac{1}{N_{\rm inj}}\left(\sum_{j=0}^{N_{\rm det}} s_j\right)^2\right]^{-1} > 5\,N_{\rm ev} \tag{2.31}$$

This ensures, when $N_{\rm eff}^{\rm inj}$ is large enough, that this Monte Carlo integral is a good approximation of the analytical one (Eq. (2.6)).

Galaxy terms: The *catalog term* p_{cat} is pre-computed once and then stored for later use. This is a crucial strategy to avoid large computational costs. The *completeness* is accounted, following the method applyed in Finke et al. (2021), applying a number of masks N_{mask} to group together healpix pixels with the implementation of k-means clustering method*. Within each mask, the completeness is evaluated by comparing the observed galaxies to the theoretical background distribution for each redshift bin. Also this term can be computed at the beginning and stored. The *background distribution* is computed by integrating the Schechter stellar–mass function over the host–mass range. Since this integral is recomputed at each likelihood evaluation, employing Just-In-Time (JIT) optimization is essential to improve efficiency of the pipeline.

Population models: `CHIMERA` provides a modular library of population models, covering the main astrophysical and cosmological components relevant for GW population inference.

The models currently included are:

- *Mass distributions* (`mass.py`): Truncated Power Law, Broken Power Law, Power Law + Peak, and Power Law + Two Peaks. These parameterizations span from simple monotonic behaviors to more structured distributions capturing an increasing number of features to better reproduce the empirical mass distribution.
- *Rate evolutions* (`rate.py`): power-law (PL) and Madau–Dickinson model. The latter describes the cosmic rise and fall of the progenitor formation rate and is widely adopted in BBH studies.

* K-means is an iterative imaging method: starting from a chosen number of clusters, it alternates between assigning each point to the nearest centroid and updating centroids to the mean of their assigned points, minimizing within cluster variance.

- *Cosmological models* (`cosmo.py`): flat ΛCDM, flat ΛCDM with modified-gravity, and $w_0 w_a$CDM.

3

Exploring the Information Content in GW and Galaxy Data

The analysis presented in this thesis relies on two primary sources of information: gravitational wave signals from binary black hole mergers and the distribution of their potential host galaxies. Both datasets are essential and complementary: gravitational waves trace the distribution of compact binary mergers, while galaxy catalogs encode the underlying matter distribution. Their combination enables the use of gravitational waves as *dark sirens* as presented in Chapter 1 and 2, allowing cosmological parameters to be statistically inferred in the absence of electromagnetic counterparts.

The following sections describe the methods used to construct the galaxy catalogs derived from the MICE simulation performed in Borghi, Mancarella et al. (2024); Borghi, Moresco et al. (2025). These catalogs provide a realistic population of potential host galaxies for simulated GW events.

Building on this foundation, the central goal of this chapter is to explore the statistical relations among the main parameters that characterize GW detections in the two catalogs. Correlations are examined through different statistical methodologies and, on this basis,

a series of sub–catalogs defined by progressively higher thresholds in SNR are constructed. This strategy allows the distribution of information across the population of events to be mapped and provides the basis for Chapter 4, where the impact of such event selections on the accuracy of cosmological and astrophysical parameter estimation is quantified.

3.1 Catalogs Generation

3.1.1 Galaxy catalogs

The mock galaxy catalogs used in this work are constructed from the MareNostrum Instituto de Ciències de l'Espacio (MICE) Grand Challenge v2 light-cone simulation (Carretero et al. (2014), Fosalba, Gaztanaga et al. (2013), Crocce et al. (2015), Fosalba, Crocce et al. (2015)). This simulation provides a synthetic galaxy distribution over $\sim 5157\, deg^2$, reaching a redshift of $z \sim 1.4$, and including galaxies down to an apparent magnitude $i < 24$. It covers one octant of the sky (see Fig. 3.1.1), corresponding to the angular ranges $0 \leq ra \leq \pi/2$, $0 \leq dec \leq \pi/2$, and assumes a flat ΛCDM cosmology with $H_0 = 70$ km s$^{-1}Mpc^{-1}$, $\Omega_m = 0.25$, and $\Omega_\Lambda = 0.75$.

Galaxy properties in the simulation are assigned using a combination of Halo Occupation Distribution (HOD) and SubHalo Abundance Matching (SHAM) techniques. This approach starts from an N-body dark matter simulation, where Dark Matter (DM) halos are identified and used as seeds to place galaxies. All halos are populated with one central galaxy and a number of satellite galaxies derived from an HOD that specifies their mean abundance as a function of halo mass. A luminosity is then assigned to the central galaxy through a halo mass–luminosity relation derived via SHAM (Crocce et al. (2015)).

The combination of large cosmological volume and fine mass resolution in the simulation allows for accurate modeling of structure formation across all relevant scales, from the linear Gaussian regime to the highly non-linear, ensuring that the resulting galaxy catalog reproduces observed distributions and clustering properties (Fosalba, Crocce et al. (2015)).

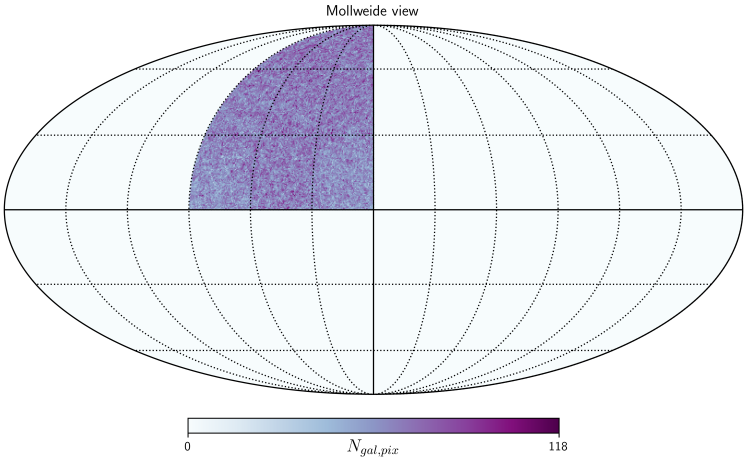


Figure 3.1.1: Galaxy distribution of the mock galaxy catalog covering one sky octant; the color scale indicates the relative number of galaxies per pixel.

From the original MICE mock, two distinct galaxy catalogs are extracted:

Uniform in Comoving Volume (UCV) catalog is obtained by subsampling the full MICEv2 catalog to reproduce a distribution uniform in comoving volume. Specifically, galaxies are extracted with probability $p(z) \propto \frac{dV_c}{dz}$, where dV_c/dz is the differential comoving volume element. This procedure, implemented in (Borghi, Mancarella et al. (2024)), ensures that the catalog provides a homogeneous sampling across cosmic volume.

To mimic a population of potential host galaxies for compact binary mergers while maintaining computational efficiency, a stellar–mass threshold of $\log_{10}(M_\star/M_\odot) > 10.5$ is applied, thereby restricting the sample to massive galaxies. This choice is motivated by both theoretical and observational arguments: binary black hole merger rates are expected to trace stellar mass, with more massive galaxies being preferential hosts. The resulting catalog contains approximately 1.6×10^6 galaxies above the adopted stellar–mass threshold.

Full MICE catalog is built starting from the MICEv2 light–cone simulation. To obtain a parent galaxy catalog suitable for cosmological inference, the stellar–mass distribution of MICE is modeled with an evolving double–Schechter function, ensuring a smooth parametric description of the redshift distribution. This approach

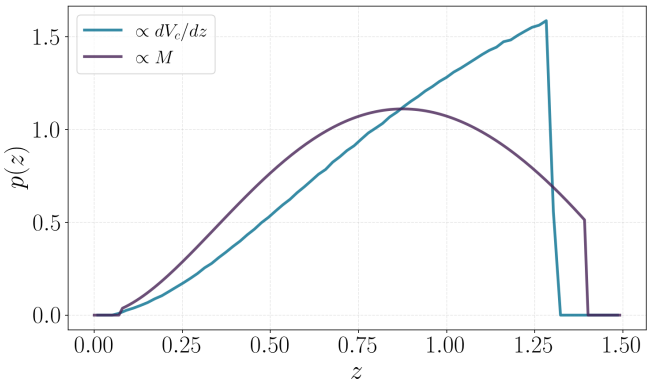


Figure 3.1.2: Redshift distribution of the galaxies in the two catalogs: UCV with $p(z) \propto dV/dz$ and Full MICE with $p(z) \propto \int \Phi(M, z)\, dM$.

is the one introduced in Borghi, Moresco et al. (2025), offering a realistic representation of an observed galaxy catalog. Following the same approach, the stellar–mass completeness limit is determined as a function of redshift, and all galaxies below this evolving threshold are discarded. The observed stellar mass functions in each redshift slice are then fitted with the parametric double–Schechter model. The resulting catalog contains about 335 million galaxies, corresponding to roughly two–thirds of the full MICEv2 sample. To assign merger hosts, each galaxy is weighted according to its stellar mass. They consider three prescriptions: a uniform weighting $\propto M_\star^0$, a linear weighting $\propto M_\star$, and a quadratic weighting $\propto M_\star^2$. These weightings progressively enhance the preference for massive galaxies as potential hosts: while the uniform case treats all galaxies equally, the linear and quadratic weightings increasingly emphasize high mass systems, reducing the contribution of low mass galaxies. In the following analysis the linear weighting scheme is adopted, which is motivated by theoretical predictions that binary merger rates scale approximately with stellar mass (Artale et al. (2020)). The consequences of adopting wrong weighting prescriptions, and the impact of these assumptions in the cosmological constraints, will be investigated in detail in Sec. 4.4.

Although both catalogs originate from the same MICEv2 light–cone, they differ substantially in construction and number density. The Full MICE catalog retains the complete galaxy population above the imposed mass and magnitude limits, yielding $\sim 3.3 \times 10^8$ objects, whereas the UCV catalog is diluted to $\sim 1.6 \times 10^6$ galaxies to reproduce a

uniform distribution in comoving volume. This disparity of more than two orders of magnitude in galaxy counts has a direct impact on their use for GW host association: localization volumes that in the UCV case contain only a few possible hosts typically enclose thousands of galaxies when the Full MICE catalog is adopted. The consequences of this difference will be further explored in Sec. 3.3, where we investigate how the number of galaxies per localization volume scales with the SNR of the events, and in Chapter 4, where we assess its impact on the inference of cosmological parameters.

Different redshift uncertainties might be associated with galaxies in both catalogs; in this work, we considered the case of a spectroscopic survey, with $\sigma_z = 0.001(1+z)$.

3.1.2 GW catalogs

For each galaxy catalog, we extracted a corresponding set of GW events by associating compact binary coalescences to potential host galaxies. Each GW catalog contains 1000 events, with 5000 posterior samples per event, emulating the typical output of the LVK parameter estimation pipeline.

The GW catalogs are generated by modeling both the astrophysical source population and the observational selection effects of Observing run 5 (O5) for the LVK detector network. We assume that the BBH merger rate follows a Madau–Dickinson redshift evolution (Fig. 2.1b). For the BBH primary mass distribution, we adopt a PLP model (Fig. 2.1a), which captures both a power-law tail and a Gaussian peak at $\sim 35\ M_\odot$.

The GW signals are simulated using `GWFAST` (Iacovelli et al. (2022)), assuming quasi–circular, non–precessing BBHs. As introduced in Sec. 2.1.1, each waveform is characterized in the detector frame by

$$\boldsymbol{\theta}^{\rm det} = \{\mathcal{M}_c, \eta, d_L, \alpha, \delta, \iota, \chi_{1,z}, \chi_{2,z}, \psi, t_c, \phi_c\}, \tag{3.1}$$

where $\mathcal{M}_c$ is the chirp mass, η the symmetric mass ratio, d_L the luminosity distance, (α, δ) the sky coordinates, ι the inclination angle, $\chi_{1,z}, \chi_{2,z}$ the aligned spin components, ψ the polarization, and (t_c, ϕ_c) the coalescence time and phase.

For each source, the matched–filtered SNR is computed with the `IMRPhenomHM` waveform model (London et al. (2018); Kalaghatgi, Hannam and Raymond (2020)), which represents a full *inspiral-merger-ringdown* waveform model that includes higher–order

modes for an accurate description of signals from BBH. The *matched filter* SNR for a single detector is computed as:

$$SNR_i^2 = 4 \int_{f_{min}}^{f_{max}} \frac{|\tilde{h}_i(f)|^2}{S_{n,i}(f)} df \tag{3.2}$$

where $\tilde{h}_i(f)$ is the GW strain in the Fourier domain for the i^{th} detector and $S_{n,i}(f)$ is the noise spectral density of the i^{th} detector. For a network, composed of multiple detectors, SNR is defined as the sum in quadrature of the SNRs of single detectors:

$$SNR^2 = \sum_i SNR_i^2 \tag{3.3}$$

In our analysis, following the work by Borghi, Moresco et al. (2025), we adopt the O5 detector network configuration, consisting of the two LIGO instruments, Virgo, KAGRA, and LIGO–India, with a low–frequency cutoff of 10 Hz, and a 100% duty cycle. Only events with SNR > 25 are retained, yielding catalogs of $N_{\rm ev} = 1000$ detections.

Posterior distributions are generated using the Fisher Information Matrix (FIM) approximation, valid for high SNR detections. In the case of high SNR and Gaussian noise, the inverse of the FIM gives the covariance of the Bayesian posterior probability distribution of the true waveform parameters. For each event, 5000 samples are drawn from a multivariate Gaussian with covariance given by the inverse FIM, imposing priors consistent with the simulated distributions.

Finally, to estimate the GW selection effects, dedicated *injection* sets have been generated with `GWFAST`, adopting the same assumptions and SNR ≥ 25 threshold as for the catalogs. These injections cover the same sky area and extend in luminosity distance up to the detector horizon. The injection set contains $\sim 2 \times 10^7$ simulated sources, of which about 10^6 are detected, and is used to compute the selection bias term entering the likelihood (see Chapter 2).

3.2 Data Analysis and Characterization

I performed a preliminary exploratory analysis to investigate the mutual relationships between key parameters in the dataset. The parameters considered include the network SNR, luminosity distance (d_L), component masses (m_1, m_2), sky localisation area (Ω), distance uncertainty (Δd_L), localisation volume ($V_{\rm loc}$), and the estimated number of galaxies within the localisation volume ($N_{\rm gal,vol}$). Most of these quantities were provided directly in the GW event catalogs, while I derived the others from the posterior samples.

I derived the luminosity distance uncertainty, Δd_L, as the 90% credible interval of the posterior distribution of d_L, computed as the difference between the 95° and the 5° percentiles: $\Delta d_L = d_L^{95^\circ} - d_L^{5^\circ}$.

The sky localization area, Ω, is computed with the `compute_localization_areas` `CHIMERA` routine. Starting from the ϕ and θ localization angles, the method estimates the 90% credible region on the sphere and reports its size in square degrees.

The localization volume, $V_{\rm loc}$, is obtained with `compute_localization_volumes` by combining the 90% sky region with the central 90% credible interval in luminosity distance. We adopt a flat ΛCDM cosmology with $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_m = 0.25$ to convert distances to redshift and comoving volume; the resulting $V_{\rm loc}$ is expressed in Gpc^3.

Finally, the number of galaxies within the localization volume, $N_{\rm gal,vol}$, is derived by intersecting $V_{\rm loc}$ with the pixelated galaxy distribution from the chosen catalog. For each event, galaxies were counted in the sky pixels overlapping the GW localization, and only those within the redshift range defined by the GW distance posterior were retained. The computation was carried out with the `CHIMERA` pixelated catalog tools, which allow an efficient match between GW posteriors and galaxy distributions. The resulting $N_{\rm gal,vol}$ thus represents the expected number of potential host galaxies consistent with the GW localization at 90% credibility.

The derived set of parameters provides a consistent basis for the analysis presented in the following sections. In Sec. 3.2.1 the mutual correlations each parameters pair are investigated, highlighting the main statistical dependencies in the catalogs. The impact of the assumed cosmology on derived quantities such as $V_{\rm loc}$ and $N_{\rm gal,vol}$ is then assessed in Sec. 3.2.2. A Principal Component Analysis (PCA) is introduced in

Sec. 3.2.3 to identify the dominant drivers of variance. Finally, in Sec. 3.3 the results of this preliminary analysis are used to construct a set of sub–catalogs, selected with progressively higher SNR thresholds, which will serve as the basis for the inference studies presented in Chapter 4.

3.2.1 Analysis of the Correlations

To verify the potential correlations between each of the parameters described above, a corner plot is generated (Fig. 3.2.1), displaying projections of the pairwise distributions. Each off–diagonal panel shows the scatter plot between two parameters, allowing a direct visual inspection of their distribution and the possible presence of correlations or degeneracies. The diagonal panels display the corresponding one–dimensional distributions, providing information on the spread of each parameter.

A complementary quantitative assessment is obtained through the computation of the Spearman correlation coefficient ρ for each parameters pair (Fig. 3.2.2) which, ranging from -1 (perfect negative correlation) to $+1$ (perfect positive correlation), provides a measure of correlation, particularly suitable for parameters whose relationships may not be strictly linear.

In Fig. 3.2.2, the lower triangle of the matrix reports the Spearman correlation coefficients obtained for the *UCV* galaxy catalog, while the upper triangle shows the difference between the UCV coefficients and those computed with the *Full MICE* catalog. This arrangement was designed to highlight any significant differences in correlation trends between the two catalogs. The comparison reveals that, for most parameter pairs, the differences between correlations in the two catalogs are negligible, indicating that the correlation structure is largely independent of the chosen catalog. The only notable variations arise in the correlations involving $N_{\mathrm{gal,vol}}$, which are affected by the intrinsic differences in galaxy number density between UCV and Full MICE. This behavior is expected, as the Full MICE catalog contains about two orders of magnitude more galaxies than UCV. This difference in number density leads to a systematic rescaling of $N_{\mathrm{gal,vol}}$ and, consequently, of its correlations with the other parameters.
On this basis, the following analysis on the correlations can be treated as independent of the considered catalog.

The combined use of the corner plot (Fig. 3.2.1) and the Spearman correlation matrix

(Fig. 3.2.2) provides both a qualitative and quantitative understanding of the dependencies in the dataset, enabling the identification of physically motivated trends. The analysis of the main correlations identified are reported in the following discussion.

Signal–to–Noise Ratio - SNR

The SNR exhibits a remarkably consistent behaviour across the correlation matrix: it is the only parameter with exclusively negative Spearman coefficients with respect to all other variables. This systematic anti–correlation pattern ($\rho_{\mathrm{UCV}} < 0$ in all cases) underscores its role as a primary driver of observational precision. Strongest correlations are observed with luminosity distance d_L ($\rho_{\mathrm{UCV}} \approx -0.40$), distance uncertainty Δd_L ($\rho_{\mathrm{UCV}} \approx -0.59$), localization area Ω ($\rho_{\mathrm{UCV}} \approx -0.53$), and localization volume V_{loc} ($\rho_{\mathrm{UCV}} \approx -0.54$). These trends are physically well–motivated: sources at closer distances yield higher strain amplitudes, leading to higher SNR and reduced parameter uncertainties (Abbott et al. (2021)). Higher SNR also enhances triangulation accuracy across the detector network, having an impact on the smaller both angular and volumetric localization regions. This directly reduces the number of potential host galaxies, producing a clear anti–correlation with $N_{\mathrm{gal,vol}}$ ($\rho_{\mathrm{UCV}} \approx -0.54$).

Luminosity Distance and Binary Masses - d_L**,** m_1**,** m_2

In contrast to SNR, the luminosity distance d_L displays uniformly positive correlations with all other parameters (except SNR), with $\rho_{\mathrm{UCV}} > 0.8$. This strong, monotonic behaviour indicates that increasing d_L systematically degrades localisation precision and increases volumetric measures, while also modifying the observed population properties. A particularly strong positive correlation is found between d_L and the component masses m_1 and m_2 (both $\rho_{\mathrm{UCV}} \approx 0.8$), a direct consequence of the GW selection function: at larger distances, only more massive binaries, producing intrinsically louder signals, remain detectable. Lower–mass systems fall below the sensitivity threshold, introducing a population bias.

Sky Localization Parameters - Ω**,** V_{loc}

The angular localization area Ω and the localization volume V_{loc} are almost perfectly correlated ($\rho_{\mathrm{UCV}} > 0.92$), reflecting their geometric relationship: the volume is obtained by combining the angular extent with the line–of–sight distance uncertainty. Both

parameters strongly increase with d_L, as faint and distant events suffer from larger distance uncertainties. These localization metrics also correlate positively with $N_{\mathrm{gal,vol}}$ ($\rho_{\mathrm{UCV}} \approx 0.9$), driven by the trivial scaling: larger localization volumes contain more galaxies, assuming a roughly homogeneous galaxy distribution at the scales considered.

Number of Galaxies in Localisation Volume - $N_{gal,vol}$

$N_{\mathrm{gal,vol}}$ follows the general trend of positive correlations with all parameters except SNR, for which the relationship is negative. The anti–correlation with SNR reflects the fact that higher SNR events are more precisely localised, hence enclosing fewer potential hosts. However, at low SNR, the correlation weakens and the scatter increases. While $N_{\mathrm{gal,vol}}$ decreases systematically with increasing SNR, the relationship is not strictly monotonic for low–SNR events. In some cases, low–SNR detections correspond to small $N_{\mathrm{gal,vol}}$ due to either fortuitously narrow localization or the event being located in a low density region of galaxy distribution.

Pixelization Parameters - n_{side}, n_{pix}

The HEALPIX resolution parameter n_{side} displays a strong anti–correlation with the localization area Ω, in line with the adaptive pixelization scheme implemented in the pipeline discussed in Sec. 2.4. By design, events with larger Ω are assigned lower angular resolutions (lower n_{side}), thereby keeping the number of pixels within computationally manageable limits while preserving sufficient angular detail for smaller–area events.

The total number of pixels per event, n_{pix}, does not exhibit any significant correlation with SNR or with the astrophysical source parameters, as expected. This behaviour reflects the pixelization strategy: a target average n_{pix} is defined prior to the process, and the algorithm then selects the optimal n_{side} for each event to achieve a close match to this target while adapting to the event's localization geometry. This is further confirmed by the discrete nature of the n_{pix} distribution, clearly visible in its histogram, which shows peaks at preferred values (with a dominant mode at $n_{\mathrm{pix}} \approx 15$) corresponding to the selected target configurations. In the scatter plots against other parameters, n_{pix} values appear uniformly distributed across the plane, confirming the independence of pixel count from SNR and other physical quantities. Together, these results validate that the adaptive pixelization achieves the intended balance between resolution and computational efficiency without introducing systematic biases in the spatial sampling.

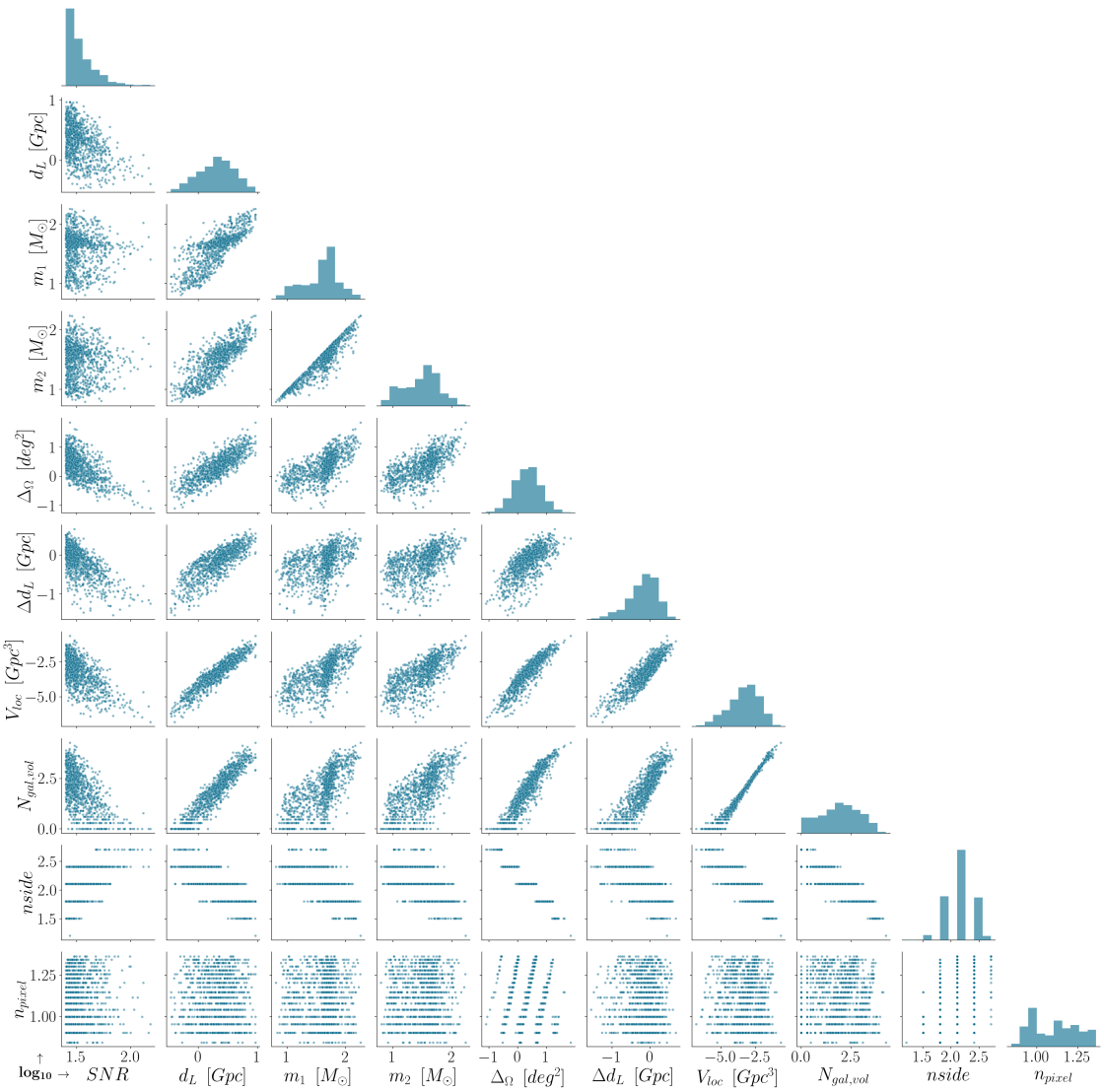


Figure 3.2.1: Matrix of scatter plots of the key GW parameters. Diagonal panels report the one-dimensional distributions of each variable, while off-diagonal panels show pairwise scatter plots. Both axes across the entire grid are shown in $\log_{10}$ scale. The parameters include the signal-to-noise ratio (SNR), the median luminosity distance (d_L), the component detector-frame masses (m_1, m_2), the sky localization area (Ω), the uncertainty on the luminosity distance (Δd_L), the localized comoving volume (V_{loc}), the number of galaxies within this volume ($N_{gal,vol}$), the HEALPix resolution parameter (n_{side}), and the number of pixels in the localization map (n_{pixel}).

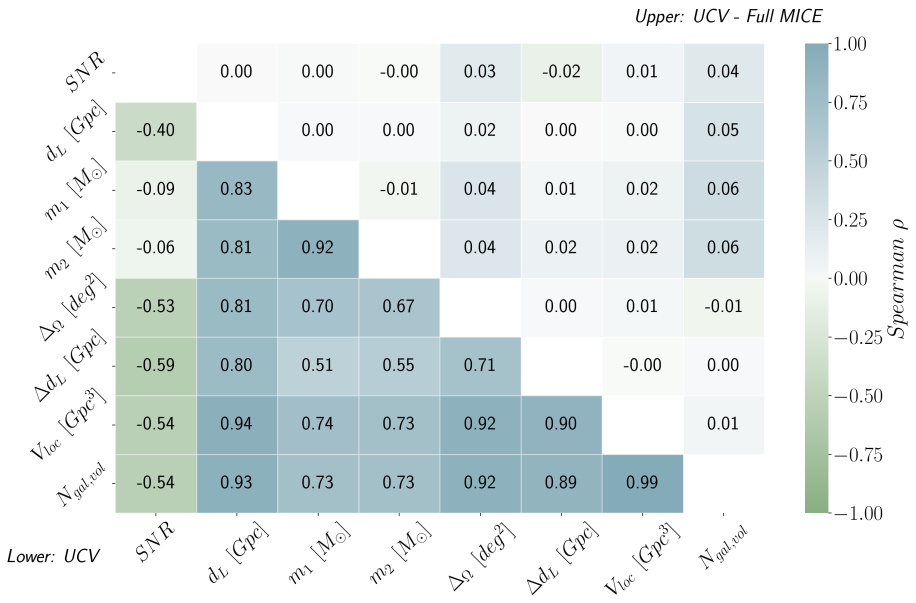


Figure 3.2.2: Spearman Matrix resulting from the analysis of correlations between parameters in Fig. 3.2.1. The matrix is divided into the upper and lower corners with respect to the diagonal. *Lower:* Spearman coefficients between the various parameters for the UCV catalog; *Upper:* Difference between the values of the coefficients in the UCV and Full MICE catalogs.

The joint use of scatter plot visualization and correlation coefficients allows the targeting of physically meaningful relationships. This diagnostic is essential for correctly interpreting parameter dependencies in hierarchical inference and assessing potential degeneracies in the hyperparameter space explored in later chapters.

3.2.2 Effect of Cosmology on $V_{\rm loc}$ and $N_{\rm gal,vol}$

The estimation of the localization volume $V_{\rm loc}$ for a GW event, and consequently the number of galaxies within this volume $N_{\rm gal,vol}$, is significantly dependent on the adopted cosmological model. The localization volume is defined by the joint uncertainty in the event's sky position (RA, Dec) and distance, the latter obtained from the luminosity distance–redshift relation, which depends on the assumed cosmology. Changes in cosmological parameters, especially in the Hubble constant H_0, directly modify the inferred redshift and, consequently, the associated comoving volumes.

To investigate the effect of varying cosmological assumptions on localization volumes

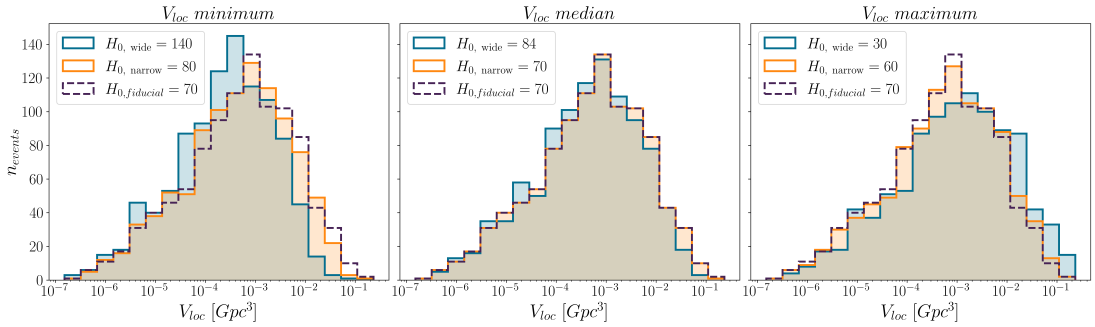


Figure 3.2.3: Histograms of localization volumes $V_{\rm loc}$ for gravitational wave events, showing the minimum, median, and maximum volume estimates across different cosmological assumptions on the Hubble constant H_0. The three panels compare distributions obtained with a fiducial fixed value $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (dashed line) against two ranges of H_0: a narrow interval $[60, 80]$ (orange) and a wide interval $[30, 140]$ (blue).

and galaxy counts, tests were conducted by varying the prior ranges of H_0. Three configurations for H_0 were considered:

Wide Range: $H_0 \in [30, 140]\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, representing a highly uninformative cosmological prior, reflecting a large uncertainty in the cosmological model.

Narrow Range: $H_0 \in [60, 80]\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, a more constrained range consistent with current tension between early and late-Universe measurements.

Fiducial Value: $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, fixed value representing the standard assumption.

For each event and each H_0 configuration, the localization volume (V_{loc}) and the number of galaxies within this volume ($N_{gal,vol}$) are recalculated. After calculating V_{loc} and $N_{gal,vol}$ for each scenario, histograms are generated (Fig. 3.2.3 and 3.2.4) to compare the differences between the various cosmological assumptions.

As illustrated in Fig. 3.2.4, by allowing H_0 to vary freely within each considered range, multiple realizations of localization volumes $V_{\rm loc}$ are generated. For each GW event, the minimum, median, and maximum values of $V_{\rm loc}$ are selected, while keeping track of the corresponding H_0 values.

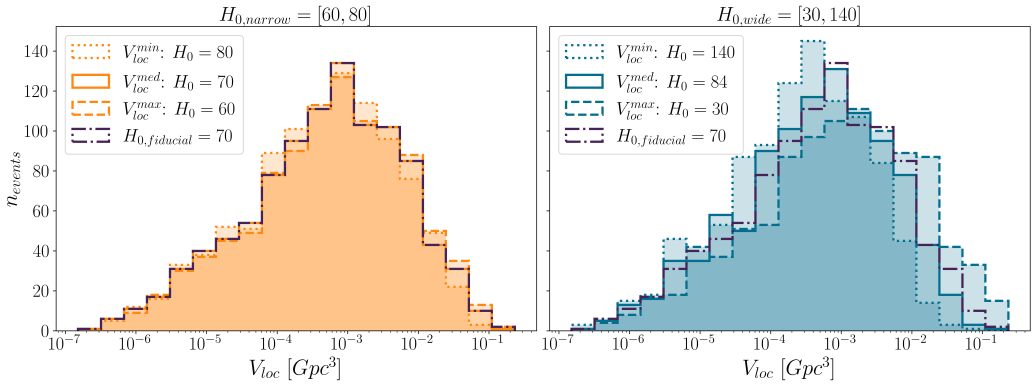


Figure 3.2.4: Histograms of localization volumes $V_{\rm loc}$ for gravitational wave events comparing two cosmological parameter ranges for the Hubble constant H_0: a narrow range $[60, 80]$ (left panel) and a wide range $[30, 140]$ (right panel). Within each panel, the minimum, median, and maximum volume estimates are shown with distinct line styles. The fiducial case with fixed $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ is overlaid as a dash-dot line.

The analysis shows that for all events, the *minimum* localization volume always corresponds exactly to the upper bound of the H_0 range ($80\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for the narrow range, $140\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for the wide range), reflecting the fact that larger H_0 values reduce inferred distances for a fixed redshift, thereby shrinking the comoving volume. Conversely, the *maximum* $V_{\rm loc}$ corresponds to the *minimum* allowed H_0, where smaller expansion rates imply larger distances and hence larger volumes The *median* $V_{\rm loc}$ in the narrow range closely aligns with the fiducial $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, whereas in the wide range it is biased toward slightly higher values ($H_0 \approx 84\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$). This consistent behavior across all events underlines the strong dependence of localization volumes on the chosen H_0 within the given ranges.

Since $N_{\rm gal,vol}$ is strongly correlated with the localization volume, its dependence on H_0 mirrors that of $V_{\rm loc}$. As seen before, for fixed luminosity distance posteriors, increasing H_0 maps the same d_L to lower redshift, which reduces both $V_{\rm loc}$ and $N_{\rm gal,vol}$. Figure 3.2.5 shows this behavior for both catalogs: the trend with H_0 is the same, while the overall normalization differs because Full MICE has a higher galaxy number density than UCV.

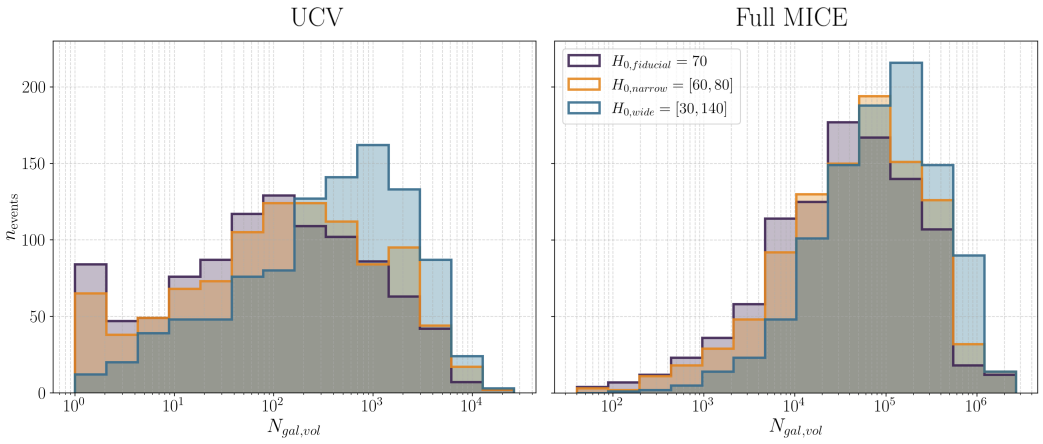


Figure 3.2.5: Distribution of the number of galaxies within the localization volumes $N_{\mathrm{gal,vol}}$ for different assumptions on the Hubble constant H_0. Each histogram corresponds to one of the three cosmological assumptions considered: the fiducial value $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, narrow range $[60, 80]$, and wide range $[30, 140]$. *Left:* UCV catalog, *Right:* Full MICE catalog.

3.2.3 Principal Component Anlysis

To further explore correlations among the key parameters involved in the cosmological and astrophysical parameters inference, we perform a PCA on a selected subset of variables.

PCA is a classical dimensionality–reduction technique originally introduced in Pearson (1901) and further formalized in Hotelling (1933). Its central idea is to represent a dataset with a new set of orthogonal variables, the *principal components* (PCs), which are linear combinations of the original variables and capture the directions of maximum variance in the data. Intuitively, PCA provides a low–dimensional representation of the data in which the main patterns of correlation are preserved, while redundant variance is compressed into higher order components.

Formally, let $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ be the data matrix with n observations and p variables. Since variance depends on measurement scale, all variables are first standardized to a zero mean and unit variance so that each variable contributes equally to the analysis

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{3.4}$$

where $\bar{x}_j$ and σ_j denote the mean and standard deviation of the j-th variable, and Z is the obtained standardized matrix . This step is essential because PCA is based on variance maximization, and variances are sensitive to the units of measurement of the variables. Without standardization, variables with larger scales would dominate the first components and bias interpretation (Jolliffe and Cadima (2016)).

The covariance matrix of Z (which coincides with the correlation matrix in the case of standardized components) is then:

$$\boldsymbol{C} = \frac{1}{n-1}\boldsymbol{Z}^\top \boldsymbol{Z}. \tag{3.5}$$

PCA relies on the eigendecomposition of the covariance matrix $\boldsymbol{C}$,

$$\boldsymbol{C}\,\mathbf{v}_k = \lambda_k\,\mathbf{v}_k, \qquad k = 1, \ldots, p, \tag{3.6}$$

yielding orthonormal eigenvectors $\mathbf{v}_k$ and eigenvalues λ_k. The k-th principal component for all observations is obtained by projecting the standardized data onto $\mathbf{v}_k$

$$\mathrm{PC}_k = Z\,\mathbf{v}_k. \tag{3.7}$$

By construction, successive components are uncorrelated and explain decreasing fractions of the total variance. The eigenvectors $\mathbf{v}_k$ provide the coefficients (weights) that define each PC as a linear combination of the standardized variables, while the eigenvalues quantify the variance explained by each component, with $\mathrm{Var}(\mathrm{PC}_k) = \lambda_k$. The explained–variance ratio (EVR) of component k is

$$\mathrm{EVR}_k = \frac{\lambda_k}{\sum_{j=1}^{p}\lambda_j}, \tag{3.8}$$

which represents the fraction of total variance captured by that component. Inspecting the EVR (and its cumulative sum) provide a standard criterion to determine how many PCs to retain in the analysis.

In the context of this thesis, PCA is used as a diagnostic tool to identify which parameters dominate the variance of the dataset and to investigate inter–parameter correlations. The variables included are the SNR, the luminosity distance d_L and its uncertainty Δd_L, the primary mass m_1, the localization area Ω, the localization volume V_{loc}, and the

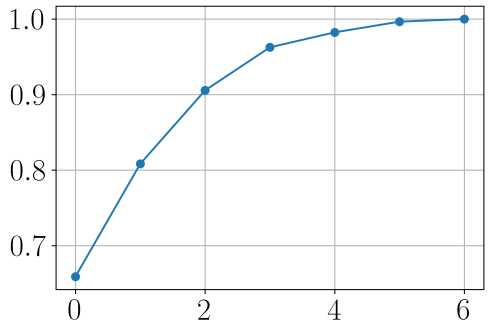


Figure 3.2.6: Cumulative explained variance as a function of the number of principal components. The plots show how the addition of each principal component contributes to the total variance of the dataset. The first three components account for nearly 90% of the total variance.

number of galaxies within the localization volume $N_{\mathrm{gal,vol}}$. We consider only the first three principal components based on cumulative explained variance (Fig. 3.2.6), since together they account for nearly 90% of the total variance (65.9%, 14.9%, and 9.7%), providing a compact, lower–dimensional representation that is useful for visual inspection and further correlation studies. The analysis is implemented using the `scikit-learn` package (Pedregosa et al. (2018)).

To visualize the PCA results, Fig. 3.2.7 shows a *correlation biplot* of the first three principal components, providing an intuitive geometric view of how variables relate to one another and to PCs. Each panel displays the data projected onto a pair of PC axes (PC1–PC2, PC2–PC3, PC1–PC3), while colored arrows depict, for each original (standardized) variable, its *loadings* (L) on the displayed PCs:

$$L_k = \mathbf{v}_k \cdot \sqrt{\lambda_k} \tag{3.9}$$

This quantifies the correlations between the standardized variable and the principal components. In a correlation biplot, the cosine of the angle between two arrows approximates the correlation between the corresponding variables within the displayed PC plane: arrows in similar directions (small angles) indicate positive association; arrows in opposite directions (angles near 180°) indicate negative association; orthogonal arrows indicate weak association. For instance, in our case the vectors associated with V_{loc}, Ω, and $N_{\mathrm{gal,vol}}$ are nearly parallel, reflecting strong mutual correlations driven by

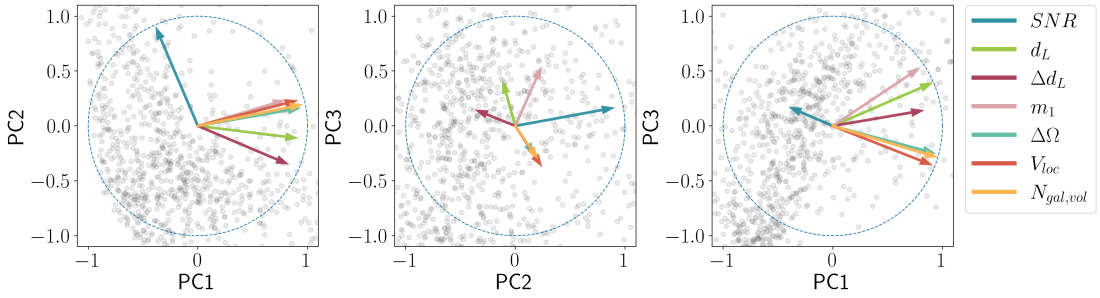


Figure 3.2.7: Correlation biplot of the first three PCs from a PCA on standardized variables. Each panel shows observation scores and variable arrows equal to the loadings (L_{jk}). Arrow orientation encodes the sign of association; arrow length reflects the magnitude of the loadings in the displayed plane.

localization accuracy and distance. Likewise, the loading of a variable on PC_k equals its correlation with that PC and is given by the coordinate of the arrow along the PC_k axis.* The *length* l of each arrow in a given panel reflects how well the variable is represented by that 2D PC subspace, and is obtained combining the loadings of the two PCs as: $\|l\| = \sqrt{L_a^2 + L_b^2}$. Longer arrows therefore indicate stronger associations with the shown PCs. This is consistent with the interpretation of loadings as correlations between standardized variables and the principal components (Jolliffe and Cadima (2016)).

To complement the geometric view and quantify how each PC is composed, we also inspect the eigenvector coefficients (weights). The three PC are defined with a linear combination of the standardized variables, each accounted for in the PC with a weight provided by the eigenvector components. So, the functional definition of the three PC can be written as follows:

$$\begin{aligned} PC_1 &= -0.17\ \mathrm{SNR} + 0.41\ d_L + 0.37\ \Delta d_L + 0.36\ m_1 + 0.42\ \Omega + 0.41\ V_{loc} + 0.43\ N_{gal,vol} \\ PC_2 &= 0.85\ \mathrm{SNR} + -0.11\ d_L - 0.33\ \Delta d_L + 0.22\ m_1 + 0.15\ \Omega + 0.22\ V_{loc} + 0.18\ N_{gal,vol} \\ PC_3 &= 0.19\ \mathrm{SNR} + 0.45\ d_L + 0.16\ \Delta d_L + 0.61\ m_1 - 0.29\ \Omega - 0.41\ V_{loc} - 0.33\ N_{gal,vol} \end{aligned} \tag{3.10}$$

where all variables have been standardized (zero mean and unit variance) before entering the principal component. Fig. 3.2.8 reports a heatmap of the weights associated with the initial variables for the first three PCs. Here, the color encodes the *signed weight* w_{kj} of the j-th variable in PC_k (rows sum of squares equal 1), which tells how much

* Because the figure shows a 2D subspace, both angles and lengths summarize contributions restricted to the chosen pair of PCs. Contributions along omitted PCs are not visible.

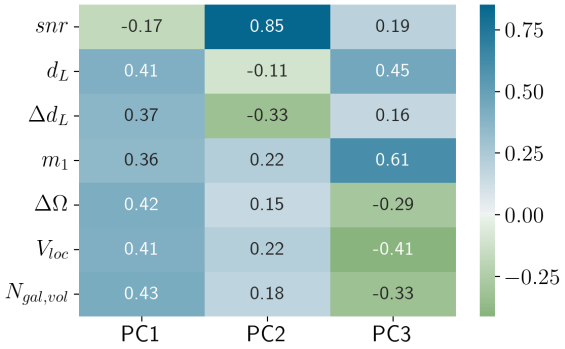


Figure 3.2.8: Heatmap of eigenvector coefficients (component weights) for the first three PCs. The color scale encodes the strength and sign of the correlation between each standardized variable and the PCs, with darker shades indicating stronger contributions.

each variable enters the linear combination defining that PC.

In addition, a *distance biplot* is provided in Fig. 3.2.9, showing how the original variables project onto the space of the principal components. This representation offers a geometric visualization of the weights: the projection length along a given PC axis reflects its relative importance in that component. It should be noted, however, that in the generated biplot the vectors are rescaled for visualization purposes; therefore, they do not provide direct numerical information on the weights, which is instead reported in the heatmap (Fig. 3.2.8). Taken together, the correlation biplot, the distance biplot, and the weight heatmap provide complementary perspectives: the former emphasizes orientations and correlations among variables, the second illustrates how variables are included in the PCs, and the latter reports the exact coefficients that define each component.

Analysing the heatmap in Fig. 3.2.8, it is evident that the SNR exhibits the strongest association with PC2, with a coefficient of 0.85. This suggest that PC2 is primarly driven by the SNR, meaning that it alone captures the majority of the variance in the dataset.
Other parameters also play a significant role in shaping the variance of the dataset. The primary mass m_1 has a notable weight coefficient of 0.61 in the third principal component (PC3), suggesting it contributes meaningfully to the third direction of variance, even though it is not as influential as SNR in PC2. Additionally we observe that parameters

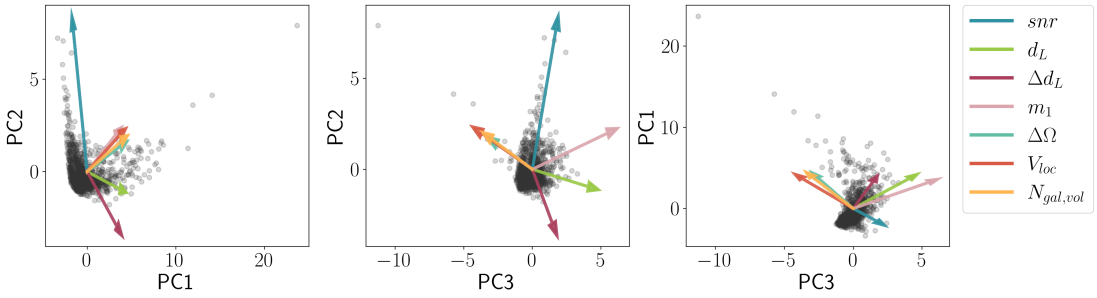


Figure 3.2.9: Distance biplot of the first three PCs from a PCA on standardized variables. Each panel shows a 2D projection of the dataset in the plane of two principal components. Points are projections of the observations onto PC planes; Arrows represent the PCA eigenvector coefficients (*weights*) of the standardized variables. This scaling emphasizes the composition of each PC; arrow lengths is proportional to the magnitude of the weights in the displayed PC plane.

related to the spatial localization of the event, such as d_L, Ω, V_{loc}, and $N_{gal,vol}$, are strongly associated with the first principal component (PC1), with similar coefficients (around 0.4), consistent with their mutual correlations. The primary mass m_1 contributes most to PC3 (coefficient ≈ 0.61), pointing to a distinct third direction of variation. Overall, PC1 explains the largest share of total variance (65.9%), followed by PC2 and PC3 (14.9% and 9.7%), so while SNR strongly *drives* PC2, PC1 remains the dominant source of explained variance in the dataset.

Taken together, our PCA results reveal two dominant and largely orthogonal sources of variability: a localization–driven axis (PC1) and a signal–strength axis (PC2) dominated by SNR. This orthogonality indicates that, to first order, variation associated with SNR is independent of localization effects. In addition, in the PC1–PC3 correlation biplot the SNR arrow points nearly opposite to the localization variables (d_L, Ω, $V_{\rm loc}$, $N_{\rm gal,vol}$), indicating a residual negative association in that subspace.

3.3 Extraction of Optimal GW Samples

The analyses of correlations and principal components presented in the previous sections clearly identified the SNR as the dominant factor driving measurement uncertainties and localization properties. In particular, SNR was shown to correlate negatively with luminosity distance and localization area, and to align in the PCA representation with the direction of variance of PC2. These results motivate a more systematic investigation of how the quality of the detected signals, as traced by their SNR, propagates into cosmological inference.

To this end, we performed a subsampling of the catalogs by imposing successive SNR thresholds. Four cuts were applied, corresponding to SNR ≥ 25, SNR ≥ 35, SNR ≥ 50, and SNR ≥ 70. Selecting events by SNR is standard practice in GW analyses: current observational catalogs typically adopt a network threshold around SNR ~ 12 to discard bad detections and ensure basic parameter estimation quality, while forecasts for O5 detectors often assume higher cuts (SNR$\geq$25) to secure cleaner samples and manageable computational loads. However, the choice of a specific SNR threshold is rarely based on quantitative evidence about its impact on cosmological and population inferences. The analysis carried out here addresses precisely this gap by selecting four cuts from 25 to 70 and measuring how gains in data quality (via SNR) trade off against losses in sample size for the parameter inference.

The immediate effect of increasing the SNR threshold is a progressive reduction in number of retained GW events: the total number of events drops from $N_{\rm ev} = 1000$ at SNR ≥ 25 to only a few tens at SNR ≥ 70 (see Tab. 4.1).

This reduction is not merely a loss of statistics; it also reshapes the sample in a systematic way. Because SNR is tightly correlated with several GW observables and source properties, imposing an SNR cut selects not only fewer events but also a different *region* of parameter space. Imposing a higher SNR threshold preferentially keeps nearby and typically higher–mass binaries. The selected sample is therefore less representative of the underlying population, which can bias or weaken constraints on quantities describing the mass distribution and the redshift evolution of the merger rate.

The redshift distribution (Fig. 3.3.1) becomes increasingly skewed toward the local Universe as the SNR threshold rises. Since high-SNR detections are dominated by strong

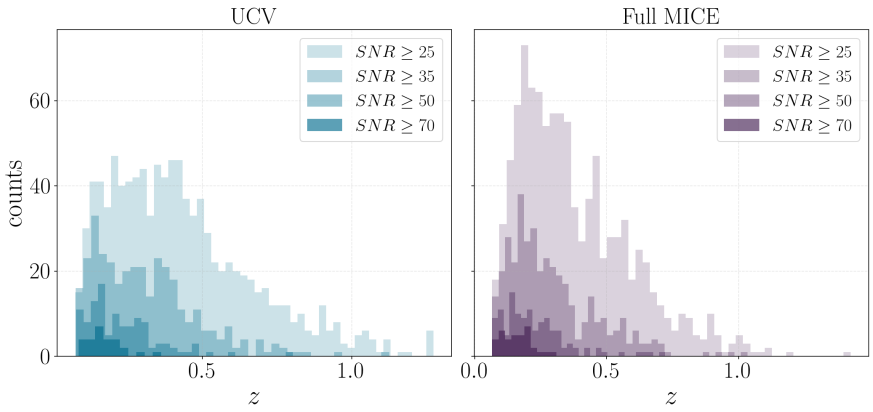


Figure 3.3.1: Redshift distributions for UCV (*left*) and Full MICE (*right*) at different SNR cuts.

signals, the selected events tend to lie at smaller luminosity distances, in line with the strong anti–correlation between SNR and d_L evidenced by the Spearman analysis. As a consequence, the sampled redshift range is progressively truncated at high z values, and the characteristic Madau–Dickinson behaviour (2.1b) mapped by the full population (initial rise, peak, and high z decline) is no longer visible at high cuts in SNR. This makes it increasingly difficult to characterize the merger rate density $\psi(z)$ (see Sec. 2.2), since the restricted redshift coverage does not properly capture the peak and the subsequent decline of the distribution. For the most stringent cut (SNR ≥ 70), the ascending branch and the peak are not observed from the considered subsample. This loss of dynamic range reduces sensitivity to the merger-rate evolution hyperparameters $\boldsymbol{\lambda}_z$ and will have an impact on the inferences (see Chapter 4).

The primary–mass distribution (see Fig. 3.3.2) does not shift systematically with increasing SNR. Instead, the high SNR subsamples retain a relatively uniform distribution even on a truncated mass range, with a preference for intermediate-mass systems. Importantly, the characteristic peak around $35\,M_\odot$ in the intrinsic mass function remains clearly visible even at the most stringent SNR cuts, in a coherent framework in which stringent cuts well represent intermediate mass population while suffer at the extremes of the whole mass distribution. This persistence indicates that, although the number of events is reduced and the statistical power weakened, the high SNR subsets still preserve the main features of the underlying population, maintaining a degree of representativity,

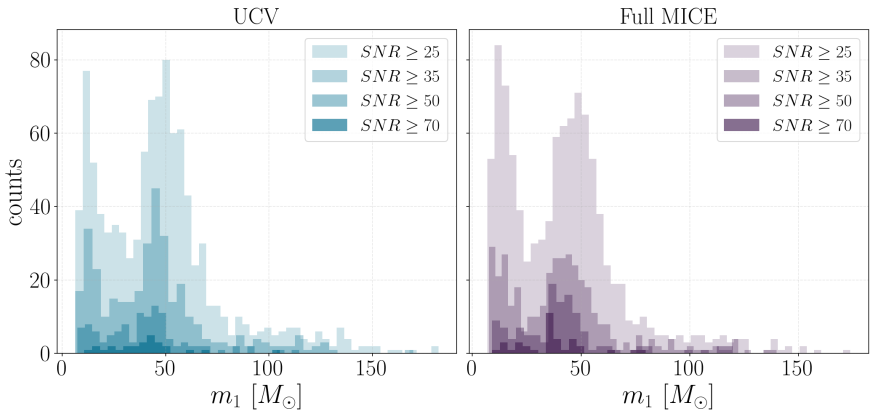


Figure 3.3.2: Mass distributions for UCV (*left*) and Full MICE (*righ*) at different SNR cuts.

especially for the peak.

The effect of these cuts is illustrated in Fig. 3.3.4 and 3.3.5, where scatter–plot matrices show the parameter space distribution of events in each SNR subsample. As anticipated from the correlation analysis, high SNR events cluster at shorter distances, with smaller fractional uncertainties in d_L and smaller Ω, thereby occupying the most informative regions of parameter space. A quantitative summary of these trends is provided in Tab. 3.1, which reports the mean values of Δd_L, Ω, $V_{\rm loc}$, and $N_{\rm gal,vol}$ for each cut. For the UCV catalog, already at SNR ≥ 35 the typical sky localization improves by more than a factor of two compared to the baseline (SNR ≥ 25), while the average localization

Catalog	SNR	$N_{\rm ev}$	$\langle \Delta d_L \rangle$ $[Gpc]$	$\langle \Omega \rangle$ $[deg^2]$	$\langle V_{loc} \rangle$ $[Gpc^3]$	$\langle N_{gal,vol} \rangle$
UCV	≥ 25	1000	0.8	3.7	4.09×10^{-3}	5.90×10^{2}
	≥ 35	347	0.4	1.7	7.11×10^{-4}	1.13×10^{2}
	≥ 50	113	0.2	0.9	1.28×10^{-4}	21.0
	≥ 70	27	0.1	0.3	7.07×10^{-6}	2.4
Full MICE	≥ 25	1000	0.8	3.1	2.89×10^{-3}	1.2×10^{5}
	≥ 35	381	0.4	1.6	7.67×10^{-4}	3.9×10^{4}
	≥ 50	122	0.2	0.7	9.75×10^{-5}	8.4×10^{3}
	≥ 70	36	0.1	0.3	6.76×10^{-6}	1.3×10^{3}

Table 3.1: Mean Values of Δd_L, Ω, V_{loc} and $N_{gal,vol}$ parameters at different SNR selection.

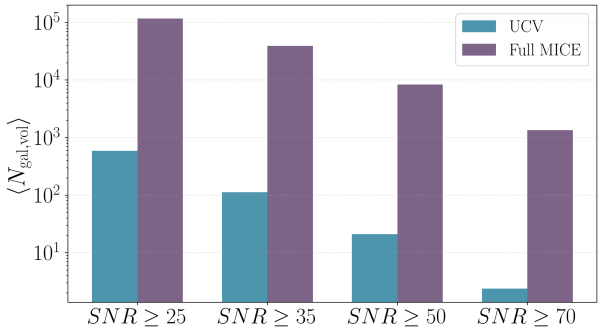


Figure 3.3.3: Mean number of galaxies within the localization volume, $\langle N_{\rm gal,vol} \rangle$, for different SNR thresholds. For each threshold, the UCV (*left bar*) and Full MICE (*right bar*) catalogs are shown side by side to facilitate comparison.

volume decreases by nearly an order of magnitude. At the most stringent threshold (SNR $\geq$ 70), the localization volume reaches $V_{\rm loc} \sim 10^{-6}$ Gpc^3, enclosing on average fewer than three galaxies and approaching the bright–siren regime. The Full MICE catalog exhibits qualitatively similar improvements, but due to its higher galaxy density, the mean number of galaxies per volume is of order 10^3 even in the highest SNR subset, precluding unique host identification.

This effect is clearly visible in Fig. 3.3.3, where the average value of $N_{\rm gal,vol}$ is plotted as a function of the applied SNR cuts for both catalogs. The comparison highlights that, while for UCV the mean number of galaxies decreases rapidly and approaches unity at the highest SNR threshold, the values for the Full MICE catalog remain systematically larger by nearly two orders of magnitude across the entire range. This systematic offset reflects the intrinsic difference in galaxy density between the two catalogs and illustrates why, even after stringent SNR cuts, unique host identification remains feasible only in the idealized UCV case.

Overall, the SNR sampling highlights two complementary aspects. On the one hand, higher thresholds systematically improve distance and sky localization, thereby affecting parameter constraints. On the other hand, stricter cuts reduce the sample size, while biasing the selection toward nearby and more massive systems, which limits how representative the sample is of the underlying astrophysical population.

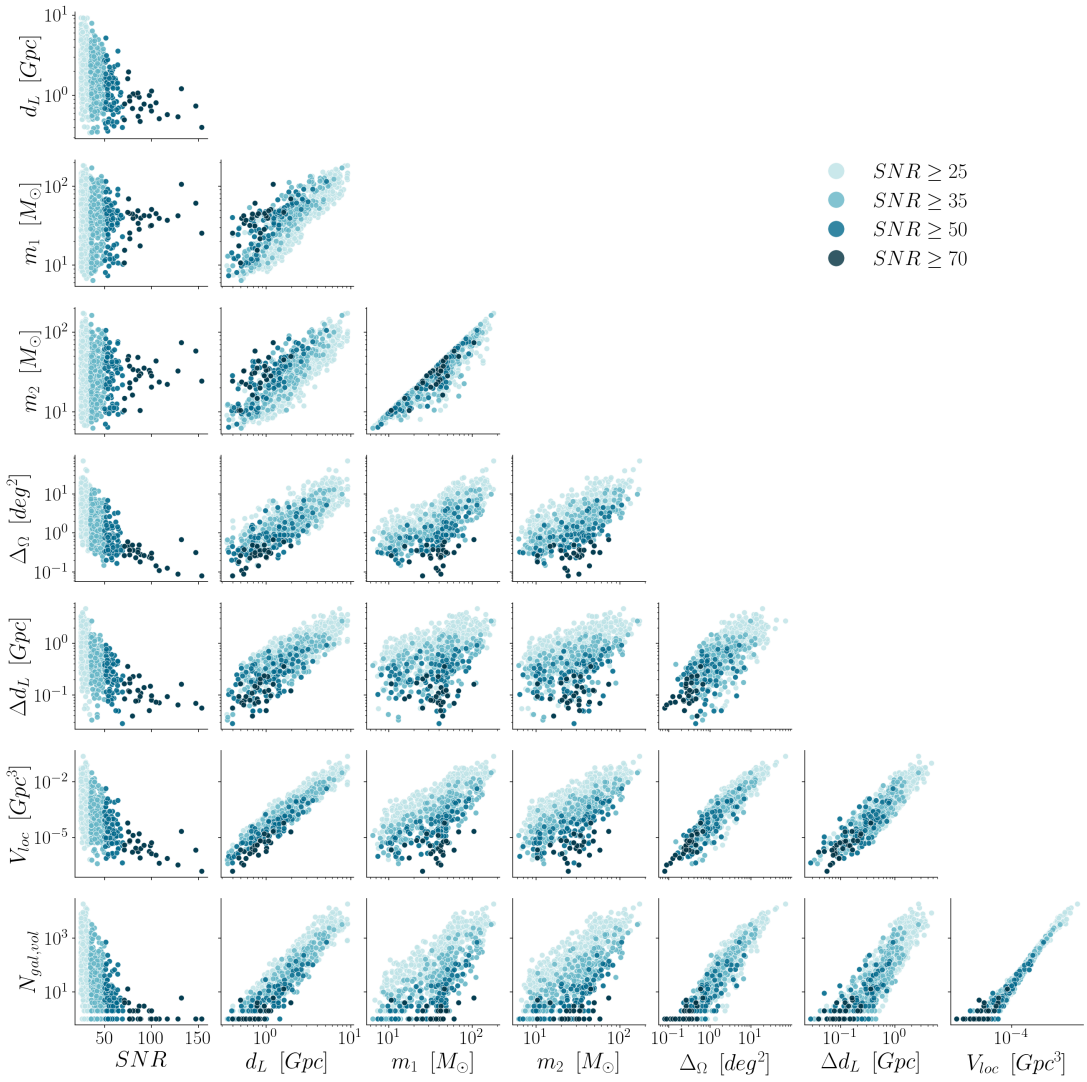


Figure 3.3.4: Corner plot showing all the pairwise scatter plots between parameters for *UCV catalog*, colored by different SNR thresholds: 25, 35, 50, and 70.

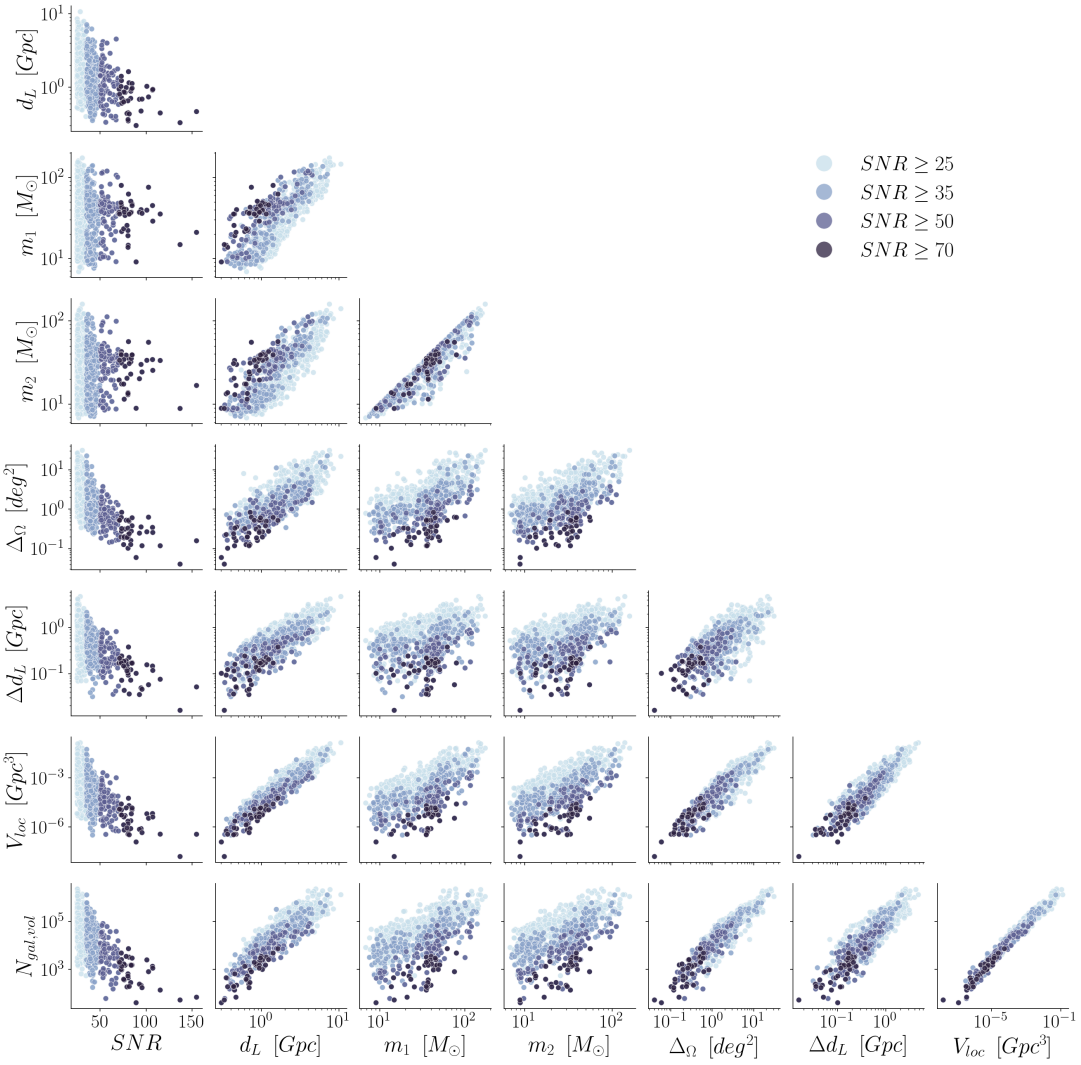


Figure 3.3.5: Corner plot showing all the pairwise scatter plots between parameters for *Full MICE catalog*, colored by different SNR thresholds: 25, 35, 50, and 70.

In what follows, both the UCV and Full MICE GW catalogs are used as starting point to generate different subsets of events by applying progressively stricter cuts on the SNR (SNR $\geq$ 25, 35, 50, and 70). The central goal is to determine which events contribute most to the extraction of cosmological and astrophysical information. These two aspects rely on different requirements. Astrophysical inference benefits from large and representative samples, covering the full distribution of binary parameters, so as to capture the global properties of the underlying population. Cosmological inference, instead, is primarily driven by events that are well localized in distance and sky position, since these provide the tightest constraints when matched with galaxy catalogs. Distinguishing between these complementary needs allows us to assess how event selection based on SNR impacts the balance between astrophysical and cosmological inference. In Chapter 4 we explore whether carefully selected subsets can provide constraints analogous to those from the entire catalog, yet with a significant reduction in computational cost. This second point is practical and particularly crucial for the analysis with future GW detectors. Analyzing a complete catalog makes full use of the information but requires substantial computational resources and wall time. Establishing that a strategically chosen subsample based on a clear and repeatable criterion, such as SNR, preserves most of the cosmological and astrophysical information could have a significant impact on the analyses of current and future datasets.

4

Results

After characterizing the GW and galaxy catalogs in terms of their statistical properties and correlations (Chapter 3), we now turn to the inference of cosmological and astrophysical population hyperparameters. The hierarchical Bayesian framework described in Chapter 2, and implemented through the `CHIMERA` pipeline, is here applied to the mock data sets presented in Sec. 3.1.

The inference is carried out on a representative set of parameters that describe both the cosmological model and the properties of the Binary Black Holes (BBH) population. The input GW catalog is analyzed using multiple SNR cuts, motivated by the analysis in Sec. 3.3, where we studied the correlation between the SNR and the main observables of the detected sources.

The introduction of SNR cuts serves a dual purpose: it allows us to quantify how much selection effects compromise the accuracy and precision of the inferred parameters, and, at the same time, it tests whether limiting the analysis to higher SNR detections can preserve most of the available information while reducing the computational cost. This perspective allows us to address a central question of the work: how much cosmological and astrophysical information is contained in the full event set, and whether a suitably

chosen SNR threshold can preserve accuracy without including a large number of low information events.

In the following, we first discuss the posterior distributions obtained for single and combined events, then present a set of targeted studies addressing the impact of SNR cuts on parameter accuracy, the computational scaling with the number of events, and the consequences of possible mismatches in the host–galaxy weighting schemes.

4.1 Astrophysical and Cosmological Constraints

In the context of hierarchical Bayesian inference, a one–dimensional (1D) posterior is obtained by fixing all hyperparameters to their fiducial values except one, which is varied over a grid. It represents first–order evaluations, providing a direct indication of how the data constrain each parameter individually. The full set of hyperparameters considered in this work was introduced in Chapter 2 and summarized in Tab. 2.1. In the present analysis, however, we restrict our attention to four parameters that capture the main cosmological and astrophysical information: the Hubble constant H_0, the mean μ_g and standard deviation σ_g of the Gaussian peak in the Black Hole (BH) mass distribution, and the peak redshift z_p of the merger rate evolution.

We distinguish between two complementary applications. First, we compute 1D posteriors for H_0 at the level of individual GW events. This step allows us to evaluate how informative single detections can be, and to investigate correlations between the sharpness of the posterior and basic event properties. By isolating the contribution of individual events, we aim to identify which ones carried the most statistical weight in the combined inference. Second, we move to the combined inference, where 1D posteriors are obtained by combining the contributions of multiple events. We focus on the four key hyperparameters (H_0, μ_g, σ_g, z_p) and apply various SNR thresholds to the GW catalogs (as presented in Sec. 3.3).

4.1.1 Results on Individual Events

We first evaluate the constraining power of single GW events by computing individul 1D posteriors for the Hubble constant H_0. The goal of this analysis is to investigate how constraining each single event is in the overall estimation of H_0.

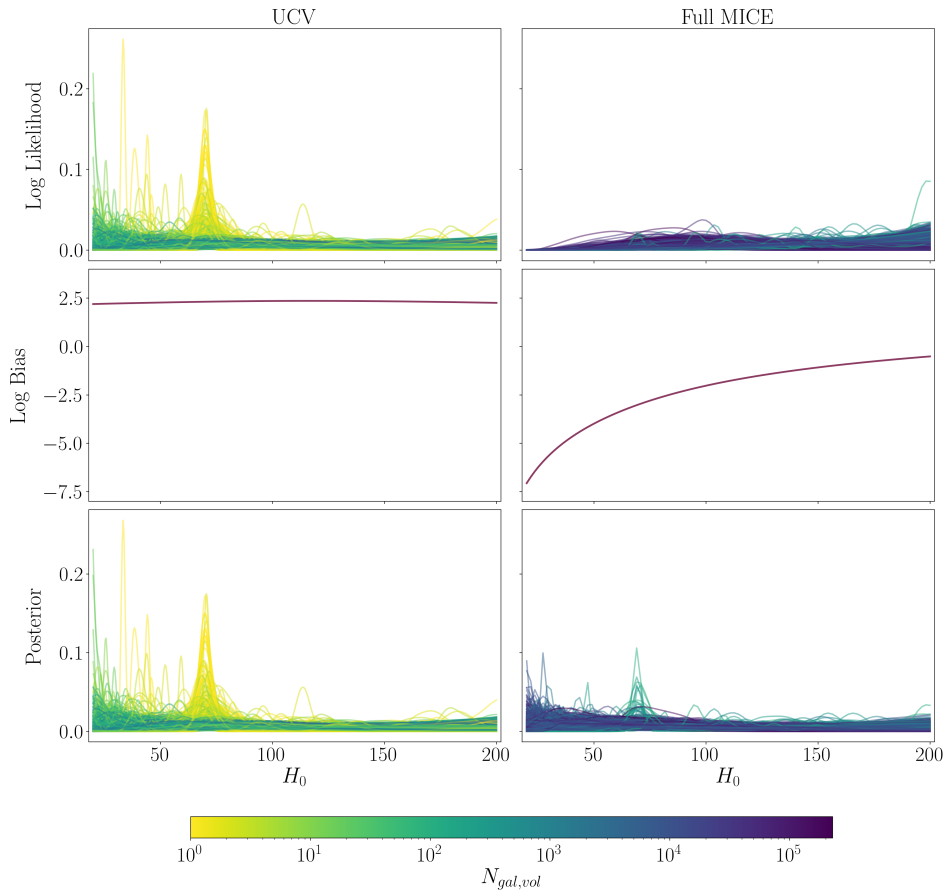


Figure 4.1.1: Log-likelihood (top panel), bias term (central panel), and posterior distribution (bottom panel) for the Hubble constant H_0 for individual GW events. Each curve is computed on a uniform grid spanning 20–200 and is colored by the number of galaxies within the localization volume $N_{\rm gal,vol}$

The results are shown in Fig. 4.1.1, where the three panels report, for each of the 1000 simulated detections of each catalog, the likelihood, the bias term, and the posterior distribution of H_0. Each curve is computed on a uniform grid spanning $H_0 \in [20, 200]$ and is colored by the number of galaxies within the localization volume, $N_{\rm gal,vol}$.

Starting from the likelihoods (top panel), the two catalogs display different behaviors. In UCV, several events fall in the *bright sirens* regime for which localization volumes contain only a few galaxies and allow association with a single host galaxy. In this case, the likelihoods are sharply peaked and centered close to the fiducial value $H_0 \simeq 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, already providing informative constraints. In contrast, no event in the

Full MICE catalog reaches the bright sirens regime, since even for the best localized events, the number of galaxies remains at least $N_{\rm gal,vol} \sim 10^2$–$10^3$. Consequently, the likelihoods are smooth, broad, and essentially featureless, reflecting the dilution of information over a much larger galaxy population.

The bias term (central panel) further highlights the contrast between the two cases. In UCV, it is nearly flat across the entire H_0 range, indicating that it has little effect on the final inference. In Full MICE, instead, the bias grows steadily with H_0, ranging from ~ -6 at low values to nearly zero at high values. These differences are directly mirrored in the posteriors (bottom panel). For UCV, the posteriors closely reproduce the likelihoods: sharp peaks are preserved, and the events that fall in the bright sirens regime remain strongly informative even after the bias correction is applied. In other words, when only a few galaxies are compatible with a GW localization, the likelihood already contains essentially all the constraining power, and the bias term is almost negligible.

In Full MICE, instead, the bias term becomes essential. Without it, the likelihoods are broad, smooth, and uninformative, reflecting the large number of potential host galaxies. Once the bias correction is introduced, however, the resulting posteriors sharpen significantly, with peaks emerging around the fiducial value $H_0 \simeq 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. This demonstrates that, in dense catalogs such as Full MICE, the bias acts as the key ingredient that compensates for the large number of possible hosts and allows individual events to contribute to the global inference.

4.1.2 Combined Analysis

After evaluating the individual contribution of GW events to the estimation of the Hubble constant, we extend the analysis to the combined GW event inference on four the four hyperparameters, H_0, μ_g, σ_g and z_p, under different SNR selections, specifically SNR $\geq 25,\ 35,\ 50,\ 70$.

The combined 1D posteriors for the four hyperparameters are shown in Fig. 4.1.2. Each row corresponds to one parameter, while the columns represent increasing thresholds in SNR. For each case, the posteriors obtained with the UCV and Full MICE catalogs are plotted together, allowing for a direct comparison of the two catalogs under the same selection.

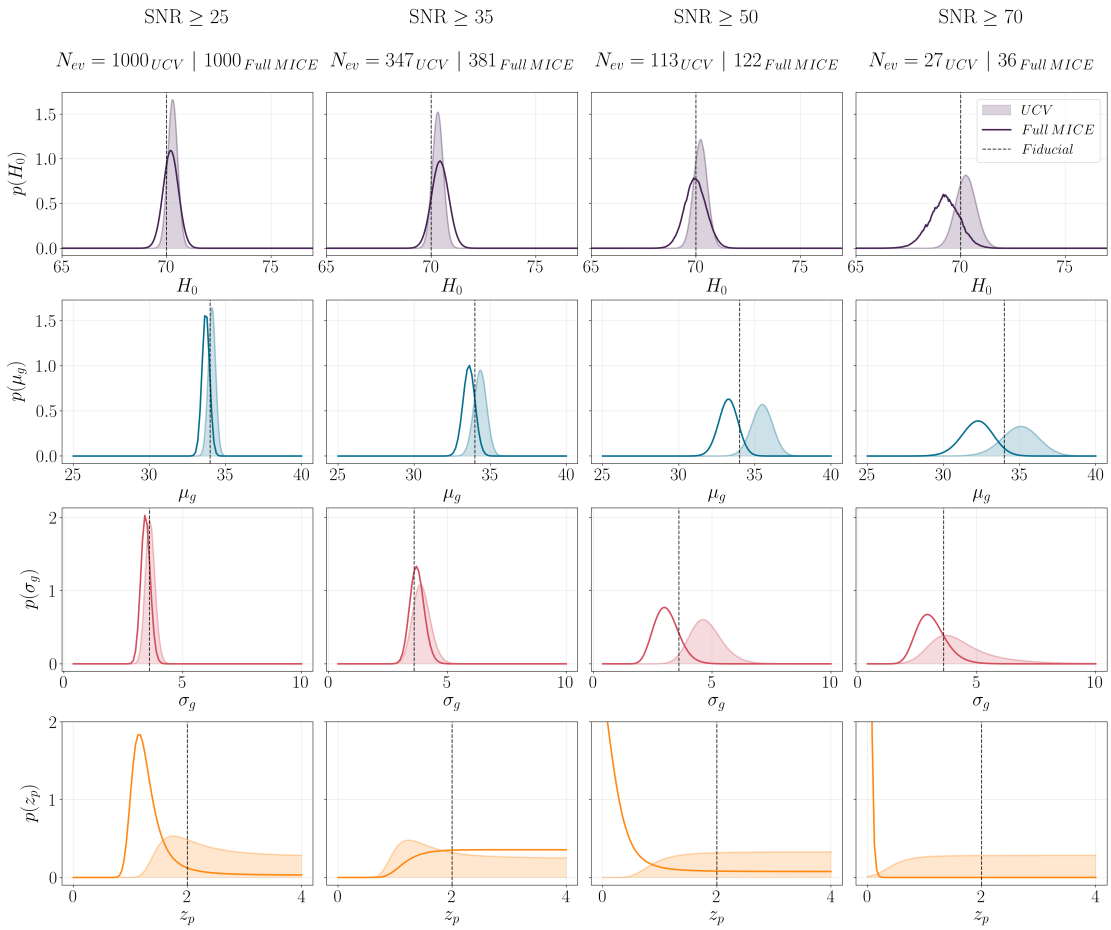


Figure 4.1.2: Grid of 1D posterior distributions for the four hyperparameters (H_0, μ_g, σ_g, z_p) obtained from the combined inference on GW events. Each row corresponds to a different parameter, while each column corresponds to a different SNR threshold. For each panel, posteriors from the two galaxy UCV (*shaded*) and Full MICE (*contour*) catalogs are shown for direct comparison.

A first general trend can be immediately noticed. When the SNR cut is progressively increased, the number of events contributing to the inference decreases, leading to broader and less peaked posteriors. It is interesting to note that the transition from SNR ≥ 25 to SNR ≥ 35 reduces the event sample size by about a factor of three, but the posteriors remain stable in both catalogs. This indicates that a moderate SNR cut can reduce the dataset size, and thus the computational cost, without significantly degrading the accuracy of the inference (further details in Sec. 4.2 and 4.3).

The behavior of the redshift evolution parameter z_p under different SNR selections highlights the limitations introduced by the restricted sampling of the source population. For the UCV catalog, the posterior distributions remain defined at all thresholds, but they become progressively broader and the peak shifts toward lower redshift values as the cut increases. This systematic trend is consistent with the expected effect of selecting only nearby, high-SNR events, which reduces sensitivity to the high-redshift portion of the merger population. In the case of the Full MICE catalog, the constraints degrade much more rapidly. For the samples with SNR ≥ 50 and 70, the posterior can only provide, as expected, a lower limit on z_p. As a result, information on the location of the merger rate peak is effectively inaccessible once the catalog is restricted to high SNR detections.

The mass distribution parameters μ_g and σ_g show a different trend. Since they trace the intrinsic BH mass spectrum rather than the redshift distribution, the main limitation is the reduced number of detections at higher thresholds. Their posteriors remain visible up to SNR ≥ 70, but progressively broaden and flatten, with σ_g in particular showing a significant increase of the associated uncertainty.
The Hubble constant H_0 is the most robust: both catalogs yield stable, narrow posteriors across all thresholds, with only a mild degradation at the most restrictive cuts.

In summary, the combined analysis highlights two complementary effects of SNR cuts. On the one hand, moderate thresholds can reduce the dataset size with little impact on the constraints, thereby improving computational efficiency. On the other hand, very aggressive cuts compromise the coverage of the population (as in the case of z_p) or strongly limit the available statistics (as for μ_g and σ_g), leading to loss of constraining power.

4.2 Accuracy Analysis

In addition to the visual inspection of the posterior grids, we extracted the 68% and 90% credible intervals for each parameter and reported them in the error-bar plots of Fig. 4.2.1. A quantitative summary of the 68% fractional uncertainties is provided in Tab. 4.1, where we report the ratio between the posterior width and the median value for each parameter and selection threshold. We also include the peak value of the posteriors and their bias with respect to the fiducial value. The behavior of the four parameters,

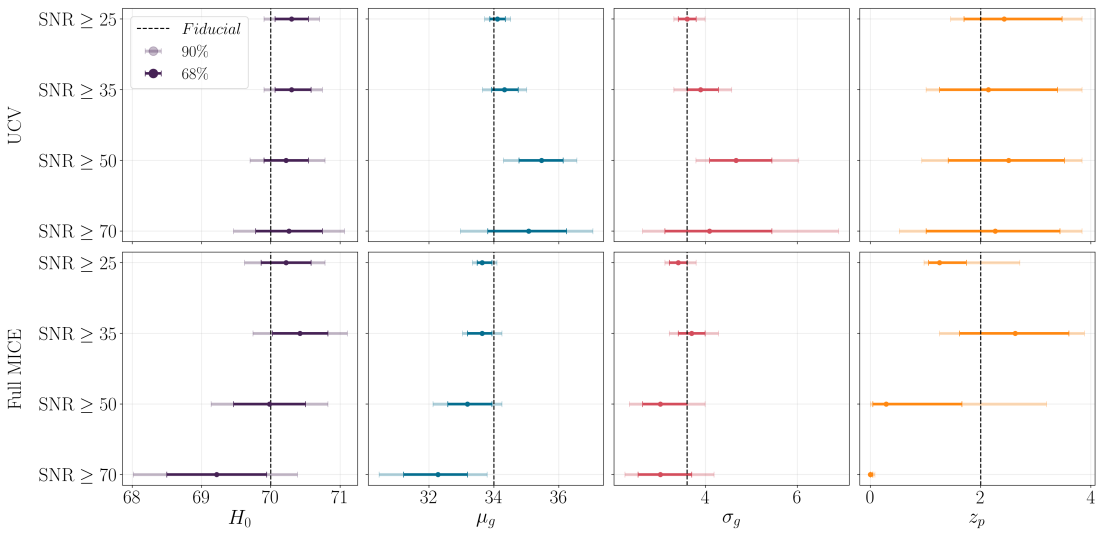


Figure 4.2.1: Bar plots displaying the 68% (*dark bars*) and 90% (*light bars*) credible intervals of the posterior distributions for each parameter. The fiducial value (*dotted line*) is indicated as a reference to assess potential biases.

considering the various SNR cuts, reveals trends in terms of both precision and bias with respect to the fiducial values ($H_0 = 70$, $\mu_g = 34$, $\sigma_g = 3.6$, $z_p = 2$). The trends highlighted by these results are discussed in the following.

H_0: Both catalogs yield stable constraints on the Hubble constant across all SNR cuts. For UCV the relative uncertainty increases only from 0.7% (SNR $\geq$ 25, $N_{\rm ev} = 1000$) to 1.4% (SNR$\geq$ 70, $N_{\rm ev} = 27$). The posterior peak remains essentially unbiased, with deviations of at most $\sim 0.4\%$ from the fiducial. Full MICE follows a similar trend, with uncertainties rising from 1.1% to 2.0% and peak values slightly shifted downward, with a bias of about 1% in the most extreme case (SNR $\geq$ 70). Overall, even when the number of events decreases by more than 95%, H_0 remains robustly constrained at the sub–percent (UCV) or percent (Full MICE) level, with negligible bias. This is a remarkable result, as the Full MICE catalog is much more realistic and densely populated, yet it still achieves constraints comparable to those obtained with UCV.

μ_g: The mean of the Gaussian peak in the BH mass distribution, μ_g, is also robust against SNR cutting. In UCV, inferred values remain clustered around the fiducial 34, with modest deviations of 3–4% and uncertainties growing from 1.4% to 6.6%

Catalog	SNR	$N_{\rm ev}$	H_0			μ_g			σ_g			z_p		
			med	*bias*%	Δ%	*med*	*bias*%	Δ%	*med*	*bias*%	Δ%	*med*	*bias*%	Δ%
UCV	≥ 25	1000	70.3	+0.4	0.7	34.1	+0.3	1.4	3.6	0.0	10.8	1.7	−13	73.3
	≥ 35	347	70.3	+0.4	0.7	34.4	+1.0	2.4	3.9	+8.1	19.9	1.3	$-37,5$	101.9
	≥ 50	113	70.2	+0.3	0.9	35.5	+4.4	3.9	4.6	+26.9	28.5	4.0	+100	88.5
	≥ 70	27	70.3	+0.4	1.4	35.1	+3.2	7.0	3.7	+2.8	54.6	4.0	+100	103.5
Full MICE	≥ 25	1000	70.1	+0.3	1.0	33.6	−1.1	1.4	3.4	−5.4	11.4	1.2	$-41,5$	51.6
	≥ 35	381	70.4	+0.6	1.1	33.6	−1.1	2.3	3.7	+2.7	15.7	3.2	+59.5	73.9
	≥ 50	122	69.9	−0.1	1.5	33.3	−2.0	4.1	3.0	−16.2	35.3	0.0	−100	-
	≥ 70	36	69.2	−1.2	2.1	32.3	−5.1	6.6	2.9	−18.9	38.6	0.0	−100	-

Table 4.1: Inference summary: medians, biases, and 1 uncertainty, for the four parameters H_0, μ_g, σ_g, and z_p as a function of the SNR threshold, reported for both catalogs (UCV and Full MICE). For each threshold, we also list the number of retained events. For every parameter the table provides: (*med*) the posterior median; (*bias* %) the fractional bias, defined as $100 \times (\text{med} - \text{fiducial})/\text{fiducial}$; and (Δ%) the fractional 1σ uncertainty, defined as $100 \times (\sigma_{68}/\text{med})$ where σ_{68} is the half-width of the 68% credible interval.

as the event count drops. Full MICE produces a similar pattern, with posteriors that broaden at higher thresholds and central estimates that fall slightly below the fiducial, the largest offset being about 5%. In this case, contrary to the H_0 constraints, which are driven by higher SNR well-localized events, precisely constraining the astrophysical population requires a larger sample of events. As expected, no significant differences emerge between the two catalogs, since the inference of population parameters is mainly governed by the overall number of detections.

σ_g: The width of the Gaussian peak, σ_g, shows a much stronger sensitivity to the number of chosen events. In UCV, the posterior peak remains close to the fiducial 3.6 at low thresholds, but then drifts upward, reaching 4.6 (with a bias of +27%) at SNR $\geq$ 50. At the highest cut, SNR $\geq$ 70, the peak falls back to 3.7 (+3%), but the fractional uncertainty grows steeply, from 11% at SNR$\geq$ 25 to more than 50% at SNR$\geq$ 70. In Full MICE, the trend is somewhat different: the degradation is accompanied by a systematic downward shift up to the highest cut where the peak value drops to ~ 2.9, nearly 18% below the fiducial.

z_p: The redshift–evolution parameter, z_p, is by far the most fragile of the set. In UCV, even at SNR$\geq$ 25 the fractional uncertainty already exceeds 70%, indicating that the parameter is only very weakly constrained. As the threshold increases, the

situation degrades: the posterior peak fluctuates widely, dropping to 1.3 (−38% bias) at SNR ≥ 35 and rising to 4.0 (+100% bias) at higher cuts. Although the posteriors remain technically defined, their widths are always larger than the signal, so the estimates carry little statistical weight and are strongly affected by the removal of high–redshift events. In Full MICE, the situation is even more severe. At SNR≥ 25 the uncertainty is already ∼ 50%, and by SNR≥ 35 the posterior peak has shifted to 3.2 (+60% bias). Beyond SNR≥ 50, the distributions effectively collapse to $z_p \simeq 0$, corresponding to a complete loss of constraining power. These results confirm that z_p is poorly measurable in both catalogs: even under the most inclusive selection, its uncertainty is too large to provide meaningful information, and discarding low–SNR events makes the estimates essentially unconstrained.

It is worth noting that, in some cases, the bar plots of Fig. 4.2.1 appear to show biases that fall outside the 1σ credible interval for some parameters. Such shifts can be explained as the manifestation of residual systematics inherent in the construction of the simulated catalogs. A first possible source of bias lies in the extraction of the catalogs themselves: here we analyze one realization of the simulated detections, but in principle the procedure should be repeated over many independent realizations (possibly also varying the cosmology) to fully capture statistical fluctuations. A second effect comes from the injections used to reproduce the population and recover selection effects. Since the true distribution is logarithmic in mass, with many low–mass and fewer high–mass binaries, a finite sample of injected events may not uniformly represent the underlying mass function, unless the injection strategy is specifically designed to do so. Finally, for each GW event the posterior samples used in the analysis are finite, and if they do not map the posterior surface densely enough, they can also introduce small biases in the recovered parameters. Taken together, these considerations suggest that the modest deviations seen in the plots do not point to fundamental issues, but rather reflect the limitations of working with finite realizations of simulated data.

To further explore the origin of the trends discussed above, we analyse how the relative parameter uncertainties scale with the average properties of the GW measurements in each subsample. For every set of events defined by an SNR threshold, we compute the mean luminosity distance error ($\langle \Delta d_L \rangle$), the mean sky–localization area ($\langle \Omega_{\rm loc} \rangle$), the mean localization volume ($\langle V_{\rm loc} \rangle$), and the mean number of galaxies within that volume ($\langle N_{\rm gal,vol} \rangle$). The fractional uncertainties on the inferred parameters are then plotted

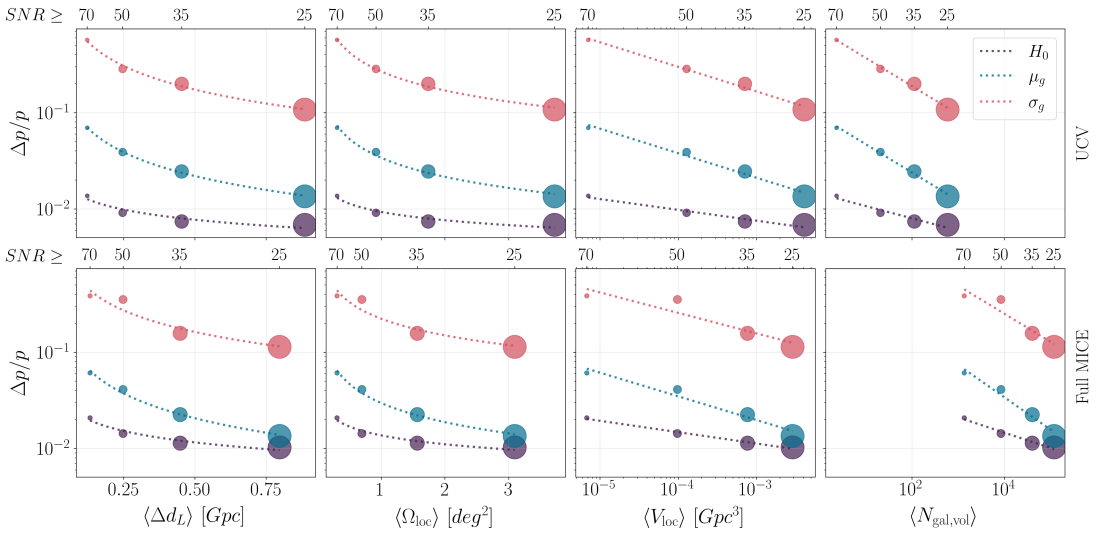


Figure 4.2.2: Fractional parameter uncertainties as a function of average GW measurement properties: luminosity distance uncertainty ($\langle \Delta d_L \rangle$), sky–localization area ($\langle \Omega_{\rm loc} \rangle$), localization volume ($\langle V_{\rm loc} \rangle$), and average number of galaxies per volume ($\langle N_{\rm gal,vol} \rangle$). Best–fit trends are shown for both catalogs, and the marker sizes are scaled to reflect the decreasing number of events at higher SNR thresholds. Note that the $\langle V_{\rm loc} \rangle$ and $\langle N_{\rm gal,vol} \rangle$ panels are plotted with log–log axes, whereas the $\langle \Delta d_L \rangle$ and $\langle \Omega_{\rm loc} \rangle$ panels use a linear x–axis. The apparent linear behavior in the last two panels is therefore a consequence of the plotting scale.

against these quantities, and linear relations are fitted to the resulting points. The z_p parameter is not included in this analysis, because of its large statistical uncertainty and unstable posterior peaks. From the fits we extract both the slope, which quantifies how quickly the uncertainty grows as the GW measurement becomes less accurate, and the normalization, which encodes the absolute level of accuracy achieved at each GW localization level.

The fitted relations, shown in Fig. 4.2.2 and summarized in Tab. 4.2, reveal two distinct regimes. When uncertainties are plotted against $\langle \Delta d_L \rangle$, $\langle \Omega_{\rm loc} \rangle$ or $\langle V_{\rm loc} \rangle$, the slopes and normalizations are very similar in UCV and Full MICE. This indicates that the scaling of uncertainties with GW localization accuracy is primarily governed by the SNR cut, through its effect on selecting overall closer events with a better localization, and on the number of retained detections, and is therefore independent of the galaxy catalog. To make this clearer, the points are plotted with marker sizes proportional to the number of events at each SNR threshold. This shows directly that the slope of the relation is linked

Catalog	Mean val.	H_0		μ_g		σ_g	
		slope (m)	*norm* (A)	*slope* (m)	*norm* (A)	*slope* (m)	*norm* (A)
UCV	$\langle \Delta d_L \rangle$	-0.35	0.006	-0.809	0.013	-0.798	0.100
	$\langle \Omega_{loc} \rangle$	-0.287	0.009	-0.648	0.034	-0.641	0.265
	$\langle V_{loc} \rangle$	-0.113	0.003	-0.252	0.004	-0.249	0.030
	$\langle N_{gal,vol} \rangle$	-0.128	0.014	-0.290	0.092	-0.286	0.704
Full MICE	$\langle \Delta d_L \rangle$	-0.415	0.009	-0.890	0.011	-0.778	0.095
	$\langle \Omega_{loc} \rangle$	-0.322	0.014	-0.688	0.030	-0.603	0.226
	$\langle V_{loc} \rangle$	-0.112	0.005	-0.258	0.003	-0.224	0.033
	$\langle N_{gal,vol} \rangle$	-0.235	0.100	-0.499	2.006	-0.435	8.806

Table 4.2: Best–fit slope (m) and normalization (A) parameters from the scaling relations of fractional uncertainties with GW measurement properties, for both the UCV and Full MICE catalogs.

to sample size: as the SNR cut becomes more restrictive, both catalogs lose events at a similar rate, and the reduced statistics broaden the posteriors in the same way.

A different behavior emerges when using the mean number of galaxies per localization volume, $\langle N_{\rm gal,vol} \rangle$. In this case, Full MICE shows both steeper slopes and significantly larger normalizations compared to UCV. This is due to a higher density of potential hosts which not only shifts the uncertainty level upward, but produces a steeper trend of uncertainty as $\langle N_{\rm gal,vol} \rangle$ increases.

4.3 Computational Time Analysis

To quantify the computational cost of the hierarchical Bayesian inference, we measure the runtime required to evaluate 1D posteriors under different SNR thresholds. The analysis is performed on the posterior of H_0, chosen because it remains tightly constrained even at high SNR thresholds. Focusing on this parameter provides a stable reference to evaluate how computation time scales with the number of events, and allows us to later compare these gains in efficiency with the corresponding loss in accuracy, as discussed in the trade–off analysis of Sec. 4.5.

The results in Fig. 4.3.1 and Tab. 4.3 are presented in a *fractional* form: we report the ratio between the runtime at a given SNR cut and the runtime at the baseline selection SNR ≥ 25. This quantity can be interpreted as the inverse of a fractional gain in efficiency, with values closer to zero corresponding to shorter execution times.

Despite the large difference in size between the two catalogs, the computational times turn out to be very similar. This is explained by the structure of the pipeline: all galaxy–related

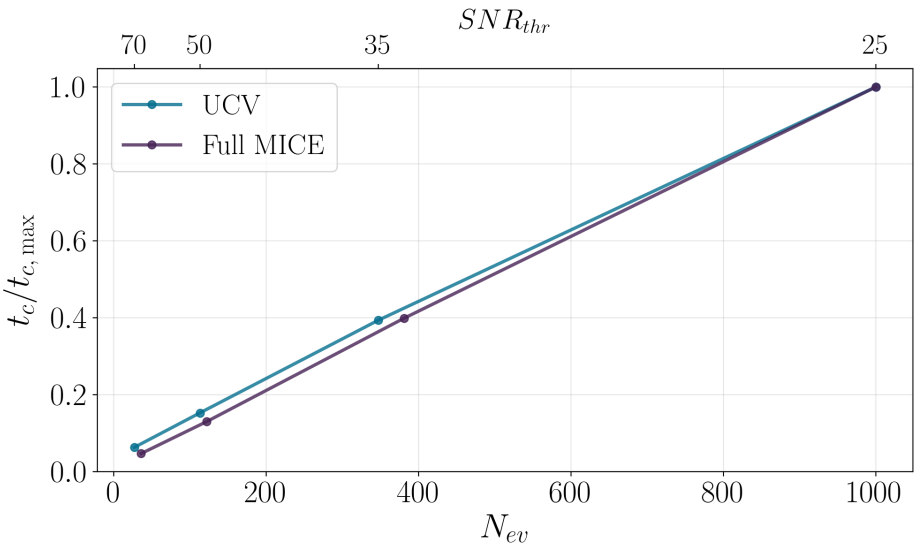


Figure 4.3.1: Relative computational time for the 1D posterior analyses as a function of the number of events included after different SNR cuts. Times are normalized to the baseline case $\mathrm{SNR}_{\mathrm{thr}} \geq 25$, such that the curves represent the fractional gain with respect to the maximum runtime. Results are shown for both catalogs: UCV (*dark purple*) and Full MICE (*light blue*).

Catalog	SNR	$N_{\rm ev}$	$t_c/t_{c,max}$
UCV	≥ 25	1000	1.00
	≥ 35	347	0.39
	≥ 50	113	0.15
	≥ 70	27	0.06
Full MICE	≥ 25	1000	1.00
	≥ 35	381	0.40
	≥ 50	122	0.13
	≥ 70	36	0.05

Table 4.3: Fractional 1D runtimes as a function of the SNR threshold for the two catalogs (UCV and Full MICE). For each cut, we list the number of retained events $N_{\rm ev}$ and the relative wall time $t_c/t_{c,\rm max}$, normalized to the baseline selection $\mathrm{SNR}_{\rm thr} \geq 25$. Values below unity indicate faster runs.

operations, such as pixelization, completeness corrections, and the interpolation of the background distribution, are performed once in a pre-processing stage and stored. As a result, the runtime during inference is driven almost entirely by the number of GW events ($N_{\rm ev}$), rather than by the size of the galaxy catalog.

The trend is clearly visible in Fig. 4.3.1: already at SNR ≥ 35 the computational cost drops to $\sim 40\%$ with respect to the baseline. At SNR ≥ 70, run times correspond to only $\sim 5\%$–6% of the initial.

The analysis of runtimes for the 1D posterior of H_0 can also be used as a reference to forecast the performance of future detectors such as ET, which is expected to deliver samples of order 10^5 events (Branchesi et al. (2023)). We perform an extrapolation of the measured times to larger event counts to obtain estimates of computational costs as the number of GW events increases. Our extrapolation is therefore explicitly constructed to bridge that gap: starting from the $N_{\rm ev} = 1000$ case and using the fitted slope, we evaluate the model at $N_{\rm ev} = 10^5$.

The scaling of the time with the number of events is modeled with a simple linear relation which model the wall time T as a function of the number of processed events $N_{\rm ev}$:

$$T(N_{\rm ev}) \;\simeq\; c\, N_{\rm ev}. \tag{4.1}$$

A small departure from perfect linearity is visible at the lowest $N_{\rm ev}$. This is expected

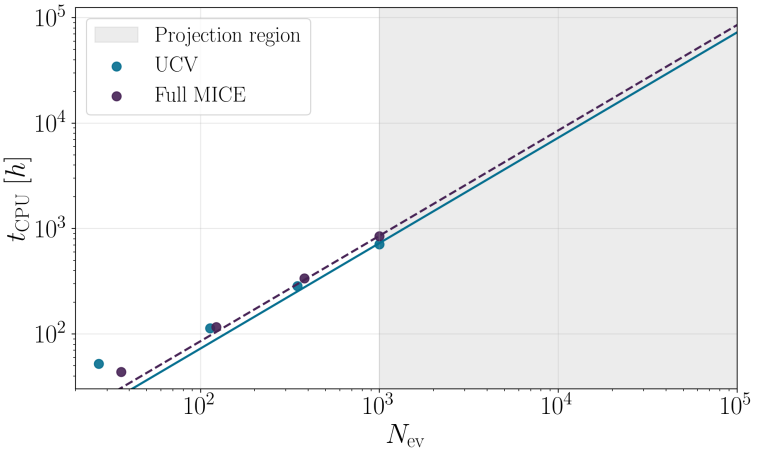


Figure 4.3.2: Projected MCMC CPU time (single core) as a function of the number of events. Points show the measured timings at $\mathrm{SNR}_{\mathrm{thr}} \in \{25, 35, 50, 70\}$ for the two catalogs, converted to MCMC–equivalent hours by rescaling the 1D timings. Solid and dashed curves are least–squares fits constrained through the origin, estimated separately for each catalog and used to extrapolate up to $M = 10^5$.

because the fit does not include an intercept, so fixed independent setup times produce a residual that is only visible when N_{ev} is small and becomes negligible at larger sizes. With this caveat in mind, the linear model captures the trend very well once N_{ev} is higher, with excellent agreement for both catalogs, confirming that the total runtime is almost entirely determined by the number of GW events.

To extend the forecast from 1D timings to a full multi-parameter Monte Carlo Markov Chain (MCMC), we apply a single multiplicative conversion that maps the iteration budget used in the 1Ds to the workload required by a reliable MCMC run. In our 1D analysis we evaluated a single-parameter posterior with 500 steps. For a realistic MCMC configuration we assume 20 000 steps per walker and 50 walkers. This leads to the conversion factor

$$F_{\mathrm{MCMC}} \;=\; \frac{20\,000}{500} \times 50 \;\approx\; 2\,000. \tag{4.2}$$

The resulting MCMC times should be read as order-of-magnitude forecasts at the ET scale, calibrated on a single-CPU run. Actual times depend on coding and analysis choices, particularly the iteration budget (number of steps and walkers) and the available computational resources (GPUs or multi-CPU parallelism), which mainly act by rescaling

the overall normalization of the timing curves.

Given this setup, and with the understanding that the calculation provides an indicative estimate rather than a precise timing, evaluating the linear fit at $N_{\rm ev} = 10^5$ and then translating the 1D timings into MCMC terms yields total runtimes of order 8×10^4 hours in the baseline configuration without additional cuts (SNR ≥ 25): specifically, we obtain $\sim 7.7 \times 10^4$ hours for UCV and $\sim 8.5 \times 10^4$ hours for Full MICE. Raising the threshold to SNR ≥ 35 reduces the expected cost to $\sim 3 \times 10^4$ hours, and a stricter selection at SNR ≥ 50 brings runtimes down to around 10^4 hours, that is about one order of magnitude below the no-cut configuration. Pushing the threshold to SNR ≥ 70, we reach a computational time of $\sim$ (2.3–2.6) $\times 10^3$ hours, approximately 3% of the baseline, corresponding to a speed-up of the order of thirty to forty times. These results are summarized by the projection curves in Fig. 4.3.2.

In summary, this analysis shows that computational time scales almost linearly with the number of events. Applying selection cuts that reduce the event sample allows to control wall time, with increasingly stringent thresholds yielding substantial time savings. In Sec. 4.5, we assess the trade-off between the loss of statistical accuracy induced by discarding low–SNR events and the gain in computational efficiency, with the goal of identifying where the reduction in cost outweighs the loss in precision.

4.4 Assessing Systematic effects

A final test was performed to assess the impact of mismodeling the dependence of merger rates on host galaxy stellar mass. Throughout the analysis, the Full MICE GW events were generated assuming the linear weighting proportional to the stellar mass: $p \propto M_\star$.

To evaluate the robustness of this assumption, we repeat the analysis using wrong background galaxy distributions with different weighting prescriptions: a uniform distribution ($p \propto M_\star^0$), where all galaxies are equally weighted, and a quadratic distribution ($p \propto M_\star^2$), where massive galaxies are preferentially selected. This wrongly assumed weighting affects the analysis since it enters into the redshift distribution of the galaxies, which is used to compute the completeness function and the background distribution (Eq. (2.26)). In all cases, the GW catalog used for inference remains fixed with the one extracted with the fiducial $p \propto M_\star$ weighting. This setup allows us to quantify the biases that arise if the true mass dependence of merger rates deviates from the linear

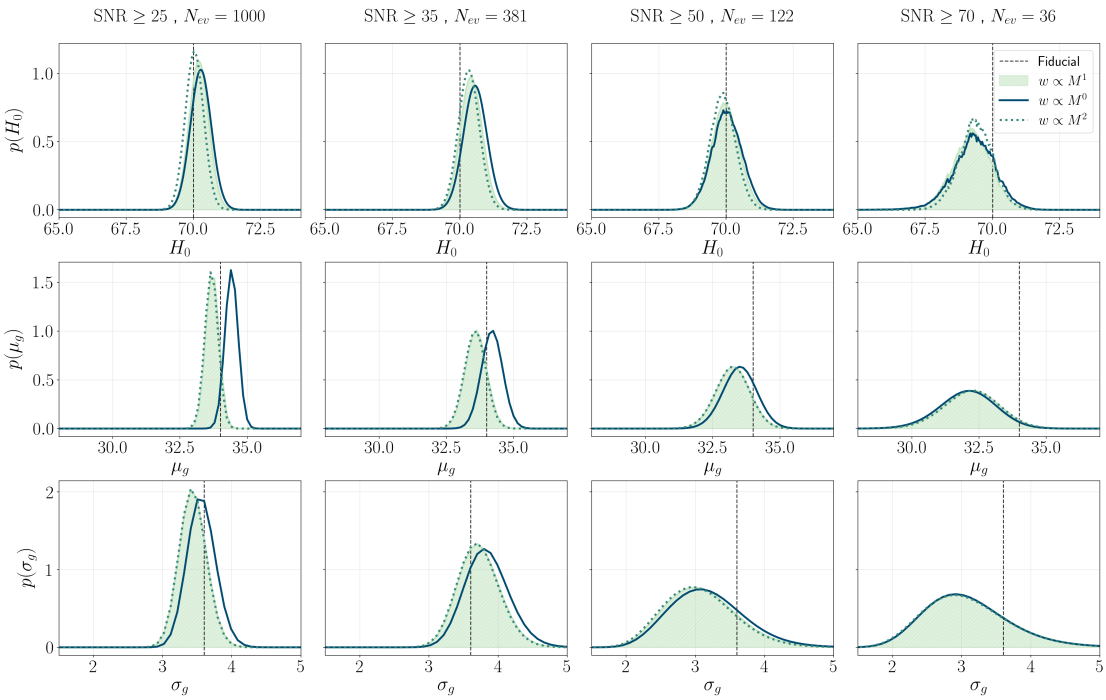


Figure 4.4.1: Posterior distributions obtained using the same mock GW events while varying the host–galaxy weighting. The fiducial weighting ($p \propto M_\star$) is contrasted with two mismatched "truth" scenarios: uniform ($M_\star^0$) and quadratic ($M_\star^2$). In all cases, the inference is performed with the fiducial weighting ($p \propto M_\star$).

model assumed in the analysis. The results are shown in Fig. 4.4.1, where posterior distributions obtained from the "true" fiducial linear mock are compared with those derived from the wrong uniform and quadratic mocks. The comparison shows that some parameters remain robust against weighting mismatches, while others exhibit significant biases or loss of constraining power. In particular, the uniform case tends to dilute the information by giving equal weight to all galaxies, whereas the quadratic case emphasizes massive galaxies.

The fiducial linear weighting ($\propto M_\star$) yields sub–percent uncertainties at low thresholds (0.97% at $\mathrm{SNR_{thr}} \geq 25$) and grows to about 2.1% at $\mathrm{SNR_{thr}} \geq 70$. The uniform case ($\propto M_\star^0$) systematically degrades the precision, with uncertainties from 1.08% at baseline to 2.2% at the highest cut. In contrast, the quadratic weighting ($\propto M_\star^2$) delivers the tightest constraints, from 0.97% at baseline to 1.73% at $\mathrm{SNR_{thr}} \geq 70$. Importantly, the bias on H_0 remains negligible in all cases, always $\sim 1\%$ or below, relative to the fiducial

Weighting	SNR	H_0			μ_g			σ_g		
		med	*bias*%	Δ%	*med*	*bias*%	Δ%	*med*	*bias*%	Δ%
$\propto M^0$	≥ 25	70.3	+0.4	1.1	34.4	+1.2	1.3	3.5	−2.7	10.8
	≥ 35	70.6	+0.8	1.2	33.2	+0.7	2.2	3.8	+5.4	17.8
	≥ 50	69.9	−0.1	1.5	33.5	−1.5	3.6	3.1	−13.5	34.2
	≥ 70	69.3	−1.1	2.2	32.1	−5.5	6.1	2.9	−18.9	41.8
$\propto M$ (*true*)	≥ 25	70.1	+0.3	1.0	33.6	−1.1	1.4	3.4	−5.4	11.4
	≥ 35	70.4	+0.6	1.1	33.6	−1.1	2.3	3.7	+2.7	15.7
	≥ 50	69.9	−0.1	1.5	33.3	−2.0	4.1	3.0	−16.2	35.3
	≥ 70	69.2	−1.2	2.1	32.3	−5.1	6.6	2.9	−18.9	38.6
$\propto M^2$	≥ 25	70.0	+0.0	1.0	33.6	−1.1	1.4	3.4	−5.4	11.4
	≥ 35	70.3	+0.5	1.1	33.6	−1.1	2.3	3.7	+2.7	15.7
	≥ 50	69.9	−0.1	1.3	33.2	−2.4	4.1	2.9	−18.9	35.3
	≥ 70	69.3	−1.0	1.7	32.3	−5.1	6.6	2.9	−18.9	41.8

Table 4.4: Posterior summaries for the three parameters H_0, μ_g, and σ_g as a function of the SNR threshold and of the host–galaxy weighting adopted (uniform $M_\star^0$, linear $M_\star$, quadratic $M_\star^2$). The label "true" marks the matching with the weighting used to generate the mock GW catalogs. For each parameter we report: *med*, the posterior central value (median); *bias* %, defined as $100 \times$ (med − fiducial)/fiducial; and Δ%, the fractional 1σ uncertainty, computed as $100 \times (\sigma_{68}/\text{med})$, where σ_{68} denotes the half–width of the 68% credible interval.

value of 70.

This improvement can be traced back to the fact that giving stronger weight to massive galaxies reduces the effective number of candidate hosts. As a result, the catalog distribution becomes less diluted by the large population of low–mass galaxies, and the inference gains precision without introducing detectable biases. Conversely, the uniform weighting tends to broaden the posteriors, since low–mass galaxies are over–represented, thereby weakening the statistical contrast between true hosts and background galaxies.

The situation is different for the population parameters. For μ_g, uncertainties lie between $\sim 1.3\%$ and $\sim 6.6\%$ depending on the SNR cut, with modest differences between weighting schemes. Biases are generally small at low thresholds ($\lesssim 1\%$ at SNR ≥ 25) but grow to $\sim -5\%$ at SNR ≥ 70, indicating a mild systematic drift towards lower values as the event sample shrinks. For σ_g, the spread is larger: fractional uncertainties increase from $\sim 11\%$ at baseline to 38–42% at the highest thresholds, and the posterior peaks exhibit biases up to $\sim -19\%$ relative to the fiducial 3.6.

Taken together, these tests indicate that the analysis is robust to moderate mismodeling of the host–galaxy weighting: biases remain negligible, yet the choice of weighting does control statistical efficiency, particularly for H_0 parameter. The scheme that upweights massive hosts ($\propto M_\star^2$) delivers the tightest posteriors even though it does not match the linear prescription used to generate the GW mocks. We interpret this behavior as the concentration of probability onto a smaller subset of high–mass galaxies, effectively transforming an almost uniform candidate catalog into one with a few prominent peaks. This concentration reduces host–association ambiguity and increases the information carried by each event, thereby strengthening cosmological and astrophysical constraints.

4.5 Trade-off Analysis

The previous sections have shown separately how *accuracy* of the inferred hyperparameters and *computational time* scale with the choice of SNR threshold, and the consequent reduction of the number of events.

From the accuracy side, we have seen that H_0 remains robust up to very aggressive cuts, with uncertainties staying at the sub–percent or percent level even when the number of events is reduced by more than an order of magnitude. The mass–distribution parameters show a moderate degradation: μ_g is relatively stable, while σ_g broadens substantially at higher thresholds. The redshift evolution parameter z_p is the most fragile, retaining only weak constraints in UCV and becoming essentially unconstrained in Full MICE beyond SNR$\geq$ 50. This confirms that H_0 is the most robust parameter, while population parameters are more sensitive to the size and coverage of the sample.

From the computational side, the cost of the inference scales nearly linearly with the number of GW events and is independent of the galaxy catalog. Moderate cuts such as SNR$\geq$ 35 already reduce the runtime by a factor of ~ 2.6, while more aggressive cuts produce strong speed–ups, reaching $\sim 20\times$ at SNR$\geq$ 70. These results were obtained from the H_0 1D computation, but are representative for all parameters since the scaling is controlled almost entirely by the number of events.

We now combine these two perspectives to identify the most effective trade–off between retaining cosmological information and reducing runtime. In what follows, the accuracy-time trade-off is quantified using the uncertainty and bias of H_0, since H_0 is the most robustly constrained parameter across all thresholds. Table 4.5 summarizes the combined

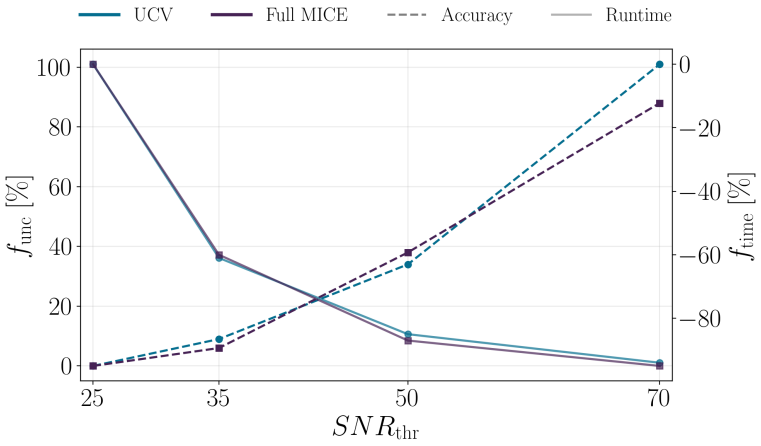


Figure 4.5.1: Effect of the SNR threshold on H_0 accuracy (dashed) and computational time (solid) for the *UCV* and *Full MICE* catalogs. Both curves show percentage changes relative to the baseline case $\mathrm{SNR}_{\rm thr} \geq 25$: positive values indicate increased H_0 uncertainty (loss of accuracy), while negative values indicate reduced runtime (gain in efficiency). This representation highlights the balance between information loss on H_0 and computational savings at different thresholds.

trends, while Fig. 4.5.1 provides a visual representation. A clear pattern emerges: at SNR $\geq$ 35, the number of events is reduced to about one–third of the baseline (347 out of 1000 for UCV, 381 for Full MICE). This cut shortens the runtime by about 60%, while the uncertainty on H_0 grows only modestly, by $\sim$ 6–9%. Moving to SNR $\geq$ 50 reduces the runtime by another factor $\sim$ 2, but at the cost of a +34–38% increase in H_0 uncertainty. At the most extreme cut, SNR$\geq$ 70, the runtime shrinks by more than 90%, while uncertainties nearly double.

Taken together, these results show that the optimal balance is achieved at SNR$\geq$ 35. In this regime the catalog size is reduced by roughly two–thirds, bringing substantial computational savings while preserving the relevant cosmological information. This compromise is particularly appealing in the context of third–generation detectors such as the ET, where analyzing the entire set of $\sim 10^5$ expected events would require larger computational resources and times. In such scenarios, a moderate SNR cut can ensure that the analysis remains both computationally feasible and scientifically robust.

A further point concerns the comparison between the two galaxy catalogs. Despite Full MICE being much larger in galaxy count (about two orders of magnitude more objects

Catalog	SNR	$N_{\rm ev}$	$f_{unc,\,H_0}$	f_{t_c}
UCV	≥ 25	1000	0 %	0%
	≥ 35	347	+9%	−61%
	≥ 50	113	+34%	−85%
	≥ 70	27	+101%	−94%
Full MICE	≥ 25	1000	0%	0%
	≥ 35	381	+6%	−60%
	≥ 50	122	+38%	−87%
	≥ 70	36	+88%	−95%

Table 4.5: Trade-off analysis between accuracy and computational cost, expressed as percent changes with respect to the baseline case $\mathrm{SNR}_{\rm thr} \geq 25$. Positive values indicate increased uncertainty (worse accuracy), while negative values indicate reduced computational time (improved efficiency).

than UCV), the H_0 constraints are remarkably similar: they remain unbiased and at the percent level in both cases. This behavior is explained by the host–galaxy weighting adopted in Full MICE catalog, proportional to the stellar mass. By concentrating probability on a smaller subset of massive galaxies, the weighting effectively turns a uniform catalog of possible hosts into one with a few high–probability peaks. In other words, even though Full MICE contains many more galaxies, the mass weighting prevents a dilution of the inference and leads to accuracy that is comparable to UCV. This outcome is far from trivial: without an informative weighting, one would expect the larger catalog to degrade constraints simply by adding many low–probability candidates.

5

Conclusions

The rapid progress of observational cosmology has led to increasingly precise measurements of the Hubble constant, but also to the emergence of a persistent discrepancy between determinations obtained in the local and early Universe. This so called Hubble tension opens to the possibility of new physics beyond the standard ΛCDM framework (Verde, Treu and Riess (2019); Abdalla et al. (2022)). In this context, gravitational waves represent a key independent cosmological probe. Their role is expected to become increasingly central with the development of new generations of detectors, which will achieve unprecedented sensitivities and thereby enable tighter constraints on cosmological parameters. At the same time, these instruments will deliver a dramatic increase in the number of detections, with third-generation observatories such as the Einstein Telescope expected to observe up to 10^5 events (Branchesi et al. (2023); Abac et al. (2025)). In this perspective, it becomes crucial to understand how cosmological and astrophysical information is distributed across the detected population. While it is often assumed that precise sky localization is most relevant for cosmological inference, and that astrophysical population studies benefit mainly from large event volumes, in practice so far event selection has typically relied on naive SNR cuts, applied just to discard poorly measured events and retain only robust detections, to maximize the already low statistics available

at the moment. However, the impact of such selection on the cosmological constraining power of GW samples has never been systematically investigated. This work focuses on assessing how the selection of subsets of events can optimize the inference of cosmological and astrophysical parameters, aiming to achieve robust constraints while simultaneously reducing the number of events and the associated computational cost.

The analysis is based on simulated dark siren catalogs extracted from the MICE simulation. Two cases are considered: a diluted catalog UCV, containing $\sim 1.6 \times 10^6$ massive galaxies (Borghi, Mancarella et al. (2024)), and a much denser *Full MICE* catalog with $\sim 3.3 \times 10^8$ galaxies (Borghi, Moresco et al. (2025)), offering a realistic representation of future surveys. Both assume a spectroscopic redshift precision of $\sigma_z = 0.001(1+z)$. The cosmological and astrophysical information is inferred within a hierarchical Bayesian framework, which naturally incorporates selection effects, enabling a robust estimation of cosmological parameters, the mass distribution of BBH, and the redshift evolution of the merger rate. The entire inference is implemented through `CHIMERA` (Borghi, Mancarella et al. (2024); Tagliazucchi et al. (2025); Borghi, Moresco et al. (2025)), a dedicated pipeline designed for efficient evaluation of high-dimensional parameter spaces and for the combined use of GW and galaxy catalogs. Before addressing parameter inference, a preliminary analysis of the simulated catalogs is performed. We study the correlations among the parameters and perform a PCA to provide a first insight into the dependencies among source properties and guide the interpretation of the results obtained from the full Bayesian inference.

The key element of the analysis are the detection thresholds. A sequence of progressively more selective cuts on the SNR is applied, with thresholds at 25, 35, 50, and 70. This strategy allows us to study how the accuracy of the inferred parameters and the computational cost of the analysis scale with the size and quality of the event sample. From this framework, several main results emerge:

- The difference of more than two orders of magnitude in number density of galaxies between the UCV and Full MICE catalog, has a direct impact on host association, with localization regions that in UCV may contain only few galaxies but that in Full MICE typically enclose hundreds to thousands. Applying SNR cuts further modifies these characteristics. The number of GW events decreases from $N_{\rm ev} = 1000$ at SNR ≥ 25 to only a few tens at SNR ≥ 70 (27 for UCV, 36 for Full MICE), while the sampled redshift range contracts from $z_{\rm max} \sim 1.4$ to $z_{\rm max} \lesssim 0.4$. In UCV, high–SNR events are

localized in volumes of the order of $\sim 10^{-6}\,\mathrm{Gpc}^3$, often containing a few galaxies and approaching the bright–siren regime. In Full MICE, although the same cuts reduce V_{loc} by similar factors, the higher galaxy density means that even the best–localized events still contain $\sim 10^3$ possible hosts, preventing unique identification.

- From the correlation analysis the SNR emerge as a dominant factor, showing correlations with all other observables. Selecting events with higher SNR naturally favors events that are better localized on the sky and associated with fewer potential host galaxies, thereby enhancing their cosmological constraining power. At the same time, however, SNR also correlates with luminosity distance and component masses, meaning that such a selection preferentially retains nearby systems and biases the sample toward a less representative subset in terms of mass and redshift distributions.

- The Principal Component Analysis (PCA) provides a complementary view of the parameter space. It shows that most of the variance is captured by two orthogonal components: one dominated by localization-related properties, such as localization area, localization volume, and number of galaxies in localization volume, and the other by signal strength through the SNR. Together, these two components explain more than 80% of the overall variance, confirming that localization and SNR are the two principal axes of variability that determine the information content of GW events.

- The combined inference over multiple events yields excellent results for the Hubble constant H_0, which is recovered with high precision in both UCV and Full MICE catalogs, with uncertainties that remain at the sub–percent to percent level even under the most restrictive SNR cuts. Remarkably, the two catalogs provide very similar uncertainties on the constrained parameters, despite the Full MICE catalog containing about 200 times more galaxies, thanks to the stellar–mass host weighting, which concentrates prior probability on massive galaxies, reducing the effective number of possible hosts. A similar scaling of uncertainties is observed for the astrophysical parameters describing the BBH population, confirming the consistency of the framework across very different galaxy densities. The sensitivity to population parameters, however, depends more strongly on the distribution of the selected events across the relevant ranges. While the mean of the BH mass distribution, μ_g, remains well

constrained, its width σ_g becomes increasingly uncertain when only a small number of events is retained. The redshift–evolution parameter z_p is the most affected: excluding low–SNR events preferentially excludes distant sources, leading to a rapid loss of information on the high–redshift tail of the merger population. These results indicate, as expected, that robust inference on population parameters requires a larger and more representative sample covering the full astrophysical parameter space, which is not guaranteed at high SNR cuts. By contrast, the inference of H_0 can rely on a smaller subset of well–localized events and still achieve percent–level precision.

- The analysis of computational time highlights the practical advantage of event selection. Thanks to the fact that galaxy terms are pre–computed, the runtime of the inference scales almost linearly with the number of events. Focusing on the posterior of H_0 as a stable reference, event selection shows substantial reductions in execution time: at SNR ≥ 35 it drops to $\sim 40\%$ of the baseline, reaching $\sim 5-6\%$ at SNR ≥ 70. Importantly, moderate thresholds achieve these savings with only minor loss of cosmological information. Extrapolating these trends to third–generation detectors with catalogs of order 10^5 events, shows that the difference between full catalogs and high–SNR selections corresponds to a reduction in MCMC runtimes from $\sim 10^5$ to $\sim 10^3$ hours. These results emphasize that efficient selection strategies will be crucial to balance computational efficiency with scientific accuracy in the era of third–generation GW detectors.

- The tests on wrong host–galaxy weighting schemes in the Full MICE catalog, performed to probe potential systematics, shows that cosmological constraints remain largely unaffected: H_0 is consistently recovered with negligible bias across all prescriptions. In contrast, population parameters display a higher sensitivity, with uncertainties growing and mild systematic shifts emerging, especially for σ_g. These results emphasize that, in realistic dense catalogs, an accurate treatment of host–galaxy weighting is essential to fully exploit the cosmological and astrophysical information carried by GW events.

- The trade–off analysis quantifies the balance between accuracy and computational efficiency. At SNR ≥ 35, the number of events is reduced by about two–thirds (from 1000 to ~ 350), reducing the runtime by $\sim 60\%$ while increasing the H_0 uncertainty

by less than 10%. At SNR ≥ 50, uncertainties grow by $\sim 35\%$ while runtime shrinks by more than 85%, and at SNR ≥ 70 the computational cost is reduced by $\sim 95\%$ but with uncertainties nearly doubling. This establishes SNR ≥ 35 as an effective compromise, preserving accuracy while delivering substantial efficiency gains. The trade–off analysis introduced here represents a crucial step in quantifying the balance between accuracy and computational cost, and it will be increasingly important given the large data volumes and demanding analyses expected from next–generation detectors.

5.1 Future Prospects

The results presented in this work demonstrate the potential of event selection strategies to balance accuracy and computational feasibility in gravitational–wave cosmology. At the same time, they highlight several directions in which the analysis can be extended and refined.

A natural next step is to replace the simplified 1D posterior analysis with a full MCMC inference, where all cosmological and astrophysical parameters are jointly varied and marginalized over. This would enable a complete assessment of parameter degeneracies and biases, and provide more realistic estimates of the constraining power of GW catalogs. In particular, this approach is required to fully exploit the correlations between cosmological parameters and population parameters, which were only partially captured in the present framework.

Another promising prospect is to explore alternative event–selection criteria beyond the SNR. The correlation analysis carried out in this thesis suggests that parameters such as the number of galaxies within the localization volume, $N_{\mathrm{gal,vol}}$, may provide a more robust way to explore the actual informativeness of individual events. Selecting a small set of "golden'' events with very few potential hosts could yield constraints comparable to much larger samples, while drastically reducing the computational cost. However, this strategy would imply the need to evaluate the full posterior consistently and to carefully account for selection effects and potential biases. A systematic investigation of such criteria would complement SNR cuts and provide a clearer picture of how to optimize GW datasets for cosmological inference.

A further natural extension will be the application of this framework to real data. The

LVK Collaboration has recently released version 4.0 of the Gravitational-Wave Transient Catalog (GWTC–4.0), containing more than an hundred of new significant GW events from the first part of the O4 observing run (O4a, May 2023 – January 2024). This dataset represents the most comprehensive sample of GW detections to date. It includes several particularly interesting events, such as the most massive binary black hole merger observed so far. Extending the present analysis to GWTC–4.0 will allow one to test the robustness of the methodology on real detections, validate the impact of event–selection strategies under realistic noise and catalog incompleteness, and understand how current samples can already inform cosmological and astrophysical population studies.

Finally, while the present work focused on spectroscopic redshift uncertainties, motivated by future large–scale surveys such as Wide-field Spectroscopic Telescope (WST) (Mainieri et al. (2024)) which will be developed in parallel to future GW detectors, it will be also important to assess the impact of photometric galaxy catalogs. In the near future, photometric surveys are expected to play a major role thanks to their wide coverage and depth, although the lower redshift precision compared to spectroscopic surveys may limit their constraining power. Repeating the analysis with photometric errors would provide valuable insight into how redshift precision affects cosmological and population inference, and would help clarify the trade–offs between completeness and accuracy in realistic survey scenarios.

Abbreviations

ΛCDM Λ Cold Dark Matter. 2, 83

CHIMERA Combined Hierarchical Inference Model for Electromagnetic and gRavitational Wave Analysis. vi, 19, 25, 31–33, 63, 84

BBH Binary Black Holes. 11, 13, 16, 30, 33, 39, 40, 63, 84, 85

BH Black Hole. 64, 68, 69, 85

BNS Binary Neutron Star. 11, 16, 26

CBC Compact Binary Coalescence. 1, 6, 11, 14, 15, 28

CDM Cold Dark Matter. 2

CMB Cosmic Microwave Background. 7

DM Dark Matter. 36

ET Einstein Telescope. 13, 17, 75, 76, 81

FIM Fisher Information Matrix. 40

FLRW Friedmann Lemaître Robertson Walker. 3, 4

GR General Relativity. 3, 8

GRB Gamma-Ray Burst. 11

GW Gravitational Waves. iv, vi, 1, 6–8, 10–12, 14–16, 19–22, 28, 31–33, 35, 39–41, 46, 47, 55, 56, 58, 60, 62–64, 66, 84

GWTC Gravitational-Wave Transient Catalog. 88

JIT Just-In-Time. 33

KAGRA Kamioka Gravitational Wave Detector. 13

KDE Kernel Density Estimate. 31, 32

LIGO Laser Interferometer Gravitational-Wave Observatory. 13

LISA Laser Interferometer Space Antenna. 11

LVK LIGO-Virgo-KAGRA. 13, 39

MCMC Monte Carlo Markov Chain. 76, 87

MICE MareNostrum Instituto de Ciències de l'Espacio. 36–39, 84

NSBH Neutron Star-Black Hole. 11, 16

O5 Observing run 5. 13, 39, 55

PCA Principal Component Analysis. 41, 49–51, 84, 85

PLP Power Law + Peak. 26, 39

PTA Pulsar Timing Arrays. 11

SNR Signal to Noise Ratio. 16, 36, 39–42, 50, 53–55, 61, 63, 83–85

UCV Uniform in Comoving Volume. 37, 38, 84

WST Wide-field Spectroscopic Telescope. 88

Bibliography

Abac, A. et al. (2025). *The Science of the Einstein Telescope*. arXiv: 2503.12263 [gr-qc]. URL: https://arxiv.org/abs/2503.12263.

Abbott, R. et al. (May 2021). *Population Properties of Compact Objects from the Second LIGO–Virgo Gravitational-Wave Transient Catalog*. DOI: 10.3847/2041-8213/abe949. URL: http://dx.doi.org/10.3847/2041-8213/abe949.

Abdalla, E. et al. (June 2022). 'Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies'. In: *Journal of High Energy Astrophysics* 34, pp. 49–211. ISSN: 2214-4048. DOI: 10.1016/j.jheap.2022.04.002. URL: http://dx.doi.org/10.1016/j.jheap.2022.04.002.

Aghanim, N. et al. (Sept. 2020). 'Planck2018 results: VI. Cosmological parameters'. In: *Astronomy & Astrophysics* 641, A6. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201833910. URL: http://dx.doi.org/10.1051/0004-6361/201833910.

Antelis, J. M. and C. Moreno (Jan. 2017). 'Obtaining gravitational waves from inspiral binary systems using LIGO data'. In: *The European Physical Journal Plus* 132(1). ISSN: 2190-5444. DOI: 10.1140/epjp/i2017-11283-5. URL: http://dx.doi.org/10.1140/epjp/i2017-11283-5.

Artale, M. C. et al. (May 2020). 'An astrophysically motivated ranking criterion for low-latency electromagnetic follow-up of gravitational wave events'. In: *Monthly Notices of the Royal Astronomical Society* 495(2), pp. 1841–1852. ISSN: 1365-2966. DOI: 10.1093/mnras/staa1252. URL: http://dx.doi.org/10.1093/mnras/staa1252.

Borghi, N., M. Mancarella et al. (2024). *Cosmology and Astrophysics with Standard Sirens and Galaxy Catalogs in View of Future Gravitational Wave Observations*. DOI: 10.48550/arXiv.2312.05302. arXiv: 2312.05302 [astro-ph.CO].

Borghi, N., M. Moresco et al. (2025). 'Echoes from the Dark: Galaxy Catalog Incompleteness in Standard Siren Cosmology'. [IN PRESS].

Bradbury, J. et al. (2018). *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. URL: http://github.com/jax-ml/jax.

Branchesi, M. et al. (July 2023). 'Science with the Einstein Telescope: a comparison of different designs'. In: *Journal of Cosmology and Astroparticle Physics* 2023(07), p. 068. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2023/07/068. URL: http://dx.doi.org/10.1088/1475-7516/2023/07/068.

Broekgaarden, F. S., S. Banagiri and E. Payne (2024). *Visualizing the Number of Existing and Future Gravitational-Wave Detections from Merging Double Compact Objects*. arXiv: 2303.17628 [astro-ph.HE]. URL: https://arxiv.org/abs/2303.17628.

Carretero, J. et al. (Dec. 2014). *An algorithm to build mock galaxy catalogues using MICE simulations*. DOI: 10.1093/mnras/stu2402. URL: http://dx.doi.org/10.1093/mnras/stu2402.

Christensen, N. and R. Meyer (Apr. 2022). 'Parameter estimation with gravitational waves'. In: *Reviews of Modern Physics* 94(2). ISSN: 1539-0756. DOI: 10.1103/revmodphys.94.025001. URL: http://dx.doi.org/10.1103/RevModPhys.94.025001.

Collaboration, T. L. S., the Virgo Collaboration and the KAGRA Collaboration (2025). *GWTC-4.0: Population Properties of Merging Compact Binaries*. arXiv: 2508.18083 [astro-ph.HE]. URL: https://arxiv.org/abs/2508.18083.

Crocce, M. et al. (Aug. 2015). *The MICE Grand Challenge lightcone simulation – II. Halo and galaxy catalogues*. DOI: 10.1093/mnras/stv1708. URL: http://dx.doi.org/10.1093/mnras/stv1708.

Farr, W. M. et al. (Oct. 2019). 'A Future Percent-level Measurement of the Hubble Expansion at Redshift 0.8 with Advanced LIGO'. In: *The Astrophysical Journal Letters* 883(2), p. L42. ISSN: 2041-8213. DOI: 10.3847/2041-8213/ab4284. URL: http://dx.doi.org/10.3847/2041-8213/ab4284.

Finke, A. et al. (Aug. 2021). 'Cosmology with LIGO/Virgo dark sirens: Hubble parameter and modified gravitational wave propagation'. In: *Journal of Cosmology and Astroparticle Physics* 2021(08), p. 026. DOI: 10.1088/1475-7516/2021/08/026. URL: https://dx.doi.org/10.1088/1475-7516/2021/08/026.

Fosalba, P., M. Crocce et al. (Mar. 2015). *The MICE grand challenge lightcone simulation – I. Dark matter clustering*. DOI: 10.1093/mnras/stv138. URL: http://dx.doi.org/10.1093/mnras/stv138.

Fosalba, P., E. Gaztanaga et al. (Dec. 2013). *The MICE Grand Challenge Lightcone Simulation III: Galaxy lensing mocks from all-sky lensing maps*. DOI: 10.1093/mnras/stu2464.

Gennari, V. et al. (2025). *Searching for additional structure and redshift evolution in the observed binary black hole population with a parametric time-dependent mass distribution*. arXiv: 2502.20445 [gr-qc]. URL: https://arxiv.org/abs/2502.20445.

Gray, R. et al. (Dec. 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. DOI: 10.1088/1475-7516/2023/12/023. URL: https://dx.doi.org/10.1088/1475-7516/2023/12/023.

Hotelling, H. (1933). *Analysis of a complex of statistical variables into principal components.* URL: https://api.semanticscholar.org/CorpusID:144828484.

Iacovelli, F. et al. (Oct. 2022). *GWFAST: A Fisher Information Matrix Python Code for Third-generation Gravitational-wave Detectors*. DOI: 10.3847/1538-4365/ac9129. URL: http://dx.doi.org/10.3847/1538-4365/ac9129.

Jolliffe, I. T. and J. Cadima (2016). *Principal component analysis: a review and recent developments*. DOI: 10.1098/rsta.2015.0202. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2015.0202. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2015.0202.

Kalaghatgi, C., M. Hannam and V. Raymond (May 2020). *Parameter estimation with a spinning multimode waveform model*. DOI: 10.1103/PhysRevD.101.103004. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.103004.

Laghi, D. et al. (Oct. 2021). 'Gravitational-wave cosmology with extreme mass-ratio inspirals'. In: *Monthly Notices of the Royal Astronomical Society* 508(3), pp. 4512–4531. ISSN: 0035-8711. DOI: 10.1093/mnras/stab2741. eprint: https://academic.oup.com/mnras/article-pdf/508/3/4512/40857606/stab2741.pdf. URL: https://doi.org/10.1093/mnras/stab2741.

London, L. et al. (Apr. 2018). *First Higher-Multipole Model of Gravitational Waves from Spinning and Coalescing Black-Hole Binaries*. DOI: 10.1103/PhysRevLett.120.161102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.161102.

Madau, P. and M. Dickinson (Aug. 2014). 'Cosmic Star-Formation History'. In: *Annual Review of Astronomy and Astrophysics* 52(1), pp. 415–486. ISSN: 1545-4282. DOI:

10.1146/annurev-astro-081811-125615. URL: http://dx.doi.org/10.1146/annurev-astro-081811-125615.

Maggiore, M. (2008). *Gravitational waves - Volume 1, Theory and experiments*. eng. Oxford University Press. ISBN: 0-19-857074-0.

Mainieri, V. et al. (2024). *The Wide-field Spectroscopic Telescope (WST) Science White Paper*. arXiv: 2403.05398 [astro-ph.IM]. URL: https://arxiv.org/abs/2403.05398.

Mancarella, M., E. Genoud-Prachex and M. Maggiore (Mar. 2022). 'Cosmology and modified gravitational wave propagation from binary black hole population models'. In: *Physical Review D* 105(6). ISSN: 2470-0029. DOI: 10.1103/physrevd.105.064030. URL: http://dx.doi.org/10.1103/PhysRevD.105.064030.

Mandel, I., W. M. Farr and J. R. Gair (Mar. 2019). 'Extracting distribution parameters from multiple uncertain observations with selection biases'. In: *Monthly Notices of the Royal Astronomical Society* 486(1), pp. 1086–1093. ISSN: 1365-2966. DOI: 10.1093/mnras/stz896. URL: http://dx.doi.org/10.1093/mnras/stz896.

Mastrogiovanni, S. et al. (Aug. 2023). 'Joint population and cosmological properties inference with gravitational waves standard sirens and galaxy surveys'. In: *Phys. Rev. D* 108 (4), p. 042002. DOI: 10.1103/PhysRevD.108.042002. URL: https://link.aps.org/doi/10.1103/PhysRevD.108.042002.

Moore, C. J., R. H. Cole and C. P. L. Berry (Jan. 2015). 'Gravitational-wave sensitivity curves'. In: *Classical and Quantum Gravity* 32(1), 015014, p. 015014. DOI: 10.1088/0264-9381/32/1/015014. arXiv: 1408.0740 [gr-qc].

Moresco, M. et al. (2022). *Unveiling the Universe with emerging cosmological probes*. DOI: 10.1007/s41114-022-00040-z. URL: http://dx.doi.org/10.1007/s41114-022-00040-z.

Pearson, K. (1901). *LIII. On lines and planes of closest fit to systems of points in space*. DOI: 10.1080/14786440109462720. eprint: https://doi.org/10.1080/14786440109462720. URL: https://doi.org/10.1080/14786440109462720.

Pedregosa, F. et al. (2018). *Scikit-learn: Machine Learning in Python*. arXiv: 1201.0490 [cs.LG]. URL: https://arxiv.org/abs/1201.0490.

Punturo, M. et al. (Apr. 2010). 'The third generation of gravitational wave observatories and their science reach'. In: *Classical and Quantum Gravity* 27(8), p. 084007. DOI: 10.1088/0264-9381/27/8/084007. URL: https://dx.doi.org/10.1088/0264-9381/27/8/084007.

Riess, A. G. et al. (July 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. ISSN: 2041-8213. DOI: 10.3847/2041-8213/ac5c5b. URL: http://dx.doi.org/10.3847/2041-8213/ac5c5b.

Roy, S. K., L. A. C. van Son and W. M. Farr (2025). *A Mid-Thirties Crisis: Dissecting the Properties of Gravitational Wave Sources Near the 35 Solar Mass Peak*. arXiv: 2507.01086 [astro-ph.HE]. URL: https://arxiv.org/abs/2507.01086.

Schöneberg, N. et al. (Nov. 2022). 'BAO+BBN revisited — growing the Hubble tension with a 0.7 km/s/Mpc constraint'. In: *Journal of Cosmology and Astroparticle Physics* 2022(11), p. 039. DOI: 10.1088/1475-7516/2022/11/039. URL: https://dx.doi.org/10.1088/1475-7516/2022/11/039.

Schutz, B. F. (Sept. 1986). *Determining the Hubble constant from gravitational wave observations*. DOI: 10.1038/323310a0.

Tagliazucchi, M. et al. (2025). *Accelerating the Standard Siren Method: Improved Constraints on Modified Gravitational Wave Propagation with Future Data*. arXiv: 2504.02034 [astro-ph.CO].

Talbot, C., R. Smith et al. (Aug. 2019). 'Parallelized inference for gravitational-wave astronomy'. In: *Physical Review D* 100(4). ISSN: 2470-0029. DOI: 10.1103/physrevd.100.043030. URL: http://dx.doi.org/10.1103/PhysRevD.100.043030.

Talbot, C. and E. Thrane (Apr. 2018). *Measuring the Binary Black Hole Mass Spectrum with an Astrophysically Motivated Parameterization*. DOI: 10.3847/1538-4357/aab34c. URL: http://dx.doi.org/10.3847/1538-4357/aab34c.

Tamanini, N. et al. (to be published). *Gravitational Waves*.

Thrane, E. and C. Talbot (2019). *An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models*. DOI: 10.1017/pasa.2019.2. URL: http://dx.doi.org/10.1017/pasa.2019.2.

Verde, L., T. Treu and A. G. Riess (Sept. 2019). 'Tensions between the early and late Universe'. In: *Nature Astronomy* 3(10), pp. 891–895. ISSN: 2397-3366. DOI: 10.1038/s41550-019-0902-0. URL: http://dx.doi.org/10.1038/s41550-019-0902-0.

Vitale, S. et al. (2021). 'Inferring the Properties of a Population of Compact Binaries in Presence of Selection Effects'. In: pp. 1–60. DOI: 10.1007/978-981-15-4702-7_45-1. URL: http://dx.doi.org/10.1007/978-981-15-4702-7_45-1.