Alma Mater Studiorum · Università di Bologna

School of Science
Department of Physics and Astronomy
Master Degree Programme in Astrophysics and Cosmology

A Flexible Mass Function for Binary Black Holes: Improving Cosmological Inference with Standard Sirens

Graduation Thesis

Presented by:
Chiara Ciapetti

Supervisor:
Prof. Michele Ennio Maria Moresco

Co-supervisors:
Dr. Nicola Borghi
Dr. Matteo Tagliazucchi

Academic Year 2024/2025
Graduation date V

Abstract

Gravitational waves (GWs) from compact binary coalescences are a novel cosmological probe, opening a new window to improve our understanding of the Universe. They act as standard sirens, since they provide an independent measurement of the luminosity distance, free from any distance-ladder assumptions. However, to fully exploit the cosmological potential of the rapidly growing number of standard sirens, a more accurate characterization of the underlying source population has become essential.

In this Thesis, which focuses on binary black hole (BBH) systems, we tackle this challenge from multiple perspectives. We develop and implement a semi-parametric model of the BBH primary mass function in a Bayesian framework, providing a more flexible alternative to previous fully parametric approaches. The model is built by combining a power-law baseline with a spline component. To maximize its versatility, we introduce different knot placement strategies, ranging from logarithmic spacing to a novel data-driven approach for informed knot placement. We then integrate this model into the `CHIMERA` code and extensively test and validate it. The model is used with mock GW data to assess its performance in a LIGO-Virgo-KAGRA (LVK) Observing Run 5 scenario, demonstrating its robustness in reconstructing the underlying mass function under both dark and spectral siren configurations. Finally, we present the first application of this model to real data, analyzing 137 BBH events from the recently released GWTC-4 catalog as spectral sirens.

We find that the new model provides a more robust reconstruction of the BBH mass distribution compared to the standard parametric ones, as verified through multiple Bayesian diagnostics. Crucially, we show that this improved reconstruction translates into tighter cosmological constraints, leading to a 12% improvement in the inference of the Hubble constant. We also investigate the impact of knot placement, demonstrating that poor choices can introduce significant biases in H_0.

This work shows that improved modeling of the BBH mass function is critical to maximizing the scientific return of GWs and provides valuable insights for the analysis of future standard sirens. These results have contributed to the submitted work Tagliazucchi M., Moresco M., Borghi N., Ciapetti C., (2026).

Contents

Introduction

Modern cosmology is based on the ΛCDM model, a successful framework that has been tested to accurately reproduce a plethora of observational data and currently serves as the standard reference for studying and interpreting the evolution of our Universe.

Despite its success, this model relies on the presence of still mysterious components, dark energy and dark matter, whose nature remains an open question in cosmology. At the same time, as observational precision has improved, several tensions between cosmological parameters have emerged. The most notable is the Hubble tension, a discrepancy between the value of the Hubble constant inferred from the early Universe via CMB observations and from the late Universe via Type Ia supernovae calibrated on Cepheids. This tension may indicate the need for new physics beyond the standard ΛCDM model and the introduction of new cosmological probes (Moresco M., et al., 2022).

In this cosmological framework, gravitational waves (GWs) generated by compact binary coalescence represent a promising avenue for the future of cosmology. GWs act as standard sirens, providing a direct and independent measurement of the luminosity distance d_L without requiring any external calibration (Schutz B. F., 1986).

However, a measurement of d_L alone does not allow cosmological inference, as the redshift information is also needed. The redshift z cannot be directly derived from GWs because of its intrinsic degeneracy with the binary detected masses. Therefore, to determine cosmological parameters, such as H_0, the degeneracy must be broken in a different way. For binaries that include at least one neutron star, redshift information could be obtained through electromagnetic counterparts (bright sirens). Nevertheless, these systems represent only a small fraction of detected events due to their lower masses (7 out of 160 total confirmed events to date, Abac A. G., et al., 2025, although so far the host has been identified only for one), as most of the observed are binary black holes (BBHs). In a different approach, the redshift can be inferred statistically by correlating the GW events with galaxy catalogs (dark sirens), while for spectral sirens it is estimated from intrinsic properties of the binaries population. For instance, the primary mass distribution can be used to infer the redshift by comparing it with expected features in the source-frame mass distribution.

As of today, it becomes crucial to develop mass models capable of identifying in detail structures within the mass distribution, to capture as best as possible the astrophysical properties of the sources, and possibly derive more accurate constraints on cosmological parameters. With the release of the GWTC-4 dataset, which contains approximately 200 GW events, we are entering an era in which the number of detections is expected to grow rapidly, both due to upgrades of existing interferometers and to the upcoming deploy-

ment of third-generation detectors. As the number of detected GW events increases, it is fundamental to employ more flexible and dynamic models, capable of adapting to the expanding dataset, rather than relying on fully parametric models which constrained the data to follow a predefined function (Edelman B., et al., 2022, Edelman B., et al., 2023. Fully parametric mass models, typically based on power laws and Gaussian peaks, offer limited flexibility and may force the data to conform to predefined structures, hindering the discovery of new features.

In this Thesis, we develop a semi-parametric data-driven model for the BBH primary mass within a Bayesian hierarchical framework, enabling a joint analysis of cosmological and astrophysical information. The aim is to investigate whether and how the semi-parametric approach, by allowing greater freedom in capturing the features of the mass distribution, can improve parameter inference compared to fully parametric models. In particular, we explore how an improved mass modeling can enhance the inference of H_0 by leveraging redshift information encoded in the mass distribution. The model is tested on simulated data and validated through forecasts for the upcoming LIGO-Virgo-KAGRA (LVK) O5 observing run, under both dark and spectral sirens configurations. Finally, the model is applied to real data with the just-released Gravitational-Wave Transient Catalog (GWTC) 4 in the spectral sirens framework to jointly infer cosmological and astrophysical parameters and assess the optimal description of the primary mass function suggested by the data.

The structure of this Thesis is outlined below.

- **Chapter 1** introduces the cosmological framework, outlining the foundations of General Relativity, Einstein's field equations, and the resulting Friedmann equations. It then discusses how GWs fit into this framework, their use as standard sirens, and the central problem of the Hubble tension. The Chapter concludes with an overview of current evidence for features in the primary mass distribution and the models developed to describe them. It also outlines how this Thesis fits within the broader context of GW cosmology.

- **Chapter 2** describes the Bayesian framework in which the semi-parametric model is implemented. It begins with a brief overview of the principles of hierarchical Bayesian inference and then moves to a more detailed discussion of the likelihood and its constituent terms. The main focus is on the parametric models of the BBH primary mass function currently implemented in the `CHIMERA` pipeline.

- **Chapter 3** introduces the semi-parametric data-driven model from both a theoretical and computational perspective. It outlines the construction of the model and its first validation tests on theoretical mass functions generated with the GWTC-4 parameter posterior estimates. Particular attention is given to the methods implemented and used to place the knots, either logarithmically spaced or manually selected according to the features of the mass distribution. The chapter concludes by explaining how cosmological information is extracted within this framework.

- **Chapter 4** presents the model applied to mock data in both dark sirens and spectral sirens analysis. The chapter begins by describing the construction of the simulated galaxy catalog and the corresponding GW catalog. In the dark sirens

case, several models in different configurations are tested, allowing us to evaluate how well the mass function is recovered and to assess the inferred parameters. The analysis is then repeated in the spectral sirens mode, only for the best performing configuration in dark sirens, to reproduce as closely as possible the procedure applied to real data, where no galaxy catalog is available.

- **Chapter 5** presents the application of the semi-parametric model to the real GWTC-4 dataset. It begins with a brief overview of the data, describing the selection and filtering procedures adopted in this work. It then introduces a data-driven method for placing the knots based on the posterior distribution of individual GW events. Several models with different configurations, some data-driven and others logarithmically spaced, are tested on the real dataset. For each of them, the reconstruction of the mass function and the inference of the cosmological parameters are analyzed, highlighting how different modeling choices lead to different results, with particular focus on the posterior distribution of H_0.

The work conducted in this Thesis contributes to the submitted publication Tagliazucchi M., Moresco M., Borghi N., Ciapetti C., (2026).

Chapter 1

Introduction to Gravitational Wave Cosmology

In 1915, Albert Einstein formulated the General Theory of Relativity, a milestone that transformed modern physics and the understanding of the Universe, particularly concerning the interplay between geometry and gravitation: geometrical properties of the Universe are related to the amount of energy and matter. The theory predicts the existence of gravitational waves (GW), ripples in the space-time propagating at the speed of light and generated by system dynamics that are not spherically or cylindrically symmetric, producing a time-varying mass quadrupole moment. The prediction remained purely theoretical despite the continuous efforts of the scientific community, up to a century later, in 2015, when the first Gravitational Wave (GW150914) was detected by the LIGO Interferometer, marking a new era in observational astrophysics. Since then, many additional GW events have been observed, with about 400 candidates reported after the fourth LIGO-Virgo-KAGRA observing run. Consequently, the study of GWs is becoming an increasingly important tool in modern cosmology. Today, GWs are emerging as a robust observational probe to constrain cosmology and fundamental physics, offering a new perspective on cosmological tensions, such as the Hubble tension.

This first chapter provides an overview of modern cosmology, starting from General Relativity and Gravitational Waves, and establishes the framework and the aim of this Thesis. Section (1.1) presents the cosmological background, including the main equations of General Relativity and the definitions of cosmological distances. In Section(1.2), we focus on gravitational waves, their generation, propagation, and interaction with test particles. Gravitational waves are then discussed as standard sirens to address current cosmological open questions, in particular the H_0 tension, alongside a review of gravitational-wave detectors and outlines future observational prospects. Finally, Section (1.3) defines the aim of this work and summarizes the current state of the art in GW astrophysics and cosmology.

1.1 An expanding Universe

Despite the numerous cosmological models developed over the past century, aiming to describe different possible evolutions of the Universe based on various geometries and

components, the scientific community currently agrees on a model that best describes the observations: the ΛCDM model. This model assumes the coexistence of three components: radiation, matter, composed of baryonic matter and cold dark matter, and the cosmological constant Λ, which accounts for the accelerated expansion of the Universe, within a flat Euclidean geometry such that $\Omega_{tot} = 1$.
At present, the Universe is in a phase dominated by the contribution of Λ, although the contribution of cold dark matter remains significant, while the contribution of radiation is effectively negligible.
The discovery of the expansion of the Universe dates back in 1929, when Hubble discovered a relation between the recession velocity (v_R) and the proper distance of galaxies (d_{pr}):

$$v_R = H(t)d_{pr} \tag{1.1}$$

where $H(t)$ is the Hubble parameter, a parameter independent of space but dependent on time that represents the expansion rate of the Universe at a given time t. In the 1990s, two independent research teams (the Supernova Cosmology Project, led by Saul Perlmutter, and the High-Z Supernova Search, led by Brian Schmidt and Adam Riess) analyzed the distances and velocities of type Ia supernovae. Their observations revealed that the Universe is undergoing an accelerated expansion.
Moreover, the study of the Cosmic Microwave Background (CMB), particularly its anisotropies and the angular power spectrum, offered independent confirmation of the accelerated expansion and flat geometry of the Universe, supporting the conclusion that the ΛCDM model accurately reproduces the observations.

1.1.1 Basic equations

Modern cosmology is based on a fundamental assumption known as the Cosmological Principle, which states that, on sufficiently large scales, the Universe is almost perfectly homogeneous and isotropic, without preferred locations or directions.
Moreover, the main force that governs the large-scale Universe is gravity and it is fully described by the theory of General Relativity.
The study of any cosmological model requires the definition of a metric that allows us to measure distances between two events in space-time. The geometry of the Universe can therefore be described by a metric tensor $g_{\mu\nu}$, which defines the infinitesimal space-time interval:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \tag{1.2}$$

where dx^μ is the infinitesimal differential of the space-time coordinates $x^\mu = (ct, x, y, z)$.
The metric that describes our Universe (ΛCDM model) is the Friedmann–Lemaître–Robertson–Walker (FLRW) metric, which, in spherical coordinates, takes the form:

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2d\theta^2 + r^2sin^2\theta d\phi^2\right] \tag{1.3}$$

where all the mixed terms in time and space vanish as a consequence of the Cosmological Principle, which ensures isotropy and homogeneity.

With this metric, each point of the space-time is characterized by four comoving coordinates (t, r, θ, ϕ), which define a comoving reference frame. In this frame, a point at rest does not change its coordinates as the Universe expands. The expansion in the metric is taken into account by the *scale factor* $a(t)$. Moreover, k is the *curvature parameter* which describes the Universe geometry: flat ($k = 0$), closed ($k = 1$) or open ($k = -1$).

The evolution of the metric in the space-time is governed by the Einstein Field Equations, which relate the geometry of the space-time to its energy and matter content:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu} + \Lambda g_{\mu\nu} \tag{1.4}$$

where $R_{\mu\nu}$ and $R = g_{\mu\nu}R^{\mu\nu}$ are respectively the Ricci tensor and the Ricci scalar, which characterize the curvature of space-time, and $T_{\mu\nu}$ is the energy-momentum tensor that accounts for energy and matter content. It can be shown that in a FLRW metric and under the assumption of a perfect fluid, the Einstein equations reduce to a system of two equations because of symmetry and isotropy, called the Friedmann Equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{1.5}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + \frac{3P}{c^2}) + \frac{\Lambda}{3}c^2 \tag{1.6}$$

By combining the Friedmann equations, it is possible to define the critical density today, as the energy density required for a spatially flat Universe at the present epoch:

$$\rho_{crit,0} = \frac{3H_0^2}{8\pi G} \tag{1.7}$$

where H_0 is the Hubble constant, representing the expansion rate of the Universe at the present epoch. Moreover, the critical density allows the introduction of the dimensionless density parameter:

$$\Omega_i = \frac{\rho_i}{\rho_{crit}} \tag{1.8}$$

which quantifies the contribution of the i-th component to the total energy density of the Universe. In the ΛCDM model, $\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{crit}}$ is the density parameter for the cosmological constant, associated to the Dark Energy and responsible for the Universe acceleration, $\Omega_m = \frac{\rho_m}{\rho_{crit}}$ is the density parameter for baryonic matter and Dark Matter, and $\Omega_R = \frac{\rho_R}{\rho_{crit}}$ is the density parameter for radiation.
For a flat Universe, the total energy density must satisfy:

$$\Omega_{tot} = \sum_i \Omega_i = 1 \tag{1.9}$$

By combining the two Friedmann equations and expressing them as a function of redshift, the resulting equation expresses a relation between the Hubble parameter $H(z)$ and the energy density of the Universe:

$$H(z) = H_0 \left[\Omega_R(1+z)^4 + \Omega_m(1+z)^3 + \Omega_\Lambda\right]^{1/2} \tag{1.10}$$

where we have used the scaling laws of the ΛCDM model components:

$$\begin{cases} \rho_R = \rho_{R,0} a^{-4} \\ \rho_m = \rho_{m,0} a^{-3} \\ \rho_\Lambda = const \end{cases} \tag{1.11}$$

and the relation between the scale factor a and the redshift z:

$$a(t) = \frac{1}{1+z}. \tag{1.12}$$

1.1.2 Distances in Cosmology

In Cosmology there is not a unique definition of distance because time and space are intrinsically connected through the theory of General Relativity. Consequently, different definitions of distance can be introduced, each appropriate for specific cosmological measurements. All distance definitions are derived using the natural units approximation $c = 1$.
Given that the Universe is undergoing accelerated expansion, one of the most fundamental quantities is the *proper distance*. The most straightforward way to parametrize it is to start from the FLRW metric in (1.3). Assuming no time variation ($dt = 0$) and considering two radially aligned objects, the proper distance is defined as:

$$d_{pr} = a(t) \int_0^r \frac{dr'}{(1 + kr'^2)^{1/2}} \tag{1.13}$$

According to General Relativity, the proper distance depends on geometry and time through the scale factor $a(t)$.
For a flat Universe ($k = 0$) as the ΛCDM model, this simplifies to:

$$d_{pr} = a(t) r \tag{1.14}$$

By differentiating the proper distance with respect to time, it is possible to recover the Hubble–Lemaître law:

$$v_R = \frac{d}{dt} d_{pr} = \frac{\dot{a}}{a} d_{pr} \tag{1.15}$$

where $H(t) = \frac{\dot{a}}{a}$ is the Hubble parameter and $H_0 = H(t_0)$ is its present value, also known as the Hubble constant, which quantifies the current expansion rate of the Universe $H_0 \approx 70$ km/s/Mpc.
It is possible to define the value of proper distance today, which is the so called comoving distance: $d_C = d_{pr}(t = t_0) = a_0 r$.
The *comoving distance* is the distance traveled by light between the source and the observer at $z = 0$ in the comoving reference frame, which is not affected by the Universe expansion:

$$d_C(z) = \int_0^z \frac{dz'}{H(z')} = \int_t^{t_0} \frac{dt}{a(t)} \tag{1.16}$$

From the *comoving distance*, two fundamental cosmological distance measures can be derived.

The first is the *angular diameter distance* d_A which quantifies the distance of a source with intrinsic size D that subtends an observed angle $\Delta\theta$. It can be derived by measuring the angular size $\Delta\theta$ and knowing the intrinsic size D of an object: $d_A = \frac{\Delta\theta}{D}$. In a comoving reference frame, the intrinsic size transforms as $\frac{L}{a}$ and the angular size becomes $\Delta\theta = \frac{L}{ad_C}$. Accounting for these relations, the angular diameter distance takes the form:

$$d_A(z) = \frac{1}{1+z} d_C(z) \tag{1.17}$$

The second is the *luminosity distance* d_L which is also derived from the *comoving distance*. It is defined as the distance where an object with intrinsic luminosity L produce the observed flux $f = \frac{L}{4\pi d_L^2}$. By including several relativistic effects, such as time dilation, cosmological redshift, and the change of the surface area $4\pi r^2$ that in a comoving reference frame becomes: $4\pi a^2(t_0) r^2$, d_L can be expressed as:

$$d_L(z) = (1+z) d_C(z) \tag{1.18}$$

1.1.3 Cracks in the standard model: the Hubble tension

The Hubble tension refers to a discrepancy in the measurement of the Hubble constant H_0 obtained from early-universe probes and those from late-universe local probes (Abdalla E., et al., 2022). In particular, the value of H_0 obtained from direct local measurements based on the distance ladder is higher than the value inferred from CMB observations under the assumption of the ΛCDM model.
Specifically, there is a $\sim 5\sigma$ disagreement between the Planck satellite CMB measurement (Aghanim N., et al., 2020):

$$H_0 = 67.4 \pm 0.5 \text{ km/s/Mpc}$$

and the measurement obtained by the 2021 SHOES Collaboration using Supernovae calibrated by Cepheids (Riess A. G., et al., 2022):

$$H_0 = 73.0 \pm 1.0 \text{ km/s/Mpc}$$

Despite significant efforts by the community to account for measurement errors and underlying assumptions, the Hubble tension has persisted and grown in statistical significance. In particular, all model-dependent estimates of H_0 (such as CMB and and Baryon Acoustic Oscillations, BAO) are mutually consistent, and the same holds for all model-independent estimates, including distance ladder and strong lensing (Abdalla E., et al., 2022).
There are two main categories of methods to infer H_0: the *late* H_0 measurements and the *early* H_0 measurements.
The *late* H_0 measurements are independent on the ΛCDM cosmological model, and allow to measure H_0 locally relying on previous steps of calibration. For very low redshift, $z \ll 1$, the luminosity distance in (1.18) can be approximated to:

$$d_L(z) = \frac{cz}{H_0} \tag{1.19}$$

Therefore, by measuring the luminosity distance and knowing the redshift, it is possible to directly obtain the H_0 value. The luminosity distance can be derived from the *distance modulus*:

$$\mu = m - M = 5log\left(\frac{d_L}{10\text{ pc}}\right) \tag{1.20}$$

where m is the apparent magnitude and M is the absolute magnitude of Type Ia Supernovae. The absolute magnitude must be calibrated using independent distance indicators, such as Cepheids in the same host galaxy. Once calibrated, the distance modulus provides the luminosity distance which can be used to infer the Hubble constant H_0.
The *early* H_0 measurements, instead, rely on cosmological assumptions such as the cosmological model ΛCDM, the properties of Dark Energy (Λ) and of the Dark Matter (CDM). The Baryonic Acoustic Oscillations (BAO) and the CMB measurements that provide standard rulers in the Universe belong to the *early* H_0 measurements. Although the 2018 Planck satellite data on CMB anisotropies confirm the consistency of the ΛCDM model, the H_0 tension between the direct and indirect measurements suggests the need for new physics.
In this context, gravitational waves represent a promising cosmological probe to provide new and independent measurements of H_0, potentially helping to resolve this discrepancy (see Section 1.2.4).

1.2 Gravitational waves as cosmological probes

The relevance of gravitational waves in modern cosmology began with the first detection of a binary black hole merger, GW150914, by the LIGO detectors in 2015, opening a new window for observing the Universe. A few years later, the first binary neutron star merger, GW170817, was observed in both gravitational and electromagnetic waves, marking the beginning of the multimessenger astronomy era (Abbott B. P., et al., 2017). GWs emitted by binary black holes and binary neutron stars represent powerful cosmological probes, as they provide an independent measurement of the luminosity distance d_L (Schutz B. F., 1986). When combined with an external measurement of the redshift, they allow to place constraints on cosmological parameters.

1.2.1 Generation of gravitational waves

Gravitational waves are small perturbations of the space time, generated by the distributions of matter and energy. To study their generation, the first step is to linearize the Einstein Field Equations in Eq. (1.4) in a region of spacetime far from the source that produces the propagating wave. Because the perturbations induced by GW are extremely small, the weak-field approximation is assumed, so that the metric $g_{\mu\nu}$ is expressed as the sum of the flat Minkowski metric and a small perturbation:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.21}$$

where $\eta_{\mu\nu} = (-1, +1, +1, +1)\delta_{\mu\nu}$ is the Minkowski metric that represents the geometry in a flat and empty space and $h_{\mu\nu}$ is a small perturbation such that $|h_{\mu\nu}| \ll 1$. The

linearized Einstein Field Equations in the Lorenz gauge takes the form:

$$\begin{cases} \Box \bar{h}^{\mu\nu} = -\frac{16\pi G}{c^4} T^{\mu\nu} \\ \partial_\mu \bar{h^{\mu\nu}} = 0 \end{cases} \tag{1.22}$$

where $\Box = -(1/c^2)\partial_t^2 + \nabla^2$ is the D'Alembertian operator.
In the compact source approximation, in which the GW perturbation is evaluated at a distance r much greater than the source dimension, the equation that describes the GW perturbation is:

$$\bar{h}^{ij}(ct, \bar{x}) = \frac{2G}{c^6 r} \left[\frac{d^2 I^{ij}(ct')}{dt'^2} \right]_{t'=t_{ret}} \tag{1.23}$$

where I^{ij} is the quadrupole moment tensor of energy density of the source which depends only on time-dependent component T^{ij}:

$$I^{ij}(ct) = \int T^{00}(ct, \vec{y}) y^i y^j d^3\vec{y} \tag{1.24}$$

As evident from Eq. (1.23), the metric perturbation produced by GW is extremely small ($\propto \frac{1}{c^6}$) making GW very faint to detect.

1.2.2 Propagation of gravitational waves

To study the propagation of gravitational waves is easier to consider regions of spacetime empty and far from the source. In vacuum, the Eq. (1.22) reduced to $\Box \bar{h}^{\mu\nu} = 0$ in the Lorenz gauge.
The general solution of Eq (1.22) is a superposition of plane waves propagating at the speed of light c:

$$\bar{h}^{\mu\nu}(\bar{x}) = \int A^{\mu\nu}(\vec{k}) e^{ik_\rho x^\rho} d^3\vec{k} \tag{1.25}$$

where $A^{\mu\nu}$ is a symmetric 4×4 amplitude tensor and $\bar{k} = \left(\frac{\omega}{c}, \vec{k}\right)$. Because of symmetry, $A_{\mu\nu} = A_{\nu\mu}$, the amplitude tensor initially has 10 independent components. The Lorenz gauge condition provide other four constraints $A^{\mu\nu}_{\ \ k\nu} = 0$ leaving six independent components.
Finally, we can exploit an additional gauge invariant, the transverse traceless gauge (TT gauge) defined by the transformation:

$$h'_{\mu\nu} = h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu \tag{1.26}$$

In the TT gauge, the metric perturbation satisfies the traceless condition $h^\mu_{\ \mu} = 0$ and the transverse condition $k^\mu h_{\mu\nu} = 0$, meaning that only the components orthogonal to the propagation direction do not vanish.

The amplitude tensor $A_{\mu\nu}$ can therefore be expressed in terms of two independent polarization states, denoted by h_+ and $h_\times$, corresponding to the two possible gravitational-wave polarizations:

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

It is convenient to decompose this tensor as a linear combination of the two polarization tensors $e_+^{\mu\nu}$ and $e_\times^{\mu\nu}$:

$$A_{TT}^{\mu\nu} = h_+ e_+^{\mu\nu} + h_\times e_\times^{\mu\nu} \tag{1.27}$$

where the polarization tensors are given by:

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

For a binary system of two spiraling objects in a circular orbit, the two GW polarization in the TT gauge are:

$$h_+(t) = \frac{4}{r}\left(\frac{GM_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{gw}(t)}{c}\right)^{2/3}\left(\frac{1+\cos^2 i}{2}\right)\cos(2\pi f_{gw} t) \tag{1.28}$$

$$h_\times(t) = \frac{4}{r}\left(\frac{GM_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{gw}(t)}{c}\right)^{2/3}\cos i\,\sin(2\pi f_{gw} t) \tag{1.29}$$

where i is a generic inclination angle between the line of sight of the observer and the normal to the orbital plane and r is the distance to the source. In the two polarization equations (1.28) and (1.29) it has been used the relations $\omega_{gw} = 2\omega_s = 2\pi f_{gw}$, reflecting the fact that GW are emitted at twice the orbital frequency of the binary system.
The parameter M_c is known as the *chirp mass*, a combination of the individual masses of the binary system:

$$M_c = \mu^{3/5} m^{2/5} = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}} \tag{1.30}$$

where m_1 and m_2 are the masses of the components and μ is the reduced mass.
The GW phase $\Phi(t)$, given by the integral of the frequency, allows to study how the GW waveform evolves in time during the inspiral of a binary:

$$\Phi(t) = 2\pi \int_{t_0}^{t} f_{gw}(t')dt' \tag{1.31}$$

where the GW frequency expresses as a function of the coalescence time $\tau = t_{coal} - t$ is:

$$f_{gw}(\tau) = \frac{1}{\pi}\left(\frac{5}{256}\frac{1}{\tau}\right)^{3/8}\left(\frac{GM_c}{c^3}\right)^{-5/8} \tag{1.32}$$

The eq. (1.31) and (1.32) allow to redefine the two GW polarizations as:

$$h_+(t) = \frac{4}{r}\left(\frac{GM_c}{c^3}\right)^{5/3}\left(\frac{\pi f_{gw}}{c}\right)^{2/3}\left(\frac{1+\cos^2 i}{2}\right)\cos[\Phi(\tau)] \tag{1.33}$$

$$h_\times(t) = \frac{4}{r}\left(\frac{GM_c}{c^3}\right)^{5/3}\left(\frac{\pi f_{gw}}{c}\right)^{2/3}\cos i\,\sin[\Phi(\tau)] \tag{1.34}$$

As the two objects spiral inward, τ decreases, and consequently both the frequency and the amplitude increase, producing the characteristic chirp signal observed by gravitational-wave detectors.
When accounting for the expansion of the Universe, we can express all quantities in the observer's frame at the time of coalescence. The GW phase is not affected by the Universe expansion, since frequency and time have an opposite contribution of $(1+z)$:

$$dt_{obs} = (1+z)dt_s \qquad\qquad f_{gw}^{obs} = \frac{f_{gw}^{s}}{1+z} \tag{1.35}$$

The amplitude, instead, is affected by the Universe expansion. By multiplying the distance r for the scale factor that is assumed to be constant $a(t_0)$, we can exploit the relation:

$$a(t_0)r = \frac{d_L}{1+z} \tag{1.36}$$

to rewrite the GW polarization in the form:

$$h_+ = \frac{4}{d_L}\left(\frac{G\mathcal{M}_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{gw}^{obs}}{c}\right)^{2/3}\left(\frac{1+\cos^2 i}{2}\right)\cos[\Phi(\tau_{obs})] \tag{1.37}$$

$$h_\times = \frac{4}{d_L}\left(\frac{G\mathcal{M}_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{gw}^{obs}}{c}\right)^{2/3}\cos i\,\sin[\Phi(\tau_{obs})] \tag{1.38}$$

where $\mathcal{M}_c = (1+z)M_c$ is the redshifted Chirp Mass and $f^{obs} = \frac{f^s}{1+z}$ the observed frequency is redshifted with respect to the source frame frequency.
Finally, the replacement of r with d_L suggests that knowing the amplitude of the polarization, it is possible to derive the luminosity distance and use GW as standard sirens (see Section 1.2.4).

1.2.3 Effect on test particles

To study the effect of a passing GW and how it influences the spacetime, let us consider the motion of a set of test particles at rest, that form a circle in the (x^1, x^2) plane with the reference frame at the center. Adopting the TT gauge, the test masses maintain constant spatial coordinates during the interaction with the GW, however their physical spatial separation l changes:

$$l^2 = g_{ij}\xi^i\xi^j = \delta_{ij}\zeta^i\zeta^j \tag{1.39}$$

where $[\xi^\mu]$ is the coordinate separation and $\zeta^i = \xi^i + \frac{1}{2}h^i_k\xi^k$ is a convenient physical distance vector including the perturbation induced by the GW. For simplicity, we consider a GW propagating only along the x^3 direction such that $h^{\mu\nu}_{TT} = A^{\mu\nu}e^{ik(x^0-x^3)}$.
In the straightforward case of a single polarization, assuming $A^{\mu\nu} = h_+e^{\mu\nu}_+$, the physical separation oscillates as a function of time as illustrated in the upper panel of Figure (1.1):

$$\vec{\zeta} = (\zeta^1, \zeta^2, 0) = (\xi^1, \xi^2, 0) + \frac{1}{2}h_+\cos[k(x^0 - x^3)](\xi^1, -\xi^2, 0) \tag{1.40}$$

In contrast, for the case of the $\times$ polarization where $A^{\mu\nu} = h_\times e^{\mu\nu}_\times$, the physical separation oscillates as shown in the middle panel of Figure (1.1):

$$\vec{\zeta} = (\zeta^1, \zeta^2, 0) = (\xi^1, \xi^2, 0) + \frac{1}{2} h_\times \cos[k(x^0 - x^3)](\xi^2, \xi^1, 0) \tag{1.41}$$

In the most general case, the GW propagates as a superposition of the two linear polarizations, resulting in a combined motion that incorporates the effects of both cases, as depicted in the bottom panel of Figure (1.1).

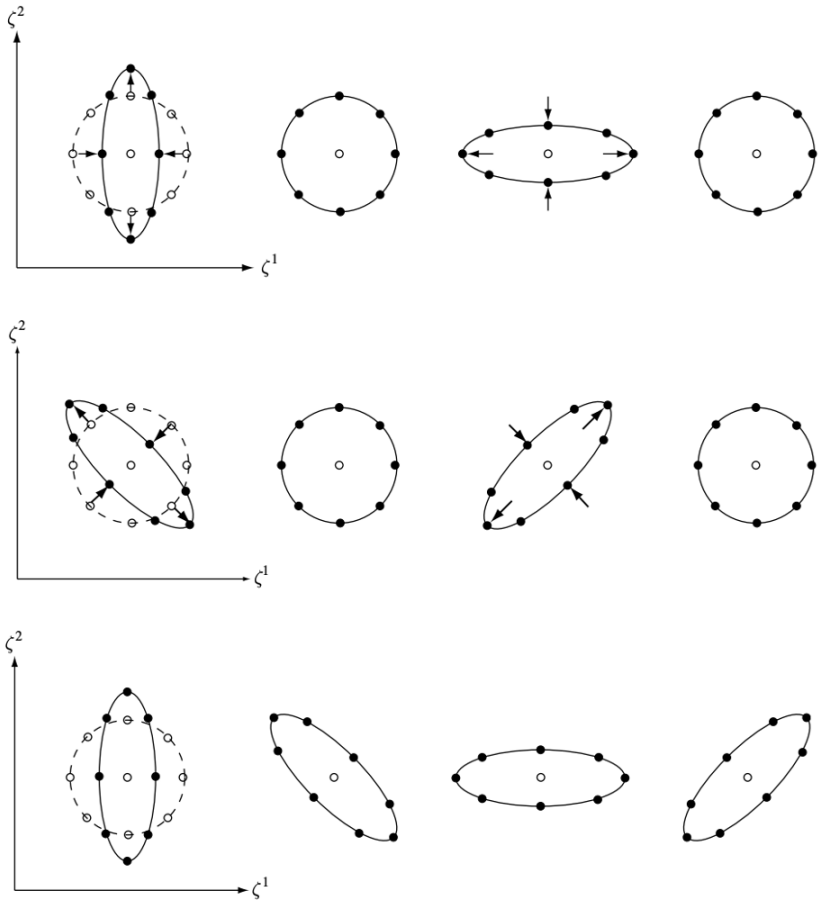

Figure 1.1: In all three panels, the open circles indicate the initial positions of the test particles, while the filled circles represent their displaced positions under the effect of the gravitational wave: the upper panel corresponds to the + polarization, the middle panel to the $\times$ polarization, and the bottom panel to a combination of the two. Illustration taken from Hobson M. P., et al., 2006

.

1.2.4 Standard sirens

Merging black holes and neutron stars, known as Compact Binary Coalescence (CBCs), emit gravitational waves as they spiral toward coalescence. These systems represent powerful cosmological probes, as from their signal it is possible to measure directly their

d_L, making them ideal *standard sirens* (Moresco M., et al., 2022).
In fact, as shown in Eqs. (1.37) and (1.38), the GW strain amplitude is inversely proportional to the luminosity distance: by knowing the amplitude and the frequency evolution of the GW signal, it is possible to derive d_L without any calibration.
However, the redshift is degenerate with detector-frame masses because higher masses at higher redshifts produce the same signal as lower mass at lower redshift. Because of this degeneracy, the redshift can not be directly derived from GW data alone. Therefore, using GW observations to perform cosmological inference through the $d_L - z$ relation is not possible without introducing additional information on z. There are three ways to include in the system a knowledge of the redshift:

- *bright sirens*: The redshift is obtained from the host galaxy of the GW event thanks to the identification of an electromagnetic (EM) counterpart (Holz E. and Hughes S., 2005). This allows for precise determination of cosmological parameters. The presence of a neutron star in the binary system is required to produce EM emission, as in the case of GW170817 (Abbott B. P., et al., 2017). When GW170817 was detected, a few seconds after the merger, a gamma-ray burst (GRB 170817A) was detected within a region of the sky consistent with the GW signal (THE LIGO and VIRGO scientific collaboration et al.) Once the region of sky was localized, optical observations made it possible to identify the host galaxy, from which the redshift could be derived through spectroscopy.
  This is the most straightforward and difficult case: currently only 1 out of ~ 300 GW events is detected as bright sirens, since it requires a good localization and a NS in the CBC.
  This combined detection of GW and EM marked the beginning of the age of multi-messenger GW astrophysics, the most promising to infer cosmological parameters in an accurate way.

- *dark sirens*: In the most cases, no direct EM counterpart is detected, because the majority of detected GW come from BBHs. In this case, the derivation of redshift becomes difficult, since the GW localization area is typically very wide and can contain numerous galaxies. A statistical approach must be used to infer the redshift, making use of galaxy catalogs (Finke A., et al., 2021). By cross-correlating a detected GW with a galaxy catalog, it is possible to statistically associate a potential host galaxy. Combining the redshift of each potential host within the GW localization volume provides separate constraints on the cosmological parameters for each event. By stacking together the information from more events, the true parameters statistically emerge, with greater weight assigned to the those estimates associated with smaller localization volumes, since they provide higher accuracy.

- *spectral sirens*: The redshift is derived relying only on GW data, making assumptions on the source-frame GW mass distribution (Mastrogiovanni S., et al., 2021). Differently from the dark sirens, spectral sirens derivation of redshift does not require necessarily the use of a galaxy catalog. This method is more reliable when one or more features are present in the mass distribution of the source population. Originally, this method was explored primarily for binary neutron stars (BNS), because of their narrow mass distribution and higher mass cutoff (Messenger C.

and Read J., 2012; Del Pozzo W., et al, 2017). However, the method is now mainly applied to BBHs, which constitute the majority of observed GW events and display evident mass distribution features whose astrophysical interpretation is still debated: an excess of detections at ~ 35 $M_\odot$ and a clear decrease of high mass BHs, producing a gap at $\sim 50 - 120$ $M_\odot$ that could be due to the pair instability supernova (PISN) (Ezquiaga J. et Holz D., 2022).
Since the masses measured by the detector are redshifted, the source-frame mass is related to the detector-frame mass by:

$$m_{1,src} = \frac{m_{1,det}}{1+z}. \tag{1.42}$$

Therefore, by measuring the detector-frame mass and assuming a model for the source-frame mass distribution, it is possible to statistically break the mass-redshift degeneracy and constrain cosmological parameters.

This thesis employs standard sirens to analyze data: it examines mock datasets using both the dark and spectral sirens methods, based on a simulated galaxy catalog, and also investigates events from the GWTC-4 catalog through the spectral sirens approach, assuming a specific form for the BBH mass function.
The main focus of this Thesis to explore how astrophysical and cosmological parameters can be jointly inferred, by recovering the redshift information from the source mass distribution, and to investigate how the accuracy of the mass model can affect these inferences.

1.2.5 Gravitational waves detectors

As already discussed in Section (1.2.3), GW induce ripples in the spacetime, producing a strain $h(t)$. Detecting the passage of a GW therefore requires measuring the stretching of space and the corresponding variation in distance. In this section, we will focus on current ground-based interferometers and future space-based interferometers, discussing their characteristics and observational capabilities.
Current ground-based interferometers are designed to detect CBCs with masses in the range $1 - 10^3$ $M_\odot$. These detectors can measure strain amplitudes of $h(t) \approx 10^{-20}$ at frequencies $\nu \approx 10^2$ Hz.
All the GW detections obtained so far have been carried out by the LVK Gravitational Wave Network, which includes the *Laser Interferometer Gravitational-Wave Observatory* (LIGO) consisting of two identical and widely separated detectors (LIGO Hanford and LIGO Livingston), the *Virgo gravitational-wave detector* (VIRGO), and the *Kamioka Gravitational Wave Detector* (KAGRA).
All four detectors are based on Michelson interferometers and share the same L-shaped configuration. Their main components include a beam splitter, which divides the incoming laser beam into two perpendicular paths; mirrors placed at the ends of each arm, which reflect the light back toward the splitter; and a photo-detector, where the two beams recombine. In the absence of a gravitational wave, the beams recombine in phase, resulting in destructive interference at the detector. The passage of a GW, instead, modifies the arm lengths, producing an interference pattern that reveals the presence of the wave. The study of the optical length to detect GW with frequency $\nu_{GW} \approx 10^2$Hz

and $\lambda_{GW} \approx 3000$ km suggests that the arms should have length such that (Maggiore M., 2008):

$$L = \frac{\lambda_{GW}}{4} = 750 \text{ km} \left(\frac{100 \text{ Hz}}{\nu_{GW}} \right). \tag{1.43}$$

This distance cannot be reached in ground based interferometers, that's why an advanced setup with a Fabry Perot cavity is introduced. A Fabry Perot cavity is a cavity identified by two reflecting mirrors, containing a transmissive substrate and a reflective coating that makes the light bounce back and forth, increasing the effective optical path. The passage of a gravitational wave deforms the length of the interferometer arms, causing a slight difference in the distances traveled by the light beams. As a result, when the beams recombine at the photo-detector, they are no longer in phase.

All currently operating interferometers are continuously upgraded between observing runs, allowing each successive run to achieve greater sensitivity and detect an increasing number of sources. These improvements allow more precise studies of the population properties of compact binaries. During the observing runs, the LVK Collaboration has implemented numerous enhancements to the interferometers, including seismic and thermal isolation. *GWTC-3.0*, which combines observations from the first three observing runs, includes 76 candidates with a false alarm rate FAR < 1 yr^{-1}: 69 BBHs, 4 NSBHs, 2 BNSs. During O4a, the first run of O4, 128 candidate signals were detected with a probability of astrophysical origin of $p_{astro} \geq 0.5$, of which 85 (84 BBHs and 1 NSBH) have $FAR < 1yr^{-1}$ (LIGO VIRGO and KAGRA scientific collaboration, 2025).
At the time of writing, the LIGO, Virgo, and KAGRA detectors have completed their fourth observing run, which concluded on 18 November 2025. Currently the interferometers are in a period of upgrades and commissioning. Before reaching the required sensitivity to start the fifth observing run O5, a six-month intermediate run IR1 is scheduled to test and validate the upgrades.

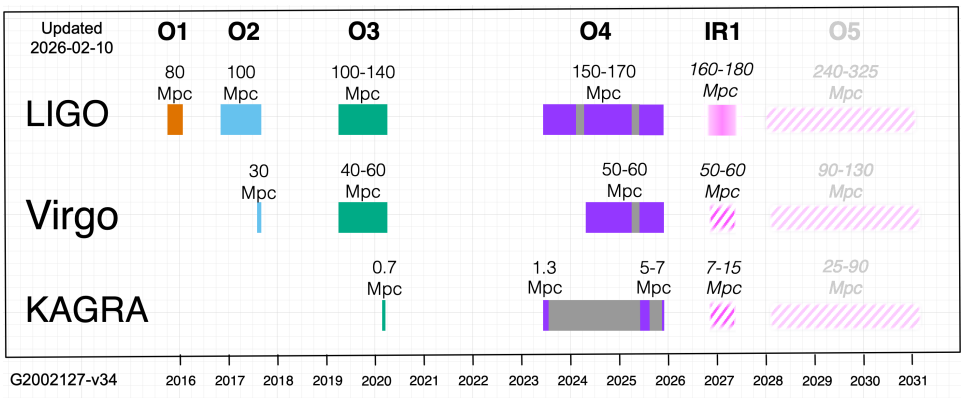

Figure 1.2: Past current and future observing runs for the LVK collaboration. The colored interval represents the period of time during which data have been collected and the future time schedule for data collection for O5. For each interval of data acquisition it is reported also the distance at which BNS have been detected with a SNR threshold of 8. Image taken from `https://observing.docs.ligo.org/plan/`.

In parallel, a third generation of GW detectors has been proposed, including the Einstein

Telescope (Punturo M., et al., 2010) and Cosmic Explorer (Reitze D., et al., 2019). These detectors aim to significantly improve sensitivity and extend the observable frequency and redshift range up to $z \sim 100$. Although the design of the Einstein Telescope is still under development, one of the key challenges is to extend the detection bandwidth to lower frequencies down to $\sim 2 - 3$ Hz, to access to earlier inspiral phases and higher redshift sources (Branchesi M., et al., 2023).
Finally, the space-based interferometers LISA (Amaro-Seoane P., et al., 2017), scheduled for launch in 2035, will probe a completely new low-frequency range $0.1\text{mHz} - 0.1\text{Hz}$ of GW, enabling the detection of mergers of massive black holes (MBHs) with masses in the range $10^4 - 10^6$ $M_\odot$.

1.3 State of the art on compact binary coalescence

In just about a decade, gravitational-wave astronomy has evolved from the detection of a single event to a catalog of more than three hundred confident events. These observations have been carried out in four main observing runs, as shown in Table (1.1).

Run	**Dates**	**Duration months**	**Total**	**BBH**	**BNS**	**NSBH**
O1	Sep 18, 2015 - Jan 12, 2016	4	3	3	0	0
O2	Nov 30, 2016 - Aug 25, 2017	9	8	7	1	0
O3	Apr 1, 2019 - Mar 26, 2020	11	79	~ 74	1	~ 4
O4	May 24, 2023 - Nov 18, 2025	30	$\sim 250^*$	$\sim 243^*$	0	$\sim 5^*$
Total			$\sim 340^*$	$\sim 327^*$	2	$\sim 9^*$

Table 1.1: Time evolution of the number of detected GW events during the observing runs, summarizing the time span, the duration, and the GW events, divided for CBC type (BBH, BNS, NSBH). * These values include also preliminary real-time candidates not yet confirmed and ambiguous events that could be misclassified. The first part of the observing run (O4a) is released with the GWTC-4.0 catalog and includes 128 confirmed events. O4b and O4c contributions are included, with ~ 2 low-latency alerts with $> 50\%$ NSBH classification probability.

With the increasing statistics, a progressively richer structure has emerged in the population of merging BBHs. Among the inferred source properties, the primary mass distribution constitutes a key observable, as its features encode information on formation channels, evolutionary processes, and may be used for cosmological purposes. The three events from the first observing run (O1, Abbott B. P., et al., 2016) show some similar source properties: the component masses fall within the range expected for stellar mass black holes, since they have masses in the range ($m_{1,2} \approx 5 - 65$ $M_\odot$). Due to the small number of detections, early population analyses adopted simple parametric models and the first developed model was a simple power law (Abbott B. P., et al., 2016) with fixed slope:

$$p(m_1) \propto m_1^{-\alpha} \qquad \alpha = 2.35 \tag{1.44}$$

At these stages, studies were mainly descriptive and focused on confirming the presence of a BBH population of stellar origin.

With the second observing run O2 and the increase of detected events, the first statistically meaningful deviations from the power law were identified. In particular, the primary mass distribution was found to exhibit two additional features: a sharp high-mass cutoff and a pronounced peak around $\sim 35 - 40$ $M_\odot$. To account for these features, Talbot and Thrane (2018) proposed a phenomenological model combining a power-law with a Gaussian peak centered near the upper end of BBH mass distribution. This was physically motivated by the expected signature of pulsational pair-instability supernovae, which limits the final mass of the BBH remnants. Stars with initial masses between $100 - 150$ $M_\odot$ are expected to form ~ 40 $M_\odot$ BH remnant due to the pulsational pair instability supernova; whereas stars in the $150 - 200$ $M_\odot$ range undergo pair-instability supernovae, producing a gap in the BBH mass distribution between 40 and 250 $M_\odot$ (Belczynski K., et al., 2016). This leads to a cutoff in the source mass distribution and an excess around 40 $M_\odot$. The physical origin of pulsational pair-instability supernovae could lie in the Helium core becoming unstable in the mass range $30 - 64$ $M_\odot$ due to the electron-positron pair production, which reduces the internal pressure. The star undergoes pulsations, losing mass until a stable remnant of ~ 35 – 40 $M_\odot$ is formed. For cores of $64 - 135$ $M_\odot$, the instability completely disrupts the star, leaving no remnant (Woosley S. E. and Heger A., 2021). Therefore, the high-mass gap in the BBH primary mass distribution could be explained by the total disruption of the most massive stars and the mass loss from pulsational pair-instability, producing the observed peak around $35 - 40$ $M_\odot$.

During the combined analysis of O1 and O2 detections, Abbott et al. (2019) explored three parametric mass function models, such as a simple power law, a broken power law and a power law plus peak, inspired by Talbot and Thrane (2018). Among these models, the power law plus peak provides the best fit to the data, supporting the presence of the bump feature.
The release of the second Gravitational Wave Transient Catalog GWTC-2, combining the observations from O1, O2 and the first half of the third observing run O3a, enabled a more robust study of the primary mass function distribution and the inference of population parameters.
Abbott et al. (2021) introduced a new model to parametrize the primary BBH mass distribution, namely *Multi peak* model, constructed as a power law with two Gaussian peaks: the first centered at ~ 33 $M_\odot$ and the second at ~ 68 $M_\odot$. With the release of GWTC-2, the shape of the primary BBH mass above 45 $M_\odot$ begins to be resolved. The *Multi peak* model was motivated by the presence of a non-zero merger rate for BBH systems with primary masses above 45 $M_\odot$. While analyses based on O1 and O2 data indicated that approximately 99% of BBH systems had primary masses below 45 $M_\odot$, the inclusion of higher mass detections in GWTC-2 reduces this fraction to about 97%. Nevertheless, it was found that the addition of the second peak is not statistically favored by the data, and the model that still best reproduces the population remains the power law plus peak.
GWTC-2 marked the first use of semi-parametric and non-parametric models to infer

the BBH mass spectrum, relaxing strong assumptions and enabling a more data driven reconstruction of features present in the population.
A semi-parametric and a non-parametric modeling of BBH primary mass distributions have been proposed through two different approaches, respectively by Edelman B., et al. (2022) and Rinaldi S. and Del Pozzo W. (2021). In the former paper, the authors implemented a cubic spline constructed over a power law, finding empirical support for the peak at 35 $M_\odot$ and for an additional feature around 10 $M_\odot$. The second paper, instead, introduces a fully non parametric framework based on a Hierarchical Dirichlet Process Gaussian Mixture Model (HDPGMM). Within this approach, the inferred mass function again exhibits the presence of two distinct peaks located at 38 $M_\odot$ and 15 $M_\odot$. Both studies represent a turning point in the investigation of BBH system masses, although the results remain affected by a clear degeneracy between the inferred mass features and other population parameters, as well as by possible fluctuations arising from the limited number of detected events.
The third gravitational wave transient catalog GWTC-3, incorporating the full O3 observing run, expanded the BBH sample and introduced new classes of sources, including neutron star - black hole (NSBH) mergers (Abbott R., et al., 2023).
With this enlarged dataset, semi-parametric models became crucial for population inference. The analyses of GWTC-3 revealed features in the BBH primary mass distribution that are not captured by parametric models. In particular, they identified a lower mass gap between $2 - 5$ $M_\odot$ and found evidence for multiple over-densities across the mass spectrum.
This gap is between the most massive neutron stars and the lightest black holes and it originates from the physics of core collapse supernovae explosion. The understanding of this gap provides not only direct information on the BBH primary mass distribution, but also indirect information on supernova (SN) mechanism. If the SN explosion develops rapidly, on timescales of about 10–100 ms, the instabilities grow quickly, leading to the ejection of the envelope before significant accretion occurs and producing few remnants in the $\sim$ 2–5 $M_\odot$ range. In contrast, if the explosion is delayed (on timescales of about $\gtrsim$ 200 ms), the proto–neutron star can continue to accrete material before the neutrino-driven explosion takes place, leading to the formation of remnants that fill the gap. The study of the primary mass distribution of NS and BH in our Galaxy provides evidence that the SN explosion develops on rapid timescales, explaining the presence of the gap (Zevin M, et al., 2020).
With the increasing number of GW events, it becomes possible to jointly infer the population and cosmological hyperparameter. In particular, spectral sirens rely on prominent features in the mass distribution to break the degeneracy between the source frame mass and redshift. Given the current uncertainties in the BBH formation channels and the emergence of unresolved features in the mass domain, a more flexible non parametric approach to spectral siren cosmology is therefore required.
The Binned Gaussian Process, was introduced by Abbott R., et. al. (2023) as one of the key non-parametric approaches, which models the mass distribution by dividing the mass space into bins and estimating the merger rate in each bin, while enforcing correlations between neighboring bins. Hernandez I. M. and Ray A. (2024) applied the BGP to GWTC-3 BBH detections, using them as spectral sirens to jointly infer both population and cosmological parameters, without assuming any specific parametric form

for the source-frame mass distribution or the underlying power law.
Another used non parametric model is VAMANA (Tiwari V., 2021), which, unlike the previous approach, removes the need for binning and reconstructs the BBH mass function as a continuous function driven by the data through sampling the function on a fine grid.
More independent non-parametric analyses have further improved this framework. In particular, Toubiana A., et al. (2023), using reversible jump Markov Chain Monte Carlo techniques, reported a statistically significant excess of BH systems around ~ 20 $M_\odot$. This feature has been interpreted as a possible signature of second-generation mergers occurring in dense stellar environments, although it is recovered by only a subset of the analyses performed on the GWTC-3 dataset, and its statistical robustness remains under investigation. Tentative evidence was also reported for a third Gaussian peak around ~ 65 $M_\odot$, which could account for the observed excess merger rate at the boundary of the PISN mass gap.
These results strengthened the idea that the BBH mass spectrum contains a complex structure, suggesting multiple formation channels and evolutionary processes. A variety of additional non parametric approaches have been explored, including autoregressive processes (Callister T. A. and Farr W. M, 2024), adaptive kernel density methods (Sadiq J., et al., 2022) and maximum likelihood methods (Payne E. et Thrane E., 2023).

The release of GWTC-4.0, including events from O1 through the first segment of the fourth observing run (O4a), has confirmed the presence of the main features identified in earlier catalogs. The peaks at 10 and 35 $M_\odot$ and the bump around 20 $M_\odot$ are consistently recovered by all the models considered despite significant differences between them.
Abac A. et al. (2025) adopted as a fiducial mass model to analize GWTC-4 a power law plus two peaks, which best captures the two prominent peaks at 10 and 35 $M_\odot$. In comparison, a semi-parametric spline model has been proposed, which identifies the same features.
The various proposed models not only provide insight into the mass features but also offer the opportunity to use them as spectral sirens, as demonstrated in Hernandez I. M. (2025). In that work, two parametric models are considered: the power law plus peak, which was the best model up to GWTC-3, and the power law plus two peaks, which is the preferred model for GWTC-4. Both models yield lower cosmological constrains on H_0 compared to a flexible non-parametric approach using Gaussian processes. This highlights the growing need to explore nonparametric, data-driven models for gravitational-wave cosmology.

1.4 Aim of the Thesis

The rapidly increasing number of GW detections is providing a large amount of information on the population of BBH. In particular, the primary mass distribution of BBH mergers is fundamental, as it contains information about stellar evolutions and BH formation. Features like peaks and gaps in the mass distribution could be linked to specific astrophysical processes, while the shape of the mass function affects the cosmological

parameters constraints. For this reason, there is a growing interest in the scientific community in developing and using flexible models that allow the data to determine the shape of the mass function, without relying on strong prior assumptions.
The aim of this Thesis is to develop a semi-parametric model for the BBH primary mass function within the Bayesian framework of `CHIMERA`, allowing the data to reveal potential features in a flexible way. This model is tested and validated using simulated data and subsequently applied to real GW observations, considering the GWTC-4 dataset.
Moreover, while significant progresses have been made in the development of semi- and non- parametric models, the interplay between the mass function and the inference of cosmological parameters is not fully explored. In this Thesis, we also perform a joint astrophysical and cosmological analysis to investigate whether and how the inferred mass distribution can influence cosmology and impact the estimation of cosmological parameters. In particular, we focus on the estimation of the Hubble constant H_0 and on its role in addressing the tension between local and early Universe measurements.

Dataset	Models	Evidences	Cosmology free	BS | DS | SS	Ref.
O1	PL	BH population of stellar origin	×	-	Abbott B. P., et al., 2016
O1–O2	PLP	High-mass cutoff Peak at 35 $M_\odot$	×	-	Talbot and Thrane (2018)
O1–O2	Model A (PL), Model C (PLP)	BBH mass and spin population inference	×	-	Abbott et al., 2019
GWTC-2	TPL, PLP, BPL, Multi-Peak	BBH shape above 45 $M_\odot$	×	-	Abbott et al. (2021)
GWTC-2	Cubic SPLINE	Peaks at 10 $M_\odot$ and 35 $M_\odot$	×	-	Edelman B., et al. (2022)
GWTC-2	non parametric HDPGMM	Peaks at 15 $M_\odot$ and 38 $M_\odot$	×	-	Rinaldi S. and Del Pozzo W. (2021)
GWTC-3	PS, FM , BGP	Gap at $2-5$ $M_\odot$	×	-	(Abbott R., et al., 2023)
GWTC-3	BGP	Gap at low masses and peak at 35 $M_\odot$	✓	SS	Hernandez I. M. and Ray A. (2024)
GWTC-3	reversible jump Markov Chain Monte Carlo	Bump at 20 $M_\odot$	×	-	Toubiana A., et al. (2023)
GWTC-4	PL2P, semi-parametric SPLINE	Peaks at 10 $M_\odot$ and 35 $M_\odot$	×	-	Abac A. et al. (2025)
GWTC-4	PLP, BP2P, GP	Cosmological inference	✓	SS	Hernandez I. M. (2025)

Table 1.2: List of the most relevant papers on BH mass distributions and their implications for gravitational-wave cosmology discussed in this work. The labels **BS**, **DS** and **SS** refer to bright sirens, dark sirens, and spectral sirens, respectively.

Chapter 2

Statistical framework

Over the past few years, several independent pipelines have been developed for standard siren analyses, such as `MGCosmoPop` (Mancarella et al., 2022), `GWPopulation` (Talbot et al., 2019), and `GWInferno` Edelman B., et al., 2023 for spectral sirens, and `DarkSirensStat` (Finke et al., 2021) for dark sirens using galaxy catalogs.
More recently, other pipelines have been publicly released, improving on the previous ones since they can provide joint astrophysical and cosmological inference in a variety of configurations, including spectral, dark, and bright sirens (e.g., `icarogw` Mastrogiovanni S., et al., 2024 and `gwcosmo` Gray R., et al., 2023).
`CHIMERA` is one of the most recently released pipelines, and it can be applied to dark, spectral, and bright sirens. Compared to the previous ones, `CHIMERA` extends the range of implemented models and introduces a new approach to model the likelihood.

The aim of this Thesis is to introduce a semi-parametric binary black hole (BBH) mass function model to obtain cosmological constraints using the spectral siren method, based on a flexible data-driven description of the population. In particular, we jointly infer cosmological and BBH population hyperparameters using the Hierarchical Bayesian Inference approach implemented in `CHIMERA` (Borghi N., et al., 2024, Tagliazucchi M., et al., 2025).
This chapter introduces the statistical framework and its implementation within the `CHIMERA` pipeline, leading to the development of the likelihood used to perform the joint inference. Particular focus is given to the population modeling, and in particular to the parametric mass function models already implemented in `CHIMERA`, which will be extended in this Thesis through the introduction of a semi-parametric model. Finally, a schematic overview of the `CHIMERA` workflow is presented to summarize the main components of the analysis pipeline.

2.1 Hierarchical Bayesian Inference

A common approach to study the population of GW events is the Hierarchical Bayesian inference, use to simultaneously infer the source parameters and the population model (Talbot C. and Thrane E., 2018). The term 'hierarchical' refers to the two level structure of the analysis, in which event-level parameters $\bar{\theta}$ are linked to the population hyperparameters $\bar{\lambda}$, allowing for simultaneously parameter estimation and model selection

(Adams M. R., et al., 2020).
In particular, each GW event is characterized by event-level parameters $\bar{\theta}$, which describe the intrinsic (e.g., masses and spins) and extrinsic (e.g., position, orientation, and distance with respect to the observer) properties of the source. Under the assumption of adiabatic and quasi-circular non precessing orbits, the event-level parameter of a BBH in the detector frame are 15:

$$\theta^{det} = \{\mathcal{M}_c,\ \eta,\ ,\vec{\chi}_1,\ \vec{\chi}_2,\ d_L,\ \theta,\ \phi,\ i,\ \psi,\ \phi_c,\ t_c\} \tag{2.1}$$

where $\mathcal{M}_c$ is the detector frame Chirp mass and η is the symmetric ratio. The vectors $\vec{\chi}_1$ and $\vec{\chi}_2$ denote the spin components of the two BHs, d_L is the luminosity distance, $\theta = \pi/2 - \mathrm{DEC}$ and $\phi = \mathrm{RA}$ are the position angles, and ι and ψ are the inclination and polarization angles, respectively. Finally, ϕ_c and t_c correspond to the phase and time at coalescence.

It is generally not possible to uniquely reconstruct the hyper-parameters vector $\bar{\lambda}$ from a limited number of GW events affected by selection effects and measurement uncertainties. Instead, to reconstruct them, it is possible to estimate their posterior probability given the data and using the Bayes' theorem:

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

where $p(\{\bar{d}_i\}|\bar{\lambda})$ is the likelihood, i.e. the probability of observing the data set $\{d_i\}$ given the population hyper-parameters $\bar{\lambda}$, $\pi(\bar{\lambda})$ is the prior on the population hyper-parameters, and $p(\{\bar{d}_i\})$ is the Bayesian evidence, which is a normalization factor computed as the integral of the numerator over all $\bar{\lambda}$, and which can be neglected for parameter inference.

2.1.1 Derivation of the likelihood form

To derive the expression of the likelihood given a dataset of GW events, we first assume that measurement uncertainties are negligible. Under this assumption, the data $\{d_i\}$ allow the direct recovery of the event-level parameters $\bar{\theta}$. For N_{ev} independent observations, the probability of obtaining them is given by:

$$p(\{\bar{\theta}_i\}|\bar{\lambda}) = \prod_{i=1}^{N_{ev}} \frac{p_{pop}(\bar{\theta}_i|\bar{\lambda})}{\int \mathrm{d}\bar{\theta}\ p_{pop}(\bar{\theta}_i|\bar{\lambda})} \tag{2.3}$$

where $p_{pop}(\bar{\theta}_i|\bar{\lambda})$ is the population distribution of the event-level parameters for given hyper-parameters $\bar{\lambda}$, and the denominator normalizes this distribution.
However, each GW event is affected by selection effects. To account for this, it is necessary to introduce the detection probability $P_{det}(\bar{\theta})$, the probability that GW events with event-level parameters $\bar{\theta}$ can be detected in the detector frame.
In general, the detection probability depends on the signal-to-noise ratio (SNR) of the GW event. Therefore, it can be expressed as:

$$P_{det}(\bar{\theta}) = \int_{d\,\in\,det} p(d'|\bar{\theta})\mathrm{d}d' \tag{2.4}$$

where $d \in det$ means, for example, that the SNR of a GW event is above a given detectability threshold. The SNR has two stochastic components: one related to the detector noise and another intrinsic component due to the variations in the source parameters. For instance, higher-mass mergers are relatively easier to detect than lower-mass ones.

Taking into account the selection effects, Eq. (2.3) becomes:

$$p(\{\bar{\theta}_i\}|\bar{\lambda}) = \prod_{i=1}^{N_{ev}} \frac{p_{pop}(\bar{\theta}_i|\bar{\lambda})}{\int \mathrm{d}\bar{\theta}\ p_{pop}(\bar{\theta}|\bar{\lambda})\ P_{det}(\bar{\theta})} \tag{2.5}$$

where, in the numerator, we have used the fact that for the detected events, the detected probability is $P_{det}(\bar{\theta}_i) = 1$; instead, the denominator normalizes the distribution over all detectable events.
Finally, by applying the Bayes theorem to Eq. (2.5), the probability of observing an event in the presence of both measurement uncertainty and selection effects becomes:

$$p(\bar{d}|\bar{\lambda}) = \frac{\int \mathrm{d}\bar{\theta}\ p(\bar{d}|\bar{\theta}) p_{pop}(\bar{\theta}|\bar{\lambda})}{\xi(\bar{\lambda})} \tag{2.6}$$

where

$$\xi(\bar{\lambda}) = \int \mathrm{d}\bar{\theta}\ p_{pop}(\bar{\theta}_i|\bar{\lambda})\ P_{det}(\bar{\theta}_i) \tag{2.7}$$

represents the fraction of all GW events that would be detected given a specific population model (Mandel I., et al., 2018).
Assume now to have a set of N_{ev} independent GW events $\{\bar{d}_i\}$.
According to Eq. (2.6), the likelihood term $p(\{\bar{d}_i\}|\bar{\lambda})$ represents the probability of observing the full dataset $\{\bar{d}_i\}$ given the population hyper-parameters λ and the model. More specifically, every single event has its own likelihood $p(d_i^{GW}|\theta_i)$, which describes the probability of detecting the event $\bar{d}_i$ with the event-level parameters θ_i.
The total likelihood is therefore obtained by combining these single-event likelihoods with the population probability, and including the selection effects, as:

$$p(d^{GW}|\lambda) \propto \frac{1}{\xi(\lambda)^{N_{ev}}} \prod_{i=1}^{N_{ev}} \int p(d_i^{GW}|\theta_i) p_{pop}(\theta_i|\lambda) d\theta_i. \tag{2.8}$$

2.2 Population modeling

The core of the standard siren method is to construct the population prior term as in Eq. (2.8). This section presents the scheme also used in `CHIMERA` (Borghi N., et al., 2024) that we aim to extend in this Thesis. More details on the numerical implementation are provided in Section (2.3).
In this study, we focus on the BBH population, as these systems constitute the great majority (about 95%) of GW detections so far (Abac A. G., et al., 2025). Under the assumption that the BBH mass distribution does not evolve in redshift, we can divide the set of hyper-parameters λ into the three vectors, the BH mass function parameters

λ_m, the redshift distribution parameters λ_z and the cosmological parameters λ_c. The population probability can be factorized as the product of two independent contributions:

$$p_{pop}(\theta, \lambda) = p(m_1, m_2|\lambda_m)p(z, \hat{\Omega}|\lambda_c, \lambda_z). \tag{2.9}$$

The first term $p(m_1, m_2|\lambda_m)$ is the mass function, which describes the distribution of primary and secondary mass components. The mass function models are described in detail in Section (2.3).

The second term $p(z, \hat{\Omega}|\lambda_c, \lambda_z)$, instead, is the redshift prior, which depends on both the source redshift z and the GW localization $\hat{\Omega}$. This term incorporates information about the distribution of the host galaxies and the BBH merger rate across cosmic time. It can be expressed as:

$$p(z, \hat{\Omega}|\lambda_c, \lambda_z) = \frac{p_{gal}(z, \hat{\Omega}|\bar{\lambda_c})p_{rate}(z|\bar{\lambda_z})}{\int p_{gal}(z, \hat{\Omega}|\bar{\lambda_c})p_{rate}(z|\bar{\lambda_z})\mathrm{d}z\mathrm{d}\hat{\Omega}} \tag{2.10}$$

where $p_{gal}(z, \hat{\Omega})$ is the probability of finding a galaxy at redshift z and localization $\hat{\Omega}$, while $p_{rate}(z)$ describes the probability that a galaxy at redshift z hosts a GW event.

The term $p_{\mathrm{rate}}(z|\bar{\lambda_z})$ characterizes the redshift evolution of the BBH merger rate, and it can be written as:

$$p_{rate}(z|\lambda_z) \propto \frac{\psi(z|\bar{\lambda_z})}{(1+z)} \tag{2.11}$$

where $\psi(z|\bar{\lambda}_z)$ describes the merger rate as a function of redshift and the factor $(1+z)^{-1}$ converts time intervals from the source frame to the detector frame. Under the assumption that the Madau–Dickinson model describes the evolution of the merger rate density with redshift:

$$\psi(z|\lambda_z) = \frac{(1+z)^{\gamma}}{1 + \left(\frac{1+z}{1+z_p}\right)^{\gamma+\kappa}} \tag{2.12}$$

where z_p is the redshift at which the rate reaches its peak, γ describes the rising slope at low redshift, and κ characterizes the declining slope beyond the peak (Madau P. and Dickinson M., 2014).

The galaxy term p_{gal} in Eq. (2.10) represents the probability of finding a galaxy in the catalog. In the ideal case of full completeness, this reduces to the catalog term, i.e. $p_{\mathrm{gal}}(z, \hat{\Omega}, \mathcal{G}|\lambda_c) = p_{\mathrm{cat}}(z, \hat{\Omega}, \mathcal{G}|\lambda_c)$, where $\mathcal{G}$ is the explicit dependencies on the host galaxy properties that influence the BBH formation, and can be computed as the weighted mean over the contribution of each galaxy of the catalog:

$$p_{cat}(z, \hat{\Omega}, \mathcal{G}\lambda_c) = \frac{\sum_g w_g p(z, \mathcal{G}|z_g, \lambda_c)\delta(\hat{\Omega} - \hat{\Omega}_g)}{\sum_g w_g} \tag{2.13}$$

where the weight w_g corresponds to a galaxy-specific property assumed to correlate with the BBH merger rate, for instance luminosity, stellar mass, or star formation rate.
In GW cosmology, a catalog is considered complete when it contains all the galaxies that

can host GW sources. In general, each observed catalog is incomplete because of the presence of selection effects. Therefore, the p_{gal} term can be expressed as the sum of two contributions:

$$p_{gal}(z, \hat{\Omega}, \mathcal{G}|\lambda_c) = f_{\mathcal{R}} p_{cat}(z, \hat{\Omega}, \mathcal{G}|\lambda_c) + (1 - f_{\mathcal{R}}) p_{miss}(z, \hat{\Omega}, \mathcal{G}|\lambda_c) \tag{2.14}$$

where p_{cat} is the probability of finding a galaxy included in the catalog, and p_{miss} accounts for galaxies that are missing due to catalog incompleteness and observational selection effects.
The term $f_{\mathcal{R}}$ quantifies the volume-weighted completeness fraction, and it is expressed as (Finke A., et al., 2021; Borghi N., et al., 2025):

$$f_{\mathcal{R}} = \int_0^{z_{max}} P_{comp}(z, \hat{\Omega}, \mathcal{G}) p_{bkg}(z, \mathcal{G}|\lambda_c) \mathrm{d}z \tag{2.15}$$

where $P_{comp}(z, \hat{\Omega}, \mathcal{G})$ is the catalog completeness function, describing the probability that a potential host galaxy is included in a catalog, and $p_{bkg}(z, \mathcal{G}|\lambda_c)$ is the background probability distribution of potential host galaxies.
The catalog completeness function is computed as the ratio between the catalog number density and the theoretical number density of galaxies we expect to find in a given region. It depends on redshift, localization, and galaxies' properties $\mathcal{G}$:

$$P_{comp}(z, \hat{\Omega}, \mathcal{G}) = min\left(\frac{n_{cat}(z, \mathcal{G})}{n_{theo}(z, \mathcal{G})}, 1\right) \tag{2.16}$$

The theoretical number density distribution can be obtained, for instance, by integrating a given Schechter function $\Phi(M, z)$ over stellar masses M (Schetcher P., 1976; Pozzetti L., et al., 2010; Weaver J. R., et al., 2023).
With the theoretical number density distribution, we can define also the background probability distribution:

$$p_{bkg}(z|\lambda_c, \mathcal{G}) \propto n_{theo}(z, \mathcal{G}) \frac{\mathrm{d}V_c}{\mathrm{d}z} \tag{2.17}$$

where $\mathrm{d}V_c/\mathrm{d}z$ is the cosmological comoving volume.
Finally, the last term to explore present in Eq. (2.14) is p_miss, the probability of missing a galaxy in the catalog that can be expressed as:

$$p_{miss}(z, \hat{\Omega}, \mathcal{G}|\lambda_C) = \frac{1 - P_{comp}(z, \hat{\Omega}, \mathcal{G})}{1 - f_{\mathcal{R}}} p_{bkg}(z, \mathcal{G}|\lambda_c) \tag{2.18}$$

To evaluate the missing probability p_{miss}, we use the homogeneous completion method, in which we assume that the missing galaxies are homogeneously distributed with respect to the background (see Finke A., et al., 2021 for the derivation).

Combining all the equations from Eq. (2.8) to (2.15), we find the full form for the Likelihood as:

$$p(d^{GW}|\lambda) \propto \frac{1}{\xi(\lambda)^{N_{ev}}} \prod_{i=1}^{N_{ev}} \int \mathrm{d}z \; \mathrm{d}\hat{\Omega} \; \mathcal{K}_{GW,i}(z, \hat{\Omega}|\lambda_c, \lambda_m) p_{gal}(z, \hat{\Omega}|\lambda_c) \frac{\psi(z|\lambda_z)}{1+z} \tag{2.19}$$

where the GW kernel $\mathcal{K}_{GW,i}(z, \hat{\Omega}|\lambda_c, \lambda_m)$ is the marginalization of the posterior over the masses m_1 and m_2, remaining dependent on redshift z and on GW localization $\hat{\Omega}$:

$$\mathcal{K}_{GW,i}(z, \hat{\Omega}|\lambda_c, \lambda_m) = \int\int \frac{p(\theta_i|d_i^{GW}, \lambda_c)}{\pi(\theta^{det})} \left|\frac{\mathrm{d}\theta_i}{\mathrm{d}\theta_i^{det}}\right| p(m_1, m_2|\lambda_m)\ \mathrm{d}m_1\ \mathrm{d}m_2 \tag{2.20}$$

By making explicit the Jacobian factors arising from the conversion between detector and source frames and by specifying the priors on θ^{det}, with flat priors for m_1, m_2, and $\pi(d_L) \propto d_L^2$, the GW kernel takes the form:

$$\mathcal{K}_{GW,i}(z, \hat{\Omega}|\lambda_c, \lambda_m) = \int\int \frac{p(z, m_1, m_2, \hat{\Omega}|d_i^{GW}, \lambda_c)}{\pi(d_L)\pi(m_1^{det})\pi(m_2^{det})} \frac{p(m_1, m_2|\lambda_m)}{\frac{\mathrm{d}d_L}{\mathrm{d}z}(z|\lambda_c)(1+z)^2}\mathrm{d}m_1\ \mathrm{d}m_2 \tag{2.21}$$

2.3 Parametric mass function models

In this Thesis, particular attention is placed on the mass function models, as one of the main goals is to develop a semi-parametric description of the mass function. Within `CHIMERA`, several parametric models are already implemented to describe the primary mass function distribution.
According to the population term introduced in Eq. (2.9), the mass function models are used to constrain the population prior on the source-frame masses $p(m_1, m_1|\boldsymbol{\lambda}_m)$.
In particular, this latter term can be factorized as:

$$p(m1, m2|\lambda_m) = p(m1|\boldsymbol{\lambda}_m)p(m2|m1, \lambda_m)$$

where $p(m1|\boldsymbol{\lambda}_m)$ is the primary mass distribution probability conditioned on the mass hyper-parameters λ_m and $p(m2|m1, \lambda_m)$ is the secondary mass distribution probability conditioned on the primary mass and the mass hyperparameters $\boldsymbol{\lambda}_m$.
In `CHIMERA`, the primary mass distribution for BBH is modeled using four different parametric models, listed below. All these methods are widely used in literature and their development is discussed in detail in Section (1.3). All these models share common characteristics: a power law behavior across the mass range and a main peak around $\sim 35 - 40\ \mathrm{M}_\odot$, probably due to the effects of pulsational pair-instability supernovae (see Section (1.3)).
In addition to these two main features, each mass function model introduces other features related to observational evidence observed up to O4. Nevertheless, these models rely on predefined distribution forms, preventing the data from fully shaping the mass distribution in a flexible way.

2.3.1 Truncated Power Law

The Truncated Power Law primary mass distribution is the simplest model, as it implements a pure power-law with sharp cutoffs at low and high masses, without explicitly accounting for the observed peaks:

$$p(m_1|\alpha, m_{\mathrm{low}}, m_{\mathrm{high}}) \propto \begin{cases} m_1^{-\alpha} & m_{\mathrm{low}} \leq m_1 \leq m_{\mathrm{high}} \\ 0 & \text{otherwise} \end{cases} \tag{2.22}$$

where α is the power law slope, m_{low} is the lowest mass bound, and m_{high} is the upper mass bound for the primary black hole mass.
The secondary mass function distribution, instead, follows a second power law with slope β, defined between m_{low} and m_1:

$$p(m_2|m_1, \lambda_m) \propto \begin{cases} m_2^{\beta} & m_{low} \leq m_2 \leq m_1 \\ 0 & \text{otherwise} \end{cases} \tag{2.23}$$

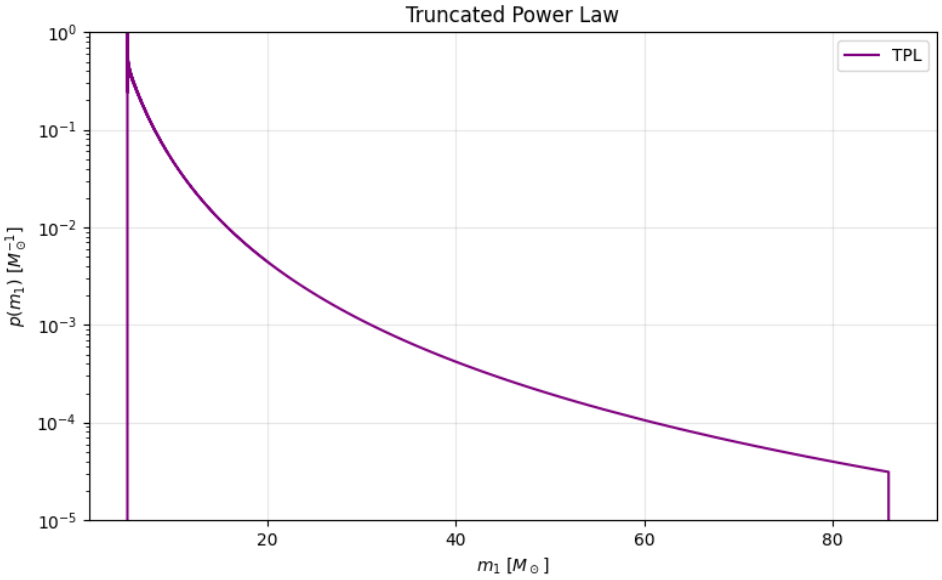

Figure 2.1: Truncated Power Law for the primary mass distribution.

The importance of this model derives from its role as a fundamental model for the development of the other parametric models.

2.3.2 Broken Power Law

The last primary mass model distribution already included in `CHIMERA` is the Broken Power Law (`BPL`). This model accounts for the $\sim 35 - 40$ $M_\odot$ peak not by introducing a Gaussian component, but by introducing a discontinuity in the power-law slope.
The Broken Power Law primary mass distribution is:

$$p(m_1|\alpha_1, \alpha_2, b, \delta_m, m_{low}, m_{high}) \propto \begin{cases} m_1^{-\alpha_1} S(m_1|m_{low,\delta_m}) & m_{\text{low}} \leq m_1 \leq m_{\text{break}} \\ m_1^{-\alpha_2} S(m_1|m_{low,\delta_m}) & m_{\text{break}} \leq m_1 \leq m_{\text{high}} \\ 0 & \text{otherwise} \end{cases} \tag{2.24}$$

where $m_{break} = m_{low} + b(m_{high} - m_{low})$ is the mass where the slope changes and $S(m_1|m_{low,\delta_m})$ is a smoothing function (Abbott R., et al., 2021).
The mass hyper-parameters that enters in the Broken Power Law are listed in Table (2.1).

Parameter	Description
α_1	Power law slope for masses below m_{break}
α_2	Power law slope for masses above m_{break}
m_{low}	Minimum mass of the power law
m_{high}	Maximum mass of the power law
b	Fraction of the distance between m_{low} and m_{high} where the slope changes
δ_m	Smoothing scale term at low masses

Table 2.1: Parameters that enter in the Broken Power Law (BPL) model for the primary mass distribution.

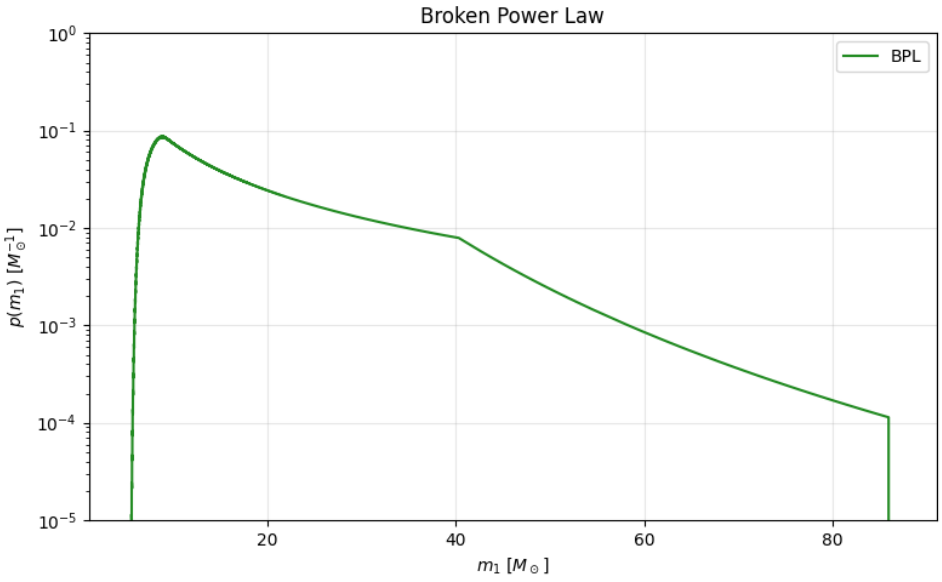

Figure 2.2: Broken Power Law for the primary mass distribution.

2.3.3 Power Law + Peak

The Power Law + Peak (PLP) model extends the Truncated Power Law model by adding a Gaussian peak and a Smoothing function. The Gaussian peak is included specifically to account for the observed feature around $\sim 35-40$ $M_\odot$ (see Section 1.3). The Smoothing function is introduced to avoid a sharp cutoff at low masses.
The Power Law + Peaks primary mass distribution is:

$$p(m_1|\alpha, \delta_m, m_{low}, m_{high}, \mu_g, \sigma_g, \lambda_g) \propto [(1-\lambda_g)P(m_1) + \lambda_g G(m_1)]S(m_1) \tag{2.25}$$

where $P(m_1)$ is the power law component; $G(m_1)$ is the normalized Gaussian distribution with mean μ_g and width σ_g and $S(m_1)$ is the smoothing function, which takes values in the range [0, 1] for masses between m_{low} and $m_{low} + \delta_m$ (Abbott R., et al., 2021).
The hyper-parameters that characterize the Power Law + Peak are listed in Table (2.2):

Parameter	Description
α	Spectral index for the power law
m_{low}	Minimum mass of the power law
m_{high}	Maximum mass of the power law
μ_g	Mean of the Gaussian peak
σ_g	Width of the Gaussian peak
λ_{peak}	Fraction of the BBH in the Gaussian peak
δ_m	Smoothing scale term at low masses

Table 2.2: Parameters that enter in the Power Law + Peak (PLP) model for the primary mass distribution.

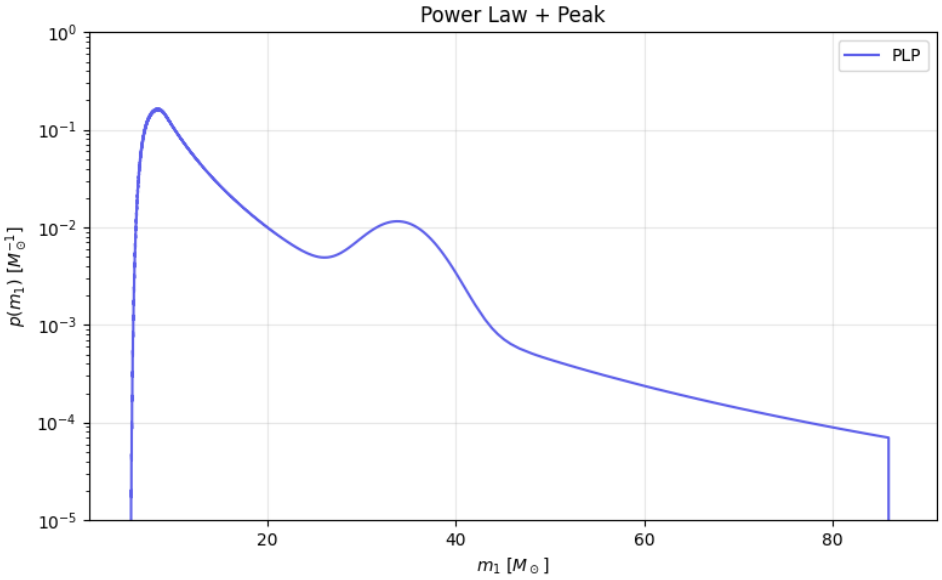

Figure 2.3: Power Law + Gaussian Peak for the primary mass distribution.

2.3.4 Power Law + Two Peaks

The Power Law + Two Peaks (PL2P) model extends the PLP distribution by introducing an additional Gaussian peak. This makes the model more structured to capture another feature in the mass function. When applied to the data, the inferred peaks are found around 10 $M_\odot$ and 35 $M_\odot$ (see Section 1.3).
The Power Law + Two Peaks primary mass distribution is:

$$\begin{aligned} &p(m_1|\alpha, \delta_m, m_{low}, m_{high}, \mu_{g1}, \sigma_{g1}, \mu_{g2}, \sigma_{g2}, \lambda_g, \lambda_{g1}) \propto \\ &\quad [(1-\lambda_g)P(m_1) + \lambda_g\lambda_{g1}G_1(m_1) + \lambda_g(1-\lambda_{g1})G_2(m_1)]S(m_1) \end{aligned} \tag{2.26}$$

where $P(m_1)$ is the power law component, $G(m_1)$ and $G(m_2)$ represent the normalized gaussian peaks for the first and second peaks, respectively, and $S(m_1)$ is the smoothing term.
The mass hyper-parameters that characterize the Power Law + Two Peaks are listed in Table (2.3).

Parameter	Description
α	Spectral index for the lower law
m_{low}	Minimum mass of the power law
m_{high}	Maximum mass of the power law
μ_{g1}	Mean of the lower-mass Gaussian peak
σ_{g1}	Width of the lower-mass Gaussian peak
μ_{g2}	Mean of the upper-mass Gaussian peak
σ_{g2}	Width of the upper-mass Gaussian peak
λ_g	Fraction of the BBH in the high-mass Gaussian peak
λ_{g1}	Fraction of the BBH in the lower-mass Gaussian peak
δ_m	Smoothing scale term at low masses

Table 2.3: Parameters that enter in the Power Law + Two Peak (PL2P) model for the primary mass distribution.

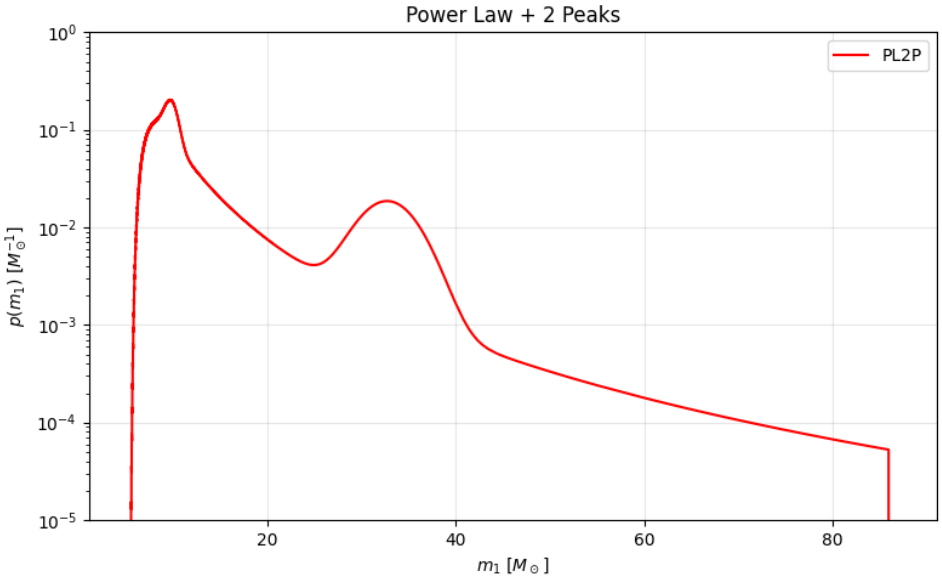

Figure 2.4: Power Law + Two Gaussian Peak for the primary mass distribution.

2.4 Implementation in CHIMERA

In this Thesis, we introduce a new semi-parametric model for the BBH primary mass distribution in `CHIMERA`. The entire analysis presented in this work was carried out using `CHIMERA` (Combined Hierarchical Inference Model for Electromagnetic and gRavitational wave Analyses), a novel and independent Python code that allows to jointly perform inference analysis on cosmological and population hyper-parameters by solving the Eq. (2.19). [1] (Borghi N., et al., 2024; Borghi N., et al., 2025; Tagliazucchi M., et al., 2025). The last release of `CHIMERA` is fully developed in JAX, which allows hyperparameters vectorization and GPU acceleration (Tagliazucchi M., et al., 2025).
Figure (2.5) shows the entire workflow of the `CHIMERA` code, presenting the main Python

[1] Available at: https://github.com/CosmoStatGW/CHIMERA

modules used to accurately compute Eq. (2.19).

3D GW Kernel The Likelihood computation is technically performed using the kernel density estimation (KDE) to model the GW data posterior. The GW kernel is represented by $\mathcal{K}_{GW,i}(z, \hat{\Omega}|\lambda_c, \lambda_m)$ in Eq. (2.21). It combines posterior samples of masses and luminosity distance obtained from single event analyses, converted into the source frame. In fact, the kernel density estimation represents the marginalized posteriors over the masses and luminosity distance assuming as priors for detector frame parameters:

$$\pi(m_1^{det}) = \pi(m_2^{det}) = 1 \qquad \pi(d_L) \propto d_L \tag{2.27}$$

The new release of `CHIMERA` provides three different KDE algorithms:

- *3D kernel*: already implemented in the first release of `CHIMERA`. It consists in dividing the 2D localization area $\mathrm{d}\hat{\Omega}_i$ into N^i_{pix} equal area pixels and discretizing the redshift interval $\mathrm{d}z_i$ into N_z equally spaced points. Finally, the two grids are combined, resulting in a 3D grid (R.A., DEC, z). `CHIMERA 2.0` improves the integration process by evaluating the KDE only over the portion of the redshift grid that contains posterior redshift samples.

- *many-1D kernel*: the pixelization of the localization area $\mathrm{d}\hat{\Omega}$ and the discretization of the redshift range remain the same as in the method *3D kernel*. However, instead of a single 3D density over (RA, DEC, z), this algorithm computes the 1D redshift distributions for each pixel by marginalizing over the sky localization. The GW kernel within each pixel is obtained by multiplying the 1D KDE, which describes the redshift distribution in that pixel, by a normalization factor that accounts for the probability of the source being located in that region of the sky. Also in this case, the KDEs are evaluated only on the portions of the redshift grids containing redshift samples.

- *single 1D kernel*: this algorithm assumes independence between sky position and redshift. Instead of computing a 1D for each redshift pixel (as done in *many 1D kernel*), all redshift samples are computed with a weighted 1D KDE. Then, the GW kernel in each pixel is obtained by multiplying the redshift distribution by a normalization factor dependent on the localization area.

In this Thesis, we employ the *many-1D kernel*, which has proven to be both efficient and accurate given the LVK data analyzed in this study (Tagliazucchi M., et al., 2025).

Selection effects The selection effects account for observational bias introduced by the limited detector sensitivity. The term related to the selection effects is defined in Eq (2.7) and in `CHIMERA`, is calculated using a set of N_{inj} injections, simulated events with parameters $\{\theta_j^d\}_{j=1}^{N_{inj}}$. The integral in Eq. (2.7) is estimated by performing the Monte Carlo integration over the injections:

$$\xi(\lambda) \approx \frac{1}{N_{inj}} \sum_{j=1}^{N_{det}} \frac{p_{pop}(\theta_j|\lambda)}{(1+z_j^3)\left|\frac{\partial d_L}{\partial z}(z_j;\lambda_c,\lambda_m)\right|} \equiv \frac{1}{N_{inj}} \sum_{j=1}^{N_d et} s_j \tag{2.28}$$

where N_{inj} and N_{det} are the numbers of injections and detected events, respectively. The numerical stability required by the Monte Carlo summation depends on the effective number of injections, defined as:

$$N_{eff}^{inj} = \left[\sum_{j=0}^{N_{det}} s_j\right]^2 \times \left[\sum_{j=0}^{N_{det}} s_j^2 - \frac{1}{N_{inj}}\left(\sum_{j=0}^{N_{det}} s_j\right)^2\right]^{-1} \tag{2.29}$$

must be larger than $5N_{det}$.

Galaxy term The galaxy term defined in Eq. (2.14) combines the contributions from the catalog term and the completeness term. In the ideal case of a complete galaxy catalog, the entire information originates from the catalog term. In the more realistic scenario of an incomplete catalog, both terms play significant roles. Conversely, when no galaxy catalog is available, all the information is derived from the completeness term and the background galaxy distribution.
The catalog term in Eq. (2.13) can be precomputed pixel by pixel in the redshift grid to increase the efficiency of the computation. Once the catalog term is computed, the completeness function in Eq. (2.16) is evaluated in N_{masks} masks, which group together healpix pixels using a k-means algorithm, according to the number of galaxies inside each pixel. To compute P_{comp} it is required the computation of the background galaxy distribution, which is obtained by integrating a given Schechter mass function over the host mass range.

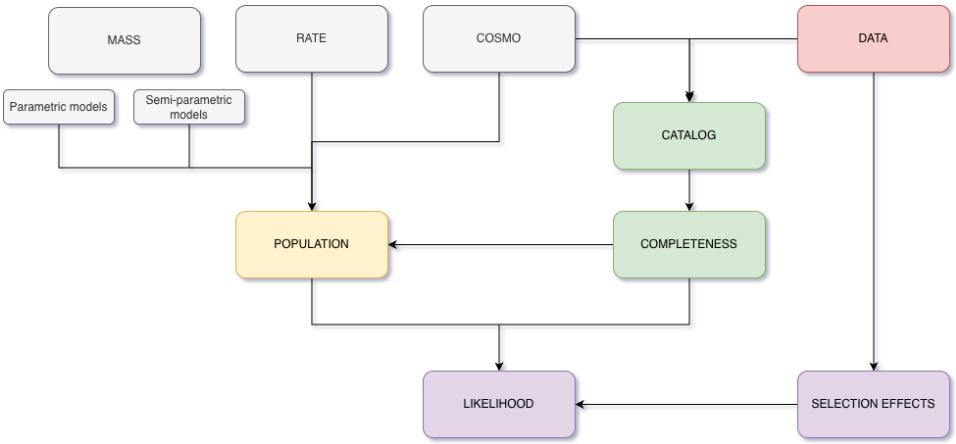

Figure 2.5: Workflow of `CHIMERA` in the *jax* branch. Modules shown in grey contain the model functions, the yellow module defines the population model that combines all grey modules, the red module handles input and output data files, the green modules compute probability corrections and the violet modules contain the functions to compute the whole likelihood with the correct normalization due to selection effects.

Chapter 3

Development and validation of semi-parametric mass function

Each new observing run of the LIGO-Virgo-KAGRA Collaboration yields an increasing number of detected GWs events from BBH mergers, enabling a deeper understanding of the astrophysical population and providing tighter cosmological constraints.
The features already identified in the mass function, such as peaks and gaps discussed in Section (1.3), provide important clues about its underlying structure and the underlying astrophysical population. When the number of detected events was still limited, the mass function was typically constrained using parametric models that imposed a specific functional form. For this reason, several parametric models were implemented within the `CHIMERA` framework, such as those discussed in Chapter 2: `TPL`, `BPL`, `PLP`, `PL2P`.
An improvement is represented by extending the mass distribution models, including semi- and non- parametric approaches, because of the difficulty of accurately recovering a priori the functional form of the data distribution when using parametric models. In fact, the assumed functional form may fail to capture additional structures in the data. While fully non-parametric models offer maximum flexibility, they are almost impractical with the current dataset sizes. In fact, with limited data, such models are inclined to overfitting, producing spurious features dominated by statistical fluctuations rather than by astrophysical distribution. Therefore, a semi-parametric approach provides a compromise: it allows sufficient flexibility to capture features while enforcing a power law trend. However, attempts to construct flexible data-driven models up to the GWTC-3 dataset did not provide additional information beyond that already known with parametric models. As shown by Toubiana A., et al. (2023), applying a Power law + Spline model to the GWTC-3 dataset generated apparent new features that were dominated by statistical fluctuations and therefore not physically meaningful.
With the recent release fo GWTC-4.0 dataset, which adds 137 BBH events with respect to GWTC-3, the larger sample size finally makes it statistically meaningful to explore the BBH population using more flexible semi-parametric approach. In the Thesis, the `CHIMERA` framework is therefore extended by introducing, in addition to the already existing parametric mass models, a semi-parametric model based on a smoothed truncated power law combined with a spline: the `PLS` model.
This Chapter presents the `PLS` model from a technical and methodological perspective, providing the necessary context before applying it to mock and real data.

3.1 Spline Function and Spline Basis

A spline function of order κ is a piecewise polynomial of polynomials of order κ, connected together at knots' position defined in the domain. Instead of fitting one single polynomial, splines divide the domain into smaller intervals separated by specific points called knots. Within each interval, a different polynomial is used, and the pieces are smoothly connected at the knots. The order of a spline is related to the degree of smoothness and continuity required at the knots' location. A spline function of order κ is built to ensure that the function itself and its first $\kappa - 1$ derivatives are continuous at each knot.
The spline function provides an efficient way to interpolate smooth functions from a finite sampling of knots' heights.
Given a set of knots $t = \{t_0, t_1, ...t_{i+\kappa}\}$, any spline function of order κ can be written as a linear combination of spline basis:

$$S(x) = \sum_{j=1}^{N_{knots}} c_j B_{j,\kappa}(x,t) \tag{3.1}$$

where N_{knots} is the chosen number of knots in which the domain is divided, $B_{j,\kappa}(x,t)$ are the B-spline basis functions with degree κ and defined on the knots positions t, and c_j are the spline coefficients that controls and weights the contribution of the corresponding basis function to the overall perturbation.
The basis functions are generally constructed using the Cox-de Boor recursion formula, starting from the base case ($\kappa = 0$) and constructing the base components of higher orders (de Boor, 1978). The zero-order relation is:

$$B_{i,0}(x|t) = \begin{cases} 1 & t_i \leq x \leq t_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

The recursive formula is:

$$B_{i,\kappa+1}(x|t) = \omega_{i,\kappa}(x|t) B_{i,\kappa}(x|t) + [1 - \omega_{i+1,\kappa}(x|t)] B_{i+1,\kappa}(x|t) \tag{3.3}$$

where ω is a weighted factor defined as:

$$\omega_{i,\kappa}(x|t) = \begin{cases} \frac{x - t_i}{t_{i+\kappa} - t_i} & t_{i+\kappa} \neq t_i \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

The spline coefficients c_j are determined from the N_{knots} resulting in a total of $N_{coeffs} = N_{knots} + \kappa - 1$. Each coefficient corresponds to the value of the $S(x)$ at its associated Greville abscissa. These points in the m_1 domain are computed as the averages of $\kappa + 1$ consecutive knots, and they indicate where the underlying power law can be modified according to the value of c_j.

3.2 A more versatile mass function model

In this work, I extend the `CHIMERA` population models by introducing the Power Law + Spline (`PLS`), a semi-parametric representation of the primary mass distribution that

superimposes a spline over a smoothed truncated power law (TPL). The TPL, introduced in Section (2.3.1), is here smoothed at low masses through the inclusion of a smoothing factor δ_m.
In this Thesis, the adopted spline is cubic (degree $\kappa = 3$), ensuring continuity up to the second derivative, which is physically reasonable.
This choice allows the spline component to freely adapt to the data, enabling the mass distribution to exhibit features such as peaks or gaps without imposing assumptions on their number, position or shape, unlike parametric components such as Gaussian peaks. This approach allows the model to closely follow the primary mass distribution and to adopt the best-fitting shape, improving the inference of astrophysical and cosmological hyperparameters.
The Power Law + Spline primary mass distribution is defined as:

$$p(m_1|\alpha, m_{low}, m_{high}, \delta_m, \{c_i\}) \propto p(m_1|\alpha, m_{low}, m_{high}, \delta_m) e^{f(m_1|\{c_i\})} \tag{3.5}$$

where $p(m_1|\alpha, m_{low}, m_{high}, \delta_m)$ is the underlying smoothed TPL primary mass distribution. This is multiplied by the exponential of the perturbation function modeled as a cubic spline $f(m_1|\{c_i\})$, which is constructed as a linear combination of the B-spline basis functions and of the spline coefficients $\{c_i\}$. The exponential form in Eq. (3.5) is used to handle the wide range of order of magnitudes of the mass function.
The parameters of the PLS model are summarized in Table (3.1).

Parameter	Description
α	Spectral index for the power law
m_{low}	Minimum mass of the power law
m_{high}	Maximum mass of the power law
δ_m	Smoothing scale term at low masses
$\{c_i\}$	Spline coefficient at the $i-$th Greville abscissa

Table 3.1: Parameters that enter in the Power Law + Spline (PLS) model for the primary mass distribution.

In the construction of the PLS, the first step is to define the knots positions and the spline basis functions. In the CHIMERA framework, knots can be placed in the mass range $[m_{\rm low},\ m_{\rm high}]\ M_\odot$, defined by the priors on m_low and m_{high}, according to three different spacings: *linear*, *logarithmic*, or *quantile*. Adopting any of these three spacing methods, the knots' positions can be determined knowing only their total number. In this Thesis, knots are placed in logarithmic space, in order to remain consistent with most of literature works, as Edelman B., et.(2022) and Edelman B., et al. (2023), and to better map the low mass regime.
The B-spline basis is then computed using the De Boor recursive formula in Eq. (28), with each basis anchored to the knots' positions and evaluated on the spline grid. In fact, the B-spline basis are defined by a set of N_{knots} knots placed in the mass domain, which determine the intervals over which the spline can vary its curvature.
Once the B-spline basis is constructed, a vector of spline coefficients, of length $N_{coeffs} = N_{knots} + \kappa - 1$, is initialized to zero. An illustrative plot showing how the power law is perturbed and how knots and spline coefficients enter into the construction of the PLS is reported in Figure (3.1).

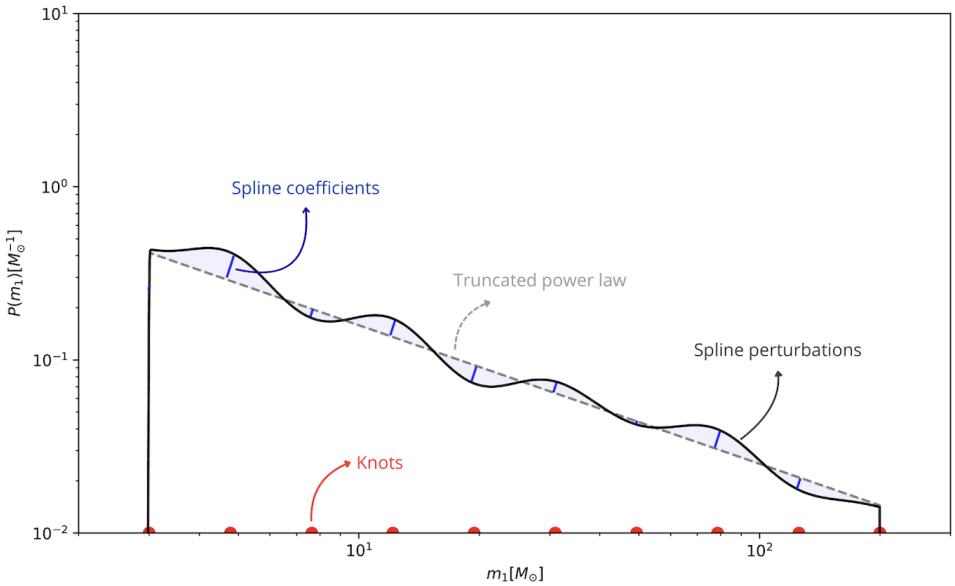

Figure 3.1: Descriptive sketch of the spline perturbation applied to a truncated power law, without the smoothing parameter ($\delta_m = 0$). The knot positions (in red) are shown along the x-axis, while the spline coefficients (in blue) indicate the amplitude of the perturbations from the underlying truncated power law (dashed line). The overall model is the black solid line.

3.2.1 Computational implementation

The computational implementation of the `PLS` model inside `CHIMERA` requires some other adjustments that are not included in the theoretical section. The `PLS` model is implemented in Python in the new code release `CHIMERA 2.0`, fully implemented in the `JAX` framework. `JAX` is a Python library which enables high performance calculations and fast array computations in the `Numpy` syntax. The `PLS` model is therefore built using `equinox` to ensure compatibility with `JAX`, as discussed in Tagliazucchi M., et al. (2025). The first computational adjustment, used when building the knot array through the `setup_spline` function, is the repetition of the knots at the domain boundaries. The first and last knots are in fact repeated $\kappa + 1$ times (four times in the cubic case), to ensure that the spline starts and ends exactly on the `TPL`, and to allow for the correct definition of the Greville abscissae at the boundaries.
The `setup_spline` methods is therefore the main method for building the `PLS`. It computes the knots, the spline grid (the grid in mass where the spline basis values are evaluated) and the spline basis functions using the De Boor formula itself.
Implementing the `PLS` class also requires modifying the `generate_dict` function. The function takes an array of parameters and a list of their corresponding names, and produce a Python dictionary with the appropriate keys and values. In fact, unlike the other models, the `PLS` does not contain only scalar parameters, but it also includes an array of spline coefficients. Because fo this, in this Thesis work, the `generate_dict` is updated to recognize all the spline coefficients, sort them, and group them in the dictionary into a single array under the key "spline_coefficients". This modification makes possible to handle the `PLS` model like all the others, without other changes in the code and ensure that the coefficients are passed correctly to the likelihood function

`hyperlike.compute_all`. All these modifications have been added to the official version of the pipeline, available on CHIMERA.

3.3 Validation of the semi-parametric mass function

In this section, we evaluate how well the PLS model can reproduce and interpolate theoretical parametric mass models before applying it to the data. In fact, the main advantage of first validating the PLS on known theoretical models, such as PLP and PL2P, is that we know their shape a priori, making it easier to identify any problems.
The two theoretical models, PLP and PL2P, are constructed by setting their mass parameters to the median posterior values of the hyperparameters inferred with `icarogw` (Data distribution of GWTC-4.0, Zenodo). These inferences are performed in spectral mode for 137 BBHs from the GWTC-4.0 dataset, assuming the ΛCDM cosmology for the PLP and MLTP models. This approach ensures that our theoretical models are aligned with the observed data, providing the most accurate possible validation. The median values used to construct the two models are listed in Table (3.2).

PLP						
α	δ_m	m_{low}	m_{high}	μ_g	σ_g	λ_{peak}
3.6	3.3	5.3	94.4	28.6	5.1	0.05

PL2P									
α	δ_m	m_{low}	m_{high}	$\mu_{g,1}$	$\sigma_{g,1}$	$\mu_{g,2}$	$\sigma_{g,2}$	λ_g	λ_{g1}
2.9	4.8	4.6	86.3	9.7	0.74	30.6	6.3	0.85	0.38

Table 3.2: Median posterior values of the mass hyperparameters for the PLP and PL2P models, inferred using `icarogw`.

In this validation, the knot positions used to represent the theoretical PLP and PL2P models are logarithmically distributed between $m_{low} = 5.1$ $M_\odot$ and $m_{high} = 87$ $M_\odot$. Figure (3.2) shows the resulting knot positions for the knots sets used in this analysis, which are uniformly distributed in logarithmic space. The numbers of knots adopted are $N_{knots} = [5, 10, 15, 20]$.

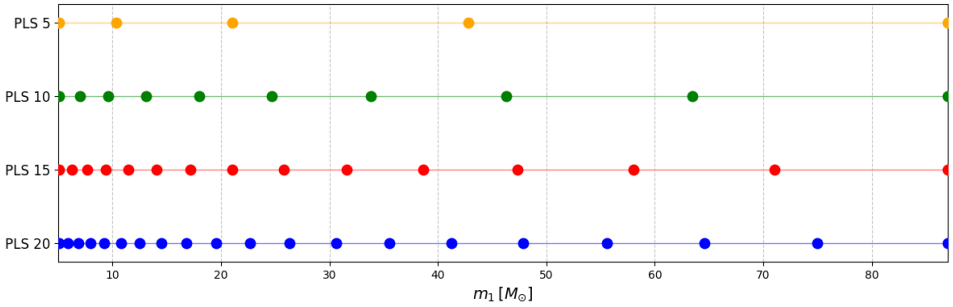

Figure 3.2: Positions of logarithmically spaced knots between $m_{low} = 5.1$ $M_\odot$ and $m_{high} = 87$ $M_\odot$ for different knots sets $N_{knots} = [5, 10, 15, 20]$. The first and last knots are fixed at the boundaries.

Once defined the knots positions and computed the spline basis functions, the internal spline coefficients are determined with an optimization process that minimize the discrepancy between the PLS and the theoretical models. The loss function to be minimized is defined as the root mean square of the relative difference between the two normalized distribution:

$$L = \sqrt{\left\langle \left(\frac{p_{\mathrm{PLS}} - p_{\mathrm{theo}}}{p_{\mathrm{theo}}} \right)^2 \right\rangle} \tag{3.6}$$

where p_{theo} denotes the chosen primary mass distribution, either PLP or PL2P.
In this analysis, the only parameters of the PLS left free to vary are the spline coefficients $\{c_i\}$. The parameters α, m_{low}, m_{high}, and δ_m, which define the contribution of the power-law component, are instead kept fixed at the default values shown in Table (3.2). The goal of this procedure is to verify that the PLS model is sufficiently flexible to reproduce standard parametric population models such as PLP and PL2P. Demonstrating that the PLS is enough flexible to accurately recover these reference models, ensures that the PLS is capable of capturing the main features reproduced by parametric models, and can therefore be applied to data.
At the beginning of the optimization, all spline coefficients are set to zero. The fit is performed using the `scipy.optimize.minimize` function, which iteratively adjusts the spline coefficients to minimize the loss function in Eq.(3.6), finding the PLS configuration that best reproduces the theoretical distribution.

3.3.1 Comparison with a standard parametric function

To validate the PLS over the theoretical primary mass distribution PLP in absence of data, the optimization procedure, described in Eq. (3.6), has been performed for each set of internal knots $N_{knots} = [5, 10, 15, 20]$. The PLS representation of the PLP is shown on the left panel of Figure (3.3) and the corresponding accuracy is reported in Table (3.3).

N_{knots}	**Mean difference (%)**	**Maximum difference (%)**
5	13.09%	38.72%
10	1.18%	6.03%
15	0.32%	5.26%
20	0.30%	5.26%

Table 3.3: Average and maximum percentage difference between PLS and the default PLP model, with logarithmically spaced knots in the mass range [5.1 $M_\odot$, 87 $M_\odot$] for $N_{knots} = [5, 10, 15, 20]$.

By studying how the PLS represents the PLP, it is evident that the accuracy improves as the number of knots increases. This is because a larger number of knots allows for a finer sampling of the domain, so that more knots lie within the mass range of the peak. The most accurate representation is obtained with 20 knots, with an average percentage difference of 0.30% and a maximum percentage difference always below 6%.
Some fluctuations, even in the 20-knots case, are mainly observed before and after the peak, where the PLS must deviate from the power-law component to reproduce the

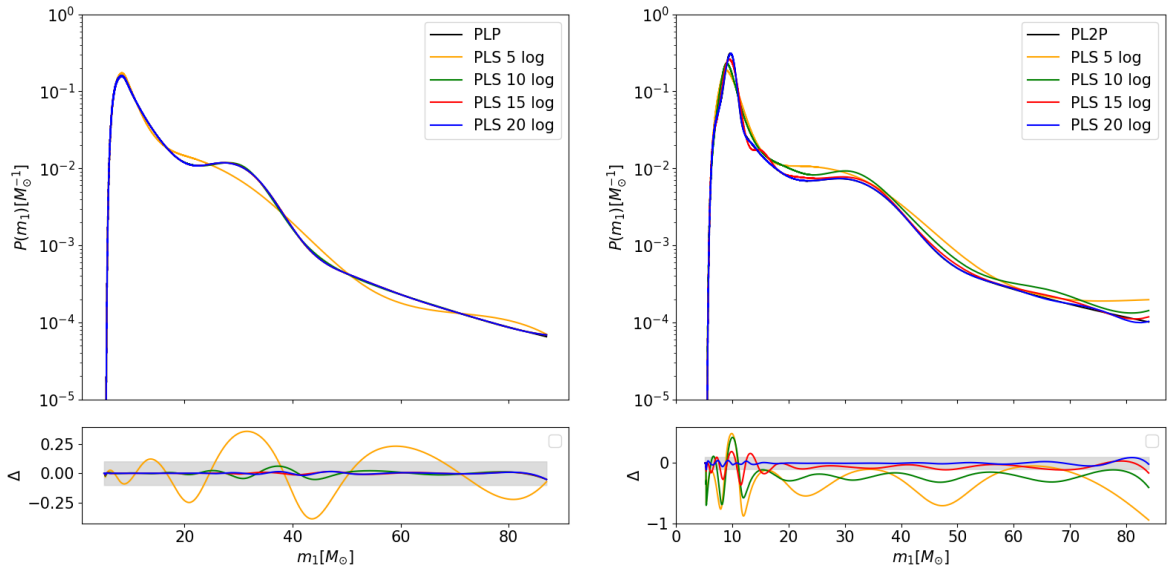

Figure 3.3: The upper panels display the PLS representations, using four different sets of knots ($N_{knots} = [5, 10, 15, 20]$), which are uniformly distributed in logarithmic space, as attempts to reproduce the PLP (*left*) and the PL2P (*right* distributions. The bottom panel shows the relative percentage difference between the PLS and the parametric models. The shaded area represent the 10% accuracy threshold. *Left:* $\Delta = (p_{plp} - p_{pls})/p_{plp}$; *right:* $\Delta = (p_{pl2p} - p_{pls})/p_{pl2p}$.

peak. The discrepancies may also be due to numerical effects associated with the fitting procedure, rather than to an intrinsic lack of flexibility of the PLS model.

3.3.2 Comparison with an advanced parametric function

In the same way as done for the PLP case we also test the PLS model on the theoretical PL2P distribution, which represents the best model to fit the GWTC-4.0 dataset with its double peak structure (Abac A. G., et al., 2025). The corresponding results are shown the the right panel of Figure (3.3), and the relative accuracy for each PLS configuration is reported in Table (3.4).
Even in the case of PL2P fit, we see that increasing the number of knots generally improves how well the model is reconstructed. However, comparing the two reconstructed PLP and PL2P models highlights an important aspect: as the mass model becomes more complex with more features, such as the peak at 10 $M_\odot$ for the PL2P, the PLS struggles to reproduce the shape accurately when the number of knots is limited. In fact, increasing the knots from 10 to 20 leads to only a small and negligible improvement when fitting the smoother PLP model (both reconstructed functions remain below a mean difference of 1.20%). Instead, for the more structured PL2P model the effect is more evident: moving from 10 to 20 knots improves the mean accuracy by more than 23%. This shows that as the number of features increases, more logarithmically spaced knots are needed to capture more details in the mass distribution.

More generally, to achieve higher accuracy in the mass reconstruction can require not only increasing the number of logarithmic knots but also optimizing where they are placed. As the next Section shows, choosing the knot positions rather than spacing them

N_{knots}	**Mean difference (%)**	**Maximum difference (%)**
5	34.75%	94.66%
10	24.72%	70.19%
15	8.67%	36.50%
20	1.44%	10.81%

Table 3.4: Average and maximum percentage difference between PLS and the standard PL2P model, with logarithmically spaced knots in the mass range [5.1 $M_\odot$, 87 $M_\odot$] for $N_{knots} = [5, 10, 15, 20]$.

logarithmically can improve the reconstruction of localized features. This becomes a key aspect of the PLS modeling used throughout this Thesis, especially because the precision with which these structures are recovered directly impacts the cosmological information in the *spectral sirens* framework.

3.4 Exploring the impact of the knots' positions

Logarithmically spaced knots often fail to capture and to reproduce key features, such as peaks, and require a larger number, and thus a higher density, of knots to achieve a good coverage. A more efficient strategy than increasing the number of knots, is to select the knot positions arbitrarily based on the posterior $p(m_1)$ of individual events, allowing better interpolation of the features in the mass function.

Figure (3.4) clearly shows that a PLS model with only 10 chosen knots can reproduce a PL2P with an average accuracy slightly lower than a PLS model with 20 logarithmically spaced knots, but with a maximum percentage difference even smaller than that of the 20 log model. The numerical values are reported in Table (3.5). This demonstrates how choosing the knot positions can improve the performances of the models, without increasing its complexity.
In this case, in absence of data, the knots are selected basing only on the observed structures in the mass function. The knots are therefore placed manually and arbitrarily. In Chapter 5, using the PLS on real data, it becomes clear that, when the underlying distribution is unknown, it is important to develop an analytical method that uses the individual mass posterior of the single GW events to suggest the knot positions.
As shown in Figure (3.4), most of the knots are concentrated around the peak at 10 $M_\odot$, which is the sharper peak in the mass function. Only three knots cover the start, the end and the central part of the smoother peak at 32 $M_\odot$, instead in the mass range between 45 $M_\odot$ and 87 $M_\odot$, where the mass function behaves like a power law, only two knots are needed.

N_{knots}	**Mean difference (%)**	**Maximum difference (%)**
10 chosen	2.13%	9.34%
20 log	1.44%	10.81%

Table 3.5: Average percentage difference between PLS and the default PL2P model, with 10 knots chosen and 20 knots linearly distributed in the mass range $[5.1, 87]$ $M_\odot$.

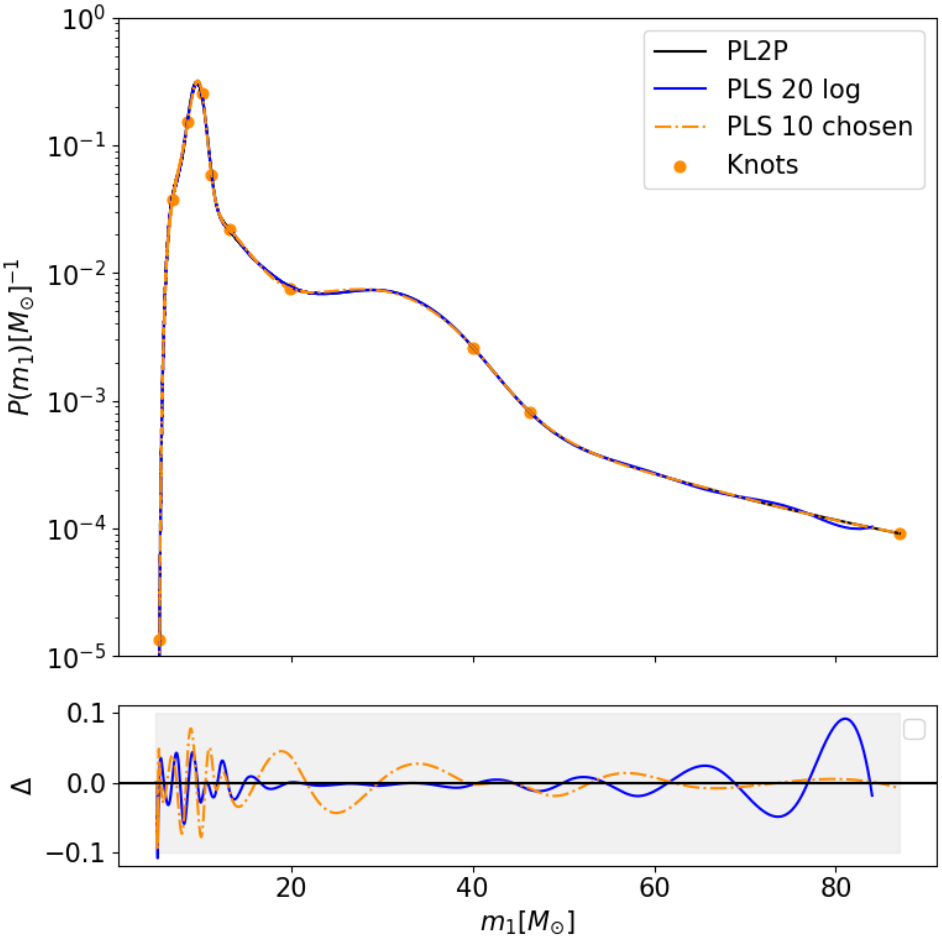

Figure 3.4: The upper panel displays the PLS representations using two different knots distributions: 20 knots logarithmically spaced in mass and 10 knots whose positions have been chosen based on the PL2P distribution. The chosen knots' position are displayed in the plot to ensure good interpolation of the peak. The bottom panel shows the relative difference $((p_{PL2P} - p_{PLS})/p_{PL2P})$ between the PLS and the PL2P models. The shaded area represent the 10% accuracy threshold.

3.5 The cosmological information in the spectral-siren approach

GWs provide an estimate of the luminosity distance d_L through Eq. (1.18), but they do not directly determine the redshift z, because of its degeneracy with detector-frame masses (see Eq. 1.42). This degeneracy can be broken by exploiting additional information, such as the mass function when it contains distinctive and identifiable structures. The ability to identify distinctive features in the mass function, which can act as references to break the mass-redshift degeneracy, is particularly important in the spectral sirens approach. In this framework, in fact, the cosmological information is inferred entirely from the mass distribution and therefore depends on the adopted mass model. For the PLP and PL2P models, the cosmological inference relies almost entirely on the locations of the Gaussian peaks, meaning that only the features associated with parametric components can be reconstructed. As the dataset grows and additional features emerge,

a fully parametric model would require more components to capture them. In contrast, the PLS model allows the mass function to adapt more flexibly to the data: the positions of the spline knots are fixed a priori, while the cosmological information is encoded primarily in the spline coefficients. Even in the PLS it is crucial the position of knots: knots located too far from the true peaks may flatten or smooth them, while simultaneously introducing artificial fluctuations and resulting in incorrect inferences. The importance of the model selection and of how accurately the model must reconstruct the underlying mass distribution to break the degeneracy is illustrated in Figure (3.5), which shows an example taken from Pierra G., et al., 2025.

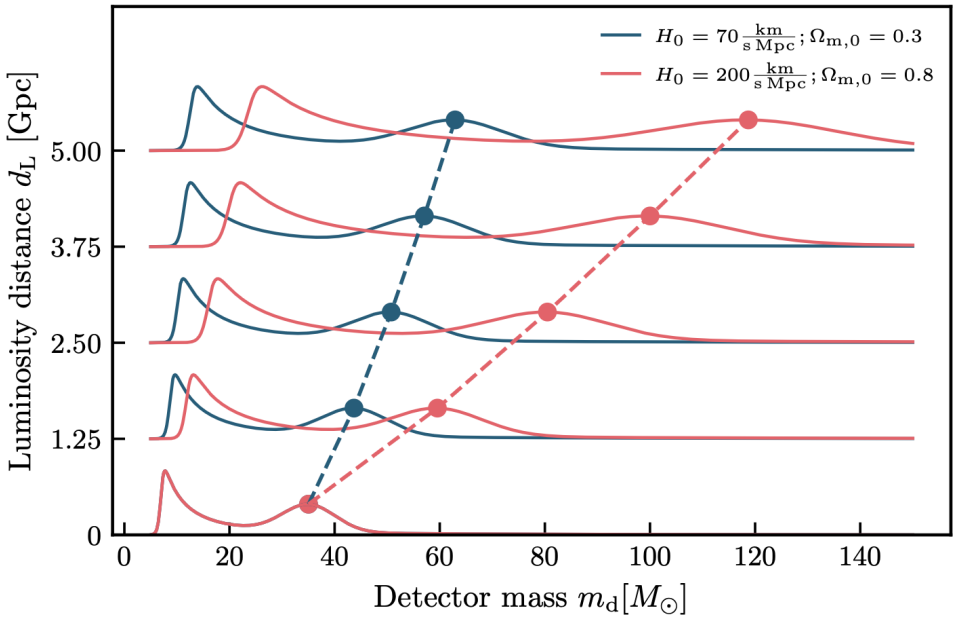

Figure 3.5: Graphical representation of the evolution of the mass function in the detector frame mass-luminosity distance d_L plane. For each value of d_L the detector frame mass spectrum is displayed for two different set of cosmological parameters (H_0, $\Omega_{m,0}$) within the ΛCDM cosmology. The dashed lines track the shifting peaks of the distributions. The illustration is taken from Pierra G., et al. (2025).

In Figure (3.5), the mass function in the detector frame is shown as a function of d_L for two fixed cosmologies. The Figure focuses on the 35 $M_\odot$ peak to illustrate that, at the same luminosity distance d_L, different cosmologies correspond to different observed mass functions. At $d_L = 0$, the two mass functions coincide because $z = 0$ and therefore $m_{det} = m_{src}$. As the distance d_L increases, the curve moves due to redshift effect, since at any given d_L the corresponding redshift depends on the assumed cosmology.
Although the plot displays only a single peak, the increasing number of detected GW events has revealed multiple features in the mass function. In addition to the 35 $M_\odot$ peak, studies on the GWTC-2 dataset identified evidence for a peak at 10 $M_\odot$ (Edelman B., et al., 2022), while the GWTC-4 dataset revealed a third peak at 20 $M_\odot$ (Tagliazucchi M., et al., 2026) and a bump around 60 $M_\odot$ (Abac A. G., et al., 2025). The more structure are observed and detected, the more cosmological information can be extracted frm the mass distribution. This is analogous to spectroscopy: having several spectral lines allows for a much more precise redshift measurement than having only one. Similarly,

additional structure in the mass function enables more accurate extraction of redshift and cosmological information.

Chapter 4

Testing the semi-parametric model on future gravitational wave observations

The main innovation of the semi-parametric `PLS` model lies in its ability to flexibly and accurately capture the structure of the data. To ensure the reliability of this approach, the `PLS` model is first validated on mock data before being applied to real data. In fact, the mock data analysis provides a framework in which the population properties are known by construction.

This Chapter presents the application of the `PLS` model to mock galaxy and GW catalogs, in order to validate the semi-parametric model and to provide forecasts for *O5* detector configuration. In Section (4.1), we present the generation of the mock galaxy and GW catalogs, created and described in Borghi N., et al., 2025. In particular, we start from a parent galaxy catalog (Sec. 4.1.1), which represents the galaxy population in a simulated universe and from which potential BBH host galaxies are identified by applying a linear weighting scheme. Then, a mock GW catalog is generated (Sect. 4.1.2) following defined population models.

The presence of a galaxy catalog enables the implementation of both dark sirens (Sect. 4.2) and spectral sirens (Sect. 4.3) analyses. In both cases, we focus on the study of the mass function, the role of the knots, and their impact on cosmological and population hyperparameters.

Although a mock galaxy catalog is available, a spectral siren analysis is also performed to reflect the conditions of real data, where no galaxy catalog is available.

4.1 Mock catalogs

4.1.1 Parent galaxy catalog and BBH host galaxy catalog

The mock galaxy catalog used in this Thesis is derived from the MICE Grand Challenge light-cone simulation (MICECATv2). This simulation covers approximately one octant of the sky ($\sim 5157\ deg^2$), including galaxies with i-band magnitudes $i < 24$ within a redshift range $z < 1.4$ and assumes a flat ΛCDM cosmology with $H_0 = 70$ km/s/Mpc, $\Omega_{m,0} = 0.25$ and $\Omega_{\Lambda,0} = 0.75$ (Fosalba P., et al., 2015, Crocce M., et al., 2015, Carretero J., et al., 2015, Hoffman K., et al., 2015)).
We use the subsample of MICECATv2 presented in Borghi N., et al., (2025), which is

constructed to ensure that the galaxy stellar mass distribution can be efficiently described by a redshift-evolving double Schechter function. This subsampling reduces the original catalog size while preserving realistic galaxy clustering and improving computational efficiency for the standard siren analysis. The resulting catalog, hereafter referred to as the parent galaxy catalog, contains 67% of the galaxies contained in the MICECATv2 ($\sim$ 335 million galaxies).

Once the parent galaxy catalog has been constructed, a crucial step is to associate the BBH mergers with their host galaxies included in the catalog. The probability that a galaxy hosts a BBH merger can depend on its physical properties, such as the stellar mass M_*, the star formation rate SFR, and the metallicity Z (Santoliquido F., et al., 2022). Previous studies have shown that the stellar mass plays a dominant role: more massive galaxies are more likely to host a BBH merger than low-mass ones. Moreover, this relation is strongly redshift dependent, shifting toward massive host galaxies at lower redshift (Artale M. C., et al., 2020).

Motivated by these results, the probability of hosting a BBH merger in our mock galaxy catalog is assigned proportional to the stellar mass of each galaxy:

$$w \propto M_*. \tag{4.1}$$

Among the three possible weighting schemes presented in Borghi N., et al., (2025), in this Thesis the linear one is adopted, as it provides the most accurate representation of the BBH host galaxy population in the local Universe for the construction of the host galaxy catalog.

4.1.2 GW catalog

The GW catalog is constructed in Borghi N., et al., (2025) by sampling the true event-parameters (RA, Dec, z) from the catalog of potential hosts. To each host galaxy is associated a BBH event with population parameters sampled from known mass and merger rate models, by fixing cosmological parameters λ_c and astrophysical population parameters λ_r and λ_m. This prior knowledge of the underlying models allows us to test and assess whether the proposed `PLS` model is able to reproduce the data accurately.

The overall BBH merger rate follows the Madau-Dickinson model, presented in Eq. (2.12). To construct the catalog, the rising slope is set to $\gamma = 2.7$, consistent with the GWTC-3 results. The redshift peak and the corresponding declining slope are instead chosen to be $z_p = 2$ and $\kappa = 3$, respectively.

The mass component follows a `PLP` model, which is the preferred mass model up to GWTC-3. The primary mass distribution consists of a truncated power law with a superimposed Gaussian peak; instead, the secondary mass distribution is modeled as a power law with index β in the mass range $[m_{low}, m_1]$. The cosmology is assumed to be the same as that used in MICECATv2.
Therefore, the GW population is described by these cosmological and astrophysical hyperparameters:

$$\begin{aligned} \lambda_c &= \{H_0\} \\ \lambda_z &= \{\gamma, \kappa, z_p\} \\ \lambda_m &= \{\alpha, \beta, \delta_m, m_{low}, m_{high}, \mu_g, \sigma_g, \lambda_g\} \end{aligned} \tag{4.2}$$

The true values are summarized in Table (4.1) along with the prior ranges used in the subsequent analysis.

Parameter	Description	True value	Prior
H_0	Hubble constant [km/s/Mpc]	70.0	$\mathcal{U}(10.0, 200.0)$
$\Omega_{m,0}$	Matter energy density	0.25	Fixed
γ	Slope at $z < z_p$	2.7	$\mathcal{U}(0.0, 12.0)$
κ	Slope at $z > z_p$	3	$\mathcal{U}(0.0, 6.0)$
z_p	Redshift Peak	2	$\mathcal{U}(0.0, 4.0)$
α	(Primary) slope of the power law	3.4	$\mathcal{U}(1.5, 12.0)$
β	(Secondary) slope of the power law	1.1	$\mathcal{U}(-4.0, 12.0)$
δ_m	(Primary) smoothing scale [$M_\odot$]	4.8	$\mathcal{U}(0.01, 120.0)$
m_{low}	Lower mass value [$M_\odot$]	5.1	$\mathcal{U}(2.0, 50.0)$
m_{high}	Upper mass value [$M_\odot$]	87.0	$\mathcal{U}(50.0, 200.0)$
μ_g	(Primary) mean of the Gaussian peak [$M_\odot$]	34.0	$\mathcal{U}(2.0, 50.0)$
σ_g	(Primary) width of the Gaussian peak [$M_\odot$]	3.6	$\mathcal{U}(0.4, 10.0)$
λ_g	(Primary) fraction of the Gaussian component	0.039	$\mathcal{U}(0.01, 0.99)$

Table 4.1: Cosmological and astrophysical population hyperparameters used to generate the mock catalogs, along with their true values and associated uniform ($\mathcal{U}$) priors used in the standard siren analysis.

We simulate the detection of our BBH population by the LIGO-Virgo-KAGRA network at O5 sensitivity, considering a O5-like network configuration which includes the 2 LIGO interferometers, Virgo, KAGRA, and a third LIGO detector located in India (Abbott B. P., et al., 2016) with a low frequency cutoff of 10 Hz and 100% duty cycle.

The detection is simulated using *GWFAST*, a Python code designed to forecast the signal-to-noise ratios and parameter estimation of GW detector networks, based on the Fisher Information Matrix method (Iacovelli F., et al., 2022). The simulation, performed assuming quasi-circular orbits and non-precessing BBH, aims to estimate the uncertainties on the intrinsic and extrinsic source parameters θ, already presented and discussed in Section (2.1), Eq. (2.1).

For each source, the GW signal is simulated using the `IMRPhenomHM`[1] waveform approximant (Kalaghatgi C, et al., 2020, London L., et al., 2018), consistently with previous works (e.g., Borghi N., et al., 2024, Borghi N., et al., 2025 and Tagliazucchi M., et al., 2025). The corresponding network signal-to-noise ratio (S/N) is then computed and a threshold of $S/N > 25$ is applied.

The final catalog of GW events considered in this Thesis consists of 1000 events at $S/N > 25$, each event with 5000 posterior samples. The number of events is chosen to ensure high statistics to accurately validate the semi-parametric mass function method in our mock analysis. We note that, for reference only, such a catalog corresponds approximately to ten years of observations under the assumed population model and LVK O5 scenario.

The selection effects, accounting for the observational biases introduced by the detector sensitivity, are evaluated by generating a set of simulated injections with *GWFAST*,

[1]The `IMRPhenomHM` approximant is a frequency domain aligned-spin waveform model

adopting the same signal-to-noise ratio threshold $S/N > 25$ used for the GW catalog. The injections population consists of $2 \cdot 10^7$ simulated sources, of which 10^6 have $S/N > 25$ in the *O5*-like scenario. We ensure that these events span the full range of detectable luminosity distances and binary masses. These injections are used to compute the selection effects term according to Eq. (2.7).

4.2 Forecasts for a LVK O5 analysis: the dark siren scenario

As a first analysis, we study the implementation of the `PLS` mass function in the *dark sirens* configuration. In this case, the inclusion of a galaxy catalog allows the redshift to be inferred from the potential host galaxies contained within the gravitational wave localization volume. Galaxy redshift measurement uncertainties are included in `CHIMERA` and modeled as Gaussian errors with a standard deviation $\sigma_z = 0.001$. In this configuration, the cosmological information is primarily constrained by the galaxy catalog (Borghi N., et al., 2024), allowing us to focus on the impact of the mass distribution. In particular, the simulated catalog is generated assuming a `PLP` model, whereas our analysis adopts a `PLS` model. The key question is, therefore, how well our model can reproduce known features of the distribution and whether this model introduces biases in other cosmological or astrophysical hyperparameters.

Hierarchical Bayesian inference is applied to the mock catalog using the `CHIMERA` pipeline, introduced in Chapter 2. The posterior distribution used in this analysis is sampled with the MCMC sampler `emcee` (Foreman-Mackey D., et al., 2013).
The `PLS` parameters that are shared with the `PLP` model use the same priors as those adopted for the `PLOP` model in the mock catalog construction. For the spline coefficients, we assign a uniform prior in the range $[-4, 4]$, which is sufficiently wide to allow the parameters to flexibly adapt to the data, while remaining sufficiently constrained to prevent excessive dispersion. This interval is chosen to encompass the full range of spline coefficient values obtained when fitting the `PLP` mass function with the `PLS` model.
The priors adopted for the `PLS` analysis on mock data are summarized in Table (4.2).

4.2.1 Mass function analysis

In Figure 4.1 , we show the reconstructed primary mass distributions for the `PLS` model using $N_{knots} = \{5, 10, 15, 20\}$. The knots are logarithmically distributed between $m_{low} = 2.0\ M_\odot$ and $m_{high} = 200.0\ M_\odot$, corresponding to the lower and upper bounds of the prior range for the low-mass and high-mass cutoffs, respectively, as reported in Table (4.1).

All reconstructed primary mass distributions, despite significant differences in the number of knots, are able to identify the peak at 35 $M_\odot$, corresponding to the Gaussian component of the `PLP` model assumed for the mock catalog. It is worth emphasizing that the `PLS` model with 5 knots, although it provides the least accurate reconstruction of the `PLP` distribution, is still able to recover the peak. This occurs despite only three knots lying within the mock data domain, with the remaining two fixed at the lower and upper mass boundaries (2 $M_\odot$ and 200 $M_\odot$).

Parameter	Description	Prior
H_0	Hubble constant [km/s/Mpc]	$\mathcal{U}(10.0, 200.0)$
$\Omega_{m,0}$	Matter energy density	Fixed
γ	Slope at $z < z_p$	$\mathcal{U}(0.0, 12.0)$
κ	Slope at $z > z_p$	$\mathcal{U}(0.0, 6.0)$
z_p	Redshift Peak	$\mathcal{U}(0.0, 4.0)$
α	(Primary) slope of the power law	$\mathcal{U}(1.5, 12.0)$
β	(Secondary) slope of the power law	$\mathcal{U}(-4.0, 12.0)$
δ_m	(Primary) smoothing scale [$M_\odot$]	$\mathcal{U}(0.01, 120.0)$
m_{low}	Lower mass value [$M_\odot$]	$\mathcal{U}(2.0, 50.0)$
m_{high}	Upper mass value [$M_\odot$]	$\mathcal{U}(50.0, 200.0)$
c_i	Spline coefficients at the knots position	$\mathcal{U}(-4.0, 4.0) \cdot N_{knots}$

Table 4.2: Summary of the PLS cosmological and astrophysical population hyperparameters and their associated prior distributions adopted in MCMC analysis.

In contrast, the reconstructed primary mass distributions obtained with 10, 15, and 20 knots are very similar and recover the same structures in the data: the main peak at 35 $M_\odot$ and an additional fluctuation around ~ 60 $M_\odot$. As suggested by the underlying posterior of single events, shown in grey, this secondary overdensity is consistent across individual event posteriors, suggesting it reflects genuine fluctuation in the mock catalog rather than a spline artifact.

The lower panel of Figure 4.1 shows the relative difference of the primary mass distribution with respect to the PLP model, while Table 4.3 reports these differences in percentage form. The PLS model with 10 knots provides the most accurate reconstruction, likely because the models with 15 and 20 knots introduce a higher degree of overfitting, resulting in spurious features and artificial fluctuations, especially at low masses where the knot density is higher.

$N_{\mathbf{knots}}$	Median relative difference
5 log	26.3%
10 log	8.5%
15 log	10.9 %
20 log	14.4%
10 chosen	7.8%

Table 4.3: Median relative percentage difference between the PLS models and the fiducial PLP model. The table compares PLS models with $N_{\text{knots}} = [5, 10, 15, 20]$ logarithmically spaced knots and a PLS model with 10 chosen knots.

The number of knots sets the number of degrees of freedom of the PLS model and determines also the knots location, until they are uniformly distributed in the log-space. The knot positions are crucial because they determine the positions where the spline is anchored to the underlying power law (see Section 3.2). Consequently, the placement of a knot can affect the reconstructed mass function: a knot located where it is expected

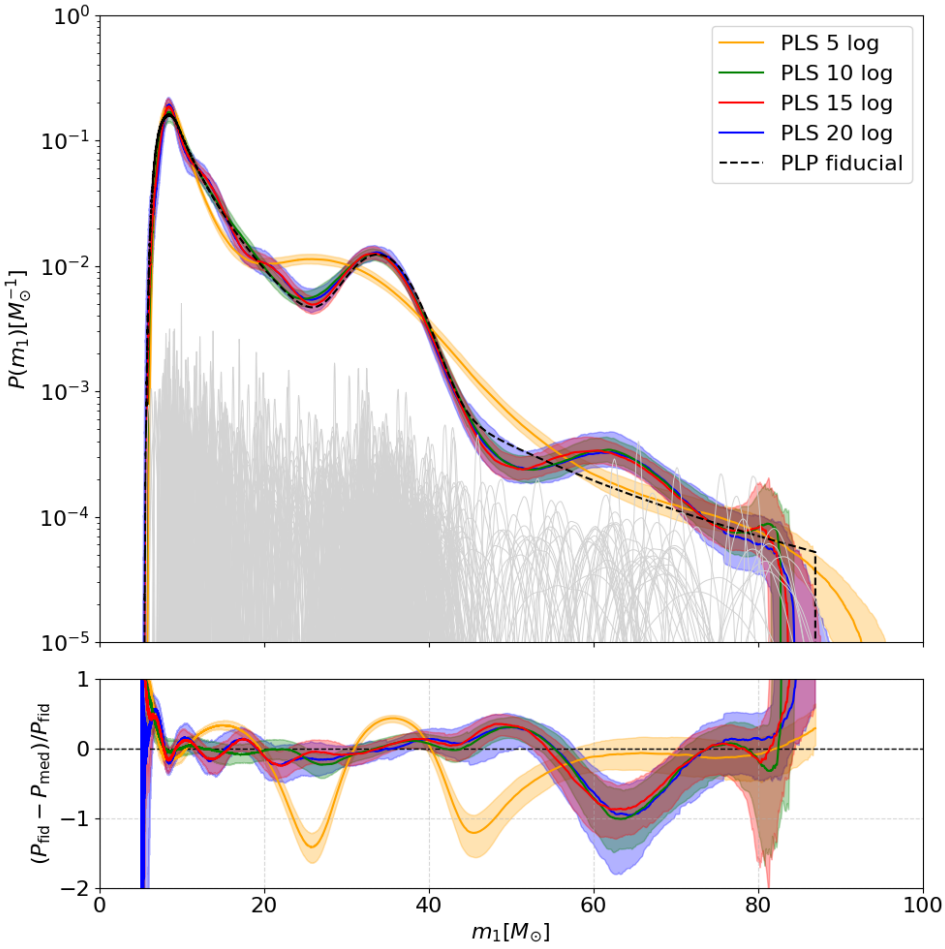

Figure 4.1: Posterior distributions of the Power Law + Spline (PLS) BBH primary mass function for different knot configurations ($N_{\rm knots} = [5, 10, 15, 20]$), uniformly spaced in log-mass. The black dashed line shows the PLP distribution at fiducial hyperparameters. Solid lines indicate median PLS values; shaded regions denote 68% credible intervals. Light-gray curves represent posteriors from 1000 individual GW mock events. The bottom panel shows the relative difference between the PLS and fiducial PLP models.

a peak could smooth or suppress that feature. For this reason, the flexibility of the PLS model depends not only on the number of knots, but also on their positions.

Starting from the PLS model with 10 knots, which provides the best overall representation of the fiducial PLP according to the median relative difference (see Table 4.3), we investigated whether the model accuracy could be further improved by modifying the knot positions. Figure (4.2) shows the comparison between two PLS models with 10 knots: one with logarithmically distributed knots and one with chosen knot positions. The latter configuration is thought to avoid placing knots on mass distribution features, to prevent their smoothness or suppression (see Section 3.4 for the properties of PLS knots). Different from the logarithmically distributed knots, the chosen knots are strategically anchored at the boundaries of the expected peaks.

In Figure (4.2), both the PLS distributions reproduce the peak at 35 $M_\odot$ well, although the PLS distribution with logarithmic knots presents small-scale fluctuations driven by

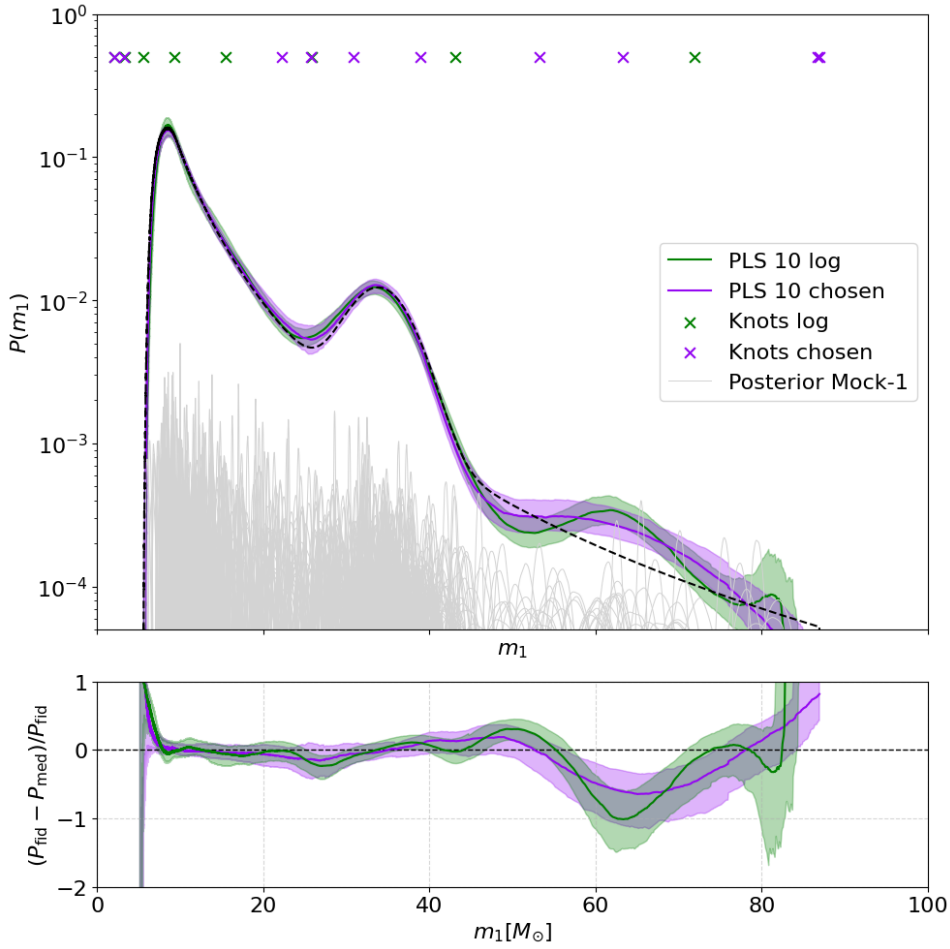

Figure 4.2: Posterior distributions of the Power Law + Spline (PLS) BBH primary mass function for two configurations: 10 logarithmically spaced knots and 10 selected knots. The black dashed line shows the PLP distribution at fiducial hyperparameters. Solid lines indicate median PLS values; shaded regions denote 68% credible intervals. Light-gray curves correspond to posteriors from 1000 individual GW mock events, and the crosses mark the knot positions for each distribution. The bottom panel shows the relative difference between the PLS and fiducial PLP models.

the specific positions of the knots. The differences in the knot positions are highlighted by the crosses shown in the upper panel of the Figure. The most significant variation between the two models is the one around the fluctuation at ~ 60 $M_\odot$, where the local accuracy between the two distributions increases by 30%.

However, from a global point of view, the increment produced by the PLS with chosen knots is lower than one percentage point, as reported in Table (4.3).

As a final step, a more consistent evaluation of the differences between the two distributions is obtained by comparing a PLS model inferred through MCMC sampling with a PLP model inferred in the same way. The comparison is now performed between the median mass function obtained by MCMC, rather than against a fixed fiducial PLP model. Therefore, we also carried out MCMC sampling for the PLP model. Figure (4.3) compares the PLS probability distributions with the PLP probability distribution. The

lower panel shows the median relative difference computed with respect to the median PLP rather than to the fiducial PLP.

At $\sim 60\ M_\odot$, as suggested by the posteriors of individual GW events, the PLP distribution deviates from the fiducial PLP by about 10%, indicating the presence of a fluctuation. This confirms that the fluctuation previously observed in the PLS models is also present in the data, suggesting that it originates from statistical fluctuations in the simulated events. Overall, the PLS model with the 10 chosen knots provides the best representation, as shown in Table 4.4.

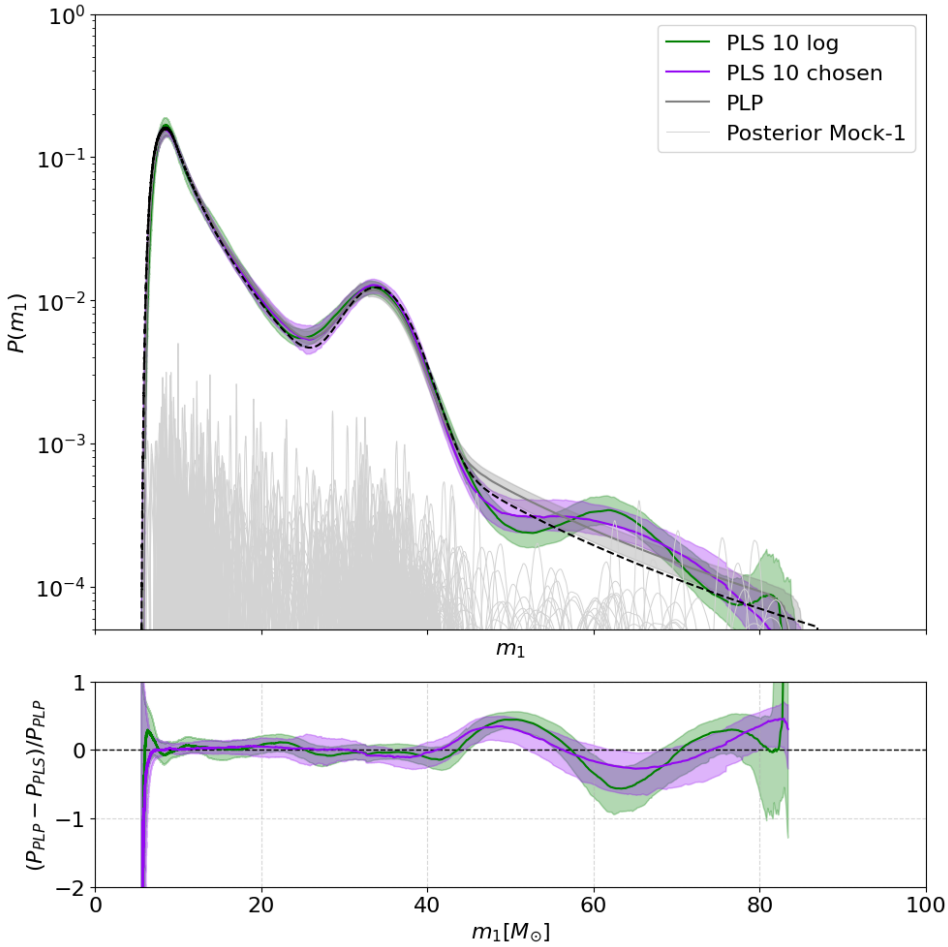

Figure 4.3: The upper panel shows the posteriors of two PLS and one PLP primary mass distributions. The two PLS models are built with 10 logarithmically spaced knots and 10 selected knots. Solid lines indicate the median distributions, and shaded regions represent the 68% credible intervals. The black dashed line shows the PLP distribution at fiducial hyperparameters, while light-gray curves correspond to posteriors from 1000 individual GW mock events. The bottom panel displays the relative difference between the PLS models and the median PLP.

4.2.2 Posterior constraints on hyperparameters

In the previous Section, we investigated the ability of the semi-parametric PLS model to accurately reconstruct mock data, showing its capability in capturing new features.

$N_{\mathbf{knots}}$	**Median relative difference**
10 log	7.7%
10 chosen	3.9%

Table 4.4: Median relative percentage difference between two PLS models — one with 10 logarithmically distributed knots and one with 10 chosen knots — and the median PLP.

In this Section, we present a detailed analysis of the impact of the PLS model on the marginalized posterior distributions of some crucial hyperparameters. This analysis evaluates the posterior distributions of a set of key parameters and examines whether they are consistent with the true values used to construct the mock catalog: a way to test the model's reliability.

The four analyzed parameters are H_0, α, m_{high} and γ. Including cosmological and rate hyperparameters, in addition to those of mass, is crucial to study how well the fiducial models are recovered and potential degeneracies between hyperparameters.

Figure (4.4) shows the constraints on the four hyperparameters, while Table (4.5) reports the median values and the errors associated to 68% credible intervals for each hyperparameter. Each column of Figure (4.4) corresponds to a different parameter, while each row illustrates a specific configuration of the analysis:

- the effect of varying the number of logarithmically spaced knots;
- a comparison between the best configuration with 10 logarithmically spaced knots and a set of 10 user-defined knots;
- a comparison between the best-performing PLS model (with 10 user-defined knots) and the PLP model.

Regarding the cosmological parameter H_0, its median value and associated uncertainties do not depend on a specific choice of knots configuration. In fact, the relative error is $\sigma_{H_0}/H_0 \approx 0.6\%$ in all cases, indicating that in the context of *dark siren* analysis, H_0 is entirely constrained by the spectroscopic redshifts of the galaxy catalog. A more detailed investigation of H_0 requires a *spectral sirens* analysis, which is presented in the following section.

For the mass hyperparameters, we focus on the power law slope α and the high mass cutoff m_{high}.
For α, the most precise result comes from the PLP model, where the median value is consistent with the fiducial value within the uncertainties and the relative error is $\sigma_\alpha/\alpha \approx 2.62\%$. Among the PLS models, the most precise result comes from the 5 knots logarithmically distributed configuration, with a relative error of 13.5%. All the other models have relative errors between 18% and 21%. It is, however, interesting to note that the median value furthest from the fiducial one is obtained from the 10-knot user-defined model. It is therefore crucial to emphasize that the PLS model that is least accurate in reproducing the mass distribution (the configuration with 5 logarithmically distributed knots) yields the most accurate inference of the parameter α, while the PLS model that best reproduces the mass distribution (the 10 knots user defined configuration) provides the least estimation. This occurs because α controls the slope of the primary mass distribution, and it is degenerate with the spline coefficients, especially at low masses.

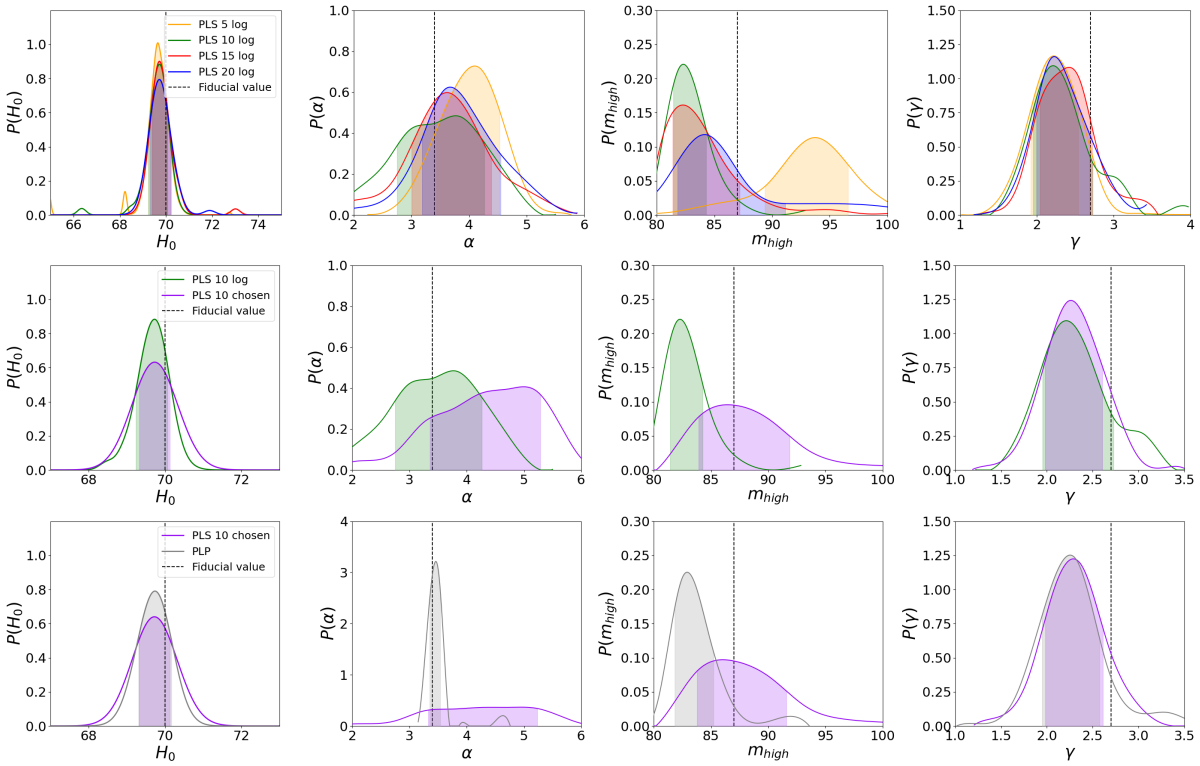

Figure 4.4: Marginalized posterior distributions for the selected hyperparameters in the *dark siren* analysis. The shaded colored regions represent the 68% credible intervals, while the dotted lines mark the fiducial values of the hyperparameters.

Instead, for m_{high}, the model that best reproduces the high mass cutoff is the PLS model with 10 chosen knots. Its median value is consistent with the fiducial one within the uncertainties with a relative precision of $\sigma_{m_{high}}/m_{high} \approx 4.5\%$. The other PLS configurations and the PLP model tend to underestimate m_{high}, with the exception of the PLS 5 knots model, which instead overestimates it. This behavior is due to a degeneracy between m_{high} and the last spline coefficients.

The final parameter analyzed is the rate parameter γ, which characterizes the rising slope. Across all the mass models investigated, γ is consistent with the fiducial value within the uncertainties, with a relative error $\sigma_\gamma/\gamma \approx 14\% - 17\%$, and shows no degeneracy with the other mass hyperparameters.

4.3 Forecasts for a LVK O5 analysis: the spectral siren scenario

As a second analysis, we investigate how and whether the PLS mass function reproduces the data in the spectral sirens configuration. In Section 4.2, we validated the mass function by inferring redshifts from a galaxy catalog, thereby obtaining tight constraints on cosmological parameters. Here, in the spectral sirens mode, we perform a joint inference analysis of astrophysical and cosmological parameters. The spectral siren method

	5 log	**10 log**	**15 log**	**20 log**	**10 user-defined**	**PLP**
$\boldsymbol{H_0}$ [km/s/Mpc]	$69.7^{+0.4}_{-0.4}$	$69.7^{+0.4}_{-0.4}$	$69.8^{+0.5}_{-0.4}$	$69.7^{+0.5}_{-0.4}$	$69.7^{+0.4}_{-0.4}$	$69.7^{+0.4}_{-0.4}$
$\boldsymbol{\alpha}$	$4.0^{+0.5}_{-0.6}$	$3.5^{+0.7}_{-0.8}$	$3.6^{+0.7}_{-0.6}$	$3.8^{+0.7}_{-0.6}$	$4.4^{+0.8}_{-1.1}$	$3.4^{+0.1}_{-0.1}$
$\boldsymbol{m_{high}}$ [$M_\odot$]	$93.4^{+3.2}_{-4.0}$	$82.7^{+1.6}_{-1.3}$	$83.3^{+3.9}_{-1.9}$	$84.5^{+6.7}_{-2.7}$	$87.5^{+4.4}_{-3.6}$	$83.3^{+1.9}_{-1.4}$
$\boldsymbol{\gamma}$	$2.2^{+0.3}_{-0.3}$	$2.3^{+0.4}_{-0.3}$	$2.4^{+0.4}_{-0.3}$	$2.3^{+0.4}_{-0.3}$	$2.3^{+0.3}_{-0.3}$	$2.2^{+0.3}_{-0.3}$

Table 4.5: Median values and 68% CI of the marginalized posterior distributions for the selected hyperparameters. The results correspond to the *dark siren* analysis for all the studied knot configurations.

provides cosmological constraints by exploiting features in the BBH mass distribution, without relying on a galaxy catalog. In this approach, redshift information is extracted from the source frame mass distribution features, which implies that the accuracy of the inferred cosmological parameters also depends on how these mass features are recovered. From this perspective, the aim is to study whether systematic difference occurs between the fiducial `PLP` model and the semi-parametric `PLS` model, focusing in particular on the cosmological information.

The reason for performing a spectral siren analysis is to mimic realistic conditions for real data, where a complete galaxy catalog is not available, and to test the performance of the `PLS` mass model under these conditions.
As in the previous Section, the results presented here are obtained through the Hierarchical Bayesian Inference using the `CHIMERA` pipeline.
The analysis begins with the inference of the mass function distribution to test whether the `PLS` model is able to accurately recover the underlying mass distribution even in the absence of a galaxy catalog.
Subsequently, we present the results of the inference on cosmological and population hyperparameters, without any galaxy catalog information.
The posterior distribution is sampled using the MCMC sampler `emcee`, and the priors adopted for the `PLS` parameters are the same as those reported in Table (4.2).

4.3.1 Mass function analysis

The mass function analysis in the spectral siren configuration is conducted using a single `PLS` model, which is compared to a `PLP` model. Specifically, we employ a `PLS` model with 10 logarithmically distributed knots. This choice is motivated by the fact that, among the `PLS` configurations tested in dark sirens with logarithmically distributed knots, this `PLS` model is the most accurate. Although the `PLS` model with 10 chosen knots provides the best fit to the mock data in the *dark siren* approach, it remains a data-driven method and is therefore less informative when preparing for the analysis of the GWTC-4 data. In this mock data study, the intent is to verify that there are unbiased measurements between the semi-parametric and parametric models. In the analysis of real GWTC-4 data, multiple logarithmic and data-driven models will then be evaluated (see Section 5.3.1), beyond the logarithmic 10 knots `PLS` model.

Figure (4.5) shows the comparison between the `PLS` and the `PLP` mass probability

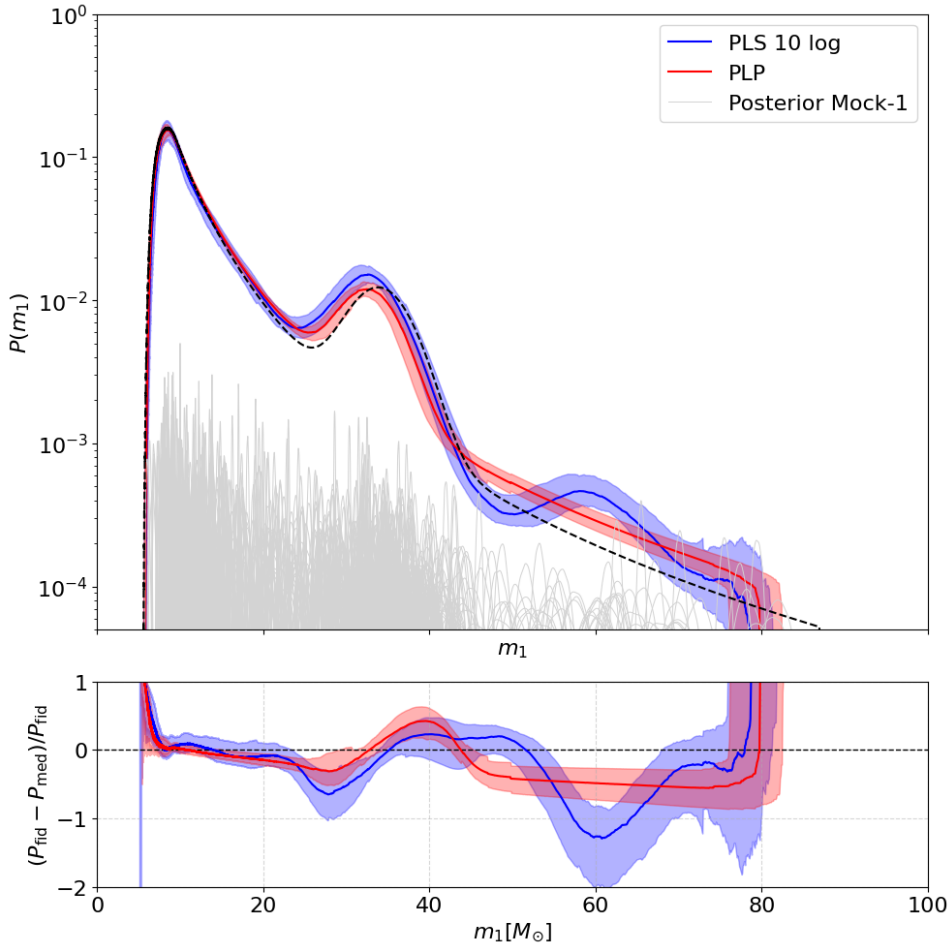

Figure 4.5: The upper panel shows the posteriors of the PLS and PLP primary mass distributions. The PLS model is constructed with 10 logarithmically spaced knots. Solid colored lines indicate the median distributions, while shaded regions denote the 68 % credible intervals. The black dashed line shows the PLP distribution at fiducial hyperparameters, and the light-gray curves represent posteriors from 1000 individual GW mock events.
The bottom panel displays the relative difference, $(p_{\rm fid} - p_{\rm PLS})/p_{\rm fid}$, between the PLS model and the fiducial PLP model.

distributions and the fiducial PLP distribution. Compared to the *dark siren* analysis, the features of the mass distribution are less accurately reconstructed in the *spectral sirens* configuration. In particular, the peak at 35 $M_\odot$ is not precisely recovered by the PLP model, and it is overestimated by the PLS model, while both the distributions show an over-density at ~ 60 $M_\odot$ consistent with the behavior found in the *dark siren* analysis. The reduced accuracy in the reconstruction of the mass distribution arises from the lack of redshift information, which introduces a degeneracy between redshift and the source frame mass distribution.

The values reported in Table (4.6) show that the PLS model provides the most accurate reconstruction of the fiducial PLP distribution compared to the PLP model. In fact, the flexibility of the semi-parametric PLS model allows it to adapt to the data and capture features in the mass distribution, leading to a more accurate inference of astrophysical and cosmological parameters (Farah A. M., et al., 2024).

$N_{\mathbf{knots}}$	**Median relative difference**
PLP	19.8%
PLS 10 log	16.3%

Table 4.6: Median relative percentage difference between the PLS and the PLP models with respect to the and the fiducial PLP in *spectral siren* configuration.

However, as discussed in the previous section, the fiducial PLP does not reproduce the over-density at ~ 60 $M_\odot$. For this reason, in order to quantify this feature, it is more appropriate to examine the relative difference between the median PLS and PLP mass distributions, which amounts to $\approx 10\%$.

4.3.2 Posterior constraints on hyperparameters

As shown in Figure (4.5), the reconstruction of the primary mass function is less accurate than in the *dark siren* case, due to the degeneracy between source frame mass and redshift. Consistently, the hyperparameters inferred in this Section are also less precisely constrained. Figure (4.6) shows the marginalized 1D and 2D posterior distributions, while Table (4.7) reports the median values and the corresponding 68% credible intervals for H_0, α, m_{high} and γ.

The parameter most affected by the absence of a redshift constraint from a galaxy catalog is H_0, which now exhibits a relative uncertainty of $\sigma_{H_0}/H_0 \approx 10\%$. While no degeneracies were observed in the *dark siren* analysis for H_0, in the *spectral siren*, the PLP model shows a clear degeneracy between H_0 and m_{high} and H_0 and μ_g. In fact, m_{high} and μ_g correspond to the most prominent mass features and are crucial to break the mass redshift degeneracy, enabling cosmological inference. Instead, for the PLS model, a significant degeneracy is observed only between H_0 and m_{high}.
The lack of degeneracy between H_0 and the spline coefficients arises because the information about the peak is not encoded in a single spline coefficient, but is instead distributed and averaged across multiple coefficients. Moreover, μ_g and the spline coefficients describe different physical quantities: μ_g parametrizes the location of the peak, whereas the spline coefficients control the amplitude of the distribution at each knot.

The mass hyperparameter α is inferred by the PLP model with a relative precision of $\sigma_\alpha/\alpha \approx 3\%$. In contrast, for the PLS model, the relative precision drops to approximately 21% and α exhibits clear degeneracies with the spline coefficients at low masses.
Instead, the mass hyperparameter m_{high} is inferred with a relative precision of $\sigma_{m_{high}}/m_{high}$ $\approx 4\%$, and this discrepancy is likely due to the degeneracies with cosmological parameters.

From the perspective of the spectral siren analysis, the crucial result is that no systematic difference is observed in the inferred parameters when comparing the semi-parametric PLS model to the fiducial model.
Finally, the absence of a galaxy catalog leads to weaker constraints, even on parameters not directly connected to the redshift.

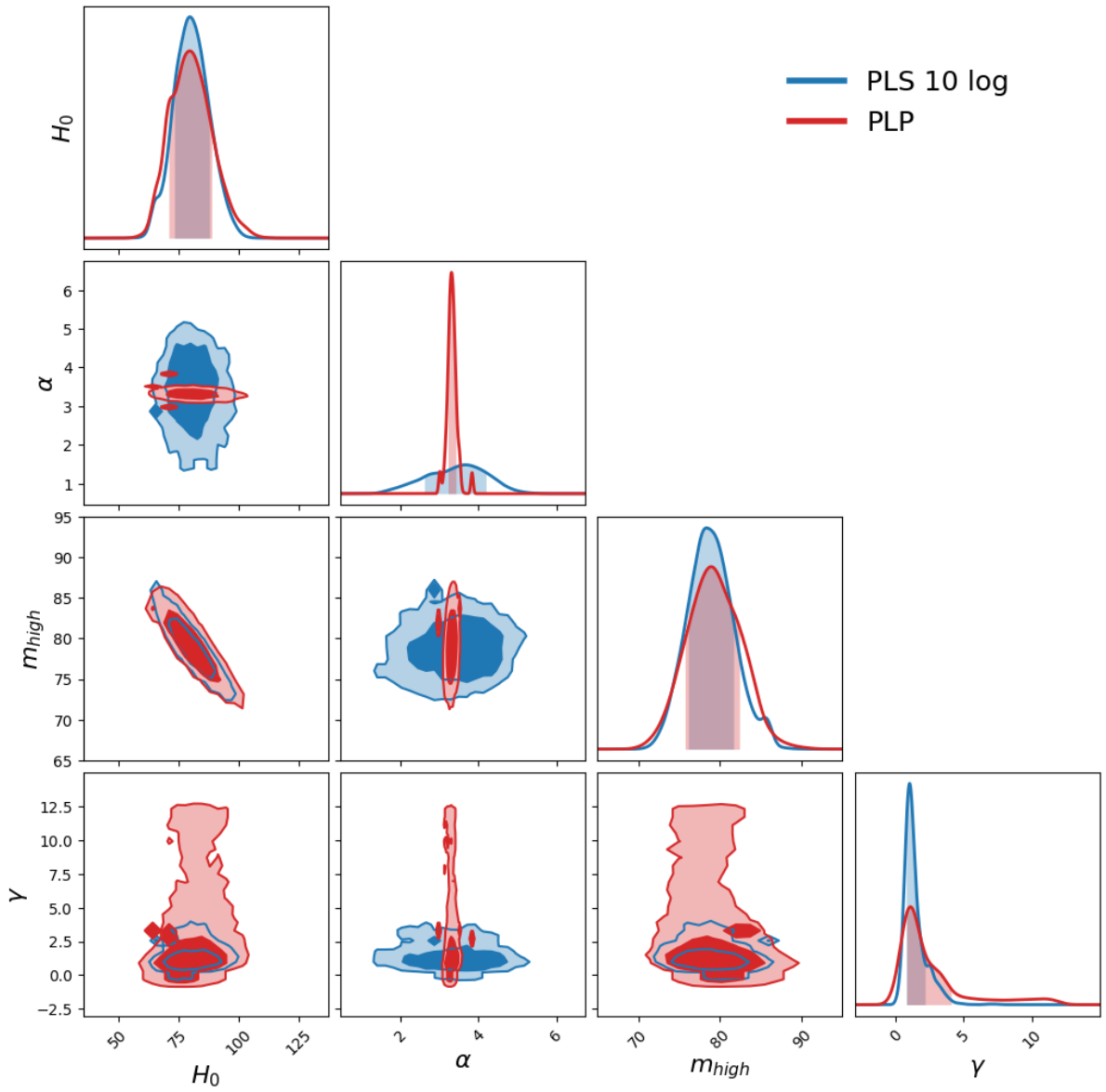

Figure 4.6: Marginalized 1D and 2D posterior distributions for the selected hyperparameters in the *spectral siren* analysis. The shaded colored regions represent the 68% credible intervals, while the dotted lines mark the fiducial values of the hyperparameters.

The rate hyperparameter γ is the worst constrained in the spectral siren analysis with relative error of $\sigma_\gamma/\gamma \approx 69\%$ and 110% for `PLS` and `PLP` respectively. Despite that, the distribution for the tested and the fiducial models are still comparable.

4.4 Summary

In this Chapter, we presented the first application of the `PLS` model to mock data, using the galaxy and GW catalogs constructed in Borghi N., et al., (2025). The importance of testing the `PLS` model on simulated data lies in the fact that the true values used to generate the data are known, providing a controlled environment to assess how the

	10 log	**PLP**
$\boldsymbol{H_0}$ [km/s/Mpc]	$79.2^{+8.0}_{-7.9}$	$80.1^{+8.8}_{-9.5}$
$\boldsymbol{\alpha}$	$3.4^{+0.7}_{-0.8}$	$3.3^{+0.1}_{-0.1}$
$\boldsymbol{m_{high}}$ [$M_\odot$]	$78.8^{+3.1}_{-3.0}$	$79.1^{+3.3}_{-3.4}$
$\boldsymbol{\gamma}$	$1.3^{+1.3}_{-0.5}$	$1.5^{+2.6}_{-0.6}$

Table 4.7: Median values and 68% CI of the marginalized posterior distributions for the selected hyperparameters. The results correspond to the *spectral siren* analysis for 10 knots PLS and PLP models.

model fits the data.

The main goal of this analysis on mock data is, therefore, to assess how well our model can reproduce known features of the distribution and whether this model introduces biases in cosmological or astrophysical hyperparameters.

The analysis of mock data is performed both in dark sirens and spectral sirens configurations. In the dark sirens case, the presence of a galaxy catalog provides tight constraints on the cosmological parameters, allowing us to focus primarily on the reconstruction of the mass function. We considered five different semi-parametric models: four characterized by logarithmically spaced knots and one with knots chosen to follow the posterior distribution of the GW events. The results show that the PLS models successfully recover the main features of the distribution used to generate the data and even reproduces the fluctuation at 60 $M_\odot$ also obtained with the fiducial model. Among the tested configurations, the model with 10 chosen knots provides the best reconstruction, while among the logarithmic configurations, the 10-log model performs best. We find that increasing the number of logarithmically spaced knots tends to increase the model complexity and introduce spurious oscillations, reducing the accuracy of the inferred parameters.

Finally, a spectral sirens analysis is performed on the same dataset, adopting the model that provided the best performance in the dark sirens analysis among the logarithmic models ($N_{knots} = 10$) for the reconstruction of the mass distribution. The main goal of the spectral sirens analysis is to assess whether the reconstruction of the mass function can introduce biases in the inferred cosmological and astrophysical parameters. By comparing the posterior distributions obtained with the PLS and PLP models, we do not find evidence of systematic differences in the inferred parameters, opening the opportunity to test the semi-parametric model on real data.

Chapter 5

Application to real data: improved constraints from the latest gravitational wave observations

The release of a new gravitational wave transient catalog (the fourth version, GWTC-4, publicly released in November 2025), significantly increased the number of detected GW events. In this Chapter, we apply the previously validated `PLS` model to real GW data to assess its performance in reconstructing the mass function and inferring hyperparameters.
First, the primary mass function is reconstructed using the `PLS` model under three different configurations and compared to the fiducial parametric model `PL2P`.
The first configuration adopts a data-driven approach, in which the spline knots are placed following the data distribution (Section 5.2). We then explore configurations with different Gaussian priors on the spline coefficients and, finally, a setup with logarithmically spaced knots. For each mass function configuration, we investigate the impact on parameter inference, considering both population and cosmological parameters. The analysis is performed using spectral sirens only, without assuming any galaxy catalog, so that the constraint on the Hubble constant H_0 relies entirely on mass features and structures in the reconstructed mass distribution.

5.1 The Gravitational-Wave Transient Catalog 4

The dataset used in this analysis is GWTC-4, the fourth gravitational wave transient catalog, fully released in November 2025. It contains GW events detected by the LIGO-Virgo-Kagra (LVK) Collaboration from the first observing run (O1) up to the first part of the fourth observing run (O4a). The dataset analyzed in this Thesis is identical to that adopted for the cosmological analyses performed by the LVK group Abac A. G., et al. (2025), then also used in Tagliazucchi M., et al. (2026). It consists of 137 BBH mergers, selected by imposing a false alarm rate (FAR) threshold of $FAR < 0.25\ \mathrm{yr}^{-1}$ in order to reduce the noise contamination. In fact, the FAR quantifies the statistical significance of GW candidates and corresponds to the expected rate of noise origin events. In this Thesis, only the FAR threshold is applied for the event selection, while the signal-to-noise (SNR) threshold is not imposed.

The greatest contribution of BBHs events comes from the O4a observing run, which includes 75 new BBH detections. The remaining events consist of 3 from O1, 7 from O2, and 52 from O3. The increasing number of detected BBH mergers has allowed the identification of more and more features in the BBH mass distribution that were not predicted by earlier parametric models.
From this dataset, the only excluded event is GW231123_135430, a GW signal originating from a BBH system with component masses of 137^{+23}_{-18} $M_\odot$ and 101^{+22}_{-50} $M_\odot$ at a redshift of $0.40^{+0.27}_{-0.25}$. Although this event could represent a key case for investigating BH formation channels beyond standard stellar collapse, it is not included in this analysis because its extreme properties lead to increased uncertainties in parameter estimation when using current waveform models (Abac A. G, et al., 2025).
Different waveform models have been employed for the parameters estimation across the observing runs, depending on the event's properties. In this analysis, for events detected during O1 - O3, the `IMRPhenomXPHM` waveform model is used, while events from O4a are analyzed with the updated `IMRPhenomXPHM_SpinTaylor`[1] model, which provides a more accurate treatment of spin precession.
The LVK pipeline samples the posterior probability distribution of the source parameters for each GW event, producing a variable number of posterior samples. These samples provide constraints on the event-level parameters that enter Eq. (2.8). For each GW event in this analysis, 5000 posterior samples are randomly selected from the release LVK samples, ensuring computational efficiency and statistical validity, while providing the same number of samples across all events. Only the parameters d_L, m_{1det}, and m_{2det} are retained for this analysis because they are the only ones that enter the population models. The distributions of luminosity distance, redshift, and primary source-frame mass for the GWTC-4 events are shown in Figure 5.1.

An important characteristic to take into account of the O4a dataset is that the sky localization of detected events is broader than expected. This is due to the fact that the Virgo detector was not operating during O4a, so the localization is based on only two detectors.

Finally, the injections used to compute the selection effects in Eq. (2.7) are taken from the public O3 - O4a set presented in (Abac A. G., et al. (2025)). In contrast, no real injections are used for O1 and O2, but the detector sensitivity is inferred from the semi-analytical model proposed by (Essick R. (2023)). Under the assumptions of Gaussian and stationary detector noise, Essick R. (2023) derives the distribution of the matched filter response and calibrates the model using the O3 data.
The event injected selection is performed by applying cuts both on the false alarm rate, requiring FAR < 0.25, and on the signal-to-noise ratio, requiring SNR > 10. In addition, spin parameters are not included in this analysis to simplify the model and to focus on the reconstruction of the mass function, therefore we have marginalized over them.

[1]The `IMRPhenomXPHM` and `IMRPhenomXPHM_SpinTaylor` approximants are theoretical waveform models describing the GW signal emitted by BBH mergers and are used to infer the physical properties of the source

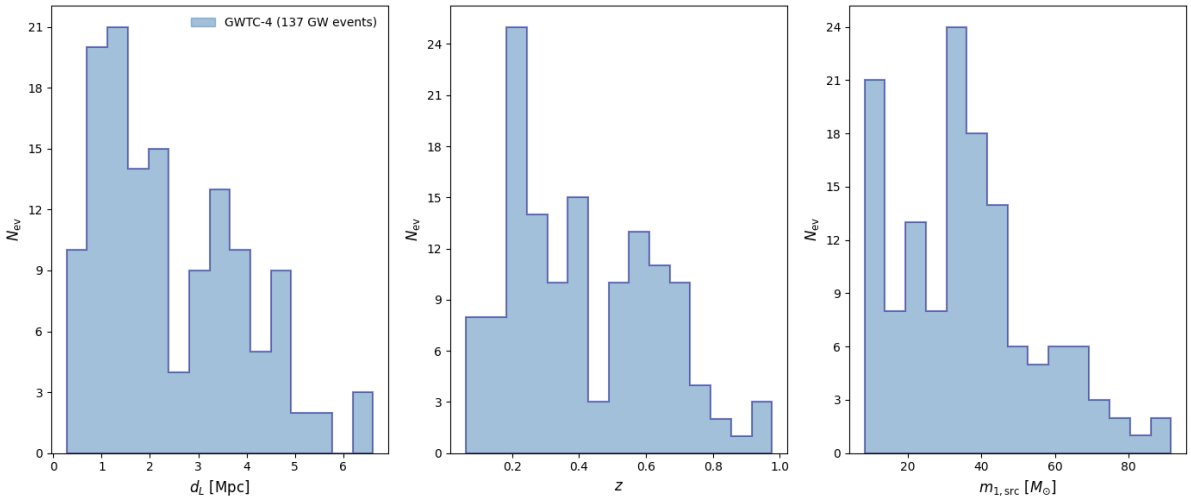

Figure 5.1: Histograms of GWTC-4 events for (from left to right) luminosity distance d_L, redshift z, and primary source-frame mass $m_{1,src}$. Source-frame masses and redshifts are calculated assuming, for illustrative purposes, $H_0 = 70$ km/s/Mpc and $\Omega_m = 0.3$.

5.2 A new data-driven knot assignment approach

Our goal is to fit the new GW catalog using the mass function model described in Section (3.2) and, following the approach of Section (4.2), to test whether the Bayesian inference on the population parameters can be improved with a more realistic modeling of the mass function. To achieve this, we both test a fit with a `PLS` model with logarithmically spaced knots, but we also introduce a new way to place the knots informed by the data distribution. In this Section, we present a data-driven approach that allows the model to better capture the features of the mass distribution (see Tagliazucchi M., et al. 2026). The data-driven method for the spline knots is designed to adapt the mass function reconstruction to the structure of the observed data. In this method, the knot positions are directly inferred by the posterior mass distribution of the detected GW events (marginalizing over cosmological parameters) rather than being fixed a priori.

To apply the approach, we start from the posterior samples of the detector-frame masses for each event. For a given value of H_0, these posterior samples are converted to the source frame through Eq. (1.42), inferring the redshift from the knowledge of the luminosity distance d_L. A Gaussian KDE (kernel density estimate) is then applied to the source-frame mass samples of each event to reconstruct a continuous functional distribution of $p(m_1)$. These event-level KDEs are subsequently summed and renormalized to produce the mean source-frame mass distribution, as shown in Figure (5.2).

This mean source-frame mass distribution is estimated for 100 different values of H_0 spanning the range [10, 200] km/s/Mpc, as shown in Figure (5.2), not to assume an explicit cosmological model, but rather to marginalize over it.

The knot positions are therefore identified in the source-frame mass domain for each value of H_0. To identify distinctive peaks in the mass distribution and construct a `PLS` model capable of representing them, knots are placed where the distribution shows the largest local variations, in correspondence to the peaks of the logarithmic derivative (Figure 5.3).

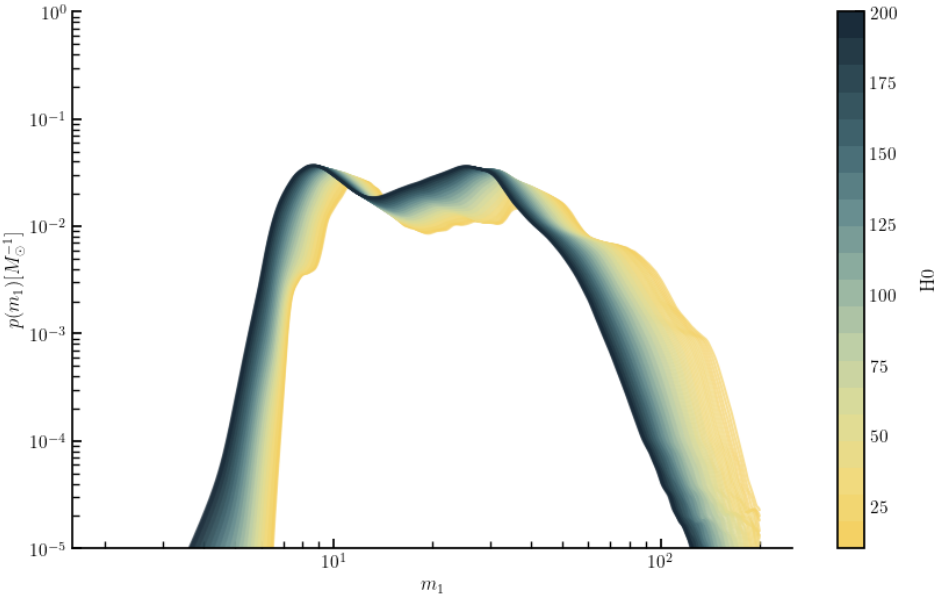

Figure 5.2: Observed source frame primary mass distribution for different values of H0, constructed by averaging over gaussian KDE.

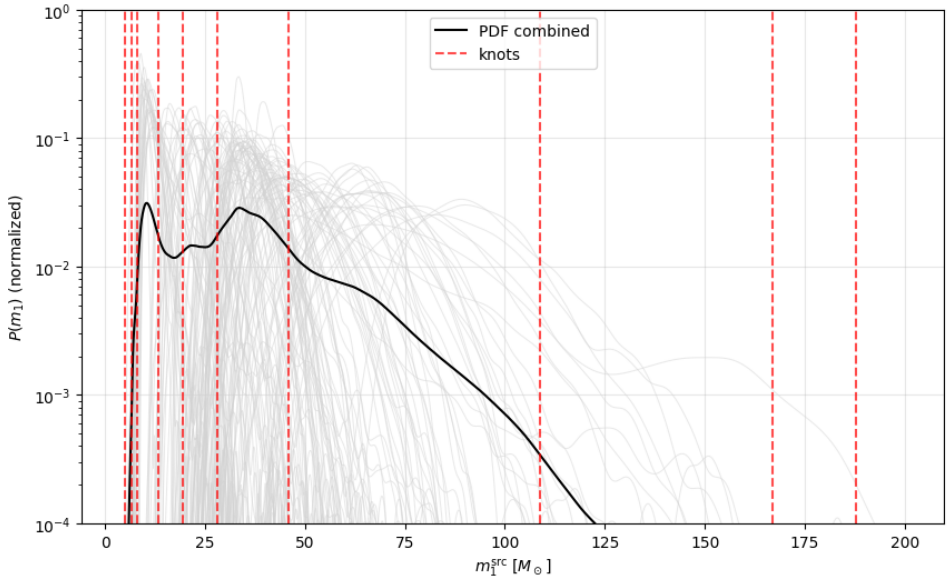

Figure 5.3: The plot shows the combined primary mass distribution (black solid line) computed as the average of the individual event KDEs of the source-frame primary mass. The knot positions are highlighted in red, while the source-frame mass posteriors of individual GW events are shown in light grey. For this illustrative example, the mass distribution is estimated assuming $H_0 = 67.9$ km/s/Mpc.

This procedure is performed for each H_0 value, and the knots identified for each mass distribution are then concatenated into a single set. Once the knots list is obtained, a clustering algorithm is applied to determine the final knot positions. The knots are grouped using Gaussian Mixture Models, and the optimal number of clusters is chosen by minimizing the Bayesian Information Criterion (BIC). The BIC is a model selection met-

ric that reduces model complexity and avoids overfitting. Models with more parameters are penalized, ensuring that the selected number of clusters captures the main features of the distribution. The final knot positions are given by the centers of the resulting clusters.
This method can be tuned through a user-defined parameter that we call *prominence*; this value corresponds to the relevance threshold for identifying peaks in the metric, therefore, by changing it, is possible to obtain a different number of possible selected knots. However, it is not possible to have an arbitrarily large number of knots a priori, as the clustering algorithm will tend to merge those placed too close together; this also allows us to avoid over-structuring the model.
In the analysis presented in Tagliazucchi M., et al., (2026), four values of the *prominence* parameter were considered: $prominence = [0.00106, 0.008, 0.004, 0.016]$, spanning a range of one order of magnitude. For each value, the variation of the knot positions was studied, and these choices resulted in an initial set of $[1914, 826, 1158, 599]$ total knots, respectively. The subsequent analysis using Gaussian Mixture Models and BIC minimization yielded the optimal number of knots: $[10, 14, 16, 12]$. These knots define the data-driven knot positions used to construct the spline basis in the data-driven configuration.

To construct the `PLS` models from the four knot configurations, it is necessary to specify also the priors for the spline coefficients. Together with different numbers of knots, we also explored the impact of different priors on the spline coefficients of the inferred mass distribution. The various considered setups are described in Section (5.3.1). In this case, unlike with the analysis performed on mock data, the posterior distribution is sampled using `pocoMC` (Karamanis M., et al., 2022). The reason behind this choice is that the new sampler offered an improved computational performance, enabling, at the same time, the estimation of the model evidence and performing a model comparison. We verified that, in terms of constraints, the different samplers were providing results in perfect agreement.

5.3 Analysis setup

In this analysis, we wanted to assess whether a more flexible and accurate reconstruction of the BBH mass distribution provides a better description of the GWTC-4 data compared to the commonly used parametric models. The goal is to understand whether, in this way, we can extract from the data more information, therefore improving the constraints on cosmological parameters, and quantifying its impact on the inference.

For this reason, we decided to explore two different functional forms of mass function models: the parametric `PL2P` model and the newly introduced semi-parametric `PLS` model. The parametric model `PL2P` has been chosen, among all the other models, because it is the reference model used in standard cosmological LVK analyses, and in particular in Abac A. G., et al. (2025) for the GWTC-4 dataset.

Below, we discuss the criteria used to build the various `PLS` models and how it is possible to identify the model that best reproduces the data.

5.3.1 Model configurations

We perform our analysis considering different semi-parametric models, depending on the number and position of the knots. In particular, we explore two different choices of knot placements: the data-driven and the logarithmically-spaced ones. In this way, we can estimate the impact of the new method on the results compared with the standard approach adopted so far in the literature (Abac A. G., et al., 2025).

All the models are constructed by placing the knots between 2 $M_\odot$ and 200 $M_\odot$, which correspond to the lower and upper bounds of the mass priors m_{low} and m_{high}, respectively (Table 5.1).
The total number of knots ranges between 10 and 20. For the data-driven case, the number of knots $N_{knots} = [10, 12, 14, 16]$ is determined through the minimization of the BIC. For the logarithmic case, instead, the number $N_{knots} = [10, 14, 15, 20]$ was chosen to explore a broader range, also taking into account the results obtained in Chapter 4 on mock data. The choice of logarithmically spacing the knots is motivated by the need to better resolve the structure of the mass spectrum at low masses compared to high masses. The model with 14 logarithmically spaced knots was included specifically to allow a direct comparison with the 14-knot data-driven configuration.

In addition to varying the number and placement of the knots, we also explore the impact of different priors on the spline coefficients. The priors are Gaussian with mean $\mu = 0$ and standard deviation σ, whose value varies depending on the considered analysis setup.

The configuration setup of our analysis can be divided into three main groups, which differ in the choice of the spline knot positions and in the prior assumptions of the spline coefficients:

- **Data-driven, σ configuration**: the number of knots is fixed $N_{knots} = 14$ and the standard deviation of the Gaussian prior on the spline coefficients varies $\sigma = [0.5, 1, 2, 3, 5]$; this analysis will allow us to choose the best prior according to the data.

- **Data-driven, knot configuration:** the number of knots varies between this range $N_{knots} = [10, 12, 14, 16]$, while keeping the Gaussian prior on the spline coefficients with fixed standard deviation $\sigma = 2$.

- **Logarithmically spaced knots:** the knots are logarithmically distributed with $N_{knots} = [10, 15, 20]$ and a fixed Gaussian prior on spline coefficients with $\sigma = 2$.

The choice of fixing $\sigma = 2$ and $N_{knots} = 14$ while varying, respectively, the number of knots and the widths of the priors, is motivated by the fact that these fixed values are preferred according to the Bayes Factor and ΔDIC (see Section 5.3.3).

For all models in these three configurations, a quantitative comparison with respect to the reference model `PL2P` is performed. The aim is to identify which configuration provides the best reconstruction of the mass function and to evaluate how and whether an improved reconstruction affects the constraints on cosmological and population parameters. The population and cosmological parameters used in this analysis and included in the models are listed in Table (5.1).

Parameter	Description	Model	Prior
H_0	Hubble constant [km/s/Mpc]	All	$\mathcal{U}(10.0, 200.0)$
$\Omega_{m,0}$	Matter energy density	All	Fixed
γ	Slope at $z < z_p$	All	$\mathcal{U}(0.0, 12.0)$
κ	Slope at $z > z_p$	All	$\mathcal{U}(0.0, 6.0)$
z_p	Redshift Peak	All	$\mathcal{U}(0.0, 4.0)$
α	Primary slope of the power law	All	$\mathcal{U}(1.5, 12.0)$
β	Secondary slope of the power law	All	$\mathcal{U}(-4.0, 12.0)$
δ_m	Primary smoothing scale [$M_\odot$]	All	$\mathcal{U}(0.01, 120.0)$
m_{low}	Lower mass value [$M_\odot$]	All	$\mathcal{U}(2.0, 50.0)$
m_{high}	Upper mass value [$M_\odot$]	All	$\mathcal{U}(50.0, 200.0)$
μ_{g1}	Position of the first Gaussian peak [$M_\odot$]	PL2P	$\mathcal{U}(5.0, 100.0)$
σ_{g1}	Width of the first Gaussian peak [$M_\odot$]	PL2P	$\mathcal{U}(0.4, 5.0)$
μ_{g2}	Position of the second Gaussian peak [$M_\odot$]	PL2P	$\mathcal{U}(5.0, 100.0)$
σ_{g2}	Width of the second Gaussian peak [$M_\odot$]	PL2P	$\mathcal{U}(0.4, 10.0)$
λ_g	Fraction of BBH in the first Gaussian peak	PL2P	$\mathcal{U}(0.0, 1.0)$
λ_{g1}	Fraction of BBH in the second Gaussian peak	PL2P	$\mathcal{U}(0.0, 1.0)$
c_i	Spline coefficients at the knots position	PLS	$\mathcal{G}(\mu = 0, \sigma = \hat{\sigma}) \cdot N_{\text{knots}}$ where $\hat{\sigma} = [0.5, 1, 2, 3, 5]$

Table 5.1: Summary of the cosmological and population parameters that enter in the analysis and in the models PLS and PL2P with their associated priors. $\mathcal{U}$ indicates a uniform prior, $\mathcal{G}$ a gaussian prior with mean μ and standard deviation σ.

5.3.2 Model selection

A common challenge in astrophysical data analysis is to identify, among many models with different numbers of parameters or different assumptions and constraints, the one that best describes the data. Since multiple models may reproduce the data distributions with comparable accuracy, it is necessary to adopt quantitative criteria that allow for assessing and comparing the quality of their fit.

Two statistical methods are used to evaluate and quantify the quality of the models: the Bayes Factor (BF) and the Deviance Information Criterion (DIC).

The Bayes Factor (BF) is a statistical indicator that quantifies how much the data favor one model with respect to another. It is defined as the ratio of the Bayesian evidence of the two models considered, in this case between a PLS model and the PL2P, and it is really sensitive to the prior amplitude. A value of BF greater than 1 (or less than 1) indicates an increase (or decrease) in the support for the first model relative to the second, given the observed data (Kass R. E., Raftery A. E., 1995; Rezaei M. and Malekjani M., 2021). Values of $BF > 20$ or $BF > 150$ indicate, respectively, strong and very strong evidence in favor of a model over the PL2P. The Deviance Information Criterion (DIC), instead, is a model selection criterion that balances the quality of the fit and the complexity of the model, penalizing the number of free parameters. Models are compared against the one achieving the minimum DIC, the preferred model, via the quantity ΔDIC: values below 6 indicate inconclusive evidence of a difference, while values above 6 indicate strong evidence that the model is worse. The list of all models used in this analysis with their BF and DIC values is reported in Table (5.2).

Another method used to evaluate models qualitatively in this Thesis is the Posterior

Predictive Check, a visual diagnostic for comparing models, already used in Edelman B., et al. (2022). This method is based on comparing the observed data with the data predicted by the considered model. Specifically, the predictive data are obtained from injections, already used to evaluate the selection effects. First, each injection is reweighted by the ratio of the model-predicted probability to the probability under which the event was originally drawn:

$$w = \frac{P_{new}}{P_{draw}} \tag{5.1}$$

After, a number of predicted events equal to the number of observed events N_{events} is drawn from the weighted injections. This weighted resampling is repeated N_{draw} times, producing N_{draw} predicted catalogs, each containing N_{events} events.
Then, the observed events are also re-weighted: in particular, the posterior samples for each event are re-weighted according to Eq. (5.1), where, instead of P_{draw}, we now use $\pi(\theta)$, the priors on redshift and masses. This re-weighting is performed N_{draw} times. In this way, both the observed and the predicted events are re-weighted according to the inferred population model. Finally, to compare the two sets of data, we compute the cumulative distribution functions (CDF) for each of the N_{draw} catalogs.

5.3.3 Model comparison

We performed the spectral siren analysis on the GWTC-4 data for all the configurations described in Sect. 5.3.1.

According to the BF criterion, all the `PLS` models have a Bayes Factor greater than one when compared to the PL2P, indicating that they all provide a better description of the data than the reference model. However, the models with the narrowest standard deviation for the Gaussian prior $\sigma = 0.5$ and with the smallest number of knots $N_{knots} = 10$ have $BF < 6$. This suggests that they are almost comparable in terms of fit to the `PL2P` model, indicating that a model that reduces or suppresses the presence of structures in the mass distribution is penalized

In contrast, stronger evidence in favor of the `PLS` model emerges for $\sigma \geq 1$ and for $N_{knots} > 10$. The best model, according to the Bayes Factor, is the `PLS-DD-14-`$\mathcal{G}_2$ with $BF = 226$. It is worth emphasizing that increasing the number of knots beyond 14 in the data-driven case does not lead to further improvements and that all the models with logarithmically spaced knots perform worse than their data-driven counterparts.

According to the ΔDIC criterion, the model that best balances the goodness of the fit and the complexity of the model is `PLS-DD-14-G5`, which allows for a more complex behaviour in the mass function with only 14 knots and a broad prior on spline coefficients ($\sigma = 5$). Among the data-driven models, all configurations with $\sigma \geq 2$ and $N_{knots} > 12$ yield the ΔDIC value below the significant threshold, showing a comparable level of evidence with the best model. In contrast, all models with logarithmically spaced knots and the `PL2P` itself exhibit ΔDIC values above the threshold and are therefore significantly disfavored. Nevertheless, among the logarithmic knot models, the configuration that performs better when combining the BF analysis with the ΔDIC evaluation is `PLS-LOG-20-`$\mathcal{G}_2$.

This analysis identifies the prior and the number of knots that best fit the data, with $\sigma = 2$ and $N_{knots} = 14$ being the optimal choices. Most importantly, this analysis

Model	Bayes factor (BF)	ΔDIC
PL2P	1.00	20.11
PLS-DD-14-$\mathcal{G}_{0.5}$	2.59	18.20
PLS-DD-14-$\mathcal{G}_1$	81.84	9.42
PLS-DD-14-$\mathcal{G}_2$	**226.14**	**2.65**
PLS-DD-14-$\mathcal{G}_3$	134.43	0.26
PLS-DD-14-$\mathcal{G}_5$	**24.07**	**0.00**
PLS-DD-10-$\mathcal{G}_2$	5.27	13.43
PLS-DD-12-$\mathcal{G}_2$	60.81	7.93
PLS-DD-16-$\mathcal{G}_2$	114.83	5.41
PLS-LOG-10-$\mathcal{G}_2$	6.02	12.79
PLS-LOG-14-$\mathcal{G}_2$	7.28	13.25
PLS-LOG-15-$\mathcal{G}_2$	8.18	13.39
PLS-LOG-20-$\mathcal{G}_2$	38.74	8.47

Table 5.2: Models with their respective BF and ΔDIC values. The best-performing models according to BF and DIC are highlighted in bold.

also shows strong evidence that the data prefer a mass function model that allows more structures and more freedom in the peaks. The preferred models perform much better than the standard parametric one, and the data-driven approach gives better results quantitatively in model comparison.

5.4 Results from the data-driven approach

The model comparison analysis showed that the data-driven configurations are the most effective in capturing the structure of the primary mass distribution, according to two quantitative criteria. For this reason, this Section focuses on the results obtained with the data-driven models, examining how they reconstruct the mass function and how the inferred features impact both population and cosmological parameters.

In Appendix (A.2), we show the full marginalized distribution for all the parameters fitted for one illustrative case. In the following section, we describe in detail the main results obtained.

5.4.1 Reconstructing the mass distribution

In Figure (5.4), we show the best-fit primary mass posterior distribution for the data-driven models as a function of the prior width σ (*left panel*) and of the number of knots N_{knots} (*right panel*). As a comparison, we also report in the plot the `PL2P` model. Here, `PL2P` is used as a reference to discuss how much additional structure it is possible to recover through the spline models.

For the configuration with varying σ, the width of the Gaussian prior controls how freely the spline coefficients can oscillate. For the narrower priors, $\sigma = [0.5, 1]$, the model is bound, by construction, to have a smooth and slowly varying behaviour, where the

peaks are strongly suppressed. Compared to the PL2P model, where the amplitudes of the peaks are free parameters with uniform priors, the semi-parametric models result in a reduced reconstruction of the two main overdensities at approximately 10 $M_\odot$ and 31 $M_\odot$ in the mass function.

Instead, with less restrictive priors, $\sigma \geq 2$, structures start to emerge in the mass distribution, including not only the two peaks previously identified by the PL2P model, but also a third additional feature at intermediate masses. In particular, the two peaks become more pronounced, the second peak slightly shifts to 33 $M_\odot$, and a third intermediate feature emerges around 18 $M_\odot$. Increasing σ also enhances the amplitude of the peaks, as expected, especially for $\sigma = 5$, where we observe the first evidence of a gap at low masses for $m < 10$ $M_\odot$.

It is worth noting that in the literature it is common to impose strong priors on the amplitudes of the peaks (i.e. Edelman B., et al., 2022 who test $\sigma = [1, 2]$). However, this produces an evident risk of suppressing part of the information that can be extracted from the mass function. As discussed in Section (5.3), allowing greater freedom in the peak amplitudes is in fact preferred by the data.

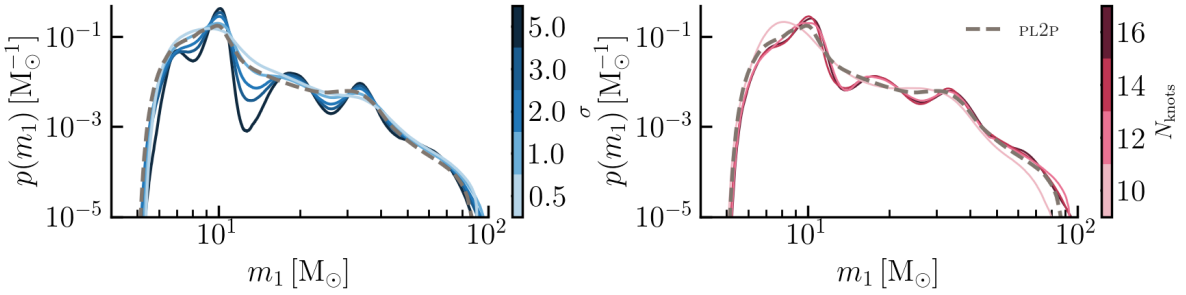

Figure 5.4: Best-fit primary mass posterior distributions for the data-driven models obtained from the GWTC-4 dataset. *Left:* data-driven configurations with fixed $N_{knots} = 14$ and varying prior width σ. *Right:* data-driven configurations with fixed $\sigma = 2$ and varying the number of knots N_{knots}. In both panels, the reference PL2P model is shown for comparison.

In contrast, varying N_{knots} while keeping $\sigma = 2$ fixed, allows us to isolate the impact of the number of knots on the recovered mass function. It is therefore interesting to highlight how the number of knots affects the reconstructed structures. When too few knots are used, the model lacks flexibility, and part of the structure in the mass function is lost, leaving only the most prominent peaks. An example is the data-driven configuration with $N_{knots} = 10$, where even the position of the first peak around 11 $M_\odot$ is not correctly captured. Conversely, the models with $N_{knots} = [14, 16]$ produce very similar results, reconstructing the three peaks suggested by the data at approximately 11 $M_\odot$, 18 $M_\odot$, and 33 $M_\odot$. An additional feature that becomes visible with a larger number of knots is the bump at ~ 60 $M_\odot$.

Overall, Figure (5.4) highlights that compared to PL2P, the PLS models are capable of extracting richer structures in the mass function, especially when the prior width or the number of knots allows sufficient flexibility. This confirms that the PLS approach can reveal features that remain hidden under more rigid and parametric assumptions, such as that of the PL2P model.

5.4.2 Comparison with the literature

To validate our results, we first compare them with those found in the literature, followed by a visual inspection through the Posterior Predictive Check (PPC).

A direct comparison with independent non-parametric approaches in the literature highlights whether and how the `PLS` model implemented in `CHIMERA` is capable of recovering the same or additional finer structure in the mass distribution. In particular, we compare our results with those obtained by Abac A. G., et al. (2025).
In Figure (5.5), we present the constraints on the primary mass functions for all data-driven models considered in this Thesis, showing the best-fit and the 68% credible interval of $p(m_1)$ for each reconstruction. These results are compared with the reconstruction obtained by Abac A. G., et al. (2025) with the `B-Spline` model.
Both approaches rely on the spline bases (B-Spline; see Section 3.1) to reconstruct the primary mass distribution, but they differ in flexibility and in the role of parameters. The `B-Spline` model used in Abac A. G., et al., 2025 is a fully data-driven approach, where the mass function is expressed as a linear combination of B-Spline bases; instead, the `PLS` model is a semi-parametric approach, where the cubic spline at the exponent perturbs and modifies an underlying parametric power-law (Eq. 3.5).
Before comparing the models, it is important to note that the datasets used for the two analyses differ slightly. While both works are based on GWTC-4, in this Thesis we select and analyze only 137 BBH events (Section 5.1), whereas Abac A. G., et al. (2025) considered 153 BBH events. This difference is due to the threshold on the false alarm rate (FAR) applied in each analysis: in this Thesis we adopt a FAR < 0.25 yr^{-1}, instead Abac A. G., et al., 2025 used FAR < 1 yr^{-1},

Moreover, the `B-Spline` model is designed to represent the full population of CBCs, and is flexible enough to include mergers of BNS and NS-BH systems as well. In contrast, `CHIMERA` currently models only the mass function of the BBHs.

Another important difference is that the analysis performed by LVK is a pure population study, in which cosmological parameters are fixed with $H_0 = 67.7$ km/s/Mpc. Instead, our work represents the first study in the literature where `PLS` data-driven models are considered to constrain jointly cosmological and astrophysical parameters, hence without fixing the cosmology. In addition, the spline knots were also selected by varying the H_0 parameter (Section 5.2).

Despite these differences, the comparison remains crucial for understanding whether the primary mass structure is recovered similarly. The `B-Spline` model resolves multiple over- and under-densities in the BH mass distribution, reproducing the peaks at 10 $M_\odot$ and 35 $M_\odot$, and providing evidence for a third feature around 20 $M_\odot$. Compared to the reference `PL2P` model, the `B-Spline` reveals additional substructure as the third peak at 20 $M_\odot$, and it identifies a bump at 60 $M_\odot$. These features are also recovered by the `PLS` reconstruction. Figure (5.5) clearly shows that increasing the number of knots improves the agreement with the `B-Spline` results, while increasing σ tends to make the peaks and gaps appear more spurious, such that they do not reflect the data distribution and deviate from the `B-Spline` reconstruction. Conversely, using a very low σ suppresses information contained in the mass distribution, reducing and smoothing features.

Finally, to visually assess how well each model reproduces the observed mass distribution, we perform the Posterior Predictive Checks (PPC). For this analysis, the

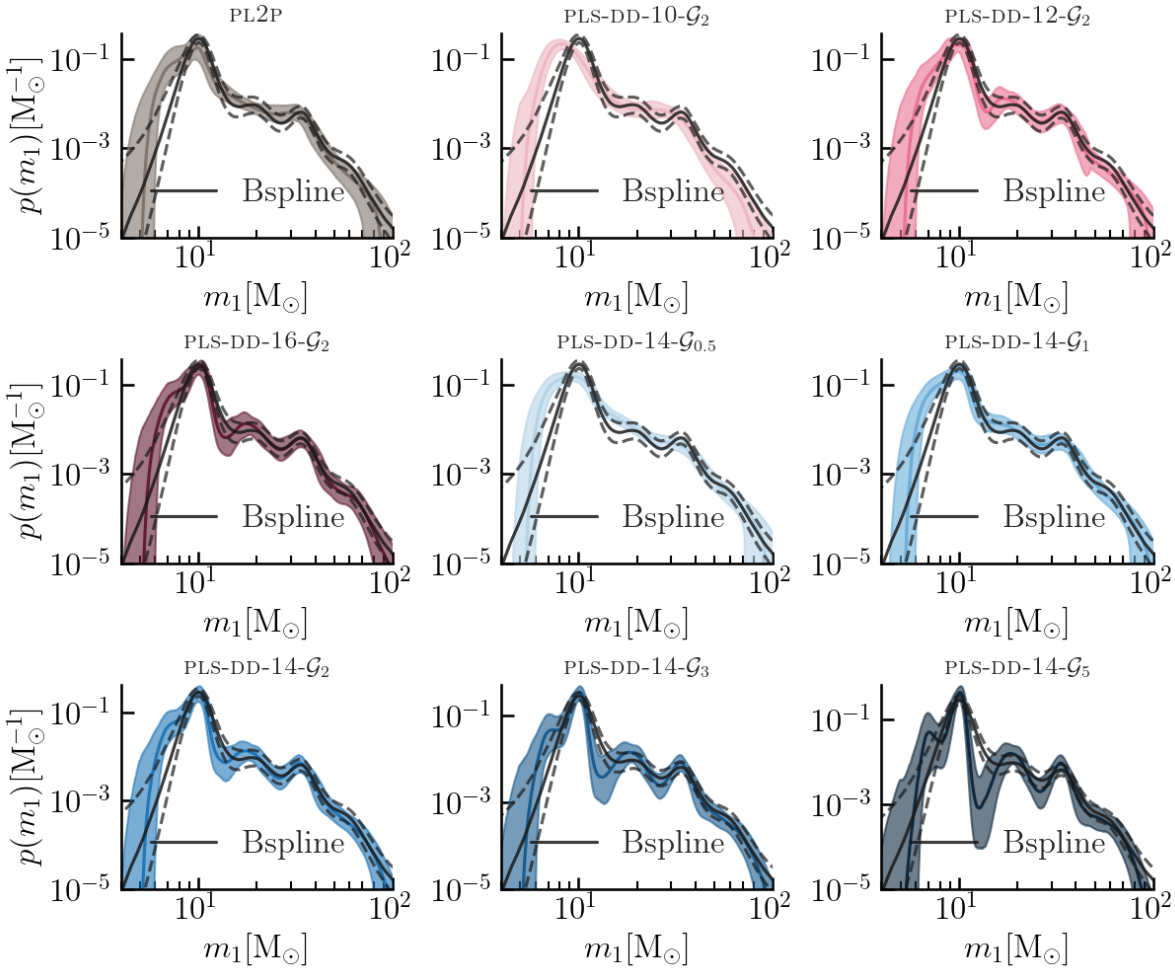

Figure 5.5: Best-fit and 68% confidence interval of the primary mass posterior distributions for all the data-driven models, compared with the results obtained by Abac A. G., et al. (2025), which use another independent weakly-parametric approach (dashed lines).

number of predicted events drawn from the weighted injections is set equal to the number of observed events $N_{events} = 137$. Both the predicted and observed events are drawn $N_{draw} = 200$ times, producing 200 catalogs for each, to ensure statistical stability. Figure (5.6) shows the cumulative distribution function (CDF) of the predicted events, compared to the observed events, for each model analyzed.
In a PPC, a model becomes inconsistent with the data when the observed band extends outside the predictive band. Overall, in this analysis, all models yield an observed distribution that is compatible with the predicted one, meaning that none of them are rejected based solely on PPC. The only model where the observed curve slightly falls outside the predicted band is the `PLS-DD-14-`$\mathcal{G}_{0.5}$. In fact, it is not a coincidence that the worst data-driven model according to BF and DIC is also the worst model according to the PPC, as this is due to the excessively narrow width of the spline coefficients prior. Where the qualitative PPC diagnostic is no longer informative, the assessment of the model fit must rely on quantitative criteria such as DIC and BF, as reported in Table (5.2).

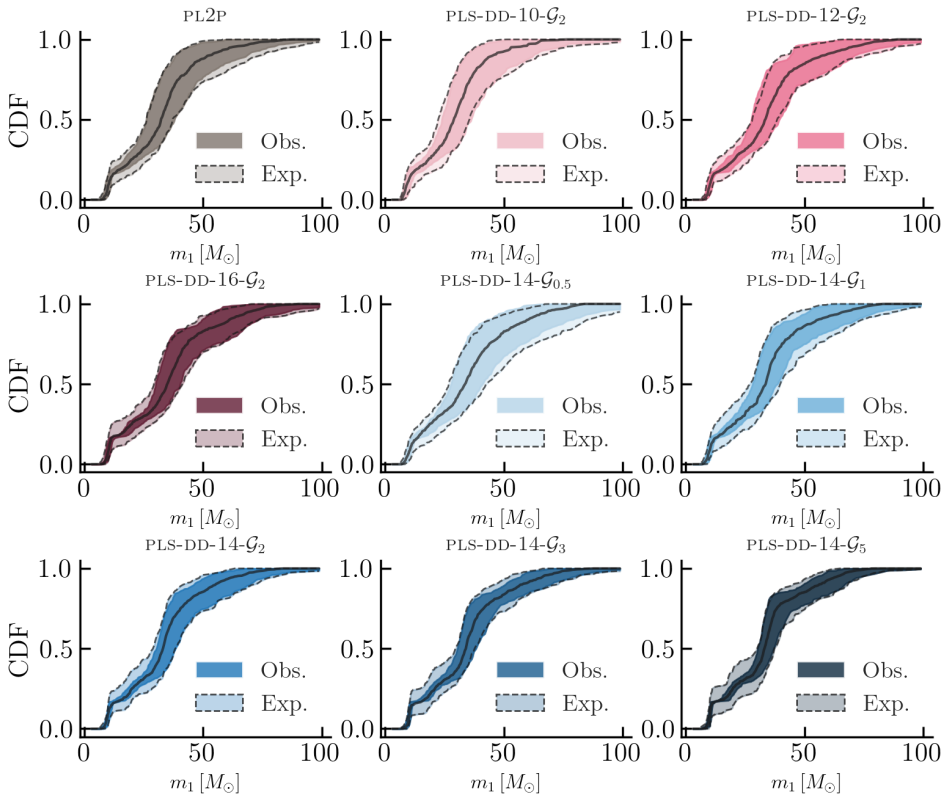

Figure 5.6: Posterior predictive checks for all data-driven models and for the logarithmic model with $N_{knots} = 14$, evaluated on GWTC-4 dataset. The darker region shows the 90% credible interval of the observed cumulative distribution of the BBH population, while the lighter region corresponds to the 90% credible interval of the predicted distribution. The solid line indicates the median of the predictive cumulative distribution.

5.4.3 Constraints on cosmological and astrophysical parameters

In Tagliazucchi M., et al. (2026), the analysis about constraints on parameters posterior focuses exclusively on the Hubble constant H_0, the cosmological parameter most sensitive to the reconstruction of the mass function. In this Section, in addition to revisiting the results already discussed for H_0, we expand those by discussing the other population parameters relevant for the analysis, such as m_{low}, m_{high}, δ_m, and α. For all these parameters, we study whether and how much increasing the information extracted from the mass function improves the constraining power.

One of the major results obtained, as shown in Figure (5.7) and summarized in Table (5.3), is that models capturing more structures in the mass distribution $p(m_1)$ lead to tighter constraints on H_0. In particular, the best model according to the Bayes Factor provides:

$$H_0 = 57.8^{+21.9}_{-20.6}\ \mathrm{km/s/Mpc} \tag{5.2}$$

which corresponds to a 12% improvement with respect to the reference PL2P model. Conversely, the model with the lowest DIC value shows an even larger improvement of

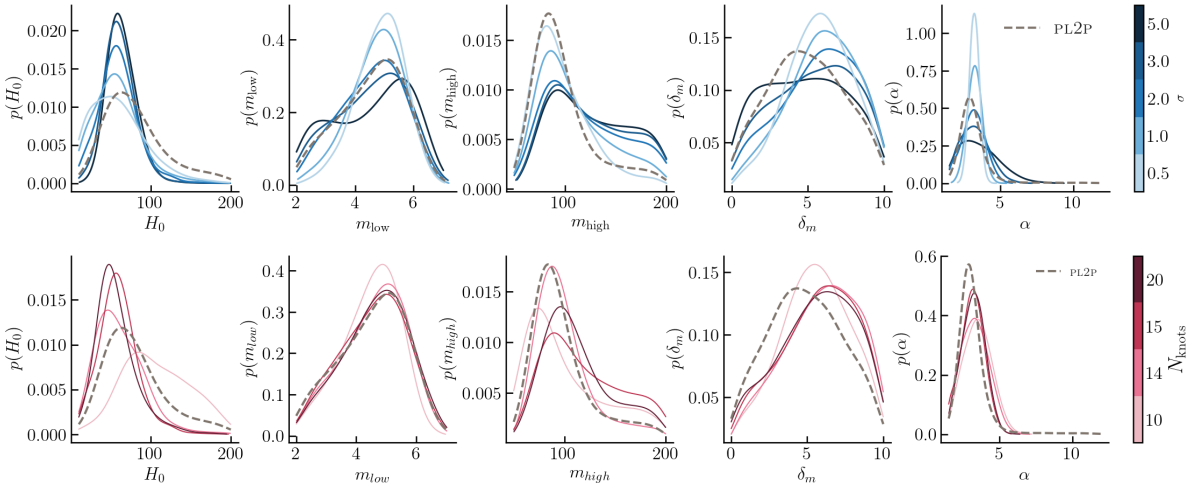

Figure 5.7: Posterior distributions of parameters $H_0, m_{\rm low}, m_{\rm high}, \delta_m, \alpha$ as a function of σ (*upper panel*) and $N_{\rm knots}$ (*bottom panel*).

21% yielding:

$$H_0 = 61.7^{+19.3}_{-14.9}\ \text{km/s/Mpc}. \tag{5.3}$$

A notable trend is that the accuracy on the H_0 parameter improves as σ increases: the relative uncertainty decreases from $\sigma_{H_0}/H_0 \approx 61.5\%$ for $\sigma = 0.5$ down to 27.7% for $\sigma = 5$. This behavior can be explained by the fact that, as discussed in Section (3.5) and shown in Fig. (3.5), the structure in the BBH mass function provide an indirect information on the redshift, and hence has a direct impact on cosmological constraints. Here, we demonstrate that each structure in the mass distribution carries relevant cosmological information, and hence models with greater freedom in the spline coefficients (i.e., a larger σ) increase the information extracted to infer H_0.
A different number of knots N_{knots} in the `PLS` model also affects the inferred value of H_0. Specifically, models with $N_{knots} \leq 12$ produce very weak constraints on the parameter, being also, as discussed, associated with worse BF and DIC values.

Regarding the population parameters, we observe opposite trends to those seen for H_0. In particular, for all four population parameters analyzed (m_{low}, m_{high}, δ_m, α), the tightest constraints come from the `PLS-DD-14-`$\mathcal{G}_{0.5}$ model, which corresponds to the model with the smallest prior width $\sigma = 0.5$.
As σ increases, the relative uncertainty grows for all population parameters. This behaviour can be understood by recalling the construction of the `PLS` model (see Section 3.5), where the primary mass distribution is modeled as an exponential perturbation of an underlying truncated power-law, where the spline components can capture deviations. As a consequence, the parameters describing the power law and the spline coefficients are interdependent: increasing the prior width on the spline coefficients allows the spline to absorb part of the information that would otherwise be constrained by the parametric component. This interplay leads to larger uncertainties on the parametric parameters as σ increases. The population parameters, in contrast to the cosmological ones, are also less sensitive to the details of the mass function reconstruction. They do not exhibit

clear degeneracies with the spline coefficients, although part of the information absorbed by the spline coefficients can reduce the information carried by the parametric power-law parameters. Nevertheless, the median values obtained for each parameter across the different models are all consistent within the uncertainties.

Here below we provide a more in-depth discussion of the main astrophysical parameters.

The lower mass value $\mathbf{m_{low}}$. For the parameter m_{low}, all models yield consistent median values within their uncertainties, indicating that the lower mass cutoff is well constrained by the data independently of the considered model. All the data-driven `PLS` and `PL2P` models return values in the range $4.6 - 4.9$ $M_\odot$ (see Table 5.3). However, what clearly changes is the precision: the relative uncertainty increases from $\sigma_{m_{low}}/m_{low} \approx 16.3\%$ for the `PLS-DD14-`$\mathcal{G}_{0.5}$ model up to $\sigma_{m_{low}}/m_{low} \approx 32.2\%$ for the most flexible model `PLS-DD14-`$\mathcal{G}_5$.
The model preferred by the Bayes Factor is the one that infers m_{low} the most similar to the reference `PL2P` model, giving:

$$m_{low} = 4.7^{+1.0}_{-1.3}\ M_\odot, \tag{5.4}$$

whereas the model with the lowest DIC value exhibits the largest uncertainty. In contrast, varying the number of knots N_{knots} at fixed $\sigma = 2$ produces almost constant median values and uncertainties, yielding an improvement of $1 - 5\%$ compared to the `PL2P` model. However, these differences are well below the inferred uncertainties on m_{low} and are therefore negligible in finding the best model.

The upper mass value $\mathbf{m_{high}}$. The upper mass cutoff m_{high} shows a stronger dependence on σ. Although the median value remains statistically consistent within the uncertainties across all models, it increases from ≈ 90 $M_\odot$ for `PL2P` and for `PLS` models with $\sigma = 0.5$ and $N_{knots} = [10, 12]$, to about 120 $M_\odot$ for the largest values of prior width $\sigma = [3, 5]$. Its uncertainties also roughly double across the models.
m_{high} is the only parameter that does not follow the general trend of increasing uncertainty with σ. The model that provides the broadest uncertainties is `PLS-DD14-`$\mathcal{G}_3$ with a relative error $\sigma_{m_{high}}/m_{high} \approx 37.5\%$, whereas the tightest constraints are obtained with `PLS-DD12-`$\mathcal{G}_2$. This behavior indicates a degeneracy between m_{high} and H_0 that we explore in Figure (5.8). For high values of the prior width ($\sigma = 3 - 5$), the degeneracy between m_{high} and H_0 becomes weaker and smoother. Instead, for lower values ($\sigma = 0.5; 1; 2$), m_{high} exhibits a clear negative correlation with H_0. The behaviour originates from the degeneracy between the detector-frame masses and the redshift: smaller values of H_0 correspond to smaller inferred redshift z at fixed luminosity distance d_L, which leads to larger source-frame masses. In order to maintain these masses compatible with the population model, the m_{high} must increase, producing the observed anti-correlation.

Because of this degeneracy, it is not possible to separate the effects of the upper mass scale and the expansion rate of the Universe using only these data and these models. In this regime, the model compensates for a lower value of the Hubble constant H_0 with a higher value of the upper mass cutoff in order to remain consistent with the observed population.

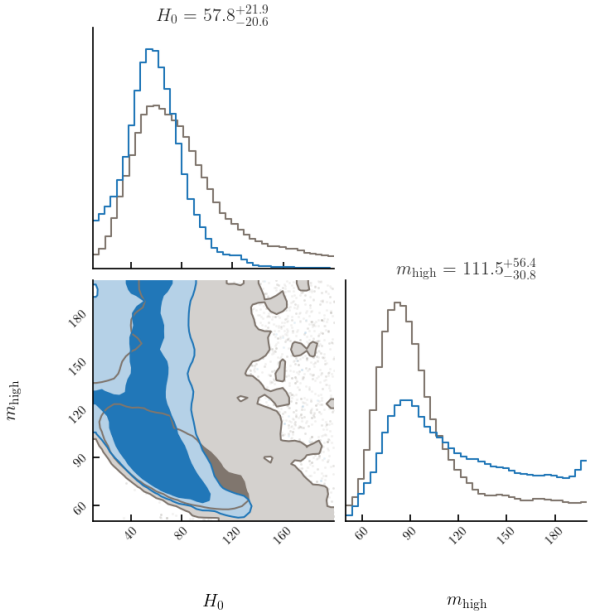

Figure 5.8: Degeneracy between H_0 and m_{high}. Here, we report the PL2P model (in gray) and the best model according to the BF (PLS-DD14-$\mathcal{G}_2$, in blue), showing both the 1D and 2D marginalized distributions.

The primary smoothing scale $\delta_{\mathbf{m}}$. The parameter δ_m follows the expected trend with σ: its uncertainty increases as σ grows. The tightest constraint (37.9%) comes from the $\sigma = 0.5$ model, while the largest uncertainty (65.6%) is found for $\sigma = 5$. An interesting feature is the strong negative prior-dependent degeneracy between δ_m and m_{low} (see Appendix A.2 and Appendix A.3): as one increases, the other decreases. This degeneracy weakens as σ increases and as the spline captures more structure in the mass distribution. The same degeneracy remains with the variation of N_{knots}. However, the variation of N_{knots} at fixed $\sigma = 2$ does not affect the parameter, nor the median value, nor its uncertainty. The improvement of PLS model with $N_{knots} = [10, 12, 14, 16]$ with respect to the reference model PL2P lies in the range $9.2\% - 13.2\%$.

The primary slope of the power law α. The median value of the power-law slope α is consistent across all models within the uncertainties. Its uncertainty grows from 9.4% for $\sigma = 0.5$ up to 39.7% for $\sigma = 5$. The most precise results come from PLS-DD14-$\mathcal{G}_{0.5}$ model in which the spline coefficients are more constrained, providing:

$$\alpha = 3.2 \pm 0.3, \tag{5.5}$$

In PLS models with low values of the prior width, the parametric components more directly capture the global slope of the mass distribution, leading to a better constraint of α.

Overall, the population parameters illustrate a complementary behavior to that of H_0: while increasing the spline flexibility enhances the recovery of the mass distribution substructures and improves the cosmological constraints, it simultaneously weakens the precision with which the parametric component of the population model is obtained. Nevertheless, the stability of the median values across all analyzed models confirms that

the global features of the mass distribution are inferred by the data.

Model	H_0 [km/s/Mpc]	m_{low} [$M_\odot$]	m_{high} [$M_\odot$]	δ_m [$M_\odot$]	α
PL2P	$72.5^{+42.7}_{-27.5}$	$4.7^{+1.0}_{-1.4}$	$90.0^{+37.4}_{-16.5}$	$4.8^{+2.9}_{-2.7}$	$2.9^{+0.7}_{-0.5}$
PLS-DD-14-$\mathcal{G}_{0.5}$	$54.6^{+36.5}_{-30.7}$	$4.9^{+0.7}_{-0.9}$	$90.5^{+34.9}_{-18.6}$	$5.8^{+2.2}_{-2.2}$	$3.2^{+0.3}_{-0.3}$
PLS-DD-14-$\mathcal{G}_1$	$55.4^{+28.0}_{-26.6}$	$4.8^{+0.8}_{-1.0}$	$99.1^{+49.6}_{-22.4}$	$6.0^{+2.3}_{-2.6}$	$3.2^{+0.5}_{-0.5}$
PLS-DD-14-$\mathcal{G}_2$	$57.8^{+21.9}_{-20.6}$	$4.7^{+1.0}_{-1.3}$	$111.5^{+56.4}_{-30.8}$	$5.9^{+2.5}_{-3.1}$	$3.1^{+0.8}_{-0.8}$
PLS-DD-14-$\mathcal{G}_3$	$59.4^{+19.2}_{-16.2}$	$4.7^{+1.1}_{-1.5}$	$118.2^{+54.4}_{-33.9}$	$5.5^{+2.8}_{-3.3}$	$3.2^{+1.0}_{-1.0}$
PLS-DD-14-$\mathcal{G}_5$	$61.7^{+19.3}_{-14.9}$	$4.8^{+1.2}_{-1.9}$	$121.3^{+52.6}_{-36.0}$	$4.8^{+3.1}_{-3.2}$	$3.4^{+1.5}_{-1.2}$
PLS-DD-10-$\mathcal{G}_2$	$102.1^{+51.2}_{-36.3}$	$4.6^{+0.8}_{-1.1}$	$88.6^{+63.8}_{-22.2}$	$5.5^{+2.5}_{-2.5}$	$3.5^{+1.0}_{-1.0}$
PLS-DD-12-$\mathcal{G}_2$	$56.8^{+32.5}_{-23.4}$	$4.7^{+0.9}_{-1.3}$	$93.6^{+32.0}_{-17.7}$	$6.1^{+2.5}_{-3.0}$	$3.3^{+1.0}_{-0.9}$
PLS-DD-16-$\mathcal{G}_2$	$52.0^{+26.2}_{-16.5}$	$4.7^{+1.0}_{-1.3}$	$105.8^{+48.6}_{-23.5}$	$5.8^{+2.6}_{-3.1}$	$3.2^{+0.8}_{-0.8}$

Table 5.3: Median and 68% C.I. of parameters H_0, m_{low}, m_{high}, δ_m, α for each data-driven model.

5.5 Exploring the impact of logarithmic knot spacing

This Section examines models with logarithmically spaced knots, which currently represent the most widely explored configuration in the literature (for instance Edelman B., et al., 2022). The motivation behind the choice of adopting the logarithmic spacing is to place an appropriate number of knots, especially at low masses, characterized by a finer structure in the BBH mass distribution. Model comparison (Section 5.3.3) indicates that, according to the Bayes Factor (see Table 5.2), the only logarithmic model that shows strong evidence in favor of itself with respect to the `PL2P` is the `PLS-LOG20-`$\mathcal{G}_2$ model. Instead, according to the DIC, all the logarithmic models appear to be disfavored models relative to the data-driven models with the lowest DIC values. This is a consequence of the logarithmic spacing: capturing additional substructures in the mass function requires increasing the number of knots, leading to more complex models. Among the models analyzed, the `PLS-LOG20-`$\mathcal{G}_2$ emerges as the preferred model according to BF and ΔDIC.

In this Section, we focus our analysis on highlighting the differences in the reconstruction of the mass function with respect to the `PLS` models, and how this reconstruction affects both the cosmological and population parameters inference with respect to the data-driven results.

5.5.1 Mass distribution analysis

In Figure (5.9), we show the best-fit primary mass posterior distribution for the log-spacing models with the corresponding 68% confidence interval, compared also in this case with the weakly parametric data-driven approach adopted by Abac A. G., et al. (2025). The main difference to keep in mind is that the `B-Spline` model used by Abac A. G., et al. (2025) is constructed assuming a fixed cosmology, with knots determined directly by the data.

For models with a number of knots $N_{knots} = [10, 14, 15]$, the reconstructed mass functions are generally smoother compared to the `PL2P` reference model. In particular,

the peak at $\sim$ 35 $M_\odot$, clearly visible in the PL2P, is much less pronounced in these logarithmic models, while the main peak at $\sim$ 10 $M_\odot$ remains evident. Both peaks, however, appear shifted to lower masses compared to the PL2P model: for this reason, a more quantitative assessment of the peak locations and their dependencies on the model parameters is discussed in detail in Section (5.5.3). Moreover, for these models, in which knot placement is not driven by the data, just increasing the number of knots does not ensure a priori a more accurate reconstruction.

The PLS-LOG20-$\mathcal{G}_2$ model identifies three peaks at approximately 9 $M_\odot$, 18 $M_\odot$ and 30 $M_\odot$. This reconstruction is the closest to the weakly parametric data-driven model of Abac A. G., et al. (2025) and to the data-driven models already discussed in Section (5.4.1), although the third peak still appears shifted to slightly lower masses. In addition, the PLS-LOG20-$\mathcal{G}_2$ model is the only logarithmic one that displays clear mass gaps.

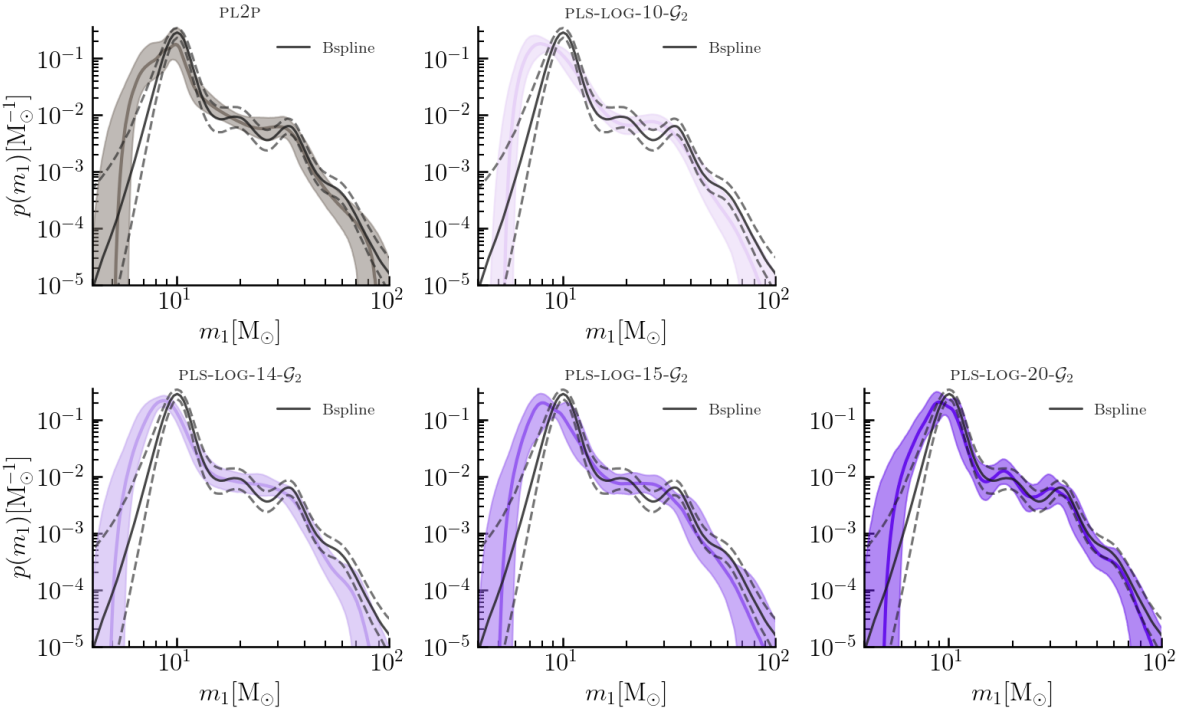

Figure 5.9: Best-fit and 68% confidence interval of the primary mass posterior distributions for the models with logarithmically distributed knots $N_{knots} = [10, 14, 15, 20]$ and for PL2P, compared with the results obtained by Abac A. G., et al. (2025).

In Figure (5.10), it is shown the median of the primary mass distribution with the considered knot positions. It clearly shows how small differences in the knot placement can shift the location of the reconstructed peak. The model that best reproduces the three peaks, together with the gaps between them, is the one whose knots fall near the expected peaks and especially in the regions where the mass function changes slope (similarly to the data-driven method reported in Section (5.2)). In contrast, the other models place their knots more broadly across these features, missing the weakest peak at 18 $M_\odot$ first and then leading to a suppression and smoothing of the other peaks.

A visual and qualitative Posterior Predictive Check analysis has also been carried out for this type of mass function model. All models exhibit observed distributions that

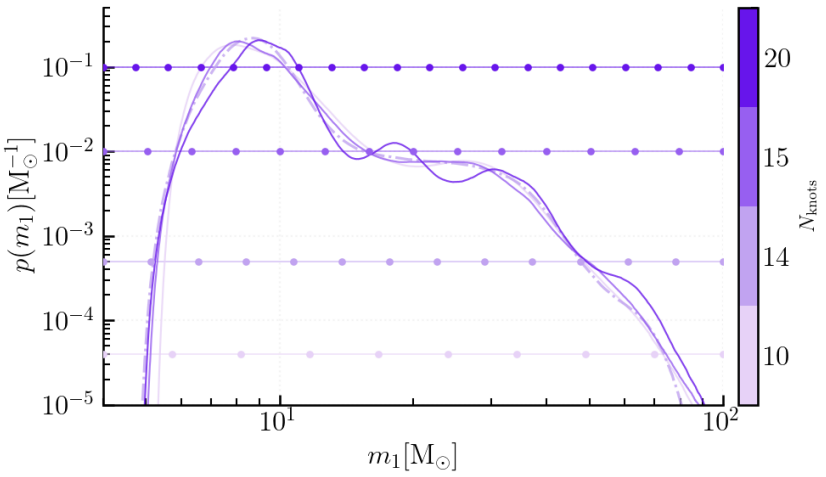

Figure 5.10: Median of the primary mass posterior distribution for the four logarithmically spaced model with $N_{knots} = [10, 14, 15, 20]$ and their corresponding knot position.

remain consistent with the corresponding predictions, and the PPC does not indicate that any model should be excluded from the analysis. Since the PPC does not provide additional information, the full set of PPC plots is reported in Appendix (A.1).

5.5.2 Constraints on cosmological and astrophysical parameters

Data-driven models have shown that capturing substructures in the mass function can improve the reconstruction of population parameters. Here, we explore the logarithmic models with different numbers of knots $N_{knots} = [10, 14, 15, 20]$ and we analyze their effect on the inferred population parameters.
The models mainly differ in how well they reproduce the peaks of the mass distribution. While the `PLS-LOG20-`$\mathcal{G}_2$ model reproduces all three main peaks, it slightly shifts them toward lower masses. Instead, models with fewer knots, such as `PLS-LOG10-`$\mathcal{G}_2$, `PLS-LOG14-`$\mathcal{G}_2$, and `PLS-LOG15-`$\mathcal{G}_2$, smooth part of the mass structure, leading to more stable parameter estimation.
The population parameters, as discussed in Section (5.4.3), are less sensitive to the reconstruction of the mass function compared to the cosmological ones. While they remain independent and do not exhibit degeneracies with the spline coefficients, some of the information captured by the splines may partially weaken the constraints on the parametric power-law parameters. Overall, the median values are mutually consistent within their uncertainties and align well with those obtained from the data-driven models, as reported in Table (5.4). In particular, they follow a trend that is similar to the data-driven population posterior as a function of knots.
The only population parameter that exhibits a strong correlation with H_0 and captures the degeneracy with H_0 independently of the number of knots is m_{high}.

The cosmological parameter H_0 is the most sensitive to the reconstruction of the mass function, especially for the spectral sirens analysis, where no galaxy catalog is available, and it is completely inferred from the shape of the primary mass distribution. In this case, we observe that the knot placement with the logarithmic approach has an almost disruptive impact on the retrieved Hubble constant. Different from the data-driven

models, here H_0 shows some unexpected features in its posterior distribution, as shown in Figure (5.11). In particular, the two logarithmic models with the largest number of knots ($N_{knots} = 15, 20$) present significant wiggles in the posterior of H_0.

We notice, strangely, that the model that best reproduces the mass function (`PLS-LOG20-`$\mathcal{G}_2$, even though the three peaks found are shifted to lower masses) presents a double-peaked posterior distribution of H_0. Instead, the best inference for H_0 comes from the `PLS-LOG14-`$\mathcal{G}_2$ model. Cosmological inference and the mass function reconstruction in the logarithmic case do not depend only on the knot positions, since their locations are independent of the data.

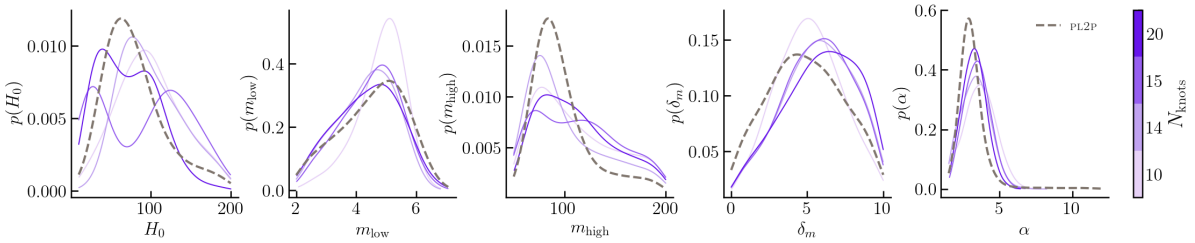

Figure 5.11: Posterior distributions of parameters H_0, m_{low}, m_{high}, δ_m, α as a function of the number of knots $N_{knots} = [10, 14, 15, 20]$ logarithmically distributed.

Model	H_0 [km/s/Mpc]	m_{low} [$M_\odot$]	m_{high} [$M_\odot$]	δ_m [$M_\odot$]	α
PL2P	$72.5^{+42.7}_{-27.5}$	$4.7^{+1.0}_{-1.4}$	$90.0^{+37.4}_{-16.5}$	$4.8^{+2.9}_{-2.7}$	$2.9^{+0.7}_{-0.5}$
PLS-LOG-10-$\mathcal{G}_2$	$95.9^{+45.8}_{-37.2}$	$4.9^{+0.6}_{-0.9}$	$102.6^{+59.1}_{-32.5}$	$5.1^{+2.3}_{-2.2}$	$3.6^{+1.0}_{-1.1}$
PLS-LOG-14-$\mathcal{G}_2$	$93.0^{+48.6}_{-30.8}$	$4.4^{+0.9}_{-1.2}$	$88.7^{+61.2}_{-20.4}$	$5.8^{+2.5}_{-2.6}$	$3.3^{+1.0}_{-1.0}$
PLS-LOG-15-$\mathcal{G}_2$	$103.2^{+47.5}_{-78.1}$	$4.6^{+0.8}_{-1.2}$	$111.6^{+50.8}_{-43.8}$	$5.7^{+2.4}_{-2.6}$	$3.4^{+0.9}_{-0.9}$
PLS-LOG-20-$\mathcal{G}_2$	$68.3^{+41.4}_{-36.4}$	$4.5^{+1.0}_{-1.3}$	$106.8^{+51.1}_{-33.6}$	$6.0^{+2.5}_{-3.0}$	$3.3^{+0.8}_{-0.8}$

Table 5.4: Median and 68% C.I. of parameters for the logarithmic models.

5.5.3 Inspecting the H_0 issue

To explore the issue in the retrieved H_0 with the logarithmically-spaced knots, we inspected in detail the parameters derived in the various fits.

It is in particular interesting to consider the logarithmic models with $N_{knots} = 15, 20$, that show two independent peaks in the H_0 posterior. Inspecting the relation between the retrieved Hubble constant and the reconstructed mass function, we hypothesize that this bimodality could arise from the observed shift in the position of the peaks. As in the case of the data-driven models, an improved reconstructed mass distribution can provide better constraints on H_0, in the same way an incorrect reconstruction can induce a bias. In particular, in the `PLS` models, degeneracies can arise between H_0 and the spline coefficients, which measure the amplitude of the structures in the mass distribution. On the other hand, we remind that the position of the knots is fixed in this approach. This is a crucial difference with respect to the parametric `PL2P` model, where both the position and amplitude of the peaks are free parameters.

In the models with $N_{knots} = 15, 20$, some spline coefficients exhibit strong degeneracies with H_0, producing the bimodal posterior distribution shown in Figure (5.11). Appendix (A.3) shows a part of the corner plot for the models `PLS-LOG15-`$\mathcal{G}_2$, explicitly highlighting H_0, the population parameters, and two relevant spline coefficients. The `PLS-LOG15-`$\mathcal{G}_2$ is in fact the model for which the H_0 bimodality is the most evident and with the strongest degeneration between H_0 and spline coefficients.

To investigate correlations between the population parameters and the cosmological parameters H_0, we divide the posterior samples into two subsamples based on the value of the spline coefficients c_5 and c_6. These coefficients correspond to the first peak (~ 10 $M_\odot$) and to the subsequent gap in the mass distribution and are strongly correlated with H_0, respectively positively and negatively.

We then extracted from the sampled posterior the mass function for two regimes:

- **Low** H_0, defined by $c_5 < 0$ and $c_6 > 0$;
- **High** H_0, defined by $c_5 > 0$ and $c_6 < 0$.

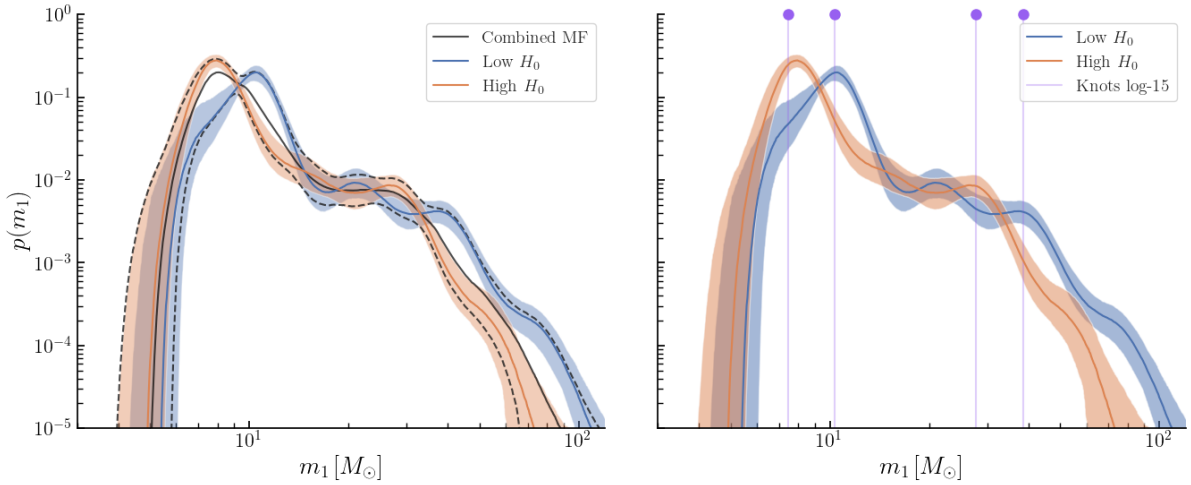

Figure 5.12: The red and blue bands show the mass function reconstructed from the two subsamples selected based on the values of the spline coefficients c_5 and c_6, which identify the low and high H_0 regimes. *Left:* the two subsample mass functions are plotted together with the fully marginalized mass function, shown as a black solid line (median) with 68% credible interval as dashed lines. *Right:* the four vertical lines mark the 4 strategic knots positions around the peaks at 10 $M_\odot$ and 35 $M_\odot$.

In Figure (5.12), the mass function is reported for the two H_0 regimes. A clear difference is observed between them, which can explain the bimodality seen in the H_0 posterior.

The mass function associated with the low H_0 regime is shifted toward higher masses, while the one corresponding to the high H_0 regime is shifted toward lower masses. These differences in the mass distributions lead to two equally probable solutions in the H_0 posterior, producing two distinct peaks. In particular, the low H_0 peak is anchored by the "low H_0" mass function, while the high H_0 peak is anchored by the "high H_0" mass function.

The two mass functions not only show different peaks and gaps, but they also cover different ranges of masses. This behaviour arises because H_0 is proportional to the

redshift through Eq. (1.19), which in turn is inversely proportional to the source-frame mass m_{src}, as shown in Eq. (1.42). As a consequence, changes in the reconstructed mass function directly affect the inferred H_0 values in the spectral sirens analysis.

In fact, because of the intrinsic nature of the PLS model, the spline is forced to pass through the imposed knots. As shown in the right panel of Figure (5.12), the knots near 10 $M_\odot$ and 35 $M_\odot$ are located on opposite sides of the peaks in the combined mass function. Consequently, the mass function is forced to place the peak in correspondence with one knot, either to the left or to the right of the combined mass function peak, depending on the H_0 regime. The model PLS-LOG15-$\mathcal{G}_2$ is not the only one that produces a bimodal posterior for the cosmological parameter; the PLS-LOG20-$\mathcal{G}_2$ model also exhibits the same behavior. The model with 20 knots has more degrees of freedom, and hence it allows the mass function to oscillate more freely. As a result, the posterior does not show a sharp double peak but rather a smoother structure in the H_0 parameter.
Moreover, as shown in the corner plot in Figure (A.3), it is noteworthy that in the case where the bimodality in the H_0 posterior is most pronounced, a corresponding double peak also appears in the posterior of m_{high}, the parameter most strongly correlated with H_0.
The clear separation between the two mass functions highlights the strong correlation between cosmological and population parameters and demonstrates that spectral sirens analysis requires extremely accurate models.

5.6 Correlation between mass function features and H_0

To identify which features of the mass function contribute most to the inference of H_0 and quantify their impact, we compute the Spearman correlation coefficient ρ between H_0 and the parameters describing the mass distribution. This coefficient ranges from -1 to 1 and quantifies the strength and direction of the correlation between variables.

For each mass function model, we evaluate the correlation between H_0 and the parameters that define the main features in the mass function. In the case of the PL2P model, these correspond to the parameters describing the low and high mass cutoffs (m_{low} and m_{high}) and the Gaussian peaks (μ_1 and μ_2). For the PLS models, the correlations are computed with the spline coefficients (c_i) and still the cutoff masses.

Each correlation coefficient, therefore, quantifies how variations in a feature of the mass functions are associated with changes in the inferred value of H_0. To visualize it, these correlations between two parameters are associated with the primary mass and shown along the m_1, allowing us to highlight which mass ranges have the most impact on the cosmological inference.

The results are shown in Figure (5.13), where the colored regions represent the value of the Spearman coefficient in that mass range, while the dashed black line shows the median mass function.

For the PL2P model, we find that both peaks in the mass distribution are anti-correlated with H_0. For the data-driven models, with the exception of PLS-DD10-$\mathcal{G}_2$, a clear pattern emerges: the coefficient preceding each peak is positively correlated with H_0, while the one following it is anti-correlated. This means that H_0 is constrained through the combination of the two coefficients that characterize the rising and falling

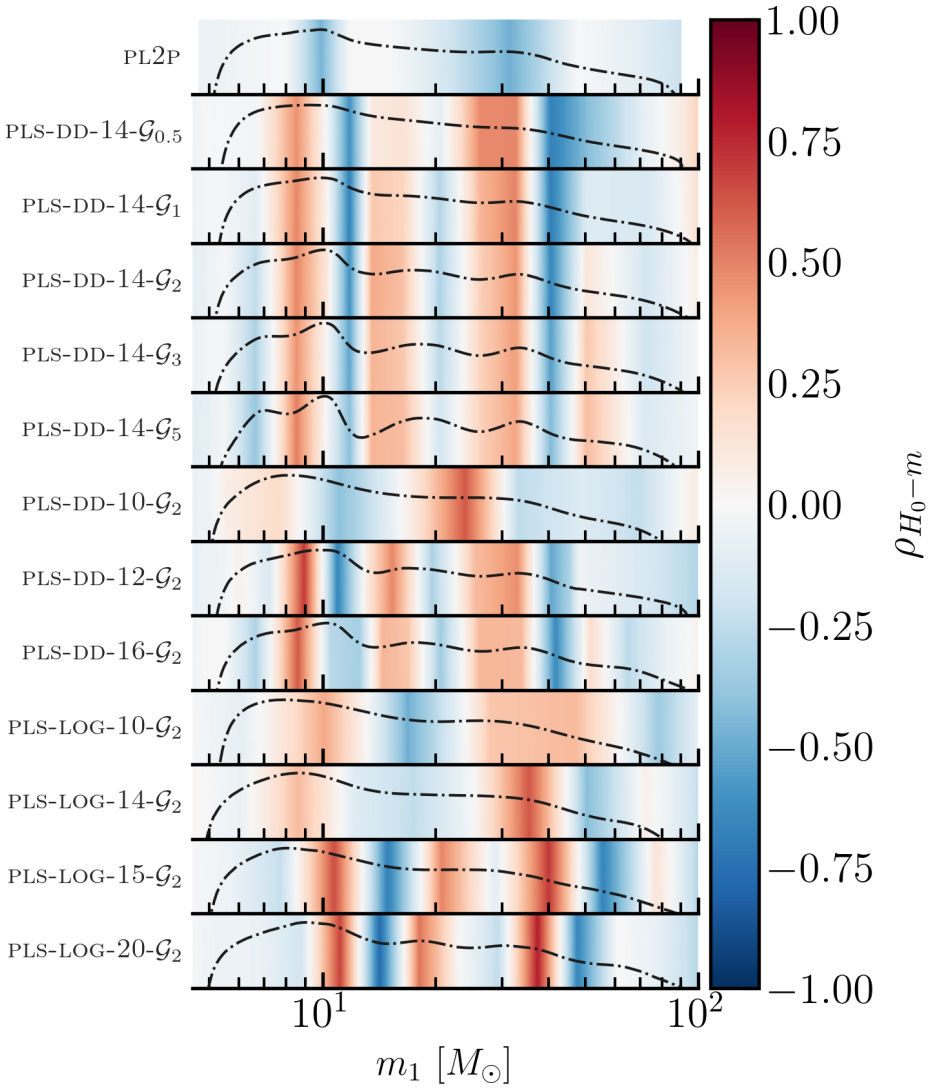

Figure 5.13: Spearman correlation coefficients ρ between H_0 and the parameters describing the mass function for each model. The correlations are interpolated along the primary mass m_1. The dashed black line represents the median of the mass function.

edges of each peak in the mass function. The PLS-DD10-$\mathcal{G}_2$ instead, due to the low number of knots and the uncertain reconstruction of structures in the mass function, does not exhibit a clear correlation or anticorrelation. It is in particular interesting to underline here the emergence of this behaviour around three peaks for the data-driven PLS, evidence of the fact that also the third intermediate peak has a non-negligible contribution to constraining the Hubble constant.

Regarding the logarithmic models, we find that PLS-LOG10-$\mathcal{G}_2$ behaves similarly to the PLS-DD10-$\mathcal{G}_2$. The PLS-LOG14-$\mathcal{G}_2$ model displays correlations associated with the two main features in the mass function, occurring at mass scales consistent with those observed in the data-driven models. This explains why PLS-LOG14-$\mathcal{G}_2$ provides an H_0 inference most similar to the data-driven approach among the logarithmic models. In contrast, PLS-LOG15-$\mathcal{G}_2$ and PLS-LOG20-$\mathcal{G}_2$, which both have two distinct peaks in the

H_0 posterior, show positive correlations just after the peaks and anticorrelation in correspondence with gaps. The correlations involve mass scales that are shifted to lower values relative to those carrying cosmological information, as evident from the plot.
Comparing the data-driven correlations with those from the logarithmic models, it becomes evident that, as the number of knots increases, the data-driven approaches yield more stable results. This greater stability arises from the fact that, even if with more knots, the data-driven reconstructions can place knots consistently around the peak structures (Section 5.2). Instead, the logarithmic models place the knots independently of the data, and in this case, the `PLS` forces the mass function to pass through those points, weirdly placed with respect to the overdensities in the mass distribution. This further stresses that the choice of knot positions is crucial, and that simply increasing their number in the logarithmic case would not necessarily lead to an improved reconstruction.

5.7 Summary

In this Chapter, we explored for the first time the `PLS` model on real data, using the GWTC-4 catalog. The analysis of the GWTC-4 dataset was already carried out by Abac A, et al. (2025) with a spline model, but only with fixed cosmology. Furthermore, a logarithmic spline configuration has already been tested by Edelman B., et al. (2022) and by Edelman B., et al. (2023) on GWTC-2 and GWTC-3 datasets, respectively, still with fixed cosmology.

In this Thesis, the analysis on GWTC-4 is performed in the spectral siren regime using both configurations, data-driven and logarithmic, for the first time, fitting at the same time cosmological and astrophysical parameters. Allowing cosmology to vary makes it possible to study not only population parameters, but also the inference of cosmological ones and how the two sets of parameters are correlated and degenerate.

Among the two configurations, the data-driven approach can improve the cosmological constraints because it recovers the features of the mass function more accurately, thanks to the freedom of choosing knot positions. This analysis demonstrates how crucial the knot placement is. In fact, a wrong placement of knots could produce artificial oscillations in the posterior distribution. This issue is clearly visible in the logarithmic case, where the posteriors of parameters could show oscillatory behavior and a bimodal shape.
Another important point stressed in this Chapter concerns the choice of parameters' priors, in particular that of m_{high}. In this analysis, the upper bound of the maximum cutoff mass prior is set to 200 $M_\odot$, allowing the parameter m_{high} to vary and to perform a completely cosmologically unconstrained study. In contrast, Edelman B., et al. (2022) and Edelman B., et al. (2023) fixed the upper bound $\sigma_{m_{high}} = 100$ $M_\odot$. Combined with the fact that their cosmology was fixed, and in the absence of the degeneracy between H_0 and m_{high}, this also restricted the mass range and the allowed knots positions below 100 $M_\odot$.

Overall, the analysis suggests that the data-driven method should be preferred over the logarithmic one, as it provides better control over where the mass function is recon-

structed. We also observed that increasing the number of knots in the data-driven case leads to a small but significant improvement in the reconstruction of the mass function, whereas in the logarithmic case, it mainly introduces spurious oscillations.

Conclusions

The advent of gravitational waves (GWs) has opened a new observational window on the Universe, impacting both astrophysics and cosmology. The first detection of GWs occurred in 2015, when the LIGO Collaboration detected GW150914 (Abbott B. P., et al., 2016), marking the beginning of the gravitational wave observations era. Two years later, in 2017, the first binary neutron star (BNS) merger event, GW170817, was detected with its electromagnetic counterpart, opening the era of multimessenger astronomy (Abbott B. P., et al., 2017). Since then, a continuous campaign of instrumental upgrades has significantly increased the sensitivity of ground-based interferometers such as LIGO, Virgo, and KAGRA, leading to a rapidly growing catalog of GW events. Today, the latest GW catalog is GWTC-4, (released in November 2025, Abac A. G., et al., 2025), which contains ≈ 200 GW events from compact binary coalescences (CBCs), the majority of which are binary black hole (BBH) mergers. The increasing number of observed events has improved our knowledge of the BBH population and its astrophysical properties, and this progress is expected to continue with future observing runs and the next generation of interferometers. At the same time, GWs have started to be used as cosmological probes. In particular, CBCs act as standard sirens (Schutz B. F., 1986, Holz E. et Hughes S., 2005), providing an independent way to measure the luminosity distance d_L without any external calibration. However, the rapidly increasing volume of data has also highlighted new challenges. The growing number of detections requires a continuous review of the models used to reconstruct the BBH population properties. In particular, the BBH mass function has revealed multiple complex structures not recovered by parametric models.

Despite the success of the ΛCDM model, modern cosmology has highlighted several open questions: the nature of dark matter and dark energy is still under debate, and with increasing precision of modern observations, tensions have emerged in the cosmological parameters measured with different probes. The most significant is the Hubble tension, a disagreement between the Hubble constant inferred from early-Universe data (such as CMB) and the value obtained from late-Universe measurements (using Type Ia supernovae calibrated with Cepheids). This discrepancy between H_0 values may suggest that additional physics is required beyond the ΛCDM model, motivating the development of new observational probes to resolve it (Moresco M., et al., 2022).

GWs are emerging in this scenario as a potentially powerful cosmological probe for the future, especially for resolving the current tension in the measured values of H_0. At present, however, the uncertainties in the available measurements of H_0 are too large to solve the discrepancy. The situation is expected to improve with future GW interferometers, which will enable the detection of $\mathcal{O}(10^5)$ events, reaching redshift up

to ~ 20.

The aim of this Thesis is to develop and implement a semi-parametric methodology for modeling the BBH mass function. This flexible modeling is crucial, especially for the future, to maximize the cosmological information extracted from GWs. In fact, it enables not only a more accurate reconstruction of the mass function itself but also a better extraction of cosmological insights encoded in its structure. Especially for the spectral sirens mode, where the redshift information is inferred from features in the mass function, a semi-parametric data-driven model is crucial for the emergence of new features.

In this Thesis, the semi-parametric model `PLS` is developed and implemented within the Bayesian framework `CHIMERA` to perform a joint hierarchical Bayesian analysis of both astrophysical and cosmological parameters, aiming to study how these parameters can be inferred according to the mass function models adopted and to explore their degeneracies and mutual dependencies. This work is included in and contributed to the publication of Tagliazucchi M., Moresco M., Borghi N., Ciapetti C. (2026).

The main results of the Thesis are summarized below.

Implementation of a semi-parametric mass function. In this Thesis, the first step was the study and implementation of a semi-parametric primary mass function (`PLS`) to model the mass distribution of BBH. The structure of the `PLS` model (taken from Edelman B., et al., 2022), is successfully implemented and validated within the `CHIMERA` pipeline. By combining a power law with a flexible spline component, the model introduces controlled deviations from the standard parametric form, allowing the mass distribution to adapt more closely to the data. The construction of the spline through the de Boor recursive algorithm, together with the flexibility in choosing knot positions, allows for capturing both global trends and localized features of the BBH mass function. We also included the possibility for an automatic or manual placement of the spline's knots, to allow for maximum versatility. This implementation required extending the internal structure of the pipeline, including a modification of the `generate_dict` function to support arrays of coefficients. With this extension, the `PLS` model is now fully integrated and can be used as the already existing population models. Overall, the introduction of a semi-parametric approach provides a more flexible and data-driven tool for mass function reconstruction, and represents a significant step forward for future hierarchical analyses within the `CHIMERA` framework.

Validation and testing. The validation process constitutes an essential preliminary test in assessing the robustness and performance of the `PLS` model, as it allows us to evaluate its reconstruction capabilities in a controlled setting, without observational data and starting from well-defined and known theoretical mass functions. The goal is to verify whether and with which accuracy the `PLS` model is capable of reproducing the distribution of BBH masses most commonly adopted in the literature to describe the observed dataset. We decided to consider as testing data the `PLP` and `PL2P` models, since they are the representative reference models considered for the GWTC-3 and GWTC-4 analyses, respectively. We set up the models adopting the parameters directly derived from the analysis of the GWTC-4 dataset, in order to make the functions as close as

possible to the observational evidence.

- The first phase of the validation explored how four PLS configurations, with an increasing number of knots, specifically $N_{knots} = [5, 10, 15, 20]$ logarithmically spaced between 5.1 $M_\odot$ and 87 $M_\odot$, can reproduce the test models. The results clearly indicate that a higher number of knots generally improves the reconstruction accuracy, particularly when the target mass function exhibits multiple features. In the case of the PLP model, however, increasing the knots from 15 to 20 yields no significant improvement, suggesting that above a certain complexity, the gained accuracy becomes negligible. In contrast, the PL2P model benefits from this increase: moving from 15 to 20 knots enhances the reconstruction accuracy by approximately 7%, enabling the model to reproduce not only the broad peak around 35 $M_\odot$ but also the secondary peak near 10 $M_\odot$.

- An interesting insight emerging from this analysis concerns the placement of knots, which often plays a more decisive role than their total number. We verify that a model with 20 logarithmically spaced knots can achieve the same reconstruction precision as a model with only 10 knots whose positions are strategically chosen to capture the main structures, supporting the importance of the versatility with which we built the code. The selection of knot positions, and consequently of the model configuration, is a crucial aspect of this and future analysis, because the placement of knots, and thus of the spline coefficients, determines which features are highlighted or smoothed, affecting the accuracy of the cosmological information extraction.

Application to mock data The PLS model is then applied and studied on mock galaxy and GW catalogs created and described in Borghi N., et al., 2025. Firstly, a parent galaxy catalog is built from the MICECATv2 and is described by a parametric redshift-evolving stellar mass function. Then, potential BBH host galaxies are identified using a linear weighting scheme on mass. Mock GW events are therefore generated according to population models (PLP for the mass distribution and Madau-Dickinson for the merger rate), and their detections from the LVK detector network are simulated using GWFAST (Iacovelli F., et al., 2022), providing a forecast for O5 observing run. Working with simulated data allows the knowledge of the true parameters and full control over the PLP model used to generate the data, enabling a clear validation of the hierarchical inference. We analyze these catalogs, exploring both the dark siren (considering the use of a galaxy catalog to provide redshift information about the GW hosts) and spectral siren (considering only the mass distribution of BBHs to derive constraints) approaches. These are the main results.

- In the dark sirens case, 5 PLS models with different knot placements are tested: 4 of them have logarithmically spaced knots, respectively $N_{knots} = [5, 10, 15, 20]$, one is built by exploring the chosen knot positions using the same number of knots that gives the best result in the logarithmic case ($N_{knots} = 10$). All the models reconstruct the mass distribution successfully, recovering the fiducial peak at 35 $M_\odot$, but models with ten or more knots also identify a fluctuation at 60 $M_\odot$. The 10-knot models provided the most accurate and stable reconstruction, avoiding artificial fluctuations at low masses. The 60 $M_\odot$ feature, absent in the fiducial PLP

model, persisted when reanalyzing the data with a PLP and a 10 user-defined knots PLS models, suggesting that it is not an artifact of the spline representation but rather depends on the specific mock realization.

- In dark sirens, the analysis of the hyperparameters (H_0, α, m_{high}, γ) shows that H_0 remains consistent across all models with a relative error $\sigma_{H_0} \approx 0.6\%$, as its inference is driven by the inclusion of a spectroscopic galaxy catalog. Instead, the mass hyperparameters exhibit degeneracies with spline coefficients, which tend to absorb the contribution of the parametric model parameters. This highlights how crucial the model choice is when reconstructing the mass function.

- In the spectral sirens approach, the cosmological inference becomes directly linked to the reconstruction of the BBH mass distribution. Only the reference PLP model and the best performing PLS model (10 logarithmically spaced knots) are compared, simulating a realistic scenario for the real data application. The mass function reconstruction is consistent and robust, recovering both the peak at 35 $M_\odot$ and a statistical fluctuation at 60 $M_\odot$, even with slightly lower accuracy with respect to dark sirens cases, due to the absence of galaxy information.

- In the spectral sirens mode, the parameters inferred (H_0, α, m_{high}, γ) are mutually consistent within the uncertainty with no evidence for systematic differences between the two models.

Overall, the results on simulated data demonstrate that the PLS model is able to reliably reconstruct the global shape and key features of the mass function. The consistency of cosmological and astrophysical parameters between PLP and PLS models indicates that the model developed is suitable for application to real GW data. We note, however, that given the available catalogs, our tests focused on a simplified mass distribution modeled with a single peak.

Application to real data. Once implemented and validated our model, the final step consisted of performing the analysis for the first time on a real dataset, and in particular the real GWTC-4 catalog (just released during the development of this Thesis), consisting of 137 BBH events. We decided to perform the analysis in spectral sirens mode, and study whether and how the PLS model, thanks to its increased flexibility compared to parametric models, is able to reconstruct the BBH primary mass distribution more accurately than the reference parametric model PL2P. This approach allows us to reveal new features and enhance the inference of cosmological and astrophysical parameters. Here are the main results.

- A data-driven method for knot selection is devised and implemented. This new method explores the observed mass distribution, and with an algorithm based on Gaussian Mixture Models provide an optimal selection of spline knots across the most relevant features derived in the BBH mass function. This procedure is performed by varying the cosmological parameters (H_0, in particular) in order not to be cosmology-dependent. This approach is verified to be the most effective and economic (in terms on number of knots found) for the knots placement, as it is driven by the posterior of GW events for different values of H_0.

- Two classes of `PLS` models are explored in the analysis: one constructed using the data-driven method, varying both the number of knots and the priors on the spline coefficients, and one with logarithmically spaced knots, the most popular choice considered for semi-parametric modeling in the literature. In both cases, the goal is to identify the best compromise between model complexity and descriptive accuracy according to model selection diagnostics; here, we considered in particular the Bayes Factor (BF) and the Deviance Information Criterion (DIC). We find that the best performing models are the ones based on the data-driven approach, and in particular with $N_{knots} = 14$ and $\sigma = 2, 5$.

- The analysis of the BBH mass distributions derived with the data-driven models shows that the prior of the spline coefficients has a stronger impact than the number of knots. As the prior width σ changes, the features of the distribution are modeled with varying levels of detail, although the same underlying structures are consistently recovered. Compared to the `PL2P` model (the standard model used in cosmological analyses in LVK), the analysis identifies not only the known peaks at 10 $M_\odot$ and 35 $M_\odot$, but also a third peak at 18 $M_\odot$ and a bump around 60 $M_\odot$. This is interesting, since it confirm another observational evidence found by population studies (Abac A. G., et al., 2025).

- For the data-driven models, the analysis on the parameter constraints includes H_0, m_{low}, m_{high}, δ_m, and α. For the cosmological parameter H_0, a clear trend emerges: as σ increases, the relative uncertainties decrease. This behavior indicates that a larger σ enables a more accurate reconstruction of the mass function, allowing additional structures in the distribution to be captured. As a consequence, we find that the constraint on H_0 improves when the mass function is reconstructed with higher accuracy. This highlights the impact of the precision of the mass function modeling and the strength of the cosmological inference. In contrast, the mass parameters exhibit the opposite trend: their uncertainties increase with σ. This is likely because, as σ grows, the spline coefficients gain more freedom to capture and absorb the information that would otherwise constrain the parametric mass parameters.

- The logarithmic models are less predictable than the data-driven ones because their knots are spaced deterministically (depending on m_{low} and m_{high}) rather than informed by the data. As a consequence, it is not surprising that the model capable of reproducing all three peaks is the one with 20 knots, simply because it allows fluctuation at the values where they are observed.

- For logarithmic models, the mass parameters do not exhibit any clear trend and remain mutually consistent. However, the inaccurate reconstruction of the mass function introduces a bias in the inferred value of H_0, which strongly depends on the knots placement and on the resulting location of the identified mass structures. This effect is particularly evident per the log-15 and log-20 models, for which the posterior of H_0 shows a double peak. We studied this effect in detail, and we identified its origin in the degeneracy between H_0 and the spline coefficients in correspondence with the peaks.

Future perspectives

The results presented in this Thesis demonstrate that the constraints on both astrophysical and cosmological parameters can be significantly enhanced through a more versatile modeling of the mass distribution, particularly when employing semi-parametric approaches and, even more, when optimally selecting the knot positions.
A future extension of this work would be to explore additional methods for studying the dependence and influence of knot positions, but also to investigate methods that do not require the knot positions a priori. Moreover, the data-driven method could be further developed, potentially by constructing an approach that does not rely on an external parameter such as the *prominence*.
Another promising direction is the implementation of a semi-parametric model based still on `BSpline` to investigate the mass distribution not only of BBH systems but also of other CBCs. In fact, although BBH systems remain the most frequently detected by the LVK detector network, the release of the GWTC-4 dataset already shows an increasing number of systems containing one or two neutron stars (NS). This suggests that in future observing runs, and even more so with the third generation of interferometers, we can expect to detect a growing number of systems involving NSs. These systems will be used not only as dark and spectral sirens, but also as bright sirens when their electromagnetic counterpart is identified.
Within the expansion of the `BSpline`, a future perspective to extend this work is to study the joint mass distribution $p(m_1, m_2)$ of primary and secondary masses instead of the primary mass distribution $p(m_1)$ alone. In fact, the joint mass distribution is a two-dimensional distribution that contains much more information: it allows us to obtain $p(m_1)$ and $p(m_2)$ by marginalization, and it also describes the correlations between primary and secondary mass distribution parameters.
This analysis has been further simplified by marginalizing the GWTC-4 data over the spin parameters, effectively neglecting their contribution. Therefore, a natural next step is to extend the analysis to explicitly include spins and assess how they impact our results.
Lastly, the spectral sirens analysis adopted in this work relies on the assumption that the source-frame mass distribution can be used to break the degeneracy between redshift and detected masses. However, it has been shown (Rinaldi S., et al., 2024) that the mass function itself may evolve with redshift. It would therefore be interesting to investigate a redshift-dependent evolving mass spectrum and to understand how to adapt the semi-parametric mass model to the case of a mass function that evolves in redshift.

With the increasing number of detections, this Thesis has demonstrated that semi-parametric methods can effectively complement traditional parametric models for deriving cosmological constraints. These approaches should be further developed in parallel with parametric models, especially in light of the large number of detections expected from third-generation detectors such as the Einstein Telescope (ET; (Punturo M., et al., 2010)) and Cosmic Explorer (CE; (Reitze D., et al., 2019)).

Appendix A

Appendix

A.1 Posterior Predictive Check for logarithmic models

In Figure (A.1), it is shown the Posterior Predictive Check (PPC) for the logarithmic models. The number of predicted events is fixed, as in the case of data-driven models, to the number of observed ones, $N_{events} = 137$, and 200 posterior draws are used. The PPC panels compare the observed CDF of the primary mass distribution with the CDF predicted by each model. Across the different models, the overall behavior is almost similar. However, the model exhibiting the poorest PPC performance is the `PLS-LOG15-`$\mathcal{G}_2$ case. It shows the largest visible discrepancy between the observed and the predicted CDFs, particularly in the intermediate mass range around $\approx 50\ M_\odot$. This is also consistent with the ΔDIC values reported in Table (5.2), which show that the `PLS-LOG15-`$\mathcal{G}_2$ model is the worst-performing logarithmic spline configuration in terms of balance between the goodness of fit and model complexity: despite introducing 15 spline coefficients, the chosen knot positions do not support the recovery of the underlying structure.

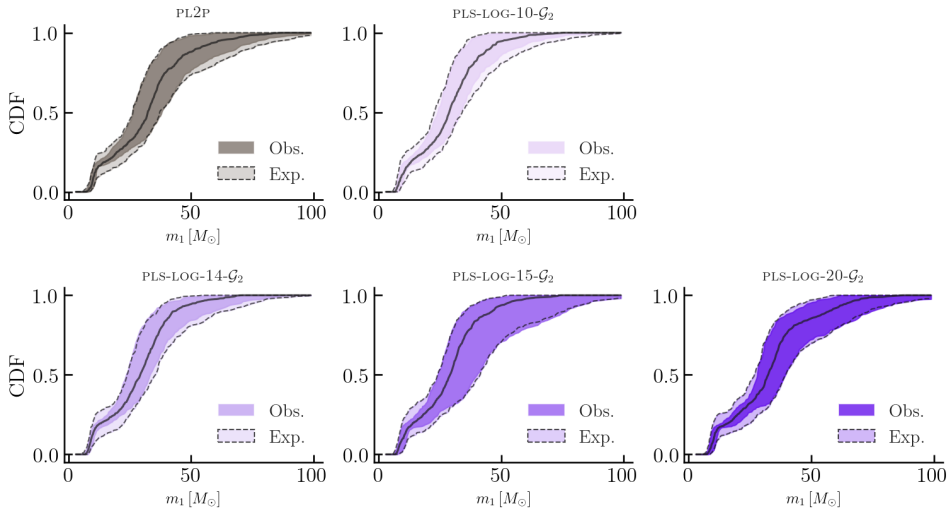

Figure A.1: Posterior predictive checks for all log-spacing models and for `PL2P` evaluated on GWTC-4 dataset. The darker region shows the 90% credible interval of the observed cumulative distribution of the BBH population, while the lighter region corresponds to the 90% credible interval of the predicted distribution. The solid line indicates the median of the predictive cumulative distribution.

A.2 Full corner plot for the PLS-DD14-$\mathcal{G}_2$ model

The full corner plot for the PLS-DD14-$\mathcal{G}_2$ model is reported. It shows the marginalized 1D and 2D posterior distributions for all the model parameters. This case is presented as an illustrative example, for the model that performs best according to the Bayesian diagnostics (see Section 5.3.3).

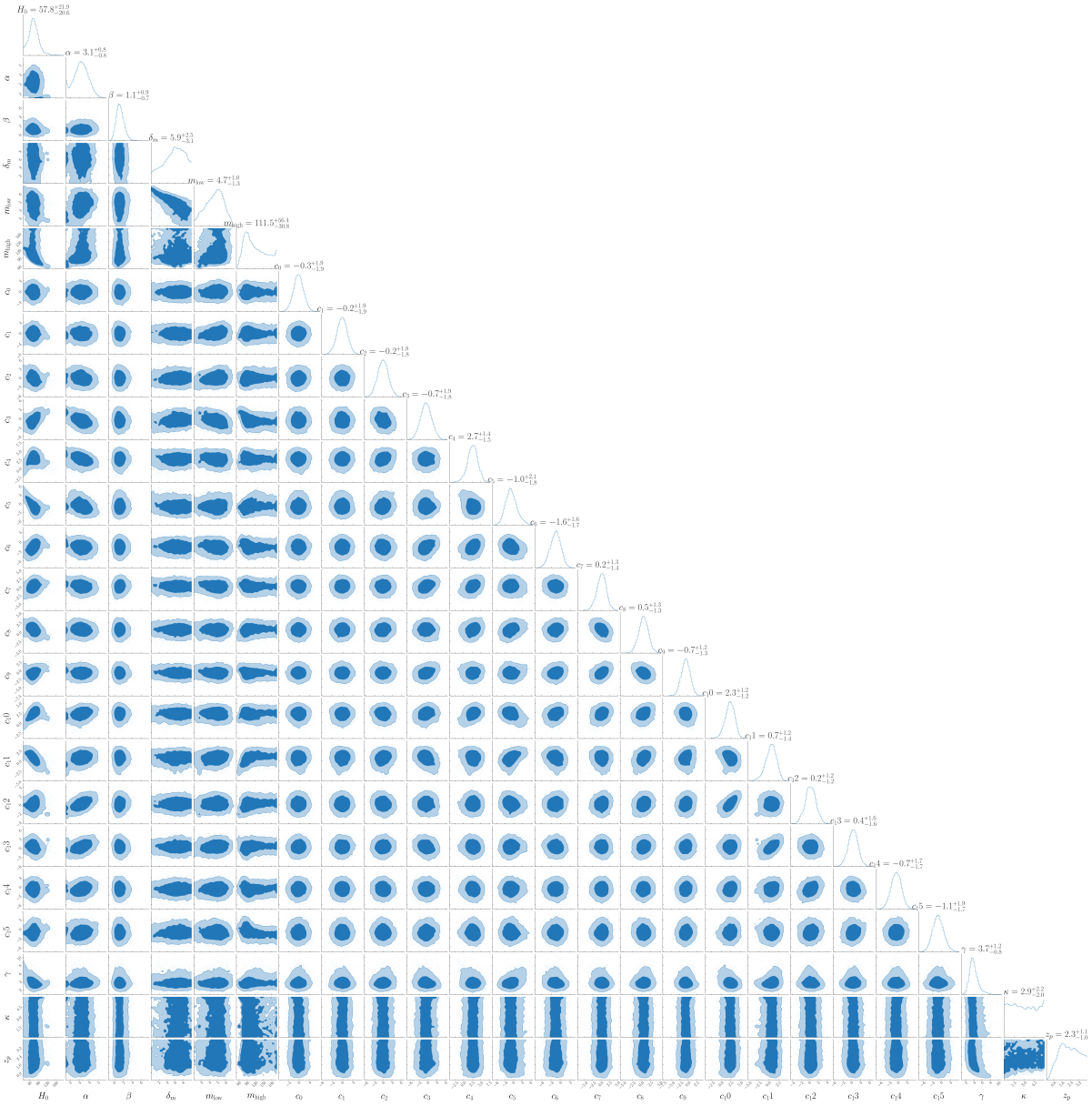

Figure A.2: Full corner plot showing the marginalized 1D and 2D posterior distributions for all parameters of the PLS-DD14-$\mathcal{G}_2$ model.

A.3 Corner plot for the PLS-LOG15-$\mathcal{G}_2$ model

In this Section we report the corner plot of the PLS-15LOG-$\mathcal{G}_2$ model, showing the marginalized 1D and 2D posterior distribution for H_0, the mass parameters (α, δ_m, m_{low}, m_{high}) analyzed in Section (5.5.2), and two spline coefficients used to investigate the double-peaked structure of the H_0 posterior distribution (see Section 5.5.3).

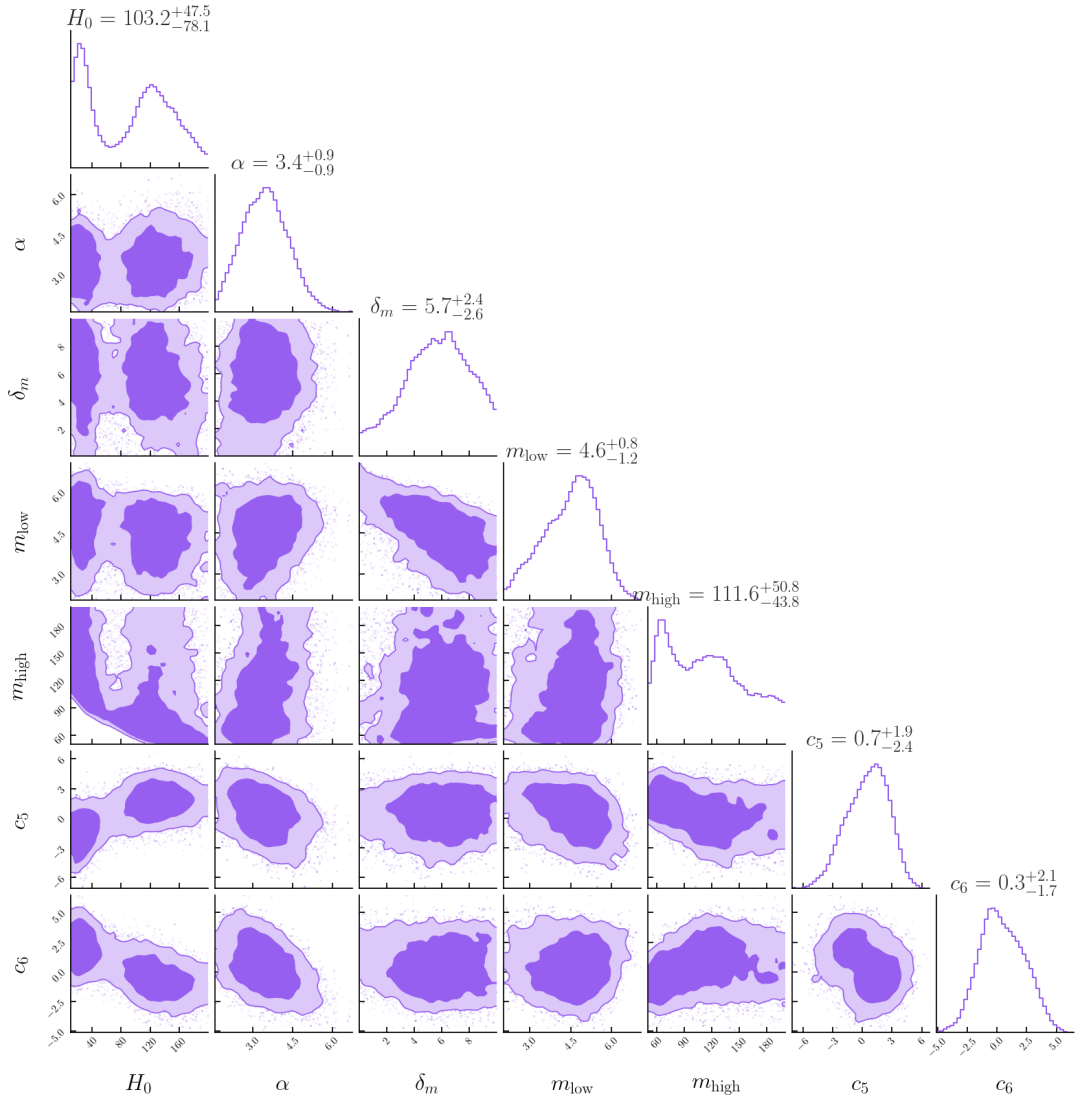

Figure A.3: Corner plot of the PLS-15LOG-$\mathcal{G}_2$ model for H_0, the four analyzed population parameters (α, δ_m, m_{low} and m_{high}) and a representative selection of spline coefficients. Specifically, c_5 and c_6 are included to show their respective positive and negative correlations with H_0.

Bibliography

[1] Abac A. G., et al., 2024, *Observation of Gravitational Waves from the Coalescence of a* 2.5–4.5 $M_\odot$ *Compact Object and a Neutron Star*, The Astrophysical Journal Letters, 970:L34 (https://doi.org/10.3847/2041-8213/ad5beb)

[2] Abac A. G., et al., 2025, *GWTC-4.0: Population Properties of Merging Compact Binaries*, arXiv, page arXiv:2508.18083 (https://arxiv.org/abs/2508.18083)

[3] Abac A. G., 2025, *GWTC-4.0: Constraints on the Cosmic Expansion Rate and Modified Gravitational-wave Propagation*, arXiv, page arXiv:2509.04348 (https://arxiv.org/abs/2509.04348)

[4] Abac A. G., 2025, *GW231123: A Binary Black Hole Merger with Total Mass* $190-265$ $M_\odot$, The Astrophysical Journal Letters, 993:L25 (37pp) (https://doi.org/10.3847/2041-8213/ae0c9c)

[5] Abac A. G., et al., 2025, *GWTC-4.0: Methods for Identifying and Characterizing Gravitational-wave Transients*, arXiv, page arXiv:2508.18081 (https://arxiv.org/abs/2508.18081)

[6] Abbott B. P., et al., 2016, *Binary Black Hole Mergers in the First Advanced LIGO Observing Run*, PHYSICAL REVIEW X (https://doi.org/10.1103/PhysRevX.6.041015)

[7] Abbott B. P., et al., 2016, *Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo*, Living Reviews in Relativity (http://dx.doi.org/10.1007/lrr-2016-1)

[8] Abbott B. P., et al., 2016, *Observation of Gravitational Waves from a Binary Black Hole Merger*, PHYSICAL REVIEW LETTERS (https://doi.org/10.48550/arXiv.1602.03837)

[9] Abbott B. P., et al., 2017, *A GRAVITATIONAL-WAVE STANDARD SIREN MEASUREMENT OF THE HUBBLE CONSTANT*, arXiv, page arXiv:1710.05835 (https://arxiv.org/abs/1710.05835)

[10] Abbott B. P., et al., 2017, *GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*, PHYSICAL REVIEW LETTERS (https://doi.org/10.1103/PhysRevLett.119.161101)

[11] Abbott R., et al., 2019, *Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo*, The Astrophysical Journal Letters (https://doi.org/10.3847/2041-8213/ab3800)

[12] Abbott R., et al., 2021, *Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog*, arXiv, page arXiv:2010.14533 (https://arxiv.org/abs/2010.14533)

[13] Abbott R., et al., 2023, *Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3*, Physical Review X, 13:011048 (https://doi.org/10.1103/PhysRevX.13.011048)

[14] Abbott R., et al., 2023, *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run*, PHYSICAL REVIEW X, 13:041039 (https://doi.org/10.1103/PhysRevX.13.041039)

[15] Abdalla E., et al., 2022, *Cosmology Intertwined: A Review of the Particle Physics, Astrophysics, and Cosmology Associated with the Cosmological Tensions and Anomalies*, arXiv, page arXiv:2203.06142 (https://arxiv.org/abs/2203.06142)

[16] Adams M. R., et Cornish N. J., 2012, *Astrophysical Model Selection in Gravitational Wave Astronomy*, arXiv, page arXiv:1209.6286 (https://arxiv.org/abs/1209.6286)

[17] Aghanim N., et al., 2020, *Planck 2018 results VI. Cosmological parameters*, Astronomy & Astrophysics, 641, A6, (http://dx.doi.org/10.1051/0004-6361/201833910)

[18] Amaro-Seoane P., et al., 2017, *Laser Interferometer Space Antenna*, arXiv, page arXiv:1702.00786 (https://arxiv.org/abs/1702.00786)

[19] Artale M. C., et al., 2020, *An astrophysically motivated ranking criterion for low-latency electromagnetic follow-up of gravitational wave events*,Monthly Notices of the Royal Astronomical Society, Volume 495, Pages 1841–1852 (https://doi.org/10.1093/mnras/staa1252)

[20] Belczynski K., et al., 2016, *The effect of pair-instability mass loss on black-hole mergers*, Astronomy & Astrophysics, 594, A97 (https://doi.org/10.1051/0004-6361/201628980)

[21] Borghi N., et al., 2024, *Cosmology and Astrophysics with Standard Sirens and Galaxy Catalogs in View of Future Gravitational Wave Observations*, The Astrophysical Journal, 964:191 (https://doi.org/10.3847/1538-4357/ad20eb)

[22] Borghi N., et al., 2025, *Echoes from the Dark: Galaxy Catalog Incompleteness in Standard Siren Cosmology*, arXiv, page arXiv:2509.18243,(https://arxiv.org/abs/2509.18243)

[23] Branchesi M., et al., 2023, *Science with the Einstein Telescope: a comparison of different designs*, arXiv, page arXiv:2303.15923,(https://arxiv.org/abs/2303.15923)

[24] Carretero J., et al., 2015, *An algorithm to build mock galaxy catalogues using MICE simulations*, Monthly Notices of the Royal Astronomical Society, Volume 447, Pages 646–670 (https://doi.org/10.1093/mnras/stu2402)

[25] Callister T. A. and Farr W. M., 2024, *A Parameter-Free Tour of the Binary Black Hole Population*, arXiv, page arXiv:2302.07289 (https://arxiv.org/abs/2302.07289)

[26] Crocce M., et al., 2015, *The MICE Grand Challenge lightcone simulation – II. Halo and galaxy catalogues*,Monthly Notices of the Royal Astronomical Society, Volume 453, Pages 1513–1530 (https://doi.org/10.1093/mnras/stv1708)

[27] Data distribution of GWTC-4.0: Constraints on the Cosmic Expansion Rate and Modified Gravitational-wave Propagation (https://zenodo.org/records/16919645)

[28] De Boor C., 1978, *A Practical Guide to Spline*, Applied Mathematical Sciences, New York: Springer, 1978, Volume 27 (https://doi.org/10.2307/2006241)

[29] Del Pozzo W., et al., 2017, *Cosmological inference using only gravitational wave observations of binary neutron stars*, PHYSICAL REVIEW D 95, 043502 (http://dx.doi.org/10.1103/PhysRevD.95.043502)

[30] Edelman B., et al., 2022, *Ain't No Mountain High Enough: Semiparametric Modeling of LIGO–Virgo's Binary Black Hole Mass Distribution*, The Astrophysical Journal, 924:101 (https://doi.org/10.3847/1538-4357/ac3667)

[31] Edelman B., et al., 2023, *Cover Your Basis: Comprehensive Data-Driven Characterization of the Binary Black Hole Population*, arXiv, page arXiv:2210.12834 (https://arxiv.org/abs/2210.12834)

[32] Essick R., et al., 2023, *Semianalytic Sensitivity Estimates for Catalogs of Gravitational-Wave Transients*, arXiv, page arXiv:2210.12834 (https://dx.doi.org/10.1103/PhysRevD.108.043011)

[33] Ezquiaga J. et Holz D., 2022, *Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries*, arXiv, page arXiv:2202.08240 (https://arxiv.org/abs/2202.08240)

[34] FarrOutLab, 2024, *GWInferno: Deep learning based gravitational-wave inference*, GitHub Repository (https://github.com/FarrOutLab/GWInferno)

[35] Farah M., et al., 2024, *No need to know: astrophysics-free gravitational-wave cosmology*, arXiv, page arXiv:2404.02210 (https://arxiv.org/abs/2404.02210)

[36] Finke A., et al., 2021, *Cosmology with LIGO/Virgo dark sirens: Hubble parameter and modified gravitational wave propagation*, arXiv, page arXiv:2101.12660 (https://arxiv.org/abs/2101.12660)

[37] Foreman-Mackey D., et al., 2013, *emcee: The MCMC Hammer*, PASP 125 306 (https://doi.org/10.1086/670067)

[38] Fosalba P., et al., 2015, *The MICE grand challenge lightcone simulation – I. Dark matter clustering*, Monthly Notices of the Royal Astronomical Society, Volume 448, Pages 2987–3000, (https://doi.org/10.1093/mnras/stv138)

[39] Fryer C. L., et al., 2012, *COMPACT REMNANT MASS FUNCTION: DEPENDENCE ON THE EXPLOSION MECHANISM AND METALLICITY*, The Astrophysical Journal, 749:91 (https://doi.org/10.1088/0004-637X/749/1/91)

[40] Gray R., et al., 2023, *Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues*, arXiv, page arXiv:2308.02281 (https://arxiv.org/abs/2308.02281)

[41] Hernandez I. M. et Ray A., 2024, *Beyond Gaps and Bumps: Spectral Siren Cosmology with Non-Parametric Population Models*, arXiv, page arXiv:2404.02522 (https://arxiv.org/abs/2404.02522)

[42] Hernandez I. M., et al., 2025, *Spectral siren cosmology from gravitational-wave observations in GWTC-4.0*, arXiv, page arXiv:2509.03607 (https://arxiv.org/abs/2509.03607)

[43] Hobson, M. P., et al., 2006, *General Relativity: An Introduction for Physicists.*, Cambridge University Press.

[44] Hoffmann K. et al., 2015, *Measuring the growth of matter fluctuations with third-order galaxy correlations*, Monthly Notices of the Royal Astronomical Society, Volume 447, Pages 1724–1745 (https://doi.org/10.1093/mnras/stu2492)

[45] Holz E. et Hughes S., 2005, *USING GRAVITATIONAL-WAVE STANDARD SIRENS*, The Astrophysical Journal (https://doi.org/10.1086/431341)

[46] Hussain A., et al., 2025, *Living on the edge: Testing for compact population features at the edges of parameter space*, arXiv, page arXiv:2510.20010 (https://arxiv.org/abs/2510.20010)

[47] Iacovelli F., et al., 2022, *Forecasting the Detection Capabilities of Third-generation Gravitational-wave Detectors Using GWFAST*, The Astrophysical Journal, Volume 941, Number 2 (https://doi.org/10.3847/1538-4357/ac9cd4)

[48] Kalaghatgi C, et al., 2020, *Parameter estimation with a spinning multimode waveform model*, Physical Review D (https://doi.org/10.1103/PhysRevD.101.103004)

[49] Karamanis M, et al., 2022, *Accelerating astronomical and cosmological inference with preconditioned Monte Carlo*, Monthly Notices of the Royal Astronomical Society, Volume 516, Issue 2, Pages 1644–1653, (https://doi.org/10.1093/mnras/stac2272)

[50] Kass R. E., Raftery A. E., 1995, *Bayes Factors*, Journal of the American Statistical Association, 90, 773–795 (https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572)

[51] Krishnan C., et al., 2021, *Running Hubble Tension and a H_0 Diagnostic*, arXiv, page arXiv:2011.02858 (https://arxiv.org/abs/2011.02858)

[52] Li Y., et al., 2024, *Multi-spectral Sirens: Gravitational-wave Cosmology with (Multi-) subpopulations of Binary Black Holes*, arXiv, page arXiv:2406.11607 (https://arxiv.org/abs/2406.11607)

[53] London L., et al., 2018, *First Higher-Multipole Model of Gravitational Waves from Spinning and Coalescing Black-Hole Binaries*, Physical Review Letters, 120 (https://doi.org/10.1103/PhysRevLett.120.161102)

[54] Madau P. and Dickinson M., 2014, *Cosmic Star-Formation History*, arXiv, page arXiv:1403.0007 (https://arxiv.org/abs/1403.0007)

[55] Maggiore M., 2008 *Gravitational waves - Volume 1, Theory and experiments.*, Oxford University Press. isbn: 0-19-857074-0.

[56] Mancarella M., et al., 2022 *Cosmology and modified gravitational wave propagation from binary black hole population models*, arXiv, page arXiv:2112.05728 (https://arxiv.org/abs/2112.05728)

[57] Mandel I., et al., 2018, *Extracting distribution parameters from multiple uncertain observations with selection biases*, arXiv, page arXiv:1809.02063 (https://arxiv.org/abs/1809.02063)

[58] Mastrogiovanni S., et al., 2021, *Cosmology in the dark: On the importance of source population models for gravitational-wave cosmology*, arXiv, page arXiv:2103.14663 (https://arxiv.org/abs/2103.14663)

[59] Mastrogiovanni S., et al., 2024, *ICAROGW: A python package for inference of astrophysical population properties of noisy, heterogeneous, and incomplete observations*, Astronomy & Astrophysics, 682, A167 (https://doi.org/10.1051/0004-6361/202347007)

[60] Messenger C. and Read J., 2012, *Measuring a cosmological distance–redshift relationship using only gravitational wave observations of binary neutron star coalescences*, Phys. Rev. Lett. 108, 091101 (http://dx.doi.org/10.1103/PhysRevLett.108.091101)

[61] Moresco M., et al., 2022, *Unveiling the Universe with emerging cosmological probes*, arXiv, page arXiv:2201.07241 (https://arxiv.org/abs/2201.07241)

[62] Payne E. et Thrane E., 2023, *Model exploration in gravitational-wave astronomy with the maximum population likelihood*, arXiv, page arXiv:2210.11641 (https://arxiv.org/abs/2210.11641)

[63] Pierra G. et al., 2025, *Non-Parametric Reconstruction of the Hubble Parameter from the Fourth Gravitational Wave Transient Catalog and DESI Baryonic Acoustic Oscillations*, arXiv, page arXiv:2511.11795 (https://arxiv.org/abs/2511.11795)

[64] Pozzetti L. et al., 2010, *zCOSMOS – 10k-bright spectroscopic sample The bimodality in the galaxy stellar mass function: exploring its evolution with redshift*, Astronomy & Astrophysics (https://doi.org/10.1051/0004-6361/200913020)

[65] Punturo M. et al., 2010, *The Einstein Telescope: a third-generation gravitational wave observatory*, Classical and Quantum Gravity (https://doi.org/10.1088/0264-9381/27/19/194002)

[66] Talbot C. and Thrane E., 2018, *Measuring the Binary Black Hole Mass Spectrum with an Astrophysically Motivated Parameterization*, The Astrophysical Journal, 856:173 (https://doi.org/10.3847/1538-4357/aab34c)

[67] THE LIGO SCIENTIFIC COLLABORATION AND THE VIRGO COLLABORATION, et al., 2017, *A GRAVITATIONAL-WAVE STANDARD SIREN MEASUREMENT OF THE HUBBLE CONSTANT*, arXiv, page arXiv:1710.05835 (https://arxiv.org/abs/1710.05835)

[68] Ray A., et al., 2023, *Nonparametric Inference of the Population of Compact Binaries from Gravitational wave Observations Using Binned Gaussian Processes*, The Astrophysical Journal, 957:37 (https://doi.org/10.3847/1538-4357/acf452)

[69] Reitze D., et al., 2019, *Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO*, arXiv, page arXiv:1907.04833 (https://arxiv.org/abs/1907.04833)

[70] Rezaei M. and Malekjani M., 2021, *Comparison between different methods of model selection in cosmology.*, Eur. Phys. J. Plus 136, 219 (https://doi.org/10.1140/epjp/s13360-021-01200-w)

[71] Riess A. G., et al., 2022, *A Comprehensive Measurement of the Local Value of the Hubble Constant with* 1 $km\ s^{-1}\ Mpc^{-1}$ *Uncertainty from the Hubble Space Telescope and the SH0ES Team*, The Astrophysical Journal Letters, 934:L7 (https://doi.org/10.3847/2041-8213/ac5c5b)

[72] Rinaldi S. and Del Pozzo W., 2021, *(H)DPGMM: A Hierarchy of Dirichlet Process Gaussian Mixture Models for the inference of the black hole mass function*, arXiv, page arXiv:2109.05960 (https://arxiv.org/abs/2109.05960)

[73] Rinaldi S., et al., 2024 *Evidence for the evolution of black hole mass function with redshift*, arXiv, page arXiv:2310.03074 (https://arxiv.org/abs/2310.03074)

[74] Sadiq J., 2022, *Flexible and Fast Estimation of Binary Merger Population Distributions with Adaptive KDE*, Phys. Rev. D 105, 123014 (https://doi.org/10.1103/PhysRevD.105.123014)

[75] Sadiq J., 2024, *Binary vision: The merging black hole binary mass distribution via iterative density estimation*, arXiv, page arXiv:2307.12092 (https://arxiv.org/abs/2307.12092)

[76] Santoliquido F., et al., 2022, *Modelling the host galaxies of binary compact object mergers with observational scaling relations*, Monthly Notices of the Royal Astronomical Society, Volume 516, Issue 3 (https://doi.org/10.1093/mnras/stac2384)

[77] Schutz B. F., 1986, *Determining the Hubble constant from gravitational wave observations*, Nature 323 , 310–311 (https://doi.org/10.1038/323310a0)

[78] Schechter P., 1976, *An analytic expression for the luminosity function for galaxies*,The Astrophysical Journal, 203:297-306,(https://doi.org/10.1086/154079)

[79] Tagliazucchi M., et al., 2025, *Accelerating the Standard Siren Method: Improved Constraints on Modified Gravitational Wave Propagation with Future Data*, arXiv, page arXiv:2504.02034 (https://arxiv.org/abs/2504.02034)

[80] Tagliazucchi M., et al., 2026, *Mind the peak: improving cosmological constraints from GWTC-4.0 spectral sirens using semiparametric mass models*, arXiv, page arXiv:2601.03347 (https://arxiv.org/abs/2601.03347)

[81] Tenorio R., et al., 2025, *Where did heavy binaries go? Gravitational-wave populations using Delaunay triangulation with optimized complexity*, arXiv, page arXiv:2509.19466 (https://arxiv.org/abs/2509.19466)

[82] Thomas C. K. Ng., et al., 2025, *Inferring cosmology from gravitational waves using non-parametric detector-frame mass distribution*, arXiv, page arXiv:2410.23541 (https://arxiv.org/abs/2410.23541)

[83] Tiwari V. and Fairhurst S., 2021, *The Emergence of Structure in the Binary Black Hole Mass Distribution*, arXiv, page arXiv:2011.04502 (https://arxiv.org/abs/2011.04502)

[84] Tiwari V., 2021, *VAMANA: Modeling Binary Black Hole Population with Minimal Assumptions*, arXiv, page arXiv:2006.15047 (https://arxiv.org/abs/2006.15047)

[85] Toubiana A., et al., 2023, *Is there an excess of black holes around* 20 $M_\odot$*? Optimising the complexity of population models with the use of reversible jump MCMC.*,arXiv, page arXiv:2305.08909 (https://arxiv.org/abs/2305.08909)

[86] Weaver J. R., et al., 2023, *COSMOS2020: The galaxy stellar mass function The assembly and star formation cessation of galaxies at* $0.2 < z \leq 7.5$, Astronomy & Astrophysics (https://doi.org/10.1051/0004-6361/202245581)

[87] Woosley S. E. and Heger A., 2021, *The Pair-instability Mass Gap for Black Holes*, The Astrophysical Journal Letters, 912:L31, (https://doi.org/10.3847/2041-8213/abf2c4)

[88] Zevin M., et al., 2020, *Exploring the Lower Mass Gap and Unequal Mass Regime in Compact Binary Evolution*, arXiv, page arXiv:2006.14573(https://arxiv.org/abs/2006.14573)